

# Bioinformatics analysis of gene expression profiles to identify causal genes in luminal B2 breast cancer

JINGUANG WANG<sup>1</sup>, QI DU<sup>2</sup> and CHEN LI<sup>3</sup>

<sup>1</sup>Department of Thyroid and Breast Surgery, Linyi People's Hospital, Linyi, Shandong 272100;

<sup>2</sup>Department of Prevention and Health Care, Lanshan District Center for Disease Control and Prevention, Linyi, Shandong 276000; <sup>3</sup>Department of General Surgery, Linyi People's Hospital, Linyi, Shandong 272100, P.R. China

Received February 26, 2016; Accepted August 3, 2017

DOI: 10.3892/ol.2017.7256

**Abstract.** Patients with the luminal B subtype of breast cancer exhibit a poor prognosis, high metastatic risk and high incidence of chemotherapy resistance. Luminal B breast cancer is sub-classified into B1 and B2. The pathophysiological mechanism of luminal B2 breast cancer (LB2BC) progression has yet to be characterized. Therefore, the present study aimed to identify the genes involved in the pathogenesis of LB2BC. The data of 117 LB2BC expression profiles were downloaded from The Cancer Genome Atlas (TCGA) and differentially expressed genes (DEGs) were identified by comparison with non-tumor tissue expression profiles. Gene Ontology enrichment analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and protein-protein interaction (PPI) networks were used to obtain insight into the functions of DEGs. Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis was performed to validate the expression level of DEGs in tissue samples. A total of 2,251 DEGs, including 759 upregulated and 1,492 downregulated genes, were identified between LB2BC and non-tumor tissues. The top 15 upregulated and downregulated genes were used to construct a PPI network: Epidermal growth factor receptor (EGFR), fibronectin-1 (FN1) and Polo-like kinase-1 had the highest connectivity degrees. KEGG analysis identified that DEGs were most significantly enriched in 'focal adhesion', 'pathways in cancer' and 'ECM-receptor interaction' pathways. The results of RT-qPCR demonstrated that EGFR was significantly downregulated in LB2BC, whereas FN1 was significantly upregulated, whereas neurotrophic receptor tyrosine kinase 2 (NTRK2) trended towards downregulation. In conclusion, the DEGs identified in the present study,

including NTRK2, FN1 and EGFR, may serve pivotal roles in the tumorigenesis of LB2BC by affecting the 'focal adhesion', 'pathways in cancer' and 'ECM-receptor interaction' KEGG pathways.

## Introduction

Breast cancer, a complex and heterogeneous malignancy, is the most frequently diagnosed cancer and the leading cause of cancer mortality among females worldwide; there were estimated to be 1.7 million cases and 521,900 deaths in 2012 (1). The incidence rates of breast cancer in economically developed regions, including North America, Australia, New Zealand, and Northern and Western Europe, are higher than those in Africa and Asia (1).

Breast cancer is classified into 4 subtypes: Luminal A, luminal B, human epidermal growth factor receptor 2 (HER2)-positive and basal-like, according to clinically and biologically relevant molecular features that were previously identified by DNA microarray-based expression profiling (2-4). These subtypes are associated with distinct pathological features, treatment responses and clinical outcomes (5). Basal-like breast cancer is predominantly triple-negative (i.e. no expression of estrogen receptor (ER), progesterone receptor (PR) or HER2). In HER2-positive breast cancer, the HER2 gene is overexpressed. The basal-like and HER2-positive subtypes are associated with a poor prognosis, and the luminal A subtype was identified as having the more favorable clinical outcome (4).

The clinical and biological behavior of the luminal B subtype is more aggressive than the luminal A subtype, with a higher metastatic risk and greater resistance to hormone therapy and conventional chemotherapy; patients with luminal B cancer exhibit increased relapse rates in the first 5 years following diagnosis (6). Luminal B-subtype cancer cases have a higher recurrence score than luminal A, based on OncotypeDX detection: In a cohort of 831 untreated lymph node metastasis-negative patients, the hazard ratio (HR) for early metastasis (<5 years) was 2.86 for luminal B relative to luminal A (6,7).

As a subtype of Luminal B breast cancer, luminal B2 breast cancer (LB2BC) is usually associated with poor prognosis, and the pathophysiological mechanism underlying luminal B2 breast cancer remains unknown. In the present study,

---

*Correspondence to:* Dr Chen Li, Department of General Surgery, Linyi People's Hospital, 27 Jiefang Road, Linyi, Shandong 272100, P.R. China

E-mail: lychen\_li@sina.com

**Key words:** luminal B2, breast cancer, differentially expressed genes, protein-protein interaction network, reverse transcription-quantitative polymerase chain reaction

bioinformatics methods were used to identify differentially expressed genes (DEGs) between LB2BC and non-tumor tissues, with the aim of providing valuable information for further pathogenesis mechanism elucidation of LB2BC.

### Subjects and methods

**TGCA data.** All 1,098 patients (as included in the database on the 27th of October, 2015) with breast cancer were retrieved from The Cancer Genome Atlas (TGCA) data portal (8). A total of 117 LB2BC expression profiles from this dataset were included in the present study, according to the inclusion and exclusion criteria below. Level 3 raw expression data of mRNA from the Illumina HiSeq\_miRNAseq V2 platform for the 117 LB2BC profiles and 18 normal controls were downloaded from TCGA data portal.

**Exclusion criteria.** The exclusion criteria for patients with LB2BC were as follows: i) History of any other malignancy; ii) history of previous treatment, including chemotherapy, radiotherapy or endocrinotherapy; iii) samples without mRNA sequence data.

**Inclusion criteria.** The inclusion criteria were as follows: i) HER2 status, as assessed by immunohistochemistry (IHC), was positive; ii) ER expression, PR expression, or both, were positive.

**Screening of differentially expressed genes.** In R (version 3.2.3), the DESeq (version 1.28.0) package was used for reads count analysis (9). The limma package (version 1.9.6) (10) was used to screen for DEGs by comparing LB2BC expression profiles to control profiles. The raw P-value was adjusted to a false discovery rate (FDR) value using the Benjamin and Hochberg method. Genes within the cut-off criteria of  $FDR < 0.001$  were designated as DEGs.

**Functional annotation of DEGs.** To identify the biological function and associated metabolic pathways for the DEGs, they were submitted to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for pathway enrichment analysis and GOrrilla online software (11) for gene ontology (GO) term enrichment analysis (12,13).  $P < 0.001$  and  $FDR < 0.05$  were set as the cut-offs for selecting significantly enriched functional GO terms and KEGG pathway, respectively.

**Construction of protein-protein interaction (PPI) network.** A PPI network can organize genes into a network to aid the understanding of their biological function (13). The BioGRID database (14) includes verified and predicted protein interactions. In the present study, DEGs were mapped into the BioGRID database to screen for interacting protein pairs (15). A PPI network of the top 15 upregulated and downregulated DEGs was constructed and visualized using Cytoscape software (version 3.3.0) (16).

**Reverse transcription-quantitative polymerase chain reaction (RT-qPCR).** A total of five tumor tissues and five paired adjacent normal tissues were obtained from five female patients with breast cancer undergoing routine surgical procedures

in Linyi People's Hospital (Linyo, China) during February to August 2015, with a mean age of 45.5 years. The tissues were frozen immediately after surgery in liquid nitrogen and were stored at  $-80^{\circ}\text{C}$  until RNA extraction. The present study was approved by the Ethics Committee of Linyi People's Hospital. Written informed consent was obtained from all patients prior to enrollment in the present study. Total RNA from tissue samples was extracted using TRIzol reagent (Invitrogen; Thermo Fisher Scientific, Inc., Waltham, MA, USA) according to the manufacturer's protocol. The SuperScript III Reverse Transcription kit (Invitrogen; Thermo Fisher Scientific, Inc.) was used for cDNA synthesis according to the manufacturer's protocol. qPCR was used to quantify differences in the mRNA expression of associated genes using SYBRGreen PCR reagent (Applied Biosystems; Thermo Fisher Scientific, Inc.) according to the manufacturer's protocol. PCR reactions were performed in triplicate and run under the following conditions: 1 cycle of  $95^{\circ}\text{C}$  for 10 min, followed by 45 cycles of  $95^{\circ}\text{C}$  for 15 sec and  $60^{\circ}\text{C}$  for 60 sec. The relative expression level was calculated using the  $2^{-\Delta\Delta C_q}$  method (17), normalized to  $\beta$ -actin. Each reaction was performed  $\geq 3$  times.

The primer sequences were as follows: Neurotrophic receptor tyrosine kinase 2 (NTRK2) forward, 5'-ATCTCCAACCTCAGACCACCACT-3' and reverse, 5'-AATCTGTTTCTCATCTTCCCATAC-3'; fibronectin-1 (FN1) forward, 5'-CAACCTACGGATGACTCGTGCTT-3' and reverse, 5'-TTTCTCCCTGACGGTCCCACTT-3'; Polo-like kinase 1 (PLK1) forward, 5'-GGCAGCGTGCAGATCAACTT-3' and reverse, 5'-CAGGAGACTCAGGCGGTATGT-3'; epidermal growth factor receptor (EGFR) forward, 5'-GTGACCGTTTGGGAGTTGATGA-3' and reverse, 5'-GGCTGAGGGAGGCGTTTCTC-3'; and  $\beta$ -actin forward, 5'-CTGAAGTACCCCATCGAGCAC-3' and reverse, 5'-ATAGCACAGCCTGGATAGCAAC-3'.

**Statistical analysis.** RT-qPCR experimental data were expressed as the mean  $\pm$  standard deviation. GraphPad Prism (version 6.0) software (GraphPad Software, Inc., La Jolla, CA, USA) was used to perform statistical analysis. Statistically significant differences between two groups were evaluated using an unpaired Student's t-test.  $P < 0.05$  was considered to indicate a statistically significant difference.

### Results

**DEGs in LB2BC.** A total of 2,251 significantly DEGs were identified in LB2BC tissues compared with non-tumor controls. The DEGs included 759 upregulated and 1,492 downregulated genes. Matrix metalloproteinase-11 was the most significantly upregulated gene; NTRK2 was the most significantly downregulated gene. The top 15 significantly upregulated and downregulated genes are listed in Table I. The pattern of expression change for the top 200 DEGs is presented in Fig. 1.

**GO annotation of DEGs.** To obtain insights into their potential biological roles, GO term enrichment analysis was performed on the DEGs from LB2BC. 'Mitotic cell cycle process' (GO, 1903047;  $P = 1.22 \times 10^{-8}$ ), 'cell cycle G2/M phase transition' (GO, 0044839;  $P = 7.43 \times 10^{-7}$ ) and 'G2/M transition of mitotic cell cycle' (GO, 0000086;  $P = 7.43 \times 10^{-7}$ ) were the most significantly enriched biological process GO terms; 'vinculin binding'

Table I. Differentially expressed genes in luminal B2 breast cancer.

A, Top 15 upregulated genes

Gene symbol	Gene ID	P-value	FDR
MMP11	4320	$1.23 \times 10^{-32}$	$7.06 \times 10^{-30}$
COL10A1	1300	$1.14 \times 10^{-31}$	$6.02 \times 10^{-29}$
NEK2	4751	$2.03 \times 10^{-22}$	$4.39 \times 10^{-20}$
COL11A1	1301	$1.07 \times 10^{-21}$	$2.02 \times 10^{-19}$
PPAPDC1A	196051	$1.04 \times 10^{-20}$	$1.79 \times 10^{-18}$
INHBA	3624	$7.51 \times 10^{-20}$	$1.17 \times 10^{-17}$
KIF4A	24137	$1.70 \times 10^{-19}$	$2.54 \times 10^{-17}$
UBE2C	11065	$3.61 \times 10^{-19}$	$5.29 \times 10^{-17}$
WISP1	8840	$7.36 \times 10^{-19}$	$1.04 \times 10^{-16}$
FN1	2335	$7.50 \times 10^{-19}$	$1.05 \times 10^{-16}$
IBSP	3381	$2.99 \times 10^{-18}$	$3.92 \times 10^{-16}$
PLK1	5347	$4.13 \times 10^{-18}$	$5.25 \times 10^{-16}$
TPX2	22974	$5.67 \times 10^{-18}$	$6.98 \times 10^{-16}$
IQGAP3	128239	$1.12 \times 10^{-17}$	$1.31 \times 10^{-15}$
HSD17B6	8630	$1.65 \times 10^{-17}$	$1.87 \times 10^{-15}$

B, Top 15 downregulated genes

Gene symbol	Gene ID	P-value	FDR
DMD	1756	$8.19 \times 10^{-72}$	$8.22 \times 10^{-68}$
SYNM	23336	$2.04 \times 10^{-68}$	$1.37 \times 10^{-64}$
FMO2	2327	$3.72 \times 10^{-63}$	$1.87 \times 10^{-59}$
SFRP1	6422	$4.64 \times 10^{-57}$	$1.86 \times 10^{-53}$
MYH11	4629	$1.16 \times 10^{-56}$	$3.87 \times 10^{-53}$
COL17A1	1308	$3.89 \times 10^{-54}$	$1.11 \times 10^{-50}$
FAT2	2196	$2.20 \times 10^{-53}$	$5.52 \times 10^{-50}$
EGFR	1956	$3.05 \times 10^{-52}$	$6.81 \times 10^{-49}$
CX3CL1	6376	$2.70 \times 10^{-50}$	$5.42 \times 10^{-47}$
SAMD5	389432	$3.28 \times 10^{-47}$	$5.98 \times 10^{-44}$
NDRG2	57447	$7.56 \times 10^{-47}$	$1.26 \times 10^{-43}$
GRIA4	2893	$1.85 \times 10^{-46}$	$2.85 \times 10^{-43}$
EPHB1	2047	$6.87 \times 10^{-45}$	$9.85 \times 10^{-42}$
MAMDC2	256691	$2.88 \times 10^{-44}$	$3.86 \times 10^{-41}$

FDR, false discovery rate.

(GO, 0017166;  $P=5.02 \times 10^{-6}$ ), 'structural constituent of muscle' (GO, 0008307;  $P=1.53 \times 10^{-5}$ ) and 'potassium channel regulator activity' (GO, 0015459;  $P=2.77 \times 10^{-5}$ ) were the most significantly enriched molecular function GO terms; 'condensed chromosome outer kinetochore' (GO, 0000940;  $P=2.34 \times 10^{-5}$ ), 'extracellular matrix' (GO, 0031012;  $P=4.08 \times 10^{-5}$ ) and 'microtubule spindle' (GO, 0005876;  $P=8.77 \times 10^{-5}$ ) were the most significantly enriched cellular component GO terms (Table II).

**Pathway enrichment analysis.** KEGG pathway enrichment analysis was performed on the DEGs from LB2BC.  $FDR < 0.05$  was used as the cut-off for pathway significance.

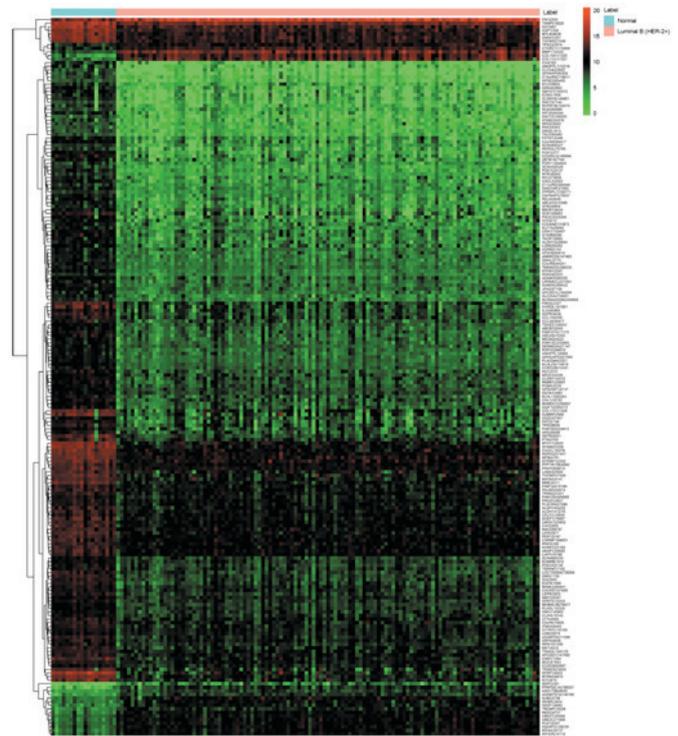


Figure 1. Heat map visualization of the patterns of expression change for the top 200 most significantly differentially expressed genes between luminal B2 breast cancer and normal controls. Red and green squares represent upregulated and downregulated genes, respectively.

The most significantly enriched pathways were 'focal adhesion' ( $FDR=1.02 \times 10^{-20}$ ), 'pathways in cancer' ( $FDR=1.50 \times 10^{-20}$ ), 'ECM-receptor interactions' ( $FDR=1.15 \times 10^{-14}$ ) and 'cytokine-cytokine receptor interaction' ( $FDR=1.50 \times 10^{-11}$ ; Table III).

**Protein-protein interaction (PPI) network construction.** PPI networks for the top 15 upregulated and downregulated DEGs were constructed using Cytoscape software. The network consisted of 1,681 nodes and 1,898 edges. The hub proteins were EGFR, FN1 and PLK1 (Fig. 2).

**RT-qPCR verification of DEGs.** The expression level of four of the identified DEGs (NTRK2, FN1, PLK1 and EGFR) were validated with RT-qPCR in five LB2BC tumor and adjacent tissue samples. The expression of NTRK2 did not differ significantly between LB2BC tumor and adjacent tissues, but trended towards downregulation in LB2BC (Fig. 3A). The expression of FN1 was significantly downregulated in LB2BC ( $P < 0.05$ ; Fig. 3B). PLK1 expression did not differ significantly between LB2BC and adjacent tumor tissues (Fig. 3C). The expression of EGFR was significantly downregulated in LB2BC compared with adjacent tissue ( $P < 0.001$  Fig. 3D).

## Discussion

NTRK2 was the most significantly downregulated DEG identified in LB2BC (Table I) and was associated with the KEGG term 'MAPK signaling pathway' (Table III). The expression of NTRK2 did not significantly differ between LB2BC and non-cancerous tissue, although it trended towards

Table II. Top enriched GO terms of differentially expressed genes in luminal B2 breast cancer.

A, Biological process			
GO ID	GO term	Genes, n	P-value
GO:1903047	Mitotic cell cycle process	75	1.22x10 <sup>-8</sup>
GO:0044839	Cell cycle G2/M phase transition	17	7.43x10 <sup>-7</sup>
GO:0000086	G2/M transition of mitotic cell cycle	17	7.43x10 <sup>-7</sup>
GO:0051302	Regulation of cell division	51	8.20x10 <sup>-7</sup>
GO:0022610	Biological adhesion	119	3.30x10 <sup>-6</sup>
GO:0032501	Multicellular organismal process	290	3.67x10 <sup>-6</sup>
GO:0044707	Single-multicellular organism process	288	3.91x10 <sup>-6</sup>
GO:0007155	Cell adhesion	118	4.79x10 <sup>-6</sup>
GO:0022402	Cell cycle process	88	7.13x10 <sup>-6</sup>
GO:0007346	Regulation of mitotic cell cycle	44	1.12x10 <sup>-5</sup>
GO:0032502	Developmental process	393	1.25x10 <sup>-5</sup>
GO:0030198	Extracellular matrix organization	67	1.64x10 <sup>-5</sup>
GO:0043062	Extracellular structure organization	67	1.64x10 <sup>-5</sup>
GO:0044699	Single-organism process	731	2.09x10 <sup>-5</sup>
GO:0044767	Single-organism developmental process	339	2.57x10 <sup>-5</sup>
B, Molecular function			
GO ID	GO term	Genes, n	P-value
GO:0017166	Vinculin binding	2	5.02x10 <sup>-6</sup>
GO:0008307	Structural constituent of muscle	3	1.53x10 <sup>-5</sup>
GO:0015459	Potassium channel regulator activity	13	2.77x10 <sup>-5</sup>
GO:0005102	Receptor binding	52	5.78x10 <sup>-5</sup>
GO:0005539	Glycosaminoglycan binding	41	2.09x10 <sup>-4</sup>
GO:0005030	Neurotrophin receptor activity	1	4.69x10 <sup>-4</sup>
GO:0060175	Brain-derived neurotrophic factor-activated receptor activity	1	4.69x10 <sup>-4</sup>
GO:0008017	Microtubule binding	28	5.28x10 <sup>-4</sup>
GO:0043394	Proteoglycan binding	7	7.72x10 <sup>-4</sup>
GO:0035173	Histone kinase activity	5	8.73x10 <sup>-4</sup>
C, Cellular component			
GO ID	GO term	Genes, n	P-value
GO:0000940	Condensed chromosome outer kinetochore	9	2.34x10 <sup>-5</sup>
GO:0031012	Extracellular matrix	64	4.08x10 <sup>-5</sup>
GO:0005876	Spindle microtubule	14	8.77x10 <sup>-5</sup>
GO:0005819	Spindle	25	2.64x10 <sup>-4</sup>
GO:0030056	Hemidesmosome	2	4.12x10 <sup>-4</sup>
GO:0016328	Lateral plasma membrane	6	4.77x10 <sup>-4</sup>
GO:0043034	Costamere	2	5.70x10 <sup>-4</sup>
GO:0044449	Contractile fiber part	7	6.02x10 <sup>-4</sup>
GO:0044421	Extracellular region part	22	6.18x10 <sup>-4</sup>
GO:0005578	Proteinaceous extracellular matrix	47	7.15x10 <sup>-4</sup>
GO:0000922	Spindle pole	19	7.55x10 <sup>-4</sup>
GO:0016013	Syntrophin complex	1	9.39x10 <sup>-4</sup>
GO, gene ontology.			

Table III. Top-15 enriched KEGG terms of differentially expressed genes in luminal B2 breast cancer.

KEGG ID	KEGG term	Count	FDR	Genes
hsa04510	Focal adhesion	57	1.02x10 <sup>-20</sup>	FLT4, CAV2, MET, COMP, LAMA3, RELN, ITGA6, TLN2, ERBB2, MYLK, IGF1, LAMA2, IBSP, ROCK2, PDGFD, CCND2, PIK3R2, COL6A6, PDGFRA, KDR, FLNC, PDGFA, AKT3, PAK4, MAPK10, PRKCA, TNXB, LAMC1, VWF, COL5A2, CAV1, ARHGAP5, LAMB3, FIGF, LAMA4, EGFR, COL3A1, COL1A2, COL1A1, PIK3R1, LAMA1, JUN, ITGA9, THBS2, LAMC3, PPP1CA, LAMC2, PAK3, COL5A1, PAK7, ITGB8, MYL9, COL11A1, ITGA11, FN1, VAV3, ITGA7
hsa05200	Pathways in cancer	74	1.50x10 <sup>-20</sup>	CDK6, CCNE2, PTCH1, FGF9, TCF7L1, MET, LEF1, STAT1, LAMA3, FOS, RARB, FGF1, WNT7B, FGF2, ITGA6, KIT, BIRC5, TGFA, FOXO1, EPAS1, ERBB2, IGF1, RUNX1T1, FZD7, LAMA2, BMP2, PLCG2, CCNE1, TGFBR2, FGFR3, PIK3R2, E2F1, RAD51, TPM3, E2F2, WNT2, PDGFRA, PDGFA, AKT3, TCF7, FZD5, MAPK10, TCEB1, PRKCA, FADD, LAMC1, LAMB3, FGFRI, PTCH2, FIGF, MECOM, LAMA4, EGFR, PIK3R1, FZD4, LAMA1, MYC, JUN, LAMC3, LAMC2, TCF7L2, FGF7, ZBTB16, RET, CEBPA, PLD1, PPARG, DAPK2, STAT5B, TCEB2, STAT5A, FN1, PTGS2, MMP1
hsa04512	ECM-receptor interaction	31	1.15x10 <sup>-14</sup>	SDC1, COMP, LAMA3, RELN, ITGA6, LAMA2, IBSP, HMMR, COL6A6, SV2B, TNXB, LAMC1, VWF, COL5A2, LAMB3, LAMA4, COL3A1, COL1A2, COL1A1, LAMA1, ITGA9, THBS2, LAMC3, LAMC2, COL5A1, CD36, ITGB8, COL11A1, ITGA11, FN1, ITGA7
hsa04060	Cytokine-cytokine receptor interaction	51	1.50x10 <sup>-11</sup>	INHBB, FLT4, MET, TNFRSF10D, KIT, CXCL9, LEPR, BMP2, CXCL2, CNTFR, TGFBR2, CXCL10, IL28RA, CXCR7, TNFRSF21, CXCL3, CX3CR1, CCR8, LIFR, EDA, IL22RA2, CXCR2, PDGFRA, KDR, IL6R, PDGFA, IL21R, EDAR, EDA2R, CSF1, CX3CL1, FIGF, EGFR, CCL13, RELT, IFNGR1, LEP, CSF3, INHBA, NGFR, GHR, CXCL11, IL11RA, CXCL12, CCL21, TNFRSF18, CCL14, TNFSF4, TPO, CCL28, TSLP
hsa04360	Axon guidance	32	4.70x10 <sup>-10</sup>	SLIT3, GNAI1, MET, ROBO3, EFNA4, SEMA5A, DPYSL2, EPHB6, SEMA7A, SEMA6D, ROCK2, CFL2, EFNB3, SEMA3G, ABLIM3, EPHB1, EFNB1, SLIT2, PAK4, ABLIM1, UNC5B, SEMA3D, NFATC2, SEMA3F, NFAT5, PAK3, CFL1, PAK7, SEMA3A, EPHA2, CXCL12, RND1
hsa04110	Cell cycle	30	3.32x10 <sup>-9</sup>	CDK6, CCNE2, ESPL1, MCM4, CCNB2, CDC20, CCNA2, PTTG1, CCNE1, CCND2, E2F1, CDC25C, E2F2, CCNB1, MAD2L1, CDKN2C, MYC, PLK1, BUB1B, ANAPC11, CDC45, CDC25A, PKMYT1, TTK, BUB1, CDC14B, CDKN1C, PCNA, MCM2, CDK1
hsa04810	Regulation of actin cytoskeleton	39	2.69x10 <sup>-8</sup>	MRAS, DIAPH3, ARHGFEF6, FGF9, FGF1, FGF2, ITGA6, ARPC1B, MYLK, EZR, ROCK2, IQGAP3, PDGFD, CFL2, FGFR3, PIK3R2, PDGFRA, PDGFA, PAK4, GSN, DIAPH2, FGFRI, EGFR, PIK3R1, ITGA9, RRA52, PPP1CA, WASF2, FGF7, PAK3, CFL1, ARHGFEF4, PAK7, ITGB8, MYL9, ITGA11, FN1, VAV3, ITGA7

Table III. Continued.

KEGG ID	KEGG term	Count	FDR	Genes
hsa04080	Neuroactive ligand- receptor interaction	45	4.29x10 <sup>-8</sup>	GRIA3, DRD2, AVPR1A, ADRB2, S1PR1, GABRE, HTR2A, CNR1, EDNRB, LEPR, HTR1D, GLP2R, PTGFR, NMUR1, SSTRI, VIPRI, ADRA2A, GABRD, ADCYAP1R1, CTSG, LHCGR, OXTR, GRIN2D, PTGER4, GABRP, NR3C1, PTGER3, FPR3, P2RX6, LEP, GRIN2A, GRIA4, GHR, NPY2R, AVPR2, THRB, CRHR1, GRM4, GRIK1, PTHIR, P2RY13, TACRI, ADRA1A, LPAR6, P2RY14
hsa05215	Prostate cancer	23	7.92x10 <sup>-8</sup>	CCNE2, TCF7L1, LEF1, TGFA, FOXO1, ERBB2, CREB3L1, IGF1, CCNE1, PDGFD, PIK3R2, E2F1, E2F2, PDGFRA, PDGFA, CREB3L4, AKT3, TCF7, CREB5, FGFRI, EGFR, PIK3R1, TCF7L2
hsa04914	Progesterone-mediated oocyte maturation	22	2.40x10 <sup>-7</sup>	GNAI1, CCNB2, CCNA2, IGF1, CPEB1, PIK3R2, CDC25C, CCNB1, MAD2L1, AKT3, MAPK10, ADCY4, PDE3B, RPS6KA2, PIK3R1, PLK1, ANAPC11, CDC25A, PKMYT1, BUB1, RPS6KA3, CDK1
hsa03320	PPAR signaling pathway	19	6.97x10 <sup>-7</sup>	ACSL1, AQP7, ACADL, SLC27A4, FABP4, PPARA, PLTP, ADIPOQ, SLC27A6, ACSL5, ANGPTL4, PLIN1, ACSL4, CD36, SORBS1, LPL, PPARG, MMP1, OLR1
hsa05222	Small cell lung cancer	21	7.22x10 <sup>-7</sup>	CDK6, CCNE2, LAMA3, RARB, ITGA6, LAMA2, CCNE1, PIK3R2, E2F1, E2F2, AKT3, LAMC1, LAMB3, LAMA4, PIK3R1, LAMA1, MYC, LAMC3, LAMC2, FNI, PTGS2
hsa04610	Complement and coagulation cascades	18	1.04x10 <sup>-6</sup>	F10, PLAUR, F3, C3, PROS1, SERPING1, CFI, C6, F12, MASP2, VWF, C4BPA, MASP1, C7, TFPI, PROC, PLAU, F7
hsa04114	Oocyte meiosis	24	1.06x10 <sup>-6</sup>	CCNE2, ESPL1, CCNB2, CDC20, PTTG1, IGF1, CPEB1, SGOL1, CCNE1, CDC25C, CCNB1, MAD2L1, PPP2R1B, ADCY4, ITPR2, RPS6KA2, PLK1, PPPICA, ANAPC11, PKMYT1, BUB1, RPS6KA3, AURKA, CDK1
hsa04010	MAPK signaling pathway	40	2.61x10 <sup>-6</sup>	NTRK2, MRAS, CACNA2D1, FGF9, MEF2C, FOS, FGF1, FGF2, MAPK8IP2, CACNG1, DUSP1, TGFB2, FGF3, CACNG4, PDGFRA, FLNC, RAPGEF2, PDGFA, AKT3, DUSP6, MAPK10, PRKCA, FGFRI, MAP3K5, RPS6KA2, MECOM, EGFR, MYC, JUN, PLA2G4A, NR4A1, RRAS2, NFATC2, FGF7, HSPB1, RPS6KA3, STMN1, CACNB3, NTF4, MAP3K8

KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, false discovery rate.

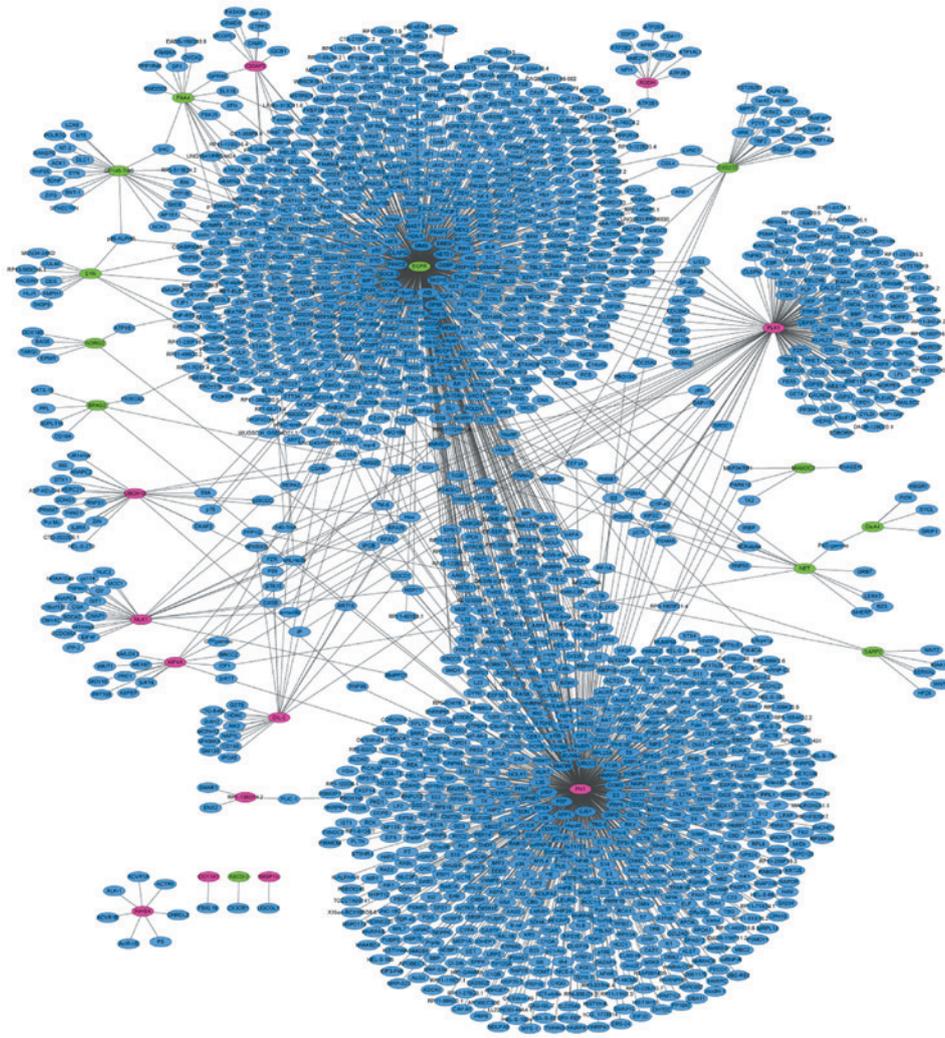


Figure 2. Protein-protein interaction network for the top 15 upregulated and downregulated DEGs. Red and green nodes denote upregulated and downregulated DEGs, respectively. Blue nodes denote products of genes predicted to interact with the DEGs. DEG, differentially expressed gene.

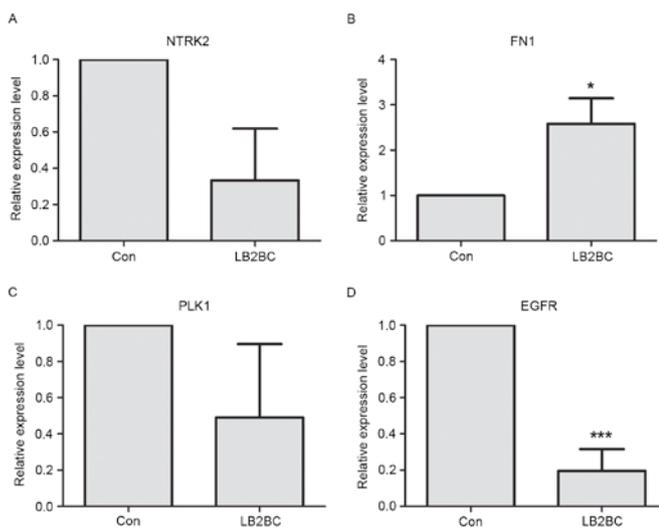


Figure 3. Reverse transcription-quantitative polymerase chain reaction validation of DEGs between LB2BC tumor and adjacent tissue samples. Relative expression of (A) NTRK2, (B) FN1, (C) PLK1, (D) EGFR. \* $P < 0.05$ ; \*\*\* $P < 0.001$  vs. Con. LB2BC, luminal B2 breast cancer; NTRK2, neurotrophic receptor tyrosine kinase 2; FN1, fibronectin-1; PLK1, Polo-like kinase 1; EGFR, epidermal growth factor receptor; CON, adjacent healthy control tissue.

downregulation in LB2BC (Fig. 3A). NTRK2 is a member of the neurotrophic tyrosine receptor kinase family. The expression of NTRK2 was previously identified through microarray analysis to be downregulated in patients with breast cancer with a relatively poor prognosis (18). NTRK2 expression leads to anoikis resistance in several types of cancer, including ovarian cancer (19,20). NTRK2 is a direct target of microRNA (miR)-200c; miR-200c decreases the endogenous expression of NTRK2 and mediated the suppression of anoikis resistance in breast cancer cells (21).

In the present study, FN1 was identified as a top-10 upregulated DEG in LB2BC (Table I). RT-qPCR analysis identified that FN1 was significantly upregulated in LB2BC (Fig. 3B), in accord with the DEG analysis. FN1 was one of the hub proteins in the PPI network (Fig. 2). FN1 is a glycoprotein present in a soluble dimeric form in plasma, and in a dimeric or multimeric form at the cell surface and the extracellular matrix (22). FN1 is associated with cell adhesion, cell migration, wound healing and cell metastasis (23). FN1 is upregulated in breast cancer; FN1 was also identified as hub protein in a previously constructed PPI network (24). Previous gene regulatory network analysis of breast cancer demonstrated that FN1 was upregulated in aggressive breast cancer cell lines and was associated with the

aggressive behavior of breast cancer cells (25). The expression of FN1 is a positive predictor for breast cancer cell line sensitivity to paclitaxel (26). The elevated expression of FN1 promotes vascular endothelial growth factor-C expression and the epithelial-mesenchymal transition, promoting lymph node metastasis in human oral squamous cell carcinoma (27). FN1 was associated with 5 significantly enriched KEGG pathways, including 'focal adhesion', 'pathways in cancer', 'ECM-receptor interaction', 'regulation of actin cytoskeleton' and 'small cell lung cancer' (Table III).

PLK1 was significantly upregulated in LB2BC (Table I). PLK1 is a key regulator of cell division and is overexpressed in numerous types of human cancer, including prostate, pancreatic and breast cancer (28). PLK1 is significantly enriched in the 'cell cycle' and 'progesterone-mediated oocyte maturation' pathways (Table III). PLK1 was identified as downregulated in LB2BC as assessed by RT-qPCR, whereas the opposite result was obtained in DEG analysis (Fig. 3C); this may be due to the small experimental sample size ( $n=5$ ) used in the present study. PLK1 was a hub protein, interacting with 153 genes in the PPI network (Fig. 2). PLK1 encodes a serine/threonine protein kinase that belongs to the cell cycle serine/threonine-protein kinase CDC5/Polo subfamily. PLK1 is significantly overexpressed in triple-negative breast cancer (TNBC), compared with other breast cancer subtypes; the inhibition of PLK1 activity induces an increase in G2/M cell cycle arrest and TNBC cell apoptosis (29). PLK1 is overexpressed in breast cancer cell lines compared with normal breast cell lines; silencing PLK1 enhances the sensitivity of breast cancer cell lines to rapamycin (30).

EGFR expression was significantly downregulated in LB2BC, as observed via DEG analysis and validated using RT-qPCR (Table I; Fig. 3D). EGFR was the hub protein with the highest connectivity degree, connecting to 781 genes in the PPI network. EGFR was significantly enriched in several pathways, including pathways in cancer and the MAPK signaling pathway (Table III). EGFR was previously identified as upregulated in various types of cancer, including lung cancer, gastric cancer and glioblastoma (31-33). EGFR is frequently overexpressed in TNBC; high EGFR gene copy number predicts a poor patient outcome (34). EGFR expression was downregulated in LB2BC in the present study, indicating that the pathogenesis of LB2BC is distinct from TNBC.

In conclusion, 2,251 DEGs were identified between LB2BC and non-tumor tissue. The top 15 upregulated and downregulated genes in LB2BC were used to construct a PPI network. A number of genes, including NTRK2, FN1, PLK1 and EGFR, were identified that potentially serve critical roles in the pathogenesis of LB2BC via the 'focal adhesion', 'pathways in cancer' and 'ECM-receptor interaction signaling' KEGG pathways. The findings of the present study may contribute to the further elucidation of the pathogenesis of LB2BC.

## References

- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A: Global cancer statistics, 2012. *CA Cancer J Clin* 65: 87-108, 2015.
- Perou CM, Sørbye T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, *et al*: Molecular portraits of human breast tumours. *Nature* 406: 747-752, 2000.
- Sørbye T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, *et al*: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98: 10869-10874, 2001.
- Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, Livasy C, Carey LA, Reynolds E, Dressler L, *et al*: The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 7: 96, 2006.
- Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J and Shi B: Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res* 5: 2929-2943, 2015.
- Ignatiadis M, Bedard P, Haibe-Kains B, Haibe-Kains B, Singhal S, Loi S, Criscitiello C, Desmedt C, Bontempi G, Piccart M, *et al*: A meta-analysis of gene expression profiling studies identifies clinically relevant oncogenic pathways in basal-like breast cancer. *Cancer Res* 69: 106, 2009.
- Tran B and Bedard PL: Luminal-B breast cancer and novel therapeutic targets. *Breast Cancer Res* 13: 221, 2011.
- Tomczak K, Czerwińska P and Wiznerowicz M: The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp Oncol (Pozn)* 19: A68-A77, 2015.
- Anders S and Huber W: Differential expression analysis for sequence count data. *Genome Biol* 11: R106, 2010.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK: Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47, 2015.
- Eden E, Navon R, Steinfeld I, Lipson D and Yakhini Z: GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10: 48, 2009.
- Kanehisa M: The KEGG database. *Novartis Found Symp* 247: 91-103, 119-128, 244-252, 2002.
- Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, *et al*: A human protein-protein interaction network: A resource for annotating the proteome. *Cell* 122: 957-968, 2005.
- Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, *et al*: The BioGRID interaction database: 2015 update. *Nucleic Acids Res* 39 (Database Issue): D470-D478, 2015.
- Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, *et al*: The BioGRID interaction database: 2015 update. *Nucleic Acids Res* 43 (Database Issue): D470-D478, 2015.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T: Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504, 2003.
- Schmittgen TD and Livak KJ: Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 3: 1101-1108, 2008.
- Li Z, Peng L, Han S, Huang Z, Shi F, Cai Z, Li X, Zhang P, Zhu H and Jin W: Screening molecular markers in early breast cancer of the same pathological types but with different prognoses using Agilent gene chip. *Nan Fang Yi Ke Da Xue Xue Bao* 33: 1483-1488, 2013 (In Chinese).
- Yu X, Liu L, Cai B, He Y and Wan X: Suppression of anoikis by the neurotrophic receptor TrkB in human ovarian cancer. *Cancer Sci* 99: 543-552, 2008.
- Kupferman ME, Jiffar T, El-Naggar A, Yilmaz T, Zhou G, Xie T, Feng L, Wang J, Holsinger FC, Yu D and Myers JN: TrkB induces EMT and has a key role in invasion of head and neck squamous cell carcinoma. *Oncogene* 29: 2047-2059, 2010.
- Howe EN, Cochrane DR and Richer JK: Targets of miR-200c mediate suppression of cell motility and anoikis resistance. *Breast Cancer Res* 13: R45, 2011.
- Chen D, Wang X, Liang D, Gordon J, Mittal A, Manley N, Degenhardt K and Astrof S: Fibronectin signals through integrin  $\alpha 5 \beta 1$  to regulate cardiovascular development in a cell type-specific manner. *Dev Biol* 407: 195-210, 2015.
- Steffens S, Schrader AJ, Vetter G, Eggers H, Blasig H, Becker J, Kuczyk MA and Serth J: Fibronectin I protein expression in clear cell renal cell carcinoma. *Oncol Lett* 3: 787-790, 2012.
- Liu X, Ma Y, Yang W, Wu X, Jiang L and Chen X: Identification of therapeutic targets for breast cancer using biological informatics methods. *Mol Med Rep* 12: 1789-1795, 2015.
- Yang X, Jia M, Li Z, Lu S, Qi X, Zhao B, Wang X, Rong Y, Shi J, Zhang Z, *et al*: Bioinformatics analysis of aggressive behavior of breast cancer via an integrated gene regulatory network. *J Cancer Res Ther* 10: 1013-1018, 2014.

26. Dorman SN, Baranova K, Knoll JH, Urquhart BL, Mariani G, Carcangiu ML and Rogan PK: Genomic signatures for paclitaxel and gemcitabine resistance in breast cancer derived by machine learning. *Mol Oncol* 10: 85-100, 2016.
27. Morita Y, Hata K, Nakanishi M, Omata T, Morita N, Yura Y, Nishimura R and Yoneda T: Cellular fibronectin 1 promotes VEGF-C expression, lymphangiogenesis and lymph node metastasis associated with human oral squamous cell carcinoma. *Clin Exp Metastasis* 32: 739-753, 2015.
28. Gutteridge RE, Ndiaye MA, Liu X and Ahmad N: Plk1 inhibitors in cancer therapy: From laboratory to clinics. *Mol Cancer Ther* 15: 1427-1435, 2016.
29. Maire V, Némati F, Richardson M, Vincent-Salomon A, Tesson B, Rigai G, Gravier E, Marty-Prouvost B, De Koning L, Lang G, *et al*: Polo-like kinase 1: A potential therapeutic option in combination with conventional chemotherapy for the management of patients with triple-negative breast cancer. *Cancer Res* 73: 813-823, 2013.
30. Ou O, Huppi K, Chakka S, Gehlhaus K, Dubois W, Patel J, Chen J, Mackiewicz M, Jones TL, Pitt JJ, *et al*: Loss-of-function RNAi screens in breast cancer cells identify AURKB, PLK1, PIK3R1, MAPK12, PRKD2, and PTK6 as sensitizing targets of rapamycin activity. *Cancer Lett* 354: 336-347, 2014.
31. Serna E, Lopez-Gines C, Monleon D, Muñoz-Hidalgo L, Callaghan RC, Gil-Benso R, Martinetto H, Gregori-Romero A, Gonzalez-Darder J and Cerda-Nicolas M: Correlation between EGFR amplification and the expression of microRNA-200c in primary glioblastoma multiforme. *PLoS One* 9: e102927, 2014.
32. Naruke A, Azuma M, Takeuchi A, Ishido K, Katada C, Sasaki T, Higuchi K, Tanabe S, Saegusa M and Koizumi W: Comparison of site-specific gene expression levels in primary tumors and synchronous lymph node metastases in advanced gastric cancer. *Gastric Cancer* 18: 262-270, 2015.
33. Yang CH, Chou HC, Fu YN, Yeh CL, Cheng HW, Chang IC, Liu KJ, Chang GC, Tsai TF, Tsai SF, *et al*: EGFR over-expression in non-small cell lung cancers harboring EGFR mutations is associated with marked down-regulation of CD82. *Biochim Biophys Acta* 1852: 1540-1549, 2015.
34. Park HS, Jang MH, Kim EJ, Kim HJ, Lee HJ, Kim YJ, Kim JH, Kang E, Kim SW, Kim IA and Park SY: High EGFR gene copy number predicts poor outcome in triple-negative breast cancer. *Mod Pathol* 27: 1212-1222, 2014.