

RESEARCH

Open Access



# Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer

Yulong Bao<sup>1,3†</sup>, Li Wang<sup>1†</sup>, Lin Shi<sup>2</sup>, Fen Yun<sup>2</sup>, Xia Liu<sup>2</sup>, Yongxia Chen<sup>3</sup>, Chen Chen<sup>2</sup>, Yanni Ren<sup>2</sup> and Yongfeng Jia<sup>1,3\*</sup> 

\* Correspondence: [jjayf0471@163.com](mailto:jjayf0471@163.com)

<sup>†</sup>Yulong Bao and Li Wang contributed equally to this work.  
<sup>1</sup>College of Basic Medicine, Inner Mongolia Medical University, Hohhot, Inner Mongolia, China  
<sup>3</sup>Tumor Molecular Diagnostic Laboratory, The Inner Mongolia Cancer Hospital, Hohhot, Inner Mongolia, China  
Full list of author information is available at the end of the article

## Abstract

**Background:** Exploration of the genes with abnormal expression during the development of breast cancer is essential to provide a deeper understanding of the mechanisms involved. Transcriptome sequencing and bioinformatics analysis of invasive ductal carcinoma and paracancerous tissues from the same patient were performed to identify the key genes and signaling pathways related to breast cancer development.

**Methods:** Samples of breast tumor tissue and paracancerous breast tissue were obtained from 6 patients. Sequencing used the Illumina HiSeq platform. All. Only perfectly matched clean reads were mapped to the reference genome database, further analyzed and annotated based on the reference genome information. Differentially expressed genes (DEGs) were identified using the DESeq R package (1.10.1) and DEGSeq R package (1.12.0). Using KOBAS software to execute the KEGG bioinformatics analyses, enriched signaling pathways of DEGs involved in the occurrence of breast cancer were determined. Subsequently, quantitative real time PCR was used to verify the accuracy of the expression profile of key DEGs from the RNA-seq result and to explore the expression patterns of novel cancer-related genes on 8 different clinical individuals.

**Results:** The transcriptomic sequencing results showed 937 DEGs, including 487 upregulated and 450 downregulated genes in the breast cancer specimens. Further quantitative gene expression analysis was performed and captured 252 DEGs (201 downregulated and 51 upregulated) that showed the same differential expression pattern in all libraries. Finally, 6 upregulated DEGs (CST2, DRP2, CLEC5A, SCD, KIAA1211, DTL) and 6 downregulated DEGs (STAC2, BTNL9, CA4, CD300LG, GPIHBP1 and PIGR), were confirmed in a quantitative real time PCR comparison of breast cancer and paracancerous breast tissues from 8 clinical specimens. KEGG analysis revealed various pathway changes, including 20 upregulated and 21 downregulated gene enrichment pathways. The extracellular matrix–receptor (ECM-receptor) interaction pathway was the most enriched pathway: all genes in this pathway were DEGs, including the THBS family, collagen and fibronectin. These DEGs and the ECM-receptor interaction pathway may perform important roles in breast cancer.

(Continued on next page)



(Continued from previous page)

**Conclusion:** Several potential breast cancer-related genes and pathways were captured, including 7 novel upregulated genes and 76 novel downregulated genes that were not found in other studies. These genes are related to cell proliferation, movement and adhesion. They may be important for research into breast cancer mechanisms, particularly CST2 and CA4. A key signaling pathway, the ECM-receptor interaction signal pathway, was also identified as possibly involved in the development of breast cancer.

**Keywords:** Breast cancer, Transcriptome, Differentially expressed genes, ECM-receptor interaction pathway

## Background

Breast cancer is the most common malignant tumor and the fifth leading cause of cancer-related death for women in China [1]. The morbidity and mortality for breast cancer patients are higher than for any other malignant tumor, and the risk increases annually worldwide [2]. Its genesis closely related to genetic mutations and abnormal epigenetic modifications [3]. Although substantial progress has been made in studies of genetic predisposition to breast cancer, few breakthroughs have been made regarding its development mechanism [4, 5]. Studying more diverse groups of breast cancer patients or samples could provide more insight into its cellular mechanisms. Transcriptome research would not only elucidate its cellular mechanisms and/or development progress, but also provide potential diagnostic targets [6].

A variety environmental factors, including living environment, habits and chemical exposure, contribute to the tumorigenesis of breast cancer [7]. Various genetic factors also play a role, with up to 20–40% of hereditary breast cancer patients showing particular gene mutations [8]. Many genes associated with breast cancer have been annotated and analyzed.

Mutations of breast cancer 1 (BRCA1), BRCA2 and TP53 are risk factors for a high incidence (40–66%) of breast cancer occurrence. The breast cancer 1 (BRCA1) and breast cancer 2 (BRCA2) genes normally behave as tumor suppressor genes and can maintain cell proliferation and differentiation [9]. A BRCA1 mutation can be detected in 52% of breast cancer patients [10] and up to 80% have a mutation in either BRCA1 or BRCA2 [11]. Non-mutant TP53 can regulate the life cycle of cells, mediate signaling pathways and play an important role in repairing DNA, preventing tumor recurrence and metastasis [12]. Gene polymorphism of TP53 is associated with the occurrence and development of breast cancer [13]. Some other genes, such as PTEN, ataxia telangiectasia mutated (ATM) [14], checkpoint kinase 2 homolog (CHEK2) [15], Rad50 DNA repair protein [16], BRCA1-interacting protein C-terminal helicase 1 (BRIP1) [17] and fibroblast growth factor receptor 2 (FGFR2) [18] can also contribute to the risk of breast cancer at a low probability.

Exploration of the genes and proteins that are abnormally expressed during the development of breast cancer is essential to provide a deeper understanding of the mechanisms involved [7]. However, differences in people's genetics background and living environments make it very hard to unequivocally identify a common cancer-related gene for the occurrence of breast cancer. It is essential that more cancer-related genes are discovered in samples from patients with different living environments.

Transcriptome sequencing and bioinformatics analysis can efficiently evaluate whole processes in one tissue globally [19]. Whole transcriptome profiling can reveal genes that are differentially expressed in related tissues (for example, breast tumor tissues and

paracancerous breast tissues). Changed genes in any cancer, including breast cancer, reflect the biological diversity of cellular phenotype and physiological function and could become important research areas for elucidating molecular mechanisms. Already, many studies have found genes or proteins strongly associated with the progress and prognosis of breast cancer, including enhancer of zeste 2 (EZH2) polycomb repressive complex 2 subunit [20] and Jab1/COPS5 [21]. In addition, the nuclear receptor-interacting protein 1 (NRIP1) and the MAPK signaling pathway can regulate the development of breast cancer cells [22].

## Materials and methods

### Study patients, tissue sample preparation and collection

Histopathological breast cancer (invasive ductal carcinoma, tumor tissue) and adjacent normal tissue (paracancerous tissues, normal tissue) samples were obtained from 14 patients with pathologically confirmed breast cancer. Six of the cases were randomly selected for transcriptome sequencing, while the other eight were selected for expression pattern studies of novel breast cancer-related genes. The samples were taken in 2016 and 2017 at the Department of Pathology of the Affiliated Hospital of Inner Mongolia Medical University in Hohhot, China. This study was approved by the Ethics Committee of Inner Mongolia Medical University. After surgery, each specimen was cut into 3–8 mm<sup>2</sup> pieces. The cut tissue was immediately placed in an RNA protectant (RNAlater, Sigma Aldrich). After being infiltrated by RNAlater for 12 h at 4 °C, all samples were then quickly placed in liquid nitrogen and stored at – 80 °C until needed for further processing and sequencing.

### Isolation of total RNA from samples

Total RNA extraction was performed with TRIzol (Takara) following the manufacturer's protocol, and isolated total RNA was stored at – 80 °C until the next step. RNA degradation and contamination was monitored using 1% agarose gel electrophoresis. A Nano-Photometer spectrophotometer (Implen) and Qubit RNA Assay Kit with a Qubit 2.0 Fluorometer (Life Technologies) were respectively used to detect the purity and concentration of total RNA. An RNA Nano 6000 Assay Kit and a Bioanalyzer 2100 system (Agilent Technologies) were used to assess the integrity of total RNA.

### Library preparation for transcriptome sequencing

At least 3 µg of total RNA per sample was used. Following the manufacturer's instructions, the NEBNext Ultra RNA Library Prep Kit (Illumina) was used to generate the 6 pairs of sequencing libraries (6 for normal tissues and 6 for tumor tissues). Random hexamer primer, M-MuLV Reverse Transcriptase (RNase H-) and DNA Polymerase I followed by RNase H were respectively used to synthesize first- and double-stranded cDNA. Any remaining overhangs were converted into blunt ends with exonuclease and polymerase. The AMPure XP system (Beckman Coulter) was used to select cDNA fragments, preferentially those that were ~ 150–200 bp in length. Three microliters of USER Enzyme was used with size-selected, adaptor-ligated cDNA at 37 °C for 15 min followed by 5 min at 95 °C, and then PCR was performed. Two pairs of cDNA libraries were constructed: one from the cDNA libraries from 6 normal tissue samples (named N1 to N6) and the other from the cDNA libraries from 6 tumor tissue samples (named

T1 to T6). The Illumina HiSeq platform (Beijing Novo Gene Biological) was used for transcriptome sequencing according to the manufacturer's instructions.

### Bioinformatics analysis

Raw (sequenced) reads were obtained first. After raw read filtering, sequencing error rate check (Q20 and Q30) and GC content profiling, the high-quality clean paired-end reads from each sample were aligned to the reference genome using TopHat version 2.0.9. The mapped genes from the reference genome were queried in databases such as UniProtKB/SwissProt and the Non-Redundant Protein Database (NRPD) with the help of BLASTX program (E-value cut-off of  $1e^{-5}$ ). The mapped read numbers in the exon and intron regions (exonic and intronic rates) and total mapping rate were analyzed independently using HTSeq version 0.6.1. The total number of mapped reads was determined and the RPKMs (reads per kilobase of exon model per million mapped reads) were calculated for each annotated gene. The software R package of DESeq was employed to capture the DEGs (differentially expressed genes) between same pair transcriptome data from the same individual and calculate the fold changes for each gene. Genes with fold changes  $> 2$ , q values  $< 0.01$  and FDRs  $< 0.01$  were defined as DEGs and captured for further analysis. All DEGs were enriched to the KEGG signal pathway based on a q value  $< 0.01$  and FDR  $< 0.01$ .

The results for DEGs obtained in this study were compared to corresponding breast cancer research transcriptome information from the GEO database (especially the latest research GSE33447 and GSE109169).

### Validation and clinical experiments with quantitative real time PCR

The validation experiment was performed with the same 6 pairs of breast cancer tissue and adjacent normal tissue used for the transcriptome sequencing. The following 12 genes were selected to perform the quantitative real time PCR: pituitary tumor-transforming 1 (PTTG1), TTK protein kinase (TTK), COL10A1, CYCS, eukaryotic translation elongation factor 1 alpha 2 (EEF1A), BUB1B, CCNB1, CDC20, karyopherin alpha 2 (RAG cohort 1 importin alpha 1; KPNA2), tetraspanin 1 (TSPAN1), tetraspanin 13 (TSPAN13) and tetraspanin 15 (TSPAN15). The group includes cancer-related genes identified in previous research. Primers were designed and are listed in Additional file 1: Table S1. 18S ribosomal RNA expression was used as an internal reference. The reaction system consisted of  $2 \times$  Super Real PreMix Plus, forward primer (10  $\mu$ M), reverse primer (10  $\mu$ M), cDNA and  $50 \times$  ROX Reference Dye, and the volumes of RNase-Free ddH<sub>2</sub>O used were 0.4, 0.6, 1, 7.4 and 10  $\mu$ l. The PCR amplifications were performed in triplicate wells with initial denaturation at 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 34 s.

The clinical experiment was prepared with the total RNA from the other different 8 pairs of breast cancer tissue and adjacent normal tissue. The first-strand cDNA was synthesized using a PrimeScript RT reagent Kit with gDNA Eraser (Perfect real time PCR). The expression levels of upregulated and downregulated genes that were selected as novel breast cancer-related genes were verified via quantitative real time PCR. The primers, PCR system and amplification conditions were the same as in the validation experiment. The data were analyzed using ABI 7500 HT SDS software 4.1 (Applied Biosystems). DEG expression levels were analyzed using the  $2^{-\Delta\Delta CT}$  analysis method.

## Results

### The sequencing and transcriptome data

The relevant parameters, including raw reads, clean reads and total mapped rate of breast cancer tissue and normal breast tissue are summarized in Table 1. Based on filtered sequenced reads, we obtained 164,352,319 clean reads in normal tissues and 166,067,405 in tumor tissues. The average Q20, Q30, exonic, intronic and total mapping rates for tumor samples were 96.18, 90.9, 80.37, 15.8 and 92.88%, respectively. All raw reads were submitted to the NCBI SRA database under the accession number PRJNA528582.

In total, 39,649 different genes were annotated from the whole transcriptome. Within these sequences, 4685 lncRNAs, 923 miRNAs, 926 misc. RNAs, 170 rRNAs and 18,800 protein-coding genes were annotated based on the various reference databases. In total, 18,013 genes were known protein-coding genes and 787 sequences were novel genes that were not mentioned in any database. These unknown protein-coding genes may be novel genes.

### Search for known cancer-related genes in breast cancer tissue

In total, 93 different previously reported cancer-related genes were annotated in this study (Additional file 2: Table S2). This included 7 breast cancer-related genes (Table 2): caspase 8 (CASP8), cadherin 1 type 1 (CDH1), estrogen receptor 1 (ESR1), ETS variant 6 (ETV6), forkhead box A1 (FOXA1), GATA-binding protein 3 (GATA3) and neurotrophic tyrosine kinase receptor type 3 (NTRK3). The expression levels of GATA3 and ESR1, which are both the breast tumor-related genes, showed upregulation in tumor tissues compared with normal tissues. The GATA3 expression level was 15,000 in tumor tissues and 5000 in normal tissues. The expression level was ESR1 4700 in tumor tissues and 1500 in normal tissues.

Of the 93 cancer-related genes, 58 were downregulated in the tumor tissue transcriptome. The WNT inhibitory factor 1 (WIF1) gene, which is related to the tumor type pleomorphic salivary gland adenoma, was the gene with the greatest downregulation (16-fold change), while a member of the ETS family of transcription factors (ETV6), which is related to non-small cell lung cancer, had the joint smallest downregulation

**Table 1** The details of the transcriptome assembly result

| Sample name | Raw reads  | Clean reads | Clean bases | Total mapping rate (%) | Q20 (%) | Q30 (%) | Exonic | Intronic |
|-------------|------------|-------------|-------------|------------------------|---------|---------|--------|----------|
| N1          | 26,062,090 | 26,062,090  | 7.82G       | 93.37                  | 96.91   | 92.34   | 81.18  | 15.28    |
| N2          | 28,333,875 | 28,333,875  | 8.5G        | 92.49                  | 96.97   | 92.61   | 75     | 20.16    |
| N3          | 23,342,517 | 23,342,517  | 7.0G        | 93.59                  | 97.11   | 92.8    | 74.14  | 20.96    |
| N4          | 31,528,594 | 30,258,328  | 9.08G       | 86.85                  | 96.09   | 90.75   | 89.19  | 7.88     |
| N5          | 23,513,213 | 22,160,567  | 6.65G       | 90.18                  | 95.2    | 88.91   | 86.13  | 10.32    |
| N6          | 31,572,030 | 29,732,384  | 8.92G       | 89.70                  | 95      | 88.59   | 85.46  | 10.98    |
| T1          | 27,667,902 | 27,667,902  | 8.3G        | 93.19                  | 96.88   | 92.33   | 83.1   | 13.85    |
| T2          | 26,369,158 | 26,369,158  | 7.91G       | 93.89                  | 96.84   | 92.3    | 72.62  | 22.07    |
| T3          | 26,583,360 | 26,583,360  | 7.98G       | 92.58                  | 97.02   | 92.73   | 74.98  | 20.07    |
| T4          | 23,947,809 | 22,670,303  | 6.8G        | 91.81                  | 95.49   | 89.41   | 84.2   | 12.96    |
| T5          | 36,079,555 | 34,114,397  | 10.23G      | 92.74                  | 95.39   | 89.25   | 84.71  | 12.06    |
| T6          | 25,419,621 | 24,060,173  | 7.22G       | 93.06                  | 95.46   | 89.39   | 82.59  | 13.8     |

**Table 2** Breast cancer-related genes

| Gene symbol | Tumor types  |
|-------------|--|
| CASP8       | Hepatocellular; oral squamous cell; breast   |
| CDH1        | Lobular breast; gastric  |
| ESR1        | Breast   |
| ETV6        | Congenital fibrosarcoma; multiple leukemia and lymphoma types including ALL; secretory breast; MDS |
| FOXA1       | Breast; prostate   |
| GATA3       | Breast   |
| NTRK3       | Congenital fibrosarcoma; secretory breast  |

(0.64-fold change). Only neurotrophic tyrosine kinase receptor type 3 (NTRK3; 5.76-fold downregulation) and ETV6 (0.64-fold downregulation) were related to breast cancer. Of the 35 upregulated genes, 5 were all related to breast tumor types: CASP8 (0.7-fold upregulation), CDH1 (1.21-fold upregulation), GATA3 (3-fold upregulation), ESR1 (3-fold upregulation) and FOXA1 (+ 2.89-fold upregulation).

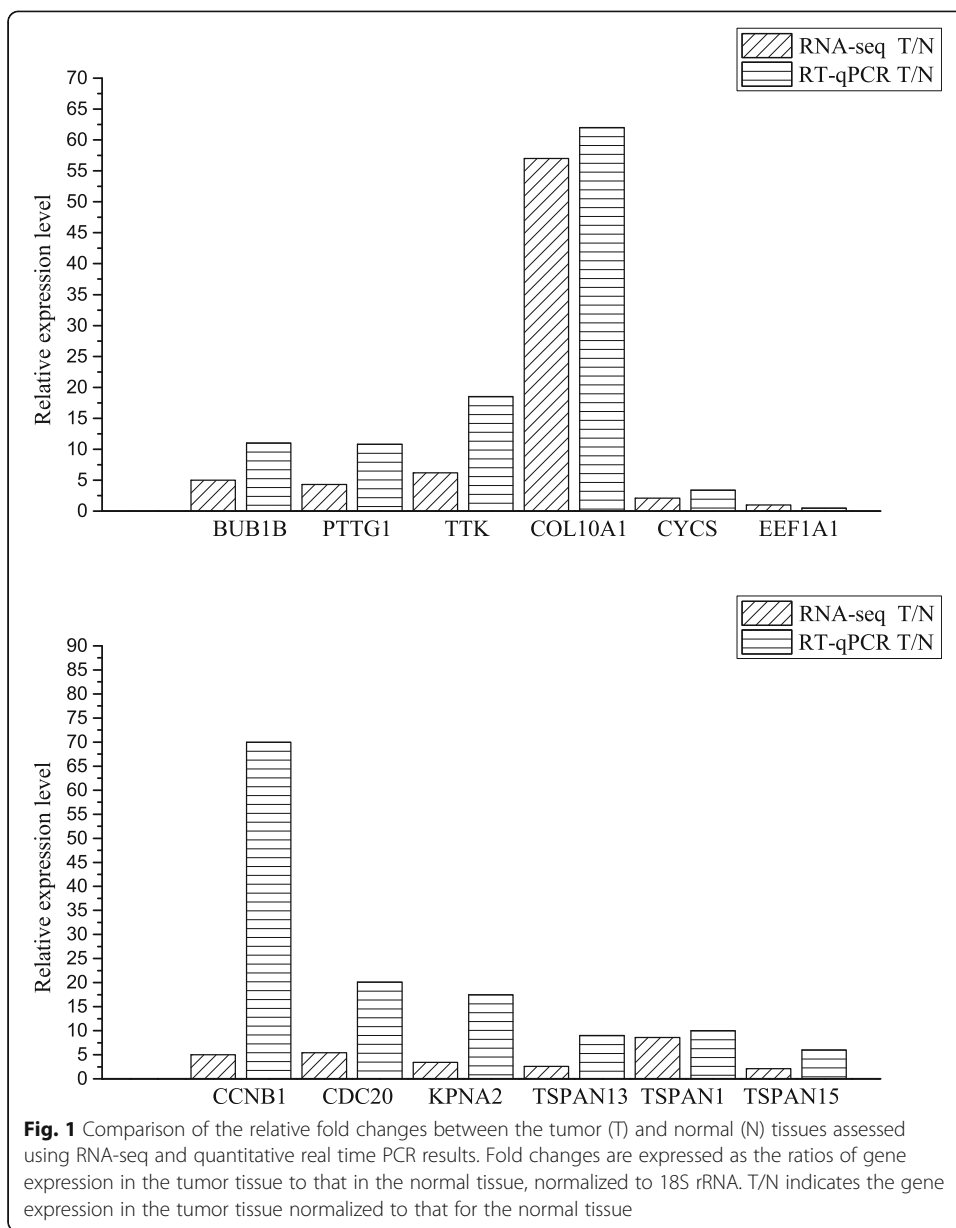
**Validate the accuracy of the transcriptome expression results using quantitative real time PCR**

To validate the accuracy of the transcriptome expression results, we measured the expression levels of 12 randomly selected genes via quantitative real time PCR: pituitary tumor-transforming 1 (PTTG1), TTK protein kinase (TTK), COL10A1, CYCS, eukaryotic translation elongation factor 1 alpha 2 (EEF1A), BUB1B, CCNB1, CDC20, karyopherin alpha 2 (RAG cohort 1 importin alpha 1) (KPNA2), tetraspanin 1 (TSPAN1), tetraspanin 13 (TSPAN13) and tetraspanin 15 (TSPAN15). The expression patterns of these 12 genes provide evidence that the transcriptome was accurate (Fig. 1). There was a significant correlation between the two methods, with coefficients ranging from 0.91 to 0.96. This result implied that the RNA-seq result could reflect gene expression levels in the tissues.

**Identification and analysis of differentially expressed genes (DEGs)**

In the tissue collections and sequencing program, the breast tumor and paracancerous tissue samples from 6 patients were treated independently. These 6 different sample pairings were sequenced, mapped, analyzed and annotated. The DESeq R package (1.10.1) and DEGSeq R package (1.12.0) were used to identify the DEGs in the different libraries from the same individual patient. Pairwise comparisons for DEG analysis were performed between the tumor tissues and paracancerous tissues in the six individual groups.

Interestingly, due to individual differences, the 6 groups' transcriptomes showed little difference in the number of DEGs (Table 3). In total, 937 DEGs (487 upregulated genes and 450 downregulated genes) were found to be differentially expressed in at least one tumor tissue compared with the paracancerous tissues within the same individual (Table 3). Further analysis shown that only 26.9% of the identified genes (252 of 937 DEGs) have a similar expression pattern among all individual groups, which indicated that the effect of individual differences must be taken into account when we define a universal cancer-related gene for breast tumors. Meanwhile, these 252 DEGs, including 51 upregulated genes and 201 downregulated genes (Fig. 2), showed the same up- or

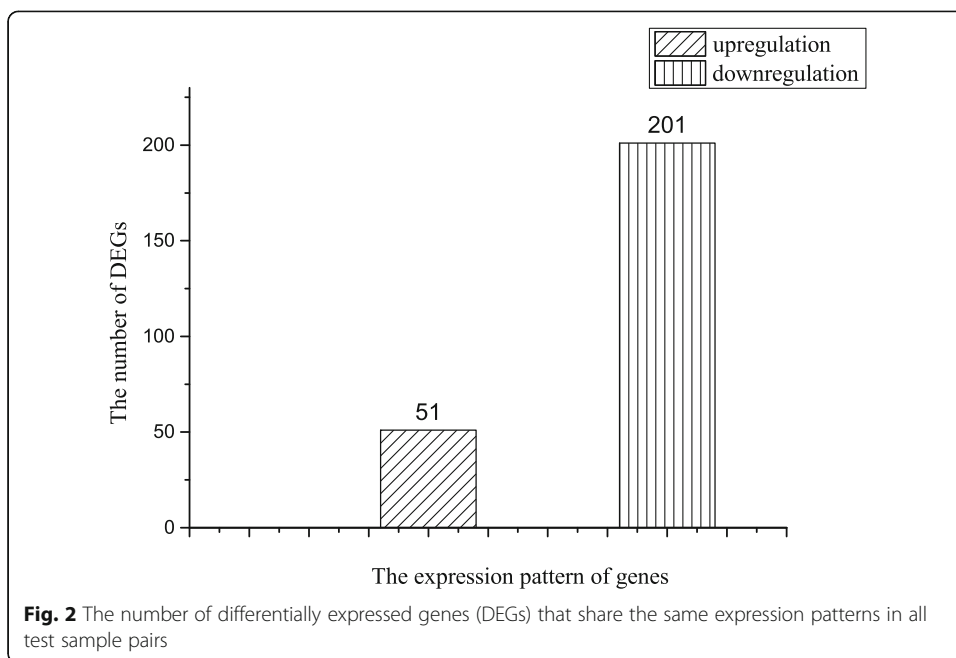


downregulation pattern in all 6 breast tumor transcriptomes with a q value < 0.01 and false discovery rate (FDR) < 0.01. Of the 51 upregulated genes, 44 were identified in the previous study (GEO result) and only 7 (CST2, DRP2, CLEC5A, SCD, KIAA1211, RP1-34B20.4, DTL) had not been studied. Of the 201 downregulated genes, 125 were identified in the previous study (GEO results) and only 76, such as cysteine rich domain 2 (STAC2), BTNL9, CA4, GPIHBP1 and PIGR, had not been studied on any cancer.

**Table 3** The differentially expressed genes in all transcriptome groups

|               | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 | Group 6 | All |
|---------------|---------|---------|---------|---------|---------|---------|-----|
| Upregulated   | 461     | 418     | 537     | 556     | 534     | 474     | 487 |
| Downregulated | 370     | 336     | 338     | 316     | 360     | 411     | 450 |
| Total         | 830     | 754     | 875     | 872     | 894     | 885     | 937 |





Of all the DEGs, 51 were upregulated in breast cancer tissues (Additional file 3: Table S3). Ffibronectin-1 (FN1) showed the highest expression level in the tumor tissue transcriptome: 71,967, which was 10-fold higher than that in the paracancerous tissues transcriptome. The vacuole membrane protein 1 (VMP1) gene exhibited the second highest expression level, followed by collagen triple helix repeat containing-1 (CTHRC1), inhibin beta A (INHBA), and matrix metalloproteinase 11 (MMP11). The relative expression levels of these DEGs were higher than 4000 in tumor tissues and less than 2000 in paracancerous tissues (Table 4). Of these 51 genes, 44 could be cancer-related based on the reference and previous research. Twenty genes were not mentioned in any study about breast cancer.

More genes in the cancer transcriptome showed a downregulated expression pattern than an upregulated one. A total of 201 DEGs were downregulated (Additional file 4: Table S4). The top 10 genes with the largest differences were delta-like 1 homolog (DLK1), CA4, phospholipid phosphatase related 1 (LPPR1), adhesion G protein-coupled receptor D2

**Table 4** Upregulated genes in breast cancer tissue

| Gene name | log2 fold change | q value  | Gene description                          |
|-----------|------------------|----------|---|
| FN1       | 3.604814         | 1.59E-18 | Fibronectin1                              |
| VMP1      | 1.625331         | 3.89E-09 | Vacuole membrane protein1                 |
| CTHRC1    | 2.77433          | 2.99E-11 | Collagen triple helix repeat containing 1 |
| INHBA-AS1 | 0.559233         | 0.348753 | INHBA antisense RNA1                      |
| MMP11     | 5.286475         | 7.26E-19 | Matrix metalloproteinase11                |
| FOXA1     | 1.699589         | 1.05E-06 | Forkhead box A1                           |
| CST2      | 5.665185         | 2.98E-12 | Cystatin SA                               |
| HIST1H2AI | 5.189283         | 1.84E-13 | Histonecluster1 H2ai                      |
| HIST1H2BF | 4.235548         | 1.39E-09 | NA  |
| DRP2      | 3.339072         | 5.25E-10 | Dystrophin-related protein 2              |
| ASPM      | 2.946803         | 5.58E-09 | Abnormal spindle microtubule assembly     |



(GPR144), CD300LG, heparanase 2 (HPSE2), solute carrier family 13 (sodium-dependent dicarboxylate transporter) member 2 (SLC13A2), heparan sulfate-glucosamine 3-sulfotransferase 4 (HS3ST4), polymeric immunoglobulin receptor (PIGR), and ciliary neurotrophic factor receptor (CNTFR). These genes were downregulated by 5- to 128-fold.

The reads per kilobase of transcript per million mapped reads (RPKM values) of these genes were not less than 2000 in normal tissues but were more than 700 in tumor tissues. Two genes, PIGR and BTNL9, showed downregulation by 32-fold and 26-fold, respectively, in tumor tissues compared to normal tissues (Table 5).

### KEGG pathway enrichment analysis

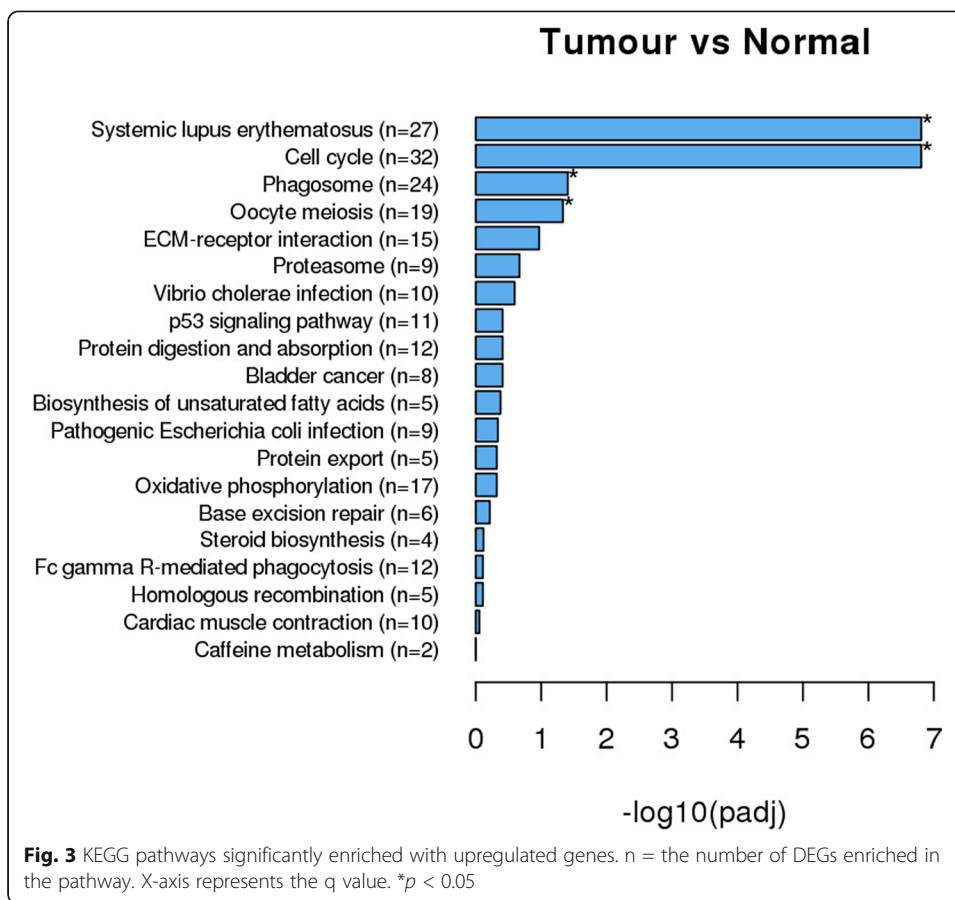
KEGG is a database for the molecular or system biology study of gene clusters. These genes perform their functions at different levels (e.g., cell and organism levels). KOBAS software was used to test the statistical enrichment of DEGs in the KEGG pathways. A total of 937 DEGs were enriched in 219 different KEGG pathways, and 41 significant DEG-enriched KEGG pathways (21 downregulated pathways and 20 upregulated pathways) were annotated.

Among the upregulated pathways, the extracellular matrix–receptor (ECM-receptor) interaction (22 DEGs), systemic lupus erythematosus (27 DEGs), phagosome (24 DEGs), oocyte meiosis (19 DEGs) and cell cycle (32 DEGs) pathways were significantly enriched in all 6 transcriptomes. All DEGs annotated in the ECM-receptor interaction pathway, including collagen, THBS, fibronectin and BSP, were upregulated in the tumor tissues (Figs. 3 and 4).

Similarly, 9 downregulated pathways were significantly enriched: the axon guidance pathway (28 DEGs), ether lipid metabolism pathway (12 DEGs), salivary secretion pathway (21 DEGs), PPAR signaling pathway (18 DEGs), metabolism of xenobiotics by cytochrome P450 pathway (16 DEGs), tyrosine metabolism pathway (12 DEGs), protein

**Table 5** Downregulated genes in breast cancer tissue

| Gene name | log2 fold change | q value  | Gene description  |
|-----------|------------------|----------|---|
| DLK1      | -5.55726         | 9.45E-17 | Delta-like1homolog ( <i>Drosophila</i> )                                      |
| CA4       | -5.33851         | 4.63E-10 | Carbonic anhydrase IV   |
| LPPR1     | -3.26132         | 1.17E-07 | Phospholipid phosphatase related1   |
| GPR144    | -6.65466         | 1.85E-32 | Adhesion G protein-coupled receptor D2  |
| CD300LG   | -4.33055         | 1.26E-09 | CD300 molecule like family member g   |
| HPSE2     | -4.70027         | 6.79E-13 | Heparanase 2 (inactive)   |
| SLC13A2   | -4.26341         | 5.06E-16 | Solute carrier family13 (sodium-dependent dicarboxylate transporter) member 2 |
| HS3ST4    | -4.99125         | 8.68E-16 | Heparan sulfate-glucosamine3-sulfo transferase 4                              |
| PIGR      | -4.83315         | 1.31E-28 | Polymeric immune globulin receptor  |
| CNTFR     | -3.40391         | 9.08E-10 | Ciliary neuro trophic factor receptor   |
| BTNL9     | -3.07588         | 4.88E-08 | Butyrophilin-like9  |
| CDH12     | -3.86619         | 4.58E-08 | Cadherin12 type 2 (N-cadherin2)   |
| GNA11     | -0.73688         | 0.00466  | Guanine nucleotide-binding protein (Gprotein) alpha11 (Gqclass)               |
| DLK1      | -5.55726         | 9.45E-17 | Delta-like 1 homolog ( <i>Drosophila</i> )                                    |
| GPIHBP1   | -3.70459         | 4.43E-12 | NA  |

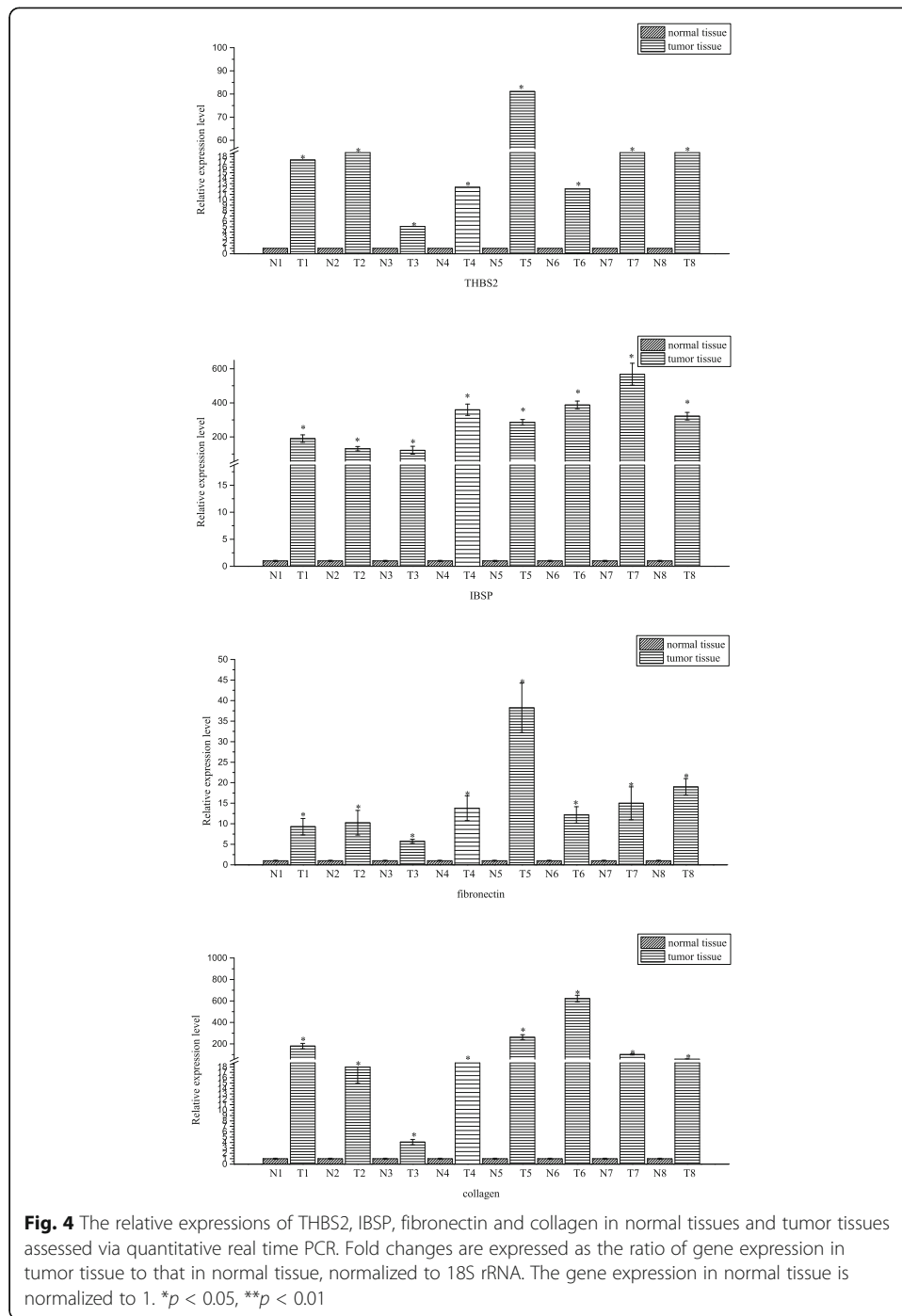


digestion and absorption pathway (18 DEGs), focal adhesion pathway (36 DEGs) and neuroactive ligand-receptor interaction pathway (43 DEGs). The PPAR signaling pathway was annotated as a downregulated DEG enrichment pathway in all different 6 transcriptomes, and the 18 DEGs, including fatty acid binding protein 7 brain (FABP7), solute carrier family 27 (fatty acid transporter) member 6 (SLC27A6), solute carrier family 27 (fatty acid transporter) member 1 (SLC27A1) and collagen domain-containing (ADIPOQ), showed downregulation by 1.5-fold to 6.7-fold in all sequencing groups (Fig. 5).

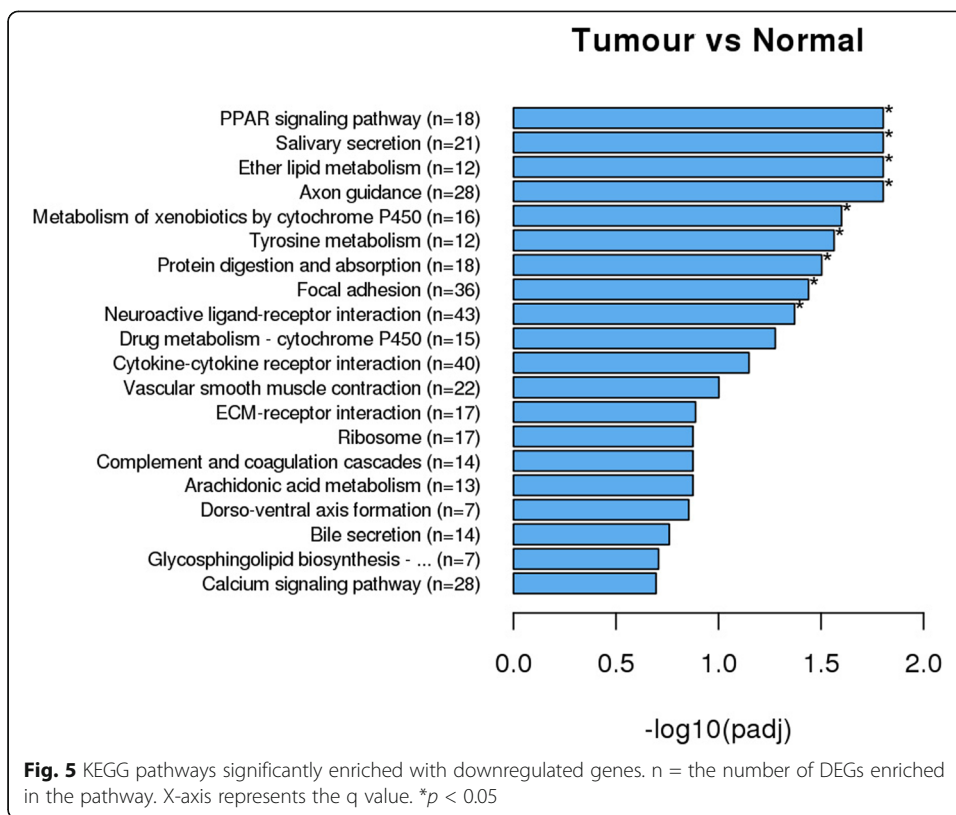
**Search for potential cancer-related genes in DEGs from breast cancer tissue**

Only genes that showed the same expression pattern in all 6 transcriptome pairs were taken into consideration. Of these 51 genes, CST2 showed the biggest expression differences between tumor tissues and paracancerous tissues (350-fold upregulation). Only ~ 1 mean relative expression level was detected in the normal tissues. Functional analysis revealed that this gene is a protein-coding gene, 748 bp in length, and located on chromosome 20. The other genes with high fold changes, dystrophin related protein 2 (DRP2) and COL10A1, were also annotated. COL10A1 showed a relative expression level of 3937 in breast tumor tissues and only 21 in paracancerous breast tissues.

Of the 201 downregulated genes, DLK1 exhibited a 128-fold downregulation in breast tumor tissues. However, the RPKM values of this gene were not very high in the



transcriptomes (37 in normal tissue and 0.3 in tumor tissue). Its low expression level may mean it is not a good cancer-related gene for breast tumors. CD300LG and BTNL9, which exhibited more than 32-fold downregulation in all the tumor transcriptomes, showed a very high differential expression patterns. The expression level of CD300LG (2343 RPKM) and BTNL9 (7326 RPKM) were very high in normal tissues but very low in the tumor transcriptome (56 RPKM and 283 RPKM, respectively). The same result was observed in the expression pattern of polymeric immunoglobulin receptor (PIGR), which demonstrated a negative 32-fold change (12,789 RPKM in normal tissues and 412 RPKM in



tumor tissues). These genes could be potential low expression level breast cancer-related genes.

**The clinical experimental with quantitative real time PCR**

To determine the clinical effects, we screened 6 high expression level and 6 low expression level genes to determine expression patterns in breast cancer tissues and adjacent tissues from 8 different patients. All quantitative real time PCR primers were designed based on the gene sequences reported in the NCBI database (Additional file 1: Table S1 primers). The results showed that the upregulated *CST2*, *GJB2*, *UBE2T*, *NUF2*, *ORC6* and *CCNB1* (Fig. 6), and the downregulated *ELF5*, cysteine-rich domain 2 (*STAC2*), *BTNL9*, *CA4*, *CD300LG* and *PIGR* (Fig. 7) showed the same result in different patients. This also verified the RNA-seq results. These 12 genes could be new cancer-related genes for the clinical treatment of breast cancer.

**Discussion**

Using next-generation sequencing technology and quantitative real time PCR, we successfully analyzed the DEGs in breast cancer tissues from patients from Inner Mongolia in China. Since the molecular changes in breast cancer tissues are dependent on tumor type, grade, size and receptor status [23–25], we limited our study to invasive cases. Transcriptome sequencing techniques play an important role in the identification of cancer-specific genes [3, 5, 6, 19]. We sequenced the transcriptome from 6 pairs of

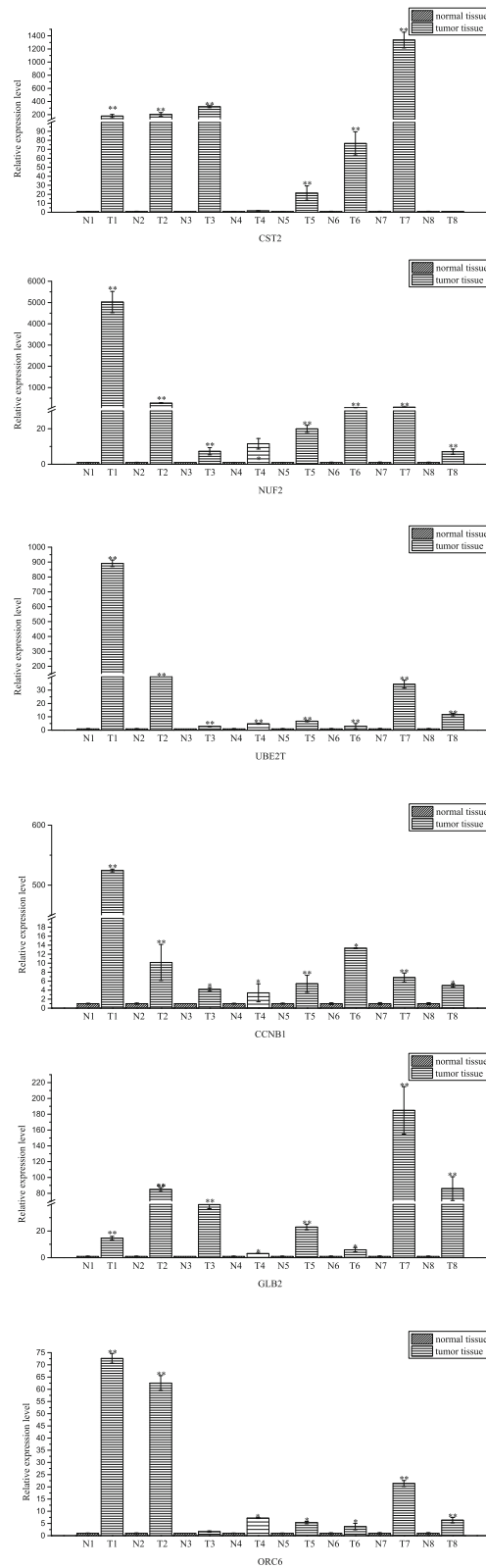


Fig. 6 (See legend on next page.)

(See figure on previous page.)

**Fig. 6** The relative expressions of CST2, GJB2, NUF2T, NUF2, ORC6 and CCNB1 in normal tissues and tumor tissues assessed via quantitative real time PCR. Fold changes are expressed as the ratio of gene expression in tumor tissue to that in normal tissue, normalized to 18S rRNA. The gene expression in normal tissue is normalized to 1. \* $p < 0.05$ , \*\* $p < 0.01$

breast cancer tissues and adjacent normal tissues and compared expressions in each pair, finding that 51 DEGs were upregulated and 201 DEGs were downregulated.

Because the gene expression patterns or the transcriptomes of cancer patients are greatly impacted by multiple factors, including living environment [26] and the severity of the disease [27], there can be considerable inter-patient variation. The DEG results in this study support the fluctuations in the number of DEGs between the breast cancer tissue and paracancerous tissue in different individuals. They also confirm that the expression levels of DEGs display significant differences between breast cancer patients.

At the same time, the statistical results of all DEGs in our study showed that each patient expressed unique DEGs (937 DEGs in total and 253 DEGs in common). The expression patterns of many DEGs in the transcriptome were not stable, which may due to the development of the disease or the genetic background of the individual [7]. This is a hindrance for researchers seeking universal cancer-related genes for breast cancer. Therefore, individual differences must be taken into account when conducting subsequent studies.

The expressions of three tetraspanin family members, TSPAN1, TSPAN13 and TSPAN15, are upregulated. They all function as transmembrane transport proteins, and TSPAN15 is also associated with the notch signaling pathway [28, 29]. TSPAN1 has been reported to regulate the progression of many human cancers, including gastric cancer, pancreatic cancer and cervical cancer [30–32]. Meanwhile, the expression of TSPAN1 was higher in ER-positive and HER2-positive breast cancer [33]. All samples in this study were collected from ER-positive patients. While the expression of TSPAN13 in prostate cancer and glioma is known to be elevated [34, 35], there is only one study on TSPAN13 in breast cancer [36]. It mentioned that TSPAN13 was upregulated in breast cancer cells. There are few studies on TSPAN15, and its effect on cancer was reported less frequently.

In our results, the expression levels of TSPAN1, TSPAN13, and TSPAN15 in breast cancer were all increased. Our TSPAN1 results are consistent with those previously reported [33], so we speculate that TSPAN13 and TSPAN15 may be potential cancer-related genes for breast cancer. This needs further study.

Our validation showed that the expression patterns of BUB1B, CCNB1, CDC20, COL10A1, CYCS, EEF1A2, gap junction protein beta 2 (GJB2), KPNA2, PTTG1, RAB31, TTK, UBE2C, ELF5 and STAC2 were the same in all individual patients. These genes have been reported as cancer-related for breast cancer [23–25, 37–45].

Previous reports [25] have shown that in invasive breast cancer, upregulated genes are related to cell proliferation and cell movement, while downregulated genes are associated with cell adhesion. Our study showed that ASPM, INHBA, NUF2, ORC6, UBE2T and PKMYT1 are associated with cell proliferation [46–52], and the expressions of these genes were also elevated in our breast cancer tissue transcriptome. The immune function-related genes CD300LG and PIGR were also detected as downregulated in breast cancer [53, 54].

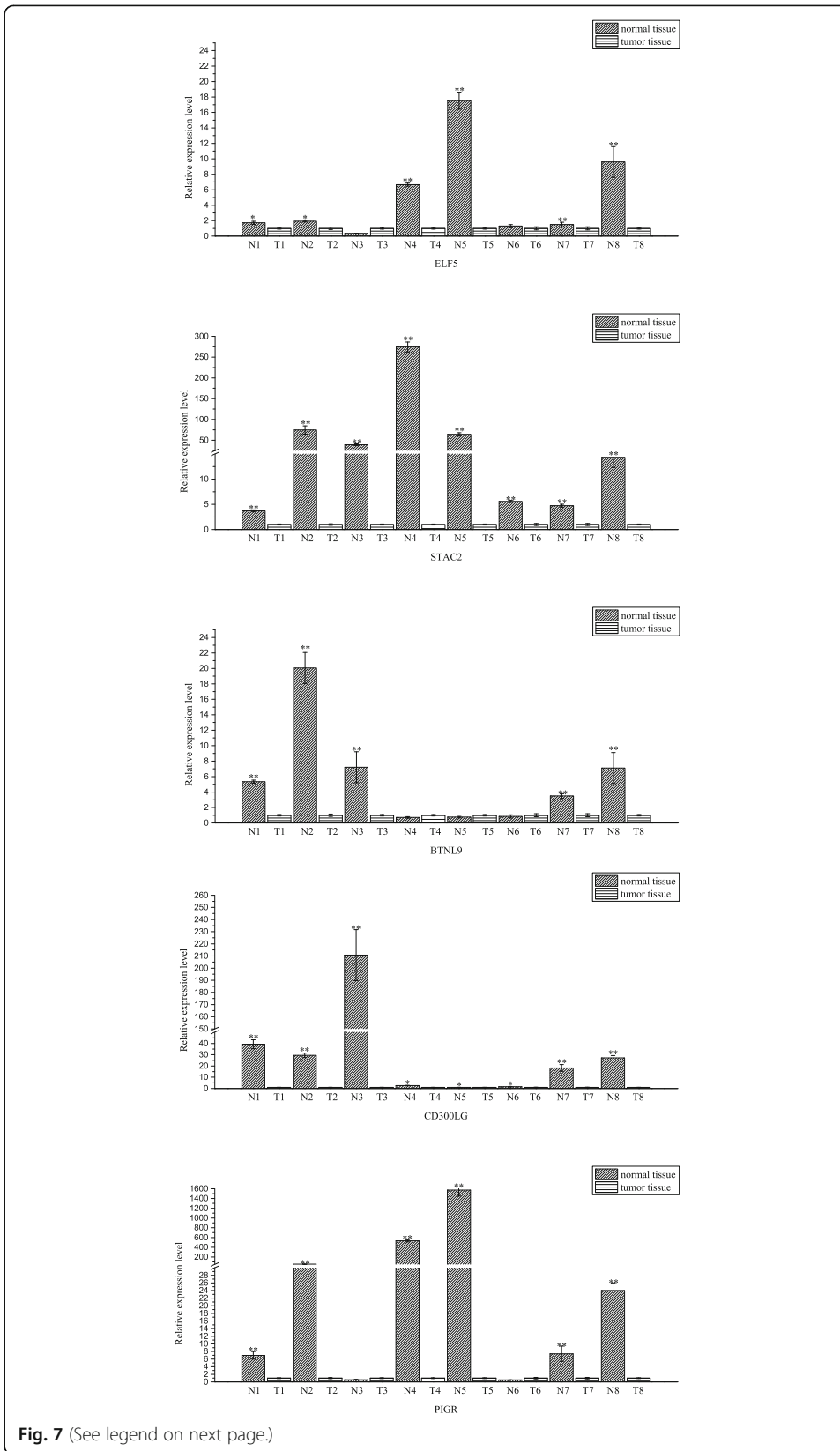


Fig. 7 (See legend on next page.)



(See figure on previous page.)

**Fig. 7** The relative expressions of BTNL9, CA4, CD300LG, ELF5, PIGR and STAC2 in normal tissues and tumor tissues assessed via quantitative real time PCR. Fold changes are expressed as the ratio of gene expression in tumor tissue to that in normal tissue, normalized to 18S rRNA. The gene expression in normal tissue is normalized to 1. \* $p < 0.05$ , \*\* $p < 0.01$

In this study, 7 upregulated and 76 downregulated DEGs were captured and may be the important genes for breast cancer research. Of the 6 upregulated genes, CST2, which belongs to the cystatin superfamily and is an active cysteine protease inhibitor [55], showed a 350-fold change compared to its expression in normal tissue. The protein of this gene is found in a variety of human fluids and secretions [55], which could provide a new detection method for breast cancer. Till now, few studies have focused on CST2 in any tumor type, except to show that is responsive to the anti-growth activity of triptolide in ovarian cancer cells [56]. More study should be performed to confirm the function and character of CST2 in the development of breast cancer.

The other high expression level gene in breast tumors was DRP2, which is associated with paranoid-type schizophrenia [57]. Some study suggests a relationship between DRP2 and lung cancer [58] and brain cancer [59]. The function of this gene in breast cancer is still unknown.

Just like the CST2, GJB2, UBE2T, NUF2 and ORC6 also showed the same high expression level in breast tumors. GJB2 is implicated in the mechanisms of invasion of ductal breast carcinoma [60] and it is a prognosis marker in pancreatic cancer [61]. The downregulation of UBE2T could inhibit the progression of gastric cancer [62] and performed the same function in prostate cancer [63]. Previous study indicated that the downregulation of NUF2 could inhibit the growth of pancreatic cancer [64]. Few studies have focused on the gene function of ORC6 in breast cancer, but single-nucleotide polymorphisms (SNPs) were detected in this breast cancer-related gene [65].

We found more genes with low expression levels in tumors: 63 with at least a 10-fold change, including BTNL9, CA4, GPIHBP1 and PIGR. In total, 6 low expression level genes were confirmed using quantitative real time PCR for 6 transcriptome specimens and 8 clinical specimens.

BTNL genes show expression pattern changes in intestinal inflammation and colon cancer [66] and may be important in tumor immunity [67]. The expression and prognostic significance of PIGR, an immunoglobulin receptor, is similar, in epithelial ovarian cancer [68]. CA4, which is involved in cell proliferation, has been shown to inhibit cell proliferation, invasion and metastasis and was downregulated in our results [69]. The glycosylphosphatidylinositol high-density lipoprotein binding protein 1 (GPIHBP1) acts to chaperone secreted LPL and interact in fatty acids and breast cancer [70]. The role of GPIHBP1 has yet to be studied in cancer.

To the best of our knowledge, the function of these genes in breast cancer has not received much coverage. More study should be performed to explore the role of these genes. An expression pattern like the one we found for these genes may imply a high risk of breast cancer.

The KEGG pathway annotation showed that all DEGs were significantly enriched in 20 pathways, including the ECM-receptor interaction pathway and the protein digestion and absorption pathway, suggesting that there are many DEGs and signaling

pathways involved in breast cancer. This is also a major reason why breast cancer is so difficult to cure. ECM-receptor interaction pathways were the most upregulated gene-enriched signaling pathways. They play an important role in the process of tumor shedding, adhesion, degradation, movement and hyperplasia. The role of ECM in other cancers has been proved. ECM is upregulated in prostate cancer tissue [71] and participates in the process of tumor invasion and metastasis in gastric cancer [72]. The ECM in colorectal cancer could promote the development of epithelial–mesenchymal transition (EMT) in cancer cells [73]. Glioblastoma is the deadliest adult brain tumor. It shows pathological features of abnormal neovascularization and diffuse infiltration of tumor cells. The interactions between the ECM and the glioblastoma microenvironment are important in this progression [74].

During tumor metastasis, tumor cells pass through the ECM, and the tumor suppressor nischarin may prevent cancer cell migration by interacting with many proteins [75]. Some studies have revealed that nischarin can prevent the migration and invasion of breast cancer cells by changing the expression patterns of key adhesive proteins [76]. The expression of nischarin could reduce the ability of cells to attach to the ECM, which would lead to a decrease in invadopodium-mediated matrix degradation [77].

Invasive metastasis is a unique biological feature of malignant tumors. The high expression level of ECM proteins or genes in breast tumor tissues may provide new ideas for cancer treatment. We think that these genes and pathways may be potential markers for breast cancer, but the mechanisms of tumorigenesis and development need to be verified in further experiments.

## Additional files

**Additional file 1: Table S1.** primers. Primers for all the differentially expressed genes for quantitative real time PCR analysis. (XLSX 11 kb)

**Additional file 2: Table S2.** Tumor type-related genes. (XLSX 42 kb)

**Additional file 3: Table S3.** The 51 upregulated genes in breast cancer tissues. (XLSX 18 kb)

**Additional file 4: Table S4.** The 201 downregulated genes in breast cancer tissues. (XLSX 43 kb)

## Abbreviations

BLAST: Basic Local Alignment Search Tool; DEGs: Differentially expressed genes; ECM: Extracellular matrix; KEGG: Kyoto Encyclopedia of Gene and Genomes

## Acknowledgements

None.

## Authors' contributions

YLB and YFJ conceived and designed this study. LW, LS and FY participated in the design of the study and performed statistical analyses. XL, YXC, CC and YNR conducted experiments and contributed to the preparation of the manuscript. YLB and YFJ analyzed the results and revised the manuscript. All authors read and approved the final manuscript.

## Funding

This project was supported by the Inner Mongolia Autonomous Region Tumor Biotherapy Collaborative Innovation Center (No. 2016ZLXT009), the Inner Mongolia Autonomous Region University Innovation Research Team project (NMGIRT-A1604), and the Inner Mongolia Autonomous Region Youth Science and Technology Talent Support plan (NJYT-14-A11).

## Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

#### Ethics approval and consent to participate

We confirm that any aspect of the study that involved human patients was approved by the Ethics Committee of Inner Mongolia Medical University. The research was conducted in accordance with the Helsinki Declaration. The Ethics Committee of Inner Mongolia Medical University made their declaration on 03/06/2018, with the number YKD2018267. This project did not contain any studies with animals performed by any of the authors.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>College of Basic Medicine, Inner Mongolia Medical University, Hohhot, Inner Mongolia, China. <sup>2</sup>Department of Pathology, Inner Mongolia Medical University, Hohhot, Inner Mongolia, China. <sup>3</sup>Tumor Molecular Diagnostic Laboratory, The Inner Mongolia Cancer Hospital, Hohhot, Inner Mongolia, China.

Received: 17 December 2018 Accepted: 21 May 2019

Published online: 06 June 2019

#### References

- Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66:115–32.
- Majoor BC, Boyce AM, Bovée JV, Smit VT, Collins MT, Cleton-Jansen AM, Dekkers OM, Hamdy NA, Dijkstra PS, Appelman-Dijkstra NM. Increased risk of breast cancer at a young age in women with fibrous dysplasia. *J Bone Miner Res.* 2018;33:84–90.
- Van't Veer LJ, Dai H, Van De Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, Van Der Kooy K, Marton MJ, Witteveen AT. Gene expression profiling predicts clinical outcome of breast cancer. *nature.* 2002;415:530.
- Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012;486:346.
- Ma X-J, Dahiya S, Richardson E, Erlander M, Sgroi DC. Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Res.* 2009;11:R7.
- Sørliie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, Van De Rijn M, Jeffrey SS. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci.* 2001;98:10869–74.
- Ponder BA. Cancer genetics. *Nature.* 2001;411:336.
- Walsh T, King M-C. Ten genes for inherited breast cancer. *Cancer Cell.* 2007;11:103–5.
- Welsh PL, King M-C. BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet.* 2001;10:705–13.
- Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, Averill D. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science.* 1994;265:2088–90.
- Kurian AW, Gong GD, John EM, Johnston DA, Felberg A, West DW, Miron A, Andrulis IL, Hopper JL, Knight JA. Breast cancer risk for noncarriers of family-specific BRCA1 and BRCA2 mutations: findings from the breast Cancer family registry. *J Clin Oncol.* 2011;29:4505.
- Bišof V, Perić Salihović M, Smolej Narančić N, Škaric-Juric T, Jakic-Razumovic J, Janicijević B, Turek S, Rudan P. TP53 gene polymorphisms and breast cancer in Croatian women: a pilot study. *Eur J Gynaecol Oncol.* 2010;31:539.
- Faghani M, Ghasemi F, Nikhbakht M, Salehi M. TP53 PIN3 polymorphism associated with breast cancer risk in Iranian women. *Indian J Cancer.* 2011;48:298.
- Chen BP, Uematsu N, Kobayashi J, Lerenthal Y, Krempler A, Yajima H, Löbrich M, Shiloh Y, Chen DJ. Ataxia telangiectasia mutated (ATM) is essential for DNA-PKcs phosphorylations at the Thr-2609 cluster upon DNA double strand break. *J Biol Chem.* 2007;282:6582–7.
- Shieh S-Y, Ahn J, Tamai K, Taya Y, Prives C. The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites. *Genes Dev.* 2000;14:289–300.
- Stracker TH, Carson CT, Weitzman MD. Adenovirus oncoproteins inactivate the Mre11–Rad50–NBS1 DNA repair complex. *Nature.* 2002;418:348.
- Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, Wahrer DC, Sgroi DC, Lane WS, Haber DA. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell.* 2001;105:149–60.
- Moffa AB, Tannheimer SL, Ethier SP. Transforming Potential of Alternatively Spliced Variants of Fibroblast Growth Factor Receptor 2 in Human Mammary Epithelial Cells11NIH grant 2RO1CA70354 and NIH through the University of Michigan's Cancer Center Support (grant 5 P30 CA46592). Note: SL Tannheimer is currently in Cytokine Biology, ZymoGenetics, Seattle, Washington. *Mol Cancer Res.* 2004;2:643–52.
- Martin JA, Wang Z. Next-generation transcriptome assembly. *Nat Rev Genet.* 2011;12:671.
- Raaphorst FM, Meijer CJ, Fieret E, Blokzijl T, Mommers E, Buerger H, Packeisen J, Sewalt RA, Ottet AP, Van Diest PJ. Poorly differentiated breast carcinoma is associated with increased expression of the human polycomb group EZH2 gene. *Neoplasia.* 2003;5:481–8.
- Hou J, Liu G, Yuan Y, Wang D, Jiao P, Xing L, Pan Y. Increased Jab1/COP55 is associated with therapeutic response and adverse outcome in lung cancer and breast cancer patients. *Oncotarget.* 2017;8:97504.
- Fang JY, Richardson BC. The MAPK signalling pathways and colorectal cancer. *Lancet Oncol.* 2005;6:322–7.
- Maire V, Baldeyron C, Richardson M, Tesson B, Vincent-Salomon A, Gravier E, Marty-Prouvost B, De Koning L, Rigault G, Dumont A. TTK/hMPS1 is an attractive therapeutic target for triple-negative breast cancer. *PLoS One.* 2013;8:e63712.
- Loussouarn D, Champion L, Leclair F, Campone M, Charbonnel C, Ricolleau G, Gouraud W, Bataille R, Jezequel P. Validation of UBE2C protein as a prognostic marker in node-positive breast cancer. *Br J Cancer.* 2009;101:166.
- Perou CM, Sørliie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature.* 2000;406(6797):747–52.

26. Drasar B, Irving D. Environmental factors and cancer of the colon and breast. *Br J Cancer*. 1973;27:167.
27. Spiegel D, Giese-Davis J. Depression and cancer: mechanisms and disease progression. *Biol Psychiatry*. 2003;54:269–82.
28. Hemler ME. Specific tetraspanin functions. *J Cell Biol*. 2001;155:1103–8.
29. Dunn CD, Sulis ML, Ferrando AA, Greenwald I. A conserved tetraspanin subfamily promotes notch signaling in *Caenorhabditis elegans* and in human cells. *Proc Natl Acad Sci*. 2010;107:5907–12.
30. Lu Z, Luo T, Nie M, Pang T, Zhang X, Shen X, Ma L, Bi J, Wei G, Fang G. TSPAN1 functions as an oncogene in gastric cancer and is downregulated by miR-573. *FEBS Lett*. 2015;589:1988–94.
31. Hou F-Q, Lei X-F, Yao J-L, Wang Y-J, Zhang W. Tetraspanin 1 is involved in survival, proliferation and carcinogenesis of pancreatic cancer. *Oncol Rep*. 2015;34:3068–76.
32. Hölters S, Anacker J, Jansen L, Beer-Grondke K, Dürst M, Rubio I. Tetraspanin 1 promotes invasiveness of cervical cancer cells. *Int J Oncol*. 2013;43:503–12.
33. Desouki MM, Liao S, Huang H, Conroy J, Nowak NJ, Shepherd L, Gaille DP, Geradts J. Identification of metastasis-associated breast cancer genes using a high-resolution whole genome profiling approach. *J Cancer Res Clin Oncol*. 2011;137:795–809.
34. Arencibia JM, Martín S, Pérez-Rodríguez FJ, Bonnín A. Gene expression profiling reveals overexpression of TSPAN13 in prostate cancer. *Int J Oncol*. 2009;34:457–63.
35. Minchenko D, Novik Y, Maslak H, Tiazhka O, Minchenko O. Expression of PFKFB, HK2, NAMPT, TSPAN13 and HSPB8 genes in pediatric glioma. *Lik Sprava.Likars' ka sprava*. 2015(7-8):43-8.
36. Matis LA, Palmer TD, Ashby WJ, Nashabi A, Chytil A, Aakre M, Pickup MW, Gorska AE, Zijlstra A, Moses HL. Lack of transforming growth factor- $\beta$  signaling promotes collective cancer cell invasion through tumor-stromal crosstalk. *Breast Cancer Res*. 2012;14:R98.
37. Mansouri N, Movafagh A, Sayad A, Heidary Pour A, Taheri M, Soleimani S, Mirzaei HR, Alizadeh Shargh S, Azargashb E, Bazmi H. Targeting of BUB1b gene expression in sentinel lymph node biopsies of invasive breast cancer in Iranian female patients. *Asian Pac J Cancer Prev*. 2016;17:317–21.
38. Khan S, Brougham CL, Ryan J, Sahrudin A, O'Neill G, Wall D, Curran C, Newell J, Kerin MJ, Dwyer RM. miR-379 regulates cyclin B1 expression and is decreased in breast cancer. *PLoS One*. 2013;8:e68753.
39. Brodsky AS, Xiong J, Yang D, Schorl C, Fenton MA, Graves TA, Sikov WM, Resnick MB, Wang Y. Identification of stromal ColXa1 and tumor-infiltrating lymphocytes as putative predictive markers of neoadjuvant therapy in estrogen receptor-positive/HER2-positive breast cancer. *BMC Cancer*. 2016;16:274.
40. Kadam CY, Abhang SA. Serum levels of soluble Fas ligand, granzyme B and cytochrome c during adjuvant chemotherapy of breast cancer. *Clin Chim Acta*. 2015;438:98–102.
41. Kulkarni G, Turbin DA, Amiri A, Jeganathan S, Andrade-Navarro MA, Wu TD, Huntsman DG, Lee JM. Expression of protein elongation factor eEF1A2 predicts favorable outcome in breast cancer. *Breast Cancer Res Treat*. 2007;102:31–41.
42. Dahl E, Kristiansen G, Gottlob K, Klamann I, Ebner E, Hinzmann B, Hermann K, Pilarsky C, Dürst M, Klinkhammer M, Schalke molecular profiling of laser-microdissected matched tumor and normal breast tissue identifies karyopherin  $\alpha 2$  as a potential novel prognostic marker in breast cancer. *Clin Cancer Res*. 2006;12:3950–60.
43. Grizzi F, Di Biccari S, Fiamengo B, Štifter S, Colombo P. Pituitary tumor-transforming gene 1 is expressed in primary ductal breast carcinoma, lymph node infiltration, and distant metastases. *Dis Markers*. 2013;35:267–72.
44. Kutzsch M, Sieuwerts AM, Gresser M, Meyre A, Fuessel S, Meijer-van Gelder ME, Smid M, Schmitt M, Baretton G, Luther T. Urokinase receptor splice variant uPAR-del4/5-associated gene expression in breast cancer: identification of rab31 as an independent prognostic factor. *Breast Cancer Res Treat*. 2008;111:229–40.
45. Ma X-J, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, Payette T, Pistone M, Stecker K, Zhang BM. Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci*. 2003;100:5974–9.
46. Capecchi MR, Pozner A. ASPM regulates symmetric stem cell division by tuning cyclin E ubiquitination. *Nat Commun*. 2015;6:8763.
47. Eliyahu E, Shtraizent N, He X, Chen D, Shalgi R, Schuchman EH. Identification of cystatin SA as a novel inhibitor of acid ceramidase. *J Biol Chem*. 2011. <https://doi.org/10.1074/jbc.M111.260372>.
48. Taylor C, Loomans HA, Le Bras GF, Koumangoye RB, Romero-Morales AI, Quast LL, Zaika AI, El-Rifai W, Andl T, Andl CD. Activin signaling regulates cell invasion and proliferation in esophageal adenocarcinoma. *Oncotarget*. 2015;6:34228.
49. Fu H, Shao L. Silencing of NUF2 inhibits proliferation of human osteosarcoma Saos-2 cells. *Eur Rev Med Pharmacol Sci*. 2016;20:1071–9.
50. Gavin EJ, Song B, Wang Y, Xi Y, Ju J. Reduction of Orc6 expression sensitizes human colon cancer cells to 5-fluorouracil and cisplatin. *PLoS One*. 2008;3:e4054.
51. Ueki T, Park J-H, Nishidate T, Kijima K, Hirata K, Nakamura Y, Katagiri T. Ubiquitination and downregulation of BRCA1 by ubiquitin-conjugating enzyme E2T overexpression in human breast cancer cells. *Cancer Res*. 2009;69(22):8752–60.
52. Liu L, Wu J, Wang S, Luo X, Du Y, Huang D, Gu D, Zhang F. PKMYT1 promoted the growth and motility of hepatocellular carcinoma cells by activating beta-catenin/TCF signaling. *Exp Cell Res*. 2017;358:209–16.
53. Støy J, Kampmann U, Mengel A, Magnusson NE, Jessen N, Grarup N, Rungby J, Stødkilde-Jørgensen H, Brandslund I, Christensen C. Reduced CD300LG mRNA tissue expression, increased intramyocellular lipid content and impaired glucose metabolism in healthy male carriers of Arg82Cys in CD300LG: a novel genomemetic cross-link between CD300LG and common metabolic phenotypes. *BMJ Open Diabetes Res Care*. 2015;3:e000095.
54. Stadtmueller BM, Huey-Tubman KE, López CJ, Yang Z, Hubbell WL, Bjorkman PJ. The structure and dynamics of secretory component and its interactions with polymeric immunoglobulins. *Elife*. 2016;5:e10640.
55. Rawlings ND, Barrett AJ. Evolution of proteins of the cystatin superfamily. *J Mol Evol*. 1990;30:60–71.
56. Li H, Takai N, Yuge A, et al. RETRACTED: Novel target genes responsive to the anti-growth activity of triptolide in endometrial and ovarian cancer cells. *Cancer Letters*. 2013;2(338):331.
57. Lubec G, Nonaka M, Krapfenbauer K, et al. Expression of the dihydropyrimidinase related protein 2 (DRP-2) in Down syndrome and Alzheimer's disease brain is downregulated at the mRNA and dysregulated at the protein level[M]//The Molecular Biology of Down Syndrome. Vienna: Springer; 1999. p. 161-77.
58. Devaraj S, Natarajan J. miRNA-mRNA network detects hub mRNAs and cancer specific miRNAs in lung cancer. *In silico biology*. 2011;11:281–95.
59. Pooladi M, Rezaei-Tavirani M, Hashemi M, Hesami-Tackallou S, Khaghani-Razi-Abad S, Moradi A, Zali AR, Mousavi M, Rakhshan A, Firozi-dalvand L. The study of "Dihydropyrimidinase related proteins (DRPs)" expression changes influence in malignant astrocytoma brain tumor. *Iran J Cancer Prev*. 2014;7:130.

60. Castellana B, Escuin D, Peiró G, García-Valdecasas B, Vázquez T, Pons C, Pérez-Olabarria M, Barnadas A, Lerma E. ASPN and GJB2 are implicated in the mechanisms of invasion of ductal breast carcinomas. *J Cancer*. 2012;3:175–83. <https://doi.org/10.7150/jca.4120>.
61. Zhu T, Gao YF, Chen YX, Wang ZB, Yin JY, Mao XY, Li X, Zhang W, Zhou HH, Liu ZQ. Genome-scale analysis identifies GJB2 and ERO1LB as prognosis markers in patients with pancreatic cancer. *Oncotarget*. 2017;8:21281–9. <https://doi.org/10.18632/oncotarget.15068>.
62. Luo C, Yao Y, Yu Z, Zhou H, Guo L, Zhang J, Cao H, Zhang G, Li Y, Jiao Z. UBE2T knockdown inhibits gastric cancer progression. *Oncotarget*. 2017;8:32639–54. <https://doi.org/10.18632/oncotarget.15947>.
63. Wen M, Kwon Y, Wang Y, Mao JH, Wei G. Elevated expression of UBE2T exhibits oncogenic properties in human prostate cancer. *Oncotarget*. 2015;6:25226–39. <https://doi.org/10.18632/oncotarget.4712>.
64. Hu P, Shanguan J, Zhang L. Downregulation of NUF2 inhibits tumor growth and induces apoptosis by regulating lncRNA AF339813. *Int J Clin Exp Pathol*. 2015;8:2638–48.
65. Koleck TA, Bender CM, Clark BZ, Ryan CM, Ghotkar P, Brufsky A, McAuliffe PF, Rastogi P, Sereika SM, Conley YP. An exploratory study of host polymorphisms in genes that clinically characterize breast cancer tumors and pretreatment cognitive performance in breast cancer survivors. *Breast Cancer (Dove Med Press)*. 2017;9:95–110. <https://doi.org/10.2147/bctt.S123785>.
66. Lebrero-Fernández C, Wenzel UA, Akeus P, Wang Y, Strid H, Simrén M, Gustavsson B, Börjesson LG, Cardell SL, Öhman L, Quiding-Järbrink M, Bas-Forsberg A. Altered expression of Butyrophilin (BTN) and BTN-like (BTNL) genes in intestinal inflammation and colon cancer. *Immun Inflamm Dis*. 2016;4:191–200. <https://doi.org/10.1002/iid3.105>.
67. Blazquez JL, Benyamine A, Pasero C, Olive D. New insights into the regulation of  $\gamma\delta$  T cells by BTN3A and other BTN/BTNL in tumor immunity. *Front Immunol*. 2018;9:1601. <https://doi.org/10.3389/fimmu.2018.01601>.
68. Berntsson J, Lundgren S, Nodin B, Uhlén M, Gaber A, Jirstrom K. Expression and prognostic significance of the polymeric immunoglobulin receptor in epithelial ovarian cancer. *J Ovarian Res*. 2014;7:26. <https://doi.org/10.1186/1757-2215-7-26>.
69. Zhang J, Tsoi H, Li X, Wang H, Gao J, Wang K, Go MY, Ng SC, Chan FK, Sung JJ. Carbonic anhydrase IV inhibits colon cancer development by inhibiting the Wnt signalling pathway through targeting the WTAP–WT1–TBL1 axis. *Gut*. 2016; 65:1482–93.
70. Kinlaw WB, Baures PW, Lupien LE, Davis WL, Kuemmerle NB. Fatty acids and breast Cancer: make them on site or have them delivered. *J Cell Physiol*. 2016;231:2128–41. <https://doi.org/10.1002/jcp.25332>.
71. Andersen MK, Rise K, Giskeødegård GF, Richardsen E, Bertilsson H, Størkersen Ø, Bathen TF, Rye M, Tessem M-B. Integrative metabolic and transcriptomic profiling of prostate cancer tissue containing reactive stroma. *Sci Rep*. 2018;8: 14269.
72. Yan P, He Y, Xie K, Kong S, Zhao W. In silico analyses for potential key genes associated with gastric cancer. *PeerJ*. 2018; 6:e6092.
73. Rahbari NN, Kedrin D, Incio J, Liu H, Ho WW, Nia HT, Edrich CM, Jung K, Daubriac J, Chen Anti I. VEGF therapy induces ECM remodeling and mechanical barriers to therapy in colorectal cancer liver metastases. *Sci Transl Med*. 2016;8: 360ra135.
74. Cui X, Morales R-TT, Qian W, Wang H, Gagner J-P, Dolgalev I, Placantonakis D, Zagzag D, Cimmino L, Snuderl M. Hacking macrophage-associated immunosuppression for regulating glioblastoma angiogenesis. *Biomaterials*. 2018;161:164–78.
75. Alahari SK. Nischarin inhibits Rac induced migration and invasion of epithelial cells by affecting signaling cascades involving PAK. *Exp Cell Res*. 2003;288:415–24.
76. Maziveyi M, Alahari SK. Breast cancer tumor suppressors: a special emphasis on novel protein nischarin. *Cancer Res*. 2015;75:4252–9.
77. Maziveyi M, Dong S, Baranwal S, Alahari SK. Nischarin regulates focal adhesion and Invadopodia formation in breast cancer cells. *Mol Cancer*. 2018;17:21.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

