Breast Cancer (2019) 26:305–316
https://doi.org/10.1007/s12282-018-0924-y

ORIGINAL ARTICLE

Identification of differentially expressed genes and typical fusion genes associated with three subtypes of breast cancer

Rong Wang[1] · Jinbin Li[2] · Chunyu Yin[2] · Di Zhao[3] · Ling Yin[2]

Received: 22 June 2018 / Accepted: 15 October 2018 / Published online: 16 November 2018

Abstract

Background This study aimed to identify the differentially expressed genes (DEGs) and the typical fusion genes in different types of breast cancers using RNA-seq.

Methods GSE52643 was downloaded from Gene Expression Omnibus, which included 1 normal sample (MCF10A) and 7 breast cancer samples (BT-474, BT-20, MCF7, MDA-MB-231, MDA-MB-468, T47D, and ZR-75-1). The transcript abundance and the DEGs screening were performed by Cufflinks. The functional and pathway enrichment was analyzed by Gostats. SnowShoes-FTD was applied to identify the fusion genes.

Results We screened 430, 445, 397, 417, 369, 557, and 375 DEGs in BT-474, BT-20, MCF7, DA-MB-231, MDA-MB-468, T47D, and ZR-75-1, respectively, compared with MCF10A. DEGs in each comparison group (such as *CD40* and *CDH1*) were significantly enriched in the functions of cell adhesion and extracellular matrix organization and pathways of CAMs and ECM receptor interaction. *UCP2* was a common DEG in the 7 comparison groups. *SFRP1* and *MMP7* were significantly enriched in *wnt*/–catenin signaling pathway in MDA-MB-231. *FAS* was significantly enriched in autoimmune thyroid disease pathway in BT-474. Besides, we screened 96 fusion genes, such as *ESR1-C6orf97* in ZR-75-1, *COBRA1-C9orf167* in BT-20, and *VAPB-IKZF3* and *ACACA-STAC2* in BT-474.

Conclusions The DEGs such as *SFRP1, MMP7, CDH1, FAS*, and *UCP2* might be the potential biomarkers in breast cancer. Furthermore, some pivotal fusion genes like *ESR1-C6orf97* with *COBRA1-C9orf167* and *VAPB-IKZF3* with *ACACA-STAC2* were found in Luminal A and Luminal B breast cancer, respectively.

Keywords Breast cancer · RNA-seq · Fusion genes · Differentially expressed gene

Introduction

Breast cancer, the leading cause of cancer death among females, possessed an approximately proportion of 23% among the total new cancer cases as well as mortality of 14% (458,400) of the total cancer deaths in 2008 [1]. Breast cancer was a collection of diseases that had distinct histopathological features including genetic and genomic variability, and multiple prognostic outcomes [2]. All these meant that clinical management for breast cancer should consider cases within the differences to arrive at appropriate therapeutic advice [3]. Gene expression studies using DNA microarrays have classified breast cancer into distinct subtypes, including basal-like, human epidermal growth factor receptor-2 positive/estrogen receptor negative (HER2+/ER–), luminal A, and luminal B [4]. However, these subtypes had different epidemiological risk factors [5, 6], different natural histories [7, 8], and different responses to systemic and local therapies [9–11], which brought a big challenge for the treatment of breast cancer.

So far, various gene biomarkers have been found for breast cancer. For example, *Erβ* could reduce the tumorigenesis by down-regulating the tumor cell cycle and transcription

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12282-018-0924-y) contains supplementary material, which is available to authorized users.

✉ Ling Yin
yinling17vip@163.com

[1] National Research Institute for Health and Family Planning, Beijing 100081, China

[2] Core Laboratory of Translational Medicine, Chinese PLA General Hospital, No. 28, Fuxing Road, Haidian District, Beijing 100853, China

[3] Dermatological of Department, The 309 Hospital of Chinese PLA, Beijing 100091, China

in T-47D [12]. In addition, the chemokine receptor *CXCR7* could promote the breast cancer cell proliferation by interacting with *EGFR* [13]. However, the mechanisms related to the different type of breast cancer were still unclear to us. In recent years, DNA and RNA microarrays have been of great value in illustrating the complex biological mechanisms, resulting in a new understanding of the development and progression of human cancer [14]. In addition, the second-generation sequencing of the transcriptome (RNA-seq) was a sensitive and efficient method for detecting gene fusions, somatic mutations, and alternatively spliced forms [15]. For all the advantages, RNA-seq was a revolutionary tool for the breast cancer treatment.

In our study, RNA-seq along with the bioinformatics analysis was used to identify the differentially expressed genes (DEGs), fusion genes, and the biological processes related to several breast cancer cell lines and to provide theoretical guidance for the future breast cancer treatment.

Materials and methods

RNA-seq data sets

The breast cancer RNA-seq data GSE52643 was downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/), which was based on the platform of Illumina HiSeq 2000 (GPL11154). GSE52643 data set was a collection of 1 normal mammary epithelial cell sample (MCF10A) and 7 breast cancer cell samples [Luminal A subtype (MCF7, T47D and ZR-75-1), Luminal B subtype (BT-474), and triple negative (TN) subtype (BT-20, MDA-MB-231 and MDA-MB-468)], as shown in Table 1. The RNA-seq data set was paired-end sequenced with the read length of 50 bp.

RNA-seq data set pre-processing

The raw data set with SRA format was downloaded from FTP in national center for biotechnology information (NCBI) and then was transferred into the fastq format with the software SRAtoolkit (version: 2.3.5-2, http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software) [16]. Then, FastQC (version: 0.11.2, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) [17] was applied to investigate the data quality. In addition, the FASTX-Toolkit (version: 0.06, http://hannonlab.cshl.edu/fastx_toolkit/index.html) was employed to screen the RNA-seq data. The short fragments whose 90% nucleotide base sequencing quality scores no less than 20 were taken as the high-quality reads. The Python script was used to re-integrate the fragments from both ends of the sequence to obtain the paired-end reads with high quality.

Alignments of paired-end reads

The paired-end reads with high quality were mapped to the human genome (version: GRCh37, which was downloaded from the official website of Tophat, http://ccb.jhu.edu/software/tophat/igenomes.shtml) using the software Tophat (version: 2.0.12) with the default parameters (with two mismatch at most) to get the BMA files [18].

Transcript abundance calculation of transcript and the screening of DEGs

The FPKM (Fragments Per Kilobase of exon per Million fragments mapped, FPKM) was introduced to evaluate the expression abundance of each transcript by the software Cufflinks (version: 2.21) [19] using the BMA profiles. The Cuffdiff of Cufflinks [20] was used to screen the DEGs. Benjamini and Hochberg method [21] was used to conduct multiple tests. The adjusted p value [that is, false discovery rate (FDR)] < 0.05 was taken as the cut-off criteria. Finally, these DEGs were represented using the heat map, and was visualized by the R package pheatmap (version: 1.0.8) [22].

Functional and pathway enrichment analysis of DEGs

The Gostats package of R language [23] was used to conduct the analysis of functional and pathway enrichment. The

Table 1 Detail information of the 8 cell lines

Cell lines	Disease state	ER	PR	HER-2	Subtypes	GEO number
BT-474	Breast cancer	+	+	+	Luminal B	SRX381532
MCF10A	Normal breast	−	−	−	Normal	SRX381533
BT-20	Breast cancer	−	−	−	Triple negative (TN)	SRX381534
MCF7	Breast cancer	+	+	−	Luminal A	SRX381535
MDA-MB-231	Breast cancer	−	−	−	TN	SRX381536
MDA-MB-468	Breast cancer	−	−	−	TN	SRX381537
T47D	Breast cancer	+	+	−	Luminal A	SRX381538
ZR-75-1	Breast cancer	+	−	−	Luminal A	SRX381539

adjusted p value (FDR) < 0.05 and DEGs counts ≥ 2 were used as the cut-off criteria for the identification of significantly enriched pathways and functions for DEGs.

The analysis of common DEGs in 7 comparison groups

For the DEGs of 7 comparison groups, the intersection analysis between samples was performed using Upset R package (version 1.3.3, https://cran.r-project.org/web/packages/UpSetR) [24]. Then, the common DEGs were selected. Meanwhile, the R package clusterprofiler (http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html) [25] was used to conduct GO-BP [26] function enrichment analysis. The p value < 0.05 was set as the cut-off criteria for identification of significantly enriched functions for common DEGs.

Fusion gene identification

The data of fusion genes were downloaded from GEO (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52643). The paired-end reads were mapped to the human genome (version: GRCh37) using the tool BWA (version: 0.6.2) [27, 28] to get the SAM profiles which could be sorted by the tool SAMtools based on the reading IDs. Then, the fusion genes were identified by Snowshoes-FTD (version: 2.0) [29] based on the Mitelman database.

Results

Alignment analysis

The alignments of sequenced reads are shown in the supplementary Table 1. Among them, sample ZR-75-1 had the highest ratio (78.19%) (22,111,223/28,279,001) of paired-end reads with high quality to total reads, while the sample MDA-MB-231 obtained a ratio of 27.25% (8,555,113/31,393,951), which was mainly due to the poor-sequencing quality of a single end (supplementary Fig. 1). For the mapping ratio among the 8 samples, the MDA-MB-231, BT-20, and MCF10A possessed the relatively high ratio of 88.83%, 88.41%, and 88.27%, respectively. The BT-474 had the highest total reads of 33,108,579, while the T47D had the lowest total reads of 27,834,336.

Screening of DEGs related to breast cancer

Compared with the MCF10A, the identified DEGs in BT-474 (164 up-regulated and 266 down-regulated DEGs), BT-20 (107 up-regulated and 308 down-regulated DEGs), MCF7 (127 up-regulated and 270 down-regulated DEGs), MDA-MB-231 (76 up-regulated and 341 down-regulated DEGs), MDA-MB-468 (81 up-regulated and 288 down-regulated DEGs), T47D (143 up-regulated and 414 down-regulated DEGs), and ZR-75-1 (124 up-regulated and 251 down-regulated DEGs) are shown in the clustering heat map (Fig. 1). These DEGs could well distinguish the breast cancer cell lines and normal breast cell line. Furthermore, after intersection analysis of DEGs in 7 comparison groups (supplementary Fig. 2, data are

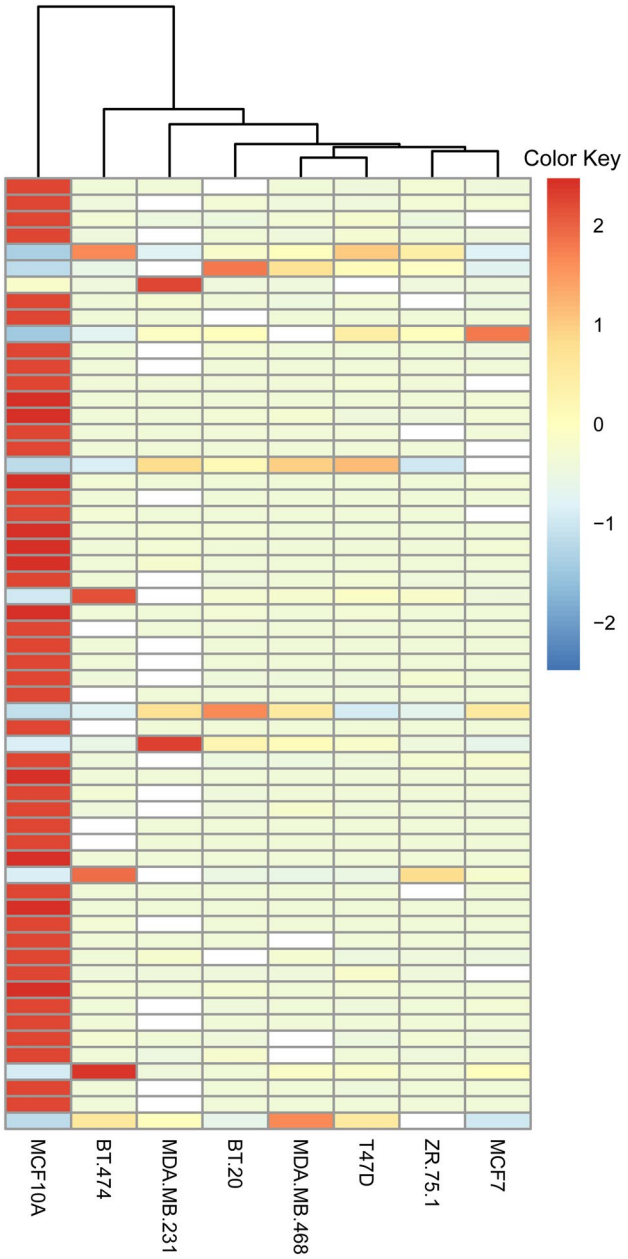

Fig. 1 Clustering heat map of the differentially expressed genes (DEGs) between the breast cancer cell lines and the normal breast cell line. The horizontal axis represents samples and vertical axis represents gene expressions

partly showed), a total of 16 common DEGs were identified, such as *UCP2, RGL3*, and *SGIP1*. It was found that *UCP2* was over-expressed among all the 7 comparison groups.

Functional and pathway enrichment analysis

According to the GO function enrichment analysis, DEGs such as *ALOX15* and *CD44* were significantly enriched in the function of cell adhesion (GO: 0007155). DEGs like *AGT* and *VCAN* were significantly related to extracellular matrix organization (GO: 0030198). Moreover, these two GO terms were significantly associated with the DEGs screened from all of the 7 breast cancer cell samples. The enriched top five GO terms for the DEGs of the 7 breast cancer cell samples are shown in Table 2. Meanwhile, as is indicated in the Fig. 2, the common DEGs from 7 comparison groups were significantly enriched in the functions of extracellular matrix organization (GO: 0030198) and cellular response to organonitrogen compound (GO:0071417).

For the KEGG pathway enrichment, DEGs such as *CD40* and *CDH1* were significantly enriched in the pathway of cell adhesion molecules (CAMs) (KEGG ID: 4514) and DEGs like *COL1A2* and *CD44* were significantly enriched in the pathway of ECM receptor interaction (KEGG ID: 4512). What's more, DEGs of the 7 breast cancer cell samples were significantly enriched in these above two pathways. In addition, in the MDA-MB-231 cell line of TN breast cancer, DEGs such as *MMP7* and *SFRP1* were significantly enriched in *wnt*-signaling pathway (KEGG ID: 4310). Besides, in BT-474 of Luminal B breast cancer, DEGs such as *FAS* and *HLA-A* were significantly enriched in the pathway of autoimmune thyroid disease, while, in the other breast cancer types, there did not exist. Furthermore, the *UCP2* was significantly enriched in the functions of regulation of cell death and regulation of apoptotic process. The top five KEGG pathways for the DEGs of the 7 breast cancer cell samples are listed in Table 3.

Fusion gene analysis

Totally, 96 fusion genes were identified with tool Snowhsoes-FTP in 8 breast cell lines (Table 4). Among them, the partner genes of 18 fusion genes were from different chromosomes (inter-chr), while the remains' partner genes came from the same chromosome (intra-chr). Searching for the Mitelman database, a total of 12 fusion genes (such as *VAPB-IKZF3* and *ACACA-STAC2*) were found to be related to cancer. Moreover, these 12 fusion genes were all from the BT-474 and MCF7 cell lines (Table 5). What's more, it was also found that these 12 fusion genes were existed in the specific cell line in our study, for example, *GLB1-CMTM7* only existed in BT-474 cell line, while *ARFGEF2-SULF2* only in MCF7. In addition, our study also discovered that some fusion genes such as *ADCK4-NUMBL, ARPC4-TTLL3, ESR1-C6orf97, COBRA1-C9orf167, FAM76A-MCM4*, and *LOC100131434-IDS* were found in two different cell lines (supplementary Table 2). However, fusion gene was not found to be expressed in three or more cell lines.

Discussion

In this study, we screened out the DEGs in the breast cancer cell lines by comparing with normal breast cell line and performed the functional analysis as well as the fusion genes analysis. *UCP2* was found to be over-expressed in all the 7 breast cancer cell lines. According to the functional enrichment, *UCP2* was significantly enriched in the functions of regulation of cell death and regulation of apoptotic process. The mitochondrial uncoupling protein 2 (*UCP2*, also known as uncoupling protein 2) is a mitochondrial anion carrier proteins. Dysfunctional mitochondria could cause retrograde response in the cell nuclear, leading to cell reprogramming [30]. Mitochondrial uncoupling and Warburg effect were the molecular basis for metabolic recombination of cancer cells [31]. Chen et al. proved that the expression of *UCP2* could help the adaptation and survival of tumor cells in an increasingly hostile tissue microenvironment and make a contribution to the tumor evolution [32]. In addition, reportedly, it has been suggested that *UCP2* was over-expressed in neck, skin, pancreas, and prostate tumor tissues, and promoted tumors growth [33]. What's more, former report has shown that *UCP2* was over-expressed in breast cancer and could promote tumorigenic properties [34]. Thus, *UCP2* could be a potential target gene for the treatment of breast cancer. According to the GO function enrichment analysis, DEGs of each comparison group were closely related to cell adhesion and extracellular matrix organization, and the common DEGs from 7 comparison groups were significantly enriched in the functions of extracellular matrix organization. To some extent, it was suggested that cell adhesion and extracellular matrix organization played a key role in breast cancer. Furthermore, the recent study has showed that the poor treatment and the high mortality of breast cancer were attributed to the metastasis of tumor cells [35]. To complete the process of tumor metastasis, the tumor cells should first enhance the interaction with neighboring cells through the cell adhesion, the ECM, and the microenvironment [36]. Consequently, cell adhesion and extracellular matrix organization might play key roles in the metastasis of breast cancer.

In addition, the pathway enrichment revealed that the DEGs in all the 7 comparison groups were significantly enriched in the pathways of CAMs and ECM–receptor interaction. CAMs were cell-surface proteins that performed the interactions for cell-to-cell or cell-to-ECM. Furthermore,

Table 2 Top five enriched function of the DEGs in the 7 comparison groups

Groups	GO IDs	FDR	Counts	Terms	**Gene symbols**
BT-20 vs. MCF10A	GO:0044707	2.99E − 19	226	Single–multicellular organism process	*SERPINA3, ACTG2, JAG1, AGT, ALOX5, ALOX15, AMPH, ANGPT1, ANK1, ANXA1 …*
	GO:0032501	1.43E − 18	230	Multicellular organismal process	*SERPINA3, ACTG2, JAG1, AGT, ALOX5, ALOX15, AMPH, ANGPT1, ANK1, ANXA1…*
	GO:0048731	1.52E − 17	163	System development	*JAG1, AGT, ALOX15, ANGPT1, ANK1, ANXA1, APBA2, APBB2, BDNF, BMPR1B …*
	GO:0007275	1.64E − 17	185	Multicellular organismal development	*JAG1, AGT, ALOX15, ANGPT1, ANK1, ANXA1, APBA2, APBB2, BDNF, BMPR1B …*
	GO:0048856	2.73E − 16	178	Anatomical structure development	*JAG1, AGT, ALOX15, ANGPT1, ANK1, ANXA1, APBA2, APBB2, BDNF, BMPR1B …*
BT-474 vs. MCF10A	GO:0030198	2.34E − 17	43	Extracellular matrix organization	*BMP1, BMP7, CD44, COL1A2, COL3A1, COL6A2, COL16A1, COL17A1, VCAN, CTSB …*
	GO:0043062	2.59E − 17	43	Extracellular structure organization	*BMP1, BMP7, CD44, COL1A2, COL3A1, COL6A2, COL16A1, COL17A1, VCAN, CTSB …*
	GO:0032501	4.15E − 14	219	Multicellular organismal process	*ACTG2, ADD2, AFP, AGTR1, ALDH1A3, AMPD2, ANGPT1, ANK2, APBA2, KLK3 …*
	GO:0044707	5.22E − 14	213	Single–multicellular organism process	*ACTG2, ADD2, AFP, AGTR1, ALDH1A3, AMPD2, ANGPT1, ANK2, APBA2, KLK3 …*
	GO:0051239	3.67E − 11	97	Regulation of multicellular organismal process	*AGTR1, ANGPT1, ANK2, KLK3, FAS, ASCL2, BDNF, BMP1, BMP7, C3 …*
MCF7 vs. MCF10A	GO:0030198	1.22E − 15	39	Extracellular matrix organization	*APBB1, BMP1, BMP7, COL1A2, COL3A1, COL5A1, COL8A2, COL16A1, CTSB, DCN …*
	GO:0044707	1.27E − 15	203	Single–multicellular organism process	*ACTA2, ACTC1, ACTG2, ADD2, AFP, ALDH1A3, ALOX15, ANGPT1, ANK2, APBA2 …*
	GO:0043062	1.34E − 15	39	Extracellular structure organization	*APBB1, BMP1, BMP7, COL1A2, COL3A1, COL5A1, COL8A2, COL16A1, CTSB, DCN …*
	GO:0032501	8.79E − 15	206	Multicellular organismal process	*ACTA2, ACTC1, ACTG2, ADD2, AFP, ALDH1A3, ALOX15, ANGPT1, ANK2, APBA2 …*
	GO:0032502	5.91E − 14	181	Developmental process	*ACTA2, ACTC1, ADD2, AFP, ALDH1A3, ALOX15, ANGPT1, ANK2, APBA2, APBB1 …*
MDA-MB-231 vs. MCF10A	GO:0007155	5.66E − 20	80	Cell adhesion	*ADAM8, AGT, ANGPT1, AZGP1, BMP1, BMP2, DDR1, CD22, CDH1, CDH6 …*
	GO:0022610	7.22E − 20	80	Biological adhesion	*ADAM8, AGT, ANGPT1, AZGP1, BMP1, BMP2, DDR1, CD22, CDH1, CDH6 …*
	GO:0048731	8.47E − 20	172	System development	*ACP5, ACTA2, ADAM8, AFP, AGT, ALDH1A3, ANGPT1, APBA2, APOE, BMP1 …*
	GO:0044707	9.68E − 19	230	Single–multicellular organism process	*SERPINA3, ACP5, ACTA2, ACTG2, ADAM8, AFP, AGT, ALDH1A3, AMPH, ANGPT1 …*

Table 2 (continued)

Groups	GO IDs	FDR	Counts	Terms	**Gene symbols**
	GO:0001501	2.19E − 18	48	Skeletal system development	*ACP5, BMP1, BMP2, CBS, CDH11, CHI3L1, COL1A1, COL1A2, COL3A1, COL13A1* …
MDA-MB-468 vs. MCF10A	GO:0048731	7.24E − 21	156	System development	*ADD2, AFP, ALOX15B, ALPL, ANGPT1, ANK2, APBB2, BDNF, C3, CACNA1C* …
	GO:0022610	2.11E − 20	74	Biological adhesion	*ANGPT1, CD22, CDH4, CDH11, CDH13, CDKN2A, CNTN1, COL6A2, COL6A3, COL16A1* …
	GO:0007155	7.02E − 20	73	Cell adhesion	*ANGPT1, CD22, CDH4, CDH11, CDH13, CDKN2A, CNTN1, COL6A2, COL6A3, COL16A1* …
	GO:0007275	3.75E − 19	172	Multicellular organismal development	*ADD2, AFP, ALOX15B, ALPL, ANGPT1, ANK2, APBB2, BDNF, C3, CACNA1C* …
	GO:0048856	1.27E − 17	165	Anatomical structure development	*ADD2, AFP, ALOX15B, ALPL, ANGPT1, ANK2, APBB2, BDNF, C3, CACNA1C* …
T47D vs. MCF10A	GO:0030198	8.49E − 19	51	Extracellular matrix organization	*A2M, BMP1, CD44, COL1A2, COL6A2, COL6A3, COL16A1, COL17A1, VCAN, DCN* …
	GO:0043062	9.58E − 19	51	Extracellular structure organization	*A2M, BMP1, CD44, COL1A2, COL6A2, COL6A3, COL16A1, COL17A1, VCAN, DCN* …
	GO:0009653	4.15E − 17	143	Anatomical structure morphogenesis	*ACTC1, ALDH1A3, ANPEP, BDNF, BMP1, C3, CA2, CD44, CDH11, CHN1* …
	GO:0044707	1.52E − 16	272	Single–multicellular organism process	*A2M, SERPINA3, ACTC1, ADD2, AFP, ALDH1A3, ALOX15, ANPEP, ANXA1, RHOH* …
	GO:0007275	2.49E − 16	224	Multicellular organismal development	*ACTC1, ADD2, AFP, ALDH1A3, ALOX15, ANPEP, ANXA1, RHOH, BDNF, BMP1* …
ZR-75-1 vs. MCF10A	GO:0044707	6.11E − 19	204	Single–multicellular organism process	*ACTC1, ADD2, AFP, ALOX5, ANK2, ANXA1, APOD, KLK3, RHOH, ATP2A3* …
	GO:0030198	1.15E − 18	42	Extracellular matrix organization	*BMP1, CD44, COL6A3, COL7A1, COL16A1, COL17A1, CTSB, DCN, FAP, FGFR4* …
	GO:0043062	1.27E − 18	42	Extracellular structure organization	*BMP1, CD44, COL6A3, COL7A1, COL16A1, COL17A1, CTSB, DCN, FAP, FGFR4* …
	GO:0032501	4.39E − 18	207	Multicellular organismal process	*ACTC1, ADD2, AFP, ALOX5, ANK2, ANXA1, APOD, KLK3, RHOH, ATP2A3* …
	GO:0048731	2.76E − 17	148	System development	*ACTC1, ADD2, AFP, ANK2, ANXA1, APOD, KLK3, RHOH, BDNF, BMP1* …

CAMs also played a key role in the metastasis in breast cancer by promoting the tumor cells detachment from the primary tumor as well as the reduction of intercellular adhesion [37]. Besides, it was found that the DEGs such as *CD40* and *CDH1* were significantly enriched in CAMs pathway. *CD40* was a member of the tumor necrosis factor receptor superfamily and was broadly expressed to inhibit the tumor growth by induction of the tumor cell apoptosis in breast cancer [38]. In addition, *CDH1* (E-cadherin 1), belonging to cadherin superfamily, was a tumor suppressor gene [39]. Former reports have demonstrated that the loss of E-cadherin could induce the mesenchymal state of breast cancer cells, and then, invasion and metastasis occurred [40, 41]. Therefore, we assumed that the *CD40* and *CDH1* might

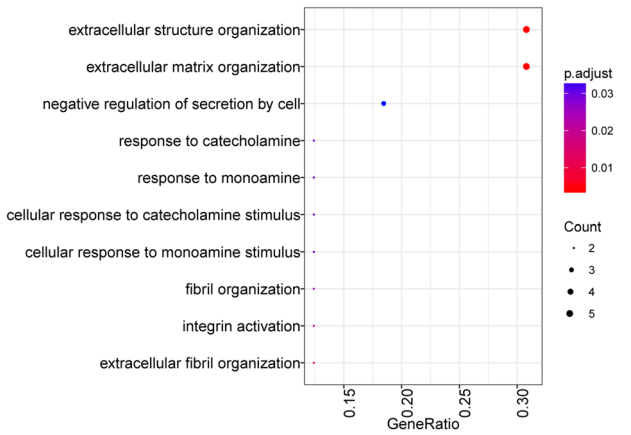

Fig. 2 Function enrichment analysis of the common DEGs in the 7 comparison groups

be key genes through regulating the CAMs pathway in the breast cancer. Apart from the common pathways, we also identified the *wnt*-signaling pathway in MDA-MB-231 of TN type breast cancer. In general, it has been found that the *wnt*-signaling pathway always recruited the cytoplasmic protein β-catenin to act as a key signaling intermediate and was known as the *wnt*/β-catenin signaling pathway [42]. Furthermore, the *wnt*/β-catenin signaling pathway was preferentially activated and played a key role in the development and progression of breast cancer cell [43]. Moreover, in our study, we also discovered that *MMP7* and *SFRP1* were significantly enriched in the *wnt*/β-catenin signaling pathway. Recent study has shown that *MMP7* expression was associated with tumor invasion, metastasis, and survival in breast cancer by degrading all the ECM components which was considered an essential step in the processes of invasion and metastasis [44, 45]. Furthermore, reports have also demonstrated that, in the TN breast cancer, the *MMP7*, which was involved in the *wnt*/β-catenin signaling pathway, was associated with the functional loss of *PTEN* gene, the most common first event associated with TN breast cancer [46, 47]. In addition, the former report has discovered that *SFRP1* (secreted frizzled-related protein 1) was a negative regulator for the *wnt* pathway and the loss of *SFRP1* expression in the early stage breast tumors was associated with poor prognosis [48]. As a result, we assumed that *MMP7* and *SFRP1* could take part in the regulation of TN breast cancer through the *wnt*/β-catenin signaling pathway. In addition, in the BT-474 cell line of Luminal B breast cancer, *FAS* was significantly enriched in the pathway of autoimmune thyroid disease. It has been proven that women affected by either benign or malignant thyroid disease had a significantly greater risk of breast cancer and the thyroid auto-antibodies appear to be protective against breast cancer. Moreover, *FAS* has been found to interact with the *FasL* which might trigger the cell death pathways in breast cancer cell and killed tumors by immune-effector cells [49, 50]. Thus, we speculated that *FAS* could trigger the immune-effector cells to kill the breast cancer cells through the autoimmune thyroid disease pathway in the BT-474 of luminal B breast cancer. However, further experiments are still necessary to verify the role of these genes in various subtypes of breast cancer cell lines.

Moreover, for the fusion gene analysis, a fusion gene *ESR1-C6orf97* was screened out in the luminal A type breast cancer including MCF7 and ZR-75-1. Former research has shown that a SNP on 6q25.1 which was associated with breast cancer lay ~ 200 kb upstream of *ESR1* in an intron of *C6orf97* [51]. Moreover, it has also been proven that *ESR1* could be co-expressed with the closely adjacent gene *C6orf97* spanning a breast cancer susceptibility locus at 6q25.1 [52]. Consequently, we speculated that the expression of *ESR1-C6orf97* could be related to the breast cancer susceptibility. In addition, in the luminal B type breast cancer, *VAPB-IKZF3* was also identified. It has been found that the knock-down of *VAPB-IKZF3* could lead to the inhibition of breast tumor cell growth [53]. When analyzing the partner genes of the fusion genes, we found that the partner genes *COBRA1* of fusion gene *COBRA1-C9orf167* in BT-20 (TN) and MCF7 (luminal A) and *ACACA* of *ACACA-STAC2* in BT-474 (Luminal B) were related to breast cancer. *COBRA1* could modulate estrogen-dependent and independent transcription, and suppresses the growth of breast cancer cells by binding to the breast cancer susceptibility gene product-*BRCA1* and its expression was significantly reduced in metastatic and recurrent breast cancer, referring to be a tumor suppressor in breast cancer development [54]. It has been also found that the *ACACA* could activate the fatty acid synthesis which was required for breast carcinogenesis [55]. However, there were a few reports on these fusion genes which were related to breast cancer. Former study has proven that the gene fusion could give rise to the activation of the silenced genes. For example, *GSDMB* could be activated by the fusion gene *TATDN1-GSDMB* in SK-BR-3 breast cancer cells and *IKZF3* could be activated by *GSDMB-IKZF3* in BT474 breast cancer cells [53]. Thus, it was speculated that the *COBRA1* and *ACACA* might be activated by the fusion gene to be involved in the breast cancer. Besides, we also discovered that the most fusion genes were only existed in one breast cancer cell line, while only 6 fusion genes were found in two kinds of cell lines. None of fusion genes was simultaneously existed in three or more cell lines. All these findings suggested that the expression of fusion genes might have something to do with the molecular subtypes of breast cancer.

Unfortunately, we did not obtain a regular relationship between fusion genes and breast cancer types, which might be due to the insufficient samples involved in our research. In addition, our results were not validated through animal

Table 3 Top five enriched pathway with the DEGs in the 7 comparison groups

Groups	KEGG IDs	Terms	FDR	Counts	Gene symbols
BT20 vs. MCF10A	4510	Focal adhesion	2.08E − 05	17	*CCND2, COL1A2, COL5A1, COL6A3, FN1, TNC, ITGA4, ITGA5, ITGB3, KDR, LAMA4, MYLK, PAK3, PDGFRB, VEGFC, PDGFD, COL6A6*
	4512	ECM–receptor interaction	7.84E − 05	10	*COL1A2, COL5A1, COL6A3, FN1, TNC, ITGA4, ITGA5, ITGB3, LAMA4, COL6A6*
	4514	Cell adhesion molecules (CAMs)	0.0008	11	*CD22, CNTN1, VCAN, HLA-B, HLA-F, ITGA4, MAG, NRXN3, PVRL3, NLGN4X, CADM3*
	4974	Protein digestion and absorption	0.001345	8	*FXYD2, COL1A2, COL5A1, COL6A3, MME, SLC8A1, SLC7A8, COL6A6*
	300	Lysine biosynthesis	0.002095	2	*ALDH7A1, AADAT*
BT474 vs. MCF10A	4512	ECM–receptor interaction	5.56E − 09	14	*CD44, COL1A2, COL3A1, COL6A2, FN1, HSPG2, ITGA5, ITGB3, ITGB4, ITGB6, LAMA4, LAMC1, LAMC2, SPP1*
	4510	Focal adhesion	2.66E − 06	17	*COL1A2, COL3A1, COL6A2, ERBB2, FN1, FYN, ITGA5, ITGB3, ITGB4, ITGB6, LAMA4, LAMC1, LAMC2, MYLK, RAC2, SPP1, PDGFC*
	4610	Complement and coagulation cascades	3.18E − 06	10	*CFB, C1R, C2, C3, CD59, F7, PLAU, PLAUR, PROS1, TFPI*
	4640	Hematopoietic cell lineage	2.91E − 05	10	*CD14, CD44, CD59, CSF2RA, CSF3R, EPOR, IL6, ITGA5, ITGB3, TPO*
	5410	Hypertrophic cardiomyopathy	0.000107	9	*CACNA2D1, DMD, IL6, ITGA5, ITGB3, ITGB4, ITGB6, TPM1, CACNA2D4*
MCF7 vs. MCF10A	4512	ECM–receptor interaction	7.88E − 07	12	*CD36, COL1A2, COL3A1, COL5A1, FN1, TNC, ITGA5, ITGB3, LAMA4, LAMB3, LAMC2, LAMA1*
	5146	Amoebiasis	8.50E − 06	12	*COL1A2, COL3A1, COL5A1, FN1, CXCL1, IL1B, LAMA4, LAMB3, LAMC2, SERPINB2, SERPINB10, LAMA1*
	4610	Complement and coagulation cascades	3.84E − 05	9	*CFB, C1R, C2, C3, F3, CFH, PLAT, PLAU, PLAUR*
	4514	Cell adhesion molecules	8.43E − 05	12	*CD22, CD40, CNTN1, CLDN3, NCAM1, NCAM2, NRCAM, PTPRM, CADM1, CNTNAP2, CADM3, PDCD1LG2*
	4510	Focal adhesion	0.000332	14	*COL1A2, COL3A1, COL5A1, FN1, TNC, ITGA5, ITGB3, LAMA4, LAMB3, LAMC2, MYLK, PDGFC, PDGFD, LAMA1*
MDA-MB-231 vs. MCF10A	53	Ascorbate and aldarate metabolism	3.76E − 09	9	*UGT1A10, UGT1A8, UGT1A7, UGT1A6, UGT1A5, UGT1A9, UGT1A4, UGT1A1, UGT1A3*
	5410	Hypertrophic cardiomyopathy	5.38E − 09	14	*CACNA1C, CACNA2D1, CACNB3, DMD, IGF1, ITGA4, ITGB6, SGCA, SGCD, SLC8A1,TNF, TNNT2, TPM1, TPM2*
	5414	Dilated cardiomyopathy	1.60E − 08	14	*CACNA1C, CACNA2D1, CACNB3, DMD, IGF1, ITGA4, ITGB6, SGCA, SGCD, SLC8A1,TNF, TNNT2, TPM1, TPM2*

Table 3 (continued)

Groups	KEGG IDs	Terms	FDR	Counts	Gene symbols
	40	Pentose and glucuronate interconversions	2.99E − 08	9	*UGT1A10, UGT1A8, UGT1A7, UGT1A6, UGT1A5, UGT1A9, UGT1A4, UGT1A1, UGT1A3*
	514	Other types of O-glycan biosynthesis	7.37E − 08	10	*UGT1A10, UGT1A8, UGT1A7, UGT1A6, UGT1A5, UGT1A9, UGT1A4, UGT1A1, UGT1A3, ST6GAL2*
MDA-MB-468 vs. MCF10A	4512	ECM–receptor interaction	1.16E − 06	11	*COL1A2, COL6A2, COL6A3, FN1, GP1BB, ITGA4, ITGA5, ITGB3, LAMA3, SDC2, SPP1*
	4640	Hematopoietic cell lineage	7.73E − 05	9	*CD22, CSF2RA, CSF3R, EPOR, GP1BB, IL1A, ITGA4, ITGA5, ITGB3*
	4514	Cell adhesion molecules	8.81E − 05	11	*CD22, CD40, CDH4, CNTN1, CLDN3, VCAN, HLA-DQA1, ITGA4, NRCAM, SDC2, NLGN4X*
	5412	Arrhythmogenic right-ventricular cardiomyopathy	0.000132	8	*CACNA1C, GJA1, ITGA4, ITGA5, ITGB3, SGCA, SLC8A1, LEF1*
	4510	Focal adhesion	0.000235	13	*COL1A2, COL6A2, COL6A3, FN1, ITGA4, ITGA5, ITGB3, LAMA3, MYLK, PAK3, SPP1, VEGFC, PDGFD*
T47D vs. MCF10A	5410	Hypertrophic cardiomyopathy	1.36E − 10	18	*ACTC1, CACNA2D1, ACE, DMD, IGF1, ITGA3, ITGA4, ITGA5, ITGB3, ITGB6, SGCA, SGCD, SLC8A1, TGFB2, TNNI3, TPM1, TPM4, CACNA2D3*
	5414	Dilated cardiomyopathy	4.37E − 09	17	*ACTC1, CACNA2D1, DMD, IGF1, ITGA3, ITGA4, ITGA5, ITGB3, ITGB6, SGCA, SGCD, SLC8A1, TGFB2, TNNI3, TPM1, TPM4, CACNA2D3*
	40	Pentose and glucuronate interconversions	4.66E − 08	10	*AKR1B1, UGT1A10, UGT1A8, UGT1A7, UGT1A6, UGT1A5, UGT1A9, UGT1A4, UGT1A1, UGT1A3*
	53	Ascorbate and aldarate metabolism	8.21E − 08	9	*UGT1A10, UGT1A8, UGT1A7, UGT1A6, UGT1A5, UGT1A9, UGT1A4, UGT1A1, UGT1A3*
	982	Drug metabolism—cytochrome P450	8.79E − 08	14	*ALDH3A1, ALDH1A3, AOX1, GSTM1, GSTM2, UGT1A10, UGT1A8, UGT1A7, UGT1A6, UGT1A5, UGT1A9, UGT1A4, UGT1A1, UGT1A3*
ZR-75-1 vs. MCF10A	4512	ECM–receptor interaction	3.29E − 07	12	*CD44, COL6A3, FN1, TNC, ITGA3, ITGA4, ITGB3, ITGB6, LAMA4, LAMB3, LAMC2, THBS1*
	4640	Hematopoietic cell lineage	4.85E − 07	12	*CD44, CD59, CSF1, CSF2RA, EPOR, IL1B, IL7R, ITGA3, ITGA4, ITGB3, MME, TPO*
	4610	Complement and coagulation cascades	2.59E − 06	10	*CFB, C1R, C2, C3, CD59, F3, FGG, PLAT, PLAUR, TFPI*
	4510	Focal adhesion	8.72E − 06	16	*CAV2, COL6A3, FLNA, FN1, FYN, TNC, ITGA3, ITGA4, ITGB3, ITGB6, LAMA4, LAMB3, LAMC2, MYLK, THBS1, PDGFC*

Table 3 (continued)

Groups	KEGG IDs	Terms	FDR	Counts	Gene symbols
	5410	Hypertrophic cardiomyopathy	1.42E − 05	10	*ACTC1, CACNA2D1, ITGA3, ITGA4, ITGB3, ITGB6, SGCA, SLC8A1, TGFB2, TPM1*

Table 4 Fusion genes screened in the normal breast cell line and the 7 breast cancer cell lines

Cell lines	Disease states	Counts	Gene symbols
MCF10A	Normal breast	6	*ARIH2-ZNF283, LOC155060-ZNF783, CNPY2- CS, ARPC4-TTLL3, GXYLT2-PPP4R2, SIDT2-TAGLN*
BT-474	Breast cancer	22	*ACACA-STAC2, CYP4F8-ILVBL, CIZ1-FAM102A, FITM2-UQCC, COL5A2-FLNB, CMTM7-GLB1, LAMP1-MCF2L, MED1-STXBP4, PIP4K2B-RAD51C, MGAT4C-PPP1R12A, MYO9B- RAB22A, RPS6KB1-SNF8, DOK5-STARD3, SYNRG-TOB1, MYO19-TRIM37, MRPL45-TRPC4AP, IKZF3-VAPB, CEP250-ZMYND8, CIZ1-FAM102A, PIP4K2B-RAD51C, IKZF3-VAPB, CEP250-ZMYND8*
BT-20	Breast cancer	15	*ACTB-RNF216, ADCK4-NUMBL, CIZ1-LCN2, C9orf167-COBRA1, GOLGB1-ILDR1, LIMA1-USP22, IDS-LOC100131434, CHD2-LOC100507217, ARHGEF4-PLEKHB2, RASSF3-TBK1, CRTC2-SLC39A1, KIAA1984-TMEM141, TYW1-VOPP1, CIZ1-LCN2, GOLGB1-ILDR1*
MCF7	Breast cancer	34	*ADAMTS19-SLC27A6, ADCK4-NUMBL, ARFGEF2-SULF2, ARPC4-TTLL3, ASL-CRCP, ATXN7L3-FAM171A2, ATXN7-BCAS3, BCAS3-REG4, BCAS3-BCAS4, BMP7-SULF2, ABCC1-C16orf45, C16orf62-IQCK, C9orf167-COBRA1, DEPDC1B-ELOVL7, EIF3H-RAD21, C6orf97-ESR1, GATAD2B-NUP210L, GCN1L1-MSI1, BCAS3-KIAA1324L, IDS-LOC100131434, FCHO1-MYO9B, AK7-PAPOLA, MATN2-POP1, CATSPER2-PPIP5K1, DIAPH3-RPS6KB1, RPS6KB1-VMP1, MIF-SLC2A11, CARM1-SMARCA4, PRICKLE2-SULF2, BRIP1-TAF4, CA4-TANC2, SYAP1-TXLNG, RPS6KB1-VMP1, PRICKLE2-SULF2*
MDA-MB-231	Breast cancer	0	
MDA-MB-468	Breast cancer	5	*ARID1A-MAST2, FAM76A-MCM4, C16orf46-GCSH, SLC25A32-UBR5, C16orf46-GCSH*
T47D	Breast cancer	4	*ARID3C-DCTN3, FAM76A-MCM4, SH3BP5-VGLL4, FAM149B1-VPS26A*
ZR-75-1	Breast cancer	10	*ARL6IP1-RPS15A, ARHGAP39-BOP1, C1orf151-NBL1-RCC2, C6orf97-ESR1, EXOSC10-KPNA6, MINOS1-RCC2, PDCD4-RBM20, SCNN1A-TNFRSF1A, HOMER1-SREK1, C6orf97-ESR1*

Table 5 Fusion genes verified in BT-474 and MCF7 breast cancer cell lines

Sample title	Fusion gene directional	Type	Fusion_ Strand	Dis-concordant read pairs	Split reads	5′ Chromosome	3′ Chromosome
BT-474	GLB1->CMTM7	Intra-chr	−	14	2	3	3
BT-474	ACACA->STAC2	Intra-chr	−	72	21	17	17
BT-474	LAMP1->MCF2L	Intra-chr	+	7	8	13	13
BT-474	RAB22A->MYO9B	Inter-chr	+	18	5	19	20
BT-474	RPS6KB1->SNF8	Intra-chr	+	162	35	17	17
BT-474	STARD3->DOK5	Inter-chr	+	21	13	20	17
BT-474	VAPB->IKZF3	Inter-chr	+	82	6	17	20
BT-474	ZMYND8->CEP250	Intra-chr	−	196	4	20	20
MCF7	ARFGEF2->SULF2	Intra-chr	+	422	156	20	20
MCF7	BCAS4->BCAS3	Inter-chr	+	1750	55	17	20
MCF7	MYO9B->FCHO1	Intra-chr	+	35	5	19	19
MCF7	PAPOLA->AK7	Intra-chr	+	18	11	14	14

or clinical experiments, which was really a limitation in this study. Therefore, further studies are still needed to confirm our findings.

Conclusions

In conclusion, according to the RNA-seq data analysis, *UCP2, CD40*, and *CDH1* might be potential target genes

for the treatment of TN, luminal A, and luminal B breast cancer. In addition, *SFRP1* and *MMP7* might be associated with the TN breast cancer through *wnt*/β-catenin signaling pathway. Besides, *FAS* could be related to the luminal B breast cancer. Moreover, the fusion *ESR1-C6orf97* and *COBRA1-C9orf167* were related to the luminal A breast cancer including MCF7 and ZR-75-1. *VAPB-IKZF3* and *ACACA-STAC2* were related to luminal B breast cancer.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61(2):69–90.
2. Geyer FC, Marchiò C, Reis-Filho JS. The role of molecular analysis in breast cancer. Pathology. 2009;41(1):77–88.
3. Goldhirsch A, Wood W, Coates A, Gelber R, Thürlimann B, Senn H-J. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer. Ann Oncol. 2011;2011:mdr304.
4. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006;295(21):2492–502.
5. Phipps AI, Chlebowski RT, Prentice R, McTiernan A, Stefanick ML, Wactawski-Wende J, Kuller LH, Adams-Campbell LL, Lane D, Vitolins M. Body size, physical activity, and risk of triple-negative and estrogen receptor–positive breast cancer. Cancer Epidemiol Biomarkers Prevent. 2011;20(3):454–63.
6. Millikan RC, Newman B, Tse C-K, Moorman PG, Conway K, Smith LV, Labbok MH, Geradts J, Bensen JT, Jackson S. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat. 2008;109(1):123–39.
7. Phipps AI, Buist DS, Malone KE, Barlow WE, Porter PL, Kerlikowske K, Li CI. Reproductive history and risk of three breast cancer subtypes defined by three biomarkers. Cancer Cause Control. 2011;22(3):399–405.
8. Dignam JJ, Dukic V, Anderson SJ, Mamounas EP, Wickerham DL, Wolmark N. Hazard of recurrence and adjuvant treatment effects over time in lymph node-negative breast cancer. Breast Cancer Res Treat. 2009;116(3):595–602.
9. Aebi S, Sun Z, Braun D, Price K, Castiglione-Gertsch M, Rabaglio M, Gelber R, Crivellari D, Lindtner J, Snyder R. Differential efficacy of three cycles of CMF followed by tamoxifen in patients with ER-positive and ER-negative tumors: long-term follow up on IBCSG Trial IX. Ann Oncol. 2011;22(9):1981–7.
10. Tang G, Shak S, Paik S, Anderson SJ, Costantino JP, Geyer CE Jr, Mamounas EP, Wickerham DL, Wolmark N. Comparison of the prognostic and predictive utilities of the 21-gene recurrence score assay and adjuvant! for women with node-negative, ER-positive breast cancer: results from NSABP B-14 and NSABP B-20. Breast Cancer Res Treat. 2011;127(1):133–42.
11. Nguyen PL, Taghian AG, Katz MS, Niemierko A, Raad RFA, Boon WL, Bellon JR, Wong JS, Smith BL, Harris JR. Breast cancer subtype approximated by estrogen receptor, progesterone receptor, and HER-2 is associated with local and distant recurrence after breast-conserving therapy. J Clin Oncol. 2008;26(14):2373–8.
12. Strom A, Hartman J, Foster JS, Kietz S, Wimalasena J, Gustafsson JA. Estrogen receptor beta inhibits 17beta-estradiol-stimulated proliferation of the breast cancer cell line T47D. Proc Natl Acad Sci USA. 2004;101(6):1566–71.
13. Salazar N, Muñoz D, Kallifatidis G, Singh RK, Jordà M, Lokeshwar BL. The chemokine receptor CXCR7 interacts with EGFR to promote breast cancer cell proliferation. Mol Cancer. 2014;13(1):198.
14. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. Clin Chem. 2009;55(4):641–58.
15. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2010;12(2):87–98.
16. Leinonen R, Sugawara H, Shumway M. The sequence read archive. Nucl Acids Res. 2010:gkq1019.
17. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
18. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genom Biol. 2013;14(4):R36.
19. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protocol. 2012;7(3):562–78.
20. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5.
21. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Methodol). 1995:289–300.
22. Kolde R. Pheatmap: pretty heatmaps. R package version 1.0.8. 2015. https://CRAN.R-project.org/package=pheatmap.
23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25(1):25–9.
24. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33(18):2938–40.
25. Yu G, Wang LG, Han Y, He QY. Cluster Profiler: an R package for comparing biological themes among gene clusters. Omics J Integr Biol. 2012;16(5):284–7.
26. Sherlock G. Gene Ontology: tool for the unification of biology. Can Inst Food Sci Technol J. 2009;22(4):415.
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
28. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010;26(5):589–95.
29. Asmann YW, Hossain A, Necela BM, Middha S, Kalari KR, Sun Z, Chai H-S, Williamson DW, Radisky D, Schroth GP. A novel bioinformatics pipeline for identification and characterization of fusion transcripts in breast cancer and normal cell lines. Nucl Acids Res. 2011;39(15):e100–0.
30. Srinivasan S, Guha M, Avadhani NG. Mitochondrial respiratory defects promote the Warburg effect and cancer progression. Mol Cell Oncol. 2016;3(2):-.
31. Samudio I, Fiegl M, Andreeff M. Mitochondrial uncoupling and the Warburg effect: molecular basis for the reprogramming of cancer cell metabolism. Can Res. 2009;69(6):2163.
32. Chen X, Wang K, Chen J, Guo J, Yin Y, Cai X, Guo X, Wang G, Yang R, Zhu L. In vitro evidence suggests that miR-133a-mediated

regulation of uncoupling protein 2 (UCP2) is an indispensable step in myogenic differentiation. J Biol Chem. 2009;284(8):5362–9.
33. Li W, Nichols K, Nathan CA, Zhao Y. Mitochondrial uncoupling protein 2 is up-regulated in human head and neck, skin, pancreatic, and prostate tumors. Cancer Biomark. 2013;13(5):377–83.
34. Ayyasamy V, Owens KM, Desouki MM, Liang P, Bakin A, Thangaraj K, Buchsbaum DJ, LoBuglio AF, Singh KK. Cellular model of Warburg effect identifies tumor promoting function of UCP2 in breast cancer and its suppression by genipin. PLoS One. 2011;6(9):e24792.
35. Lawson DA, Bhakta NR, Kessenbrock K, et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature. 2015;526(7571):131.
36. Li D-M, Feng Y-M. Signaling mechanism of cell adhesion molecules in breast cancer metastasis: potential therapeutic targets. Breast Cancer Res Treat. 2011;128(1):7–21.
37. Saadatmand S, de Kruijf E, Sajet A, Dekker-Ensink N, van Nes J, Putter H, Smit V, van de Velde C, Liefers G, Kuppen P. Expression of cell adhesion molecules and prognosis in breast cancer. Br J Surg. 2013;100(2):252–60.
38. Tong AW, Papayoti MH, Netto G, Armstrong DT, Ordonez G, Lawson JM, Stone MJ. Growth-inhibitory effects of CD40 ligand (CD154) and its endogenous expression in human breast cancer. Clin Cancer Res. 2001;7(3):691–703.
39. Wong AS, Gumbiner BM. Adhesion-independent mechanism for suppression of tumor cell invasion by E-cadherin. J Cell Biol. 2003;161(6):1191–203.
40. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat Rev Cancer. 2009;9(4):265–73.
41. Hajiabadi FR, Tavassoli M, Hemmati S: The study of STR in E-cadherin gene: a new era of breast cancer research. In: 1st Tabriz International Life Science Conference and 12th Iran Biophysical Chemistry Conference, Tabtiz University of Medical Sciences; 2013.
42. Miller JR. The wnts. Genome Biol. 2002;3(1):1–15.
43. Howe LR, Brown AM. Wnt signaling and breast cancer. Cancer Biol Therap. 2004;3(1):36–41.
44. Bucan V, Mandel K, Bertram C, Lazaridis A, Reimers K, Park-Simon TW, Vogt PM, Hass R. LEF-1 regulates proliferation and MMP-7 transcription in breast cancer cells. Gene Cell. 2012;17(7):559–67.
45. Kan JS, DeLassus GS, D'Souza KG, Hoang S, Aurora R, Eliceiri GL. Modulators of cancer cell invasiveness. J Cell Biochem. 2010;111(4):791–6.
46. Dey N, Young B, Abramovitz M, Bouzyk M, Barwick B, De P, Leyland-Jones B. Differential activation of Wnt-β-catenin pathway in triple negative breast cancer increases MMP7 in a PTEN dependent manner. PLoS One. 2013;8(10):e77425.
47. Martins FC, De S, Almendro V, Gonen M, Park SY, Blum JL, Herlihy W, Ethington G, Schnitt SJ, Tung N, et al. Evolutionary pathways in BRCA1-associated breast tumors. Cancer Discov. 2012;2(6):503–11.
48. Castanos-velez E, Singer G, Stöhr R, Simon R, Sauter G, Leibiger H, Weber E, Hermann BK, Rosenthal A: Loss of SFRP1 is associated with breast cancer progression and poor prognosis in early stage tumors. Int J Oncol. 2004;25:641–9.
49. Hashemi M, Fazaeli A, Ghavami S, Eskandari-Nasab E, Arbabi F, Mashhadi MA, Taheri M, Chaabane W, Jain MV, Łos MJ. Functional polymorphisms of FAS and FASL gene and risk of breast cancer–pilot study of 134 cases. PloS one. 2013;8(1):e53075.
50. Villa-Morales M, Fernández-Piqueras J. Targeting the Fas/FasL signaling pathway in cancer therapy. Exp Opin Therap Target. 2012;16(1):85–101.
51. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet. 2010;42(6):504–7.
52. Stacey SN, Sulem P, Zanon C, Gudjonsson SA, Thorleifsson G, Helgason A, Jonasdottir A, Besenbacher S, Kostic JP, Fackenthal JD. Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. PLoS Genet. 2010;6(7):e1001029.
53. Edgren H, Murumagi A, Kangaspeska S, Nicorici D, Hongisto V, Kleivi K, Rye IH, Nyberg S, Wolf M, Borresen-Dale A-L. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. Genome Biol. 2011;12(1):R6.
54. Aiyar SE, Cho H, Lee J, Li R. Concerted transcriptional regulation by BRCA1 and COBRA1 in breast cancer cells. Int J Biol Sci. 2007;3(7):486.
55. Chajès V, Cambot M, Moreau K, Lenoir GM, Joulin V. Acetyl-CoA carboxylase α is essential to breast cancer cell survival. Cancer Res. 2006;66(10):5287–94.