69987_Auto_Edited-check.docx

Clinical and Translational Research

Eight hub genes as potential biomarkers for breast cancer diagnosis and prognosis: A TCGA-based study

Liu N *et al*. Hub genes as potential biomarkers for breast cancer

Nan Liu, Guo-Duo Zhang, Ping Bai, Li Su, Hao Tian, Miao He

Abstract

BACKGROUND

Breast cancer (BC) is the most common malignant tumor in women.

AIM

To investigate BC associated hub genes for better understanding BC tumorigenesis

METHODS

1203 BC samples were downloaded from The Cancer Genome Atlas database, which include 113 normal samples and 1090 tumor samples. The limma package of R software was used to analyze the differentially expressed genes (DEGs) in tumor tissues compared with the normal tissues. The cluster Profiler package was used to perform the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of up-regulated genes and down-regulated genes, respectively. Univariate COX regression was conducted to explore the DEGs which process the statistical significance with prognosis. The protein-protein interaction (PPI) network analysis was employed to investigate the hub-genes through cytohubba algorithm by the Cytoscape software.

Survival analysis of the hub-genes were carried out through the Kaplan-Meier database. The expression level of these hub-genes was validated in the Gene Expression Profiling Interactive Analysis database and the Human Protein Atlas database, respectively.

RESULTS

A total of 1317 DEGs (fold change > 2, P < 0.01) were confirmed through bioinformatics analysis, which include 744 up-regulated and 573 down-regulated genes in BC samples. KEGG enrichment analysis indicated that up-regulated genes mainly enriched in cytokine-cytokine receptor interaction, cell cycle and P53 signaling pathway (P < 0.01), besides, the down-regulated genes were mainly enriched in cytokine-cytokine receptor interaction, PPAR signaling pathway and AMPK signaling pathway (P < 0.01).

CONCLUSION

In view of the results of PPI analysis which verified by survival and expression analysis, we conclude that *MAD2L1*, *PLK1*, *SAA1*, *CCNB1*, *SHCBP1*, *KIF4A*, *ANLN* and *ERCC6L* may act as biomarkers for diagnosis and prognosis in BC patients.

Key Words: Breast cancer; Bioinformatics; Hub gene; The Cancer Genome Atlas

Core Tip: This study identified 1317 DEGs related to the occurrence and development of breast cancer (BC), 165 DEGs related to prognosis, and 8 hub genes (*MAD2L1*, *PLK1*, *SAA1*, *CCNB1*, *SHCBP1*, *KIF4A*, *ANLN* and *ERCC6L*). Each of these eight hub genes has different expression in BC and is significantly related to prognosis. The results of this study indicate that the study of these DEGs would help us for full understanding on the molecular mechanisms for BC of pathogenesis and progression. Moreover, these hub

genes may serve as potential prognostic markers and therapeutic targets, which provides reference for more in-depth and extensive prospective clinical research.

INTRODUCTION

Breast cancer (BC) is the most common malignant tumor in women. In 2019, 268600 new BC patients and 41760 new BC deaths were reported, which accounting for 30% of all new cancer cases and 15% of cancer-related deaths respectively. And the mortality of BC is second only to lung cancer[1]. In recent years, the BC outcome has been significantly improved, and treatment strategies such as surgery, chemotherapy, radiotherapy, endocrine therapy and targeted therapy have achieved fine clinical benefits[2], while patients with distant metastases are almost incurable[3]. In addition, even after resection of the primary tumor, 30% of early BC is prone to recurrence in distant organs[4]. In clinical practice, treatment and prognosis of different molecular subtypes of BC are significantly different: estrogen receptor positive (ER +) patients preferred endocrine therapy, human epidermal growth factor receptor-2 positive (HER2 +) patients preferred targeted therapy, and poorly differentiated tumors are usually associated with poor prognosis[5-7].

Recent studies have found that the occurrence and development of BC are related to many molecular markers. For example, the expression of CD82 is significantly decreased in BC and is associated with disease progression and metastasis[8]. In addition, a study on triple negative BC suggests that multiple lncRNA are associated with prognosis, including MAGI2-AS3, GGTA1P, NAP1L2, CRABP2, SYNPO2, MKI67 and COL4A6[9]. Advances in microarray and high-throughput sequencing technology provide strong support for the development of more reliable prognostic markers[10,11]. Genome wide expression profiling can reveal the molecular changes in the process of tumorigenesis and development, and has proved to be an efficient method to identify key genes[12]. Therefore, it is particularly important to explore more sensitive and specific biomarkers for further understanding the pathogenesis of BC and the choice of treatment strategies. This public database-based study aims to explore the potential hub

genes in the occurrence and development of BC through bioinformatics analysis of gene expression profile and clinical characteristics of BC, so as to provide new biological targets and directions for clinical diagnosis and treatment of BC.

MATERIALS AND METHODS

Data sources and processing

The Cancer Genome Atlas (TCGA) database is a cancer research project established by National Cancer Institute and National Human Genome Research Institute. It aims to understand the mechanism of carcinogenesis and development of cancer cells and develop new diagnosis and treatment methods by collecting various kinds of cancer-related omics data. In this study, 1203 breast samples (Fragments Per Kilobase Million, FPKM format) were downloaded from TCGA database (https://portal.gdc.cancer.gov/), including 1090 tumor samples and 113 normal samples. For more accurate compare of gene expression, FPKM format data was converted to TPM (Transcripts Per Million). At the same time, 1097 tumor samples containing clinical information were downloaded, and the data that did not match the expression samples were excluded. The remaining 1089 tumor samples were included in the univariate Cox regression analysis. The overall survival (OS) was taken as the endpoint event, and the gene expression in TPM format was converted into $\log_2 (x + 1)$.

DEGs acquisition

Limma package of R software (version 3.6.3) was employed for differential gene analysis[13], using the adjusted P-value (adj P-value) can avoid false-positive results. The inclusion criteria of DEGs were: | log2 fold change (FC) | > 2 and the adjusted P value < 0.01. The ggplot2 package of R software was used to generate volcano plot to visualize these differential genes.

Functional enrichment analysis

DEGs were converted into gene ID through org.Hs.eg.db package of R software, and then KEGG enrichment analysis was carried out by R software's clusterProfiler and enrichlot program package. ggplot2 program package was used to display the top ten enrichment items, adjusted P value < 0.05 was considered to be statistically significant.

Univariate Cox regression analysis

The survival package of R software was used to carry out univariate Cox regression analysis on 1089 BC samples with survival information. The median value of expression was set as the cut-off point between the high expression group and low expression group, and the differential genes related to prognosis were obtained for subsequent analysis. P < 0.05 was considered to be statistically significant.

Construction of PPI

STRING database (https://string-db.org/) is a search tool for searching interacting genes, which aims to construct protein-protein interaction (PPI) networks of different genes based on known and predicted PPI, and analyze the proteins that interact with each other[14]. Based on the online tool STRING, PPI of prognosis related DEGs was constructed, and the confidence score ≥ 0.4. Then, PPI network is visualized by Cytoscape software (version 3.7.2). In addition, using the CytoHubba plug-in of Cytoscape software to calculate the gene degree through the "degree" method, the top ten genes are taken as the hub genes for subsequent analysis and verification.

Survival analysis of hub genes

Kaplan-Meier plotter (http://kmplot.com/analysis/) can use 18,674 cancer samples to evaluate the impact of 54675 genes on survival. These studies included the recurrence-free survival and overall survival information of 5143 cases of BC, 1816 cases of ovarian cancer, 2437 cases of lung cancer, 1065 cases of gastric cancer, and 364 cases of liver cancer, which are mainly based on GEO, TCGA and EGA databases. The role of the tool is to benefit patients in clinical decision-making, health care policy and resource allocation through meta-analysis of biomarker assessment[16]. In this study, we analyzed

the overall survival rate of 10 hub genes in BC through Kaplan Meier plotter. According to the median expression of each hub gene in Kaplan Meier plotter, the patients were divided into two groups to present the difference of survival probability between the high expression group and the low expression group. A total of 14 dataset were enrolled in our analysis according to the Kaplan-Meier webtool and the detailed clinical information could retrospective in http://kmplot.com/analysis/. $P < 0.05$ was considered to be statistically significant.

In addition, to further investigate the prognostic value of hub genes selected above, we performed the log-rank test of these hub genes in molecular subtypes of BC based on TCGA cohort. Through PAM50 algorithm, the TCGA cohort were separated into five major subtypes, named Luminal A, Luminal B, Her-2 enriched, Basal-like and Normal-like. This method was completed through utilizing the "genefu" R package according to detailed operation protocol.

Expression analysis of hub genes

Gene Expression Profiling Interactive Analysis (GEPIA) database was employed to verify the mRNA expression levels of 10 hub genes in normal breast tissues and cancer tissues. GEPIA database contains data from 9736 tumor samples and 8587 normal samples, which were used to display the mRNA expression levels of each key gene in cancer and non-cancer tissues[17]. The protein expression levels of 10 hub genes in human normal tissues and BC tissues were analyzed using the human protein atlas database (HPA), which contains immunohistochemical expression data covering about 20 of the most common types of cancer[18].

RESULTS

Identification and functional analysis of DEGs

After DEGs analysis of 113 normal breast samples and 1090 BC samples, we found that there were 1317 DEGs, among which 744 genes were up-regulated and 573 genes were down regulated in BC. As shown in Figure1A, red represents high expression and blue

represents low expression. At the same time, volcano plot was used to present the distribution of DEGs (Figure1B), the red dots represent up-regulated genes and the blue dots represent down regulated genes.

In order to further understand the biological function of these 1317 DEGs, the clusterProfiler and enrichplot packages of R software were used to perform KEGG enrichment analysis on these DEGs.

The enrichment analysis results of up-regulated genes and down-regulated genes are shown in Figure 1C and D, respectively. The top ten up-regulated genes were Cytokine-cytokine receptor interaction, Neuroactive ligand-receptor interaction, Cell cycle, Oocyte meiosis, IL-17 signaling pathway, Cellular senescence, Progesterone-mediated oocyte maturation, P53 signaling pathway, Nicotine addiction and Bladder cancer. And the top ten down regulated genes were Cytokine-cytokine receptor interaction, PPAR signaling pathway, AMPK signaling pathway, Retinol metabolism, Tyrosine metabolism, Adipocytokine signaling pathway, Drug metabolism - cytochrome P450, ABC transporters, Regulation of lipolysis in adipocytes and Fatty acid degradation.

Screening of hub genes

In order to screen the differential genes related to the prognosis of BC, we used the survival package of R software to perform univariate Cox regression analysis on 1317 differential genes, and found that the prognosis of 165 genes was statistically significant (Supplementary Table 1). Further analysis of the PPI of these 165 genes revealed that there are a total of 164 nodes and 156 interactions (edges), and the confidence score adopted default value ≥ 0.4. The CytoHubba algorithm of Cytoscape software is used to calculate the degree scoreof each node. The top ten genes are MAD2L1, PLK1, SAA1, CCNB1, SHCBP1, KIF4A, ANLN, ERCC6L, CXCL2 and WT1 (Figure 3). The up-regulated genes were represented by red and round nodes, and the down regulated genes were represented by blue and diamond nodes. The node size represented the level, and most of the hub genes were up-regulated DEGs. Gene annotation and grade scores are shown in Table 1.

Survival analysis of hub genes

Kaplan - Meier plotter was used to explore the prognostic value of 10 hub genes in BC. The results showed that, except for CXCL2 (HR 0.86 [0.69-1.07], P = 0.170) and WT1 (HR 1.03 [0.83-1.28], P = 0.760), the highly expressed MAD2L1 (HR 2.02 [1.62-2.51], P = 1.8e-10), PLK1 (HR 1.42 [1.15-1.76], P = 0.0012), CCNB1 (HR 1.42 [1.04-1.94], P = 0.028]), SHCBP1 (HR 1.76 [1.42-2.19], P = 2.1 e-07), KIF4A (HR 1.8 [1.44-2.23], P = 8.8e-08]), ANLN (HR 1.48 [1.08-2.03], P = 0.014) and ERCC6L (HR 1.68 [1.35-2.09], P = 2e-06) were related to the poor overall survival rate of BC patients. In contrast, the high expression of SAA1 (HR 0.71 [0.57-0.88], P = 0.018) was associated with a better overall survival rate for BC patients (Figure 4).

Besides, we also conducted the survival analysis of these ten hub genes in the TCGA molecular subtypes. As a result, the TCGA cohort was successfully divided into five subtypes based PAM50 identifier, in which contains 563 of Luminal A, 215 of Luminal B, 82 of Her-2 enriched, 189 of Basal-like and 39 of Normal-like. And then, the survival analysis of these ten genes were performed in each subtype group. The results indicated that, CXCL2 (HR = 0.45, P < 0.05) and SAA1 (HR = 0.53, P < 0.05) were protective factors in Luminal A subtype (Figure 5). ANLN (HR = 2.12, P < 0.05), ERCC6L (HR = 3.04, P < 0.05), KIF4A (HR = 2.50, P < 0.05), PLK1 (HR = 2.40, P < 0.05) and SHCBP1 (HR = 2.42, P < 0.05) were hazard factors in Luminal B subtype, while the CXCL2 (HR = 0.45, P < 0.05) showed protective effect. At last, KIF4A (HR = 4.31, P < 0.05) act as a risk factor in Her-2 enriched patients and the CXCL2 played a satisfactory role among Basal-like patients (HR = 0.46, P < 0.05).

Expression analysis of hub genes

In order to verify the expression differences of key genes in BC, GEPIA was employed to analyze the mRNA expression levels of MAD2L1, PLK1, SAA1, CCNB1, SHCBP1, KIF4A, ANLN, ERCC6L, CXCL2 and WT1 between BC and non-cancerous tissues (Figure 5). Compared with non-cancerous tissues, MAD2L1 (Figure 5A), PLK1 (Figure

5B), CCNB1 (Figure 5D), SHCBP1 (Figure 5E), KIF4A (Figure 5F), ANLN (Figure 5G) and ERCC6L (Figure 5H) in BC tissues were significantly up-regulated ($P < 0.01$); SAA1 (Figure 5C) and CXCL2 (Figure 5I) were significantly down-regulated in BC ($P < 0.01$); and WT1 (Figure 5J) tended to increase in BC tissues. After verifying the mRNA expression level of hub genes, we used HPA database to verify the protein expression level of these hub genes in BC. It is worth noting that MAD2L1 (Figure 6A), PLK1 (Figure 6B), CCNB1 (Figure 6C), SHCBP1 (Figure 6D), ANLN (Figure 6F), ERCC6L (Figure 6G) and WT1 (Figure 6H) were not expressed in normal breast tissues, but expressed in different levels in BC tissues. Besides, KIF4A was found to be moderately expressed in normal breast tissues and highly expressed in BC tissues. In short, the expressions of hub genes were consistent with the results of differential analysis at both mRNA and protein levels.

DISCUSSION

In this study, we used bioinformatics analysis to screen and verify potential biomarkers associated with BC. After comparing the gene expression matrix of breast tissue retrieved from the TCGA database, 744 up-regulated DEGs and 573 down-regulated DEGs were successfully identified. Combined with the survival data, 165 prognostic-related DEGs were analyzed. And according to PPI network analysis, the top ten node genes were ranked, including *MAD2L1*, *PLK1*, *SAA1*, *CCNB1*, *SHCBP1*, *KIF4A*, *ANLN*, *ERCC6L*, *CXCL2* and *WT1*. After subsequent survival analysis and expression analysis verification, the expression and prognosis of MAD2L1, PLK1, SAA1, CCNB1, SHCBP1, KIF4A, ANLN and ERCC6L in BC were finally confirmed. These 8 hub genes may play a vital role in the occurrence and development of BC.

Among the 1317 identified DEGs, significant gene expression dysregulations were observed in cell cycle, PPAR signaling pathway and AMPK signaling pathway. Cell cycle is a highly conserved process in human evolution and is essential for the normal growth of cells. Abnormal cell cycle is a hallmark of human cancer[19]. Studies reported in recent years have also identified several genes related to cell cycle, including CCNB1,

ANLN, MAD2L1 and PLK1. For example, CCNB1 may be a biomarker for the prognosis of ER-positive BC patients and monitoring the efficacy of hormone therapy[20]. Recent studies have found that the occurrence and proliferation of gastric cancer cells induced by ISL1 is mediated by the expression and regulation of CCNB1, CCNB2 and C-MYC[21]. In addition, the high expression of ANLN in BC cell nuclei is significantly related to tumor tissue size, histopathological grade, high proliferation rate and worse prognosis[22]. MAD2L1 is a mitotic spindle checkpoint gene. In patients with primary BC, compared with patients with ER +, PR + and low-grade tumors, patients with ER -, PR - and high-grade tumors have higher expression of MAD2L1, and high expression of MAD2L1 is associated with poor overall survival[23]. It has been reported that PLK1 is a key oncogene that can regulate the transition of cells in the G2-M phase, thus promoting the growth and metastasis of tamoxifen resistant BC[24]. These studies are consistent with our current conclusion that CCNB1, ANLN, MAD2L1 and PLK1, as key genes, are overexpressed in BC tissues, and their overexpression is correlated with poor prognosis. Meanwhile, PPAR signaling pathway may be an important predictor of BC response to neoadjuvant chemotherapy[25], and the activation of AMPK signaling pathway can inhibit the activity of Wnt/β-catenin signaling pathway, thereby inhibiting the growth of BC cells[26]. These studies have shown that the identified DEGs play a critical role in the occurrence and development of BC, and the hub genes among them may serve as prognostic markers and are worthy of further exploration.

Except CCNB1, ANLN, MAD2L1 and PLK1, the gene combination model of CD74, MMP9, RPA3 and SHCBP1 in tumor microenvironment (TME) can effectively predict the prognosis and disease risk of BC patients[27], while their potential mechanism remains unknown. In addition, the circKIF4A-miR-375-KIF4A axis can regulate the development of triple-negative BC through competitive endogenous RNA (ceRNA), and circKIF4A can act as a prognostic biomarker and therapeutic target for triple-negative BC[28].

SAA1 is a serum amyloid protein family member. It has been reported to be highly expressed in non-small cell lung cancer, and is associated with poor prognosis and

tyrosine kinase inhibitors[29]. However, the latest research reported that SAA1 is low expressed in hepatocellular carcinoma, and the high expression of SAA1 is associated with better prognosis[30]. Up to now, it has not been reported in BC, the specific role and function of this gene in BC requires further experimental exploration and clinical specimen verification. ERCC6L is a newly discovered DNA helicase. In human BC cell line MDA-MB-231, exogenous interference with the expression of ERCC6L can inhibit the growth of BC cells[31]. However, its role and specific mechanism in clinical specimens are still unknown. Meanwhile, it has been reported that the expression of ERCC6L is up-regulated in clear cell renal cell carcinoma, and the highly expressed ERCC6L can promote the proliferation of clear cell renal cell carcinoma cells by regulating the MAPK signaling pathway[32]. In this study, we found that SAA1 and ERCC6L may be used as prognostic markers for BC, while there are few reports on these two genes, and further research is necessary.

In this study, we found that differential expressions of the eight hub genes are related to the occurrence and development of BC, and are significantly related to the overall survival rate, which indicate that these hub genes may be utilized as potential prognostic biomarkers and therapeutic targets for BC. Nevertheless, we have to admit that there are some limitations in this study. Firstly, due to the complexity of data set in public database, it is difficult to consider some important confounding factors, such as different ages, races, regions and tumor stages when analyzing DEGs. Secondly, according to the results, 7 key genes were up-regulated in BC and 1 key gene was down-regulated, but the mechanism of their differential expression is still unclear, and more studies are needed to confirm their biological basis. Finally, this study focused on the expression level and overall survival rate of the eight hub genes, and whether these key genes can be used as biomarkers and whether they can improve the diagnostic accuracy and specificity of BC requires further research.

CONCLUSION

In conclusion, based on comprehensive bioinformatics analysis, this study identified 1317 DEGs related to the occurrence and development of BC, 165 DEGs related to prognosis, and 8 hub genes (*MAD2L1, PLK1, SAA1, CCNB1, SHCBP1, KIF4A, ANLN* and *ERCC6L*). Each of these eight hub genes has different expression in BC and is significantly related to prognosis. The results of this study indicate that the study of these DEGs would help us to have a deeper understanding on the molecular mechanisms of the pathogenesis and progression of BC. Moreover, these hub genes may serve as potential prognostic markers and therapeutic targets for BC, which provides reference for more in-depth and extensive prospective clinical research.

ARTICLE HIGHLIGHTS

Research background
Breast cancer (BC) is the most common malignant tumor in women. In 2019, 268600 new BC patients and 41760 new BC deaths were reported, which accounting for 30% of all new cancer cases and 15% of cancer-related deaths respectively. Therefore, it is particularly important to explore more sensitive and specific biomarkers for further understanding the pathogenesis of breast cancer and the choice of treatment strategies.

Research motivation
Explore more valuable therapeutic targets would be helpful to treat with high efficacy.

Research objectives
This study aims to identify novel biomarkers for BC.

Research methods
Bioinformatics.

Research results

Up-regulated genes mainly enriched in cytokine-cytokine receptor interaction, cell cycle and P53 signaling pathway ($P < 0.01$), besides, the down-regulated genes were mainly enriched in cytokine-cytokine receptor interaction, PPAR signaling pathway and AMPK signaling pathway ($P < 0.01$).

Research conclusions

MAD2L1, PLK1, SAA1, CCNB1, SHCBP1, KIF4A, ANLN and *ERCC6L* may act as biomarkers for diagnosis and prognosis in BC patients.

Research perspectives

Proper validations must be made in future studies.

REFERENCES

1 **Siegel RL**, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019; **69**: 7-34 [PMID: 30620402 DOI: 10.3322/caac.21551]

2 **Shi H**, Zhang L, Qu Y, Hou L, Wang L, Zheng M. Prognostic genes of breast cancer revealed by gene co-expression network analysis. *Oncol Lett* 2017; **14**: 4535-4542 [PMID: 29085450 DOI: 10.3892/ol.2017.6779]

3 **Redig AJ**, McAllister SS. Breast cancer as a systemic disease: a view of metastasis. *J Intern Med* 2013; **274**: 113-126 [PMID: 23844915 DOI: 10.1111/joim.12084]

4 **McAllister SS**, Gifford AM, Greiner AL, Kelleher SP, Saelzler MP, Ince TA, Reinhardt F, Harris LN, Hylander BL, Repasky EA, Weinberg RA. Systemic endocrine instigation of indolent tumor growth requires osteopontin. *Cell* 2008; **133**: 994-1005 [PMID: 18555776 DOI: 10.1016/j.cell.2008.04.045]

5 **Zhang Y**, Lv F, Yang Y, Qian X, Lang R, Fan Y, Liu F, Li Y, Li S, Shen B, Pringle GA, Zhang X, Fu L, Guo X. Clinicopathological Features and Prognosis of Metaplastic Breast Carcinoma: Experience of a Major Chinese Cancer Center. *PLoS One* 2015; **10**: e0131409 [PMID: 26115045 DOI: 10.1371/journal.pone.0131409]

6 **Clarke C**, Madden SF, Doolan P, Aherne ST, Joyce H, O'Driscoll L, Gallagher WM, Hennessy BT, Moriarty M, Crown J, Kennedy S, Clynes M. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis* 2013; **34**: 2300-2308 [PMID: 23740839 DOI: 10.1093/carcin/bgt208]

7 **Krishnamurti U**, Silverman JF. HER2 in breast cancer: a review and update. *Adv Anat Pathol* 2014; **21**: 100-107 [PMID: 24508693 DOI: 10.1097/PAP.0000000000000015]

8 **Wang X**, Zhong W, Bu J, Li Y, Li R, Nie R, Xiao C, Ma K, Huang X, Li Y. Exosomal protein CD82 as a diagnostic biomarker for precision medicine for breast cancer. *Mol Carcinog* 2019; **58**: 674-685 [PMID: 30604894 DOI: 10.1002/mc.22960]

9 **Tian T**, Gong Z, Wang M, Hao R, Lin S, Liu K, Guan F, Xu P, Deng Y, Song D, Li N, Wu Y, Dai Z. Identification of long non-coding RNA signatures in triple-negative breast cancer. *Cancer Cell Int* 2018; **18**: 103 [PMID: 30026672 DOI: 10.1186/s12935-018-0598-8]

10 **Kulasingam V**, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol* 2008; **5**: 588-599 [PMID: 18695711 DOI: 10.1038/ncponc1187]

11 **Chen B**, Tang H, Chen X, Zhang G, Wang Y, Xie X, Liao N. Transcriptomic analyses identify key differentially expressed genes and clinical outcomes between triple-negative and non-triple-negative breast cancer. *Cancer Manag Res* 2019; **11**: 179-190 [PMID: 30613165 DOI: 10.2147/CMAR.S187151]

12 **Li G**, Liu Y, Liu C, Su Z, Ren S, Wang Y, Deng T, Huang D, Tian Y, Qiu Y. Genome-wide analyses of long noncoding RNA expression profiles correlated with radioresistance in nasopharyngeal carcinoma *via* next-generation deep sequencing. *BMC Cancer* 2016; **16**: 719 [PMID: 27599611 DOI: 10.1186/s12885-016-2755-6]

13 **Diboun I**, Wernisch L, Orengo CA, Koltzenburg M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. *BMC Genomics* 2006; **7**: 252 [PMID: 17029630 DOI: 10.1186/1471-2164-7-252]

14 **Szklarczyk D**, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C.

STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015; **43**: D447-D452 [PMID: 25352553 DOI: 10.1093/nar/gku1003]

15 **Lánczky A**, Nagy Á, Bottai G, Munkácsy G, Szabó A, Santarpia L, Győrffy B. miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. *Breast Cancer Res Treat* 2016; **160**: 439-446 [PMID: 27744485 DOI: 10.1007/s10549-016-4013-7]

16 **Lacny S**, Wilson T, Clement F, Roberts DJ, Faris P, Ghali WA, Marshall DA. Kaplan-Meier survival analysis overestimates cumulative incidence of health-related events in competing risk settings: a meta-analysis. *J Clin Epidemiol* 2018; **93**: 25-35 [PMID: 29045808 DOI: 10.1016/j.jclinepi.2017.10.006]

17 **Tang Z**, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 2017; **45**: W98-W102 [PMID: 28407145 DOI: 10.1093/nar/gkx247]

18 **Asplund A**, Edqvist PH, Schwenk JM, Pontén F. Antibodies for profiling the human proteome-The Human Protein Atlas as a resource for cancer research. *Proteomics* 2012; **12**: 2067-2077 [PMID: 22623277 DOI: 10.1002/pmic.201100504]

19 **Dominguez-Brauer C**, Thu KL, Mason JM, Blaser H, Bray MR, Mak TW. Targeting Mitosis in Cancer: Emerging Strategies. *Mol Cell* 2015; **60**: 524-536 [PMID: 26590712 DOI: 10.1016/j.molcel.2015.11.006]

20 **Ding K**, Li W, Zou Z, Zou X, Wang C. CCNB1 is a prognostic biomarker for ER+ breast cancer. *Med Hypotheses* 2014; **83**: 359-364 [PMID: 25044212 DOI: 10.1016/j.mehy.2014.06.013]

21 **Shi Q**, Wang W, Jia Z, Chen P, Ma K, Zhou C. ISL1, a novel regulator of CCNB1, CCNB2 and c-MYC genes, promotes gastric cancer cell proliferation and tumor growth. *Oncotarget* 2016; **7**: 36489-36500 [PMID: 27183908 DOI: 10.18632/oncotarget.9269]

22 **Magnusson K**, Gremel G, Rydén L, Pontén V, Uhlén M, Dimberg A, Jirström K, Pontén F. ANLN is a prognostic biomarker independent of Ki-67 and essential for cell cycle progression in primary breast cancer. *BMC Cancer* 2016; **16**: 904 [PMID: 27863473 DOI: 10.1186/s12885-016-2923-8]

23 **Wang Z**, Katsaros D, Shen Y, Fu Y, Canuto EM, Benedetto C, Lu L, Chu WM, Risch HA, Yu H. Biological and Clinical Significance of MAD2L1 and BUB1, Genes Frequently Appearing in Expression Signatures for Breast Cancer Prognosis. *PLoS One* 2015; **10**: e0136246 [PMID: 26287798 DOI: 10.1371/journal.pone.0136246]

24 **Jeong SB**, Im JH, Yoon JH, Bui QT, Lim SC, Song JM, Shim Y, Yun J, Hong J, Kang KW. Essential Role of Polo-like Kinase 1 (Plk1) Oncogene in Tumor Growth and Metastasis of Tamoxifen-Resistant Breast Cancer. *Mol Cancer Ther* 2018; **17**: 825-837 [PMID: 29437878 DOI: 10.1158/1535-7163.MCT-17-0545]

25 **Chen YZ**, Xue JY, Chen CM, Yang BL, Xu QH, Wu F, Liu F, Ye X, Meng X, Liu GY, Shen ZZ, Shao ZM, Wu J. PPAR signaling pathway may be an important predictor of breast cancer response to neoadjuvant chemotherapy. *Cancer Chemother Pharmacol* 2012; **70**: 637-644 [PMID: 22903535 DOI: 10.1007/s00280-012-1949-0]

26 **Zou YF**, Xie CW, Yang SX, Xiong JP. AMPK activators suppress breast cancer cell growth by inhibiting DVL3-facilitated Wnt/β-catenin signaling pathway activity. *Mol Med Rep* 2017; **15**: 899-907 [PMID: 28035400 DOI: 10.3892/mmr.2016.6094]

27 **Wang J**, Yang Z, Zhang C, Ouyang J, Zhang G, Wu C. A four-gene signature in the tumor microenvironment that significantly associates with the prognosis of patients with breast cancer. *Gene* 2020; **761**: 145049 [PMID: 32791092 DOI: 10.1016/j.gene.2020.145049]

28 **Tang H**, Huang X, Wang J, Yang L, Kong Y, Gao G, Zhang L, Chen ZS, Xie X. circKIF4A acts as a prognostic factor and mediator to regulate the progression of triple-negative breast cancer. *Mol Cancer* 2019; **18**: 23 [PMID: 30744636 DOI: 10.1186/s12943-019-0946-x]

29 **Milan E**, Lazzari C, Anand S, Floriani I, Torri V, Sorlini C, Gregorc V, Bachi A. SAA1 is over-expressed in plasma of non small cell lung cancer patients with poor outcome after treatment with epidermal growth factor receptor tyrosine-kinase inhibitors. *J Proteomics* 2012; **76 Spec No.**: 91-101 [PMID: 22771314 DOI: 10.1016/j.jprot.2012.06.022]

30 **Zhang W**, Kong HF, Gao XD, Dong Z, Lu Y, Huang JG, Li H, Yang YP. Immune infiltration-associated serum amyloid A1 predicts favorable prognosis for

hepatocellular carcinoma. *World J Gastroenterol* 2020; **26**: 5287-5301 [PMID: 32994688 DOI: 10.3748/wjg.v26.i35.5287]

31 **Liu J**, Sun J, Zhang Q, Zeng Z. shRNA knockdown of DNA helicase ERCC6L expression inhibits human breast cancer growth. *Mol Med Rep* 2018; **18**: 3490-3496 [PMID: 30066865 DOI: 10.3892/mmr.2018.9317]

32 **Zhang G**, Yu Z, Fu S, Lv C, Dong Q, Fu C, Kong C, Zeng Y. ERCC6L that is up-regulated in high grade of renal cell carcinoma enhances cell viability *in vitro* and promotes tumor growth *in vivo* potentially through modulating MAPK signalling pathway. *Cancer Gene Ther* 2019; **26**: 323-333 [PMID: 30459398 DOI: 10.1038/s41417-018-0064-8]

69987_Auto_Edited-check.docx

ORIGINALITY REPORT

11%

SIMILARITY INDEX

PRIMARY SOURCES

1   www.frontiersin.org
    Internet   61 words — 2%

2   assets.researchsquare.com
    Internet   31 words — 1%

3   www.ncbi.nlm.nih.gov
    Internet   27 words — 1%

4   ss.bjmu.edu.cn
    Internet   25 words — 1%

5   Saroj Thapa, Mandika Chetry, Kaiyu Huang, Yangpei
    Peng et al. "Significance of aquaporins' expression in   20 words — 1%
    the prognosis of gastric cancer", Bioscience Reports, 2018
    Crossref

6   Wanli Yang, Wei Zhou, Xinhui Zhao, Xiaoqian
    Wang et al. "Prognostic biomarkers and   18 words — < 1%
    therapeutic targets in oral squamous cell carcinoma: a study
    based on cross-database analysis", Hereditas, 2021
    Crossref

7   bmccancer.biomedcentral.com
    Internet   18 words — < 1%

8   pureadmin.qub.ac.uk
    Internet   18 words — < 1%

9 "High-efficient Screening Method for Identification of Key Genes in Breast Cancer Through Microarray and Bioinformatics", Anticancer Research, 2017
Crossref

17 words — < 1%

10 Tianyi Wang, Minwen Ha. "Silencing ARHGAP9 correlates with the risk of breast cancer and inhibits the proliferation, migration, and invasion of breast cancer", Journal of Cellular Biochemistry, 2018
Crossref

15 words — < 1%

11 www.oncotarget.com
Internet

15 words — < 1%

12 Anupama Modi, Purvi Purohit, Ashita Gadwal, Shweta Ukey, Dipayan Roy, Sujoy Fernandes, Mithu Banerjee. "In-silico analysis of differentially expressed genes and their regulating microRNA involved in lymph node metastasis in invasive breast carcinoma", Cold Spring Harbor Laboratory, 2020
Crossref Posted Content

14 words — < 1%

13 Biao Huang, Wei Han, Zu-Feng Sheng, Guo-Liang Shen. "Identification of new biomarkers and pathways associated with treatment and prognosis in melanoma patients", Research Square, 2020
Crossref Posted Content

14 words — < 1%

14 Camargo, A.. "Identification of dilated cardiomyopathy signature genes through gene expression and network data integration", Genomics, 200812
Crossref

14 words — < 1%

15 Ye Cheng, Hanzhi Sun, Hanjin Wang, Wei Jiang, Weiwei Tang, Chen Lu, Wenling Zhang, Ziyi Chen,

14 words — < 1%

Chengyu Lv. "<p>Star Circular RNAs In Human Cancer: Progress And Perspectives</p>", OncoTargets and Therapy, 2019
Crossref

16 pergamos.lib.uoa.gr
Internet
14 words — < 1%

17 Hailin Tang, Xiaojia Huang, Jin Wang, Lu Yang, Yanan Kong, Guanfeng Gao, Lijuan Zhang, Zhe-Sheng Chen, Xiaoming Xie. "circKIF4A acts as a prognostic factor and mediator to regulate the progression of triple-negative breast cancer", Molecular Cancer, 2019
Crossref
13 words — < 1%

18 Zeng-hong Wu, Dong-liang Yang. "Identification of a protein signature for predicting overall survival of hepatocellular carcinoma: a study based on data mining", BMC Cancer, 2020
Crossref
13 words — < 1%

19 tcr.amegroups.com
Internet
13 words — < 1%

20 Junyi Hu, Lijie Zhou, Zhengshuai Song, Ming Xiong, Youpeng Zhang, Yu Yang, Ke Chen, Zhaohui Chen. "The identification of new biomarkers for bladder cancer: A study based on TCGA and GEO datasets", Journal of Cellular Physiology, 2019
Crossref
12 words — < 1%

21 Kuan Hu, Lei Yao, Zhijie Xu, Yuanliang Yan, Juanni Li. "Prognostic Value and Therapeutic Potential of CBX Family Members in Ovarian Cancer", Frontiers in Cell and Developmental Biology, 2022
Crossref
12 words — < 1%

22 Www.dovepress.com
Internet
12 words — < 1%