Identification of Subtype-Specific Prognostic Biomarkers in Breast Cancer

Farinaz Vafadar Esfahani

Nour Danesh Institute of Higher Education

Mansoureh Ghezlou

Islamic Azad University

Marziyeh Mazaheri Moghaddam

Tabriz University of Medical Sciences

Madiheh Mazaheri Moghaddam (✉ madiheh_mazaheri@yahoo.com)

Zanjan University of Medical Sciences

Hossein Hozhabri

University of Tehran

Research Article

Keywords: Breast cancer, Molecular subtypes, Bioinformatics analysis, Differentially expressed genes, Prognosis

Posted Date: February 17th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1330135/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Breast cancer (BC) is a heterogeneous disease divided into four molecular subtypes that display different prognoses. Molecular alterations in breast cells contribute to breast carcinogenesis. However, the molecular characteristics involved in developing BC subtypes remain largely unknown. Further identification of molecular mechanisms involved in different BC subtypes might help improve treatment strategies and prognosis of patients. In the present study, two microarray datasets (GSE57297 and GSE65194) containing four BC subtypes were downloaded from the Gene Expression Omnibus (GEO) database. Comparative analyses were performed to identify specific differentially expressed genes (DEGs) for each BC subtype, as well as overlapped (core) between all subtypes. bc-GenExMiner and Kaplan-Meier plotter databases were respectively utilized to investigate the differential expression and prognostic value of DEGs. A total of 25 DEGs (12 specific and 13 core) were found to be significantly associated with prognosis. We found that a high level of C9orf116 predicted better OS in the Luminal A subtype. Also, increased transcription levels of FAM13A and RASIP1 were associated with shorter OS in the Luminal B subtype. High mRNA expression of PDE7A, and COTL1 conferred longer OS, while elevated expression of TM4SF1 and AREG predicted shorter OS in the TNBC subtype. Increased mRNA levels of GSR, and DAPK2 conferred better OS in the HER2 subtype; however, increased expression levels of HOTAIR, PLEKHG4, and POU6F1 predicted poor OS. For the core DEGs, increased expression of IFI6 and UBE2S predicted shorter OS. Meanwhile, increased mRNA levels of AK5, C2orf40, CNN1, HOXA5, NTRK2, PAMR1, PROS1, SCARA5, SDPR, TGFBR3, and TSHZ2 conferred better OS. In conclusion, this study strictly identified specific genes that may help to select precision prognostic biomarkers in different BC subtypes.

Introduction

Breast cancer (BC) is the most commonly diagnosed cancer globally, with approximately 2.3 million new cases in 2020, accounting for 11.7% of all cancer cases¹. BC is an uncontrolled proliferation of breast cells, beginning in different areas of the breast and leading to malignancy². Based on receptor status and the gene expression pattern, BC is divided into four molecular subtypes including, Luminal A, Luminal B, human epidermal growth factor 2 (HER2), and Triple-negative breast cancer (TNBC). The Luminal A subtype is considered to be mostly estrogen receptor (ER)-positive and has a low histological grade. Luminal-B tumors are defined as being ER-positive and/or progesterone receptor (PR)-positive³. HER2, a tyrosine kinase, is identified to be overexpressed or amplified in about 20-25% of BC patients. The overexpression of this oncoprotein implicates in excessive cell growth, migration, and angiogenesis^{4,5}. TNBC is a type of BC with negative expression of ER, PR, and HER2. TNBC is considered highly invasive and has greater metastatic potential, higher relapse rates, and a worse prognosis⁶. Each subtype displays different biological behavior and prognosis and has distinct treatment strategies⁷. Despite advancements in the therapeutic approaches of BC, due to the high heterogeneity, the effectiveness of treatment methods is still not very satisfactory, leading to a poor prognosis of patients^{8,9}.

It is well documented that the molecular alterations in breast cells implicate in breast carcinogenesis⁷. However, the molecular characteristics and pathways involved in developing BC subtypes are still largely unknown¹⁰. Investigating the abnormally expressed genes in BC might help unravel the possible molecular mechanisms involved¹¹. Whole transcriptome analysis is considered to be useful in identifying differentially expressed genes (DEGs) contributing to BC development and progression^{12,13}. The combination of molecular biology and bioinformatics tools and web servers provides great convenience for identifying potential therapeutic, diagnostic, and prognostic biomarkers^{11,14,15}. The current study aimed to identify core and subtype-specific DEGs in BC and elucidate their prognostic values. Therefore, we initially analyzed DEGs between BC subtypes and normal samples using two datasets obtained from Gene Expression Omnibus (GEO). Subsequently, we conducted bioinformatics analysis to screen and verify key core and subtype-specific genes as potential prognostic biomarkers in BC.

Methods

Data collection

In order to identify DEGs in different subtypes of BC, two gene expression profiles, including GSE57297 (GPL17077 platform) and GSE65194 (GPL570 platform), were retrieved from the GEO database (<http://www.ncbi.nlm.nih.gov/geo>). GEO is an international genomics database in which scientists and researchers submit the data of microarray/gene profiles. The datasets contained 178 human BC tissue samples with different subtypes and 18 normal breast tissue samples. A flowchart of the current study and essential details of the GEO datasets are shown in Fig. 1. All methods were performed in accordance with relevant guidelines and regulations.

Data preprocessing

R software, a free statistical computing and graphics environment, was used to analyze and standardize the raw data¹⁶. Data processing, background adjustment, and quantile normalization of all the raw data files were performed using the robust multi-array average (RMA) algorithm of the Affy package¹⁷.

Identification of DEGs

The linear models for microarray data (limma) R package was applied to identify DEGs between different BC subtypes and normal samples¹⁸. Adjusted P values < 0.05 and $|\log_2 \text{fold change} (\log_2\text{FC})| > 1$ were considered as cut-off values for identifying DEGs. Core and specific DEGs among subtypes were determined and visualized by an R package called VennDiagram¹⁹.

KEGG pathway enrichment analysis

Enrichr (<https://maayanlab.cloud/Enrichr/>) is an online analysis tool for gene functional annotation. We performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of DEGs through the Enrichr database²⁰. $P < 0.05$ was defined as significant.

bc-GenExMiner v4.8

The expression trends of subtype-specific and core DEGs in BC were verified using the Breast Cancer Gene-Expression Miner v4.8 (bcGenExMiner v4.8) (<http://bcgenex.ico.unicancer.fr>) database²¹. This database contains cancer transcriptomic data obtained from The Cancer Genome Atlas (TCGA). The "expression" module of the bcGenExMiner was used to assess the expression of DEGs according to the Nature of the tissue determined by the Prediction Analysis of Microarray 50 (PAM50) test. The Dunnett-Tukey-Kramer's test and Welch's t-test were performed to calculate the *p*-value. A value of $P < 0.05$ was considered to be significant.

Kaplan–Meier plotter

Kaplan-Meier plotter (www.kmplot.com) database, which contains gene expression profiles and survival information of cancer patients, was applied to evaluate the prognostic value of the core and subtype-specific DEGs²². All BC patient samples were divided into high and low expression groups based on the median mRNA expression in order to analyze the overall survival (OS). The log-rank $P < 0.05$ was considered as the cutoff.

cBioPortal

cBioPortal (<https://www.cbioportal.org/>), a web resource, provides data exploration and analysis of cancer genomics data²³. In order to analyze the genomic profile changes of the core and subtype-specific DEGs, the breast invasive carcinoma (TCGA, Firehose legacy) dataset, which included data from 1108 samples were selected. The genomic profile changes parameters included mutations, copy number alterations (CNAs) from GISTIC, and mRNA expression Z scores (microarray).

Results

Identification of DEGs

Two datasets containing 18 normal samples and 178 BC samples were analyzed. Genes with the adjusted P -value < 0.05 and $|\log_2\text{FC}| > 1$ were considered as DEGs. Totally, 1456 (492 up-regulated and 964 down-regulated) and 1599 (1039 up-regulated and 560 down-regulated) DEGs were identified in the Luminal A subtype within GSE57297 and GSE65194, respectively. Moreover, 1079 (510 up-regulated and 569 down-regulated) and 2161 (1420 up-regulated and 741 down-regulated) genes were differentially expressed in the Luminal B subtype within GSE57297 and GSE65194 datasets, respectively. There were 543 (249 up-regulated and 294 down-regulated) and 2791 (1854 up-regulated and 937 down-regulated) DEGs in the TNBC subtype within GSE57297 and GSE65194, respectively. In addition, 2524 DEGs (1643 up-regulated and 881 down-regulated) were identified in the HER2 subtype within the GSE65194 dataset. All the DEGs are listed in Supplementary Table S1. Volcano plots of DEGs are illustrated in Fig. 2a-g. After the integration of above-mentioned DEGs in each subtype between the two datasets and obtaining one DEG profile for each subtype, we combined the DEG profiles of the four subtypes by the Venn diagram to

identify core and specific DEGs among the subtypes. A total of 2253 core and specific DEGs were identified: 31 core DEGs, 44 specific DEGs for the Luminal A subtype, 57 specific DEGs for the Luminal B subtype, 2076 specific DEGs for the HER2 subtype, and 45 specific DEGs for the TNBC subtype (Fig. 2h). The details of the core and subtype-specific DEGs are summarized in the Supplementary Table S2.

Enrichment analysis of DEGs

The KEGG pathway enrichment analysis revealed that DEGs in Luminal A, Luminal B, TNBC, and HER2 samples were mainly enriched in caffeine metabolism, pentose phosphate pathway, peroxisome, and protein processing in the endoplasmic reticulum, respectively (Fig. 3a-d). Moreover, core DEGs were most commonly enriched in protein digestion and absorption (Fig. 3e).

DEGs screening

To validate the accuracy of results and obtain key core and subtype-specific genes, we analyzed the expression of the identified core and specific DEGs using the bc-GenExMiner v4.8. To that end, we screened genes that significantly expressed in each subtype compared to other subtypes as well as normal samples. As a result, 181 genes among 2253 DEGs were verified (Supplementary Table S3). There were 9 specific DEGs for the Luminal A subtype (all down-regulated), 20 specific DEGs for the Luminal B subtype (9 up-regulated and 11 down-regulated), 33 specific DEGs for the TNBC subtype (17 up-regulated and 16 down-regulated), and 90 specific DEGs for the HER2 subtype (63 up-regulated and 27 down-regulated). Meanwhile, 29 core DEGs were identified among the subtypes (4 up-regulated and 25 down-regulated).

Survival analysis

The Kaplan-Meier plotter was used to determine the relationship between the selected DEGs and OS in BC patients. The association between the identified core DEGs and OS was analyzed in all BC patients. However, investigating the association between the expression of specific genes with the OS was restricted to the distinct subtypes, determined by the StGallen and PAM50 criteria. As a result, 36 genes (23 specific and 13 core genes) among 181 DEGs were significantly associated with the OS (Supplementary Table S4). Thereafter, we investigated the correlation between the expression of those 23 subtype-specific genes and OS in other subtypes to exclude unspecific biomarkers. Subsequently, the genes which were significantly associated with OS in other subtypes were excluded to screen genes for each subtype in a stricter manner. Finally, 25 DEGs (12 specific and 13 core) were found to be significantly associated with OS (Table 1). High expression of C9orf116 was related to better OS in the Luminal A subtype. Moreover, elevated expression of FAM13A and RASIP1 was associated with shorter OS in the Luminal B subtype. High mRNA expression of PDE7A, and COTL1 was notably correlated with longer OS, while elevated expression of TM4SF1 and AREG was associated with shorter OS in the TNBC subtype. In addition, increased mRNA levels of GSR, and DAPK2 predicted better OS in the HER2 subtype. On the other hand, high transcription levels of HOTAIR, PLEKHG4, and POU6F1 notably correlated with poor OS in the HER2 subtype. For the core DEGs, high expression of IFI6 and UBE2S was related to shorter OS; however, increased mRNA levels of AK5, C2orf40, CNN1, HOXA5, NTRK2, PAMR1, PROS1,

SCARA5, SDPR, TGFB3, and TSHZ2 were significantly correlated with better OS. Figures 4 and 5 respectively illustrate the expression pattern and prognostic value of final selected core and subtype-specific DEGs.

Table 1

A panel of 25 gene signatures (12 unique genes to each subtype) are served as prognostic biomarkers in breast cancer patients. (Kaplan-Meier plotter)

	Up-regulated	Down-regulated
Luminal A	C9orf116	
Luminal B		FAM13A, RASIP1
Triple-Negative	PDE7A, TM4SF1, COTL1	AREG
HER2	GSR, HOTAIR	DAPK2, PLEKHG4, POU6F1
Core	IFI6 (6-16), UBE2S	AK5, C2orf40 (ECRG4), CNN1, HOXA5, NTRK2, PAMR1, PROS1 (PROS), SCARA5, SDPR (CAVIN2), TGFB3, TSHZ2

Genomic alteration analysis

Genomic alterations of the 13 core and 12 subtype-specific DEGs were analyzed using the cBioPortal database. The results revealed that these genes were altered in 551 (50%) of 1101 BC patients. Notably, the PDE7A gene was mutated in 14% of BC patients (Fig. 6).

Discussion

Breast cancer is considered as a major human health problem with the highest incidence among cancers and an increased mortality rate¹. Therefore, elucidating molecular mechanisms underlying different BC subtypes is inevitable for choosing effective treatment methods and improving the prognosis of patients. Here we found 13 core and 12 subtype-specific DEGs associated with prognosis in BC, potentially indicating the clinical application of these DEGs.

C9orf116, also known as PIERCE1, was discovered to be involved in the cell cycle process in mice²⁴. C9orf116 was found to induce cell proliferation in rat liver cell line BRL-3A²⁵. Moreover, C9orf116 ablation inhibited cell growth in KRAS-mutant non-small cell lung cancer (NSCLC), which might have therapeutic potential in KRAS-mutant NSCLC²⁶. However, the role of C9orf116 in BC is unclear at present. Our analysis showed that C9orf116 was specifically up-regulated in the Luminal A subtype, and this high expression predicted better OS in the Luminal A subtype BC.

We revealed that FAM13A and RASIP1 expressions were specifically down-regulated in the Luminal B subtypes. Also, elevated expression of FAM13A and RASIP1 was correlated with shorter OS in the Luminal B subtype. FAM13A single-nucleotide polymorphism (SNP) rs1059122 was found to be associated with BC risk in the Chinese female population²⁷. It has been demonstrated that RASIP1 is essential for blood vessel formation and angiogenesis²⁸. RASIP1 was identified to play a role in NSCLC metastasis²⁹, whereas its role in BC is unclear.

In the selected specific DEGs for TNBC subtype BC, expression of PDE7A, TM4SF1, COTL1, and AREG was notably associated with prognosis. We found that PDE7A was specifically up-regulated in TNBC. It has been demonstrated that inhibition of PDE3, PDE4, PDE7, or PDE8 suppresses migration in triple-negative MDA-MB-231 cells through stimulating cyclic adenosine monophosphate (cAMP) signaling pathway³⁰. We revealed that high mRNA expression of PDE7A significantly conferred longer OS. Also, we observed that the PDE7A gene was altered in 14% of BC patients. Several lines of evidence indicate that TM4SF1 functions as an oncogene in various cancers^{31–33}. Fan et al. reported that TM4SF1 implicates in BC cell migration and invasion³⁴. Interestingly, the TM4SF1 expression level was identified to be obviously high in the TNBC subtype compared to the other subtypes^{31,34}, which is in conformance with our findings. We identified that elevated expression of TM4SF1 was significantly associated with shorter OS in TNBC patients. COTL1 was reported to suppress BC cell growth³⁵, while in a recent study, COTL1 was identified to be up-regulated in BC tissues, promoting cell migration and tumor metastasis³⁶. We observed that COLT1 was specifically highly expressed in TNBC, and its overexpression conferred longer OS in TNBC patients. It was demonstrated that knock-down of AREG in SUM-149 cells, a TNBC cell line, inhibited cellular invasion in BC³⁷. We found that the expression level of AREG was specifically down-regulated in TNBC tissues. Moreover, we identified that overexpression of AREG was related to the shorter OS in TNBC patients.

In the selected specific DEGs for HER2 subtype BC, expression of GSR, HOTAIR, DAPK2, PLEKHG4, and POU6F1 was notably associated with prognosis. GSR encodes glutathione reductase, providing a protective role against oxidative stress³⁸. GSR was reported to be a potential biomarker that could predict response to treatment in primary ER+ ductal breast carcinoma³⁹. Copy number loss of GSR was identified in primary lung adenocarcinoma and squamous cell carcinoma⁴⁰. Our results showed that GSR was specifically highly expressed in the HER2 subtype. Also, overexpression of GSR was associated with better OS in the HER2 subtype. Compared to normal breast tissues, overexpression of Long noncoding RNA HOTAIR has been detected in BC tissues and cells^{41–44}. A growing body of evidence indicates that HOTAIR might display a role in cell proliferation, migration, invasion, metastasis, and apoptosis in BC^{44–47}. We revealed that HOTAIR expression was specifically up-regulated in the HER2 subtype, which was notably correlated with worse OS in the HER2 subtype. DAPK2, a pro-apoptotic protein, was suppressed in BC cells by miR-520h, which protected BC cells from drug-induced apoptosis⁴⁸. We found that DAPK2 was specifically down-regulated in the HER2 subtype. Moreover, we observed that enhanced expression of DAPK2 conferred favorable OS in the HER2 subtype. PLEKHG4, also known as FLJ00068, was

identified to play an essential role in skeletal muscle and adipocytes^{49,50}, while its role in tumors is unknown. Our results showed that PLEKHG4 was specifically down-regulated in the HER2 subtype. We found that elevated expression of PLEKHG4 was related to poor OS. POU6F1, a transcription factor, was indicated to play a role in stem cell proliferation in HER2 positive BC tissues⁵¹. We revealed that POU6F1 was specifically down-regulated in the HER2 subtype. Additionally, we found that overexpression of POU6F1 predicted a shorter OS in the HER2 subtype.

In the selected core DEGs for BC, expression of IFI6, UBE2S, AK5, C2orf40, CNN1, HOXA5, NTRK2, PAMR1, PROS1, SCARA5, SDPR, TGFBR3, and TSHZ2 was notably associated with prognosis. IFI6, also known as 6-16, is an antiapoptotic protein that enhances the metastatic potential of BC cells via mtROS⁵². Notably, high expression of IFI6 was remarkably associated with shorter OS in ER⁺ BC patients⁵³. Also, we observed that IFI6 was up-regulated in BC patients, and its high expression was associated with poor OS. UBE2S was reported to be highly expressed in BC tissues and associated with BC cell migration and invasion⁵⁴. Similarly, we observed up-regulation of UBE2S in BC tissues, which was significantly correlated with worse OS. AK5 induces proliferation and autophagy and inhibits apoptosis in gastric cancer. Also, AK5 was suggested as a potential prognostic marker as well as a therapeutic target for gastric cancer⁵⁵. AK5 was down-regulated in colorectal cancer due to hypermethylation, and it is involved in cell migration and invasion in colorectal cancer⁵⁶. In a previous study, AK5 was indicated to be aberrantly hypermethylated in BC cells⁵⁷. We identified that all four BC subtypes displayed lower expression of AK5 than normal breast tissues. Moreover, we observed that an increased level of AK5 predicted better OS in BC. The role of AK5 in the development and progression of BC needs further exploration.

It has been demonstrated that HOXA5 mRNA and protein expression was remarkably decreased in the BC tissues and cells^{58–61}. Until now, few studies have reported the tumor-suppressive function of HOXA5 in inducing apoptosis in BC cells through p53-dependent and caspase-dependent pathways^{58,59}. Teo et al. indicated that loss of HOXA5 in mammary cells leads to reduced CDH1 and CD24 expression and consequently results in loss of epithelial phenotypes, enhanced stemness, and acquisition of invasion and migration characteristics⁶². Consistent with previous findings, we observed remarkably decreased HOXA5 mRNA in the BC tissues compared to normal breast tissues. Moreover, we identified that overexpression of HOXA5 conferred better OS in BC.

In line with our findings, CNN1 was identified to be down-regulated in BC tissues and cells compared to normal breast tissues and cells, and it was suggested as a potential gene involved in BC development^{63,64}. Wang et al. clarified that overexpression of CNN1 suppressed proliferation, migration, and invasion of BC cells and promoted cell apoptosis⁶³. Our findings also showed that high expression of CNN1 conferred better OS in BC patients. Accumulating evidence has shown that C2orf40, also known as ECRG4, expression was down-regulated in BC tissues and cells^{64–67}, confirming our results. Additionally, Lu et al. found that promoter hypermethylation leads to a decrease in the expression of C2orf40 in BC. They also indicated that C2orf40 overexpression inhibits cell proliferation, migration, and invasion in BC

cells by blocking cell cycle progression at the M phase through inhibiting UBE2C expression, a mitosis-regulating gene. According to the abovementioned findings, C2orf40 was proposed as a tumor suppressor gene in BC⁶⁶. Notably, higher expression of C2orf40 was identified to be associated with better OS⁶⁷, which is in conformance with our findings. NTRK2 has been reported to be involved in the mechanism underlying invasive breast cancer⁶⁸. In line with our results, it was demonstrated that high expression of NTRK2 was notably related to better OS. Therefore, it was supposed to be a candidate prognostic marker of OS in BC^{68,69}.

PAMR1 expression was reported to be frequently lost in BC tissues and cell lines^{70,71}. In line with these findings, we observed down-regulation of PAMR1 in BC patients compared to normal breast tissues. Additionally, we showed that up-regulation of PAMR1 in BC patients was notably correlated with longer OS. PROS1 acts as a ligand of the TAM receptors. Abnormally high expression of PROS1 has been reported to be involved in tumorigenesis and the development of human cancers, including oral squamous cell carcinoma, glioblastoma, papillary thyroid carcinoma, prostate, and colorectal cancer⁷²⁻⁷⁶. According to the available literature, PROS1 expression and prognostic significance in BC are still unknown. In the current study, we identified that PROS1 expression was decreased in the BC patients compared to normal samples. In addition, we found that high expression of PROS1 was notably associated with longer OS. It was reported that down-regulation of SCARA5 is implicated in BC tumorigenesis via promoter methylation⁷⁷. Similarly, You et al. revealed that BC tissues displayed lower SCARA5 expression than para-carcinoma tissues and significantly exhibited a correlation with histological grade, tumor size, and lymph node metastasis⁷⁸. In line with these findings, we identified that SCARA5 expression was significantly decreased in BC tissues and that conferred worse OS in BC cohorts.

SDPR, also known as CAVIN2, is reported to be down-regulated in BC tissues and cell lines⁷⁹⁻⁸⁵, and its expression level correlated with the tumor stage⁸⁴. It was previously demonstrated that decreased level of SDPR was remarkably associated with poor OS^{83,84}, RFS, and DMFS⁸¹, representing the prognostic potential of SDPR in BC^{81,84}. Dong et al. found a reduction in TGFBR3 expression during BC progression and determined that decreased expression of TGFBR3 was associated with unfavorable RFS. They provided evidence that loss of TGFBR3 expression induces BC invasion and metastasis⁸⁶. He et al. also reported that low expression of TGFBR3 was associated with a worse OS⁸⁷, which is in line with our finding. TSHZ2 was proposed to act as a tumor suppressor in BC⁸⁸, and lower expression of TSHZ2 leads to mammary tumorigenesis⁸⁹. We observed significant down-regulation of TSHZ2 in the four subtypes of BC patients compared to the normal breast samples that was correlated with worse OS.

In summary, we conducted bioinformatics analysis to identify key core and specific genes among the BC subtypes and evaluate their potential prognostic values. Although further experimental studies are required to validate our findings, the current study provides a new perspective into the molecular characteristics of each subtype of BC and might help to select precision prognostic biomarkers in different BC subtypes in the future.

Declarations

Author contributions

H.H. designed and supervised the study. F.V.E., M.G., and Madiheh.M.M. drafted the manuscript and revised it. F.V.E., Marziyeh.M.M., and Madiheh.M.M. implemented the methods, and analyzed the data. All authors participated in improving the writing of the manuscript.

Availability of data and material

Data are available on reasonable request from corresponding authors.

Competing interests

Authors declare that they do not have any conflict of interests.

Data availability

Data are available on reasonable request from corresponding authors.

Ethics approval

Not applicable

References

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
2. Majeed, W. *et al.* Breast cancer: Major risk factors and recent developments in treatment. *Asian Pacific J. Cancer Prev.* **15**, 3353–3358 (2014).
3. Sotiriou, C. & Pusztai, L. Gene-Expression Signatures in Breast Cancer. *N. Engl. J. Med.* **360**, 790–800 (2009).
4. Nasir, A., Holzer, T. R., Chen, M., Man, M. Z. & Schade, A. E. Differential expression of VEGFR2 protein in HER2 positive primary human breast cancer: Potential relevance to anti-angiogenic therapies. *Cancer Cell Int.* **17**, 1–12 (2017).
5. Schlam, I. & Swain, S. M. HER2-positive breast cancer and tyrosine kinase inhibitors: the time is now. *npj Breast Cancer* **7**, 56; 10.1038/s41523-021-00265-1 (2021).
6. Sporikova, Z., Koudelakova, V., Trojanec, R. & Hajduch, M. Genetic Markers in Triple-Negative Breast Cancer. *Clin. Breast Cancer* **18**, e841–e850 (2018).
7. Harbeck, N. *et al.* *Breast cancer*. *Nat. Rev. Dis. Primers* **5**, 66; 10.1038/s41572-019-0111-2 (2019).
8. Baliu-Piqué, M., Pandiella, A. & Ocana, A. Breast Cancer Heterogeneity and Response to Novel Therapeutics. *Cancers (Basel)*. **12**, 3271; 10.3390/cancers12113271 (2020).

9. Esteva, F. J., Hubbard-Lucey, V. M., Tang, J. & Pusztai, L. Immunotherapy and targeted therapy combinations in metastatic breast cancer. *Lancet Oncol.* **20**, e175–e186 (2019).
10. Li, X. *et al.* Uncovering the Subtype-Specific Molecular Characteristics of Breast Cancer by Multiomics Analysis of Prognosis-Associated Genes, Driver Genes, Signaling Pathways, and Immune Activity. *Front. Cell Dev. Biol.* **9**, 1–13 (2021).
11. Sun, N. *et al.* Screening and Identification of Key Common and Specific Genes and Their Prognostic Roles in Different Molecular Subtypes of Breast Cancer. *Front. Mol. Biosci.* **8**, 1–14 (2021).
12. Kim, G. E., Kim, N. I., Lee, J. S., Park, M. H. & Kang, K. Differentially expressed genes in matched normal, cancer, and lymph node metastases predict clinical outcomes in patients with breast cancer. *Appl. Immunohistochem. Mol. Morphol.* **28**, 111–122 (2020).
13. Hozhabri, H. *et al.* Comparative analysis of protein-protein interaction networks in metastatic breast cancer. *PLoS One* **17**, e0260584; 10.1371/journal.pone.0260584 (2022).
14. Zheng, H. *et al.* Comprehensive Review of Web Servers and Bioinformatics Tools for Cancer Prognosis Analysis. *Front. Oncol.* **10**, 68; 10.3389/fonc.2020.00068 (2020).
15. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10869–10874 (2001).
16. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, <http://www.R-project.org/> (2013).
17. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
18. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47; 10.1093/nar/gkv007 (2015).
19. hanbo chen & paul c Boutros. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **57**, 35; 10.1186/1471-2105-12-35 (2011).
20. Kuleshov, M. V *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-W97 (2016).
21. Jézéquel, P. *et al.* Bc-GenExMiner 4.5: New mining module computes breast cancer differential gene expression analyses. *Database* **2021**, 1–10 (2021).
22. Györfy, B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Comput. Struct. Biotechnol. J.* **19**, 4101–4109 (2021).
23. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, 269; 10.1126/scisignal.2004088 (2013).
24. Sung, Y. H., Kim, H. J. & Lee, H. W. Identification of a novel Rb-regulated gene associated with the cell cycle. *Mol. Cells* **24**, 409–415 (2007).
25. Zhang, C. *et al.* The novel protein C9orf116 promotes rat liver cell line BRL-3A proliferation. *PLoS One* **12**, 8–10 (2017).

26. Roh, J. il *et al.* Impaired AKT signaling and lung tumorigenesis by PIERCE1 ablation in KRAS-mutant non-small cell lung cancer. *Oncogene* **39**, 5876–5887 (2020).
27. Wei, Y. *et al.* Role of Polymorphisms of FAM13A, PHLDB1, and CYP24A1 in Breast Cancer Risk. *Curr. Mol. Med.* **19**, 579–588 (2019).
28. Xu, K., Chong, D. C., Rankin, S. A., Zorn, A. M. & Cleaver, O. Rasip1 is required for endothelial cell motility, angiogenesis and vessel formation. *Dev. Biol.* **329**, 269–279 (2009).
29. Chen, Y. *et al.* Rasip1 is a RUNX1 target gene and promotes migration of NSCLC cells. *Cancer Manag. Res.* **10**, 4537–4552 (2018).
30. Dong, H., Claffey, K. P., Brocke, S. & Epstein, P. M. Inhibition of breast cancer cell migration by activation of cAMP signaling. *Breast Cancer Res. Treat.* **152**, 17–28 (2015).
31. Xing, P. *et al.* Upregulation of transmembrane 4 L6 family member 1 predicts poor prognosis in invasive breast cancer. *Medicine (Baltimore)* **96**, 4–9 (2017).
32. Cao, J. *et al.* TM4SF1 regulates pancreatic cancer migration and invasion in vitro and in vivo. *Cell. Physiol. Biochem.* **39**, 740–750 (2016).
33. Ye, L. *et al.* Transmembrane-4 L-six family member-1 (TM4SF1) promotes non-small cell lung cancer proliferation, invasion and chemo-resistance through regulating the DDR1/Akt/ERK-mTOR axis. *Respir. Res.* **20**, 1–10 (2019).
34. Fan, C., Liu, N., Zheng, D., Du, J. & Wang, K. MicroRNA-206 inhibits metastasis of triple-negative breast cancer by targeting transmembrane 4 l6 family member 1. *Cancer Manag. Res.* **11**, 6755–6764 (2019).
35. Xia, L. *et al.* Coactosin-like protein CLP/Cotl1 suppresses breast cancer growth through activation of IL-24/PERP and inhibition of non-canonical TGF β signaling. *Oncogene* **37**, 323–331 (2018).
36. Pei, B. *et al.* Downregulation of microRNA-30c-5p was responsible for cell migration and tumor metastasis via COTL1-mediated microfilament arrangement in breast cancer. *Gland Surg.* **9**, 747–758 (2020).
37. Baillo, A., Giroux, C. & Ethier, S. P. Knock-down of amphiregulin inhibits cellular invasion in inflammatory breast cancer. *J. Cell. Physiol.* **226**, 2691–2701 (2011).
38. Kim, S. J. *et al.* Glutathione reductase plays an anti-apoptotic role against oxidative stress in human hepatoma cells. *Biochimie* **92**, 927–932 (2010).
39. Albogami, S. M., Asiri, Y., Asiri, A., Alnefaie, A. A. & Alnefaie, S. Effects of neoadjuvant therapies on genetic regulation of targeted pathways in ER+ primary ductal breast carcinoma: A meta-analysis of microarray datasets. *Saudi Pharm. J.* **29**, 656–669 (2021).
40. Baity, M. *et al.* Glutathione reductase (GSR) gene deletion and chromosome 8 aneuploidy in primary lung cancers detected by fluorescence in situ hybridization. *Am. J. Cancer Res.* **9**, 1201–1211 (2019).
41. Hu, X., Ding, D., Zhang, J. & Cui, J. Knockdown of lncRNA HOTAIR sensitizes breast cancer cells to ionizing radiation through activating miR-218. *Biosci. Rep.* **29**, 1–9 (2019).

42. Wu, D., Zhu, J., Fu, Y., Li, C. & Wu, B. LncRNA HOTAIR promotes breast cancer progression through regulating the miR-129-5p/FZD7 axis. *Cancer Biomarkers* **30**, 203–212 (2021).
43. Zhao, W., Geng, D., Li, S., Chen, Z. & Sun, M. LncRNA HOTAIR influences cell growth, migration, invasion, and apoptosis via the miR-20a-5p/HMGA2 axis in breast cancer. *Cancer Med.* **7**, 842–855 (2018).
44. Yu, Y. *et al.* HOTAIR may regulate proliferation, apoptosis, migration and invasion of MCF-7 cells through regulating the P53/Akt/JNK signaling pathway. *Biomed. Pharmacother.* **90**, 555–561 (2017).
45. Wang, Y. *et al.* Long noncoding RNA HOTAIR promotes breast cancer development by targeting ZEB1 via sponging miR-601. *Cancer Cell Int.* **20**, 1–13 (2020).
46. Ding, W., Ren, J., Ren, H. & Wang, D. Long Noncoding RNA HOTAIR Modulates MiR-206-mediated Bcl-w Signaling to Facilitate Cell Proliferation in Breast Cancer. *Sci. Rep.* **7**, 1–9 (2017).
47. Zhang, H. *et al.* MiR-7, inhibited indirectly by LincRNA HOTAIR, directly inhibits SETDB1 and reverses the EMT of breast cancer stem cells by downregulating the STAT3 pathway. *Stem Cells* **32**, 2858–2868 (2014).
48. Su, C. M. *et al.* MiR-520h is crucial for DAPK2 regulation and breast cancer progression. *Oncogene* **35**, 1134–1142 (2016).
49. Takenaka, N., Nakao, M., Hasegawa, K., Chan, M. P. & Satoh, T. The guanine nucleotide exchange factor FLJ00068 activates Rac1 in adipocyte insulin signaling. *FEBS Lett.* **594**, 4370–4380 (2020).
50. Machin, P. A., Tsonou, E., Hornigold, D. C. & Welch, H. C. E. Rho family gtpases and rho gefs in glucose homeostasis. *Cells* **10**, 915; 10.3390/cells10040915 (2021).
51. Rahmatpanah, F. B. *et al.* A class of genes in the HER2 regulon that is poised for transcription in breast cancer cell lines and expressed in human breast tumors. *Oncotarget* **6**, 1286–1301 (2015).
52. Cheriya, V. *et al.* G1P3 (IFI6), a mitochondrial localised antiapoptotic protein, promotes metastatic potential of breast cancer cells through mtROS. *Br. J. Cancer* **119**, 52–64 (2018).
53. Cheriya, V. *et al.* G1P3, an interferon- and estrogen-induced survival protein contributes to hyperplasia, tamoxifen resistance and poor outcomes in breast cancer. *Oncogene* **31**, 2222–2236 (2012).
54. Akter, K. A. *et al.* Erratum to: UBE2S is associated with malignant characteristics of breast cancer cells. *Tumor Biol.* **37**, 6999, 10.1007/s13277-015-3863-7 (2016).
55. Zhang, L. H. *et al.* AK5, a novel prognosis marker, inhibits apoptosis and promotes autophagy as well as proliferation in human gastric cancer. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 9900–9906 (2019).
56. Ahn, B. *et al.* Identification of novel DNA hypermethylation of the adenylate kinase 5 promoter in colorectal adenocarcinoma. *Sci. Rep.* **11**, 1–13 (2021).
57. Miyamoto, K. *et al.* Identification of 20 genes aberrantly methylated in human breast cancers. *Int. J. Cancer* **116**, 407–414 (2005).
58. Raman, V. *et al.* Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nature* **405**, 974–978 (2000).

59. Chen, H., Chung, S. & Sukumar, S. HOXA5-Induced Apoptosis in Breast Cancer Cells Is Mediated by Caspases 2 and 8. *Mol. Cell. Biol.* **24**, 924–935 (2004).
60. Henderson, G. S., Van Diest, P. J., Burger, H., Russo, J. & Raman, V. Expression pattern of a homeotic gene, HOXA5, in normal breast and in breast tumors. *Cell. Oncol.* **28**, 305–313 (2006).
61. Wang, Y. *et al.* miR-224-5p Carried by Human Umbilical Cord Mesenchymal Stem Cells-Derived Exosomes Regulates Autophagy in Breast Cancer Cells via HOXA5. *Front. Cell Dev. Biol.* **9**, 1–12 (2021).
62. Teo, W. W. *et al.* HOXA5 determines cell fate transition and impedes tumor initiation and progression in breast cancer through regulation of E-cadherin and CD24. *Oncogene* **35**, 5539–5551 (2016).
63. Wang, Z., Li, T. E., Chen, M., Pan, J. J. & Shen, K. W. miR-106b-5p contributes to the lung metastasis of breast cancer via targeting CNN1 and regulating Rho/ROCK1 pathway. *Aging (Albany. NY)*. **12**, 1867–1887 (2020).
64. Hou, L. *et al.* Systematic analyses of key genes and pathways in the development of invasive breast cancer. *Gene* **593**, 1–12 (2016).
65. Li, C., Zhang, P., Jiang, A., Mao, J. H. & Wei, G. A short synthetic peptide fragment of human C2ORF40 has therapeutic potential in breast cancer. *Oncotarget* **8**, 41963–41974 (2017).
66. Lu, J. *et al.* C2ORF40 suppresses breast cancer cell proliferation and invasion through modulating expression of M phase cell cycle genes. *Epigenetics* **8**, 571–83 (2013).
67. Sabatier, R. *et al.* Down-regulation of ECRG4, a candidate tumor suppressor gene, in human breast cancer. *PLoS One* **6**, 2–9 (2011).
68. Gao, C. *et al.* The construction and analysis of ceRNA networks in invasive breast cancer: A study based on the cancer genome atlas. *Cancer Manag. Res.* **11**, 1–11 (2019).
69. Wang, J. J. *et al.* Comprehensive analysis of the lncRNA-associated competing endogenous RNA network in breast cancer. *Oncol. Rep.* **42**, 2572–2582 (2019).
70. Makoukji, J. *et al.* Gene expression profiling of breast cancer in Lebanese women. *Sci. Rep.* **6**, 1–13 (2016).
71. Lo, P. H. Y., Tanikawa, C., Katagiri, T., Nakamura, Y. & Matsuda, K. Identification of novel epigenetically inactivated gene PAMR1 in breast carcinoma. *Oncol. Rep.* **33**, 267–273 (2015).
72. Abboud-Jarrous, G. *et al.* Protein S drives oral squamous cell carcinoma tumorigenicity through regulation of AXL. *Oncotarget* **8**, 13986–14002 (2017).
73. Che Mat, M. F. *et al.* Silencing of PROS1 induces apoptosis and inhibits migration and invasion of glioblastoma multiforme cells. *Int. J. Oncol.* **49**, 2359–2366 (2016).
74. Sierko, E., Wojtukiewicz, M. Z., Zawadzki, R., Zimnoch, L. & Kisiel, W. Expression of protein C (PC), protein S (PS) and thrombomodulin (TM) in human colorectal cancer. *Thromb. Res.* **125**, e71–e75 (2010).
75. Saraon, P. *et al.* Proteomic profiling of androgen-independent prostate cancer cell lines reveals a role for protein S during the development of high grade and castration-resistant prostate cancer. *J. Biol.*

- Chem. **287**, 34019–34031 (2012).
76. Wang, J., Lei, M. & Xu, Z. Aberrant expression of PROS1 correlates with human papillary thyroid cancer progression. *PeerJ* **9**, e11813; 10.7717/peerj.11813 (2021).
77. Ulker, D., Ersoy, Y. E., Gucin, Z., Muslumanoglu, M. & Buyru, N. Downregulation of SCARA5 may contribute to breast cancer via promoter hypermethylation. *Gene* **673**, 102–106 (2018).
78. You, K. *et al.* SCARA5 plays a critical role in the progression and metastasis of breast cancer by inactivating the ERK1/2, STAT3, and AKT signaling pathways. *Mol. Cell. Biochem.* **435**, 47–58 (2017).
79. Bai, L. *et al.* Down-regulation of the cavin family proteins in breast cancer. *J. Cell. Biochem.* **113**, 322–328 (2012).
80. Li, X. *et al.* Coordinate suppression of Sdpr and Fhl1 expression in tumors of the breast, kidney, and prostate. *Cancer Sci.* **99**, 1326–1333 (2008).
81. Ozturk, S. *et al.* SDPR functions as a metastasis suppressor in breast cancer by promoting apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 638–643 (2016).
82. Li, Y. & Wu, C. LINC00261/microRNA-550a-3p/SDPR axis affects the biological characteristics of breast cancer stem cells. *IUBMB Life* **73**, 188–201 (2021).
83. Wang, Y., Song, Z., Leng, P. & Liu, Y. A systematic analysis reveals gene expression alteration of serum deprivation response (SDPR) gene is significantly associated with the survival of patients with cancer. *Oncol. Rep.* **42**, 1161–1172 (2019).
84. Tian, Y., Liu, X., Hu, J., Zhang, H. & Wang, B. Integrated Bioinformatic Analysis of the Expression and Prognosis of Caveolae-Related Genes in Human Breast Cancer. **11**, 1–16 (2021).
85. Tian, Y. *et al.* Serum deprivation response inhibits breast cancer progression by blocking transforming growth factor- β signaling. *Cancer Sci.* **107**, 274–280 (2016).
86. Dong, M. *et al.* The type III TGF- β receptor suppresses breast cancer progression. *J. Clin. Invest.* **117**, 206–217 (2007).
87. He, Y. *et al.* Identification of Hub Genes to Regulate Breast Cancer Spinal Metastases by Bioinformatics Analyses. *Comput. Math. Methods Med.* **9**, e11813; 10.7717/peerj.11813 (2021).
88. Yamamoto, M., Cid, E., Bru, S. & Yamamoto, F. Rare and frequent promoter methylation, respectively, of TSHZ2 and 3 genes that are both downregulated in expression in breast and prostate cancers. *PLoS One* **6**, 1–10 (2011).
89. Riku, M. *et al.* Down-regulation of the zinc-finger homeobox protein TSHZ2 releases GLI1 from the nuclear repressor complex to restore its transcriptional activity during mammary tumorigenesis. *Oncotarget* **7**, 5690–5701 (2016).

Figures

Figure 1

A flowchart of the current study.

Figure 2

Identification of DEGs between different BC subtypes and normal tissues. Panels a-g illustrate volcano plots of DEGs for Luminal A within GSE57297 (**a**) and GSE65194 (**b**), Luminal B within GSE57297 (**c**) and GSE65194 (**d**), TNBC within GSE57297 (**e**) and GSE65194 (**f**), and HER2 within GSE65194 (**g**). The differential gene expression profile is represented according to the negative logarithm to the base 10 of the Adj. *P*-value and $\log_2$ FC. The red color dots denote DEGs. The black dots represent genes that were

not differentially expressed. The adjusted P -value <0.05 and $|\log_2FC| > 1$ were considered as cut-off criteria. Panel h shows the Venn diagram of the overlapping and specific DEGs among the subtypes.

Figure 3

KEGG pathway enrichment analysis of DEGs. Panels a-e illustrate KEGG enrichment scatter plots of DEGs in Luminal A (a), Luminal B (b), TNBC (c), and HER2 (d) subtypes, and core DEGs (e) in BC patients. A value of $P < 0.05$ was set as a threshold.

Figure 4

The expression of 13 core and 12 specific DEGs in each subtype (bcGenExMiner v4.8). Panels a-y illustrate the expression level of specific DEGs for the Luminal A (a), Luminal B (b, c), TNBC (d-g), and HER2 (h-l) subtypes and core DEGs (m-y). The Dunnett-Tukey-Kramer's test and Welch's t-test were used to compare mRNA expression differences between groups. C2orf40, namely ECRG4. SDPR, namely CAVIN2. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

Figure 5

The correlation between the expression of 12 subtype-specific genes and OS in Luminal A (a), Luminal B (b, c), TNBC (d-g), and Her2 (h-l). The association of mRNA expression of 13 core DEGs with OS (m-y) in BC patients. IFI6, namely 6-16. PROS1, namely PROS.

Figure 6

Genomic alteration of core and subtype-specific DEGs in BC patients (cBioPortal).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS1.xlsx](#)
- [SupplementaryTableS2.xlsx](#)
- [SupplementaryTableS34.docx](#)