

ANALYSIS

Open Access



# Integrated single-cell and spatial mapping coupled with machine learning unveils core stemness landscapes and regulatory drivers in triple-negative breast cancer

Zhenzhong Huo<sup>1</sup>, Weibo Sun<sup>2</sup>, Chun Lou<sup>2\*</sup> and Tiansong Yang<sup>3\*</sup>

\*Correspondence:

Chun Lou

drlou999@126.com

Tiansong Yang

Yangtiansong2006@163.com

<sup>1</sup>Heilongjiang University of Chinese Medicine, Harbin 150040, Heilongjiang, China

<sup>2</sup>Department of Breast Surgery, Harbin Medical University Cancer Hospital, Harbin 150081, Heilongjiang, China

<sup>3</sup>Department of Rehabilitation II, The First Affiliated Hospital of Heilongjiang University of Chinese Medicine, Harbin 150040, Heilongjiang, China

## Abstract

**Objective** Triple-negative breast cancer (TNBC) exhibits pronounced intratumoral heterogeneity, and cancer stem cells (CSCs) are thought to play a pivotal role in this process. However, the molecular regulatory mechanisms linking CSC-associated stemness features to tumor progression remain insufficiently elucidated.

**Methods** We integrated single-cell RNA sequencing (scRNA-seq), spatial transcriptomics, and bulk transcriptomic data to identify high-stemness cell populations using the inferCNV and CytoTRACE algorithms. Stemness-related genes were evaluated for feature importance through hdWGCNA combined with machine learning approaches, and an XGBoost-based risk prediction model was constructed. Cellular differentiation trajectories were inferred using Monocle3 and scTour, while the effects of core genes on stemness pathways and malignant biological behaviors were assessed via CellChat analysis, SHAP attribution, and scTenifoldKnn-based virtual knockdown experiments.

**Results** We successfully established a predictive model comprising five core stemness-related genes (*CALD1*, *ANP32B*, *FIS1*, *CD82*, and *APLP2*), with the high-stemness score group exhibiting poorer prognosis and enhanced immune evasion. Trajectory analysis confirmed that the high-stemness subpopulation resided at the initiation stage of differentiation. Enrichment analyses revealed highly active Notch signaling communication, and virtual knockdown of hub genes effectively suppressed the expression of stemness markers such as *NOTCH1*. In addition, drug sensitivity analysis identified BI.2536 and related compounds as exhibiting higher therapeutic sensitivity in the high-risk group.

**Conclusion** Our predictive model offers a novel perspective on the stemness landscape of TNBC. These core genes play key roles in maintaining stemness and also serve as potential molecular targets for personalized therapies aimed at TNBC stem-like cells.

**Keywords** Triple-negative breast cancer, Cancer stem cells, Single-cell RNA sequencing, Spatial transcriptomics, Machine learning



## 1 Introduction

Breast cancer is the leading cause of cancer-related mortality among women world-wide, ranking first in both incidence and mortality among female malignancies [1, 2]. Despite continuous advances in diagnostic and therapeutic strategies in recent years, the clinical management of breast cancer remains a major challenge due to pronounced tumor heterogeneity and therapeutic resistance. Triple-negative breast cancer (TNBC) is a distinct subtype characterized by the absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) expression [3]. Compared with other subtypes, TNBC lacks actionable molecular targets, which severely limits available treatment options. Accumulating evidence indicates that TNBC exhibits unique and highly complex molecular features [4, 5]. Although classification systems based on PAM50 gene expression profiles have been established [6], their ability to accurately predict patient survival outcomes and therapeutic responses remains limited. Therefore, an in-depth dissection of the molecular regulatory mechanisms underlying TNBC, along with the development of robust and reliable biomarkers, is of critical importance for advancing personalized diagnosis and treatment strategies.

Cancer stem cells (CSCs) represent a functional subpopulation of tumor cells and are widely regarded as a primary source of intrinsic intratumoral heterogeneity, continuously driving tumor progression during both primary tumor maintenance and distant organ metastasis [7, 8]. These cells possess pronounced self-renewal capacity, can differentiate into multiple tumor cell lineages, and demonstrate the ability to reconstitute complete tumor architecture in *in vivo* transplantation assays. Traditionally, CSCs have been defined by a set of markers enriched in specific tumor types or biological contexts [9]. However, recent studies have further identified a subset of cells with high stemness scores that do not necessarily conform to classical CSC phenotypic definitions, yet still exhibit enhanced tumor-initiating capacity, invasiveness, and therapeutic resistance [10]. Therefore, a systematic dissection of the regulatory networks governing high-stemness cells and their interactions with the tumor microenvironment is essential for deepening our understanding of stemness heterogeneity in TNBC and its microenvironmental regulatory landscape [11–13].

Building upon these findings, this study integrated single-cell RNA sequencing, spatial transcriptomics, and clinical cohort data to identify high-stemness cell subpopulations that drive tumor progression. By combining weighted gene co-expression network analysis with ensemble machine learning algorithms, we further delineated a set of stemness-associated signature genes. Based on these signature genes, we constructed a robust and high-accuracy machine learning model for predicting high-stemness cells. Collectively, these results not only elucidate the molecular characteristics of high-stemness cell populations from a bioinformatics perspective but also provide a theoretical framework for prognostic stratification and precision targeting of stem-like cells in TNBC.

## 2 Results

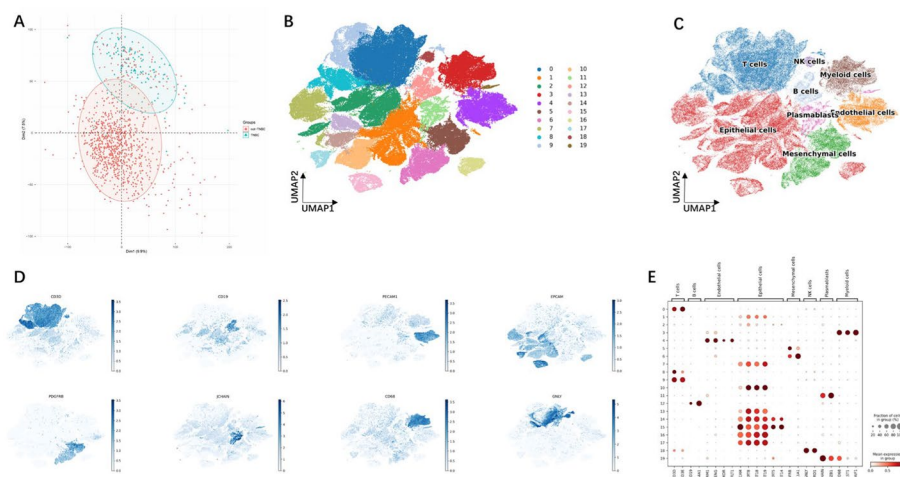
### 2.1 Comparison between TNBC and non-TNBC highlights the importance of stem-like cells

Compared with non-TNBC, TNBC exhibits extensive intratumoral heterogeneity and more pronounced stemness features. Therefore, to identify TNBC-specific potential targets and determine whether the tumor microenvironment differs from that of non-TNBC, we analyzed bulk RNA-seq data from the TCGA-BRCA cohort, including 967

non-TNBC and 114 TNBC samples. Principal component analysis (PCA) revealed a clear separation between non-TNBC and TNBC samples (Fig. 1A). In addition, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of differentially expressed genes (DEGs) between non-TNBC and TNBC demonstrated significant enrichment of stemness-related pathways in TNBC (Supplementary Fig. 1A–C). Collectively, these findings indicate that differential distribution of stemness features represents a defining characteristic distinguishing TNBC from non-TNBC.

## 2.2 Cell type identification

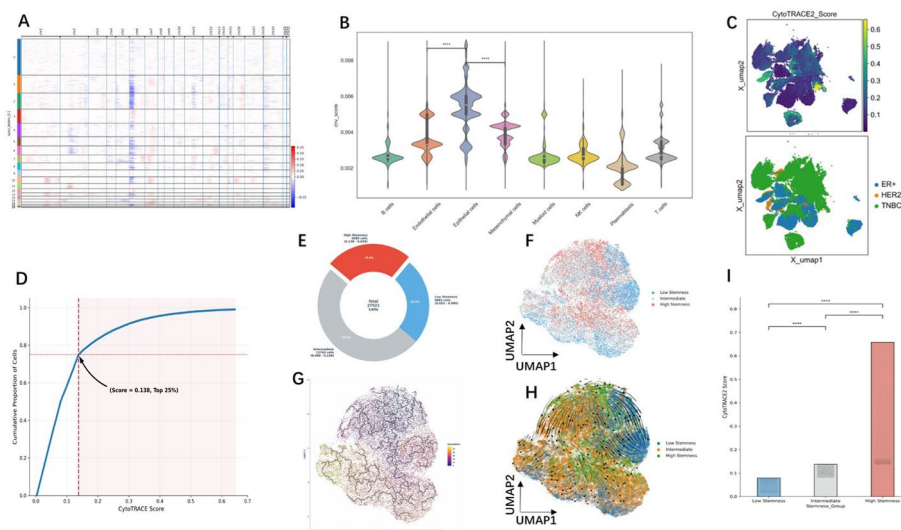
To dissect stemness heterogeneity in TNBC, we performed single-cell RNA sequencing analysis on 31 primary tumor samples, including 15 TNBC and 16 non-TNBC cases. Unsupervised clustering and dimensionality reduction were conducted using the Scanpy pipeline, revealing transcriptionally distinct cell populations (Fig. 1B). Uniform Manifold Approximation and Projection (UMAP) visualization clearly delineated the boundaries among major cell types (Fig. 1C). Based on clustering annotation using canonical marker genes, the following cell types were identified: epithelial cells, natural killer (NK) cells, T cells, B cells, myeloid cells, plasmablasts, mesenchymal cells, and endothelial cells (Fig. 1D). Dot plot analysis confirmed the unique marker gene expression patterns of each cell population (Fig. 1E). Heatmap visualization of differentially expressed genes across clusters further validated the accuracy of cell type annotation (Supplementary Fig. 1D). Additional analyses revealed differences in cell type distributions between TNBC and non-TNBC samples (Supplementary Fig. 1E), as well as variability across individual samples (Supplementary Fig. 1F), further underscoring distinct tumor micro-environmental features among breast cancer subtypes.



### 2.3 Identification and characterization of high-stemness TNBC cells

In breast cancer, a prototypical epithelial-derived malignancy, tumor initiation and progression primarily originate from malignant transcriptionally reprogrammed epithelial cells. Accordingly, we performed copy number variation (CNV) analysis on the identified cell subpopulations (Fig. 2A), revealing that epithelial cells exhibited significantly higher CNV scores than other cell types (Fig. 2B), indicative of widespread and pronounced chromosomal structural alterations. To ensure the precision of subsequent analyses, a threshold was set whereby cells with CNV scores greater than 0.3 were defined as malignant epithelial cells and designated as the core population for investigating stemness heterogeneity. To evaluate stemness, we applied the CytoTRACE algorithm, which quantifies stemness based on transcriptional diversity. As shown in Fig. 2C, high-stemness epithelial cells were predominantly concentrated in TNBC samples. To delineate the boundaries of cells with distinct stemness levels, we plotted an elbow curve based on the distribution trend of CytoTRACE scores. By identifying the inflection point at the 25% mark of the score curve (Fig. 2D), we systematically classified tumor epithelial cells into high-stemness, transitional, and low-stemness groups, corresponding to the top 25%, middle 50%, and bottom 25% of the stemness score range, respectively (Fig. 2E). We then extracted the tumor epithelial cells from the single-cell TNBC samples. UMAP visualization revealed clear spatial segregation among these groups (Fig. 2F), with high-stemness cells exhibiting significantly higher scores than low-stemness cells (Fig. 2I).

To further delineate the continuous distribution of stemness states and developmental potential within TNBC, we integrated the deep learning-based scTour algorithm with the classical Monocle3 trajectory inference method. Low-stemness cells (Low-Stem) were defined as the relative differentiation endpoint, and dynamic progression of malignant epithelial cells along a developmental timeline was inferred computationally. Monocle3 trajectory analysis revealed a clear branching structure of malignant epithelial



**Fig. 2** Identification and characterization of high-stemness TNBC cells. **A** Heatmap of CNV signals across TNBC cell clusters. **B–C** CNV scores across cell clusters and tumor types ( $n=44322$  cells). **D** Determination of stemness subgroup thresholds via elbow plot. **E** Pie chart showing subgroup classification based on CytoTRACE scores. A total of 27,521 epithelial cells were included. **F** UMAP plot colored by stemness groups. **G** Low-dimensional representation of stemness cells constructed using the Monocle 3 method. **H** Differentiation trajectory of stemness cells inferred using scTour. **I** Bar plot showing CytoTRACE scores across stemness groups. Statistical significance was assessed using the Wilcoxon signed-rank test. (\* $p<0.05$ , \*\* $p<0.01$ , \*\*\* $p<0.001$ , \*\*\*\* $p<0.0001$ )

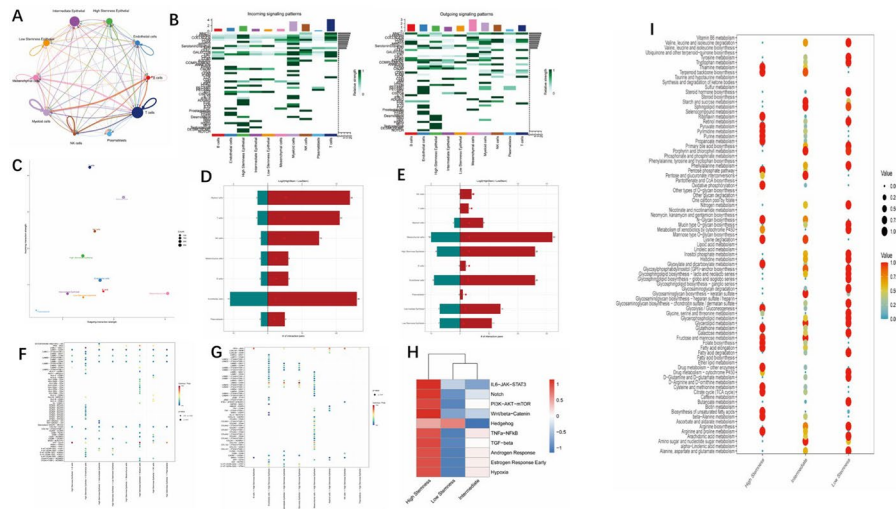
cells (Fig. 2G). Cells positioned at the trajectory origin corresponded closely to the previously identified HighStem subpopulation, while cells gradually transitioned toward the more differentiated LowStem subpopulation as pseudotime increased. Concurrently, vector field analysis by scTour provided additional evidence of developmental dynamics (Fig. 2H), with vectors consistently pointing from HighStem to LowStem, confirming the unidirectionality and temporal sequence of this differentiation trajectory.

**2.4 HighStem TNBC cells exhibit pronounced heterogeneity and complex signaling networks**

To further dissect the ligand-receptor interaction logic of HighStem TNBC cells within the tumor microenvironment, we systematically evaluated their communication patterns using the CellChat framework. The analysis revealed that HighStem TNBC cells establish a tightly connected bidirectional communication network with stromal and immune cells in the microenvironment, exhibiting both significant outgoing signal transmission and incoming signal integration functions (Fig. 3A–C). Compared with more differentiated cells, HighStem TNBC cells demonstrated higher activity in signaling pathways, reflecting their enhanced capacity to remodel and regulate the microenvironment.

Quantitative analysis further confirmed that, whether acting as ligand providers or receptor responders, HighStem TNBC cells exhibited significantly higher communication frequency and intensity compared with LowStem TNBC cells (Fig. 3D–E). Notably, among the key ligand-receptor interactions involving the HighStem subpopulation, collagen (COLs) and laminin (LAMs)-related pathways were particularly prominent. These cells mediated adhesion to the extracellular matrix by secreting abundant matrix components and binding to receptors such as CD44, and induced tumor cell dedifferentiation through integrin signaling (Fig. 3F–G).

AUCell pathway scoring further revealed the coordinated activation of key stemness-regulating signals in HighStem cells. The results showed that the HighStem

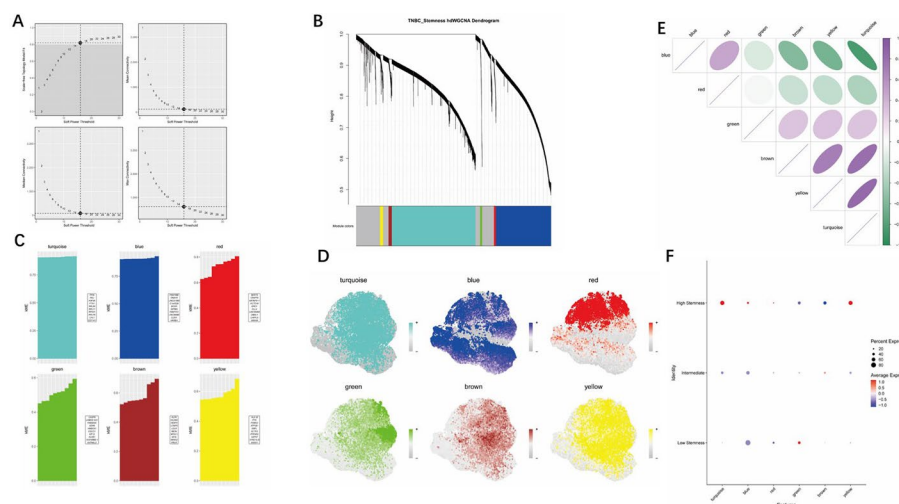


**Fig. 3** HighStem TNBC cells exhibit high heterogeneity and complex signaling communication networks. **A–B** Interaction network strength and number between cells. **C** Interaction intensity of signals sent and received by each cell type. **D–E** Logarithm of ligand-receptor interactions involving HighStem cells compared with LowStem cells. **F–G** Key ligand-receptor pairs enriched in HighStem cells. **H** AUCell pathway activity analysis of tumor stemness cells (The legend represents the Spearman correlation coefficient). **I** Metabolic activity analysis of tumor stemness cells

subpopulation exhibited significant enrichment in classical stemness signaling pathways, including JAK-STAT3, Notch, PI3K-AKT, WNT, and Hedgehog (Fig. 3H). Notably, this molecular profile is highly consistent with previously reported mechanisms for maintaining breast cancer stem cells [14]. Moreover, HighStem cells demonstrated widespread metabolic enrichment (Fig. 3I). This subpopulation upregulated key synthetic metabolic pathways, including glutathione, fatty acid, and steroid biosynthesis, while glycosaminoglycan synthesis was also significantly increased. Such extensive metabolic activity is closely associated with their highly malignant behavior, reflecting molecular-level compensation by the high-stemness subpopulation to sustain rapid proliferation and migration capabilities.

## 2.5 hdWGCNA identifies co-expression modules in HighStem cells

To analyze gene co-expression features associated with high stemness cells, we performed a systematic analysis using the hdWGCNA method. Based on scale-free topology characteristics and network connectivity evaluation, a soft-thresholding power of 16 was determined (Fig. 4A). Hierarchical clustering identified six co-expression modules, which were color-coded for distinction (Fig. 4B). Module eigengene (ME) analysis revealed the key genes within each module (Fig. 4C). Projecting module expression onto UMAP space validated the specific distribution of each module within the HighStem cell population (Fig. 4D). Further module correlation analysis showed strong co-expression associations among the turquoise, brown, and yellow modules (Fig. 4E). Notably, the turquoise and brown modules exhibited significant expression enrichment and higher cell proportions in high-stemness cells (Fig. 4F). These two modules collectively contain 951 genes, suggesting they may form the core transcriptional regulatory network of the high-stemness phenotype.

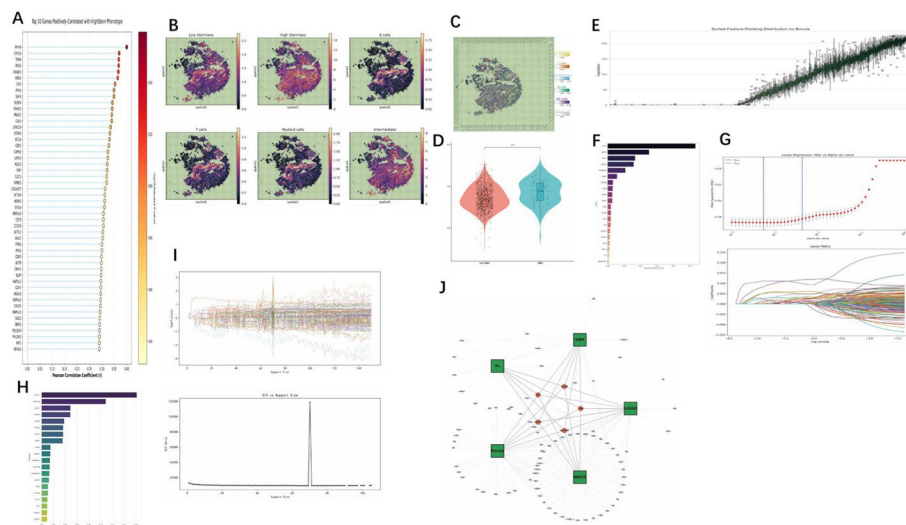


**Fig. 4** hdWGCNA identifies co-expression modules in HighStem cells. **A** Selection of soft-thresholding power based on scale-free topology and connectivity. **B** Dendrogram showing six gene co-expression modules. **C** Bar plot of module eigengenes, highlighting the most contributive genes. **D** UMAP visualization of module expression in TNBC epithelial cells. **E** Correlation matrix among modules. **F** Enrichment of yellow and turquoise modules in HighStem cells

## 2.6 Selection of signature genes in HighStem cells

To obtain a robust HighStem gene set, we first selected two phenotype-associated modules (turquoise and brown) from 6880 HighStem cells, then performed Pearson correlation analysis between genes and phenotype, resulting in 399 highly correlated genes, collectively defined as the HighStem activity gene set (Fig. 5A). Spatial transcriptomics indicated that, compared with LowStem cells, the tissue environment enriched with HighStem activity showed relatively lower immune cell abundance, potentially linked to immune evasion (Fig. 5B–C). RNA-seq data from the TCGA cohort revealed that the HighStem activity gene set was significantly elevated in TNBC compared with other breast cancer tissues (Fig. 5D).

In machine learning-based feature selection, the Boruta algorithm assessed feature importance by constructing shadow features, preliminarily eliminating noise (Fig. 5E). Gradient Boosting Machine (GBM) captured nonlinear relationships among genes in the feature space using iterative residual optimization (Fig. 5F). To address multicollinearity and overfitting, LASSO regression applied an L1 penalty to shrink coefficients of non-contributive genes to zero, maintaining model sparsity while retaining the most discriminative variables (Fig. 5G). Random Forest (RF) validated the linear discriminative ability of genes from a probabilistic perspective (Fig. 5H). Additionally, the ABESS algorithm searched the full-subset space for the combination minimizing the loss function, ensuring reliability and global optimality of the feature subset (Fig. 5I). Integrating results from the five algorithms, we ultimately identified five consensus core genes: CALD1, ANP32B, FIS1, CD82, and APLP2 (Fig. 5J, Supplementary Table 3).



**Fig. 5** Selection of HighStem cell signature genes. **A** Top 50 genes in the yellow and turquoise modules positively correlated with the HighStem phenotype. **B–C** Spatial transcriptomics showing the spatial distribution of cells with different stemness levels. The spatial transcriptomics TNBC samples consisted of Visium data (GSE210616,  $n = 1$ ). **D** Differences in stemness gene ( $n = 399$ ) scores across different tumor types. **E** Feature importance ranking by the Boruta algorithm. **F** Feature importance ranking by the Gradient Boosting Machine (GBM) algorithm ( $n = 20$ ). **G** Feature selection performed using LASSO regression. **H** Feature ranking analyzed by Random Forest ( $n = 20$ ). **I** Feature selection performed using the ABESS algorithm. **J** Interaction network of feature genes from five machine learning algorithms, highlighting the five core hub genes. Statistical significance was assessed using the Wilcoxon signed-rank test. (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ )

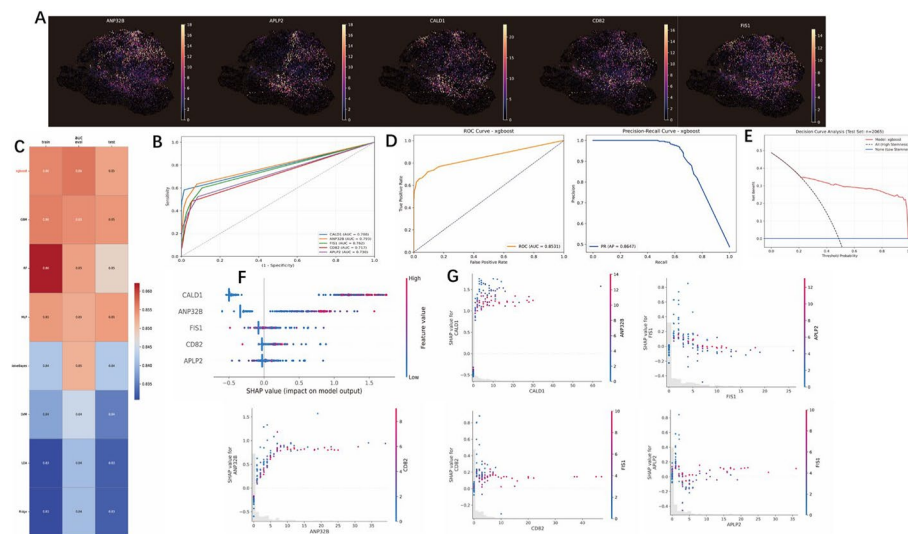
## 2.7 Validation of hub genes and construction of the machine learning model

To explore the spatial distribution of the core genes, we visualized the expression density across all TNBC cells. The results showed that CALD1, ANP32B, FIS1, CD82, and APLP2 exhibited higher signal intensity in specific cell populations (Fig. 6A). Additionally, we evaluated the predictive ability of each gene at the single-cell level for distinguishing malignant from benign cells. ROC curve analysis indicated that all five genes had moderate to good discriminatory power, with CALD1 (AUC = 0.788) and ANP32B (AUC = 0.793) showing the highest precision (Fig. 6B).

Subsequently, the hub genes were integrated into multiple machine learning frameworks, implementing a comprehensive approach involving eight different algorithms, including Support Vector Machine (SVM), Naive Bayes, Multi-Layer Perceptron (MLP), and XGBoost. Stemness cells were divided into training (70%), test (15%), and validation (15%) sets, and baseline analyses demonstrated that XGBoost exhibited stable predictive performance across both the test and validation sets (Fig. 6C–D). Decision curve analysis further confirmed the clinical utility of the XGBoost model, showing higher net benefit across a wide range of decision thresholds (Fig. 6E). SHAP analysis identified CALD1, ANP32B, FIS1, CD82, and APLP2 as the main contributors to the model output (Fig. 6F). Scatter plots revealed interactions between features; for example, the CALD1-ANP32B pair showed variable effects on predicted outcomes. CALD1 acted as a key positive predictor, with its low expression having a significant impact on the model, and its effect was modulated by ANP32B levels, being maximal when ANP32B expression was low (Fig. 6G).

## 2.8 Validation of hub gene expression

To validate the high-stemness hub genes, we evaluated their expression and clinical relevance using TCGA bulk RNA-seq data. The five genes *CALD1*, *ANP32B*, *FIS1*, *CD82*, and *APLP2* showed upregulated AUCell scores in TNBC tissues compared to other



**Fig. 6** Validation of hub genes and construction of the machine learning model. **A** Expression density in tumor epithelial cells. **B** Single-cell classification performance of each gene. **C** Baseline AUCs of eight machine learning models. **D** Precision-recall curve of XGBoost. **E** Decision curve analysis of the final classifier. **F** SHAP plot showing the contribution of each gene. **G** SHAP dependence plot illustrating the relationship between hub gene expression levels and their SHAP values

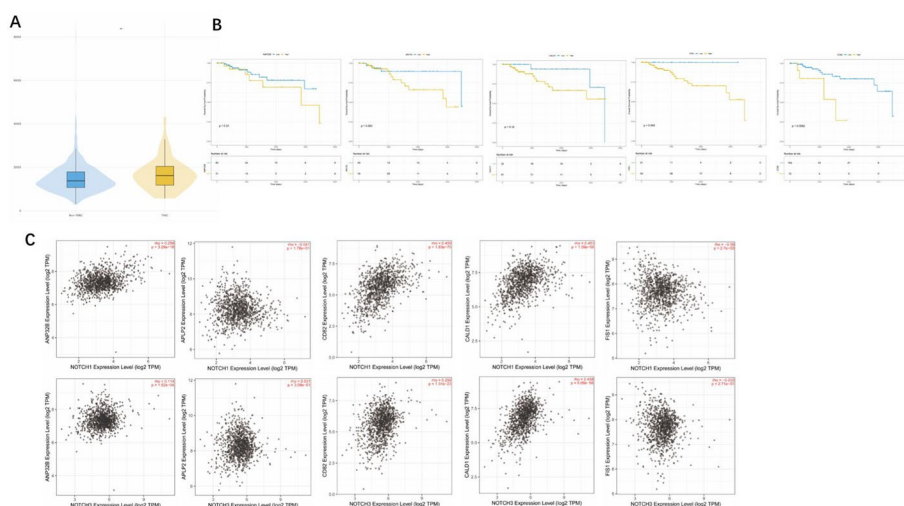
tumor types (Fig. 7A). We then explored the prognostic value of these genes. Kaplan–Meier survival analysis indicated that high expression of *CD82* ( $P=0.0082$ ) and *FIS1* ( $P=0.043$ ) was significantly associated with poorer overall survival in TNBC patients (Fig. 7B). *APLP2* showed borderline significance ( $P=0.083$ ), whereas *ANP32B* and *CALD1* did not reach statistical significance ( $P=0.33$  and  $P=0.12$ , respectively). These results highlight the potential of these five hub genes as diagnostic and prognostic biomarkers for high-stemness TNBC cells.

## 2.9 Hub genes exhibit expression trends similar to stemness pathway markers

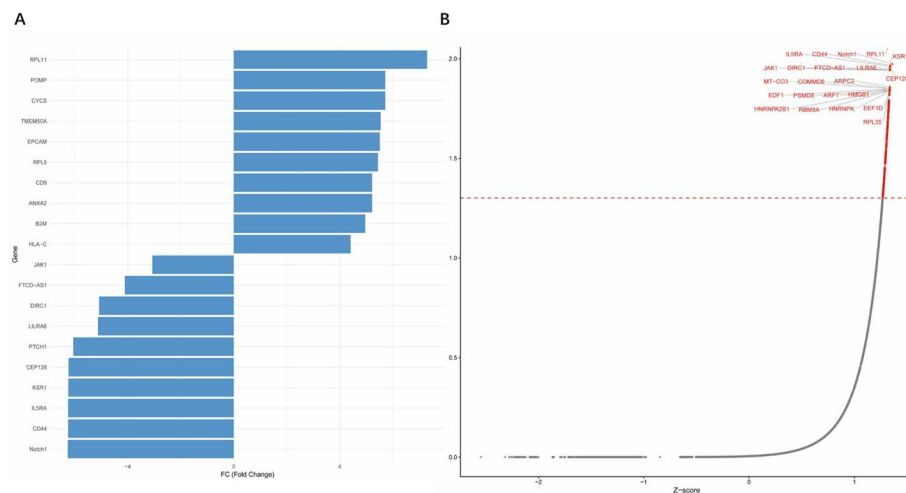
To further validate the correlation among stemness hub genes (*CALD1*, *ANP32B*, *FIS1*, *CD82*, and *APLP2*), we first analyzed their associations with stemness-enriched pathway markers, particularly the Notch signaling pathway, using the TIMER 2.0 platform. The analysis revealed that *ANP32B*, *CD82*, *CALD1*, and *APLP2* were positively correlated with the expression levels of Notch1 or Notch3. In contrast, *FIS1* showed weaker correlation with Notch1 or Notch3 and exhibited a trend toward negative association (Fig. 7C).

We also investigated the overall functional effects of these hub genes in TNBC stem cells. Using the scTenifoldKnk R package, we performed simulated knockdown of the hub genes and evaluated changes in high-stemness feature gene expression. The results showed a significant downregulation of high-stemness feature genes (Fig. 8A–B). The most notable changes included the downregulation of core stemness transcription factors and the receptor NOTCH1. The NOTCH pathway, by repressing differentiation genes and promoting transcription of core stemness factors, is recognized in TNBC as essential for maintaining cancer stem cell (CSC) self-renewal and an undifferentiated state. The downregulation of NOTCH1 suggests that these hub genes may act as upstream regulators of the NOTCH signaling pathway in TNBC cells.

Additionally, CD44, a surface marker of breast cancer stem cells (BCSCs), was reduced, directly indicating loss of stemness characteristics. The concurrent downregulation of



**Fig. 7** Clinical validation of hub genes based on TCGA and expression trends of stemness pathway markers. **A** Differences in hub gene scores across various tumor types. **B** Kaplan–Meier survival curves for each hub gene. **C** Correlation analysis between hub genes and the NOTCH pathway using the TIMER 2.0 platform. Statistical significance: survival analyses were performed using the Log-rank test, and gene correlation analyses were conducted using the Wilcoxon rank-sum test. (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ )



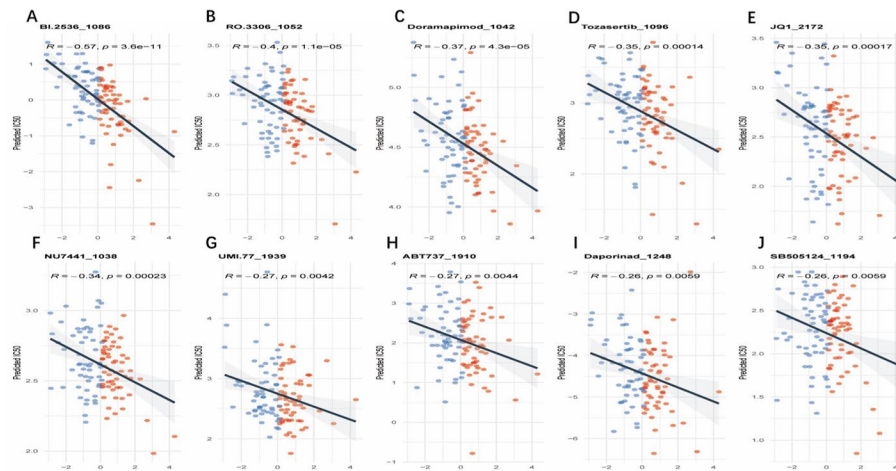
**Fig. 8** Virtual gene knockout in the single-cell cohort ( $n = 27521$  cells). **A** Upregulated and downregulated genes following knockout of hub genes. **B** Z-scores of the upregulated and downregulated genes. Statistical significance was assessed using the Wilcoxon signed-rank test

JAK1, KSR1, and PTCH1 suggests that knockdown of the hub genes disrupts key pathways (JAK-STAT3 and Hedgehog) that maintain self-renewal. Corresponding to the suppression of stemness genes, epithelial markers EPCAM and exosome-associated proteins CD9 and ANXA2 were upregulated, indicating a mesenchymal-to-epithelial transition (MET). The increased expression of major histocompatibility complex (MHC) molecules B2M and HLA-C implies that loss of hub genes may enhance the immunogenicity of TNBC cells, reversing the immune evasion characteristic of stemness cells.

## 2.10 Drug sensitivity in the TCGA-TNBC dataset

To explore mRNA-based therapeutic strategies for patients with high hub gene expression, the training set was constructed using drug sensitivity data extracted from the Genomics of Drug Sensitivity in Cancer (GDSC) database to predict sensitivity in the TCGA-TNBC dataset for high hub gene expression. Subsequently, the Wilcoxon rank-sum test was used to analyze differences in anticancer drug susceptibility between high and low hub gene expression groups in the TCGA-TNBC dataset. Significant differences were observed between the high- and low-score groups, and the top ten drugs with higher sensitivity in the high-score group were retained: BI.2536\_1086, RO.3306\_1052, Doramapimod\_1042, Tozasertib\_1096, JQ1\_2172, NU7441\_1038, UMI.77\_1939, ABT737\_1910, Daporinad\_1248, and SB505124\_1194 (Fig. 9A–J).

Among them, BI.2536 primarily functions as a highly selective inhibitor of Polo-like kinase 1 (PLK1) [15], a core regulator of mitotic initiation and progression. By disrupting spindle assembly and chromosome segregation, BI.2536 induces mitotic arrest and triggers apoptosis, particularly in cancer stem cells that rely on PLK1 for asymmetric division and genome stability maintenance. Notably, our computational predictions are consistent with previous experimental studies demonstrating that PLK1 is a key regulator of stemness in triple-negative breast cancer (TNBC), and that its inhibitor BI.2536 effectively depletes CD44<sup>+</sup>/CD24<sup>-</sup> tumor-initiating cells [16]. JQ1 exerts antitumor effects by competitively binding to the bromodomains of BET family proteins, displacing them from acetylated chromatin [17]. This action leads to transcriptional silencing of key oncogenic drivers and stemness-related factors, effectively disrupting the



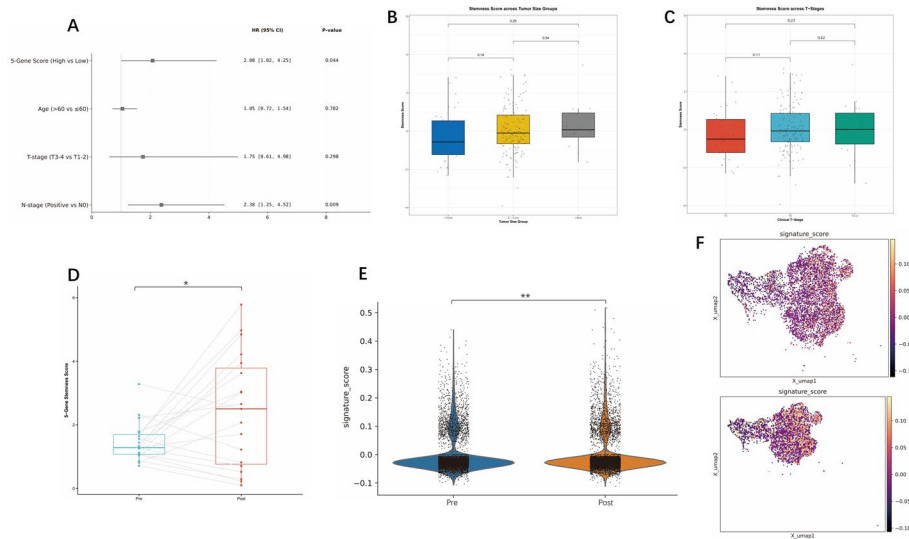
**Fig. 9** Drug sensitivity analysis. **A–J** Scatter and line plots showing IC50 values in relation to hub gene scores, displays the correlation coefficient **R**. Statistical significance was assessed using the Wilcoxon signed-rank test

transcriptional program that maintains TNBC cells in an undifferentiated state. The sensitivity of the high-stemness subpopulation to BI.2536 and JQ1 indicates that targeting mitotic dynamics and epigenetic remodeling may provide an improved therapeutic strategy for high-stemness cells in TNBC.

### 2.11 Clinical independent prognostic value and chemotherapy response characteristics of the stemness core gene score.

To evaluate the stability of the predictive performance of the stemness core genes and determine whether it is independent of conventional clinicopathological factors, we constructed a multivariate Cox proportional hazards model using the TCGA-TNBC cohort (Fig. 10A). After systematically adjusting for potential confounding factors, including age, pathological T stage, and N stage, the results showed that the stemness core genes remained an independent predictor of overall survival (OS) in TNBC patients (HR = 2.08,  $P=0.044$ ). Subsequently, we analyzed the association between the stemness score and clinical phenotypes using the GSE76275 [18] dataset. The results demonstrated a positive correlation between stemness core gene expression and both increasing tumor diameter (Fig. 10B) and advancing clinical T stage (Fig. 10C). Although statistical significance was not achieved with the current sample size, this increasing trend accompanying greater malignancy suggests a potential role of highly stem-like cell subpopulations in tumor expansion and local invasion.

In view of the high recurrence risk characteristic of TNBC, we further investigated the remodeling effects of chemotherapy-induced pressure on the stemness landscape. Analysis of paired samples from the GSE260693 dataset revealed that residual tumor tissues following adjuvant chemotherapy exhibited significantly higher stemness scores compared with pre-treatment samples (Fig. 10D). This phenomenon was further validated at the single-cell transcriptomic level, where post-chemotherapy samples showed a marked upregulation of stemness-related gene expression in single-cell profiles (Fig. 10E–F). These findings are consistent with our earlier drug sensitivity analysis, which indicated that the high-score group exhibited relatively lower sensitivity to conventional chemotherapeutic agents, such as paclitaxel. These results suggest that conventional chemotherapy regimens may be insufficient to effectively eliminate cancer cells with high



**Fig. 10** Clinical independent prognostic value and chemotherapy response characteristics of the stemness core gene score. **A** Forest plot of multivariate Cox regression analysis in the TCGA-TNBC cohort ( $n=90$ ). **B–C** Correlation between the stemness genes and clinical characteristics. (Data include GSE76275,  $n=186$ ). **D** Comparison of stemness scores between paired pre-chemotherapy and post-chemotherapy samples (bulk RNA-sequencing dataset GSE260693,  $n=31$ ). **E–F** Effects of chemotherapy on stemness gene expression at the single-cell transcriptomic level (GSE169246,  $n=5$ ). On the UMAP plot, the upper portion represents pre-treatment, and the lower portion represents post-treatment. Statistical significance: multivariate Cox analyses were assessed using the Wald test, while all other comparisons were evaluated using the Wilcoxon signed-rank test. (\* $p < 0.05$ , \*\* $p < 0.01$ )

stemness characteristics and may promote their enrichment through selective pressure, highlighting the potential clinical translational value of targeted therapies against stemness core genes.

### 3 Discussion

Tumor cell stemness has recently attracted extensive attention in the scientific community. The enrichment of CSCs is considered a central factor driving intratumoral heterogeneity and therapeutic resistance. Previous studies indicate that CSCs occupy the apex of the tumor cell hierarchy, possessing the capacity for self-renewal and reconstruction of primary tumor heterogeneity, with their biological behavior precisely regulated by both intrinsic transcriptional programs and microenvironmental signals [19]. Therefore, in-depth elucidation of these complex regulatory mechanisms has significant clinical implications.

Significant progress has been made in research on TNBC stemness markers. In addition to classical surface markers such as CD44, CD24, and ALDH1 [14], the widespread application of single-cell omics tools, including CytoTRACE [20] and SCENT [21], has facilitated the discovery of additional key regulatory factors. For example, Alessandro Sarcinella et al. [22] demonstrated that IL-3 can maintain TNBC stem-like properties through the miR-155-5p axis. Meng-Yuan Cai et al. [23] reported that IGF2BP3, as a core m6A reader, activates the  $\beta$ -catenin signaling pathway by stabilizing FZD1/7 transcripts, thereby sustaining TNBC stemness, suggesting that IGF2BP3 may serve as a biomarker or therapeutic target. Nevertheless, existing studies lack a systematic characterization of the TNBC stemness landscape and robust predictive models. To address this gap, our study first employed the CytoTRACE algorithm to identify high-stemness cell subpopulations and combined this with hdWGCNA to determine gene co-expression modules

highly associated with stemness traits. Subsequently, through the integrative screening of multiple machine learning algorithms, a predictive model with high discriminative performance was constructed. This work not only provides a novel strategy for identifying high-stemness TNBC cells but also offers a reference for clinical prognosis evaluation and targeted stemness therapies.

Our study identified five hub genes (*CALDI*, *ANP32B*, *FIS1*, *CD82*, and *APLP2*). *CALDI* is a canonical cytoskeletal regulatory protein that interacts with actin and calmodulin to mediate cell contraction, motility, and morphological remodeling [24–26]. Previous studies have demonstrated that *CALDI* plays a significant role in promoting epithelial–mesenchymal transition (EMT) during the malignant progression of various tumors. EMT enables polarized epithelial cells to acquire a mesenchymal phenotype, thereby endowing tumor cells with enhanced migratory capacity and differentiation plasticity, which represents a key prerequisite for the acquisition of stemness [27, 28]. Notably, this mechanism is highly consistent with the stemness-maintaining pathways, including Wnt/ $\beta$ -catenin [29], Notch [30] and JAK-STAT3 [31], identified in our study through CellChat analysis, collectively forming a regulatory network that drives state transitions in TNBC cells.

*ANP32B* belongs to the acidic nuclear phosphoprotein 32 kDa (ANP32) family and structurally consists of an N-terminal leucine-rich repeat (LRR) domain and a C-terminal low-complexity acidic region (LCAR). As a multifunctional regulatory protein, ANP32B primarily mediates transcriptional regulation and cytoplasmic mRNA transport, serving as a critical node in tumor cell growth signaling networks [32]. Although ANP32B has shown context-dependent tumor-suppressive roles in certain cancers [33, 34], recent studies have demonstrated its markedly elevated expression in breast cancer, with expression levels increasing alongside histopathological grading [35]. High ANP32B expression can sustain and enhance the stability of core nodes in the PI3K/AKT signaling pathway, and its phosphorylation activation further triggers downstream pro-proliferative effects. Such precise regulation of survival signaling pathways may represent one of the key mechanisms driving the maintenance of the high-stemness phenotype in TNBC.

*FIS1* is a core protein regulating mitochondrial dynamics, playing a pivotal role in mediating mitochondrial fission and metabolic reprogramming [36]. Cancer stem cells typically rely on specific mitochondrial metabolism to cope with microenvironmental stress. Studies have shown that FIS1 can activate the AMPK pathway to drive mitophagy, timely removing damaged mitochondria to maintain low intracellular reactive oxygen species (ROS) levels. This process effectively prevents premature differentiation and cell cycle arrest caused by GSK3 $\beta$  inactivation, thereby preserving CSC self-renewal capacity and stemness traits [37]. In this study, FIS1 was identified as a key hub gene, suggesting that the high-stemness TNBC cell subpopulation may utilize FIS1-mediated mitochondrial quality control to optimize metabolic homeostasis and evade therapy-induced cellular damage and apoptosis. As demonstrated by our analysis, the significant enrichment of CellChat communication networks and metabolic pathways highlights the complex interplay between intrinsic genetic programs and the external microenvironment. These metabolites, together with secreted factors, are essential for niche maintenance. For example, a recent review on ZG16B (Chen et al.) [38] emphasized how such secreted regulatory factors drive tumor progression by modulating both the tumor and immune

microenvironments. Similar to the regulatory pattern of ZG16B, metabolic intermediates produced by HighStem cells (such as glutathione and lipids) may function as signaling molecules that trigger intracellular cascade reactions, thereby sustaining the continuous expression of core genes.

*CD82* encodes a protein that is a key component in maintaining the stability of cell membrane receptor complexes. Although *CD82* has long been considered a marker inhibiting tumor migration [39, 40], this appears paradoxical given its role in highly malignant TNBC. Studies have found that, in addition to restraining excessive cell motility, *CD82* can maintain cellular homeostasis by enhancing adhesion between cells and their microenvironment (niche) [41] and by suppressing the activation of aberrant differentiation signals [42]. This “low motility, high adhesion” feature precisely aligns with the biological requirements of CSCs to maintain an undifferentiated state and signaling quiescence [43, 44]. Therefore, the high expression of *CD82* in HighStem cells may contribute to TNBC stemness maintenance by cooperatively establishing this specialized physical barrier.

*APLP2* belongs to the amyloid precursor-like protein family, and its high expression can significantly enhance cancer cell proliferative capacity and invasive phenotype, correlating positively with the progression and metastatic risk of various malignancies [45]. Studies have found that in renal cell carcinoma, *APLP2* expression levels are significantly positively associated with higher clinical stage and distant lymph node metastasis [46]. In hepatocellular carcinoma, *APLP2* knockdown induces apoptosis and delays cell cycle progression [47]. Although direct reports of *APLP2* in breast cancer are currently limited, secreted products of the same family member amyloid precursor protein (APP), such as protease nexin-2, have been shown to be cleaved by  $\gamma$ -secretase, which is a potential therapeutic target in breast cancer [48]. Based on functional similarities within the family, *APLP2* may act as a key candidate driver in maintaining TNBC stemness and malignant progression, promoting CSC homeostasis through regulation of intracellular signaling, warranting further investigation.

The Notch signaling pathway is a core regulator of cell fate determination and the maintenance of tumor stemness. As a highly conserved signaling network, Notch triggers the release of its intracellular domain through ligand–receptor binding, thereby activating the expression of downstream stemness-associated transcription factors [49–51]. Our analysis indicates that *CALD1*, *ANP32B*, *CD82*, and *APLP2* are positively correlated with certain Notch pathway marker genes, suggesting that these genes may act as upstream regulators of Notch and maintain TNBC stemness by activating this pathway. Notably, *FIS1* shows little or even negative correlation with Notch signaling, likely because *FIS1* primarily mediates stemness maintenance through mitochondrial dynamics and metabolic reprogramming rather than direct transcriptional regulation. These core genes may indirectly influence cellular differentiation and proliferation via complex cross-talk between pathways. Furthermore, virtual gene knockout experiments confirm that suppression of these hub genes significantly reduces the activity of related stemness pathways, computationally validating their potential role as Notch signaling regulators. Drug sensitivity analysis reveals that the high-stemness score group is more sensitive to agents such as BI-2536, a PLK1 inhibitor, implying that these hub genes may participate in the regulation of drug response mechanisms.

#### 4 Limitations

Although this study provides new insights at the multi-omics level, there are still several notable limitations that should be interpreted with caution. This study primarily relied on retrospective databases such as TCGA and GEO, and heterogeneity in sample collection standards, sequencing platforms, and clinical annotations may introduce bias affecting the generalizability of the conclusions. However, no in vitro experiments were conducted to validate the biological functions of these genes in TNBC stemness, specifically how they affect the Notch signaling pathway and metabolic reprogramming in TNBC. Systematic experimental validation, such as gene knockdown experiments and Western blot assays, is urgently needed. Third, potential therapeutic agents identified through drug sensitivity analysis, including BL2536 and JQ1, were predicted solely based on cell line data, which cannot fully recapitulate the influence of the tumor microenvironment and spatial heterogeneity on drug efficacy; therefore, their effectiveness in high-stemness TNBC patients requires further evaluation through in vitro drug intervention experiments. Finally, although the stemness core gene risk model established in this study demonstrated robust performance across multiple validation cohorts, its reliability as a clinical decision-making tool still requires confirmation in prospective, large-scale, multicenter clinical trials.

#### 5 Conclusion

This study delineated the stemness architecture of TNBC through the integration of multi-omics data. We identified a high-stemness cellular subpopulation that drives tumor progression and established a robust predictive model based on five core hub genes—*CALD1*, *ANP32B*, *FIS1*, *CD82*, and *APLP2*. These genes play critical roles in maintaining the undifferentiated state of cancer stem cells, potentially through regulation of the Notch signaling pathway and metabolic reprogramming. Furthermore, our findings indicate that high stemness scores are associated with poor prognosis, immune evasion, and resistance to conventional chemotherapy. These findings provide new insights for subsequent mechanistic studies and therapeutic development.

#### 6 Materials and methods

##### 6.1 Data collection and processing

Bulk RNA-seq data were obtained from TCGA as well as GSE76275 [18] and GSE260693. Single-cell RNA sequencing datasets (GSE169246 [52], GSE176078 [6],) and spatial transcriptomics data (GSE210616 [53]) were retrieved from the GEO database. All single-cell analyses were performed using primary biopsy tumor samples; core gene selection was conducted on treatment-naïve samples, and clinical correlation analyses utilized paired pre- and post-chemotherapy samples. Cells with fewer than 150 expressed genes or with mitochondrial gene content exceeding 25% were removed, and genes detected in fewer than three cells were excluded. To eliminate doublet interference, scrublet was applied to each sample. Ultimately, 123,910 high-quality cells from 26 samples were retained for analysis. The data were processed using the Scanpy workflow, including normalization, principal component analysis (PCA), UMAP dimensionality reduction, and clustering (resolution set to 0.1). To determine the optimal integration method for our dataset, we benchmarked several widely used Python-based tools: BBKNN, Harmony, and the deep learning-based scVI. scVI performed best and was therefore used for dataset integration.

Cell types were annotated based on known markers, and annotation accuracy was assessed via differentially expressed genes (DEGs;  $\log_{2}FC > 1$ ,  $FDR < 0.05$ ) for each cell type. Spatial transcriptomics data were deconvoluted using cell2location with a negative binomial regression model applied to reference scRNA-seq data using default parameters to estimate cell subtype reference profiles. The reference matrix derived from the integrated dataset was used to estimate the spatial abundance of each cell type in each sample.

## 6.2 Copy number variation (CNV) analysis

To distinguish tumor epithelial cells from normal breast epithelial cells, we applied the inferCNV algorithm, using a sliding window of 250 genes to estimate chromosome-level copy number variation (CNV) signals at the single-cell level [54]. Prior to analysis, genes with an average expression below 0.1 were filtered out, and normalization and noise reduction were performed. Immune cells were used as a reference background to infer the CNV profiles, while epithelial cells were set as the observation target. Based on the inferred results, epithelial cells were successfully classified into non-tumor and tumor categories. Ultimately, a total of 44,322 tumor epithelial cells and 1196 normal epithelial cells were identified across all tumor samples.

## 6.3 Stemness scoring of tumor epithelial cells

To identify high-stemness cells, we applied the CytoTRACE algorithm [20] to the malignant epithelial cells filtered by inferCNV. This method assigns each cell a continuous stemness score (ranging from 0 to 1) based on the inverse correlation between transcriptional diversity and differentiation potential. According to the distribution of CytoTRACE scores, the tumor cell population was divided into three subgroups: the top 25% of cells were defined as the high-stemness group, cells ranked between 25 and 75% as the transitional group, and the bottom 25% as the low-stemness group.

## 6.4 hdWGCNA analysis

Given that CytoTRACE scores primarily provide a quantitative measure of cellular differentiation potential and cannot directly reveal underlying core driver genes, we employed the hdWGCNA algorithm to further dissect the transcriptomic regulatory features of high-stemness tumor cells. By constructing a high-dimensional weighted gene co-expression network, genes with similar expression patterns were clustered into distinct functional modules [55, 56]. Subsequently, module–trait association analysis was performed to identify co-expression modules significantly positively correlated with the high-stemness phenotype. Hub genes within these modules were identified based on intramodular connectivity, highlighting key regulatory genes that may drive stemness heterogeneity and tumor progression in TNBC epithelial cells.

## 6.5 Cellchat analysis

To infer intercellular interactions, we applied the R package CellChat [57]. Potential ligand–receptor interactions were predicted using the computeCommunProb function, and the results were visualized with netVisual\_chord\_gene or netVisual\_bubble functions.

### 6.6 Pseudotime analysis

To elucidate the differentiation trajectory and developmental potential of malignant epithelial cells in TNBC, we performed trajectory analysis using the R package Monocle 3 [58] and the Python package scTour [59]. TNBC tumor cells were re-integrated, and highly variable genes (HVGs) selected during integration were used to run scTour, with all analyses performed using default settings. The resulting vector field was projected onto the recalculated UMAP embedding. Simultaneously, Monocle 3 was used for trajectory analysis, designating the cluster with the lowest stemness score as the root and determining the trajectory endpoint toward low-stemness phenotypes using diffusion mapping.

### 6.7 Selection of high-stemness cell feature genes

To identify key marker genes of the HighStem subpopulation, we integrated five machine learning algorithms—Boruta, GBM, LASSO, ABESS, and Random Forest—leveraging their complementary strengths in feature space exploration and bias control. Random Forest [60] and its derivative Boruta [61] were used for preliminary genome-wide feature weighting and importance ranking. LASSO [62] regression applied a norm-based penalty to perform feature sparsification, effectively removing redundant signals. The ABESS [63] algorithm conducted optimal path searches across the full subset space to ensure global optimality of feature combinations while avoiding selection bias. Additionally, GBM [64], an additive model, captured complex nonlinear interactions among genes. The hub genes identified by these five methods were intersected and visualized using a node diagram.

### 6.8 Construction of a high-stemness predictive model using machine learning and SHAP interpretation

To evaluate the key cell identification performance of high-stemness genes, we tested eight machine learning algorithms using the Python package sklearn, including Support Vector Machine (SVM), Naive Bayes, Multilayer Perceptron (MLP), XGBoost, Random Forest, Gradient Boosting Machine (GBM), LDA, and Ridge. High-stemness and low-stemness cells were divided into a training set (70%), a test set (15%), and a validation set (15%), and models were optimized through manual hyperparameter tuning. XGBoost, which demonstrated stable performance, was ultimately selected as the optimal framework. To further quantify the impact of individual genes on HighStem predictions, SHAP (SHapley Additive exPlanations) attribution analysis [65] was applied. Rooted in game theory, SHAP calculates the marginal contribution of each feature across all possible feature combinations, assigning unified attribution values to complex model predictions. Compared with traditional feature importance metrics, SHAP accurately captures the positive or negative regulatory effects of gene expression fluctuations on model outputs, enabling in-depth single-cell-level analysis of the drivers of high-stemness phenotypes.

### 6.9 Computational simulation of high-stemness hub gene knockout

To further validate the key regulatory roles of hub genes in TNBC, the virtual knock-down tool scTenifoldKnk [66], based on single-cell RNA sequencing data, was used to

perform in silico knockdown of hub genes to infer their downstream and upstream functional impacts.

### 6.10 Drug sensitivity analysis

In this study, we utilized the GDSC2 database and the R package oncoPredict [67] to construct a ridge regression model aimed at translating pharmacological information from cell lines into predictions of therapeutic responses in clinical samples. By applying this model to the bulk RNA-seq data of the TCGA-TNBC cohort, we estimated the half-maximal inhibitory concentration (IC50) for each sample against small-molecule drugs, where lower scores indicate higher drug sensitivity. Candidate drugs showing a significant negative correlation with the hub gene scores were subsequently identified.

### 6.11 Statistical analysis

All computations and statistical analyses were performed in the R (v4.3.0) and Python (v3.9) programming environments. Differences in continuous variables between two groups were evaluated using the Wilcoxon rank-sum test, while comparisons among multiple groups were performed using the Kruskal–Wallis test. Expression correlations among hub genes and associations between gene expression and drug sensitivity were quantified using Pearson correlation coefficients. Survival analyses were conducted using Kaplan–Meier curves with group differences assessed by the log-rank test. Multivariate Cox analysis was performed using the Wald test. All statistical tests were two-sided, with  $P < 0.05$  considered statistically significant. Significance levels were defined as follows:  $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ , and  $****p < 0.0001$ . The association between continuous variables was assessed using Pearson correlation analysis.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s12672-026-04824-5>.

Supplementary Material 1 .Fig.1: Highlights the importance of stem-like cells in the comparison between TNBC and non-TNBC. A Volcano plot of differentially expressed genes (DEGs) between bulk non-TNBC and TNBC samples. B–C GO and KEGG enrichment analyses of DEGs. D Differentially expressed genes in each single-cell cluster. E Histogram showing the distribution of cells in non-TNBC versus TNBC at the single-cell level. F Histogram showing cell distribution across individual single-cell samples.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

## Acknowledgements

Acknowledgements We would like to express our heartfelt gratitude to the TCGA database and GEO database for providing valuable resources for data analysis.

## Author contributions

Author contributions statement Z.H. and W.S. contributed equally to this work and are co-first authors. Z.H. and W.S. jointly conceived and designed the study. Z.H. performed the data analysis, and W.S. drafted the original manuscript. C.L. provided project supervision and funding support. T.Y. and C.L. reviewed and revised the manuscript. All authors read and approved the final version of the manuscript.

## Funding

This study was supported by the Natural Science Foundation of Heilongjiang Province (grant number. PL2024H181).

## Data availability

Data availability The datasets used in this study were obtained from the BRCA project of the TCGA database (<https://portal.gdc.cancer.gov/cart>) and the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), including GSE169246, GSE176078, GSE210616, GSE76275, and GSE260693.

## Declarations

### Ethics approval and consent to participate

There are no ethical considerations applicable to our article.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 23 January 2026 / Accepted: 9 March 2026

Published online: 23 March 2026

## References

1. Xiong X, Zheng LW, Ding Y, Chen YF, Cai YW, Wang LP, et al. Breast cancer: pathogenesis and treatments. *Signal Transduct Target Ther.* 2025;10(1):49.
2. Xu X, Zhang M, Xu F, Jiang S. Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities. *Mol Cancer.* 2020;19(1):165.
3. Verma A, Singh A, Singh MP, Nengroo MA, Saini KK, Satrusal SR, et al. EZH2-H3K27me3 mediated KRT14 upregulation promotes TNBC peritoneal metastasis. *Nat Commun.* 2022;13(1):7344.
4. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol.* 2016;13(11):674–90.
5. Bai X, Ni J, Beretov J, Graham P, Li Y. Triple-negative breast cancer therapeutic resistance: where is the Achilles' heel? *Cancer Lett.* 2021;497:100–11.
6. 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.
7. Chu X, Tian W, Ning J, Xiao G, Zhou Y, Wang Z, et al. Cancer stem cells: advances in knowledge and implications for cancer therapy. *Signal Transduct Target Ther.* 2024;9(1):170.
8. Prager BC, Xie Q, Bao S, Rich JN. Cancer stem cells: the architects of the tumor ecosystem. *Cell Stem Cell.* 2019;24(1):41–53.
9. Zeng Z, Fu M, Hu Y, Wei Y, Wei X, Luo M. Regulation and signaling pathways in cancer stem cells: implications for targeted therapy for cancer. *Mol Cancer.* 2023;22(1):172.
10. Pan G, Liu Y, Shang L, Zhou F, Yang S. EMT-associated microRNAs and their roles in cancer stemness and drug resistance. *Cancer Commun (Lond).* 2021;41(3):199–217.
11. Salvadori G, Zanardi F, Iannelli F, Lobefaro R, Vernieri C, Longo VD. Fasting-mimicking diet blocks triple-negative breast cancer and cancer stem cell escape. *Cell Metab.* 2021;33(11):2247–2259.e2246.
12. Caliendo A, Camorani S, Ibarra LE, Pinto G, Agnello L, Albanese S, et al. A novel CD44-targeting aptamer recognizes chemoresistant mesenchymal stem-like TNBC cells and inhibits tumor growth. *Bioact Mater.* 2025;50:443–60.
13. Bu J, Zhang Y, Wu S, Li H, Sun L, Liu Y, et al. KK-LC-1 as a therapeutic target to eliminate ALDH(+) stem cells in triple negative breast cancer. *Nat Commun.* 2023;14(1):2602.
14. Hua Z, White J, Zhou J. Cancer stem cells in TNBC. *Semin Cancer Biol.* 2022;82:26–34.
15. Zhang J, Liu X, Hou P, Lv Y, Li G, Cao G, et al. BRCA1 orchestrates the response to BI-2536 and its combination with alisertib in MYC-driven small cell lung cancer. *Cell Death Dis.* 2024;15(7):551.
16. Hu K, Law JH, Fotovati A, Dunn SE. Small interfering RNA library screen identified polo-like kinase-1 (PLK1) as a potential therapeutic target for breast cancer that uniquely eliminates tumor-initiating cells. *Breast Cancer Res.* 2012;14(1):R22.
17. Wang L, Xu M, Kao CY, Tsai SY, Tsai MJ. Small molecule JQ1 promotes prostate cancer invasion via BET-independent inactivation of FOXA1. *J Clin Invest.* 2020;130(4):1782–92.
18. den Hollander P, Rawls K, Tsimelzon A, Shepherd J, Mazumdar A, Hill J, et al. Phosphatase PTP4A3 Promotes Triple-Negative Breast Cancer Growth and Predicts Poor Patient Survival. *Cancer Res.* 2016;76(7):1942–53.
19. Nassar D, Blanpain C. Cancer stem cells: basic concepts and therapeutic implications. *Annu Rev Pathol.* 2016;11:47–76.
20. Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, Berger MJ, et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science.* 2020;367(6476):405–11.
21. Teschendorff AE, Enver T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat Commun.* 2017;8:15599.
22. Sarcinella A, Kholia S, Femminò S, Cedrino M, Tapparo M, Wen X, et al. IL-3/STAT5/miR-155-5p axis supports stem-related pathway reprogramming in TNBC. *Breast Cancer Res.* 2025;27(1):195.
23. Cai MY, Yin P, Wang ZW, Huang YZ, Wang YX, Wu KF, et al. Dual regulation of FZD1/7 by IGF2BP3 enhances stem-like properties and carboplatin resistance in triple-negative breast cancer. *Cancer Lett.* 2025;632:217944.
24. Li Y, Yang Z, Xu M, Guan H, Wu Z, Li S. CALD1 inhibits invasion of human ovarian cancer cells by affecting cytoskeletal structure and the number of focal adhesion. *Transl Cancer Res.* 2025;14(2):1323–35.
25. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinform.* 2016;54(1):1–30.
26. Mayanagi T, Sobue K. Diversification of caldesmon-linked actin cytoskeleton in cell motility. *Cell Adh Migr.* 2011;5(2):150–9.
27. Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016. *Cell.* 2016;166(1):21–45.
28. Liu M, Zhang Y, Yang J, Zhan H, Zhou Z, Jiang Y, et al. Zinc-Dependent regulation of ZEB1 and YAP1 coactivation promotes epithelial-mesenchymal transition plasticity and metastasis in pancreatic cancer. *Gastroenterology.* 2021;160(5):1771–1783.e1771.
29. Song P, Gao Z, Bao Y, Chen L, Huang Y, Liu Y, et al. Wnt/ $\beta$ -catenin signaling pathway in carcinogenesis and cancer therapy. *J Hematol Oncol.* 2024;17(1):46.
30. Guo R, Han D, Song X, Gao Y, Li Z, Li X, et al. Context-dependent regulation of Notch signaling in glial development and tumorigenesis. *Sci Adv.* 2023;9(45):eadi2167.

31. Wang T, Fahrman JF, Lee H, Li YJ, Tripathi SC, Yue C, et al. JAK/STAT3-regulated fatty acid  $\beta$ -oxidation is critical for breast cancer stem cell self-renewal and chemoresistance. *Cell Metab.* 2018;27(1):136–150.e135.
32. Yang S, Zhu XN, Zhang HL, Yang Q, Wei YS, Zhu D, et al. ANP32B-mediated repression of p53 contributes to maintenance of normal and CML stem cells. *Blood.* 2021;138(24):2485–98.
33. Björck E, Ek S, Landgren O, Jerkeman M, Ehinger M, Björkholm M, et al. High expression of cyclin B1 predicts a favorable outcome in patients with follicular lymphoma. *Blood.* 2005;105(7):2908–15.
34. Yang Q, Liu HR, Yang S, Wei YS, Zhu XN, Zhi Z, et al. ANP32B suppresses B-cell acute lymphoblastic leukemia through activation of PU.1 in mice. *Cancer Sci.* 2023;114(7):2882–94.
35. Yang S, Zhou L, Reilly PT, Shen SM, He P, Zhu XN, et al. ANP32B deficiency impairs proliferation and suppresses tumor progression by regulating AKT phosphorylation. *Cell Death Dis.* 2016;7(2):e2082.
36. Suh J, Kim NK, Shim W, Lee SH, Kim HJ, Moon E, et al. Mitochondrial fragmentation and donut formation enhance mitochondrial secretion to promote osteogenesis. *Cell Metab.* 2023;35(2):345–360.e347.
37. Pei S, Minhajuddin M, Adane B, Khan N, Stevens BM, Mack SC, et al. AMPK/FIS1-mediated mitophagy is required for self-renewal of human AML stem cells. *Cell Stem Cell.* 2018;23(1):86–100.e106.
38. Chen XM, Liang YB, Zuo JX, Yang ZS, Zhang LY, Zhang XY, et al. ZG16B: A key regulator of tumor progression and immune microenvironment modulation in cancer (Review). *Int J Mol Med.* 2026. <https://doi.org/10.3892/ijmm.2026.5729>.
39. Viera M, Yip GWC, Shen HM, Baeg GH, Bay BH. Targeting CD82/KAI1 for Precision Therapeutics in Surmounting Metastatic Potential in Breast Cancer. *Cancers (Basel).* 2021;13(17):4486.
40. Reddi KK, Zhang W, Shahrabi-Farahani S, Anderson KM, Liu M, Kakhniashvili D, et al. Tetraspanin CD82 correlates with and may regulate S100A7 expression in oral cancer. *Int J Mol Sci.* 2024. <https://doi.org/10.3390/ijms25052659>.
41. Ji H, Chen L, Dai Y, Sun X, Li X, Wang Q, et al. Aberrant expression of CD133 and CD82 in patients with pediatric acute lymphoblastic leukemia and the clinical significance. *Oncol Lett.* 2017;14(5):5811–8.
42. Hur J, Choi JI, Lee H, Nham P, Kim TW, Chae CW, et al. CD82/KAI1 maintains the dormancy of long-term hematopoietic stem cells through interaction with DARC-expressing macrophages. *Cell Stem Cell.* 2016;18(4):508–21.
43. Nevi L, Di Matteo S, Carpino G, Zizzari IG, Samira S, Ambrosino V, et al. DCLK1, a putative stem cell marker in human cholangiocarcinoma. *Hepatology.* 2021;73(1):144–59.
44. Djirackor L, Kalirai H, Coupland SE, Petrovski G. CD166high uveal melanoma cells represent a subpopulation with enhanced migratory capacity. *Invest Ophthalmol Vis Sci.* 2019;60(7):2696–704.
45. Pandey P, Sliker B, Peters HL, Tuli A, Herskovitz J, Smits K, et al. Amyloid precursor protein and amyloid precursor-like protein 2 in cancer. *Oncotarget.* 2016;7(15):19430–44.
46. Gao L, Zhao H, Zhang D, Zhou C, Wang H, Ren C, et al. Role of APLP2 in the prognosis and clinicopathology of renal cell carcinoma. *Oncol Lett.* 2019;17(1):508–13.
47. Tao Z, Huang J, Li J. Comprehensive intratumoral heterogeneity landscaping of liver hepatocellular carcinoma and discerning of APLP2 in cancer progression. *Environ Toxicol.* 2024;39(2):612–25.
48. McHenry PR, Prosperi JR. Proteins found in the Triple-Negative Breast Cancer secretome and their therapeutic potential. *Int J Mol Sci.* 2023. <https://doi.org/10.3390/ijms24032100>.
49. Volk A. NOTCH-ing up nucleotide metabolism in T-cell acute lymphoblastic leukemia. *Commun Biol.* 2021;4(1):809.
50. Ferrandino F, Grazioli P, Bellavia D, Campese AF, Screpanti I, Felli MP. Notch and NF- $\kappa$ B: Coach and Players of Regulatory T-Cell Response in Cancer. *Front Immunol.* 2018;9:2165.
51. Tsokos GC. Notch notches lupus. *Kidney Int.* 2020;97(2):251–3.
52. Zhang Y, Chen H, Mo H, Hu X, Gao R, Zhao Y, et al. Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. *Cancer Cell.* 2021;39(12):1578–1593.e1578.
53. Bassiouni R, Idowu MO, Gibbs LD, Robila V, Grizzard PJ, Webb MG, et al. Spatial Transcriptomic Analysis of a Diverse Patient Cohort Reveals a Conserved Architecture in Triple-Negative Breast Cancer. *Cancer Res.* 2023;83(1):34–48.
54. Oketch DJA, Giulietti M, Piva F. A comparison of tools that identify tumor cells by inferring Copy Number Variations from single-cell experiments in Pancreatic Ductal Adenocarcinoma. *Biomedicines.* 2024. <https://doi.org/10.3390/biomedicines12081759>.
55. Dai L, Fan G, Xie T, Li L, Tang L, Chen H, et al. Single-cell and spatial transcriptomics reveal a high glycolysis B cell and tumor-associated macrophages cluster correlated with poor prognosis and exhausted immune microenvironment in diffuse large B-cell lymphoma. *Biomark Res.* 2024;12(1):58.
56. Li Z, Qiao W, Yu S, Fan B, Yang M, Wu M, et al. Integrating computational pathology and multi-transcriptomics to characterize lung adenocarcinoma heterogeneity and prognostic modeling. *Int J Surg.* 2025;111(8):5162–81.
57. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using Cell Chat. *Nat Commun.* 2021;12(1):1088.
58. 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.
59. Li Q. scTour: a deep learning architecture for robust inference and accurate prediction of cellular dynamics. *Genome Biol.* 2023;24(1):149.
60. Tan Y, Cahan P. SingleCellNet: a computational tool to classify single cell RNA-Seq data across platforms and across species. *Cell Syst.* 2019;9(2):207–213.e202.
61. Yan F, Chen X, Quan X, Wang L, Wei X, Zhu J. Association between the stress hyperglycemia ratio and 28-day all-cause mortality in critically ill patients with sepsis: a retrospective cohort study and predictive model establishment based on machine learning. *Cardiovasc Diabetol.* 2024;23(1):163.
62. Mullah MAS, Hanley JA, Benedetti A. LASSO type penalized spline regression for binary data. *BMC Med Res Methodol.* 2021;21(1):83.
63. Li HF, Wang JT, Zhao Q, Zhang YM. BLUPmrMLM: A fast mrMLM algorithm in Genome-wide Association Studies. *Genomics Proteomics Bioinform.* 2024. <https://doi.org/10.1093/gpbjnl/qzae020>.
64. Lai H, Huang H, Keshavjee K, Guergachi A, Gao X. Predictive models for diabetes mellitus using machine learning techniques. *BMC Endocr Disord.* 2019;19(1):101.
65. Qi X, Wang S, Fang C, Jia J, Lin L, Yuan T. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. *Redox Biol.* 2025;79:103470.

66. Osorio D, Zhong Y, Li G, Xu Q, Yang Y, Tian Y, et al. scTenifoldKnk: an efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation. *Patterns (N Y)*. 2022;3(3):100434.
67. 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. <https://doi.org/10.1093/bib/bbab260>.

### **Publisher's Note**

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