



OPEN

Integrated machine learning algorithms identify *KIF15* as a potential prognostic biomarker and correlated with stemness in triple-negative breast cancer

Qiaonan Guo^{1,3}, Pengjun Qiu^{1,3}, Kelun Pan^{1,3}, Huikai Liang¹, Zundong Liu²✉ & Jianqing Lin¹✉

Cancer stem cells (CSCs) have the potential to self-renew and induce cancer, which may contribute to a poor prognosis by enabling metastasis, recurrence, and therapy resistance. Hence, this study was performed to identify the association between CSC-related genes and triple-negative breast cancer (TNBC) development. Stemness gene sets were downloaded from StemChecker. Based on the online databases, a consensus clustering algorithm was conducted for unsupervised classification of TNBC samples. The variations between subtypes were assessed with regard to prognosis, tumor immune microenvironment (TIME), and chemotherapeutic sensitivity. The stemness-related gene signature was established and random survival forest analysis was employed to identify the core gene for validation experiments and tumor sphere formation assays. 499 patients with TNBC were classified into three subgroups and the Cluster 1 had a better OS than others. After that, WGCNA study was performed to identify genes important for Cluster 1 subtype. Out of all 8 modules, the subtype of Cluster 1 and the yellow module with 103 genes demonstrated the largest positive association. After that, a four-gene stemness-related signature was established. Based on the yellow module, the 39 potential pivotal genes were subjected to the random forest survival analysis to find out the gene that was relatively important for OS. *KIF15* was confirmed as the targeted gene by LASSO and random survival forest analyses. In vitro experiments, the downregulation of *KIF15* promoted the stemness of TNBC cells. The expression levels of stem cell markers Nanog, SOX2, and OCT4 were found to be elevated in TNBC cell lines after *KIF15* inhibition. A stemness-associated risk model was constructed to forecast the clinical outcomes of TNBC patients. The downregulation of *KIF15* expression in a subpopulation of TNBC stem cells may promote stemness and possibly TNBC progression.

Keywords Triple-negative breast cancer (TNBC), Cancer stem cell (CSC), *KIF15*, Tumor microenvironment (TME), Prognosis, Risk model

Breast cancer (BC) is one of the most commonly diagnosed malignancies, responsible for approximately 31% of new cancer cases in women and is listed as the second cause of cancer-related deaths in women in the 2023 Cancer Statistics¹. TNBC is a special subtype of BC that is histologically distinguished by the absence of human epidermal growth factor receptor 2 (HER-2), estrogen receptor (ER), and progesterone receptor (PR). Current available treatments for TNBC include a mastectomy or breast-conserving procedure, chemotherapy, radiotherapy, immunotherapy, and other newly developed therapies². Despite a variety of treatment strategies, the incidence of metastasis, recurrence, and drug resistance remains high in TNBC patients. Accumulating evidence indicates that numerous malignancies involving BC have populations of cells that resemble stem cells³. CSCs are a dynamic subgroup of tumor cells that have the ability to self-renew, be pluripotent, and proliferate indefinitely⁴. Given their relative resistance to cytotoxic chemotherapy and radiotherapy, these BC stem cells (BCSCs) may

¹Department of Breast and Thyroid Surgery, Second Affiliated Hospital of Fujian Medical University, Quanzhou, China. ²Stem Cell Laboratory, Second Affiliated Hospital of Fujian Medical University, Quanzhou, China. ³These authors contributed equally: Qiaonan Guo, Pengjun Qiu and Kelun Pan. ✉email: harryblood1989@126.com; lj13905977336@163.com

play a vital role in drug resistance and recurrence^{5,6}. Hence, it is crucial to discover the vital BCSC stemness regulators for theoretical research as well as practical application.

In 2003, Al-Hajj initially described a subtype of cell fraction isolated from BC tissues. This subpopulation of BC cells was identified with the CD44⁺/CD24^{-/low} phenotype and was able to reproduce the tumor burden in mice⁷. After that, in 2007, a subgroup of BC cells was found to have the ability to initiate malignancy in vitro and in vivo by Ginestier et al., who determined that they had high aldehyde dehydrogenase (ALDH) activity⁸. Hence, human BCSCs are described on the basis of the expression of the cell surface marker CD44, the absence of CD24, and high ALDH activity⁹. Additionally, evidence is mounting suggesting BCSCs with these traits are involved in tumor initiation, development, and metastasis¹⁰. BCSCs are the main contributors to therapeutic resistance and recurrence, despite the fact that radiation therapy and chemotherapy can successfully eradicate the majority of tumor cells^{11–14}. Since CSCs possess the ability to escape from strict regulation while relying on aberrant activation of some signaling pathways, they are thought to be an early event in oncogenesis and allow these cells to resist regular chemotherapies, leading to the recurrence of tumors⁹. Recently, a number of clinical trials based on the theory of relative tolerance of BCSCs to cytotoxic chemotherapy have been designed to find ways to address TNBC resistance^{3,9}. Alongside the intrinsic signals that regulate BCSCs, these cells are also controlled by the components of the TME^{15–19}. Numerous biological elements, such as endothelial cells, fibroblasts, inflammatory cells, and mesenchymal stem cells, are present in the microenvironment surrounding stem cells and may interact with CSCs via cytokine loops¹⁵. Some inflammatory factors, such as IL-8 and IL-6, have been found to promote the proliferation of CSCs after chemotherapeutic treatment^{20,21}. Besides, some studies have demonstrated that serum levels of IL-8 and IL-6 in individuals suffering from advanced BC are linked to metastatic development and a poor prognosis^{22,23}. Hence, a further comprehensive description of the BCSCs and the TME landscape surrounding them would make contributions to addressing the problems of therapy resistance and tumor recurrence, as well as developing CSC-targeted treatment strategies and novel immunotherapies.

This study aims to investigate the relationship between CSC-related genes and the prognosis of TNBC patients, as well as to elucidate the underlying molecular mechanisms. By classifying TNBC from the perspective of cancer stem cells, this research seeks to identify prognostic biomarkers and novel targeted therapeutic strategies for TNBC. In the current study, on the basis of the public stemness gene sets, the unsupervised clustering method, and weighted gene correlation network analysis (WGCNA) were applied to identify the stemness subgroups with discrete stemness and determine the genes associated with the stemness subgroups and prognostic outcomes. After that, the stemness-related signature was established by least absolute shrinkage and selection operator (LASSO) Cox regression and random survival forest analysis. Subsequently, the prognostic outcomes, TME landscapes, and chemotherapeutic sensitivities were analyzed in TNBC based on the stemness-related risk model. According to the result of random survival forest analysis, kinesin family member 15 (*KIF15*) was identified as the vital gene associated with prognosis and TNBC cell sphere, and a comprehensive series of validation experiments were further conducted in vitro.

Methods

Acquisition and pre-processing of TNBC datasets

The RNA-sequencing (RNA-seq) expression data and corresponding information of TNBC samples were extracted from The Cancer Genome Atlas (TCGA, <http://cancergenome.nih.gov/>), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC, <http://www.cbioportal.org/>), and GSE103091^{24,25} in the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) databases, respectively. The TNBC samples with insufficient clinical profiles and an overall survival (OS) of less than 60 days were excluded. Ultimately, a total of 499 samples were included in the study. Then, the log₂ transformation was performed on the TCGA and GEO datasets. The Combat function from the “sva” R package was adopted to perform batch correction of the high-throughput sequencing results²⁶. Each database was considered a separate batch, and the “mod” parameter was set to “global”. After that, 26 stemness gene sets were collected from StemChecker (<http://stemchecker.sysbiolab.eu/>), a web-based tool with comprehensive genetic information on published stemness^{27,28}, which are provided in Supplemental Table 1.

Stemness features and consensus clustering for TNBC stemness subgroups

The “GSVA²⁹ and GSEABase” R packages were employed to calculate the stemness enrichment fraction of the stemness genes on a sample-by-sample basis for each TNBC and present them as ssGSEA scores. Subsequently, Pearson’s correlation analysis was performed on the 26 stemness gene sets by using the “corrplot” R package. After that, the “ConsensusClusterplus” R package was used to conduct the consensus clustering algorithm (K-means clustering) for unsupervised classification on TNBC samples with 10,000 repetitions³⁰. In the consensus clustering analysis, the maximum number of clusters was set to 6 (maxK = 6), the sampling proportion was 0.8 (pltem = 0.8), and the distance matrix used was Euclidean.

Analysis of immune cell type components and the tumor immune microenvironment (TIME)

CIBERSORT (<http://cibersort.stanford.edu/>) was employed to quantify the components of 22 immune cells through the RNA-seq expression profiles of normalized bulk samples³¹. The immune cell infiltration of each TNBC tissue was analyzed by the “CIBERSORT” R package with the LM22 annotated immune cell gene signature. In the analysis that followed, the assumption regarding the type of immune cell was deemed true and statistically significant at $P < 0.05$. Subsequently, the proportion of immune-stromal components of TME was assessed through the Estimation of Stromal and Immune Cells in Malignant Tumor Tissues Using Expression (ESTIMATE) algorithm with the “estimate” R package³². The results were shown with a stromal score, an immune score, and an ESTIMATE score.

Drug sensitivity analysis

On the basis of the website tool, Genomics of Drug Sensitivity in Cancer (GDSC) (<https://www.cancerrxgene.org/>), the IC50 of six classical chemotherapeutic drugs for TNBC patients, including Palbociclib, Paclitaxel, Epirubicin, Docetaxel, Cyclophosphamide, and Cisplatin, were computed in each TNBC sample by the “oncoPredict” R package³³. The ridge regression was used to determine the significance of variations in the IC50 Z-score among various subgroups with tenfold cross-validation to assess the accuracy of drug sensitivity prediction.

Weighted gene co-expression network analysis (WGCNA)

The “WGCNA” R package was performed on overall gene expression to establish the co-expression network. In order to evaluate the correlation of gene expression patterns among patients with TNBC, paired Pearson correlation coefficient matrices were further developed³⁴. After that, the similarity matrix was processed into the adjacency matrix by using the formula as follows: $S_{ij} = |\text{cor}(x_i, x_j)| a_{ij} = S_{ij}^\beta$. The soft-threshold power β was determined through a “WGCNA” package feature called soft connection based on the scale-free network. The gene co-expression network’s topology was regarded as scale-free, and batch effects were absent if the Topology Fit Index (TFI) had a low power (< 30) value of 0.85 or above. As a result, the power of β was selected as 14. Subsequently, the network modules were analyzed through the Topological Overlap Measure (TOM) and correlative dissimilarity matrix (1-TOM). The minimum module size was set to 50, and the default values for the other parameters were fixed. Following the calculation of the module eigengene (ME), which indicated the first principal component of each module, connections between modules and each stemness subgroup were established. Hub genes for a given module were defined by their gene significance (GS, Pearson’s correlation between each gene and clinical trait) and module membership (MM, correlation between each gene and module) parameters, both of which were set at > 0.5 and > 0.8 ^{35,36}, respectively. As a result, the genes with significant MM and high GS were selected as hub genes for subsequent analysis. The “clusterProfiler” R package was employed to conduct the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses^{37–39} on the identified hub genes of the co-expression module⁴⁰.

Establishment of a stemness-related prognostic signature

The prognostic stemness-related genes were screened out from the hub genes in the selected model by univariate Cox regression analysis. Afterward, the LASSO Cox regression analysis was adopted to avoid model overfitting through redundant gene reduction. The LASSO coefficient profiles of the prognostic genes were generated and tenfold cross-validation results were produced to determine optimal values of the penalty parameter λ ($\lambda = 0.08529673$). Accordingly, independent prognostic stemness-associated genes were ultimately screened, and their expression was included in the risk model formula. $\text{RiskScore} = \sum_{i=1}^n (\text{Exp}_i * \text{Coe}_i)$ ($n = 6$, the expression of each gene is shown as Exp_i , and the corresponding Cox regression coefficient of each gene is shown as Coe_i). Based on the median value of the risk score, the entire TNBC population was separated into high-risk and low-risk groups, and a survival analysis was conducted. Subsequently, the stemness-related risk scores of the 3 clusters were calculated, respectively, and $P < 0.05$ was considered statistically significant (Kruskal–Wallis test). Notably, the prognostic value of the stemness-related risk score was assessed within the clusters obtained by unsupervised clustering. In order to identify the critical genes for further study, the random forest survival analysis was performed on the 39 prognostic stemness-related genes selected by univariate Cox regression analysis using the R package “randomForestSRC”.

Enrichment analysis of hallmark gene sets

Based on the median value of the stemness-related score, 499 TNBC samples were divided into high-risk and low-risk cohorts. The hallmark gene sets were obtained from the online database Molecular Signatures Database (MSigDB, <http://software.broadinstitute.org/gsea/msigdb/>). The R package “clusterProfiler” was adopted to conduct the gene set enrichment analysis (GSEA)⁴¹. As a result, the relevant subsets of hallmark gene sets and functional enrichment pathways were identified.

Cell culture

The human breast epithelial cell lines MCF-10A and TNBC cell lines SUM159, CAL-148, MDA-MB-231, and MDA-MB-453 were obtained from the American Type Culture Collection (ATCC). According to the vendor’s guidelines, each cell line was kept in good condition. All the cell lines were placed in the incubator with 5% CO₂ and 37 °C. MCF-10A cell lines were cultivated in specialized medium purchased from Procell (Wuhan, China; CM-0525). Basically, SUM159, MDA-MB-231, and MDA-MB-453 were maintained in Dulbecco’s modified Eagle’s medium (DMEM; Gibco BRL, USA). CAL-148 was cultured in RPMI 1640 medium (Gibco BRL, USA). And all medium supplemented with 10% fetal bovine serum (FBS; Gibco, Grand Island, NY, USA) and 1% penicillin–streptomycin (Beyotime, Shanghai, China).

SiRNA transfection

Following the manufacturer’s instructions, cells were seeded in 6-well plates and transfected with either target KIF15 (siKIF15) or negative control (NC) using Lipofectamine 3000 (Invitrogen, USA). The NC sequence is a widely cited sequence in the literature that lacks homology with all mammals, making it suitable for the study of different genes in rats, mice, humans⁴². After 48 h transfection, cells were extracted and further examined. The sequence details for the siRNAs, which were specially produced from GenePharma (Suzhou, China), were as follows:

NC:

Sense 5'-UUCUCCGAACGUGUCACGUTT-3'

Antisense 5'-ACGUGACACGUUCGGAGAATT-3'

siKIF15-#1:

Sense 5'-CCGAGAGGAUCAAAUAAUATT-3'

Antisense 5'-UAUUAUUUGAUCCUCUCGGTT-3'

siKIF15-#2:

Sense 5'-GGAUGAAGAAGAGCAUAAATT-3'

Antisense 5'-UUUAUGCUCUUCUUAUCCTT-3'

Quantitative reverse transcription-PCR (qRT-PCR)

Total RNA was isolated using NucleoZOL Reagent (Macherey–Nagel, Düren Germany) according to the manufacturer's instructions. The SYBR Green method (Vazyme, Shanghai, China) was employed to examine the expression of targeted genes in triplicate. The data was calculated by using the cycle threshold (CT) ($2^{-\Delta\Delta CT}$) approach, and the data analysis was performed by the QuantStudio™ 5 Real-Time PCR System (Thermo Fisher, MA, USA). The assay was done three times for each sample. Using the comparative CT approach, the expression levels were standardized to those of b-actin. The relevant primers were provided in Supplemental Table 2.

Western blot (WB)

A RIPA lysis buffer supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) and protease/phosphatase inhibitors (Roche, Basel, Switzerland) were used to extract the total protein from cell lysates. The total protein content was determined using a BCA assay kit (Beyotime, Shanghai, China), and 20 mg of denatured proteins were submitted to 10% SDS-PAGE before being electroblotted onto PVDF membranes (Millipore, Burlington, MA, United States). Following this, the membranes were blocked via 5% milk in TBST for 1 h at room temperature before being incubated overnight at 4 °C with primary antibodies: KIF15(55,407–1-AP), Nanog(14,295–1-AP), SOX2(11,064–1-AP), OCT4(11,263–1-AP) all from proteintech, vinculin(sc-73614, santa cruz), and β -actin(AB0035, abways). Eventually, blots were treated with the relevant secondary antibody labeled HRP (1:1000, Beyotime, Shanghai, China) for 1 h at room temperature before exposing the blots to the Bio-Rad imaging system.

Tumor sphere formation assay

The sphere culture was performed as document described⁴³, briefly, The TNBC cells (1×10^4 cells/well) were seeded into the ultralow-attachment six-well plate with DMEM-F12 medium (Gibco, Milan, Italy) containing 2% B27 supplement (no vitamin A; Invitrogen, Paisley, UK; 12,587) and Recombinant Human (20 ng/ml; Sigma Aldrich, Poole, UK; E-9644). After 7 days cultured, the tumor cells sphere was examined and captured.

Statistical analysis

All statistical analyses were conducted in R software (version 4.2.0) (<https://www.r-project.org/>). The survival analysis was conducted using the Kaplan–Meier curve. The Wilcoxon test and Kruskal–Wallis test were employed for pairwise comparisons between two groups and several groups, respectively. Pearson's correlation analysis was employed in the comparison between ssGSEA scores of stemness-related gene sets. The comparison of PCR results was performed using a t-test. And $p < 0.05$ was considered statistically significant. All methods were carried out in accordance with the relevant laws and regulations.

Results

The enrichment analysis of stem-related gene sets and construction of stemness subsets

The relationship among the ssGSEA enrichment scores of the stem-related gene sets was assessed and shown in Fig. 1a. The 499 TNBC samples were then divided into 3 different clusters using unsupervised clustering ($k = 3$, Cluster 1–3, Fig. 1b). In Fig. 1c, the enrichment degrees of the stemness gene sets in the three clusters were presented by heatmap. Subsequently, according to the survival analysis, the TNBC patients in Cluster 1 were indicated with better OS than those in Cluster 2 and Cluster 3 ($P < 0.05$, Fig. 1d). In order to identify the characteristics of TME among the stemness-related subsets, the components of immune cells were analyzed by CIBERSORT. The TME fractions of the three clusters were shown in Fig. 1e. The TNBC samples in Cluster 1 manifested a subtype of anti-tumor immune activation with abundant macrophages M0 and macrophages M1. Cluster 2 exhibited a moderate infiltration of tumor immune cells, for instance B cells, T cells, (natural killer) NK cells and monocytes, but the infraction of macrophages M2 was higher that was related to worse prognosis. As for Cluster 3, some immune cells associated with tumor promotion and immunosuppression were shown higher infiltrations like regulatory T cells (Tregs) and CD8 T cells. After that, we further investigate the drug sensitivity of six classical chemotherapeutic drugs for TNBC, including Palbociclib, Paclitaxel, Epirubicin, Docetaxel, Cyclophosphamide and Cisplatin (Fig. 1f). The IC50 for Paclitaxel, Docetaxel, Cyclophosphamide and Cisplatin was lower in Cluster 1, suggesting that the subtype of Cluster 1 may be potentially more sensitive to these chemotherapeutic agents. Besides, the subtypes of Cluster 2 and Cluster 3 were indicated more sensitive to Palbociclib and Epirubicin.

Identification of the cluster 1-related stemness module and hub genes in TNBC

Since the prognostic outcomes of patients with TNBC in Cluster 1 were better than others, the WGCNA study was performed to identify genes important for Cluster 1 subtype. The 499 samples were clustered to filter outliers

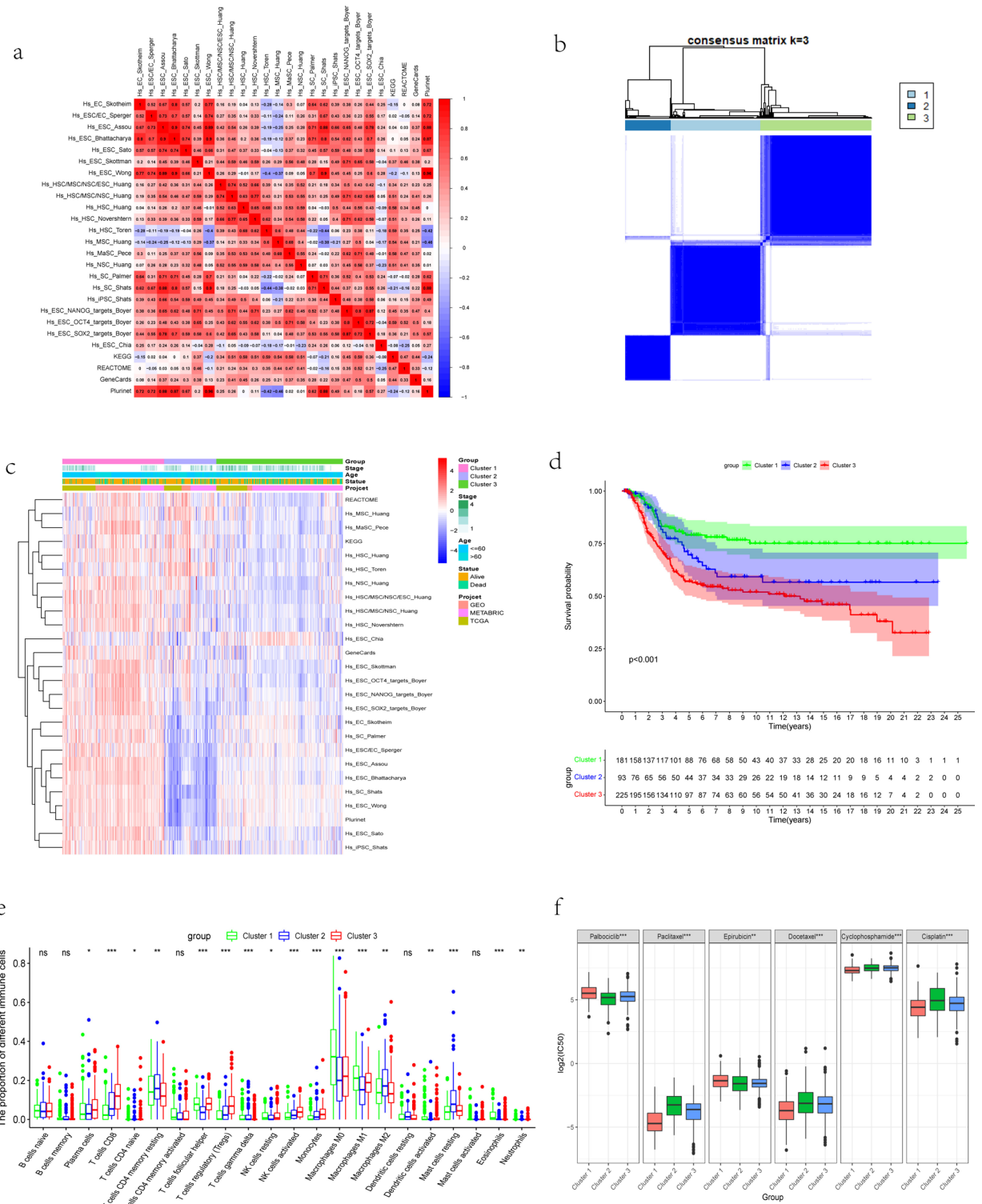


Fig. 1. The enrichment analysis of stem-related gene sets and identification of stemness subgroups. **(a)** The relationship among the ssGSEA enrichment scores of the stem-related gene sets. Red indicates positive correlations, and blue indicates negative correlations. The darker the color, the stronger the correlation. **(b)** A total of 499 TNBC patients were classified into three clusters according to the consensus clustering matrix ($k=3$). **(c)** Heatmap and the clinicopathologic characters of the three clusters classified through the enrichment of the stemness gene sets (stage, AJCC tumor stage I/II/III/IV). The scale of color in the heatmap represents the ssGSEA score of each sample. **(d)** Kaplan–Meier curves for the three clusters. **(e)** The infiltration fractions of different immune cells among the three clusters (* indicates $P < 0.05$, ** indicates $P < 0.01$, *** indicates $P < 0.001$, ns indicates $P > 0.05$). **(f)** The IC50 for Palbociclib, Paclitaxel, Epirubicin, Docetaxel, Cyclophosphamide, and Cisplatin in three clusters (** indicates $P < 0.01$ and *** indicates $P < 0.001$).

for subsequent analysis, and 2 outlier samples from TCGA database were removed out. Then, the trait heatmap was drawn on the basis of the 3 Clusters provided by unsupervised clustering. The heatmap showing the relationships between sample distribution and cluster grouping is presented in Fig. 2a. The gene expression profiles varied among different clusters. Each cluster was associated with a distinct prognosis, and differentially expressed genes exhibited diverse distributions across these clusters. Afterwards, for purpose of establishing a scale-free network, a power of $\beta = 14$ (scale-free $R^2 = 0.85$) was chosen as the soft threshold parameter (Fig. 2b). Following the setting of 50 as the minimum number of genes for every single module, the genes with comparable expression patterns were grouped into 8 modules, as illustrated by clustering dendrogram (Supplemental Fig. 1). Out of all 8 modules, the subtype of Cluster 1 and the yellow module demonstrated the largest positive association (Fig. 2c, $ME = 0.62$, $P < 0.05$). Total 103 genes in yellow module were provided in the Supplemental Table 3. As a result, the yellow module was selected as the crucial module, and 39 crossover potential pivotal genes were then screened from this module using the screening parameters of $MM > 0.8$ and $GS > 0.5$ for further analysis (Fig. 2d).

Subsequently, GO and KEGG enrichment analysis were conducted on the hub genes in the yellow module to investigate the associated biofunction activities. The top ten GO terms of molecular function (MF), cellular component (CC), and biological process (BP) were shown in Fig. 2e, which were mainly enriched in the activities of chromosomes. Besides, the results of KEGG indicated that the yellow module was principally involved in cell cycle, progesterone-mediated oocyte maturation, oocyte meiosis, cellular senescence, human T-cell leukemia virus 1 infection, microRNAs in cancer, viral carcinogenesis, FoxO signaling pathway, DNA replication, and p53 signaling pathway (Fig. 2f).

Establishment of stemness-related gene risk model

The prognostic stemness-associated gene signature was developed based on the hub genes in the yellow module. The univariate Cox regression analysis was performed on the 39 hub genes in the first instance to identify the prognosis-related ones (Supplemental Fig. 2). After that, the LASSO regression analysis was performed on the prognostic stemness-related genes. Figure 3a showed the LASSO coefficient profiles of the 39 genes. In order to find the optimal value for the penalty parameter λ , tenfold cross-validation results were created (Fig. 3b, $\lambda = 0.08529673$). Consequently, 4 candidate genes were identified to construct the stemness-related risk model: thyroid hormone receptor interacting protein 13 (*TRIP13*), Forkhead box M1 (*FOXM1*), kinesin family member 15 (*KIF15*), and centrosomal protein 55 (*CEP55*). Subsequently, the expression levels of the 4 genes and the relevant regression coefficients were used to establish the risk model: risk score = $(-0.076 \times \text{expression level of } TRIP13) + (-0.020 \times \text{expression level of } FOXM1) + (-0.016 \times \text{expression level of } KIF15) + (-0.017 \times \text{expression level of } CEP55)$. Based on the stemness-related gene signature, all TNBC sample were divided into high-risk and low-risk cohorts according to the median risk score. Then, the survival analysis was performed on the two cohorts and visualized by drawn Kaplan–Meier curves. As shown in Fig. 3c, the clinical outcomes of TNBC patients in the high-risk cohort were worse than those in the low-risk cohort ($P < 0.05$). Afterwards, GSEA analysis was conducted to identify the hallmark gene sets enriched differently between the high-risk and low-risk cohorts. The genes with high expression level in the stemness-related high-risk cohort were considerably enriched in a number of hallmark gene sets, including androgen response, DNA repair, E2F targets, estrogen response, G2M checkpoint, glycolysis, heme metabolism, mitotic spindle, mtorc1 signaling, MYC targets 1, MYC targets 2, Notch signaling, oxidative phosphorylation, protein secretion, reactive oxygen species pathway, TGF- β signaling, TNFA signaling via NF- κ B, unfolded protein response, UV response, and WNT/ β -catenin signaling (Fig. 3d).

According to the stemness-related risk model, the stemness-related risk score was calculated for Cluster 1, Cluster 2, and Cluster 3, respectively. As shown in Fig. 3e, Cluster 1 had a lower stemness risk score than the other two ($P < 0.05$).

The association of stemness-related risk model and TIME

Total 499 TNBC samples were involved in the CIBERSORT analysis to investigate the differences of immune characteristics between stemness high-risk and low-risk cohorts. As shown in Fig. 4a, the infiltrations of some immune cells were significantly higher in the high stemness-risk cohort, for instance CD8 + T cells, resting memory CD4 + T cells, Tregs, activated NK cells, monocytes, and resting mast cells. By contrast, a large number of other anti-tumor immune cells were observed in the low-risk cohort, including activated memory CD4 + T cells, follicular helper T cells, macrophages M0, and macrophages M1. After that, the ESTIMATE scores were calculated in the high and low stemness-risk cohorts and shown in Fig. 4b–d. The high-risk cohort was found to have higher values of Stromal score, Immune score, and ESTIMATE score than the low-risk cohort ($P < 0.05$).

The relevance of drug sensitivity and stemness-related risk score

The "oncoPredict" tool in R software was used to determine the relationship between the stemness-related risk score and the half maximal inhibitory concentration (IC50) of several anticancer medicines. As illustrated by Fig. 4e, TNBC anticancer agents commonly used in clinical practice, including paclitaxel, docetaxel, cyclophosphamide, and cisplatin, have been shown to have superior efficacy in high-risk populations.

Identification and validation of targeted gene

Based on the yellow module, the 39 potential pivotal genes were subjected to the random forest survival analysis to find out the gene that was relatively important for OS (Fig. 5a). As a result, *KIF15* was selected. Subsequently, the expression level of *KIF15* was detected in the different unsupervised clustering subtypes, respectively. As shown in Fig. 5b, the expression of *KIF15* was higher in the Cluster 1, which had a better prognosis, than Cluster 2 and Cluster 3 ($P < 0.05$). Similarly, the expression of *KIF15* was furtherly assessed in the high-risk and low-risk cohorts on the basis of the stemness-related signature. As shown in Fig. 5c, in contrast to the high-risk cohort,

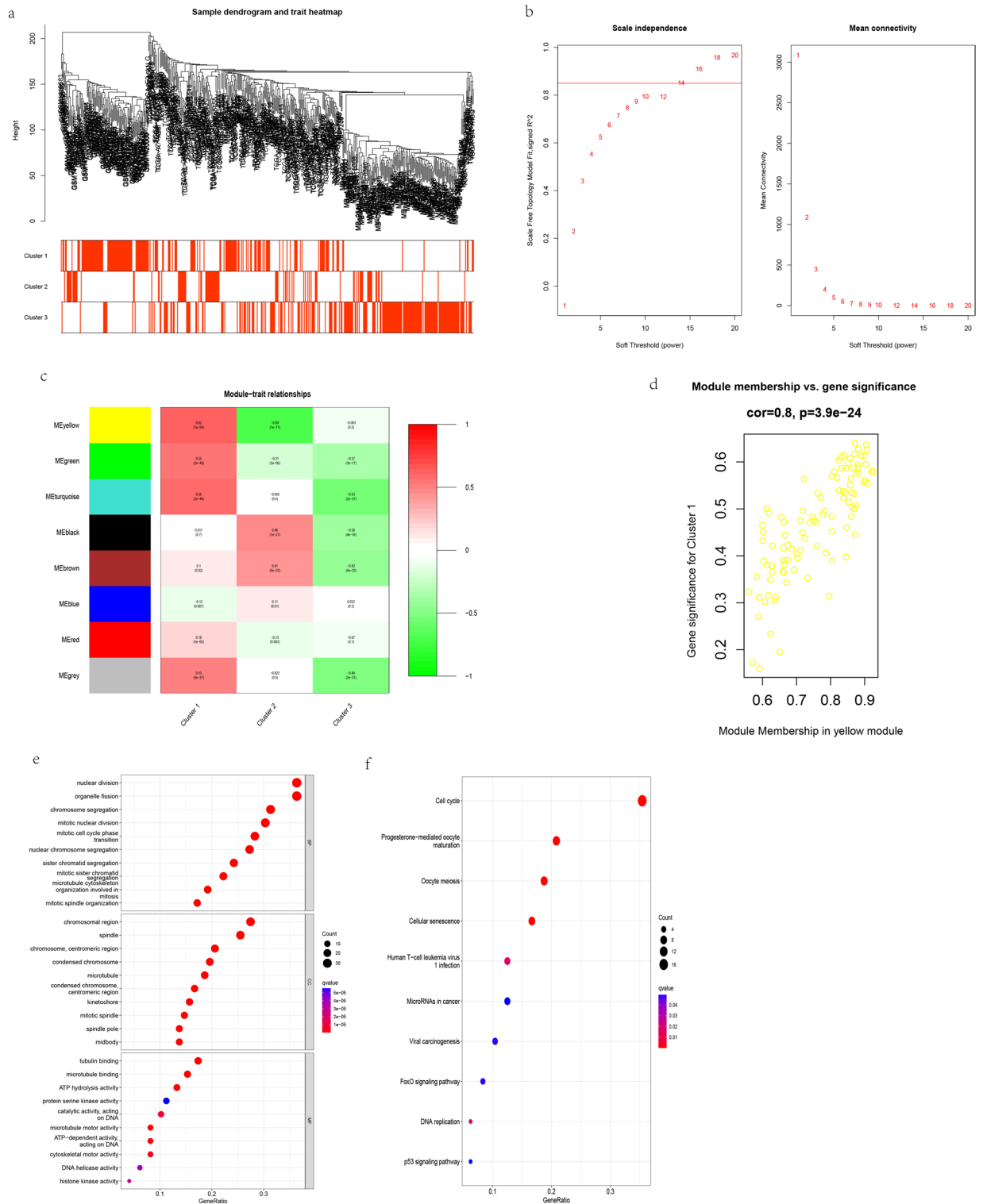


Fig. 2. Identification of the Cluster 1-related stemness module and hub genes in TNBC. **(a)** Clustering dendrogram of 499 samples. The trait heatmap based on the three clusters provided by unsupervised clustering shows the relationship between sample distribution and cluster grouping. **(b)** Scale-free fit index (left) and the mean connectivity (right) for soft-thresholding powers. When b was set at 14, the scale-free network was constructed. **(c)** Analysis of module-trait relationships in TNBC based on the 497 TNBC patients. A positive correlation between the yellow module (containing 103 genes) and the Cluster 1 subtype was indicated with a $P < 0.05$ (correlation coefficient = 0.62, $P = 1E-54$). **(d)** A scatter plot of GS for Cluster 1 subtype and the MM in the yellow module (correlation coefficient = 0.8, $P = 3.9E-24$). **(e)** GO analysis on the hub genes in the yellow module, showing the top ten GO terms of ME, CC, and BP. **(f)** The 10 enriched KEGG pathways in the hub genes.

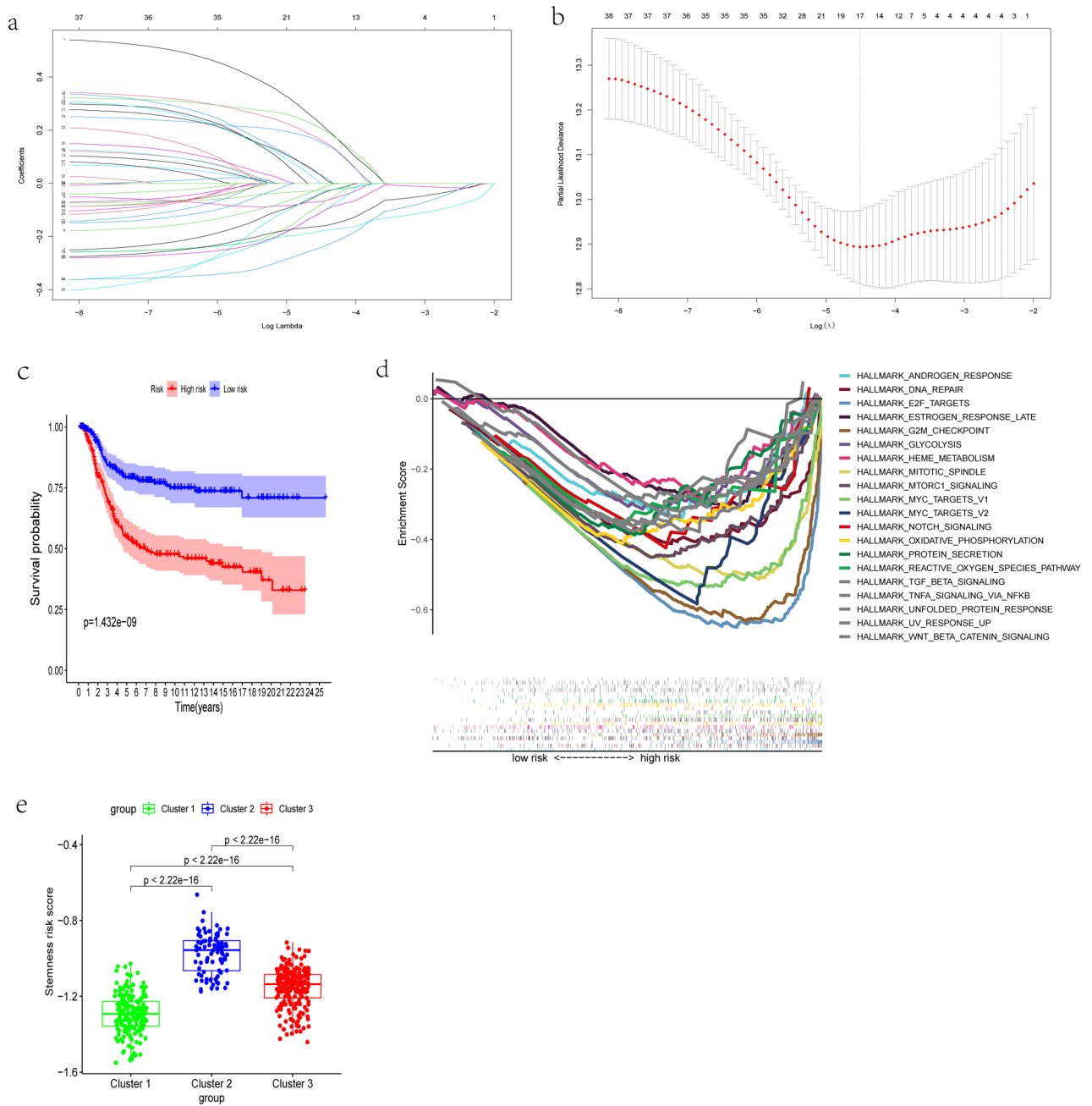


Fig. 3. Establishment of a stemness-related gene risk model. **(a)** LASSO coefficient profiles of 39 stemness-related hub genes. **(b)** The results of the tenfold cross-validation determined the optimal value of the penalty parameter λ ($\lambda = 0.08529673$). Four independent prognostic genes for the signature establishment were identified. **(c)** Kaplan–Meier survival curves of 499 TNBC patients suggest that the OS of the high-risk group is lower than the low-risk group ($P = 1.4321E - 09$). **(d)** GSEA analysis shows that the hallmark gene sets are enriched differently between the high-risk and low-risk groups. **(e)** The stemness-related risk score was calculated for Cluster 1, Cluster 2, and Cluster 3.

KIF15 expression was higher in the low-risk cohort ($P < 0.05$). Besides, the expression of *KIF15* was found negatively associated with the stemness-related risk score in TNBC (Fig. 5d, $R = -0.81$, $P < 0.05$). According to the median value of the expression level of *KIF15*, 499 TNBC patients were divided into *KIF15* high-expression group and low-expression group. The survival analysis was performed on these two groups, indicating that TNBC patients with lower expression of *KIF15* had better prognosis (Fig. 5e, $P < 0.05$). The Sankey diagram was drawn to display the association among the expression of *KIF15*, unsupervised clustering stemness subtypes, stemness-related risk types, and the survival outcomes (Fig. 5f).

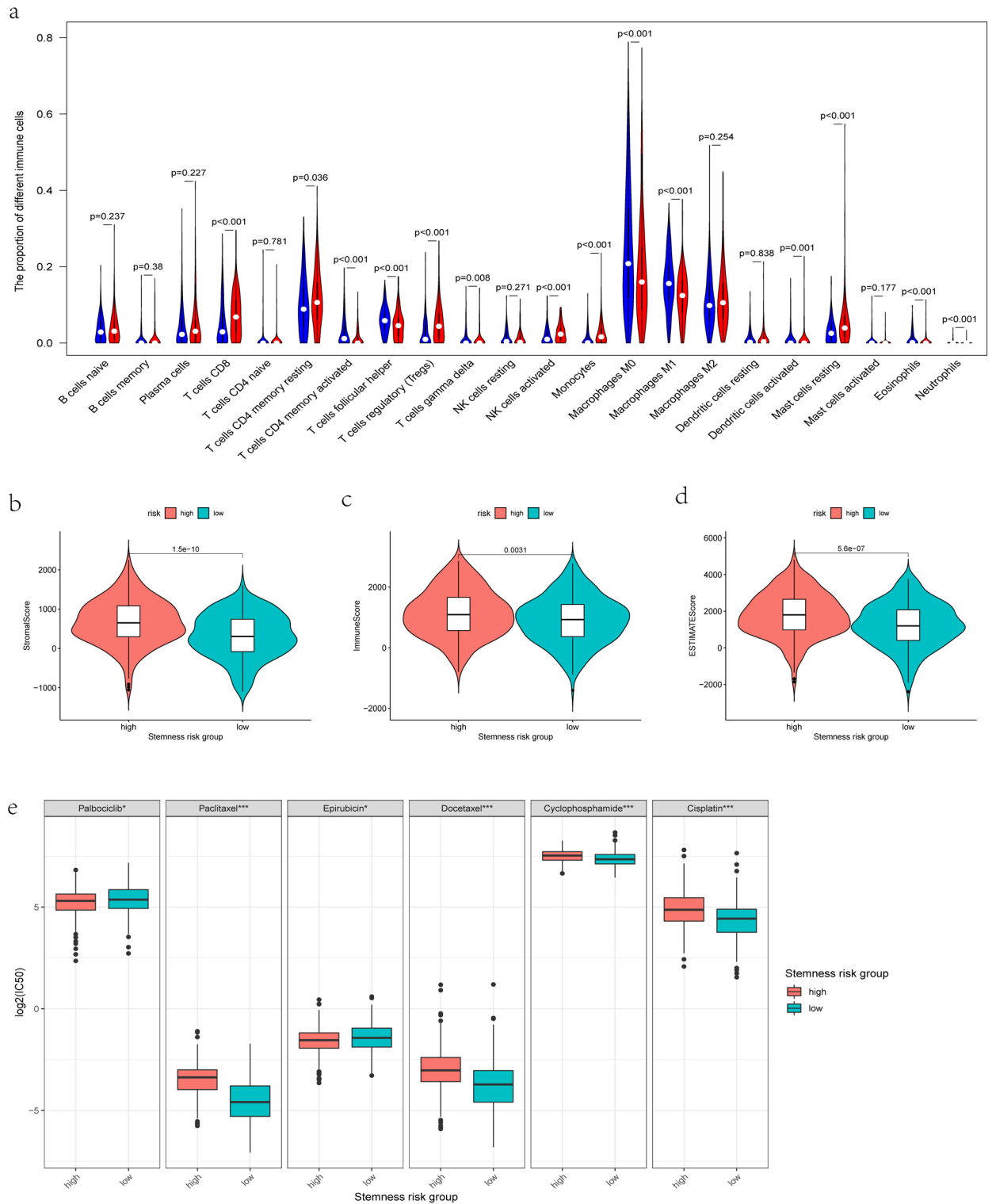


Fig. 4. The association between stemness-related risk model, immune cell infiltration, and chemotherapeutic sensitivity. **(a)** The characteristics of immune cell infiltration between low-risk and high-risk groups. **(b-d)** The Stromal scores, Immune scores, and ESTIMATE scores of high-risk and low-risk groups. **(e)** The IC₅₀ for Palbociclib, Paclitaxel, Epirubicin, Docetaxel, Cyclophosphamide, and Cisplatin in high-risk and low-risk groups (* indicates $P < 0.05$, ** indicates $P < 0.01$, and *** indicates $P < 0.001$).

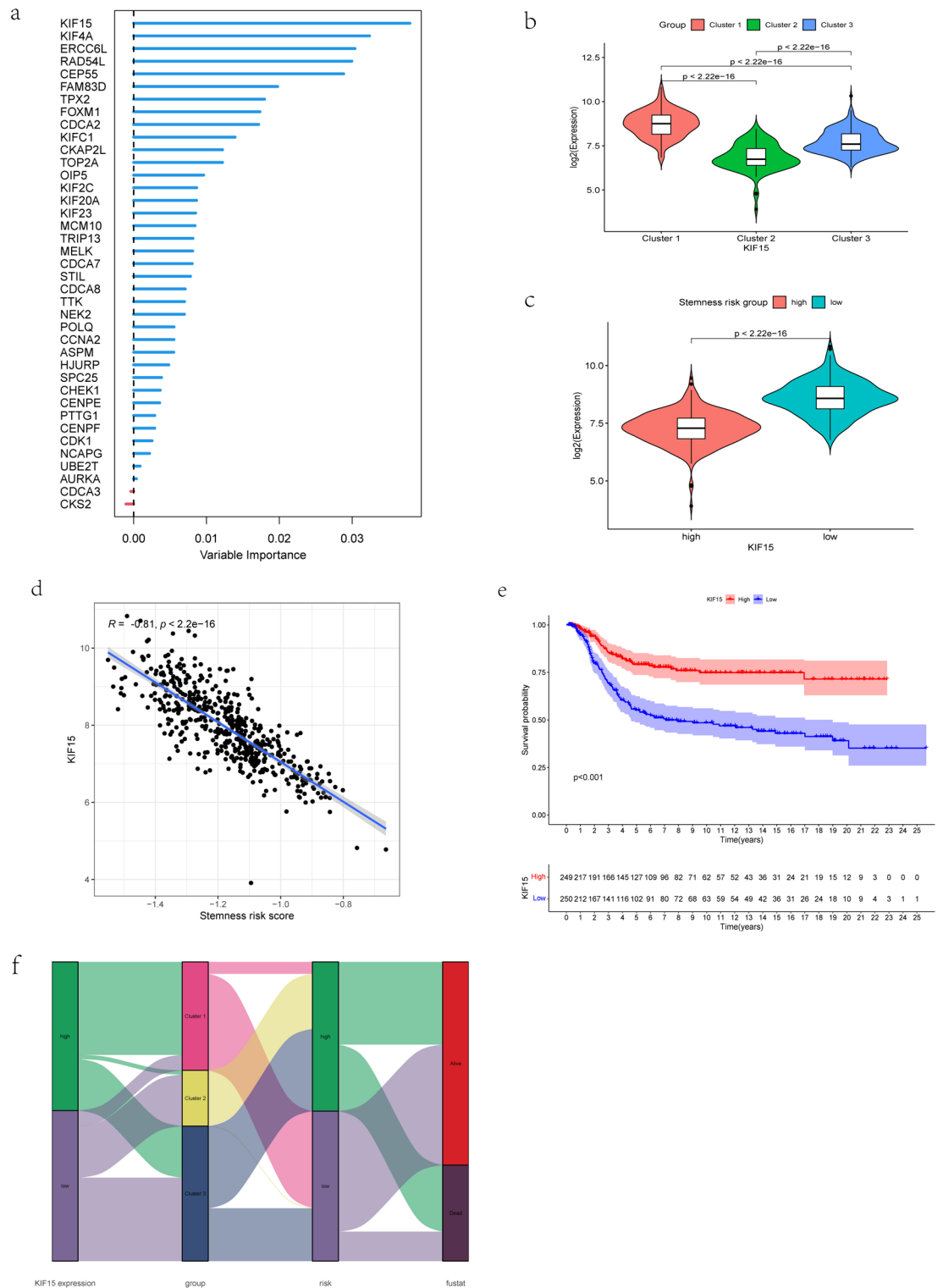


Fig. 5. Identification of the targeted gene. **(a)** The random forest survival analysis was conducted on the 39 potential pivotal genes to identify the gene that was relatively important for OS. The x-axis refers to variable importance. It represents the contribution of each selected prognosis-related gene to the prognostic model. **(b)** The expression of *KIF15* in Cluster 1, Cluster 2, and Cluster 3. **(c)** The expression of *KIF15* in stemness-related high-risk group and low-risk group. **(d)** The correlation between the expression of *KIF15* and the stemness-related risk score in TNBC ($R = -0.81, P < 0.05$). **(e)** Kaplan–Meier survival curves indicate that TNBC patients with high expression of *KIF15* have a better OS than those with low expression ($P < 0.001$). **(f)** The Sankey diagram indicated the correlation among *KIF15*, different clusters, stemness-related risk score, and survival states (alive or dead).

Knockdown of *KIF15* strengthens CSC traits of TNBC cells

From the candidate genes in the yellow module, *KIF15* was the most significant gene ascertained by LASSO and random forest survival analysis, thus the potential role of *KIF15* in TNBC characterization was further validated in terms of function. The mRNA expression level of *KIF15* and the protein level of *KIF15* were detected by qRT-PCR and WB, respectively. The mRNA level of *KIF15* was significantly higher in MDA-MB-453 and MDA-MB-231 cell lines than that in others (Fig. 6a). The protein level of *KIF15* was markedly elevated in MDA-MB-453 and MDA-MB-231 cell lines (Fig. 6b). As illustrated in Fig. 6c and d, the *KIF15* was inhibited successfully in MDA-MB-453 and MDA-MB-231 cell lines by specific siRNAs (si-*KIF15*-1 and si-*KIF15*-2). *KIF15* was considered to be associated with CSC. Consequently, the expressions of stem cell markers *Nanog*, *SOX2* and *OCT4* were evaluated through qRT-PCR and WB. The expression levels of *Nanog*, *SOX2* and *OCT4* were found to be elevated after si-*KIF15*-1 and si-*KIF15*-2 transfections in MDA-MB-453 as well as MDA-MB-231 (Fig. 6e–f). Subsequently, the expression of *KIF15* mRNA was validated in the sphere and adherence cells via qRT-PCR, demonstrating that the expression of *KIF15* was decreased in the sphere TNBC cells than that in the adherence cells (Supplemental Fig. 3). Finally, we conducted the sphere formation assay to further evaluate the stemness characteristics of *KIF15*. As shown in Fig. 6g, the sphere numbers and sizes were significantly increased after si-*KIF15*-1 and si-*KIF15*-2 transfections both in MDA-MB-453 and MDA-MB-231. The above results shown that knockdown *KIF15* promotes TNBC stemness.

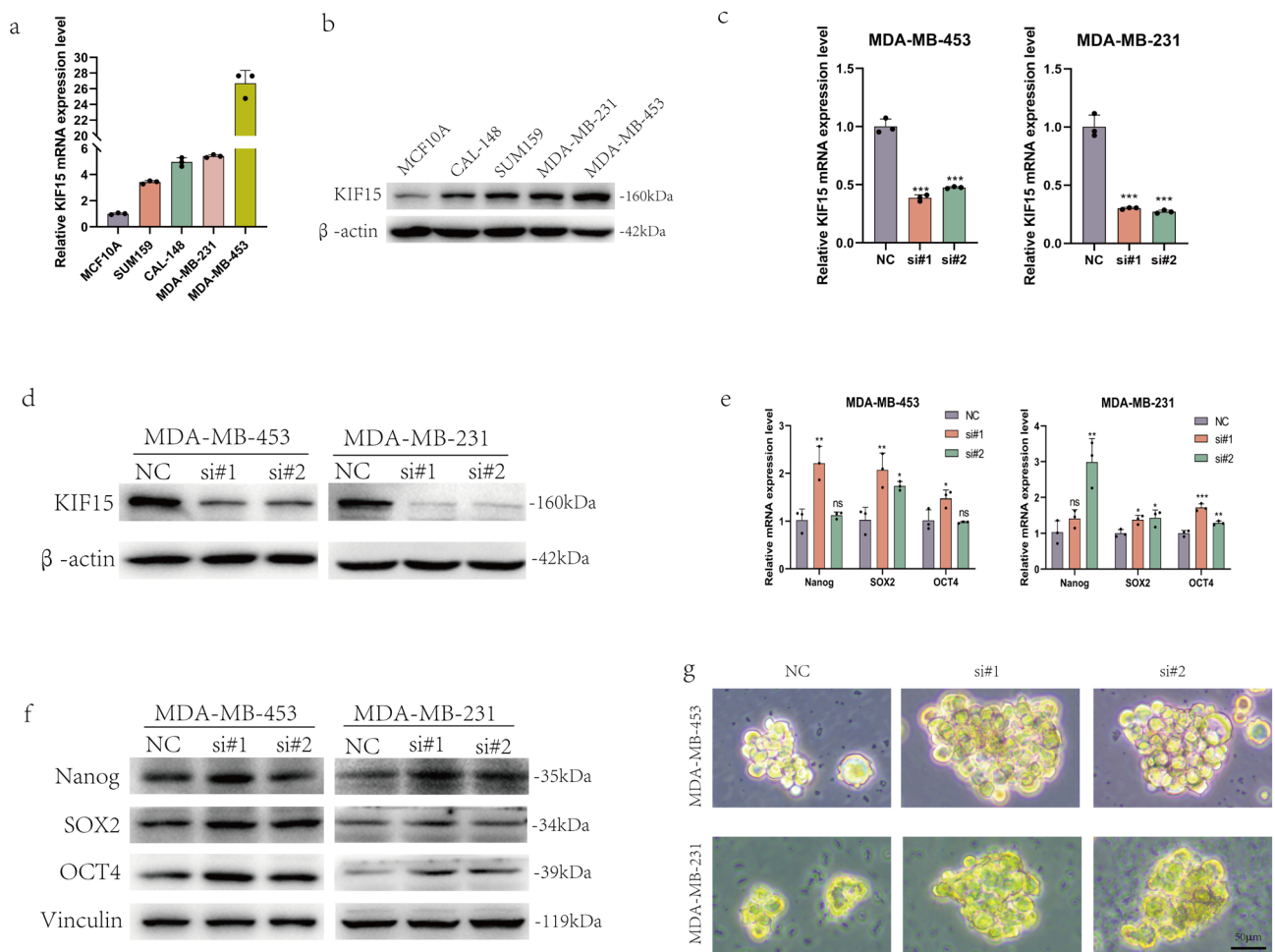


Fig. 6. Knockdown of *KIF15* strengthens the CSC traits of TNBC cells. **(a)** The qRT-PCR analysis of *KIF15* mRNA levels in normal mammary epithelial cell lines (MCF10A) and TNBC cell lines (SUM159, CAL-148, MDA-MB-231, and MDA-MB-453). **(b)** Western blot analysis of *KIF15* protein levels in normal mammary epithelial cell lines (MCF10A) and TNBC cell lines (SUM159, CAL-148, MDA-MB-231, and MDA-MB-453). **(c)** *KIF15* mRNA expression was successfully inhibited by specific siRNAs (si-*KIF15*-1 and si-*KIF15*-2) compared to the negative control (NC) in MDA-MB-453 and MDA-MB-231. **(d)** *KIF15* protein expression was inhibited by specific siRNAs (si-*KIF15*-1 and si-*KIF15*-2) compared to NC in MDA-MB-453 and MDA-MB-231. **(e)** The qRT-PCR analysis shows the mRNA expression of stem-related markers *Nanog*, *SOX2*, and *OCT4* in si-*KIF15*-1 and si-*KIF15*-2 transfected TNBC cells. **(f)** Western blot analysis shows the protein expression of stem-related markers *Nanog*, *SOX2*, and *OCT4* in si-*KIF15*-1 and si-*KIF15*-2 transfected TNBC cells. **(g)** Representative images of TNBC cell spheres after transfection with NC, si-*KIF15*-1, and si-*KIF15*-2 for 7 days, respectively.

Discussion

Recent studies confirm the concept that stem cells are essential for the creation of multicellular creatures as well as the growth of malignancies⁴⁴. It has been shown in several studies that CSCs may arise from mutations in normal stem cells yet may also be derived from mutant progenitor cells and be highly heterogeneous^{44–47}. Given the evidence, CSCs possess the ability to self-renew, the power to differentiate into any type of tumor cell, and the proliferative capacity to promote further growth of the malignant cell population^{14,48,49}. In the current research, a large cohort of multicenter TNBC patients were subjected to thorough bioinformatic studies to elucidate the molecular properties of 26 CSCs gene sets. A cancer stem cell (CSC)-related risk model was constructed to predict the prognosis as well as guide the diagnosis and treatment of TNBC. Some studies have shown that the stemness of tumor cells is associated with signaling pathways such as E2F, G2M, WNT, and NOTCH. This was consistent with our findings, where these pathways were enriched in the high-risk group related to CSC associated genes^{50–52}. Based on the CSC-related signature, *KIF15* was identified as a critical gene correlated with stemness of TNBC cells by means of univariate regression analysis and random forest approach, which was validated by in vitro experiments.

KIFs are a group of molecular motors that hydrolyze adenosine triphosphate (ATP) to create energy for material transport^{53–55}. As a member of the kinesin-12 subfamily, *KIF15* is another tetrameric spindle motor associated with the progress of mitosis and cytokinesis^{53,56,57}. Previous studies indicated that *KIF15* played a significant role in the development of several tumors and resulting a poor prognosis^{58–62}. According to the results of these research, *KIF15* was identified as an oncogene which served an essential role in the proliferation, migration, and metastasis of some malignant tumors, including lung cancer⁶³, hepatocellular cancer⁶⁴, and pancreatic cancer⁶⁵. Besides, Sheng et al. found that *KIF15* was markedly upregulated in TNBC and the downregulation of *KIF15* could suppress the invasion of TNBC cells through inducing G2/M arrest and apoptosis³⁴. However, in our study, survival analysis suggested that *KIF15* was positively associated with prognosis, as well as its expression was upregulated in TNBC. Based on the CSC-related signature, the expression of *KIF15* was higher in the low-risk group which was identified with a better clinical outcome. In addition, the Kaplan–Meier curves indicated a better OS for Cluster 1, which was found with a higher expression level of *KIF15* than the other clusters. Therefore, the findings demonstrated that increased *KIF15* acted in a defensive manner in the progression of TNBC. Interestingly, a study about gastric cancer (GC) stem cells found that GC patients with high expression of *KIF15* had longer survival times⁶⁶. It's consistent with our findings and contrary to other previous studies. The paradoxical opposite role of *KIF15* expression may be attributed to the particular way in which *KIF15* modulates cellular processes through diverse regulating networks⁶⁷. The results of our study indicated that *KIF15* may play an important role in the CSC development.

In the current research, as a gene consisting of the CSC-related signature, *KIF15* was found to be significantly upregulated in the TNBC cells, especially in MDA-MB-453 and MDA-MB-231 cell lines. After that, the *KIF15* was knocked down successfully in MDA-MB-453 and MDA-MB-231 cell lines. Notably, some stem cell markers were indicated elevated after si-*KIF15*-1 and si-*KIF15*-2 transfections in TNBC cells, including *Nanog*, *SOX2* and *OCT4*^{68–70}. These outcomes were observed simultaneously at the level of mRNA expression and protein translation. Consequently, *KIF15* was considered to play a significant role in the progression of tumor cells through CSC development regulation. Besides, the sphere numbers and sizes were markedly increased after *KIF15* was knocked down in the sphere formation assay, suggesting that *KIF15* may suppress the stemness of TNBC cells. On the other hand, to further confirm our conjecture, the expression of *KIF15* mRNA was validated in the sphere and adherence TNBC cells through qRT-PCR. The down-regulation expression of *KIF15* in the sphere cells supported our inference. In future research, we will continue to investigate the role of *KIF15* in tumor cell stemness and its specific mechanisms affecting the prognosis of TNBC. *KIF15* may potentially become one of the tumor markers for assessing TNBC prognosis in clinical practice. During the treatment follow-up of TNBC patients, it could serve as a factor for evaluating therapeutic efficacy. Furthermore, the role of CSCs in the progression of TNBC remains unclear. Further research into the relationship between CSC-related genes and TNBC prognosis could lead to a more refined classification of TNBC, aiding in the identification of new targets for targeted therapy and achieving precision treatment.

Naturally, this study has a number of limitations that must be discussed and further validated. Firstly, the data involved in this study were collected from the public databases without prospective clinical data from our medical center. The data in the database originate from different laboratories and studies, with variations in experimental conditions, sample handling, and data standardization methods. Additionally, we excluded samples with incomplete information and those with short survival times, which may affect the comprehensiveness and accuracy of the analysis. Despite the sufficient TNBC samples as training and validation sets to provide support for our findings, additional validation of the predictive effect of the CSC-related risk model on prognosis through a large cohort of sample sizes from our medical center is still important with future studies. Secondly, the CSC related genes to establish the risk model were identified through bioinformatics. Different clustering algorithms and parameter choices can affect the final consensus results. Additionally, the outcomes of consensus clustering are dependent on the quality of the initial clustering results. In some cases, determining the optimal number of clusters (k) remains a challenge. Although consensus clustering can provide stability assessments, it does not directly address the issue of selecting the number of clusters. Furthermore, our study included a relatively small dataset, and despite employing various analytical methods to construct the risk model, the issue of model overfitting still persists. It is of significant importance to conduct functional experiments to validate the association among the CSC-related risk model, TME of stemness cell, anti-tumor immunotherapies as well as the chemotherapeutic resistance. Thirdly, according to the result of our research, *KIF15* played a defensive role in the development of TNBC stem cells. However, the specific function of *KIF15* in tumor stem cells and

the mechanism of regulation of tumor cell development are not clear. These will become the focus of our future research in TNBC as well.

In conclusion, three stemness-associated subgroups with varying prognoses, immune cell infiltration, and chemotherapeutic responses were systematically discovered by unsupervised clustering on stem cell gene sets. A four-gene stemness-associated signature was established with the potential ability to predict the prognosis of TNBC patients. *KIF15* plays diverse roles in different tumors and even in different subpopulations within the same tumor. The upregulation of *KIF15* expression in a subpopulation of TNBC stem cells may lead to a loss of stem cell stemness and may inhibit the progression of TNBC.

Data availability

The datasets analyzed during the current study are available in the TCGA database (<http://cancergenome.nih.gov/>), GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse103091>), and METABRIC database (<http://www.cbioportal.org/>). The stemness-related gene sets were downloaded from StemChecker (<http://stemchecker.sysbiolab.eu/>). The data of PCR analysis were provided in the supplementary files.

Received: 17 April 2024; Accepted: 6 September 2024

Published online: 13 September 2024

References

1. Siegel, R., Miller, K., Wagle, N. & Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **73**, 17–48. <https://doi.org/10.3322/caac.21763> (2023).
2. Liedtke, C. *et al.* Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **26**, 1275–1281. <https://doi.org/10.1200/jco.2007.14.4147> (2008).
3. Liu, S. & Wicha, M. Targeting breast cancer stem cells. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **28**, 4006–4012. <https://doi.org/10.1200/jco.2009.27.5388> (2010).
4. Dick, J. Breast cancer stem cells revealed. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3547–3549. <https://doi.org/10.1073/pnas.0830967100> (2003).
5. Bai, X., Ni, J., Beretov, J., Graham, P. & Li, Y. Cancer stem cell in breast cancer therapeutic resistance. *Cancer Treat. Rev.* **69**, 152–163. <https://doi.org/10.1016/j.ctrv.2018.07.004> (2018).
6. Dontu, G., Al-Hajj, M., Abdallah, W., Clarke, M. & Wicha, M. Stem cells in normal breast development and breast cancer. *Cell Prolif.* <https://doi.org/10.1046/j.1365-2184.36.s.1.6.x> (2003).
7. Al-Hajj, M., Wicha, M., Benito-Hernandez, A., Morrison, S. & Clarke, M. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3983–3988. <https://doi.org/10.1073/pnas.0530291100> (2003).
8. Ginestier, C. *et al.* ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1**, 555–567. <https://doi.org/10.1016/j.stem.2007.08.014> (2007).
9. Park, S., Choi, J. & Nam, J. Targeting Cancer Stem Cells in Triple-Negative Breast Cancer. *Cancers (Basel)* <https://doi.org/10.3390/cancers11070965> (2019).
10. Li, W. *et al.* Unraveling the roles of CD44/CD24 and ALDH1 as cancer stem cell markers in tumorigenesis and metastasis. *Sci. Rep.* **7**, 13856. <https://doi.org/10.1038/s41598-017-14364-2> (2017).
11. Bartucci, M. *et al.* TAZ is required for metastatic activity and chemoresistance of breast cancer stem cells. *Oncogene* **34**, 681–690. <https://doi.org/10.1038/ncr.2014.5> (2015).
12. Woodward, W. *et al.* WNT/beta-catenin mediates radiation resistance of mouse mammary progenitor cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 618–623. <https://doi.org/10.1073/pnas.0606599104> (2007).
13. Li, X. *et al.* Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. *J. Natl. Cancer Inst.* **100**, 672–679. <https://doi.org/10.1093/jnci/djn123> (2008).
14. Nalla, L., Kalia, K. & Khairnar, A. Self-renewal signaling pathways in breast cancer stem cells. *Int. J. Biochem. Cell Biol.* **107**, 140–153. <https://doi.org/10.1016/j.biocel.2018.12.017> (2019).
15. Albini, A. & Sporn, M. The tumour microenvironment as a target for chemoprevention. *Nat. Rev. Cancer* **7**, 139–147. <https://doi.org/10.1038/nrc2067> (2007).
16. Colegio, O. *et al.* Functional polarization of tumour-associated macrophages by tumour-derived lactic acid. *Nature* **513**, 559–563. <https://doi.org/10.1038/nature13490> (2014).
17. Mu, X. *et al.* Tumor-derived lactate induces M2 macrophage polarization via the activation of the ERK/STAT3 signaling pathway in breast cancer. *Cell cycle* **17**, 428–438. <https://doi.org/10.1080/15384101.2018.1444305> (2018).
18. Cascone, T. *et al.* Increased tumor glycolysis characterizes immune resistance to adoptive T cell therapy. *Cell Metabolism* **27**, 977–987.e974. <https://doi.org/10.1016/j.cmet.2018.02.024> (2018).
19. Sugiura, A. & Rathmell, J. Metabolic barriers to T cell function in tumors. *J. Immunol.* **200**, 400–407. <https://doi.org/10.4049/jimmunol.1701041> (2018).
20. Bholia, N. *et al.* TGF- β inhibition enhances chemotherapy action against triple-negative breast cancer. *J. Clin. Invest.* **123**, 1348–1358. <https://doi.org/10.1172/jci65416> (2013).
21. Ibrahim, S. *et al.* Syndecan-1 is a novel molecular marker for triple negative inflammatory breast cancer and modulates the cancer stem cell phenotype via the IL-6/STAT3, Notch and EGFR signaling pathways. *Mol. Cancer* **16**, 57. <https://doi.org/10.1186/s12943-017-0621-z> (2017).
22. Inoue, K. *et al.* Interleukin 8 expression regulates tumorigenicity and metastases in androgen-independent prostate cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **6**, 2104–2119 (2000).
23. Korkaya, H., Liu, S. & Wicha, M. Regulation of cancer stem cells by cytokine networks: attacking cancer's inflammatory roots. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **17**, 6125–6129. <https://doi.org/10.1158/1078-0432.ccr-10-2743> (2011).
24. Jézéquel, P. *et al.* Gene-expression molecular subtyping of triple-negative breast cancer tumours: importance of immune response. *Breast Cancer Res. BCR* **17**, 43. <https://doi.org/10.1186/s13058-015-0550-y> (2015).
25. Jézéquel, P. *et al.* Identification of three subtypes of triple-negative breast cancer with potential therapeutic implications. *Breast cancer research : BCR* **21**, 65. <https://doi.org/10.1186/s13058-019-1148-6> (2019).
26. Leek, J., Johnson, W., Parker, H., Jaffe, A. & Storey, J. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics (Oxford, England)* **28**, 882–883. <https://doi.org/10.1093/bioinformatics/bts034> (2012).
27. Pinto, J. *et al.* StemChecker: a web-based tool to discover and explore stemness signatures in gene sets. *Nucl. Acids Res.* **43**, W72–77. <https://doi.org/10.1093/nar/gkv529> (2015).
28. Zheng, H. *et al.* Characterization of stem cell landscape and identification of stemness-relevant prognostic gene signature to aid immunotherapy in colorectal cancer. *Stem Cell. Res. Ther.* **13**, 244. <https://doi.org/10.1186/s13287-022-02913-0> (2022).

29. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7. <https://doi.org/10.1186/1471-2105-14-7> (2013).
30. Wilkerson, M. & Hayes, D. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics (Oxford, England)* **26**, 1572–1573. <https://doi.org/10.1093/bioinformatics/btq170> (2010).
31. Newman, A. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457. <https://doi.org/10.1038/nmeth.3337> (2015).
32. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612. <https://doi.org/10.1038/ncomms3612> (2013).
33. Maeser, D., Gruener, R. & Huang, R. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief. Bioinform.* <https://doi.org/10.1093/bib/bbab260> (2021).
34. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2008).
35. Hu, M., Ge, M.-R., Li, H.-X., Zhang, B. & Li, G. Identification of DAPK1 as an autophagy-related biomarker for myotonic dystrophy type 1. *Front. Genet.* **13**, 1022640 (2022).
36. Ai, D., Wang, Y., Li, X. & Pan, H. Colorectal cancer prediction based on weighted gene co-expression network analysis and variational auto-encoder. *Biomolecules* **10**, 1207 (2020).
37. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucl. Acids Res.* **51**, D587–D592 (2023).
38. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* **28**, 27–30 (2000).
39. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
40. Yu, G., Wang, L., Han, Y. & He, Q. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **16**, 284–287. <https://doi.org/10.1089/omi.2011.0118> (2012).
41. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550. <https://doi.org/10.1073/pnas.0506580102> (2005).
42. Li, K. *et al.* MicroRNA-877-5p alleviates ARDS via enhancing PI3K/Akt path by targeting CDKN1B both in vivo and in vitro. *Int. Immunopharmacol.* **95**, 107530 (2021).
43. Shaw, F. *et al.* A detailed mammosphere assay protocol for the quantification of breast stem cell activity. *J. Mammary Gland Biol. Neoplasia* **17**, 111–117. <https://doi.org/10.1007/s10911-012-9255-3> (2012).
44. Jordan, C., Guzman, M. & Noble, M. Cancer stem cells. *N. Engl. J. Med.* **355**, 1253–1261. <https://doi.org/10.1056/NEJMra061808> (2006).
45. Huntly, B. *et al.* MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *Cancer Cell* **6**, 587–596. <https://doi.org/10.1016/j.ccr.2004.10.015> (2004).
46. Cozzio, A. *et al.* Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes Dev.* **17**, 3029–3035. <https://doi.org/10.1101/gad.1143403> (2003).
47. Jamieson, C. *et al.* Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. *N. Engl. J. Med.* **351**, 657–667. <https://doi.org/10.1056/NEJMoa040258> (2004).
48. Pardal, R., Clarke, M. & Morrison, S. Applying the principles of stem-cell biology to cancer. *Nat. Rev. Cancer* **3**, 895–902. <https://doi.org/10.1038/nrc1232> (2003).
49. Kreso, A. & Dick, J. Evolution of the cancer stem cell model. *Cell stem cell* **14**, 275–291. <https://doi.org/10.1016/j.stem.2014.02.006> (2014).
50. Xie, D., Pei, Q., Li, J., Wan, X. & Ye, T. Emerging role of E2F family in cancer stem cells. *Front. Oncol.* **11**, 723137 (2021).
51. Fan, G., Xie, T., Tang, L., Han, X. & Shi, Y. 179P Integrative analysis revealed the signature of cancer stem cells and its immunosuppressive role in lung adenocarcinoma. *J. Thorac. Oncol.* **18**, S138 (2023).
52. Takebe, N. *et al.* Targeting Notch, Hedgehog, and Wnt pathways in cancer stem cells: clinical update. *Nat. Rev. Clin. Oncol.* **12**, 445–464 (2015).
53. Gao, X. *et al.* KIF15 contributes to cell proliferation and migration in breast cancer. *Hum. Cell* **33**, 1218–1228. <https://doi.org/10.1007/s13577-020-00392-0> (2020).
54. Sheng, J., Li, C., Dong, M. & Jiang, K. Identification by comprehensive bioinformatics analysis of KIF15 as a candidate risk gene for triple-negative breast cancer. *Cancer Manag. Res.* **12**, 12337–12348. <https://doi.org/10.2147/cmar.s262017> (2020).
55. Kim, H. *et al.* Human hepatocellular carcinomas with “Stemness”-related marker expression: keratin 19 expression and a poor prognosis. *Hepatology* **54**, 1707–1717. <https://doi.org/10.1002/hep.24559> (2011).
56. Liu, M. *et al.* Kinesin-12, a mitotic microtubule-associated motor protein, impacts axonal growth, navigation, and branching. *J. Neurosci. Off. J. Soc. Neurosci.* **30**, 14896–14906. <https://doi.org/10.1523/jneurosci.3739-10.2010> (2010).
57. Hirokawa, N., Noda, Y., Tanaka, Y. & Niwa, S. Kinesin superfamily motor proteins and intracellular transport. *Nat. Rev. Mol. Cell Biol.* **10**, 682–696. <https://doi.org/10.1038/nrm2774> (2009).
58. Ding, L. *et al.* KIF15 facilitates gastric cancer via enhancing proliferation, inhibiting apoptosis, and predict poor prognosis. *Cancer Cell Int.* **20**, 125. <https://doi.org/10.1186/s12935-020-01199-7> (2020).
59. Wang, J. *et al.* KIF15 knockdown suppresses gallbladder cancer development. *Eur. J. Cell Biol.* **100**, 151182. <https://doi.org/10.1016/j.ejcb.2021.151182> (2021).
60. Wang, J., Guo, X., Xie, C. & Jiang, J. KIF15 promotes pancreatic cancer proliferation via the MEK-ERK signalling pathway. *Br. J. Cancer* **117**, 245–255. <https://doi.org/10.1038/bjc.2017.165> (2017).
61. Mi, J. *et al.* Integrative pan-cancer analysis of KIF15 reveals its diagnosis and prognosis value in nasopharyngeal carcinoma. *Front. Oncol.* **12**, 772816. <https://doi.org/10.3389/fonc.2022.772816> (2022).
62. Quan, G. *et al.* KIF15 is essential for USP10-mediated PGK1 deubiquitination during the glycolysis of pancreatic cancer. *Cell Death Dis.* **14**, 137. <https://doi.org/10.1038/s41419-023-05679-2> (2023).
63. Qiao, Y. *et al.* Increased KIF15 Expression Predicts a Poor Prognosis in Patients with Lung Adenocarcinoma. *Cellular physiology and biochemistry: international journal of experimental cellular physiology, biochemistry, and pharmacology* **51**, 1–10. <https://doi.org/10.1159/000495155> (2018).
64. Chen, J. *et al.* Kinesin superfamily protein expression and its association with progression and prognosis in hepatocellular carcinoma. *J. Cancer Res. Ther.* **13**, 651–659. https://doi.org/10.4103/jcrt.JCRT_491_17 (2017).
65. Wang, Q., Han, B., Huang, W., Qi, C. & Liu, F. Identification of KIF15 as a potential therapeutic target and prognostic factor for glioma. *Oncol. Rep.* **43**, 1035–1044. <https://doi.org/10.3892/or.2020.7510> (2020).
66. Huang, C., Hu, C., Ning, Z., Huang, J. & Zhu, Z. Identification of key genes controlling cancer stem cell characteristics in gastric cancer. *World J. Gastrointest. Surg.* **12**, 442–459. <https://doi.org/10.4240/wjgs.v12.i11.442> (2020).
67. Cao, Y. *et al.* CXCL11 Correlates With Antitumor Immunity and an Improved Prognosis in Colon Cancer. *Front. Cell Dev. Biol.* **9**, 646252. <https://doi.org/10.3389/fcell.2021.646252> (2021).
68. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655. [https://doi.org/10.1016/s0092-8674\(03\)00392-1](https://doi.org/10.1016/s0092-8674(03)00392-1) (2003).
69. Hägerstrand, D. *et al.* Identification of a SOX2-dependent subset of tumor- and sphere-forming glioblastoma cells with a distinct tyrosine kinase inhibitor sensitivity profile. *Neuro Oncol.* **13**, 1178–1191. <https://doi.org/10.1093/neuonc/nor113> (2011).

70. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391. [https://doi.org/10.1016/s0092-8674\(00\)81769-9](https://doi.org/10.1016/s0092-8674(00)81769-9) (1998).

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Qiaonan Guo, Pengjun Qiu, Kelun Pan, Huikai Liang, Zundong Liu, and Jianqing Lin. The first draft of the manuscript was written by Qiaonan Guo, Pengjun Qiu, and Kelun Pan, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by Natural Science Foundation of Fujian Province (2023J01103) and Doctoral Special Foundation of the Second Affiliated Hospital of Fujian Medical University (2022BD0905).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-72406-y>.

Correspondence and requests for materials should be addressed to Z.L. or J.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024