Annals of Diagnostic Pathology 50 (2021) 151675

Contents lists available at ScienceDirect

Annals of Diagnostic Pathology

journal homepage: www.elsevier.com/locate/anndiagpath

Original Contribution

# Identification of potential prognostic biomarkers for breast cancer using WGCNA and PPI integrated techniques

Guili Zheng [a,1], Cong Zhang [b,1], Chen Zhong [a,*]

[a] Department of Oncology, No. 960 Hospital of PLA, Jinan, China
[b] Department of Radiation Oncology, No. 960 Hospital of PLA, Jinan, China

ARTICLE INFO

*Keywords:*
Breast cancer
Bioinformatics
Prognosis
Biomarker

ABSTRACT

In this study, we aimed to detect promising prognostic factors of breast cancer and interpreted the relevant mechanisms using an integrated bioinformatics analysis. RNA sequencing profile of breast cancer was downloaded from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) databases, which were combined as a group (TCGA_GTEx). GSE70947 dataset was from Gene Expression Omnibus. Blue and turquoise modules, respectively identified in TCGA_GTEx database and GSE70947 dataset using weighted co-expression network analysis (WGCNA), were both notably associated with breast cancer. By comparing genes in the two significant modules with differentially expressed genes (DEGs), we obtained a set of 40 shared genes, which were mainly enriched in chromosome segregation and mismatch repair pathway. After protein-protein interaction (PPI) network and overall survival analysis, two hub genes EXO1 and KIF4A were extracted from the set of 40 shared genes, which were up-regulated and associated with the dismal outcome of breast cancer patients. There was a notable negative correlation between EXO1 and KIF4A expression and age of breast cancer patients, whereas a positive relationship with two another clinical traits stage and tumor category was detected. Univariate and multivariate Cox regression analysis revealed that the two hub genes could be independent prognostic factors of breast cancer. Mechanistically, gene correlation analysis suggested that EXO1 and KIF4A exerted their oncogenic role via promoting breast cancer cell proliferation. Overall, our findings identify two promising individual prognostic predictors of breast cancer and pave the new way for diagnosis and therapy strategy of breast cancer.

## 1. Introduction

Breast cancer has led to more than two million new cases (11.6% of total cancer cases) and 626,000 deaths (6.6% of total cancer deaths) globally in 2018, becoming the second most common cancer representing the fourth largest cancer deaths [1]. About 70% patients with breast cancer at early stage are completely curable, whereas advanced breast cancer is commonly fatal [2]. With high risks of drug resistance, relapse and metastasis, the median overall survival of advanced breast cancer is extremely short at only 31.1 months after treatments [3]. Recently, more and more research focus on the identification of novel molecules crucial to breast cancer to improve the diagnosis, prognosis and therapy strategies of the cancer.

Comprehensive bioinformatics analysis has emerged as a powerful technique for identification of cancer-related biomarkers, pathways and drug targets. The Cancer Genome Atlas (TCGA), Genotype-Tissue Expression (GTEx) and Gene Expression Omnibus (GEO) databases are three commonly used publicly accessible repositories, in which genomic, epigenomic, transcriptomic and proteomic data provide foundation for integrated bioinformatics analysis [4-6]. Weighted gene co-expression network analysis (WGCNA) is a systematical bioinformatics tool to identify the modules of highly relevant genes, which has been widely applied in the detection of disease-related biomarkers, such as cancers, neurodegenerative diseases, and immune disease [7,8]. In this study, we performed an integrated analysis using transcriptomic profiles of breast cancer patients from TCGA, GTEx and GEO databases. WGCNA, protein-protein interaction (PPI) network and univariate and multivariate Cox regression analyses were conducted to identify prognostic molecules in breast cancer. Moreover, gene functional annotation and correlation analysis were performed to explore the biological roles

* Corresponding author at: Department of Oncology, No. 960 Hospital of PLA, NO. 25 Shifan Road, Tianqiao District, Jinan, Shandong 250031, China.
*E-mail address:* zhongchen0504@sina.com (C. Zhong).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.anndiagpath.2020.151675

Available online 3 December 2020

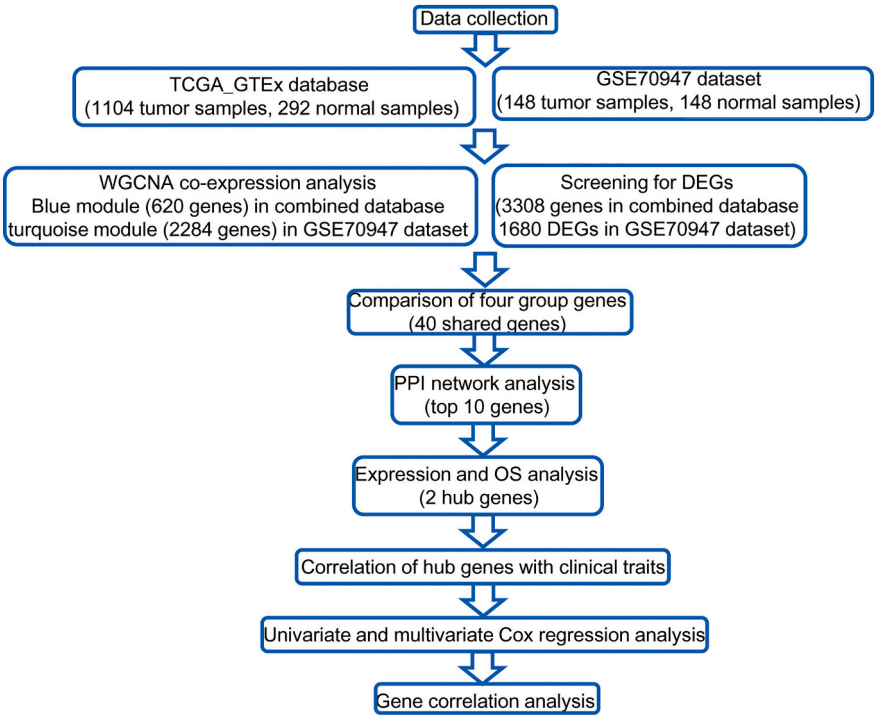


**Fig. 1.** Workflow of research path used in the present study.

of potential hub genes and relevant mechanisms in breast cancer.

## 2. Materials and methods

### 2.1. Data collection

RNA sequencing (RNA-seq) data of breast cancer from The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression (GTEx) databases were downloaded from UCSC Xena (http://xena.ucsc.edu/), an online exploration tool. We combined 1104 breast cancer samples and 113 normal samples from TCGA with 179 normal breast samples from GTEx to obtain 1396 samples including 1104 breast cancer samples and 292 normal samples, referred as a new group TCGA_GTEx database. Besides, GSE70947 dataset containing 148 breast cancer samples and 148 normal samples were downloaded from Gene Expression Omnibus (GEO) database. Clinical data including overall survival, stage, grade, age, tumor node metastasis (TNM) categories and gender were downloaded from TCGA.

### 2.2. Construction of weighted gene co-expression network

A scale-free weighted gene co-expression network was constructed using "WGCNA" package in R to mine co-expressed genes and modules [9]. First, the data were clustered to detect the outliers. Second, the thresholding power β was set to 3. In this way, the network was closely scale-free. And the hierarchical clustering tree was constructed to detect gene modules. Third, gene modules were detected by average-linkage hierarchical clustering according to a topological overlap matrix (TOM)-based dissimilarity measure. MEDissThres was set to 0.25 to automatically merge the similar modules. Finally, Pearson's correlation coefficient was calculated to evaluate the correlation between the modules and clinical features.

### 2.3. Screening of differentially expressed genes (DEGs) in breast cancer

The edgeR (Empirical Analysis of Digital Gene Expression Data in R) package in R was utilized to screen DEGs in response to breast cancer. DEGs must meet the criterion: the FDR value $< 0.05$ and $|\log FC| > 1$.

### 2.4. Gene functional annotation analysis

To explore the underlying mechanisms of interested genes, the "clusterProfiler" package in R was downloaded to perform gene ontology (GO) including gene molecular function (MF), biological process (BP), cellular component (CC) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis.

### 2.5. Analysis of protein-protein interaction (PPI)

The Search Tool for the Retrieval of Interacting Genes (STRING) database (https://string-db.org/) was utilized to analyze the interaction between proteins with the confidence score $> 0.7$. Cytoscape plugin cytoHubba (version 0.1) was used to construct PPI network. Top 10 genes were selected as the candidates of hub genes according to Maximal Clique Centrality (MCC) algorithm.

### 2.6. Overall survival (OS) analysis

The samples with survival information were divided into two groups based on the median expression of hub gene. Then "survival" package in R was downloaded to evaluate the correlation between the hub genes and OS. Besides, prognostic value of hub genes was analyzed by univariate and multivariate Cox regression analysis. $P < 0.05$ was seen as significant.

### 2.7. Analysis of correlation between clinical features and the expression of hub genes

The "cor" package in R was downloaded to analyze the correlation between clinical features and the expression of hub genes. In the present study, clinical features including age, stage, TNM categories and gender were downloaded from TCGA database. $P < 0.05$ was seen as significant.

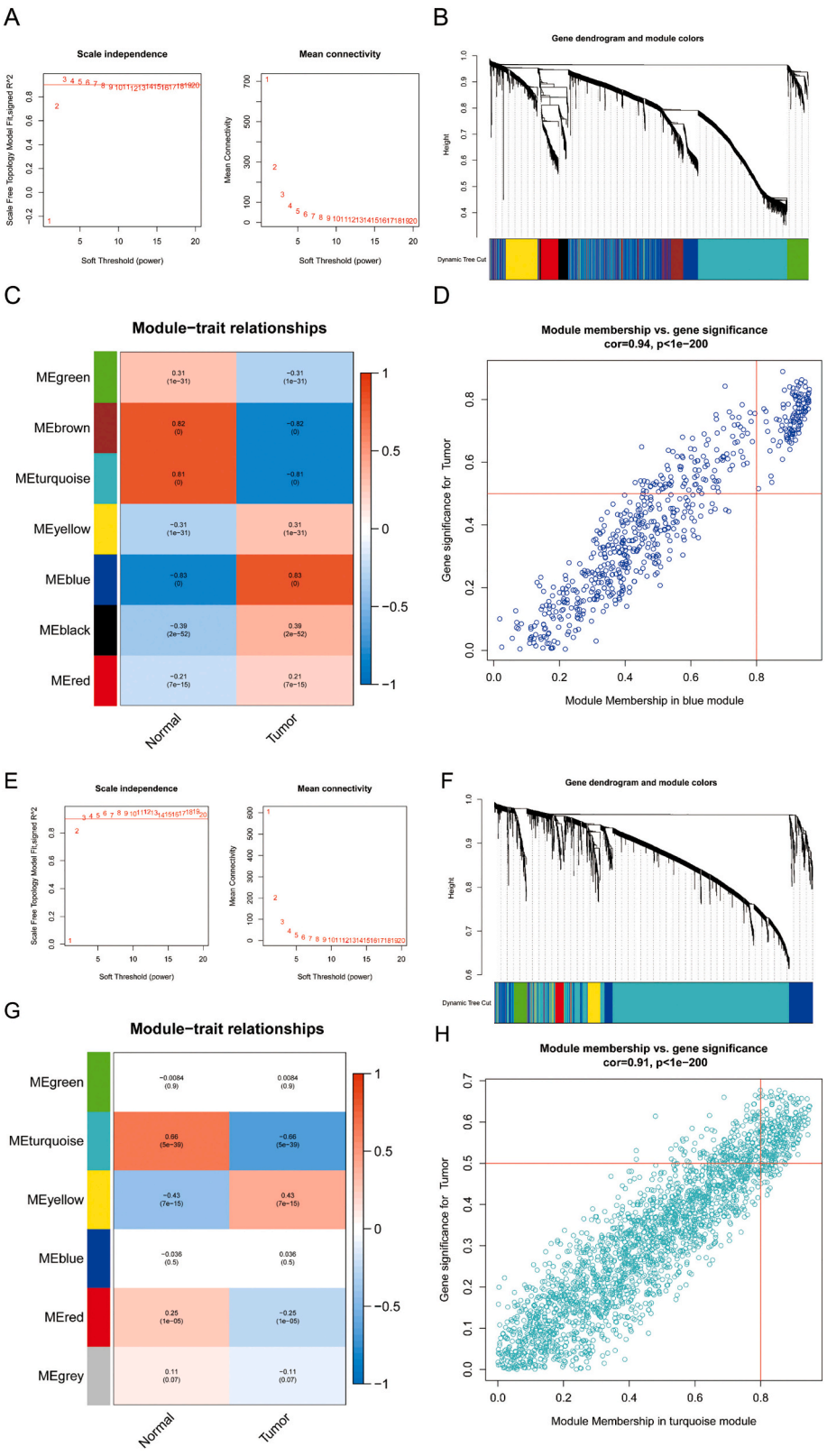


**Fig. 2.** Identification of significant gene modules. (A) When β was set at 3, scale-free network was constructed in TCGA_GTEx database. (B) Clustering dendrograms of genes based on dissimilarity topological overlap and module colors in TCGA_GTEx database. (C) Heatmap showing the correlation of gene modules with breast cancer samples or normal samples in TCGA_GTEx database. (D) Scatter plots of gene significance relative to module membership in TCGA_GTEx database (D). (E) When β was set at 3, scale-free network was constructed in GSE70947 dataset. (F) Clustering dendrograms of genes in GSE70947 dataset. (G) Heatmap in GSE70947 dataset. (H) Scatter plots in GSE70947 dataset. $P < 0.05$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 2.8. Gene correlation analysis

To explore the possible mechanism of how hub genes regulated the progression of breast cancer, we evaluated the correlation between the expression of hub genes and proliferation markers including CCNA2, MKI67 and PCNA using Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/index.html) with TCGA database. $P < 0.05$ was seen as significant.

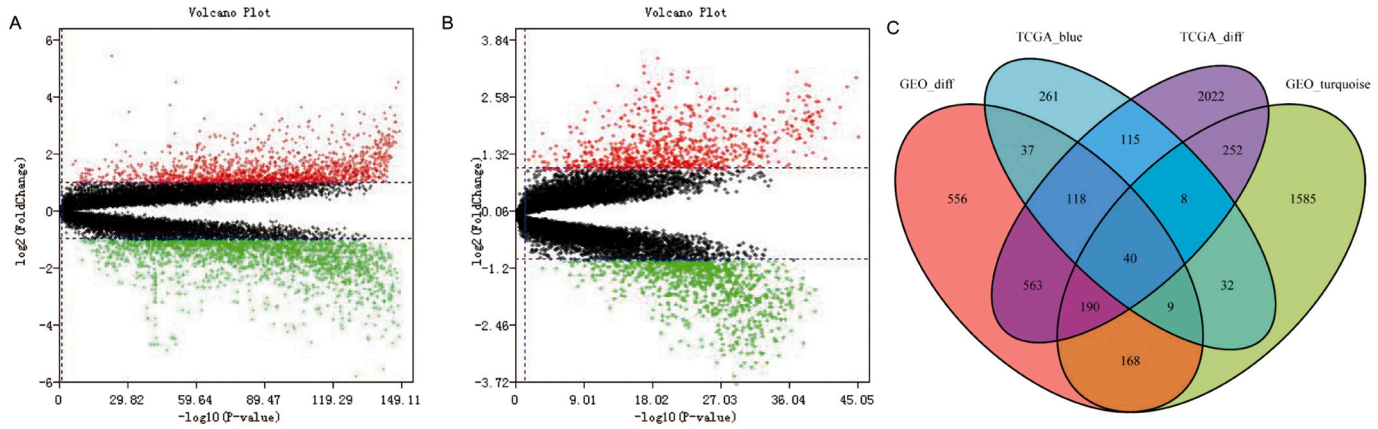


**Fig. 3.** Analysis of differentially expressed genes (DEGs). (A) Volcano plot of gene expression in TCGA_GTEx database. (B) Volcano plot of gene expression in GSE70947 dataset. (C) Venn diagram displaying the number of genes in different groups. GEO_diff, DEGs in GSE70947 dataset; TCGA_blue, genes of blue module in TCGA_GTEx database; TCGA_diff, DEGs in TCGA_GTEx database; GEO_turquoise, genes of turquoise module in GSE70947 dataset.

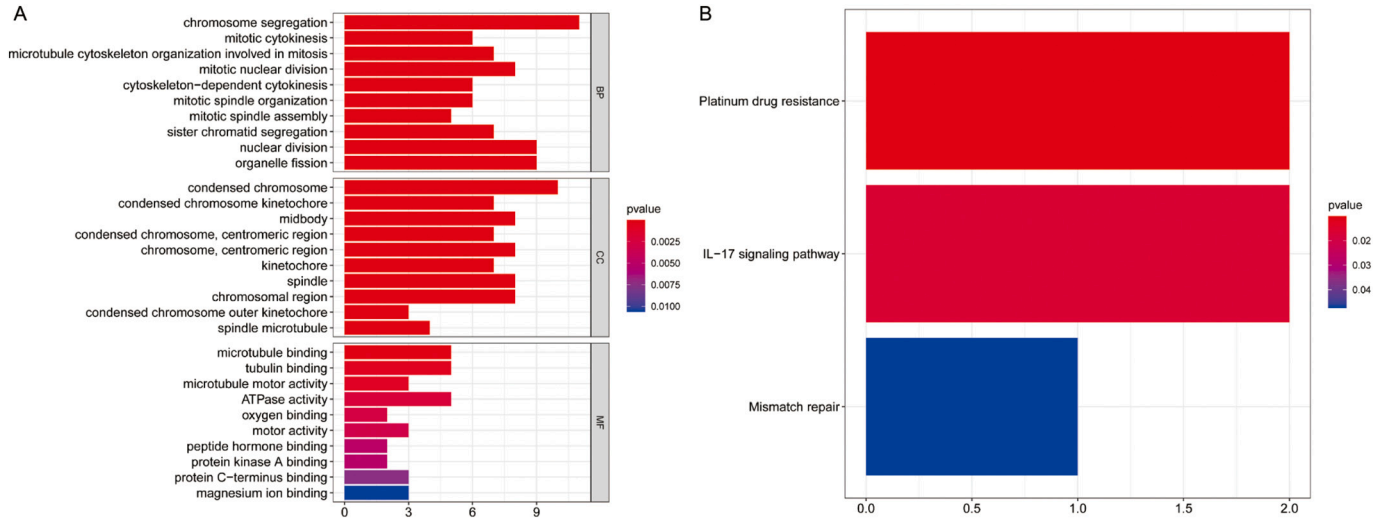


**Fig. 4.** Gene functional annotation of the set of 40 shared genes. (A) GO analysis of the 40 shared genes including biological process (BP), cellular component (CC) and molecular function (MF). (B) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the 40 shared genes. $P < 0.05$.

## 3. Results

### 3.1. Identification of significant gene modules

Fig. 1 displayed the research path in the present study. We first constructed a WGCNA network with breast samples data from the TCGA_GTEx database, and pickSoftThreshold function in WGCNA was used to calculate soft thresholding power β [10]. β was set at 3 with scale independence at 0.9 and relative high mean connectivity (Fig. 2A). Seven co-expressed gene modules were clustered, among which the blue module (620 genes) was the one most relevant to breast cancer (R = 0.83, $P \ll 0.05$, Fig. 2B and C). Importantly, there was a significant correlation between the blue module and module-relevant genes (R = 0.94, Fig. 2D, $P \ll 0.05$).

Moreover, the same analysis was performed with GSE70947 dataset. When β = 3, the scale-free network was constructed (Fig. 2E). Six modules with co-expressed genes were clustered (Fig. 2F and G). The strongest correlation was found between the turquoise module (2284 genes) and breast cancer (R = 0.83, $P \ll 0.05$, Fig. 2G). Besides, Fig. 2H revealed that genes involved in turquoise module was strongly associated with this module (R = 0.91, $P \ll 0.05$).

### 3.2. The analysis of DEGs

Next, we screened the DEGs in combined databases and GSE70947 dataset with the "edgeR" package. In TCGA_GTEx database, the volcano plot displayed that 3308 DEGs including 1523 up-regulated and 1785 down-regulated genes were obtained by applying the FDR value < 0.05 and |logFC| > 1 criterion (Fig. 3A). Meanwhile, in GSE70947 dataset, 1680 DEGs including 832 up-regulated and 848 down-regulated genes were obtained with the same criterion (Fig. 3B). Following, we compared the co-expressed genes in blue and turquoise modules with the DEGs, then a set of 40 shared genes were obtained (Fig. 3C, Supplementary Table 1).

### 3.3. Gene ontology (GO) and pathway enrichment analysis

To interpret the possible biological roles of the 40 shared genes, we subjected them to GO and KEGG analysis (Fig. 4A and B). The 40 common genes were enriched in 375 GO terms totally. As for biological process (BP), the genes were mainly enriched in chromosome segregation and mitotic cytokinesis (Fig. 4A). In regard to cellular component (CC), condensed chromosome and midbody were mainly enriched (Fig. 4A). Regarding molecular function (MF), the genes were mainly

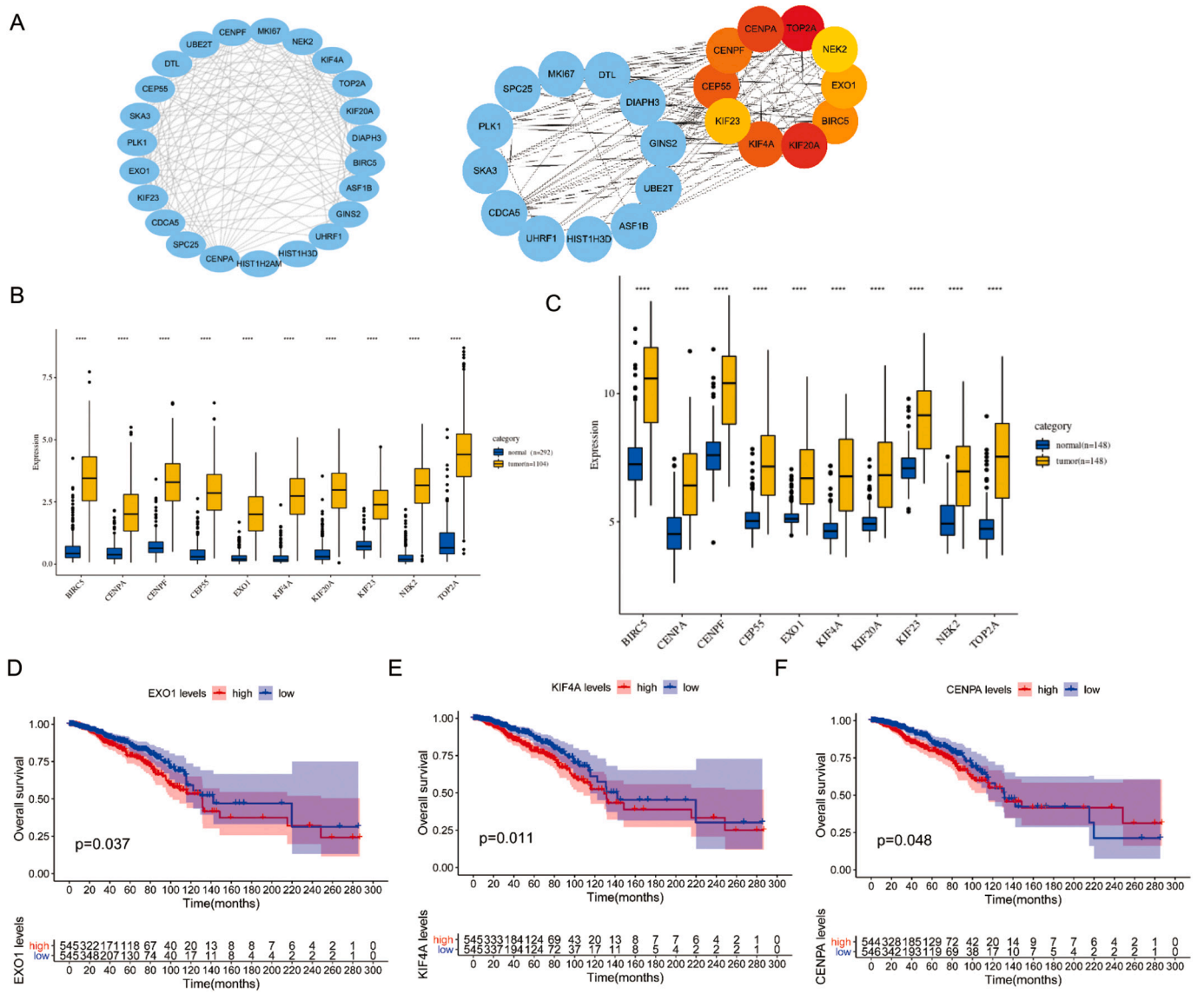


**Fig. 5.** Identification of hub genes. (A) PPI network analysis. 23 nodes and 147 edges were obtained with confidence score ≥0.7 (left). Top ten genes were selected as candidate hub genes based on MCC algorithm (right). (B) Expression of the 10 potential hub genes in breast cancer tissues and normal samples in TCGA_GTEx database. (C) Expression of the 10 potential hub genes in breast cancer tissues and normal samples in GSE70947 dataset. (D) Overall survival plot in response to EXO1 expression. (E) Overall survival plot in response to KIF4A expression. (E) Overall survival plot in response to CENPA expression. $P < 0.05$.

enriched in microtubule binding and ATPase activity (Fig. 4A). In addition, the 40 common genes were enriched in 3 KEGG pathways, including platinum drug resistance, IL-17 signaling pathway and mismatch repair (Fig. 4B, $P < 0.05$). As most of the major enriched GO terms and KEGG pathways were involved in cell proliferation [11-13], we hypothesized that the set of 40 shared genes impacted the progression of breast cancer via regulating this biological behavior of cancer cell.

### 3.4. Identification of hub genes

To extract hub genes from the set of 40 common genes tightly associated with the progression of breast cancer, we performed the following integrated bioinformatics analysis. First, the 40-gene set were subjected to String analysis, and the PPI network was constructed with Cytoscape as shown in Fig. 5A. 23 nodes and 147 edges were obtained with confidence score ≥ 0.7. CytoHubba was used to evaluate the importance of nodes or subnetworks with Maximal Clique Centrality (MCC) algorithm [14]. Here, top ten genes were selected as candidates of hub genes based on MCC algorithm. Fig. 5A showed that the 10 candidate hub genes were CENPA, TOP2A, NEK2, EXO1, BIRC5, KIF20A, KIF4A, KIF23, CEP55 and CENPF.

Second, to confirm the reliability of the 10 aforementioned genes, we investigated the expression of them in breast cancer samples. As for TCGA_GTEx database, the 10 potential hub genes were all highly expressed in breast cancer tissues (n = 1104) compared to the normal tissues (n = 292, Fig. 5B, $P < 0.05$). And the same results were obtained in GSE70947 (Fig. 5C, $P < 0.05$).

Finally, Kaplan Meier (KM) curves were plotted to evaluate the relationship of the expression of 10 potential hub genes with the OS of breast cancer patients. Fig. 5D–F revealed that there was a significant correlation between the expression of three genes including EXO1, KIF4A and CENPA with the OS of breast cancer patients ($P < 0.05$). Patients with high expression of EXO1 or KIF4A displayed notably shorter survival time than those with low expression of EXO1 or KIF4A (Fig. 5D and E). However, KM curves for CENPA high- and low-expression groups overlapped and interacted with each other in the last potion of KM curves (Fig. 5F). In addition, no notable relationship of

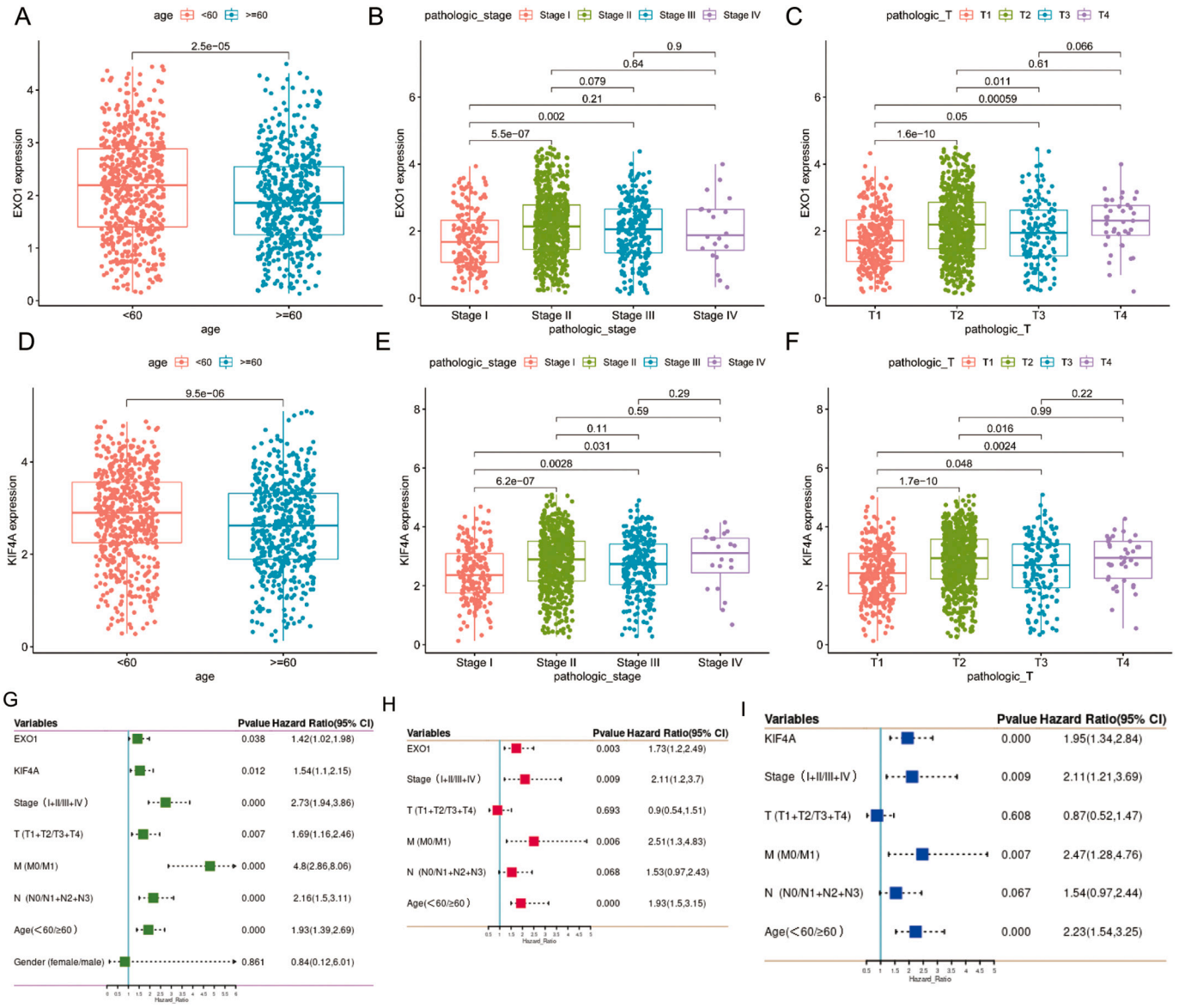


**Fig. 6.** Analysis of prognostic value of hub genes. (A–C) The expression of EXO1 in response to age (A), stages (B) and T category (C) of breast cancer patients. (D–F) the expression of KIF4A in response to age (D), stages (E) and T category (F) of breast cancer patients. (G) Univariate Cox analysis of the correlation of clinical features including stages, TNM categories, age, gender and two hub genes with the survival of breast cancer patients. (H) Multivariate Cox analysis of the correlation of clinical features including stages, TNM categories, age, gender and EXO1 expression with the survival of breast cancer patients. (I) Multivariate Cox analysis of the correlation of clinical features including stages, TNM categories, age, gender and KIF4A expression with the survival of breast cancer patients. $P < 0.05$.

the other potential hub genes with the outcome of breast cancer patients was observed (Supplementary Fig. 1A–G, $P > 0.05$). Therefore, EXO1 and KIF4A were selected as hub genes for further analysis.

### 3.5. Analysis of prognostic value of hub genes

In order to investigate the roles of two hub genes in the progression of breast cancer, we analyzed the correlation of clinical features including age, stage, TNM categories with the expression of EXO1 and KIF4A. As shown in Fig. 6A, the younger patients (age < 60 years old) displayed higher EXO1 expression level than older patients (age ≥60 years old, $P = 2.5E{-}05$); The expression of EXO1 was obviously elevated in stages II and III compared to stage I (Fig. 6B, $P < 0.05$); As for T category, patients in T2, T3 and T4 exhibited significant higher EXO1 expression than those in T1 (Fig. 6C, $P < 0.05$). Whereas, no significant correlations were found between EXO1 expression and N and M categories (Supplementary Fig. 2A and B, $P > 0.05$). In addition, similar results were obtained for KIF4A. There was a notable elevation of KIF4A in younger patients (age < 60 years old) compared to older patients (age ≥60 years old, Fig. 6D, $P = 2.5E{-}05$); Patients in stages II, III and IV showed higher KIF4A expression than those in stage I (Fig. 6E, $P < 0.05$); KIF4A expression in T2, T3 and T4 was higher than that in T1 (Fig. 6F, $P < 0.05$). Similarly, there was no significant change of KIF4A expression in different N and M status (Supplementary Fig. 2C and D, $P > 0.05$). Therefore, the expression of EXO1 and KIF4A was negatively correlated with age but positively correlated with stages and T category of breast cancer.

Furthermore, to confirm the prognostic independence of the two hub genes, clinical features including stages, TNM categories, age, gender and two hub genes were systematically analyzed (Fig. 6G–I). Univariate Cox regression analysis revealed that there was a significant correlation between the prognosis of breast cancer patients and the expression of EXO1 and KIF4A, age, stage, and TNM categories (Fig. 6G, $P < 0.05$). The corresponding multivariate Cox regression analysis revealed that

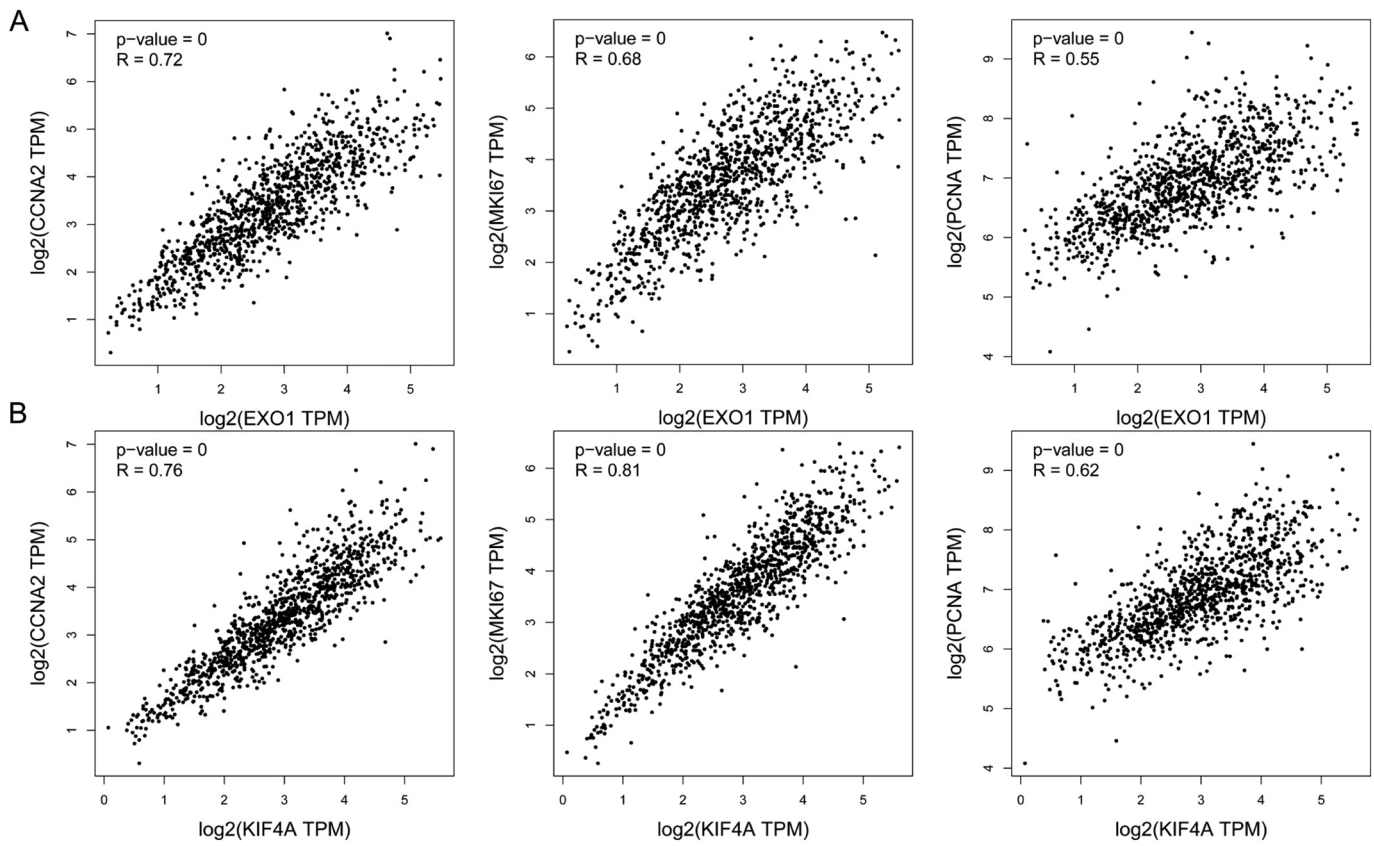


**Fig. 7.** Correlation of hub genes with proliferation markers. (A–C) The correlation between EXO1 expression and the three indexes of cell proliferation including CCNA2 (A), MKI67 (B) and PCNA (C) was analyzed using Gene Expression Profiling Interactive Analysis (GEPIA). (D–F) The correlation between KIF4A expression and the three indexes of cell proliferation including CCNA2 (D), MKI67 (E) and PCNA (F) was analyzed using GEPIA. $P < 0.01$.

the expression of EXO1 and KIF4A, stage, M category and age were independently associated with the prognosis of breast cancer (Fig. 6H and I). Collectively, EXO1 and KIF4A were hub genes and associated with the prognosis of breast cancer patients independently.

### 3.6. Correlation of hub genes with proliferation markers

Next, according to the result of gene functional annotation, we further characterized the role of the two hub genes in breast cancer cell proliferation to explore the underlying mechanism of how EXO1 and KIF4A regulated the progression of breast cancer. In the present study, we selected three indexes of cell proliferation including CCNA2, MKI67 and PCNA, which were commonly used as proliferation markers [15-17]. The correlation of EXO1 and KIF4A expression with CCNA2, MKI67 and PCNA was analyzed using GEPIA. Fig. 7 A revealed that the expression of EXO1 was positively correlated with CCNA2 ($R = 0.72$, $P < 0.001$), MKI67 ($R = 0.68$, $P < 0.001$) and PCNA ($R = 0.55$, $P < 0.001$). And similar results were obtained for KIF4A (Fig. 7B, $P < 0.001$). Therefore, EXO1 and KIF4A play oncogenic role possibly via promoting cell proliferation in breast cancer.

## 4. Discussion

In our study, we detected 7 modules in TCGA_GTEx database and 6 modules in GSE70947 dataset, of which blue and turquoise were the most highly correlated with breast cancer. By comparing the DEGs with co-expressed genes in blue and turquoise modules, we obtained a set of 40 shared genes in the intersection of Venn diagram. GO analysis revealed that the 40 genes were mainly enriched in chromosome segregation, mitotic cytokinesis, condensed chromosome, midbody, microtubule binding and ATPase activity. Besides, the 40 genes were mainly enriched in platinum drug resistance, IL-17 signaling pathway and mismatch repair. There are various studies about the critical roles of these GO terms and pathways in breast cancer. For example, Huang et al. found that PICH, a DNA-dependent ATPase, could promote proliferation of triple-negative breast cancer cells (TNBC) via ensuring the segregation of chromosomes [18]. Verma et al. used spectrofluorometry and immunofluorescence imaging to show that 9-PAN, a noscapine analogue, exerted antiproliferative role in TNBC via binding to and then damaging microtubule network [19]. Drug resistance of cancers is reported to be associated with the proliferation of cancer stem cells [20]. Kostrhunova et al. synthesized a platinum pro-drug that could figure out drug resistance and inhibit proliferation of HER2-positive breast cancer cells [21]. Combined with these studies, we hypothesized that the 40 genes were involved in breast cancer cell proliferation.

PPI network analysis of the set of 40 genes extracted 10 potential hub genes which were all up-regulated in breast cancer tissues. After OS analysis, we detected two hub genes EXO1 and KIF4A that negatively associated with OS of breast cancer patients. In addition, there was an elevation of EXO1 and KIF4A expression in patients under 60 years old compared to the patients over 60. The expression of EXO1 and KIF4A were both positively correlated with stage and T category in breast cancer. However, no significant relationship between EXO1 and KIF4A expression and N and M categories were detected. Univariate and multivariate Cox regression analysis revealed that EXO1 and KIF4A could be individual predictors of the prognosis of breast cancer. By gene correlation analysis using GEPIA, we found that the expression of EXO1 and KIF4A was positively correlated with three indexes of cell proliferation including CCNA2, MKI67 and PCNA.

Exonuclease 1 (EXO1) was identified as 5′ to 3′ nuclease and plays important roles in DNA replication, mismatch repair and recombination [22]. Loss of EXO1 can cause instability of genome and even result in

defect mismatch repair and meiosis [23]. Consistent with our results, many studies have reported that up-regulation of EXO1 is involved in the progression and prognosis of various cancers, including breast cancer. Sousa et al. demonstrated that EXO1 was highly expressed and independently associated with the poor prognosis of glioma patients [24]. And the same role of EXO1 was detected in hepatocellular carcinoma (HCC) [25]. Besides, Dai et al. revealed that loss of EXO1 inhibited HCC cell proliferation via impaired cell cycle [25]. It is worth noting that EXO1 is reported to be aberrantly highly expressed in breast cancer. For example, EXO1 was up-regulated in ductal and invasive breast carcinoma tissues and mice models [26]. By integrated bioinformatics analysis, EXO1 was screened out to be up-regulated and hypomethylated and could be an independent prognostic factor of breast cancer [27].

Kinesin Family Member 4A (KIF4A), a member of kinesin superfamily, is a microtubule-dependent motor protein which participates in the condensation of chromosome and then plays a role in cell division [28]. Abnormal expression of KIF4A was reported in several cancers, such as gastric [29], lung [30], cervical [31] and breast cancers [32]. In a study of Matsumoto et al., KIF4A was highly expressed in colorectal cancer and loss of KIF4A inhibited cell proliferation [33]. Li et al. screened out 16 up-regulated Kinesin genes including KIF4A, which was experimentally validated in breast cancer tissues and KIF4A was associated with the dismal survival of breast cancer patients [34]. Besides, Zou et al. demonstrated that KIF4A could be induced by estrogen and the high level of KIF4A was correlated with the shorter relapse-free survival of breast cancer patients [32]. Moreover, knockdown of KIF4A dramatically hindered cell proliferation and triggered cell apoptosis [32]. All the aforementioned studies suggest oncogenic roles of EXO1 and KIF4A, which strongly supported our findings. Of note, our studies firstly explored the association of EXO1 and KIF4A expression with other clinical characteristics of breast cancer patients including stage, grade, age and TNM categories. Combining our results with these previous investigations, we infer that EXO1 and KIF4A promote the progression of breast cancer possibly via triggering cancer cell proliferation, which awaits experimentally confirmed in vitro and in vivo.

## 5. Conclusion

In summary, we utilized integrated bioinformatics analysis to identify novel predictors for progression and prognosis of breast cancer. Two hub genes EXO1 and KIF4A were up-regulated and associated with the poor prognosis of breast cancer patients. Besides, there was a correlation between EXO1 and KIF4A expression and clinical traits including age, stage and T category of breast cancer patients. Moreover, EXO1 and KIF4A could be individual prognostic factors for breast cancer. Therefore, our findings interpreted biological function of EXO1 and KIF4A in breast cancer and provide new idea for diagnosis and therapy strategy of breast cancer patients.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.anndiagpath.2020.151675.

## Declaration of competing interest

There is no conflict of interest to disclose.

## Acknowledgements

Not applicable.

## References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68(6):394–424. https://doi.org/10.3322/caac.21492. PMID.

[2] Testa U, Castelli G, Pelosi E. Breast cancer: a molecularly heterogenous disease needing subtype-specific treatments. Med Sci (Basel) 2020;8(1). https://doi.org/10.3390/medsci8010018. PMID: Pmc7151639.

[3] Saad ED, Katz A, Buyse M. Overall survival and post-progression survival in advanced breast cancer: a review of recent randomized clinical trials. J Clin Oncol 2010;28(11):1958–62. https://doi.org/10.1200/jco.2009.25.5414. PMID.

[4] Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn) 2015;19(1a):A68–77. https://doi.org/10.5114/wo.2014.47136. PMID: Pmc4322527.

[5] Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30(1):207–10. https://doi.org/10.1093/nar/30.1.207. PMID: Pmc99122.

[6] The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45(6):580–5. https://doi.org/10.1038/ng.2653. PMID: Pmc4010069.

[7] Niu J, Huang Y, Liu X, Zhang Z, Tang J, Wang B, et al. Single-cell rna-seq reveals different subsets of non-specific cytotoxic cells in teleost. Genomics 2020. https://doi.org/10.1016/j.ygeno.2020.09.031. PMID.

[8] Zhu N, Hou J, Ma G, Guo S, Zhao C, Chen B. Co-expression network analysis identifies a gene signature as a predictive biomarker for energy metabolism in osteosarcoma. Cancer Cell Int 2020;20(259). https://doi.org/10.1186/s12935-020-01352-2. PMID: Pmc7310058.

[9] Langfelder P, Horvath S. Wgcna: an r package for weighted correlation network analysis. BMC Bioinformatics 2008;9(559). https://doi.org/10.1186/1471-2105-9-559. PMID: Pmc2631488.

[10] Li Q, Chen W, Song M, Chen W, Yang Z, Yang A. Weighted gene co-expression network analysis and prognostic analysis identifies hub genes and the molecular mechanism related to head and neck squamous cell carcinoma. Cancer Biol Ther 2019;20(6):750–9. https://doi.org/10.1080/15384047.2018.1564560. PMID: Pmc6605996.

[11] Kunkel TA, Erie DA. Eukaryotic mismatch repair in relation to DNA replication. Annu Rev Genet 2015;49:291–313. https://doi.org/10.1146/annurev-genet-112414-054722. PMID: Pmc5439269.

[12] Llinas P, Isabet T, Song L, Ropars V, Zong B, Benisty H, et al. How actin initiates the motor activity of myosin. Dev Cell 2015;33(4):401–12. https://doi.org/10.1016/j.devcel.2015.03.025. PMID: Pmc4476657.

[13] Pióro M, Jakimowicz D. Chromosome segregation proteins as coordinators of cell cycle in response to environmental conditions. Front Microbiol 2020;11(588). https://doi.org/10.3389/fmicb.2020.00588. PMID: Pmc7174722.

[14] Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. Cytohubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014;8(Suppl. 4):S11. https://doi.org/10.1186/1752-0509-8-s4-s11. PMID: Pmc4290687.

[15] Xiong XY, Hu XJ, Li Y, Liu CM. Inhibitory effects of enterolactone on growth and metastasis in human breast cancer. Nutr Cancer 2015;67(8):1324–32. https://doi.org/10.1080/01635581.2015.1082113. PMID.

[16] Li J, Zhu SC, Li SG, Zhao Y, Xu JR, Song CY. Tktl1 promotes cell proliferation and metastasis in esophageal squamous cell carcinoma. Biomed Pharmacother 2015;74:71–6. https://doi.org/10.1016/j.biopha.2015.07.004. PMID.

[17] Carbonel AA, Calió ML, Santos MA, Bertoncini CR, Sasso Gda S, Simões RS, et al. Soybean isoflavones attenuate the expression of genes related to endometrial cancer risk. Climacteric 2015;18(3):389–98. https://doi.org/10.3109/13697137.2014.964671. PMID.

[18] Huang Y, Li W, Yan W, Wu J, Chen L, Yao X, et al. Loss of pich promotes chromosome instability and cell death in triple-negative breast cancer. Cell Death Dis 2019;10(6):428. https://doi.org/10.1038/s41419-019-1662-6. PMID: Pmc6547724.

[19] Verma P, Nagireddy PKR, Prassanawar SS, Nirmala JG, Gupta A, Kantevari S, et al. 9-PAN promotes tubulin- and ROS-mediated cell death in human triple-negative breast cancer cells. 2020. https://doi.org/10.1111/jphp.13349. PMID.

[20] Huang R, Rofstad EK. Cancer stem cells (CSCs), cervical CSCs and targeted therapies. Oncotarget 2017;8(21):35351–67. https://doi.org/10.18632/oncotarget.10169. PMID: Pmc5471060.

[21] Kostrhunova H, Zajac J, Markova L. A multi-action PT(IV) conjugate with oleate and cinnamate ligands targets human epithelial growth factor receptor HER2 in aggressive breast cancer cells. 2020. https://doi.org/10.1002/anie.202009491. PMID.

[22] Tishkoff DX, Amin NS, Viars CS, Arden KC, Kolodner RD. Identification of a human gene encoding a homologue of saccharomyces cerevisiae exo1, an exonuclease implicated in mismatch repair and recombination. Cancer Res 1998;58(22):5027–31 (PMID, DOI:).

[23] Yang G, Dong K, Zhang Z, Zhang E, Liang B, Chen X, et al. Exo1 plays a carcinogenic role in hepatocellular carcinoma and is related to the regulation of foxp3. J Cancer 2020;11(16):4917–32. https://doi.org/10.7150/jca.40673. PMID: Pmc7330697.

[24] de Sousa JF, Torrieri R, Serafim RB, Di Cristofaro LF, Escanfella FD, Ribeiro R, et al. Expression signatures of DNA repair genes correlate with survival prognosis of astrocytoma patients. Tumour Biol 2017;39(4). https://doi.org/10.1177/1010428317694552. 1010428317694552. PMID.

[25] Dai Y, Tang Z, Yang Z. Exo1 overexpression is associated with poor prognosis of hepatocellular carcinoma patients 2018;17(19–20):2386–97. https://doi.org/10.1080/15384101.2018.1534511. PMID.

[26] Kretschmer C, Sterner-Kock A, Siedentopf F, Schoenegg W, Schlag PM, Kemmner W. Identification of early molecular markers for breast cancer. Mol Cancer 2011;10(1):15. https://doi.org/10.1186/1476-4598-10-15. PMID: Pmc3045364.

[27] Qi L, Zhou B, Chen J, Hu W, Bai R, Ye C, et al. Significant prognostic values of differentially expressed-aberrantly methylated hub genes in breast cancer. J Cancer 2019;10(26):6618–34. https://doi.org/10.7150/jca.33433. PMID: Pmc6856906.
[28] Nakagawa T, Tanaka Y, Matsuoka E, Kondo S, Okada Y, Noda Y, et al. Identification and classification of 16 new kinesin superfamily (KIF) proteins in mouse genome. Proc Natl Acad Sci U S A 1997;94(18):9654–9. https://doi.org/10.1073/pnas.94.18.9654. PMID: Pmc23244.
[29] Gao J, Sai N, Wang C, Sheng X, Shao Q, Zhou C, et al. Overexpression of chromokinesin KIF4 inhibits proliferation of human gastric carcinoma cells both in vitro and in vivo. Tumour Biol 2011;32(1):53–61. https://doi.org/10.1007/s13277-010-0090-0. PMID.
[30] Taniwaki M, Takano A, Ishikawa N, Yasui W, Inai K, Nishimura H, et al. Activation of KIF4A as a prognostic biomarker and therapeutic target for lung cancer. Clin Cancer Res 2007;13(22 Pt 1):6624–31. https://doi.org/10.1158/1078-0432.ccr-07-1328. PMID.
[31] Narayan G, Bourdon V, Chaganti S, Arias-Pulido H, Nandula SV, Rao PH, et al. Gene dosage alterations revealed by cdna microarray analysis in cervical cancer: identification of candidate amplified and overexpressed genes. Genes Chromosomes Cancer 2007;46(4):373–84. https://doi.org/10.1002/gcc.20418. PMID.
[32] Zou JX, Duan Z, Wang J, Sokolov A, Xu J, Chen CZ, et al. Kinesin family deregulation coordinated by bromodomain protein ANCCA and histone methyltransferase MLL for breast cancer cell growth, survival, and tamoxifen resistance. Mol Cancer Res 2014;12(4):539–49. https://doi.org/10.1158/1541-7786.mcr-13-0459. PMID: Pmc4139106.
[33] Matsumoto Y, Saito M, Saito K, Kanke Y, Watanabe Y, Onozawa H, et al. Enhanced expression of KIF4A in colorectal cancer is associated with lymph node metastasis. Oncol Lett 2018;15(2):2188–94. https://doi.org/10.3892/ol.2017.7555. PMID: Pmc5776904.
[34] Li TF, Zeng HJ, Shan Z, Ye RY, Cheang TY, Zhang YJ, et al. Overexpression of kinesin superfamily members as prognostic biomarkers of breast cancer. Cancer Cell Int 2020;20(123). https://doi.org/10.1186/s12935-020-01191-1. PMID: Pmc7161125.