

Analysis

# Identification of circulating tumor cells marker genes as prognostic signature in triple-negative breast cancer

Jia Hu<sup>1</sup> · Kai-Ming Zhang<sup>1</sup> · Xi Wang<sup>1</sup>

Received: 8 January 2025 / Accepted: 6 May 2025

Published online: 24 May 2025

© The Author(s) 2025 [OPEN](#)

## Abstract

**Background** Breast cancer represents a significant contributor to cancer-related mortality among women worldwide, with triple-negative breast cancer (TNBC) often exhibiting more aggressive clinical features and a heightened lethality rate. The emergence of malignant progression, along with issues of drug resistance, poses substantial challenges in the clinical management of this disease.

**Methods** The analysis of gene expression profiles at the single-cell level was conducted on circulating tumor cells (CTCs) obtained from TNBC patients, with the objective of identifying specific marker genes associated with CTCs. The TCGA database served as the training cohort for the development of a prognostic CTCs signature model, while the METABRIC dataset was utilized as the validation cohort to assess the robustness of the CTCs signature model. Furthermore, we investigated the differences in prognosis, immune scores, tumor mutational burden, and responses to immunotherapy and chemotherapy across various risk groups established based on the CTCs signature model. Colony formation and transwell assays were conducted to assess the influence of CTCs signature genes on cellular proliferation and invasive capabilities.

**Results** Seven marker genes associated with CTCs (BLOC1S3, FOXD2, GZMB, KCNJ13, NTRK3, SOAT2, and ZNF589) were identified and incorporated into a CTCs signature model. The risk score derived from this model stratified TNBC patients into high-risk and low-risk groups. Notably, the overall survival (OS) rate for the low-risk group was significantly higher than that of the high-risk group. Furthermore, the low-risk cohort exhibited more favorable prognostic outcomes and demonstrated heightened sensitivity to both immunotherapy and chemotherapy. Finally, knockdown experiments conducted in TNBC cell lines demonstrated that CTCs signature genes play a crucial role in the regulation of cellular proliferation and invasion.

**Conclusion** The CTCs signature model offers novel insights into the prognostic significance of CTC marker genes in TNBC. This understanding may serve as a valuable reference for predicting responses to immunotherapy and chemotherapy, as well as for revealing the molecular mechanisms and therapeutic targets of TNBC.

**Keywords** Triple-negative breast cancers · Circulating tumor cells · Marker genes · Prognostic model · Immunotherapy · Chemotherapy response

---

✉ Xi Wang, wangxi@sysucc.org.cn | <sup>1</sup>Department of Breast Surgery, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, SunYat-sen University Cancer Center, Guangzhou 510060, China.



## 1 Introduction

Breast cancer is the most prevalent form of cancer among women, with the highest incidence rate among female tumors [1]. While a majority of breast cancers are hormone receptor-positive (HR-positive), encompassing estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor 2 protein (HER2), approximately 10% globally are classified as triple-negative breast cancers (TNBC) [2]. TNBC is distinguished by its progression independent of ER, PR, and HER2, and a statistical analysis of 10-year mortality rates among various breast cancer subtypes revealed that TNBC exhibited the lowest 10-year recurrence-free survival rate and 10-year overall survival rate [3]. Consequently, TNBC is typically associated with a poor prognosis, characterized by frequent drug resistance, metastasis, and recurrence, which are often fatal outcomes [4].

Circulating tumor cells (CTCs) are cells originating from primary tumors that enter the bloodstream and are implicated in the metastasis and recurrence of various cancers, such as breast cancer [5, 6]. In general, a heterogeneous population of CTCs is present, with only a small subset exhibiting the ability to endure the immune and pharmacological challenges of the bloodstream and to establish metastases in distant organs, thereby playing a significant role in mortality [7]. Consequently, gaining a deeper understanding of the heterogeneity and distinctive biological markers of CTCs is imperative for informing novel approaches to prognostic prediction and evaluating clinical interventions in TNBC [8].

In recent times, the introduction of single-cell RNA-seq technologies has facilitated a comprehensive exploration of cellular heterogeneity to an unparalleled extent [9]. Initial findings from single-cell profiling have revealed substantial heterogeneity within tumors and a wide array of tumor cell subpopulations [10–12]. This study focused on analyzing the single-cell transcriptome of TNBC CTCs to enhance understanding of CTCs heterogeneity and characteristics. Through clinical correlation analysis, specific marker genes were identified and evaluated to discern CTCs with distinct biological attributes. Furthermore, this research underscores the promise of leveraging CTCs characterization to forecast prognosis and assess therapeutic response for TNBC patients.

## 2 Materials and methods

### 2.1 Data acquisition

Single-cell RNA sequencing data of TNBC CTCs was obtained from the Gene Expression Omnibus database (GEO; accession: GSE144495). This data was utilized to identify CTCs marker genes. The Cancer Genome Atlas (TCGA) dataset was acquired from the UCSC Xena platform (<https://xenabrowser.net/datapages/>) and the Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov>), while the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset was retrieved from the cBioPortal (<http://www.cbioportal.org/>). The bulk RNA sequencing data of TNBC was employed for further screening of prognosis-related CTCs marker genes, as well as for the construction and analysis of a prognostic CTCs signature.

### 2.2 Identification of CTCs marker genes

Single-cell RNA sequencing analysis was conducted utilizing the R package "monocle3" [13]. Non-linear dimensional reduction was carried out using the uniform manifold approximation and projection (UMAP) method. Identification of cell clusters and marker genes for each cluster was performed. Genes meeting the criteria of  $|\log_2(\text{fold change})| \geq 1$  and  $p \leq 0.05$  were identified as marker genes for CTCs. Gene ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway enrichments analysis were performed using the R package "ClusterProfiler" [14].

### 2.3 Construction and prognostic analysis of CTCs signature

The gene expression data of CTCs was integrated with information on overall survival (OS) time and status for each individual. With the aid of R packages "survival" [15], univariate Cox regression analysis was employed to pinpoint genes associated with prognosis. Genes exhibiting a  $p < 0.05$  in the Cox regression analysis were deemed as prognosis-related CTCs marker genes. Then, the prognosis-related CTCs marker genes were analyzed using a multivariate Cox stepwise

regression model, employing a bidirectional approach, in order to develop a CTC signature. The assumptions for the proportional hazards modeling were evaluated using the R package "survival" [15]. Subsequently, a CTCs signature model was developed with regression coefficients based on the expression of CTCs marker genes related to prognosis and patient survival data. The risk scores for each patient were calculated using the formula:  $\text{risk score} = \sum_{i=1}^n (\text{coef}_i * \text{Expression}_i)$ , where "coef" represents the regression coefficient of the gene from the multivariate Cox regression, and "Expression" denotes the gene expression level. Patients were then categorized into high-risk and low-risk groups based on the median cut-off value. The prognostic value of the risk score was assessed through Kaplan–Meier analysis. Furthermore, the prognostic capacity of the risk score was evaluated using receiver operating characteristic (ROC) curve analysis and the area under the curve (AUC) through the R package "survivalROC" [16].

## 2.4 Tumor clinical pathological characteristics analysis

The study evaluated the predictive utility of the risk score in conjunction with clinicopathological variables. The Wilcoxon test was employed to examine the association between the risk score and clinical characteristics.

## 2.5 Tumor immune microenvironment analysis

The immunoscore was calculated using the ESTIMATE algorithm within the R package "estimate" [17]. The R package "cibersort" was employed to determine the proportion of 22 types of immune cells and analyze the extent of immune cell infiltration in the immune microenvironment [18].

## 2.6 Tumor mutational burden (TMB) analysis

The mutation data was extracted from TCGA and subsequently analyzed utilizing the R package "maftools" [19]. The TMB score was assessed using the following formula:  $\text{TMB (mut/mb)} = \text{total number of mutations/size of target region coding area}$ .

## 2.7 Immunotherapy response prediction

The Immunophenoscore (IPS), which can be derived using a machine learning approach based on two gene categories (PD-1 and CTLA4) that are intricately linked to immune cell function, was utilized to examine the response to immunotherapy in patients with varying risk scores. IPS was assessed by z-scores of genes related to immunity which was extracted from the Cancer Immunome Atlas (TCIA, <https://tcia.at/home>). Furthermore, we assessed the expression levels of immune checkpoint-associated genes in order to determine the sensitivity to immunotherapy across two distinct risk groups.

## 2.8 Chemotherapy response prediction

The therapeutic efficacy of chemotherapy was evaluated through the Genomics of Drug Sensitivity in Cancer (GDSC, <https://www.cancerrxgene.org>). The half maximal inhibitory concentration (IC50) of chemotherapeutic agents in individuals belonging to various risk groups was determined utilizing a predictive algorithm implemented via the R package "oncoPredict" [20]. A total of 198 GDSC drugs were selected for the analysis of IC50 values across high-risk and low-risk groups. It is important to note that lower IC50 values indicate a greater sensitivity to specific chemotherapeutic agents, while higher values suggest reduced sensitivity.

## 2.9 Cell culture and transfection

The normal breast cell line MCF-10A, along with the triple-negative breast cancer cell lines MDA-MB-231 and MDA-MB-468, were maintained in RPMI (Roswell Park Memorial Institute) 1640 medium (Gibco, USA) enriched with 10% fetal bovine serum at 37 °C. Small interfering RNAs (siRNAs) were introduced into the cells via transfection using Lipofectamine 3000 (Invitrogen, USA).

## 2.10 Quantitative real-time polymerase chain reaction (qRT-PCR)

Total RNA was extracted from cells utilizing Trizol (Invitrogen, USA). The qRT-PCR assay was conducted using the ABI ViiATM7Dx Real-Time PCR System (Life Technologies, USA). For the detection of mRNA, 1 µg of total RNA was utilized for complementary DNA (cDNA) synthesis via a Reverse Transcription Kit (Takara, Japan). Subsequently, the synthesized cDNA was subjected to qRT-PCR with specific primers and the SYBR Green Real-time PCR Master Mix Kit (Toyobo, Japan). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) served as an internal reference, and the expression levels of the target genes were quantified using the  $2^{-\Delta\Delta Ct}$  method.

The RT-PCR primers were designed as the following:

BLOC1S3: Forward 5'-TTCCAGAACTGCCTTACCC-3'; Reverse 5'-TAGAACCAGCACACGGAACC-3'.

FOXD2: Forward 5'-CCAAAGCCTTCTACGCGGC-3'; Reverse 5'-CTGAGGAGTGCGGACCTAAC-3'.

GZMB: Forward 5'-GATCATCGGGGACATGAGG-3'; Reverse 5'-GGAGGCATGCCATTGTTTCG-3'.

KCNJ13: Forward 5'-GCAAAAGAACTGAGAAATACAGCCT-3'; Reverse 5'-TGCAGCTGTGAAACTGGTGA-3'.

NTRK3: Forward 5'-TTCTCTTCTTCTCCTCGGGC-3'; Reverse 5'-TCGCTGCTTCTTTGAAACGC-3'.

SOAT2: Forward 5'-TGGAAACACTGAGACGCACA-3'; Reverse 5'-TCATCAAGCAGGGACTTGCG-3'.

ZNF589: Forward 5'-CAGAAGGCAGTCACAGCAGA-3'; Reverse 5'-GTTGCGGAGGACTGACTCAA-3'.

GAPDH: Forward 5'-GAAAGCCTGCCGGTGACTAA-3'; Reverse 5'-TTCCCCTTCTCAGCCTTGAC-3'.

## 2.11 Colony formation assay

Cells were subjected to trypsinization and subsequently plated to evaluate clonogenic survival. Over a period of seven days, the cells were permitted to form colonies in a medium maintained at 37 °C. The colonies were then fixed using 4% formaldehyde for a duration of 15 min, followed by staining with crystal violet (Beyotime) for 20 min. After washing and capturing photographic documentation, the visible colonies were quantified utilizing ImageJ software.

## 2.12 Transwell assay

Cell invasion was assessed utilizing transwell-chamber culture systems (Becton Dickinson). Cells were introduced into the upper chamber of the transwell, which was coated with matrigel, using serum-free RPMI-1640 medium. Concurrently, RPMI-1640 medium supplemented with 10% fetal bovine serum was placed in the lower chamber. After a 24-h incubation period, non-invading cells adhering to the upper surface of the upper chamber were removed with cotton swabs. Subsequently, the cells that had migrated to the lower surface of the filters were fixed and stained with crystal violet (Beyotime). The number of invaded cells was quantified using a light microscope (Leica).

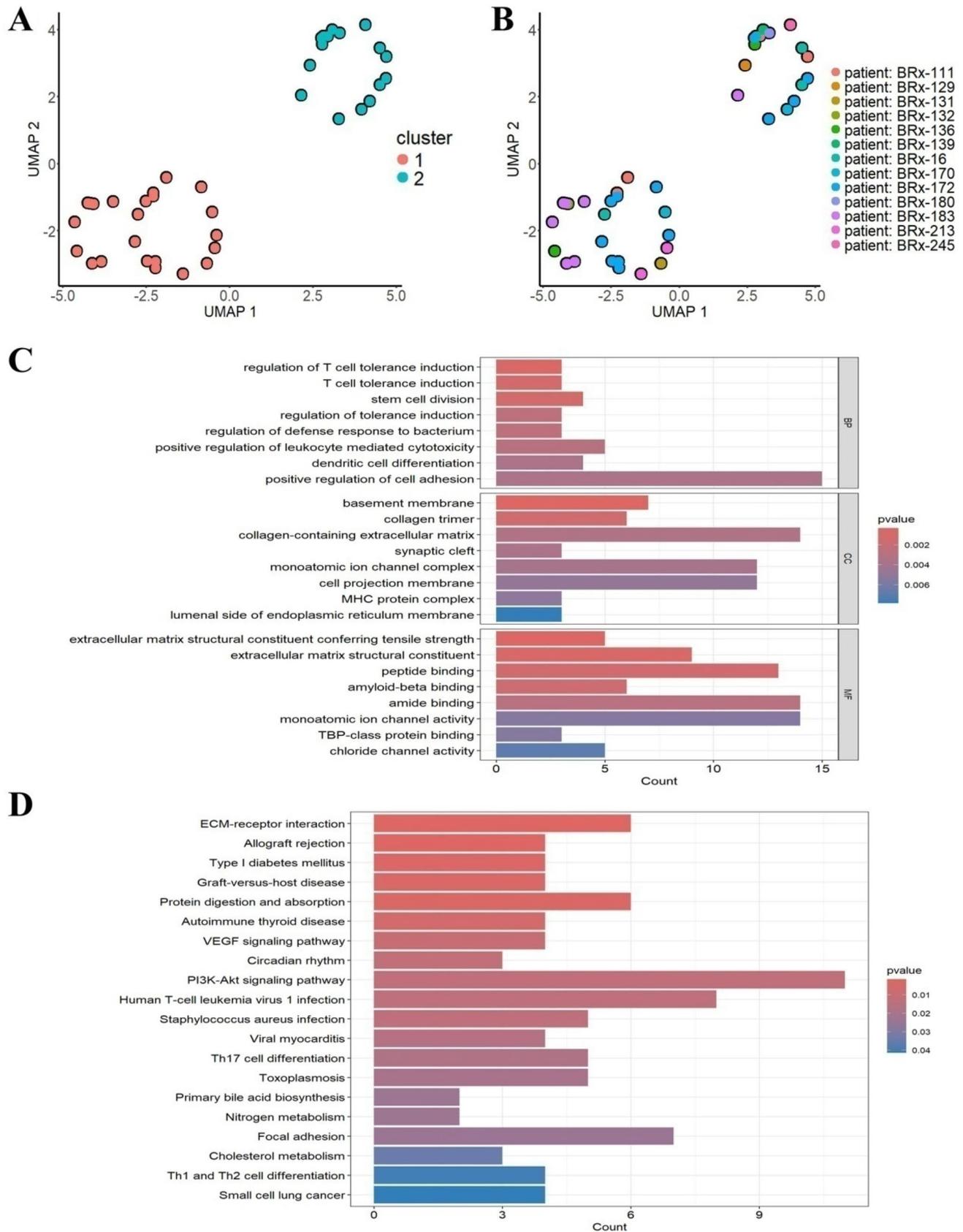
## 2.13 Statistical analysis

Statistical analyses were conducted utilizing R version 4.3.0. The Wilcoxon test was employed to ascertain statistical variances in categorical variables. Survival curves were generated using the Kaplan–Meier method, and the significance of differences was evaluated through the log-rank test. Statistical significance was established at a significance level of  $p < 0.05$ .

## 3 Results

### 3.1 Cell heterogeneity in CTCs of TNBC

Initially, the GSE144495 dataset was utilized to generate a single-cell RNA-seq dataset comprising 39 CTCs obtained from 13 patients with TNBC. Through the application of UMAP for dimension reduction, two distinct cell clusters were identified, indicating that these clusters can represent the similarities and variances among CTCs at the individual cell level (Fig. 1A). Furthermore, the presence of CTCs from the same TNBC patient within both identified cell clusters suggests a



**Fig. 1** Identification CTCs marker genes by single-cell RNA sequencing analysis. **A** UMAP plot colored by various cell clusters. **B** UMAP plot colored by various patients. **C** GO functional enrichment analysis of CTCs marker genes. **D** KEGG functional enrichment analysis of CTCs marker genes

notable transcriptional heterogeneity among CTCs from the same patient (Fig. 1B). Consequently, the heterogeneity of CTCs was delineated through the analysis of single-cell RNA-seq data.

In order to examine the differential expression of genes across cell clusters, a set of 366 marker genes that exhibit variation between two CTCs clusters in UMAP was subjected to screening based on specific criteria. Subsequently, GO and KEGG functional enrichment analyses were conducted on the 366 marker genes to investigate the heterogeneity of cell functions within CTCs. The results of the GO functional enrichment analysis indicated significant enrichment of these genes in biological processes such as immune system response, stem cell functions, cell adhesion, and molecular functions related to the extracellular matrix structure (Fig. 1C). Notably, these genes were predominantly enriched in cellular components such as the basement membrane, extracellular matrix, and major histocompatibility complex (MHC) protein complex (Fig. 1C). Additionally, the KEGG enrichment analysis revealed a significant association of these genes with extracellular matrix interactions and T cell immune responses (Fig. 1D).

### 3.2 Construction of prognostic CTCs signature

In order to investigate the prognostic significance of CTCs marker genes in TNBC, we conducted univariate Cox regression analysis and Kaplan–Meier survival analysis for each CTCs marker gene with respect to OS. By applying screening criteria based on ER, HER2, and PR status, we identified 132 TNBC samples from TCGA dataset. In the TCGA dataset, the univariate Cox regression analysis identified 14 CTCs marker genes significantly associated with OS ( $p < 0.05$ ) (Fig. 2A). Subsequently, These 14 CTCs marker genes were integrated into a multivariate Cox proportional hazards regression model to identify the relevant genes and their associated coefficients. Ultimately, 7 significant genes (BLOC1S3, FOXD2, GZMB, KCNJ13, NTRK3, SOAT2 and ZNF589) were incorporated into the CTCs signature, and the risk score for each patient in the TCGA dataset was then computed using this CTCs signature model (Fig. 2B).

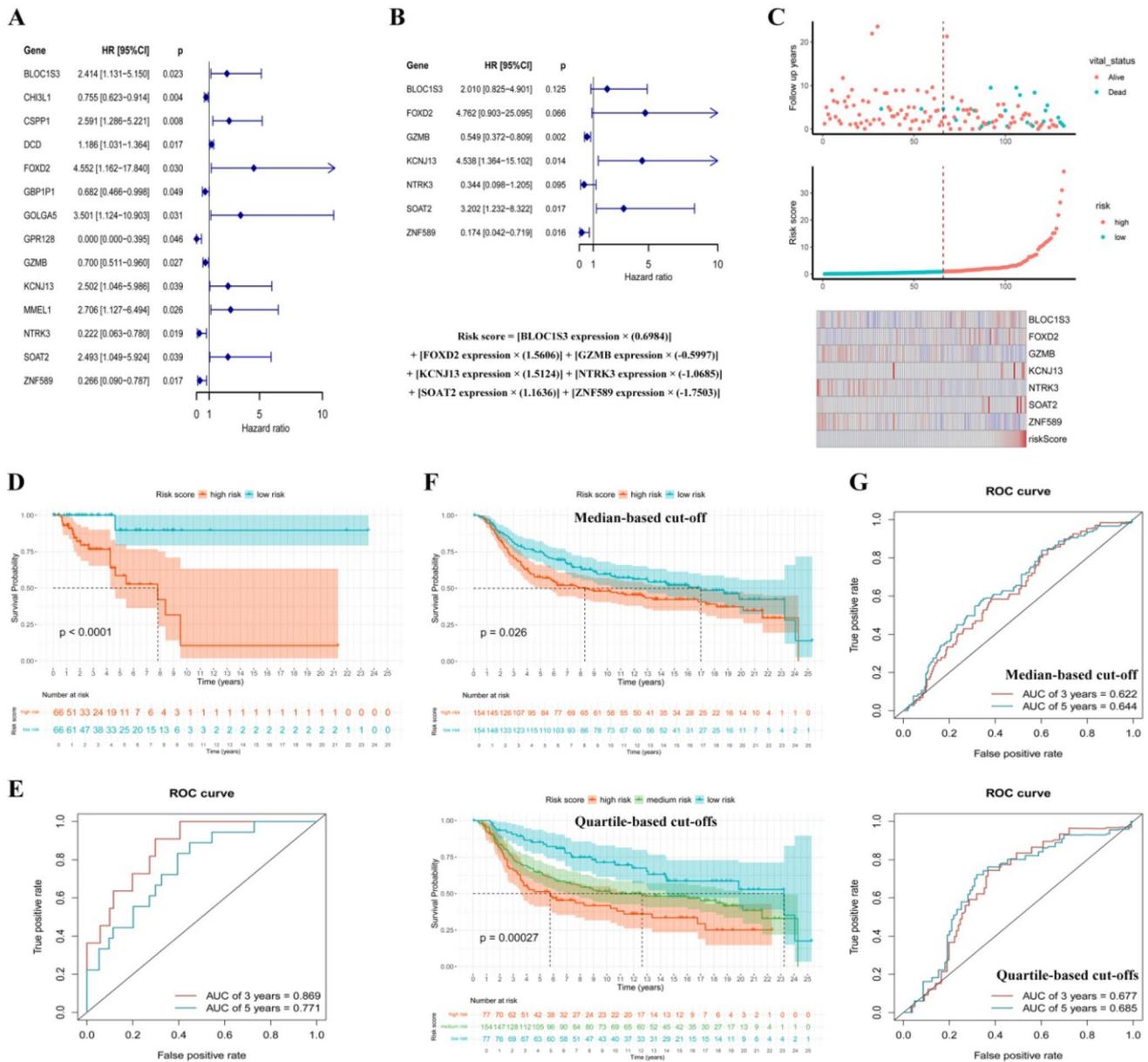
In the TCGA dataset, TNBC patients were stratified into high-risk and low-risk groups based on a median cut-off risk score, and the relationship between risk score, survival outcomes, and CTCs signature expression levels was visually represented (Fig. 2C). Analysis indicated that patients with higher risk scores were associated with increased mortality events. Kaplan–Meier survival analysis demonstrated that high-risk patients had significantly shorter OS compared to low-risk patients ( $p < 0.0001$ ) (Fig. 2D). Time-dependent ROC analysis was conducted to evaluate the predictive performance of the CTCs signature, yielding AUC values of 0.869 and 0.771 at 3 and 5 years separately (Fig. 2E).

The CTCs signature was validated in the METABRIC dataset, comprising 308 TNBC patients, by employing screening criteria based on ER, HER2, and PR status. CTCs signature risk score was computed for each patient, leading to their classification into high-risk and low-risk groups. As demonstrated by Kaplan–Meier analysis, high-risk patients exhibited significantly worse prognoses compared to low-risk individuals using the median-based cut-off ( $p = 0.026$ ) or the quartile-based cut-offs ( $p = 0.00027$ ) (Fig. 2F). A time-dependent ROC curve analysis was performed, resulting in AUC values of 0.622 at 3 years and 0.644 at 5 years when employing a median-based cut-off, and the analysis yielded AUC values of 0.677 at 3 years and 0.685 at 5 years when utilizing quartile-based cut-offs (Fig. 2G). These results suggest an association between the CTCs signature (comprising BLOC1S3, FOXD2, GZMB, KCNJ13, NTRK3, SOAT2 and ZNF589 genes) and the survival prognosis of TNBC patients.

### 3.3 Clinical features and mutation patterns of different risk groups of CTCs signature

In conjunction with clinicopathological characteristics associated with prognosis, an examination was conducted to assess the relationship between CTCs signature and clinicopathological variables such as TNM stage, recurrence, tumor size, and lymph node status. The distribution of CTCs signature risk scores in TNBC patients from the TCGA dataset was calculated and stratified based on each clinical factor. The findings indicated higher risk scores in the stage III/stage IV, T3/T4, N2/N3, and recurred groups ( $p < 0.05$ ) (Fig. 3A). Additionally, we conducted a comparable analysis utilizing the METABRIC dataset, and it was also showed that TNBC patients exhibiting elevated risk scores were associated with advanced TNM staging, increased recurrence rates, and larger tumor dimensions ( $p < 0.05$ ) (Fig. 3B). These results indicated a potential association between the CTCs signature and tumor progression.

We also illustrated the mutational landscape, utilizing waterfall plots to present mutation data for genes across the various risk groups (Fig. 3C). Notably, TP53 and TTN emerged as the two most frequently mutated genes in both risk groups. Additionally, USH2A and ZFH4 were identified as unique high-frequency mutated genes in high-risk patients, ranking fifth and ninth, respectively. Conversely, low-risk patients exhibited a higher mutation frequency of CREBBP (sixth) and FCGBP (eighth).



**Fig. 2** Screen and construction of the prognostic CTCs signature model. **A** Univariate Cox analysis of CTCs marker genes that were associated with overall survival of TNBC in the TCGA dataset. **B** CTCs marker genes related risk score model was developed by multivariate Cox stepwise analysis. **C** Distribution of risk scores, the survival of patients and the heatmap of expression of CTCs signature genes in the TCGA dataset. **D** Kaplan–Meier curves of risk scores for the overall survival of TNBC patients in the TCGA dataset. **E** ROC curves of risk scores to predict the 3 years and 5 years overall survival in the TCGA dataset. **F** Kaplan–Meier curves of risk scores for the overall survival of TNBC patients in the METABRIC dataset. **G** ROC curves of risk scores to predict the 3 years and 5 years overall survival in the METABRIC dataset

### 3.4 Immune features and therapy response of different risk groups of CTCs signature

In order to elucidate the impact of CTCs signature on the immune microenvironment of patients with TNBC, an assessment was conducted on the immunoscore and levels of immune infiltration. In the TCGA dataset, high-risk patients exhibited a decrease in immune, stromal, and estimate scores, while tumor purity was elevated in this group (Fig. 4A). Subsequent analysis compared the proportions of 22 immune cell types between high-risk and low-risk groups. In the TCGA dataset, high-risk patients demonstrated heightened levels of infiltration by CD4 memory resting T cells and M0 macrophages, alongside reduced levels of CD8T cells and CD4 memory activated T cells (Fig. 4B). In a comparable manner,



T cells, CD4 memory activated T cells, M1 macrophages, and resting mast cells (Fig. 4D). These findings underscore the association between distinct immune cell infiltration patterns and the CTCs signature.

Then, we conducted an investigation to determine whether there are differences in the potential for immunotherapy response among various risk groups. Initially, we calculated the TMB within the TCGA cohort to assess the likelihood of immunotherapy response across different risk groups, as genomic alterations may significantly influence tumor immunity. As results, low-risk patients exhibited a higher TMB compared to high-risk patients, indicating that low-risk individuals may have a more favorable response to immunotherapy (Fig. 5A). Subsequently, the IPS analysis demonstrated that low-risk patients had an increased likelihood of responding to anti-PD-1 and anti-CTLA4 treatments, further suggesting that low-risk group may respond more effectively to immune checkpoint inhibitors (ICIs) (Fig. 5B). Additionally, we evaluated the expression levels of immune checkpoint-associated genes within the two risk groups, revealing that genes such as CTLA4 and PDCD1 were expressed at significantly higher levels in the high-risk group compared to the low-risk group (Fig. 5C). Furthermore, findings from the METABRIC cohort corroborated that the expression levels of immune checkpoint-associated genes were linked to a risk signature based on CTCs marker genes, reinforcing the notion that low-risk patients are likely to exhibit a better response to ICIs (Fig. 5D).

Finally, the chemotherapy drug sensitivity of each patient with TNBC was evaluated using the GDSC database, revealing variations in IC50 values between two distinct risk groups. In the TCGA dataset, 24 drugs from the GDSC exhibited a negative correlation between risk scores and IC50 values, and notable examples of these drugs include Olaparib, Ruxolitinib, Niraparib, and Axitinib (Fig. 6A). Therefore, it was observed that TNBC patients classified as low-risk demonstrated enhanced responsiveness to chemotherapy. Similarly, individuals classified in the low-risk group exhibited heightened sensitivity to a range of drugs across the METABRIC datasets (Fig. 6B). In summary, the CTCs signature demonstrates predictive capabilities for determining the efficacy of immunotherapy and chemotherapy treatments in TNBC patients.

### 3.5 Validation of CTCs signature genes

To assess the expression levels of CTCs signature genes, we employed qRT-PCR to evaluate the relative differential expression of BLOC1S3, FOXD2, GZMB, KCNJ13, NTRK3, SOAT2, and ZNF589 in TNBC cell lines (MDA-MB-231 and MDA-MB-468 cells) compared to normal breast epithelial cells (MCF-10A cells). Among the seven CTC signature genes analyzed, FOXD2 and KCNJ13 exhibited elevated expression levels in MDA-MB-231 cells relative to MCF-10A cells, while GZMB and NTRK3 demonstrated reduced expression in MDA-MB-231 cells compared to MCF-10A cells (Fig. 7A). Furthermore, BLOC1S3, FOXD2, and SOAT2 were expressed at higher levels in MDA-MB-468 cells than in MCF-10A cells, whereas NTRK3 and ZNF589 showed lower expression levels in MDA-MB-468 cells compared to MCF-10A cells (Fig. 7B). These findings indicate that CTC signature genes exhibit differential expression in TNBC cells, although the specific differentially expressed genes vary across different TNBC cell lines.

To assess the functional roles of CTCs signature genes, siRNAs targeting BLOC1S3, FOXD2, GZMB, KCNJ13, NTRK3, SOAT2, and ZNF589 were transfected into MDA-MB-231 cells to create knockdown cell lines (Fig. 8A). The results from the colony formation assay indicated that the knockdown of GZMB, NTRK3, and ZNF589 enhanced the capacity for colony formation, whereas the knockdown of BLOC1S3, FOXD2, KCNJ13, and SOAT2 diminished this capacity (Fig. 8B). Furthermore, findings from the transwell assay demonstrated that the knockdown of GZMB, NTRK3, and ZNF589 facilitated increased cell invasion, while the knockdown of BLOC1S3, FOXD2, KCNJ13, and SOAT2 resulted in a reduction of cell invasion (Fig. 8C). Collectively, these results suggest that CTCs signature genes are instrumental in the processes of proliferation and invasion in TNBC cells.

## 4 Discussion

Notwithstanding advancements in early detection and therapeutic interventions, breast cancer continues to be deemed incurable once metastasis and recurrence occur [21, 22]. TNBC, which accounts for approximately 20% of all breast cancer cases, is distinguished by its more aggressive clinical characteristics relative to other breast cancer subtypes [23]. TNBC cases exhibit certain common traits, including a basal-like phenotype, histological features indicative of a pushing border of invasion and a central necrotic area, as well as a propensity for high rates of metastasis [24]. Furthermore, TNBC is marked by a rapidly increasing recurrence rate within the initial two years post-diagnosis, peaking at 2 to 3 years [25]. In comparison to patients with non-TNBC, those diagnosed with TNBC

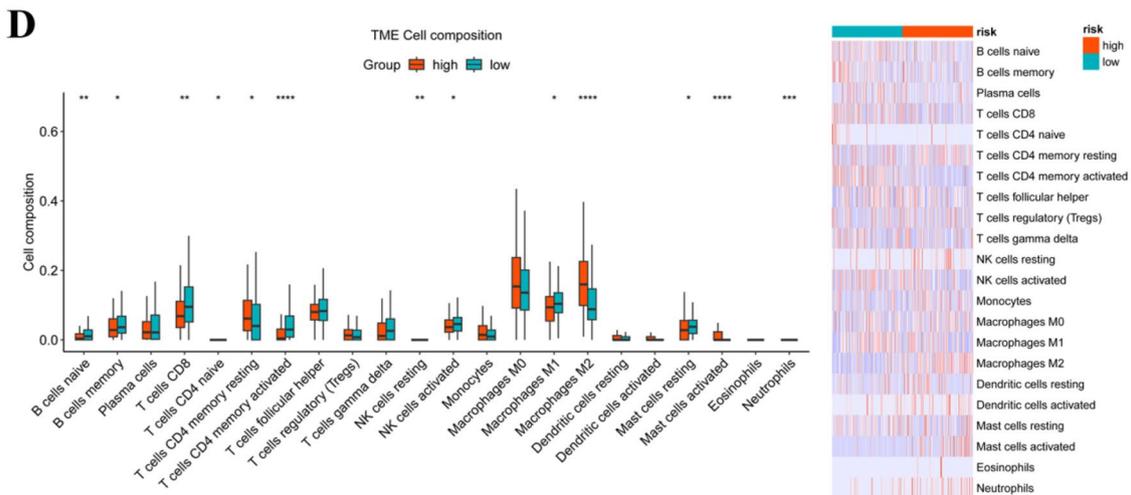
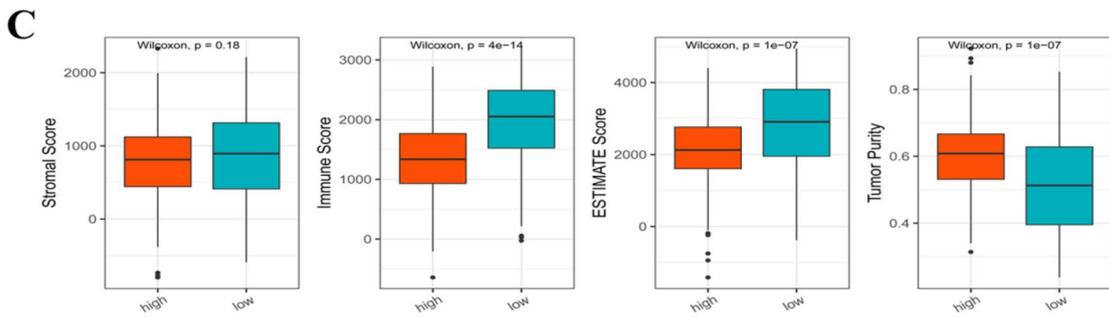
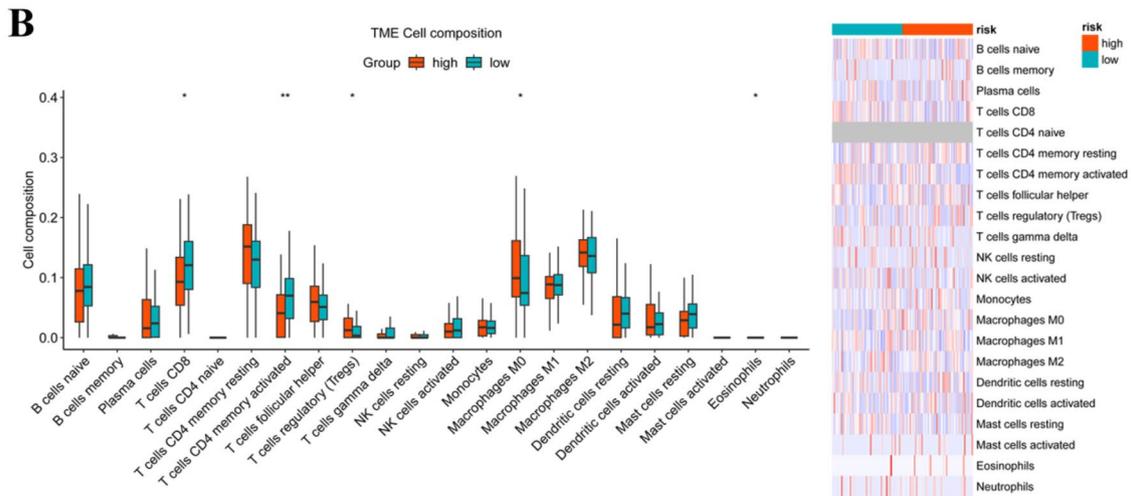
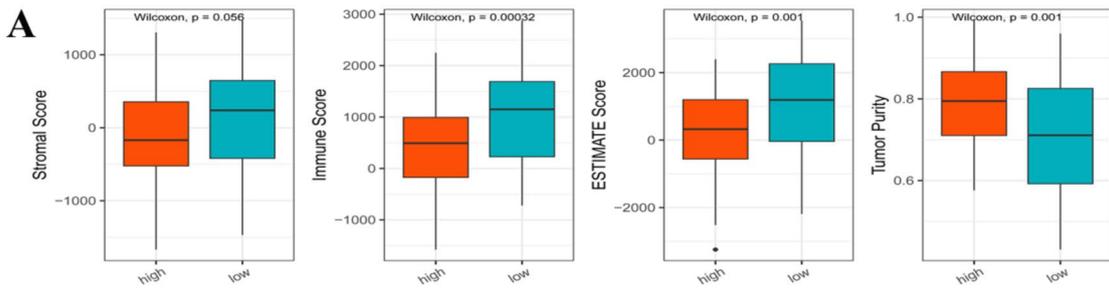
**Fig. 4** Correlation between risk scores and the level of immune infiltration. **A** The relation between risk score and immune score, stromal score, estimate score, and tumor purity of TNBC patients in the TCGA dataset. **B** Box plot and heatmap of the number of immune cells in the TCGA dataset by CIBERSORT. **C** The relation between risk score and immune score, stromal score, estimate score, and tumor purity of TNBC patients in the METABRIC dataset. **D** Box plot and heatmap of the number of immune cells in the METABRIC dataset by CIBERSORT. (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ )

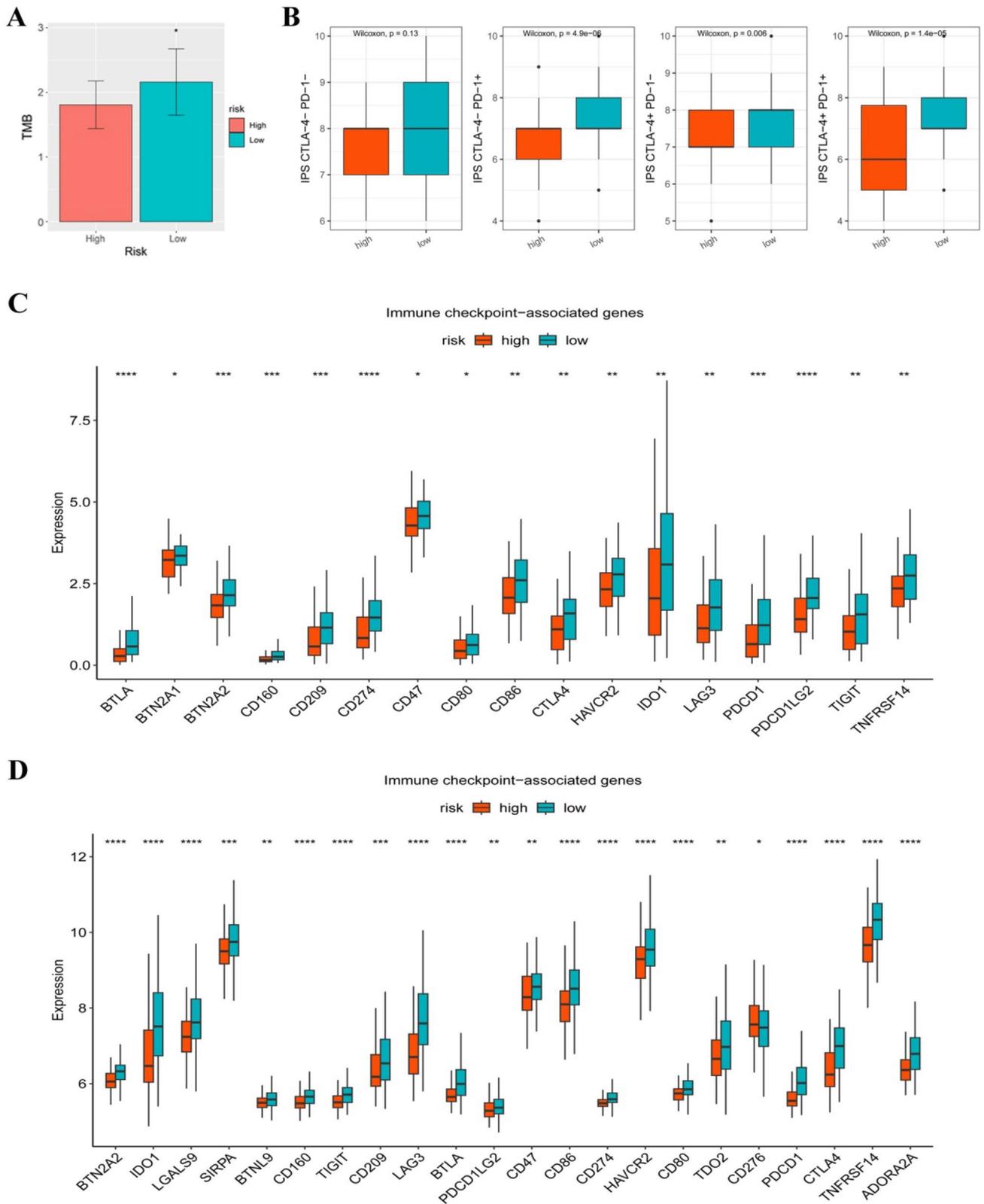
experience significantly shorter relapse-free survival (RFS), characterized by a reduced interval from diagnosis to relapse and from relapse to mortality [26]. Consequently, metastasis and recurrence continue to be the predominant causes of mortality among patients with TNBC.

In order to clarify the essential genes associated with TNBC, we conducted a single-cell gene expression profile analysis on CTCs obtained from TNBC patients. This analysis facilitated the identification of CTCs marker genes, which were subsequently examined for their prognostic relevance. Following rigorous screening, we identified seven CTCs marker genes (BLOC1S3, FOXD2, GZMB, KCNJ13, NTRK3, SOAT2, and ZNF589) that were utilized to develop a prognostic CTCs signature model, which demonstrated a correlation with survival, progression and therapy (immune and drug) response in TNBC patients. Research has demonstrated that a subset of CTCs can detach from primary tumors and enter the bloodstream, where they endure various physical, oxidative, pharmacological, and environmental stresses prior to disseminating to distant organs [27, 28]. CTCs have the capacity to initiate metastasis and contribute to disease recurrence, and the biological characteristics of these CTCs are closely associated with the metastatic and recurrent behavior of cancers, including TNBC [29]. Consequently, in contrast to the tissue signature models for TNBC, the CTCs signature model is applicable for liquid biopsy, facilitating the monitoring of metastasis and recurrence in TNBC cases [30].

Given that the CTCs signature model is correlated with poor survival probability and malignant progression in patients with TNBC, we hypothesized that the seven genes (BLOC1S3, FOXD2, GZMB, KCNJ13, NTRK3, SOAT2, and ZNF589) play critical roles in the biological characteristics and behaviors of CTCs. Our investigation demonstrated that GZMB, NTRK3, and ZNF589 serve as protective risk factors influencing the prognosis of TNBC patients. Furthermore, the expression levels of GZMB, NTRK3, and ZNF589 were found to be reduced in TNBC cell lines, where they exhibited inhibitory effects on cell proliferation and invasion. GZMB (granzyme B) encodes a protein that is part of the granzyme subfamily, with the preproprotein being secreted by natural killer (NK) cells and cytotoxic T lymphocytes (CTLs). This protein undergoes proteolytic processing to yield the active protease, which induces apoptosis in target cells [31]. GZMB is implicated in chronic inflammation and the tumor immune microenvironment, and its low expression may correlate with poor prognostic outcomes in various cancers [32, 33]. NTRK3 (neurotrophic receptor tyrosine kinase 3) encodes a member of the neurotrophic tyrosine receptor kinase family, functioning as a membrane-bound receptor that, upon binding with neurotrophins, undergoes autophosphorylation and activates members of the MAPK pathway [34]. Signaling through NTRK3 is associated with cellular differentiation and may contribute to the development of proprioceptive neurons responsible for sensing body position; furthermore, mutations in NTRK3 have been linked to medulloblastomas, breast carcinomas, and other malignancies [35, 36]. ZNF589 (zinc finger protein 589) is part of the extensive family of Krüppel-associated box domain-zinc finger transcription factors and plays a significant role in the differentiation of hematopoietic stem cells [37]. Research has indicated that ZNF589 may serve as a potential prognostic biomarker and a tumor suppressor in breast cancer [38].

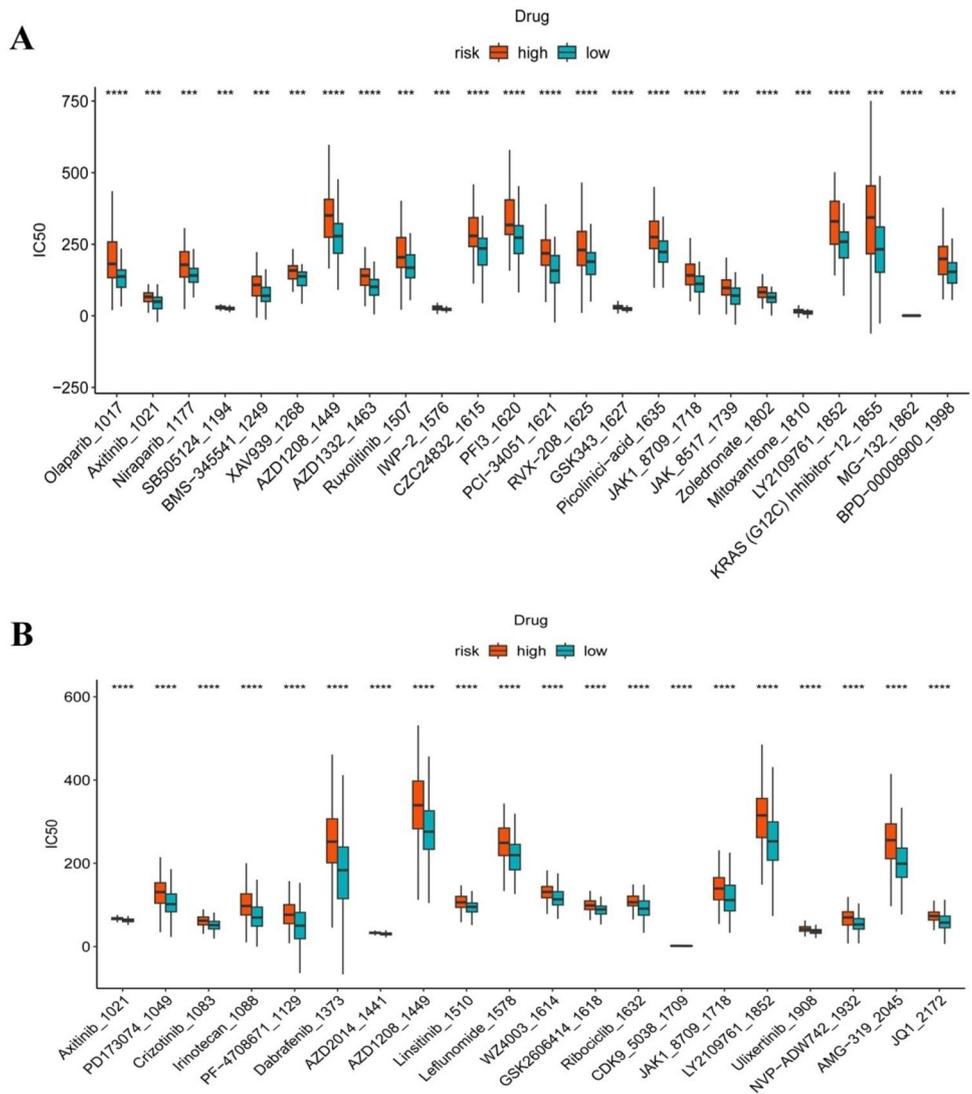
Conversely, the genes BLOC1S3, FOXD2, KCNJ13, and SOAT2 have been linked to negative clinical outcomes. Furthermore, these genes exhibited elevated expression levels in TNBC cell lines, where they facilitated cellular proliferation and invasion. BLOC1S3 (biogenesis of lysosomal organelles complex 1 subunit 3) encodes a protein that is a constituent of the BLOC1 multi-subunit protein complex, which is essential for the biogenesis of specialized organelles within the endosomal-lysosomal system, including platelet dense granules and melanosomes [39]. Mutations in BLOC1S3 are implicated in Hermansky-Pudlak syndrome, a condition characterized by lysosomal storage defects, and the expression levels of BLOC1S3 have been shown to predict outcomes in cancer patients [40–42]. FOXD2 (forkhead box D2) is a member of the forkhead family of transcription factors, which play a critical role in lymphocyte development and immunoregulation; it has also been suggested that FOXD2 is involved in T cell activation [43, 44]. Numerous studies have reported elevated expression of FOXD2 in various cancers, underscoring its significant role in cancer progression [45, 46]. KCNJ13 (potassium inwardly rectifying channel subfamily J member 13) encodes a protein that is part of the inwardly rectifying potassium channel family, which forms ion channel pores facilitating the influx of potassium ions into cells; mutations in this gene are associated with snowflake vitreoretinal degeneration [47, 48]. SOAT2 (sterol O-acyltransferase 2) encodes a membrane-bound enzyme located in the endoplasmic reticulum that synthesizes intracellular cholesterol esters from long-chain fatty acyl-CoA and cholesterol, with these esters subsequently stored as cytoplasmic lipid droplets [49]. SOAT2



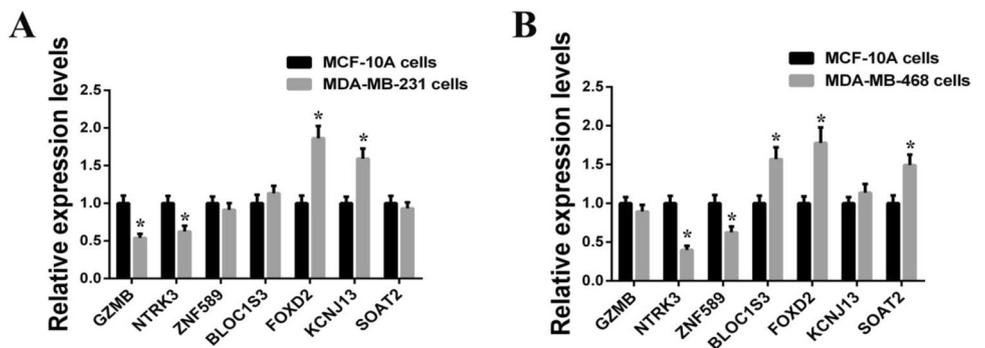


**Fig. 5** Correlation between risk scores and the sensitivity to immunotherapy. **A** The relation between risk score and TMB levels of TNBC patients in the TCGA dataset. **B** The relation between risk score and IPS of TNBC patients in the TCGA dataset. **C** The relation between risk score and immune checkpoint-associated genes of TNBC patients in the TCGA dataset. **D** The relation between risk score and immune checkpoint-associated genes of TNBC patients in the METABRIC dataset. (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ )

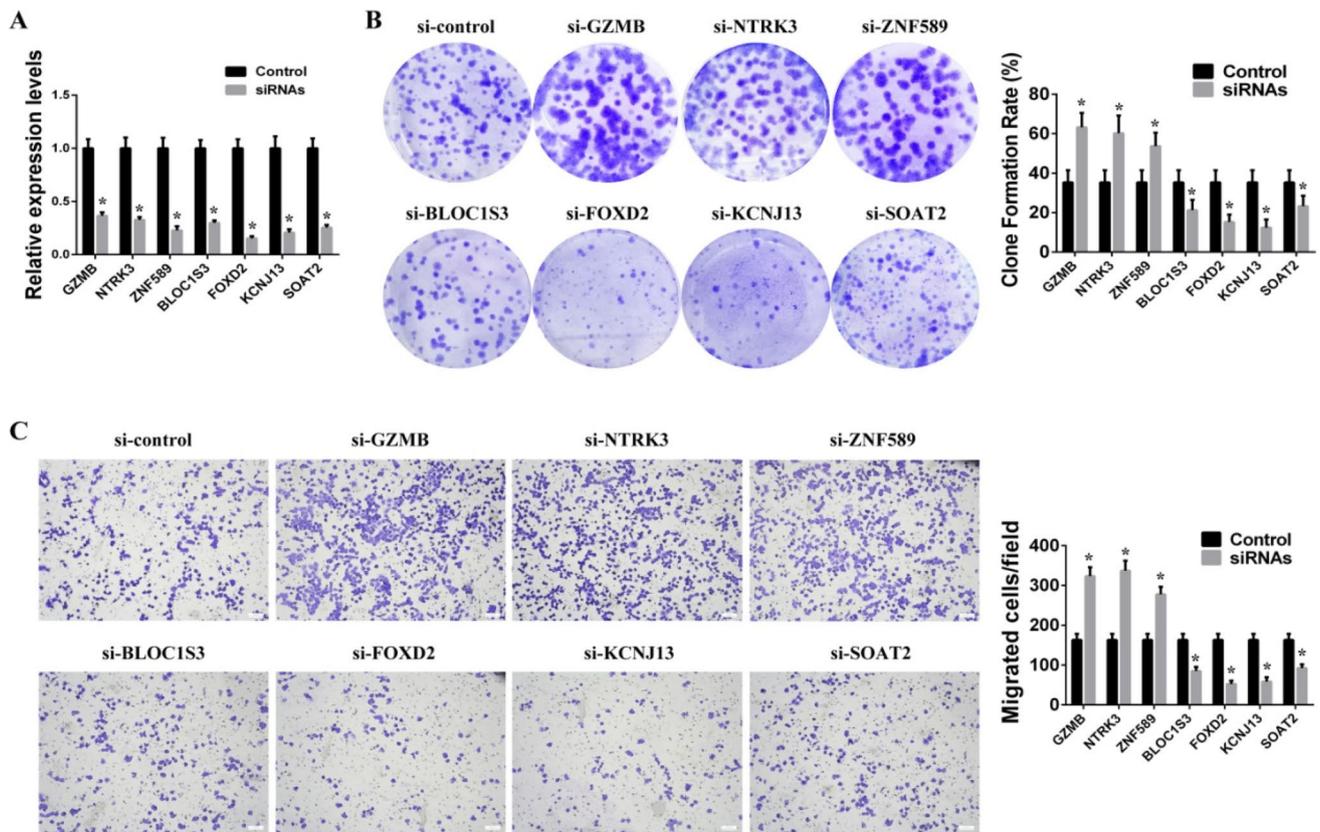
**Fig. 6** Correlation between risk scores and the sensitivity to chemotherapy. **A** The relation between risk score and the estimated IC50 of chemotherapy drugs in the TCGA dataset. **B** The relation between risk score and the estimated IC50 of chemotherapy drugs in the METABRIC dataset. (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ )



**Fig. 7** Verification of CTCs signature genes expression. **A** Expression of CTCs signature genes in normal breast cell line MCF-10A and TNBC cell line MDA-MB-231. **B** Expression of CTCs signature genes in normal breast cell line MCF-10A and TNBC cell line MDA-MB-468. (\*  $p < 0.05$ )



has demonstrated independent prognostic value regarding survival outcomes and may serve as a predictive biomarker in breast and lung cancers [50, 51]. Thus, the CTCs signature genes are closely related to various aggressive cancer traits, including immune evasion, chemotherapy resistance, and metastatic behavior.



**Fig. 8** Verification of CTCs signature genes function. **A** The efficiency of siRNA targeting CTCs signature genes. **B** Colony formation assay on TNBC cells following the knockdown of CTCs signature genes. **C** Transwell assay on TNBC cells following the knockdown of CTCs signature genes. (\*  $p < 0.05$ )

## 5 Limitations

However, there are several limitations of this study. Firstly, the predictive efficacy of the CTCs signature model may be influenced by the characteristics of the dataset utilized, indicating a necessity for further optimization to enhance its generalizability and clinical relevance. Secondly, this investigation is deficient in validation of CTCs signature genes at protein levels within tissue samples derived from TNBC patients, encompassing primary tumors, metastatic sites, and circulating tumor cells. Furthermore, *in vivo* functional studies are essential to clarify the mechanistic roles of these genes in the progression of TNBC. Prospective validation through clinical trials is also required to bolster the transparency and future direction of this study.

## 6 Conclusion

In conclusion, this study presents a transcriptional analysis of CTCs in TNBC, which has led to the identification of specific CTCs marker genes and the establishment of a prognostic CTCs signature associated with TNBC survival and progression. Moreover, the genes comprising this CTCs signature play a role in the modulation of cellular proliferation and invasion in TNBC cells. Consequently, the findings of this research may offer potential gene targets for prognostic assessment, thereby contributing to the enhancement of clinical outcomes and informing novel therapy strategies in patients with TNBC.

**Author contributions** HJ, ZKM, and WX: Methodology design. HJ and ZKM: Collection and curation of pathological and clinical data. HJ and ZKM: Data analysis and interpretation. WX: Project supervision. HJ: Manuscript writing. All the authors contributed to the article and approved the final manuscript.

**Funding** The study was supported by grants from the National Natural Science Foundation of China (Grant Nos. 81772826).

**Data availability** The datasets used and analyzed during the current study are available from GEO (<https://www.ncbi.nlm.nih.gov/geo/>), TCGA (<https://portal.gdc.cancer.gov/>), cBioPortal (<http://www.cbioportal.org/>). Further inquiries can be directed to the corresponding author.

## Declarations

**Competing interests** The authors declare no competing interests.

**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/>.

## References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin.* 2024;74(1):12–49.
2. Tsang JYS, Tse GM. Molecular classification of breast cancer. *Adv Anat Pathol.* 2020;27(1):27–35.
3. Li Y, Zhang H, Merkhher Y, Chen L, Liu N, Leonov S, et al. Recent advances in therapeutic strategies for triple-negative breast cancer. *J Hematol Oncol.* 2022;15(1):121.
4. Bianchini G, De Angelis C, Licata L, Gianni L. Treatment landscape of triple-negative breast cancer—expanded options, evolving needs. *Nat Rev Clin Oncol.* 2022;19(2):91–113.
5. Lawrence R, Watters M, Davies CR, Pantel K, Lu YJ. Circulating tumour cells for early detection of clinically relevant cancer. *Nat Rev Clin Oncol.* 2023;20(7):487–500.
6. Ring A, Nguyen-Sträuli BD, Wicki A, Aceto N. Biology, vulnerabilities and clinical applications of circulating tumour cells. *Nat Rev Cancer.* 2023;23(2):95–111.
7. Castro-Giner F, Aceto N. Tracking cancer progression: from circulating tumor cells to metastasis. *Genome Med.* 2020;12(1):31.
8. Lisencu LA, Trancă S, Bonci EA, Pașca A, Mihu C, Irimie A, et al. The role of circulating tumor cells in the prognosis of metastatic triple-negative breast cancers: a systematic review of the literature. *Biomedicines.* 2022;10(4):769.
9. Lawson DA, Kessenbrock K, Davis RT, Pervolarakis N, Werb Z. Tumour heterogeneity and metastasis at single-cell resolution. *Nat Cell Biol.* 2018;20(12):1349–60.
10. Papalexli E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol.* 2018;18(1):35–45.
11. Kinker GS, Greenwald AC, Tal R, Orlova Z, Cuoco MS, McFarland JM, et al. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat Genet.* 2020;52(11):1208–18.
12. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet.* 2021;53(9):1334–47.
13. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol.* 2014;32(4):381–6.
14. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7.
15. Therneau TM, Ou FS. Using multistate models with clinical trial data for a deeper understanding of complex disease processes. *Clin Trials.* 2024;2:17407745241267862.
16. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics.* 2000;56(2):337–44.
17. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.
18. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12(5):453–7.
19. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018;28(11):1747–56.
20. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform.* 2021;22(6):bbab60.
21. Riggio AI, Varley KE, Welm AL. The lingering mysteries of metastatic recurrence in breast cancer. *Br J Cancer.* 2021;124(1):13–26.
22. Kimbung S, Loman N, Hedenfalk I. Clinical and molecular complexity of breast cancer metastases. *Semin Cancer Biol.* 2015;35:85–95.
23. Popovic LS, Matovina-Brko G, Popovic M, Punie K, Cvetanovic A, Lambertini M. Targeting triple-negative breast cancer: a clinical perspective. *Oncol Res.* 2023;31(3):221–38.

24. Hua Z, White J, Zhou J. Cancer stem cells in TNBC. *Semin Cancer Biol.* 2022;82:26–34.
25. Karim AM, Eun Kwon J, Ali T, Jang J, Ullah I, Lee YG, et al. Triple-negative breast cancer: epidemiology, molecular mechanisms, and modern vaccine-based treatment strategies. *Biochem Pharmacol.* 2023;212:115545.
26. Tariq K, Farhangi A, Rana F. TNBC vs non-TNBC: A retrospective review of differences in mean age, family history, smoking history, and stage at diagnosis. *Clin Adv Hematol Oncol.* 2014;12(6):377–81.
27. Zhan Q, Liu B, Situ X, Luo Y, Fu T, Wang Y, et al. New insights into the correlations between circulating tumor cells and target organ metastasis. *Signal Transduct Target Ther.* 2023;8(1):465.
28. Deng Z, Wu S, Wang Y, Shi D. Circulating tumor cell isolation for cancer diagnosis and prognosis. *EBioMedicine.* 2022;83:104237.
29. Pereira-Veiga T, Schneegans S, Pantel K, Wikman H. Circulating tumor cell-blood cell crosstalk: biology and clinical relevance. *Cell Rep.* 2022;40(9):111298.
30. Piñeiro R. Introduction—biology of breast cancer metastasis and importance of the analysis of CTCs. *Adv Exp Med Biol.* 2020;1220:1–10.
31. Le MK, Oishi N, Satou A, Miyaoka M, Kawashima I, Mochizuki K, et al. Molecular and clinicopathological features of granzyme B-negative extranodal NK/T-cell lymphoma. *Hum Pathol.* 2024;143:10–6.
32. Du T, Gao Q, Zhao Y, Gao J, Li J, Wang L, et al. Long Non-coding RNA LINC02474 affects metastasis and apoptosis of colorectal cancer by inhibiting the expression of GZMB. *Front Oncol.* 2021;11:651796.
33. Liang Z, Pan L, Shi J, Zhang L. C1QA, C1QB, and GZMB are novel prognostic biomarkers of skin cutaneous melanoma relating tumor microenvironment. *Sci Rep.* 2022;12(1):20460.
34. Rogers C, Morrissette JJD, Sussman RT. NTRK point mutations and their functional consequences. *Cancer Genet.* 2022;262–263:5–15.
35. Luo Y, Kaz AM, Kannurn S, Welsch P, Morris SM, Wang J, et al. NTRK3 is a potential tumor suppressor gene commonly inactivated by epigenetic mechanisms in colorectal cancer. *PLoS Genet.* 2013;9(7):e1003552.
36. Khotskaya YB, Holla VR, Farago AF, Mills Shaw KR, Meric-Bernstam F, Hong DS. Targeting TRK family proteins in cancer. *Pharmacol Ther.* 2017;173:58–66.
37. Venturini L, Stadler M, Manukjan G, Scherr M, Schlegelberger B, Steinemann D, et al. The stem cell zinc finger 1 (SZF1)/ZNF589 protein has a human-specific evolutionary nucleotide DNA change and acts as a regulator of cell viability in the hematopoietic system. *Exp Hematol.* 2016;44(4):257–68.
38. Fan J, Zhang Z, Chen D, Chen H, Yuan W, Zhou L, et al. Bioinformatic analysis of the expression and prognosis of ZNF589 in human breast cancer. *Transl Cancer Res.* 2021;10(5):2286–304.
39. Starcevic M, Dell'Angelica EC. Identification of snapin and three novel proteins (BLOS1, BLOS2, and BLOS3/reduced pigmentation) as subunits of biogenesis of lysosome-related organelles complex-1 (BLOC-1). *J Biol Chem.* 2004;279(27):28393–401.
40. Pennamen P, Tingaud-Sequeira A, Michaud V, Morice-Picard F, Plaisant C, Vincent-Delorme C, et al. Novel variants in the BLOC1S3 gene in patients presenting a mild form of Hermansky-Pudlak syndrome. *Pigment Cell Melanoma Res.* 2021;34(1):132–5.
41. Xu C, Pei D, Liu Y, Yu Y, Guo J, Liu N, et al. Identification of a novel tumor microenvironment prognostic signature for bladder urothelial carcinoma. *Front Oncol.* 2022;12:818860.
42. Feng Z, Qian H, Li K, Lou J, Wu Y, Peng C. Development and validation of a 7-gene prognostic signature to improve survival prediction in pancreatic ductal adenocarcinoma. *Front Mol Biosci.* 2021;8:676291.
43. Jonsson H, Peng SL. Forkhead transcription factors in immunology. *Cell Mol Life Sci.* 2005;62(4):397–409.
44. Johansson CC, Dahle MK, Blomqvist SR, Grønning LM, Aandahl EM, Enerbäck S, et al. A winged helix forkhead (FOXD2) tunes sensitivity to cAMP in T lymphocytes through regulation of cAMP-dependent protein kinase R1alpha. *J Biol Chem.* 2003;278(19):17573–9.
45. Fei T, Zhou EC, Wang XJ. FOXD2 regulations IQGAP3 mediated Ca(2+) signaling pathway to facilitate gastric adenocarcinoma cell promotion. *Kaohsiung J Med Sci.* 2023;39(11):1087–95.
46. He H, Yuan F, Li Y, Pi G, Shi H, Li Y, et al. Comprehensive analysis and experimental validation of FOXD2 as a novel potential prognostic biomarker associated with immune infiltration in head and neck squamous cell carcinoma. *Curr Comput Aided Drug Des.* 2024. <https://doi.org/10.2174/0115734099302492240405065505>.
47. Yin W, Kim HT, Wang S, Gunawan F, Wang L, Kishimoto K, et al. The potassium channel KCNJ13 is essential for smooth muscle cytoskeletal organization during mouse tracheal tubulogenesis. *Nat Commun.* 2018;9(1):2815.
48. Ashkenazy N, Sengillo JD, Iyer PG, Negron CI, Yannuzzi NA, Berrocal AM. Phenotypic expansion of KCNJ13-associated snowflake vitreo-retinal degeneration. *Ophthalmic Genet.* 2023;44(5):505–8.
49. Bhattacharjee P, Rutland N, Iyer MR. Targeting sterol O-acyltransferase/acyl-CoA:cholesterol acyltransferase (ACAT): a perspective on small-molecule inhibitors and their therapeutic potential. *J Med Chem.* 2022;65(24):16062–98.
50. Xiao Y, Huang Y, Jiang J, Chen Y, Wei C. Identification of the prognostic value of Th1/Th2 ratio and a novel prognostic signature in basal-like breast cancer. *Hereditas.* 2023;160(1):2.
51. Wang Z, Cao Z, Dai Z. ACAT2 may be a novel predictive biomarker and therapeutic target in lung adenocarcinoma. *Cancer Rep.* 2024;7(2):e1956.

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