Identification of NUF2 and FAM83D as Potential Biomarkers in Triple-negative Breast Cancer

Xiuming Zhai [1], **Zhaowei Yang** [2], **Xiji Liu** [1], **Zihe Dong** [3], **Dandan Zhou** [Corresp. 3]

[1] Department of Laboratory Medicine, The Third Affiliated Hospital of Chongqing Medical University, Chongqing, China

[2] Department of Breast and Thyroid, Chongqing Hospital of Traditional Chinese Medicine, Chongqing, China

[3] Department of Laboratory Medicine, Chongqing Hospital of Traditional Chinese Medicine, Chongqing, China

Corresponding Author: Dandan Zhou
Email address: zhoudandancq@163.com

Background. Breast cancer is a heterogeneous disease. Compared with other subtypes of breast cancer, triple-negative breast cancer (TNBC) is easy to metastasize and has a short survival time, less choice of treatment options. Here, we aimed to identify the potential biomarkers to TNBC diagnosis and prognosis. Material/Methods. Three independent data sets (GSE45827, GSE38959, GSE65194) were downloaded from the Gene Expression Omnibus (GEO). The R software packages were used to integrate the gene profiles and identify differentially expressed genes (DEGs). A variety of bioinformatics tools were used to explore the hub genes, including the DAVID database, STRING database and Cytoscape software. Reverse transcription quantitative PCR (RT-qPCR) was used to verify the hub genes in 14 pairs of TNBC paired tissues. Results. In this study, we screened out 161 DEGs between 222 non-TNBC and 126 TNBC samples, of which 105 genes were up-regulated and 56 were down-regulated. These DEGs were enriched for 27 GO terms and 2 pathways. GO analysis enriched mainly in “cell division”, “chromosome, centromeric region” and “microtubule motor activity”. KEGG pathway analysis enriched mostly in “Cell cycle” and “Oocyte meiosis”. PPI network was constructed and then 10 top hub genes were screened. According to the analysis results of the Kaplan-M eier survival curve, the expression levels of only NUF2, FAM83D and CENPH were associated with the recurrence-free survival in TNBC samples ($P<0.05$). RT-qPCR confirmed that the expression levels of NUF2 and FAM83D in TNBC tissues were indeed up-regulated significantly. Conclusions. The comprehensive analysis showed that NUF2 and FAM83D could be used as potential biomarkers for diagnosis and prognosis of TNBC.

Title page

The names of the authors:

Xiuming Zhai[1], Zhaowei Yang[2], Xiji Liu[1], Zihe Dong[3], Dandan Zhou[3]*

Article title:

Identification of NUF2 and FAM83D as Potential Biomarkers in Triple-negative Breast Cancer

Running head:

Potential Biomarkers in TN Breast Cancer

The affiliation and address of the authors:

[1]Department of Laboratory Medicine, The Third Affiliated Hospital of Chongqing Medical University, No. 1 Shuanghu Branch Road, Yubei District, Chongqing 401120, China.

[2]Department of Breast and Thyroid, Chongqing Hospital of Traditional Chinese Medicine, No. 6, Qizhi Road, Panxi District, Jiangbei District, Chongqing 400021, China.

[3]Department of Laboratory Medicine, Chongqing Hospital of Traditional Chinese Medicine, No. 6, Qizhi Road, Panxi District, Jiangbei District, Chongqing 400021, China.

***Correspondence author:**

Dandan Zhou: Department of Laboratory Medicine, Chongqing Hospital of Traditional Chinese Medicine, No. 6, Qizhi Road, Panxi District, Jiangbei District, Chongqing 400021, China.

E-mail address: zhoudandancq@163.com

Tel.: (+86) 023-67063765

Abstract

Background. Breast cancer is a heterogeneous disease. Compared with other subtypes of breast cancer, triple-negative breast cancer (TNBC) is easy to metastasize and has a short survival time, less choice of treatment options. Here, we aimed to identify the potential biomarkers to TNBC diagnosis and prognosis.

Material/Methods. Three independent data sets (GSE45827, GSE38959, GSE65194) were downloaded from the Gene Expression Omnibus (GEO). The R software packages were used to integrate the gene profiles and identify differentially expressed genes (DEGs). A variety of bioinformatics tools were used to explore the hub genes, including the DAVID database, STRING database and Cytoscape software. Reverse transcription quantitative PCR (RT-qPCR) was used to verify the hub genes in 14 pairs of TNBC paired tissues.

Results. In this study, we screened out 161 DEGs between 222 non-TNBC and 126 TNBC samples, of which 105 genes were up-regulated and 56 were down-regulated. These DEGs were enriched for 27 GO terms and 2 pathways. GO analysis enriched mainly in “cell division”, “chromosome, centromeric region” and “microtubule motor activity”. KEGG pathway analysis enriched mostly in “Cell cycle” and “Oocyte meiosis”. PPI network was constructed and then 10 top hub genes were screened. According to the analysis results of the Kaplan-Meier survival curve, the expression levels of only NUF2, FAM83D and CENPH were associated with the recurrence-free survival in TNBC samples (P<0.05). RT-qPCR confirmed that the expression levels of NUF2 and FAM83D in TNBC tissues were indeed up-regulated significantly.

Conclusions. The comprehensive analysis showed that NUF2 and FAM83D could be used as potential biomarkers for diagnosis and prognosis of TNBC.

Key words Biomarker; Triple-negative Breast Cancer; Bioinformatics; RT-qPCR

Introduction

There were approximately 18.1 million new cancer cases worldwide in 2018, including 2.1 million breast cancers (Bray et al. 2018). Breast cancer is the highest incidence among new morbidity and mortality in females with cancer (Cao et al. 2019). According to variations in the expressions of the estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2), breast cancer were defined as four major intrinsic molecular subtypes: luminal A, luminal B, HER2-positive and triple-negative breast cancer (TNBC) (Sorlie et al. 2001). TNBC is characterized by a lack of expression of the ER and PR as well as HER2 (Serra et al. 2014). TNBC that occurs mostly in premenopausal young women represents approximately 15-20% of all invasive breast cancers(Foulkes et al. 2010). TNBC is a highly heterogeneous disease, not only at the molecular level, but also in terms of its pathology and clinical manifestation. Its prognosis is worse than other types of breast cancer as well as the risk of death is higher (Metzger-Filho et al. 2012). Chemotherapy is currently the primary adjuvant treatment, due to the lack of effective molecular targets, it is not only insensitive to endocrine therapy and HER-2 targeted therapy, but also easily causes chemo-resistant (Wein & Loi 2017). TNBC has become an intractable problem for clinical treatment.

Current researchers are focusing on personalized treatment based on the multi-gene assays (Pan et al. 2019). With the continuous development of high-throughput sequencing technology, bioinformatics analysis plays a key role in the diagnosis, prognosis and screening of tumors (Goldfeder et al. 2017; Ma et al. 2020). Many genes have been identified as signatures for diagnosis and prognosis of triple negative breast cancer (Dai et al. 2019; Stovgaard et al. 2020). Recent study found that CHD4-β1 integrin axis may be a prognostic marker for TNBC using next-generation sequencing and bioinformatics analysis (Ou-Yang et al. 2019). The computational analysis of complex biological networks could help research scholars identify potential genes related to TNBC (Li et al. 2020).

In this study, we first identified a group of differentially expressed genes (DEGs) associated with TNBC from the Gene Expression Synthesis (GEO) database. Then, based on bioinformatics

analysis, three candidate genes related to TNBC diagnosis and prognosis were successfully identified. Finally, reverse transcription quantitative PCR (RT-qPCR) was used to verify the candidate biomarkers in TNBC tissues and adjacent tissues. The current research aimed to explore potential biomarkers that may be highly correlated with the prognostic and diagnostic value of triple negative breast cancer.

Materials and methods

Data source

Triple-negative breast cancer gene expression data sets in this study were obtained from the publicly available GEO databases (https://www.ncbi.nlm.nih.gov/geo/) (Barrett et al. 2013). Three independent data sets from GSE45827 (Gruosso et al. 2016), GSE38959 (Komatsu et al. 2013), GSE65194 (Maire et al. 2013) were included. GSE45827 consists of 100 non-triple-negative breast cancer (non-TNBC) samples and 41 TNBC samples, GSE65194 consists of 109 non-TNBC and 55 TNBC samples, both GSE65194 and GSE45827 are based on the platform GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. GSE38959 consists of 13 non-TNBC and 30 TNBC samples, and the platform is GPL4133 Agilent-014850 Whole Human Genome Microarray 4x44K G4112F. All of the data sets were available online.

A total of 14 TNBC patients were collected in Chongqing Traditional Chinese Medicine Hospital. All patients were diagnosed with triple negative breast cancer (ER-negative, PR-negative, HER-2-negative) by histopathological examination, excluding other malignant tumors and no important organ diseases, such as severe cardiovascular, liver disease as well as renal insufficiency. A total of 28 frozen tissue specimens contained 14 tumor tissues and 14 matched adjacent non-tumor tissues were obtained. All tissues were collected immediately after surgical resection, and snap-frozen in liquid nitrogen until RNA extraction. Clinical information were obtained for all patients by the investigator from medical records. The more detailed clinical information are shown in Supplemental file 1. This study has been approved by the Chongqing Hospital of Traditional Chinese Medicine ethics committee and written informed consent was obtained from all patients.

Data processing of DEGs

R software (v3.6.2; http://www.r-project.org) was used for bioinformatics analysis. First, the gene expression profiles of three data sets were downloaded by using GEOquery package. Subsequently, background adjustment were performed by using the dplyr package. Finally, we utilized log2 transformation to normalize the data using the limma package. The RobustRankAggreg package was used to screen the differentially expressed genes, using adjust *P* value < 0.01 and |logFC|≥ 2 as cut-off criteria. The VennDiagram package was used to present significant co-expression genes.

GO enrichment and KEGG pathway analysis of DEGs

Gene ontology (GO)(The Gene Ontology 2019) is a tool for annotating genes from various ontologies, including biological processes (BP), cellular components (CC), molecular functions (MF). The Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2019) is famous for "understanding the advanced functions and utility resource library of biological systems", KEGG pathway mainly presents intermolecular interactions and intermolecular networks. GO enrichment and KEGG pathway analysis for DEGs were performed through the DAVID database (v6.8; http://david.abcc.ncifcrf.gov) (Jiao et al. 2012) with “after FDR” (corrected *P*-Value < 0.01, gene count ≥ 5) set as statistically significant. The ggplot2 package in R was used to visualize the GO functional enrichment results.

Protein-protein Interaction (PPI) networks and hub gene analysis

The online STRING database (v11.0; https://string-db.org/) collects and integrates information on the correlation between known and predicted proteins from multiple species (Szklarczyk et al. 2019). PPI network analysis could systematically study the molecular mechanisms of disease and discover new drug targets. The DEGs screened previously were mapped via the STRING database. Subsequently, visual analysis of the PPI network was matched to Cytoscape (v3.7.2; https://cytoscape.org), and hub genes were analyzed with the Cytoscape plugin CytoHubba (Chin et al. 2014). The DMNC algorithm was used to identify the top 10 hub genes.

Survival analysis

The Kaplan Meier plotter, an online survival analysis tool, could rapidly assess the effect of 54k genes on survival in 21 cancer types (http://kmplot.com/analysis/), including the effect of 22,277 genes on breast cancer prognosis (Gyorffy et al. 2012; Gyorffy et al. 2010). In this study, TNBC patients were only screened out based on the intrinsic sub-type (basis: n = 879). Probes of genes were selected "only JetSet best probe set" (Li et al. 2011). Recurrence-free survival (RFS) was selected for survival analysis the candidate hub genes, $P < 0.05$ was considered to be statistically significant.

Validation of hub genes

RT-qPCR were used to further verify the mRNA expression of the candidate hub genes in TNBC tissues and adjacent tissues. Total RNA of TNBC patients' tissues samples were isolated by TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Total RNA quantity was evaluated by a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). RNA was reverse transcribed into cDNA according to the instructions of the Takara kit (Takara Bio Inc., Japan). RT-qPCR reactions were performed using the SYBR Green PCR Master Mix System (Tiangen Biotech, Beijing, China). GAPDH was used as a control to compare the relative expression of NUF2, FAM83D and CENPH mRNA in 14 pairs of triple negative breast cancer paired tissues. Three replicate holes were performed for target genes in the RT-qPCR experiment, and the primer sequences are shown in Table 1. The primers of the target genes and the internal reference gene were synthesized by Sangon Biotech (Shanghai) Co., Ltd.

Statistical analysis

Statistical analyses of this study were analyzed with R software v3.6.2 and GraphPad Prism 5.0. Two-tailed Student's t-test was used to significance of differences between two groups, and $P < 0.05$ was considered statistically significant. The RT-qPCR results were calculated and evaluated using the $2^{-\triangle\triangle Ct}$ method.

Results

DEGs in non-TNBC and TNBC samples

Three series of matrix files, for a total of 222 non-TNBC samples and 126 TNBC samples,

were selected to identify DEGs ($P < 0.01$, $|\log FC| \geq 2$). A total of 488 genes were identified after analyzing GSE45827, of which 259 genes were up-regulated and 229 genes were down-regulated. In gene chip GSE38959, 794 DEGs were identified, 478 genes were up-regulated, and 316 genes were down-regulated. And from GSE65194, 531 DEGs including 282 up-regulated genes and 249 down-regulated genes were identified. The Venn diagrams showed that a total of 161 DEGs overlapped, in which 105 genes were up-regulated and 56 genes were down-regulated (**Fig. 1**). The more detailed results are shown in Supplemental file 2.

<Fig. 1 >

GO and KEGG pathway analysis of DEGs

Next, we attempted to identify the biological function of the 161 common DEGs. GO enrichment and KEGG pathway analysis were performed through the DAVID database. Terms with matching the filter criteria were collected and grouped into clusters according to their membership similarities. As shown in **Figure 2**, the top 5 functions for biological processes were as follows: cell division, mitotic nuclear division, chromosome segregation, sister chromatid cohesion and cell proliferation. The top 5 functions for cellular components were as follows: chromosome centromeric region, midbody, nucleus, condensed chromosome kinetochore and kinetochore. The molecular functions enriched were associated with microtubule motor activity, microtubule binding, ATP binding and protein binding. The KEGG analysis showed that the main enriched signaling pathways were related to the cell cycle and oocyte meiosis. The more detailed results are shown in Supplemental file 3.

<Fig. 2 >

PPI network construction and hub genes detection

In order to better understand which of these DEGs were most likely to be the central regulatory genes for TNBC, PPI network was constructed through the online STRING platform and Cytoscape software (**Fig. 3A**). Subsequently, according to the DMNC algorithm, the top 10 hub genes were screened through the cytoHubba and are sequentially ranked as follows: ANLN, FAM64A, CDCA2, NUF2, FAM83D, CENPH, KIF14, MKLP-1, KIF15, DEPDC1 (**Fig. 3B**). The

expression of 10 hub genes were all significantly increased in the PPI network. We initially speculate that 10 candidate hub genes may be related to tumor occurrence.

<Fig. 3 >

Survival analysis and validation of hub genes

In order to examine whether the candidate hub genes expression levels were associated with the outcome of TNBC patients. Next, the correlation between these genes and the recurrence-free survival of TNBC patients were analyzed by the Kaplan Meier plotter. According to the analysis results of the Kaplan-Meier survival curve, we found that TNBC patients with higher expression levels of NUF2, FAM83D, CENPH have significantly decreased recurrence-free survival (P<0.05), but not ANLN, FAM64A, CDCA2, KIF14, MKLP-1, KIF15, DEPDC1 (P>0.05). More specific information about these survival-related hub genes is shown in **Figure 4**.

<Fig. 4 >

Finally, we validated the expression levels of NUF2, FAM83D and CENPH in 14 pairs of triple negative breast cancer paired tissues by using RT-qPCR. **Figure 5** showed that the expression levels of NUF2 and FAM83D were significantly higher in TNBC tissues than adjacent tissues (P<0.001), but not CENPH (P=0.68). Combined with the above analysis, we preliminarily concluded that NUF2 and FAM83D may be potential biomarkers to TNBC diagnosis and prognosis. The more detailed results are shown in Supplemental file 4.

<Fig. 5 >

Discussion

TNBC is considered as an aggressive subtype of breast cancer. Compared with other types of breast cancer, TNBC is characterized by high malignancy rate, easier recurrence (Dent et al. 2007), and low survival rate (Carey et al. 2006). Despite advances in the targeted therapies of TNBC, including the approval of poly-ADP-ribose polymerase (PARP) and immune check-point inhibitors for the treatment of BRCA germ cell mutated breast cancers, there is still a lack of clinical evidence to evaluate their efficacy for TNBC patients (Vagia et al. 2020). Therefore, it is necessary to identify effective molecular therapeutic targets for TNBC.

In the present study, we screened out 161 DEGs between 222 non-TNBC and 126 TNBC samples by analyzing three datasets, of which 105 were up-regulated and 56 were down-regulated. The GO enrichment analysis and KEGG pathways showed that the screened DEGs were enriched for 27 GO terms and 2 pathways. To further investigate the interrelationship of 161 DEGs, PPI network was first constructed and then 10 top hub genes were screened out, including ANLN, FAM64A, CDCA2, NUF2, FAM83D, CENPH, KIF14, MKLP-1, KIF15, DEPDC1. The analysis results of the Kaplan-Meier survival curve showed that the expression levels of NUF2, FAM83D and CENPH were associated with the recurrence-free survival in TNBC samples (P<0.05). Finally, we found that the expression levels of only NUF2 and FAM83D did increase significantly in TNBC tissues by using RT-qPCR.

NUF2 is an essential component of the kinetochore-associated NDC80 complex, which plays a regulatory role in chromosome segregation and spindle checkpoint activity (Liu et al. 2007; Zhang et al. 2015). Several studies have shown that NUF2 was associated with the development of multiple cancers. The results showed that the expression of NUF2 was associated with poor prognosis in patients with colorectal cancer (Kobayashi et al. 2014) and oral cancer (Thang et al. 2016), which may be related to the regulation of tumor cell apoptosis involved in the NUF2. Sugimasa H et al (Sugimasa et al. 2015) demonstrated that the NUF2 gene could be directly trans-activated by the heterogeneous ribonucleoprotein K (hnRNP K), and that the hnRNP K-NUF2 axis affected the growth of colon cancer cells by participating in processes of mitosis and proliferation. Recent studies have shown that NUF2 was also closely related to breast cancer. Xu W et al (Xu et al. 2019) confirmed that NUF2 was indeed up-regulated in breast cancer tissue by bioinformatics analysis and RT-qPCR assay, and that NUF2 may regulate the carcinogenesis and progression of breast cancer via cell cycle-related pathways. However, the expression level changes of NUF2 in triple-negative breast cancer have not yet been studied. In this study, we found that the expression level of NUF2 was higher in triple-negative breast cancer than in non-triple negative breast cancer and TNBC patients with higher NUF2 expression level had significantly reduced the recurrence-free survival. GO enrichment analysis shows that NUF2 is mainly involved in cell division, mitotic

nuclear division, chromosome segregation and sister chromatid cohesion, their dysregulation impact significantly on development of cancer (Bakhoum et al. 2018; Guo et al. 2013; Lopez-Lazaro 2018). Based on the above analysis, we speculate that NUF2 plays an important role in tumor progression, and NUF2 may be serve as a biomarker for diagnosis and prognosis of triple-negative breast cancer. Certainly, the specific molecular mechanism of NUF2 expression level changes in TNBC still need to be further studied.

FAM83D belongs to the FAM83 family, which could regulate cell proliferation, growth, migration and epithelial to mesenchymal transition (Li et al. 2018; Santamaria et al. 2008). The studies have found that FAM83D could not only affect cell proliferation and motility through the tumor suppressor gene FBXW7 (Mu et al. 2017) or ERK1/ERK2 signaling cascade (Wang et al. 2015), but also affect breast cancer cell growth and promote epithelial cell transformation through MAPK signaling (Cipriano et al. 2013; Cipriano et al. 2014; Lee et al. 2012). The expression of FAM83D was significantly increased in primary breast cancer and the high expression level of FAM83D was closely related to the adverse clinical outcomes and distant metastasis in breast cancer patients (Wang et al. 2013). In our study, we found that the expression of FAM83D was significantly increased in TNBC patients and TNBC patients with higher FAM83D expression level had significantly reduced the recurrence-free survival. GO enrichment analysis shows that FAM83D is mainly involved in cell division, mitotic nuclear division and cell proliferation, their dysregulation have a major impact on the development of cancer (Bakhoum et al. 2018; Lopez-Lazaro 2018; Wu et al. 2019). We speculated that FAM83D might play a role in the progression and prognosis of triple-negative breast cancer.

Centromere protein H (CENP-H) is a component of the kinetochore and plays an essential role in mitotic processes (Lu et al. 2017), accurate chromosome segregation (Zhu et al. 2015) as well as appropriate kinetochore assembly (Zhao et al. 2012). Many studies have shown that CENPH is closely associated with human cancers, including colorectal cancer (Wu et al. 2017), renal cell carcinoma (Wu et al. 2015), non-small cell lung cancer (Liao et al. 2009) as well as breast cancer (Walian et al. 2016). However, there is no current evidence on the correlation

between CENPH and triple negative breast cancer. In this study, we found that there is no significant correlation between the mRNA expression of CENPH and triple negative breast cancer.

It is worth noting that protein-coding genes are not the sole drivers for cancer. Breast cancer is also related to the expressions of non-coding RNAs, include repetitive DNA (Yandim & Karakulah 2019), transposable element (Karakulah et al. 2019), micro RNA (Aslan et al. 2020) and Long non-coding RNA (Riahi et al. 2020),etc. In this study, we have found that the expressions of NUF2 and FAM83D are associated with triple-negative breast cancer. Next, we will further investigate whether the expression changes of NUF2/FAM83D in triple-negative breast cancer are caused by non-coding RNA.

Conclusion

In summary, we firstly demonstrated that the mRNA levels of NUF2/ FAM83D have changed significantly in TNBC tissues compared to adjacent tissues. The mRNA expression levels of NUF2/FAM83D are significantly up-regulated in TNBC tissues. NUF2/FAM83D might serve as potential molecular biomarkers for diagnosis and prognostic indicators of TNBC. However, the functional mechanisms of NUF2 and FAM83D in TNBC patients are still to be further studied, including the expression of their protein levels and their relationship with the clinical characteristics of TNBC patients and so on. Therefore, we still need to do more experiments before clinical trials.

References

Aslan C, Maralbashi S, Kahroba H, Asadi M, Soltani MS, Javadian M, Shanehbandi D, Baradaran B, Darabi M, and Kazemi T. 2020. Docosahexaenoic acid (DHA) inhibits pro-angiogenic effects of breast cancer cells via down-regulating cellular and exosomal expression of angiogenic genes and microRNAs. *Life Sci*:118094. 10.1016/j.lfs.2020.118094

Bakhoum SF, Ngo B, Laughney AM, Cavallo JA, Murphy CJ, Ly P, Shah P, Sriram RK, Watkins TBK, Taunk NK, Duran M, Pauli C, Shaw C, Chadalavada K, Rajasekhar VK, Genovese G, Venkatesan S, Birkbak NJ, McGranahan N, Lundquist M, LaPlant Q, Healey JH, Elemento O, Chung CH, Lee NY, Imielenski M, Nanjangud G, Pe'er D, Cleveland DW, Powell SN, Lammerding J, Swanton C, and Cantley LC. 2018. Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature* 553:467-472. 10.1038/nature25432

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, and Soboleva A. 2013. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 41:D991-995. 10.1093/nar/gks1193

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, and Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68:394-424. 10.3322/caac.21492

Cao J, Eshak ES, Liu K, Muraki I, Cui R, Iso H, Tamakoshi A, and Group JS. 2019. Television Viewing Time and Breast Cancer Incidence for Japanese Premenopausal and Postmenopausal Women: The JACC Study. *Cancer Res Treat* 51:1509-1517. 10.4143/crt.2018.705

Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, and Millikan RC. 2006. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 295:2492-2502. 10.1001/jama.295.21.2492

Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, and Lin CY. 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 8 Suppl 4:S11. 10.1186/1752-0509-8-S4-S11

Cipriano R, Miskimen KL, Bryson BL, Foy CR, Bartel CA, and Jackson MW. 2013. FAM83B-mediated activation of PI3K/AKT and MAPK signaling cooperates to promote epithelial cell transformation and resistance to targeted therapies. *Oncotarget* 4:729-738. 10.18632/oncotarget.1027

Cipriano R, Miskimen KL, Bryson BL, Foy CR, Bartel CA, and Jackson MW. 2014. Conserved oncogenic behavior of the FAM83 family regulates MAPK signaling in human cancer. *Mol Cancer Res* 12:1156-1165. 10.1158/1541-7786.MCR-13-0289

Dai X, Cheng H, Chen X, Li T, Zhang J, Jin G, Cai D, and Huang Z. 2019. FOXA1 is Prognostic of Triple Negative Breast Cancers by Transcriptionally Suppressing SOD2 and IL6. *Int J Biol Sci* 15:1030-1041. 10.7150/ijbs.31009

Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, and Narod SA. 2007. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res* 13:4429-4434. 10.1158/1078-0432.CCR-06-3045

Foulkes WD, Smith IE, and Reis-Filho JS. 2010. Triple-negative breast cancer. *N Engl J Med* 363:1938-1948. 10.1056/NEJMra1001389

Goldfeder RL, Wall DP, Khoury MJ, Ioannidis JPA, and Ashley EA. 2017. Human Genome Sequencing at the Population Scale: A Primer on High-Throughput DNA Sequencing and Analysis. *Am J Epidemiol* 186:1000-1009. 10.1093/aje/kww224

Gruosso T, Mieulet V, Cardon M, Bourachot B, Kieffer Y, Devun F, Dubois T, Dutreix M, Vincent-Salomon A, Miller KM, and Mechta-Grigoriou F. 2016. Chronic oxidative stress promotes H2AX protein degradation and enhances chemosensitivity in breast cancer patients. *EMBO Mol Med* 8:527-549. 10.15252/emmm.201505891

Guo G, Sun X, Chen C, Wu S, Huang P, Li Z, Dean M, Huang Y, Jia W, Zhou Q, Tang A, Yang Z, Li X, Song P, Zhao X, Ye R, Zhang S, Lin Z, Qi M, Wan S, Xie L, Fan F, Nickerson ML, Zou X, Hu X, Xing L, Lv Z, Mei H, Gao S, Liang C, Gao Z, Lu J, Yu Y, Liu C, Li L, Fang X, Jiang Z, Yang J, Li C, Zhao X, Chen J, Zhang F, Lai Y, Lin Z, Zhou F, Chen H, Chan HC, Tsang S, Theodorescu D, Li Y, Zhang X, Wang J, Yang H, Gui Y, Wang J, and Cai Z. 2013. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat Genet* 45:1459-1463. 10.1038/ng.2798

Gyorffy B, Benke Z, Lanczky A, Balazs B, Szallasi Z, Timar J, and Schafer R. 2012. RecurrenceOnline: an online analysis tool to determine breast cancer recurrence and hormone receptor status using microarray data. *Breast Cancer Res Treat* 132:1025-1034. 10.1007/s10549-011-1676-y

Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, and Szallasi Z. 2010. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* 123:725-731. 10.1007/s10549-009-0674-9

Jiao X, Sherman BT, Huang da W, Stephens R, Baseler MW, Lane HC, and Lempicki RA. 2012. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* 28:1805-1806. 10.1093/bioinformatics/bts251

Kanehisa M, Sato Y, Furumichi M, Morishima K, and Tanabe M. 2019. New approach for understanding genome variations in KEGG. *Nucleic Acids Res* 47:D590-D595. 10.1093/nar/gky962

Karakulah G, Arslan N, Yandim C, and Suner A. 2019. TEffectR: an R package for studying the potential effects of transposable elements on gene expression with linear regression model. *PeerJ* 7:e8192. 10.7717/peerj.8192

Kobayashi Y, Takano A, Miyagi Y, Tsuchiya E, Sonoda H, Shimizu T, Okabe H, Tani T, Fujiyama Y, and Daigo Y. 2014. Cell division cycle-associated protein 1 overexpression is essential for the malignant potential of colorectal cancers. *Int J Oncol* 44:69-77. 10.3892/ijo.2013.2177

Komatsu M, Yoshimaru T, Matsuo T, Kiyotani K, Miyoshi Y, Tanahashi T, Rokutan K, Yamaguchi R, Saito A, Imoto S, Miyano S, Nakamura Y, Sasa M, Shimada M, and Katagiri T. 2013. Molecular features of triple negative breast cancer cells by genome-wide gene expression profiling analysis. *Int J Oncol* 42:478-506. 10.3892/ijo.2012.1744

Lee SY, Meier R, Furuta S, Lenburg ME, Kenny PA, Xu R, and Bissell MJ. 2012. FAM83A confers EGFR-TKI resistance in breast cancer cells and in mice. *J Clin Invest* 122:3211-3220. 10.1172/JCI60498

Li J, Shima H, Nishizawa H, Ikeda M, Brydun A, Matsumoto M, Kato H, Saiki Y, Liu L, Watanabe-Matsui M, Iemura K, Tanaka K, Shiraki T, and Igarashi K. 2018. Phosphorylation of BACH1 switches its function from transcription factor to mitotic chromosome regulator and promotes its interaction with HMMR. *Biochem J* 475:981-1002. 10.1042/BCJ20170520

Li Q, Birkbak NJ, Gyorffy B, Szallasi Z, and Eklund AC. 2011. Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics* 12:474. 10.1186/1471-2105-12-474

Li XX, Wang LJ, Hou J, Liu HY, Wang R, Wang C, and Xie WH. 2020. Identification of Long Noncoding RNAs as Predictors of Survival in Triple-Negative Breast Cancer Based on Network Analysis. *Biomed Res Int* 2020:8970340. 10.1155/2020/8970340

Liao WT, Wang X, Xu LH, Kong QL, Yu CP, Li MZ, Shi L, Zeng MS, and Song LB. 2009. Centromere protein H is a novel prognostic marker for human nonsmall cell lung cancer progression and overall patient survival. *Cancer* 115:1507-1517. 10.1002/cncr.24128

Liu D, Ding X, Du J, Cai X, Huang Y, Ward T, Shaw A, Yang Y, Hu R, Jin C, and Yao X. 2007. Human NUF2 interacts with centromere-associated protein E and is essential for a stable spindle microtubule-kinetochore attachment. *J Biol Chem* 282:21415-21424. 10.1074/jbc.M609026200

Lopez-Lazaro M. 2018. The stem cell division theory of cancer. *Crit Rev Oncol Hematol* 123:95-113. 10.1016/j.critrevonc.2018.01.010

Lu G, Hou H, Lu X, Ke X, Wang X, Zhang D, Zhao Y, Zhang J, Ren M, and He S. 2017. CENP-H regulates the cell growth of human hepatocellular carcinoma cells through the mitochondrial apoptotic pathway. *Oncol Rep* 37:3484-3492. 10.3892/or.2017.5602

372 Ma X, Zhou L, and Zheng S. 2020. Transcriptome analysis revealed key prognostic genes and microRNAs in
373 hepatocellular carcinoma. *PeerJ* 8:e8930. 10.7717/peerj.8930
374 Maire V, Nemati F, Richardson M, Vincent-Salomon A, Tesson B, Rigaill G, Gravier E, Marty-Prouvost B, De Koning L,
375 Lang G, Gentien D, Dumont A, Barillot E, Marangoni E, Decaudin D, Roman-Roman S, Pierre A, Cruzalegui F,
376 Depil S, Tucker GC, and Dubois T. 2013. Polo-like kinase 1: a potential therapeutic option in combination
377 with conventional chemotherapy for the management of patients with triple-negative breast cancer.
378 *Cancer Res* 73:813-823. 10.1158/0008-5472.CAN-12-2633
379 Metzger-Filho O, Tutt A, de Azambuja E, Saini KS, Viale G, Loi S, Bradbury I, Bliss JM, Azim HA, Jr., Ellis P, Di Leo A,
380 Baselga J, Sotiriou C, and Piccart-Gebhart M. 2012. Dissecting the heterogeneity of triple-negative breast
381 cancer. *J Clin Oncol* 30:1879-1887. 10.1200/JCO.2011.38.2010
382 Mu Y, Zou H, Chen B, Fan Y, and Luo S. 2017. FAM83D knockdown regulates proliferation, migration and invasion of
383 colorectal cancer through inhibiting FBXW7/Notch-1 signalling pathway. *Biomed Pharmacother* 90:548-554.
384 10.1016/j.biopha.2017.03.073
385 Ou-Yang F, Pan MR, Chang SJ, Wu CC, Fang SY, Li CL, Hou MF, and Luo CW. 2019. Identification of CHD4-beta1 integrin
386 axis as a prognostic marker in triple-negative breast cancer using next-generation sequencing and
387 bioinformatics. *Life Sci* 238:116963. 10.1016/j.lfs.2019.116963
388 Pan X, Hu X, Zhang YH, Chen L, Zhu L, Wan S, Huang T, and Cai YD. 2019. Identification of the copy number variant
389 biomarkers for breast cancer subtypes. *Mol Genet Genomics* 294:95-110. 10.1007/s00438-018-1488-4
390 Riahi A, Hosseinpour-Feizi M, Rajabi A, Akbarzadeh M, Montazeri V, and Safaralizadeh R. 2020. Overexpression of a
391 Novel Long Non-Coding RNA MCM3AP-AS1 in breast cancer tissues compared to adjacent non-tumour
392 tissues. *Br J Biomed Sci*.
393 Santamaria A, Nagel S, Sillje HHW, and Nigg EA. 2008. The spindle protein CHICA mediates localization of the
394 chromokinesin Kid to the mitotic spindle. *Curr Biol* 18:723-729. 10.1016/j.cub.2008.04.041
395 Serra KP, Ramalho S, Torresan R, Vassallo J, Sarian LO, Silva GR, and Derchain S. 2014. [The new classification of
396 breast cancers: finding the luminal A]. *Rev Bras Ginecol Obstet* 36:575-580. 10.1590/so100-
397 720320140005158
398 Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T,
399 Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, and Borresen-Dale AL. 2001. Gene expression
400 patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U*
401 *S A* 98:10869-10874. 10.1073/pnas.191367098
402 Stovgaard ES, Bokharaey M, List-Jensen K, Roslind A, Kumler I, Hogdall E, Nielsen D, and Balslev E. 2020. PD-L1
403 diagnostics in the neoadjuvant setting: implications of intratumoral heterogeneity of PD-L1 expression in
404 triple negative breast cancer for assessment in small biopsies. *Breast Cancer Res Treat*. 10.1007/s10549-
405 020-05655-w
406 Sugimasa H, Taniue K, Kurimoto A, Takeda Y, Kawasaki Y, and Akiyama T. 2015. Heterogeneous nuclear
407 ribonucleoprotein K upregulates the kinetochore complex component NUF2 and promotes the
408 tumorigenicity of colon cancer cells. *Biochem Biophys Res Commun* 459:29-35. 10.1016/j.bbrc.2015.02.043
409 Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P,
410 Jensen LJ, and Mering CV. 2019. STRING v11: protein-protein association networks with increased coverage,
411 supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47:D607-D613.
412 10.1093/nar/gky1131

Thang PM, Takano A, Yoshitake Y, Shinohara M, Murakami Y, and Daigo Y. 2016. Cell division cycle associated 1 as a novel prognostic biomarker and therapeutic target for oral cancer. *Int J Oncol* 49:1385-1393. 10.3892/ijo.2016.3649

The Gene Ontology C. 2019. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* 47:D330-D338. 10.1093/nar/gky1055

Vagia E, Mahalingam D, and Cristofanilli M. 2020. The Landscape of Targeted Therapies in TNBC. *Cancers (Basel)* 12. 10.3390/cancers12040916

Walian PJ, Hang B, and Mao JH. 2016. Prognostic significance of FAM83D gene expression across human cancer types. *Oncotarget* 7:3332-3340. 10.18632/oncotarget.6620

Wang D, Han S, Peng R, Wang X, Yang XX, Yang RJ, Jiao CY, Ding D, Ji GW, and Li XC. 2015. FAM83D activates the MEK/ERK signaling pathway and promotes cell proliferation in hepatocellular carcinoma. *Biochem Biophys Res Commun* 458:313-320. 10.1016/j.bbrc.2015.01.108

Wang Z, Liu Y, Zhang P, Zhang W, Wang W, Curr K, Wei G, and Mao JH. 2013. FAM83D promotes cell proliferation and motility by downregulating tumor suppressor gene FBXW7. *Oncotarget* 4:2476-2486. 10.18632/oncotarget.1581

Wein L, and Loi S. 2017. Mechanisms of resistance of chemotherapy in early-stage triple negative breast cancer (TNBC). *Breast* 34 Suppl 1:S27-S30. 10.1016/j.breast.2017.06.023

Wu W, Wu F, Wang Z, Di J, Yang J, Gao P, Jiang B, and Su X. 2017. CENPH Inhibits Rapamycin Sensitivity by Regulating GOLPH3-dependent mTOR Signaling Pathway in Colorectal Cancer. *J Cancer* 8:2163-2172. 10.7150/jca.19940

Wu X, Lin Y, Shi L, Huang Y, Lai C, Wang Y, Zhang M, Wang S, Heng B, Yu G, Du X, Fang L, Fu Y, Chen J, Guo Z, Su Z, and Wu S. 2015. Upregulation of centromere protein H is associated with progression of renal cell carcinoma. *J Mol Histol* 46:377-385. 10.1007/s10735-015-9635-2

Wu Y, Qin J, Li F, Yang C, Li Z, Zhou Z, Zhang H, Li Y, Wang X, Liu R, Tao Q, Chen W, and Chen C. 2019. USP3 promotes breast cancer cell proliferation by deubiquitinating KLF5. *J Biol Chem* 294:17837-17847. 10.1074/jbc.RA119.009102

Xu W, Wang Y, Wang Y, Lv S, Xu X, and Dong X. 2019. Screening of differentially expressed genes and identification of NUF2 as a prognostic marker in breast cancer. *Int J Mol Med* 44:390-404. 10.3892/ijmm.2019.4239

Yandim C, and Karakulah G. 2019. Dysregulated expression of repetitive DNA in ER+/HER2- breast cancer. *Cancer Genet* 239:36-45. 10.1016/j.cancergen.2019.09.002

Zhang T, Zhou Y, Qi ST, Wang ZB, Qian WP, Ouyang YC, Shen W, Schatten H, and Sun QY. 2015. Nuf2 is required for chromosome segregation during mouse oocyte meiotic maturation. *Cell Cycle* 14:2701-2710. 10.1080/15384101.2015.1058677

Zhao WF, Wang HB, Xie B, Hu LJ, Xu LH, Kuang BH, Li MZ, and Zhang X. 2012. Sp1 and Sp3 are involved in the full transcriptional activity of centromere protein H in human nasopharyngeal carcinoma cells. *FEBS J* 279:2714-2726. 10.1111/j.1742-4658.2012.08654.x

Zhu L, Wang Z, Wang W, Wang C, Hua S, Su Z, Brako L, Garcia-Barrio M, Ye M, Wei X, Zou H, Ding X, Liu L, Liu X, and Yao X. 2015. Mitotic Protein CSPP1 Interacts with CENP-H Protein to Coordinate Accurate Chromosome Oscillation in Mitosis. *J Biol Chem* 290:27053-27066. 10.1074/jbc.M115.658534

Table 1(on next page)

Primers sequence

Primers sequence of target gene and internal reference gene

1 Table1 Primers sequence of target gene and internal reference gene

Gene	Primers
NUF2	Forward Primer: 5'-TACCATTCACCAATTTAGTTACT-3'
	Reverse Primer: 5'-TAGAATATCAGCAGTCTCAAAG-3'
FAM83D	Forward Primer: 5'-AGTTCCGAATCCTGTATGCC-3'
	Reverse Primer: 5'-GCTCCTTGGACTGTGGTTT-3'
CENPH	Forward Primer: 5'-CCTTATTTTGGGGAGTAAAGTCAAT-3'
	Reverse Primer: 5'-ACAAATGCACAGAAGTATTCCAAAT-3'
GAPDH	Forward Primer: 5'- AGGTCGGTGTGAACGGATTTG -3'
	Reverse Primer: 5'- GGGGTCGTTGATGGCAACA -3'

2

Figure 1

Venn diagrams of the differentially expressed genes (DEGs)

Venn diagrams of the differentially expressed genes (DEGs) from three independent data sets (GSE45827, GSE38959, GSE65194). (A)105 up-regulated DEGs were identified and (B) 56 down-regulated DEGs were identified, using adjust *P* value < 0.01 and |logFC|≥ 2 as cut-off criteria.

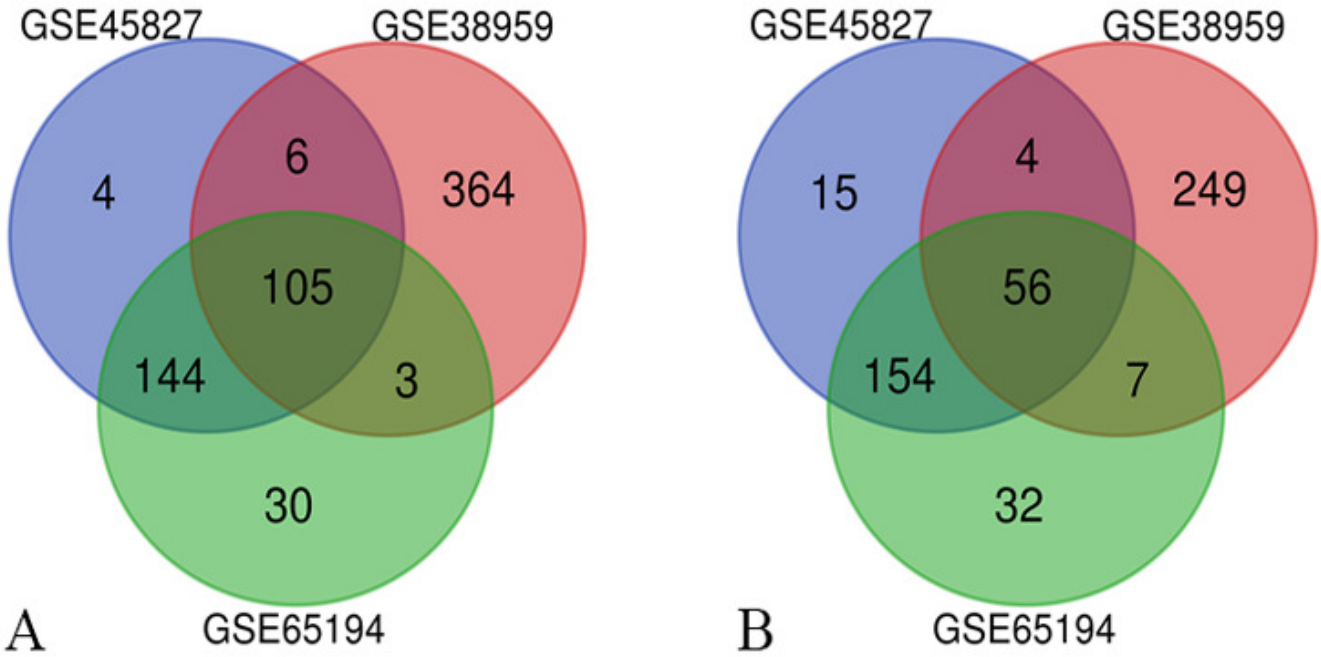

Figure 2

GO enrichment analysis of the differentially expressed genes (DEGs)

GO enrichment analysis of the differentially expressed genes (DEGs). BP: biological processes, CC: cellular components, MF: molecular functions.

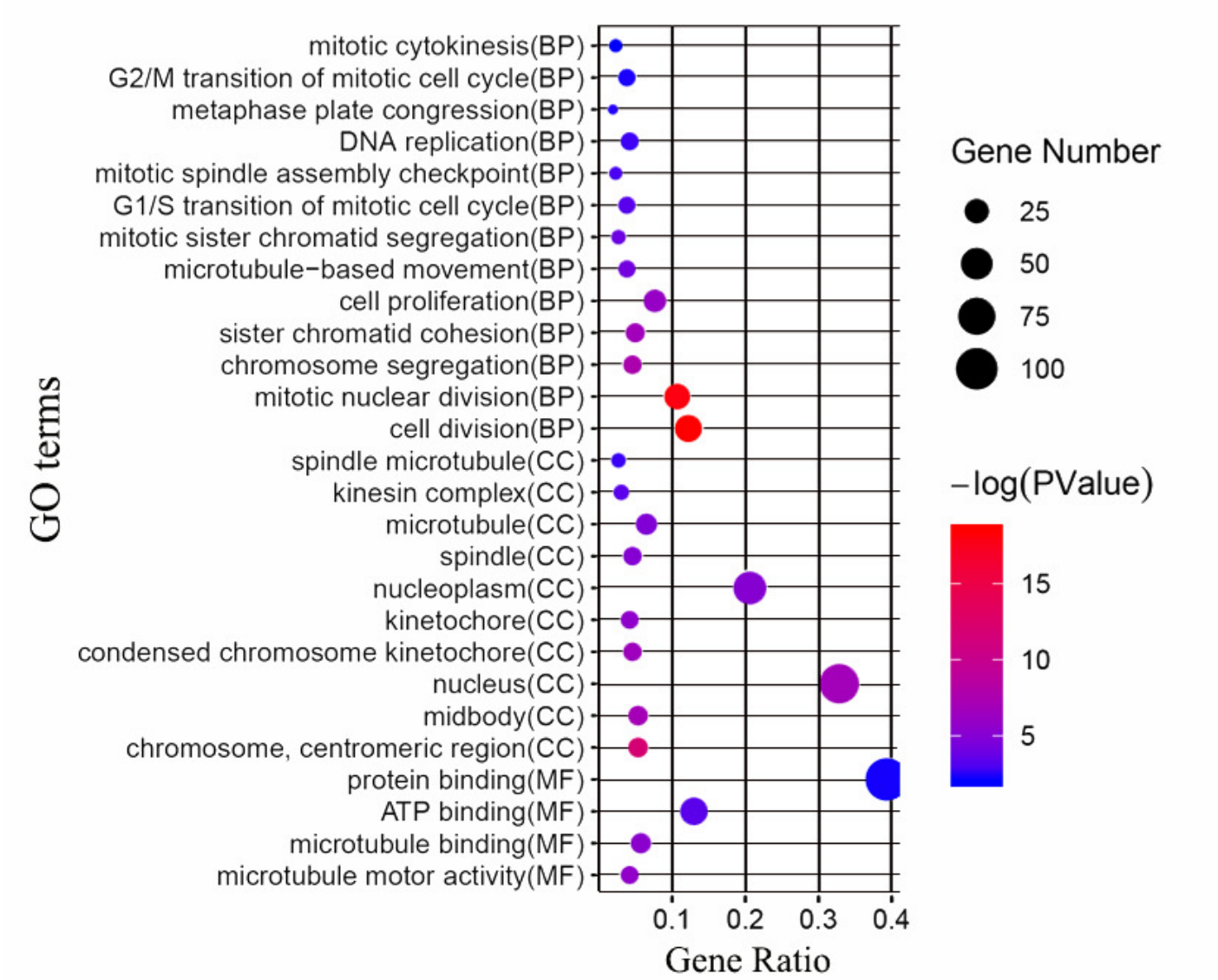

Figure 3

Protein–protein interaction network of the differentially expressed genes (DEGs)

(A) Protein–protein interaction network of the differentially expressed genes (DEGs). Red color represents up-regulated genes, Green color represents down-regulated genes. (B) Identification of the top 10 hub DEGs by cytoHubba plugin. The rank is represented by different degrees of color (from red to yellow).

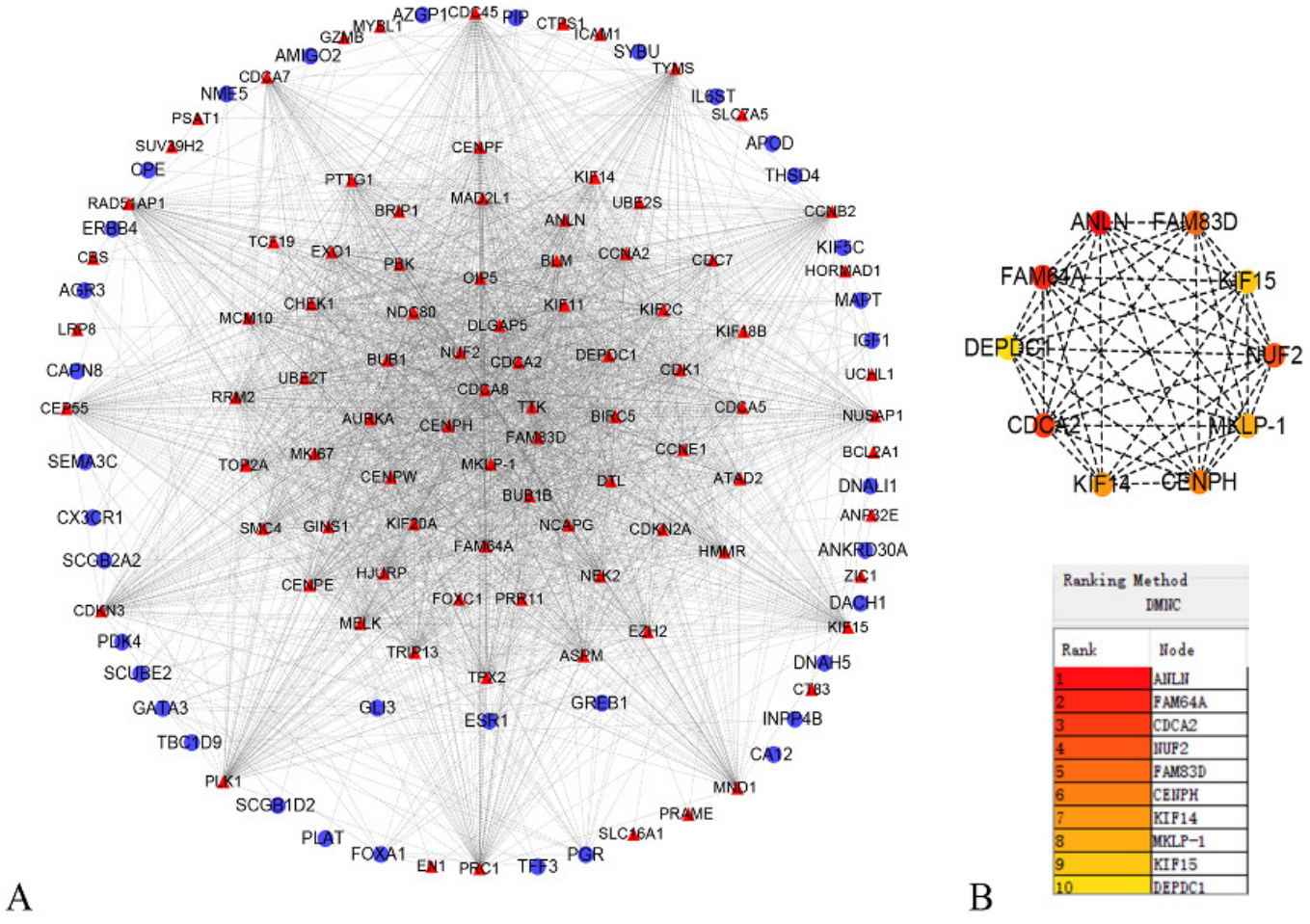

Figure 4

The correlation of 10 hub genes expression levels with the recurrence-free survival of triple-negative breast cancer (TNBC) samples.

The correlation of 10 hub genes expression levels with the recurrence-free survival of triple-negative breast cancer (TNBC) samples. TNBC patients with higher expression levels of NUF2, FAM83D, CENPH have significantly decreased recurrence-free survival ($P<0.05$), but not ANLN, FAM64A, CDCA2, KIF14, MKLP-1, KIF15, DEPDC1 ($P>0.05$).

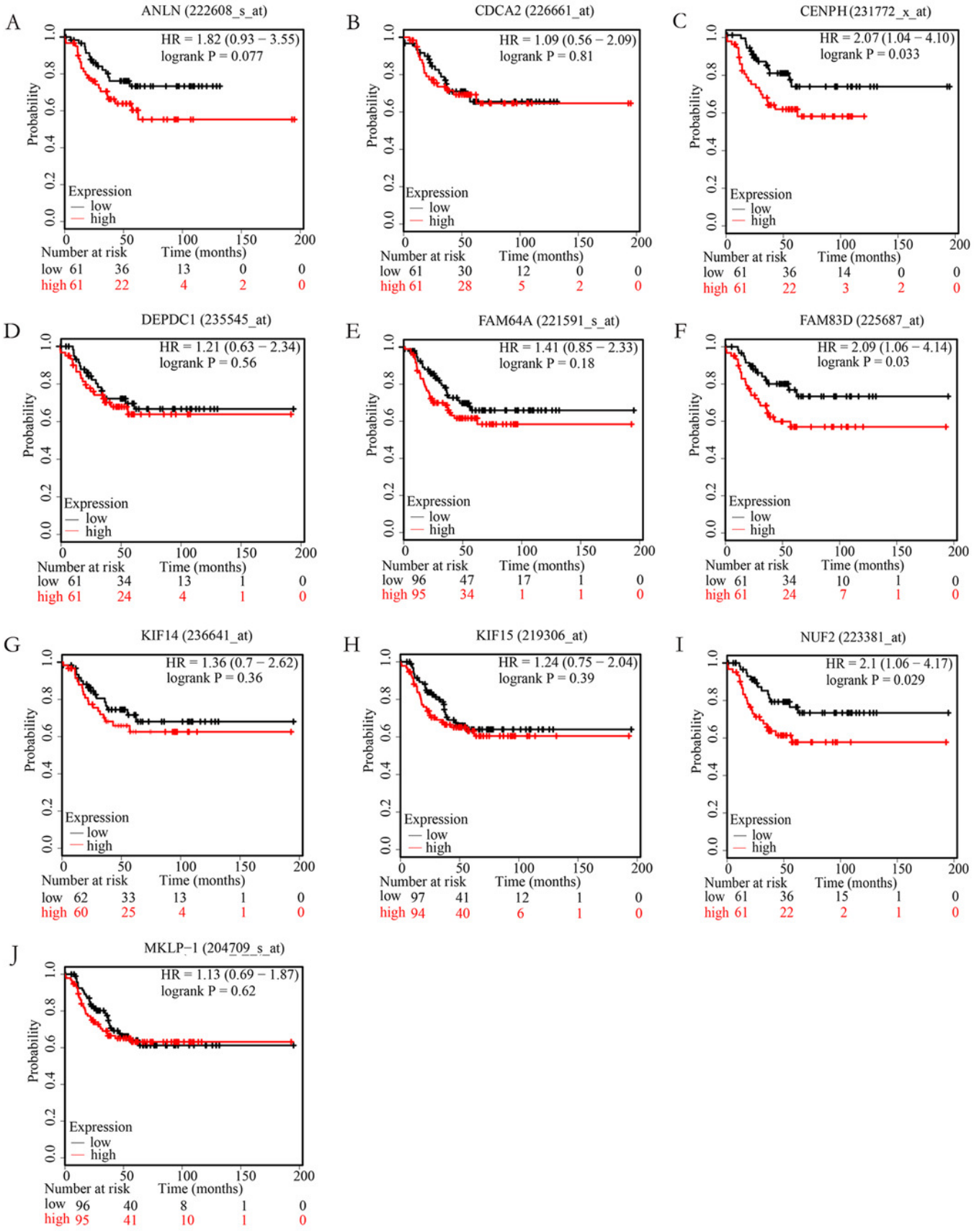
A
ANLN (222608_s_at)
HR = 1.82 (0.93 − 3.55)
logrank P = 0.077
Probability
Expression
low
high
Number at risk
Time (months)
low 61 36 13 0 0
high 61 22 4 2 0
B
CDCA2 (226661_at)
HR = 1.09 (0.56 − 2.09)
logrank P = 0.81
Probability
Expression
low
high
Number at risk
Time (months)
low 61 30 12 0 0
high 61 28 5 2 0
C
CENPH (231772_x_at)
HR = 2.07 (1.04 − 4.10)
logrank P = 0.033
Probability
Expression
low
high
Number at risk
Time (months)
low 61 36 14 0 0
high 61 22 3 2 0
D
DEPDC1 (235545_at)
HR = 1.21 (0.63 − 2.34)
logrank P = 0.56
Probability
Expression
low
high
Number at risk
Time (months)
low 61 34 13 1 0
high 61 24 4 1 0
E
FAM64A (221591_s_at)
HR = 1.41 (0.85 − 2.33)
logrank P = 0.18
Probability
Expression
low
high
Number at risk
Time (months)
low 96 47 17 1 0
high 95 34 1 1 0
F
FAM83D (225687_at)
HR = 2.09 (1.06 − 4.14)
logrank P = 0.03
Probability
Expression
low
high
Number at risk
Time (months)
low 61 34 10 1 0
high 61 24 7 1 0
G
KIF14 (236641_at)
HR = 1.36 (0.7 − 2.62)
logrank P = 0.36
Probability
Expression
low
high
Number at risk
Time (months)
low 62 33 13 1 0
high 60 25 4 1 0
H
KIF15 (219306_at)
HR = 1.24 (0.75 − 2.04)
logrank P = 0.39
Probability
Expression
low
high
Number at risk
Time (months)
low 97 41 12 1 0
high 94 40 6 1 0
I
NUF2 (223381_at)
HR = 2.1 (1.06 − 4.17)
logrank P = 0.029
Probability
Expression
low
high
Number at risk
Time (months)
low 61 36 15 1 0
high 61 22 2 1 0
J
MKLP−1 (204709_s_at)
HR = 1.13 (0.69 − 1.87)
logrank P = 0.62
Probability
Expression
low
high
Number at risk
Time (months)
low 96 40 8 1 0
high 95 41 10 1 0

Figure 5

The relative expression levels of NUF2, FAM83D and CENPH mRNA in 14 pairs of triple-negative breast cancer (TNBC) paired tissues

The mRNA expression levels of NUF2 and FAM83D were increased significantly in most TNBC lesions compared with para-adjacent tissues, but not CENPH. ***P<0.001.