



## OPEN ACCESS

## EDITED BY

Benjamin Kidder,  
Wayne State University, United States

## REVIEWED BY

Francesco De Francesco,  
Independent Researcher, Ancona, Italy  
Vinagolu K. Rajasekhar,  
Memorial Sloan Kettering Cancer Center,  
United States  
Subhadeep Das,  
Purdue University, United States

## \*CORRESPONDENCE

Ting Ye

✉ yeting1103@163.com

<sup>†</sup>These authors have contributed  
equally to this work

RECEIVED 09 April 2025

ACCEPTED 22 July 2025

PUBLISHED 13 August 2025

## CITATION

Liu L, Pei Q, Qadir J, Chen Y, Li J,  
Luo Y, Xian J, Du R and Ye T (2025)  
Application of artificial intelligence-based  
stemness index in cancer.  
*Front. Oncol.* 15:1608712.  
doi: 10.3389/fonc.2025.1608712

## COPYRIGHT

© 2025 Liu, Pei, Qadir, Chen, Li, Luo, Xian, Du  
and Ye. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Application of artificial intelligence-based stemness index in cancer

Liyuan Liu<sup>1†</sup>, Qin Pei<sup>1†</sup>, Javeria Qadir<sup>2†</sup>, Yiyu Chen<sup>1</sup>, Jingyuan Li<sup>1</sup>,  
Yanan Luo<sup>1</sup>, Jiawen Xian<sup>1</sup>, Rongrong Du<sup>1</sup> and Ting Ye<sup>1\*</sup>

<sup>1</sup>Department of Laboratory Medicine, the Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan, China, <sup>2</sup>Department of Biosciences, COMSATS University Islamabad, Islamabad, Pakistan

Cancer stem cells (CSCs) exhibit self-renewal and multidirectional differentiation capacities. The stemness of CSCs is the fundamental cause of tumor progression and treatment resistance. The stemness index, evaluating the number and activity of CSCs, is a crucial indicator predicting various aspects of tumor behavior such as growth, metastasis, and prognosis. With the advancements in artificial intelligence (AI), particularly in data analysis and machine learning, the identification and understanding of CSCs' stemness characteristics have improved. The AI-based analysis allows for processing vast datasets and recognizing patterns that assist in comprehending the role of CSCs in cancer development. The utilization of AI to analyze and compute the stemness index holds significant clinical relevance in tumor diagnosis and treatment. This approach provides more precise and personalized information, potentially influencing treatment strategies. Therefore, tailoring treatments specifically targeting CSCs is highly imperative and may enhance therapeutic efficacy and outcomes in cancer patients.

## KEYWORDS

cancer stem cells, artificial intelligence, MDNAs, mRNAs, stemness

## 1 Introduction

Cancer stem cells (CSCs) refer to the existence of a small subset of cells within a tumor that exhibit stem cell-like properties. These CSCs possess the ability to self-renew and differentiate into various cell types within the tumor, contributing to tumor growth and heterogeneity. This property of self-renewal allows CSCs to sustain their population within the tumor, acting as a source of cells that propagate and regenerate the tumor mass. Furthermore, CSCs have the capacity to prompt multidirectional differentiation, thus, generating different cell types within the tumor, often referred to as cancer-initiating cells (1–3). CSCs exhibit characteristics encompassing self-renewal, multidirectional differentiation, multi-drug resistance and radiation resistance, and signaling pathways that are common to both tumor and normal stem cells (4–6), hence, rendering them highly drug-resistant, evading conventional treatment, and more prone to cancer relapse and

metastasis. Comprehending the molecular regulation of CSCs self-renewal is crucial for advancing cancer biology research and revolutionizing cancer treatment. However, there are numerous challenges that impede extensive apprehension and effective targeting of CSCs in cancer. For instance, the absence of highly effective and specific techniques for CSCs identification and isolation presents a significant challenge, which may be attributed to their small subset representation within tumors (7). Additionally, the intricate molecular mechanisms governing CSCs self-renewal are not yet fully elucidated, thus, hampering the development of targeted therapies that specifically address CSCs populations (8). Therefore, a clearer understanding of CSCs and their role in driving cancer progression is necessary, as imprecise knowledge of the molecular underpinnings of CSCs behavior may obstruct the development of effective anti-cancer treatments (9). Moreover, advanced tools and breakthroughs are highly required in regards to CSCs isolation and characterization, that would facilitate efficient identification, isolating and assessment of CSCs, thus, propelling advancements in the niche of CSCs research (10). Regardingly, state of the art techniques such as single-cell sequencing, advanced imaging techniques, and more refined biomarker identification methods, may aid in optimum CSCs identification and characterization (11, 12). Moreover, collaborative efforts between researchers from various disciplines will be essential to unravel the complex nature of CSCs, thus, paving way for innovative clinical treatments targeting these cells. In 1994, Dick et al. used stem cell surface antigen labeling and flow cytometry for the first time to isolate and identify leukemic CSCs with stem cell markers (CD34+/CD38-) from human leukemia cells, which have potential to self-renew in acute myeloid leukemia (7). Subsequent studies used a similar approach to isolate and identify CSCs in different cancers such as breast (8), brain (9, 13), head and neck squamous cell carcinoma (10), pancreatic (11) and lung cancer (12) and prostate cancer (14). Most of the limitations of current methods for isolating and identifying cancer stem cells (CSCs) include insufficient precision, low yield of isolated cells (13), inability to mimic the complex tumor microenvironment (TME) (15), high-cost and procedural complexities, and inability to establish stable cellular models for subsequent future studies (16).

Integration of Artificial Intelligence (AI) in oncology has shown promising prospects in characterizing cancer stem cell properties, explicitly 'stemness'. The complexity and high dimensionality of cancer-related data, including genomics, transcriptomics, and epigenomics, underscore the need for Artificial Intelligence (AI) in oncology (17). Traditional analytical methods often fall short in capturing intricate, nonlinear patterns within large datasets. AI, particularly machine learning and deep learning models, has enabled the identification of hidden biomarkers, classification of tumor subtypes, and prediction of treatment responses with improved accuracy. In cancer stem cell (CSC) biology, AI facilitates the integration of multi-omics data to define stemness signatures and identify rare CSC populations (18). It also aids in modeling CSC dynamics, predicting resistance mechanisms, and uncovering novel therapeutic targets. Clinically, AI is expected to support the development of personalized therapies by stratifying

patients based on CSC-related risk and treatment sensitivity. As AI tools continue to evolve, they hold significant promise in bridging the gap between CSC research and real-world clinical applications (19). Rapid evolution of the high-throughput detection technologies entailing advanced imaging, genomic sequencing, and other -omics techniques, has enabled extensive data generation in regards to tumor histology (17, 18). This typically includes comprehensive tumor databases {TCGA (The cancer genome atlas), ICGC (International Cancer Genome Consortium), COSMIC, UCSC Cancer Genomics Browser, canEvolve, CGWB (Cancer Genome Workbench)}, Genome Database {Array Map, BioMuta, Cancer Hotspots, Mitelman Database, SomamiR, CGP (The Cancer Genome Project)}, DNA methylation database {MethHC, MethyCancer, MethDB, NGSmethDB, PubMeth, SurvivalMeth, DiseaseMeth2, MethSurv, MethBank, Lnc2Meth, MEXPRESS}, Transcriptome Database {Oncomine, GEO (Gene Expression Omnibus), ArrayExpress, ChiTaRS, miRCancer, OncomiRDB, UALCAN, CRN (Cancer RNA-Seq Nexus)}, Proteome Database {Cancer3D, CancerPPD, Cancer Proteome Variation Database (CanProVar), Clinical Proteomic Tumor Analysis Consortium (CPTAC), DbDEPC}, Database of tumor-related genes {DriverDB, Network of Cancer Genes (NCG), TP53MULTload, UMDTP53} and oncology and drug databases {CancerDR, CancerResource, canSAR, Genomics of Drug Sensitivity in Cancer (GDSC), Platinum}. This synergy between AI technologies and high-throughput detection methods holds tremendous promise in advancing our understanding of CSCs and their role in cancer biology. The amalgamation of AI-driven technologies with tumor-related databases presents a plausible tool for uncovering characteristic gene expression patterns, signaling pathways, and molecular markers associated with CSCs, thus, enhancing our ability to comprehensively and systematically understand the stemness characteristics of (CSCs) (20, 21). Moreover, this would significantly advance both basic tumor research and subsequent clinical treatment strategies, hence, fueling discoveries that may be translated into practical applications, potentially revolutionizing cancer treatment and patient care.

This review presents a novel synthesis by integrating cancer stem cell (CSC) biology with emerging artificial intelligence (AI)-driven analytical approaches, offering a unique perspective on stemness indices across various cancers. While numerous studies have explored CSCs or AI separately, few have critically examined their intersection, particularly in the context of mRNA<sub>si</sub>, mDNA<sub>si</sub>, DMP<sub>si</sub>, and ENH<sub>si</sub>. By comparing methodologies, highlighting limitations, and evaluating translational potential, this review bridges a critical knowledge gap. Additionally, the compilation of databases and tools provides a practical guide for researchers, making this review both timely and valuable for advancing precision oncology and CSC-targeted therapeutic strategies.

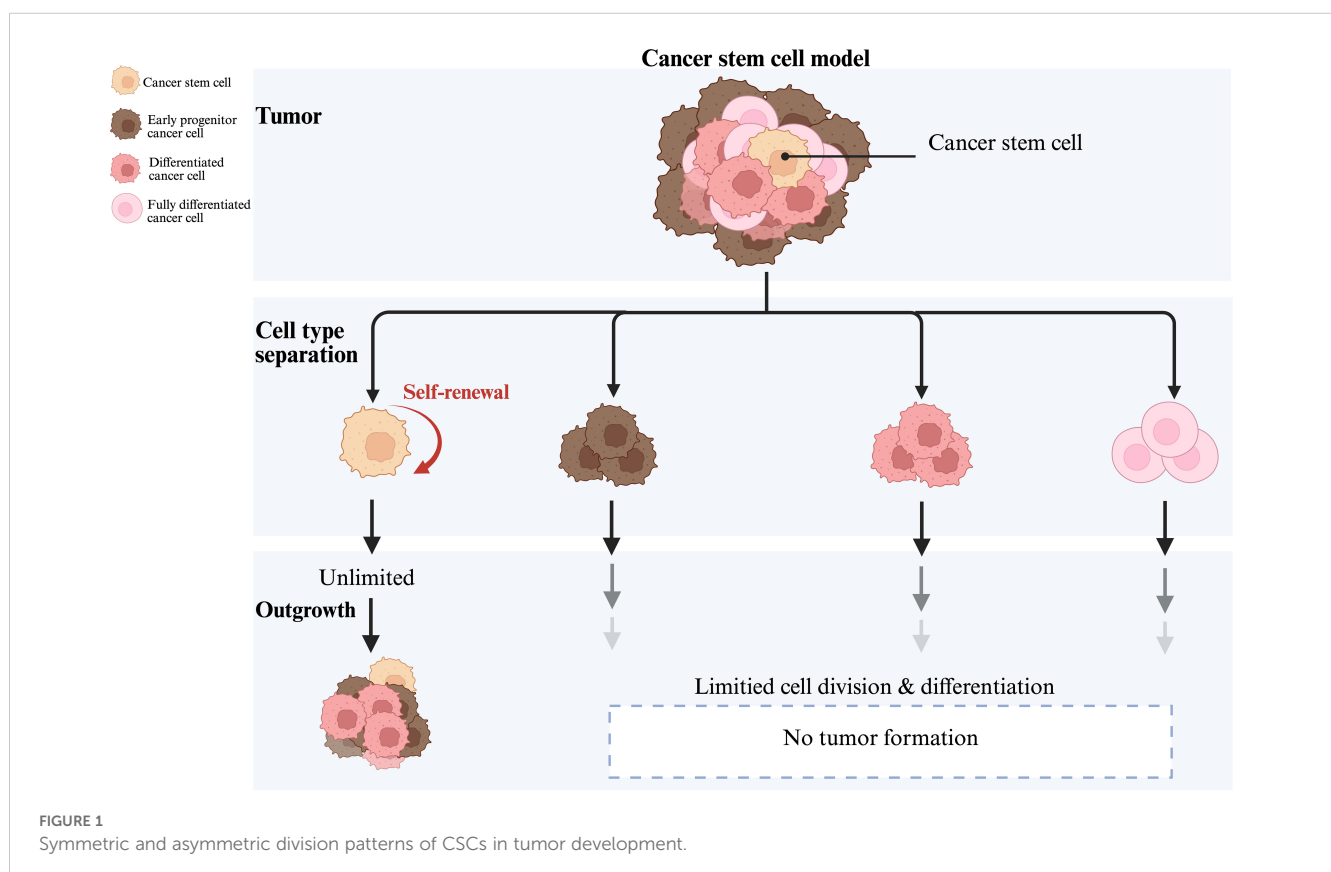
## 2 Biological basis of stemness of CSCs

Stem-like characteristics of CSCs include self-renewal capacity, differentiation potential, and high tumorigenicity (4, 5), as indicated

in the [Figure 1](#). Most distinguishing feature of CSCs is the capacity for self-renewal, generating specialized mitoses of one (asymmetric) or two (symmetric) daughter stem cells (4). When CSCs predominantly undergo symmetric divisions, they give rise to two identical CSCs, thus, culminating into an increased pool of CSCs within the tumor. Consequently, there is increased self-renewal capacity and an expansion of the stem cell population, contributing towards tumor aggressiveness, increased heterogeneity, and higher chances of malignancy. In contrast, asymmetric divisions generate one CSC and one differentiated progenitor or non-stem cell, maintaining a balance by replenishing the CSC pool while simultaneously producing cells that contribute to the tumor's differentiated cell population, therefore, resulting in a more stable tumor phenotype (22). Once this balance between the symmetric and asymmetric CSCs divisions is attained, the tumor tends to stabilize; whereas, when the balance is tilted towards symmetric divisions, the proportion of CSCs increases and the tumor manifests itself as highly malignant (23).

Some researchers have begun working on developing methods aiming to control the malignant transformation of tumors using asymmetric division regulatory mechanisms (24). CSCs exhibit high stemness in a variety of tumors such as leukemia (25), breast (26), brain (9, 13), colon (27) and lung cancer (28), and are involved in tumor growth, maintenance, and progression. Furthermore, the plasticity of CSCs (29) suggests that tumor cells can activate the stemness of CSCs by dedifferentiating and obtaining specific stimuli, leading to tumor recurrence. Additionally, CSCs stemness has a significant impact on cancer initiation, proliferation, metastasis, and

therapy resistance (30, 31). [Figure 2](#) provides a comprehensive overview of how these stemness characteristics contribute to these processes. Firstly, the stemness of CSCs affects the direction and difficulty of cancer initiation and influences augmented cellular carcinogenesis. It is closely related to cell proliferation, whereby, the stronger the stemness, the higher the proliferation ability, as shown in [Figure 2A](#). Tumor proliferative capacity and malignancy can be influenced by genetic, epigenetic, and proliferative modes of division regulation in CSCs. Secondly, the stemness of CSCs is also closely related to the ability of tumor metastasis, and their self-renewal and differentiation are often accompanied by cell motility and migration, allowing tumor cells to metastasize (32, 33). [Figure 2B](#) shows that the stemness genes of cancer stem cells mainly affect cancer metastasis through epithelial-mesenchymal transition (EMT)-related pathways (34). Studies have shown that EMT and the metastatic process of CSCs are quite similar, and their genes and transcriptomes have great overlap, suggesting that EMT cells and primary CSCs may be highly overlapping concepts (34). Although the same genetic markers are found in metastatic EMT and CSCs models, CSCs with high metastatic potential may not be a completely new type of cancer cell, but rather a subtype of CSCs or the result of a cellular gene mutation (35). For instance, in D133+ pancreatic CSCs, the migratory ability of the cells with high expression of CXC-chemokine receptor 4 (CXCR4) is significantly higher than those cells with low expression. As well, the patients with a high proportion of CD133+ and CXCR4+ cells in the cancerous tissues have higher probability of cancer metastasis (36). Briefly, the higher the stemness, the stronger is the metastatic ability. Finally, the characteristics of CSCs are closely linked to tumor therapy



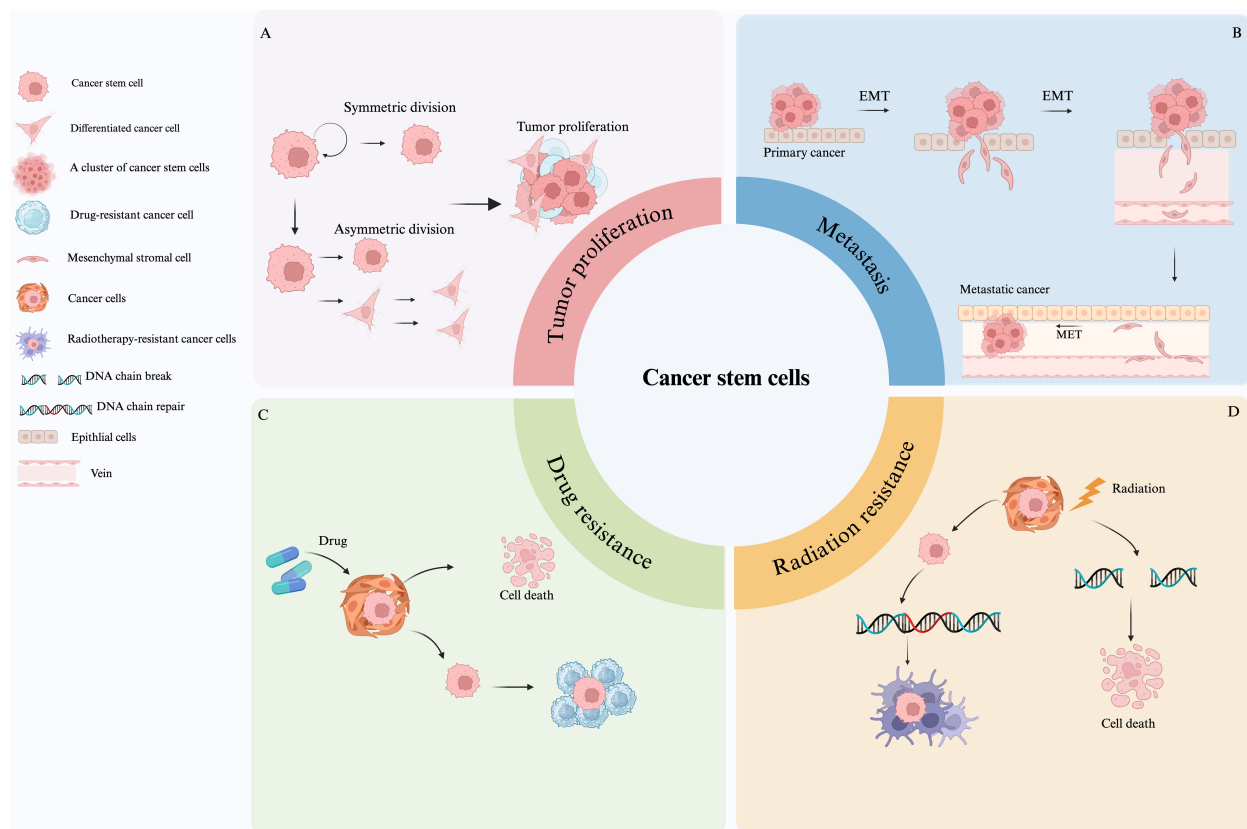


FIGURE 2

Roles of cancer stem cells in tumor proliferation, metastasis, and therapy resistance. (A) Mechanisms of CSC-driven tumor proliferation. (B) CSCs promotes metastasis primarily through EMT-related pathways. (C) CSCs express high levels of drug transporters, leading to chemoresistance. (D) CSCs resist radiation through enhanced DNA repair and microenvironmental protection.

resistance. Specifically, when the stemness of CSCs is in a dormant state, they are insensitive to external physicochemical factors that kill tumor cells and can evade treatment (37, 38), leading to therapy resistance (39). Mechanisms of resistance vary with treatment and primarily include high expression of drug transporters that assist in the transfer of intracellular toxic chemicals, as seen in Figure 2C. Additionally, CSCs possess strong DNA repair capacity, enabling them to resist radiation genome disruption (Figure 2D). Furthermore, CSCs recruit non-cancerous stem cells to form a protective microenvironment that further contributes to resistance for CSCs (40). Comprehensively, identification of the stemness features of CSCs is crucial to gain insight into the molecular mechanisms of tumor progression. These stemness features are not only relevant to tumor progression and therapeutic resistance, but also important for timely diagnosis, selection of the therapeutic strategies, and patient prognosis monitoring.

### 3 Methodology for assessment of stemness indices

With the rapid development of gene sequencing technology and the improvements in data processing technologies, AI has become a hot spot in tumor research (41, 42). Of these, Network Analysis

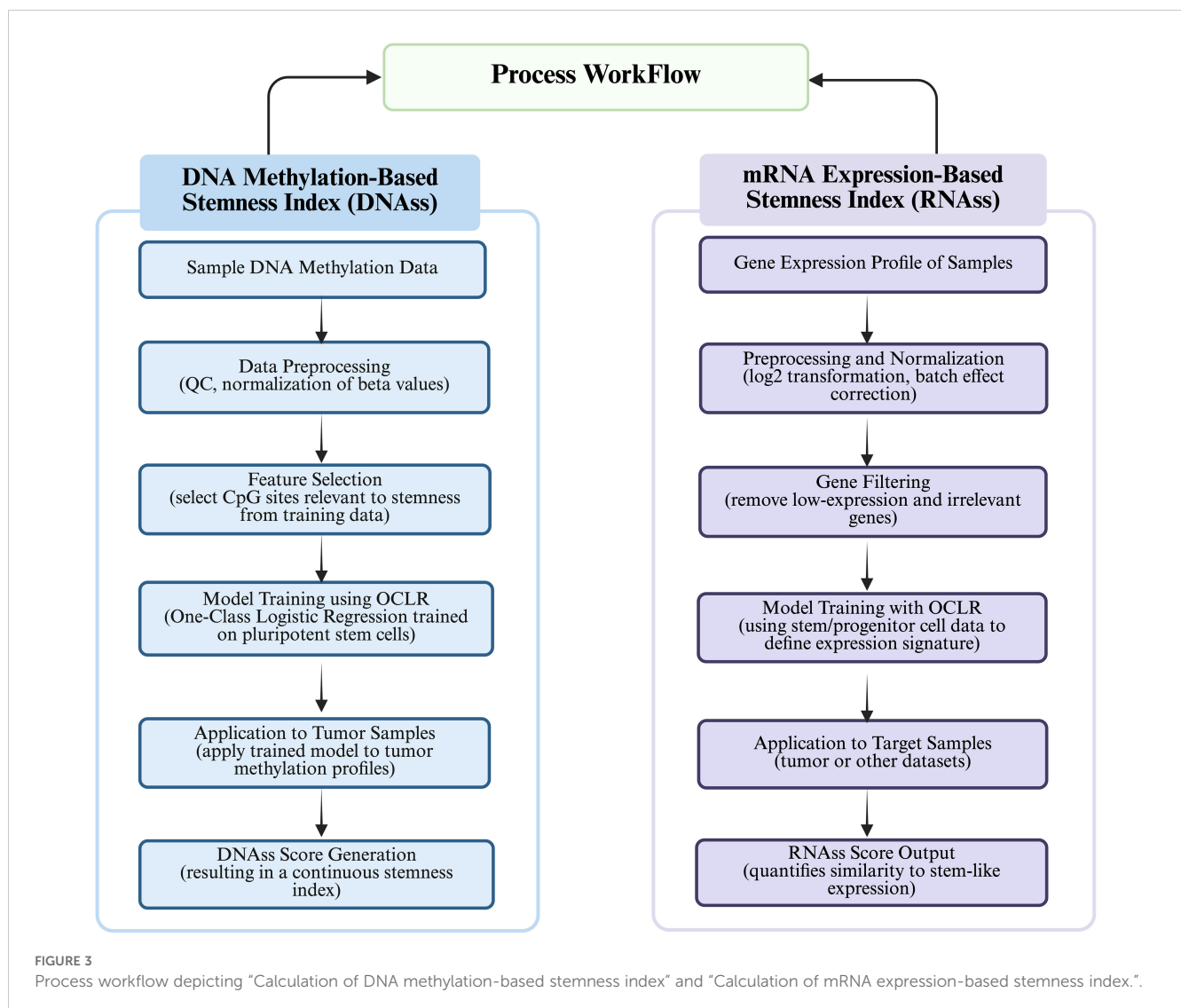
(NA) is an analytical method that uses genes or proteins as nodes and their interactions as edges (43). By constructing gene co-expression networks or protein interaction networks, a collection of genes related to a specific biological process, known as functional modules, can be identified (44), and ultimately the functional genes related to stemness can be identified by the relevance of these modules (45). Data mining through network analysis generates large-scale gene expression data, which can be utilized to identify stemness genes using machine learning with high accuracy and reliability (46). Currently, the methods employed for recognizing stemness genes based on machine learning algorithms are mainly categorized into (i) supervised learning, and (ii) unsupervised learning (47). One representative implementation of stemness index modeling is the work by Malta et al. (48), where the authors trained a one-class logistic regression (OCLR) model using stem cell-specific transcriptomic and epigenomic profiles obtained from the Progenitor Cell Biology Consortium (PCBC). The model was trained to capture the gene expression features of pluripotent stem cells and then applied to bulk tumor data from TCGA to generate a stemness index—referred to as (i) mRNAsi (based on transcriptome), derived via Spearman correlation between OCLR weights and tumor gene expression; (ii) mDNAsi (based on DNA methylation) (49), constructed by integrating three types of features: differentially methylated probes (DMPs) between

stem cells and progenitors, methylation markers of stem cell-specific enhancers (via Roadmap Epigenomics ChromHMM data), and epigenetically regulated genes identified by ELMER, as comprehended in Figure 3. Following the establishment of stemness indices by Malta et al., subsequent studies—such as the development of the TS score in bladder urothelial carcinoma (BLCA)—have leveraged tumor stemness quantification to classify subtypes, predict prognosis, and assess immunotherapy responsiveness based on CSC and EMT features (50). These indices quantitatively reflect the similarity between a tumor sample and the stem-like transcriptional or epigenetic phenotype. Importantly, high mRNasi scores and mDNasi scores were found to correlate with dedifferentiation, poor prognosis, and therapy resistance across multiple cancer types, supporting the relevance of stemness-based metrics in oncology. Since then, many studies have been conducted to develop and refine stemness indices using machine learning algorithms and gene expression signatures (17, 50, 51) to assist in the prediction of tumor growth, metastasis, and prognostic information, which are clinically significant in tumor diagnosis and treatment (5, 24, 42).

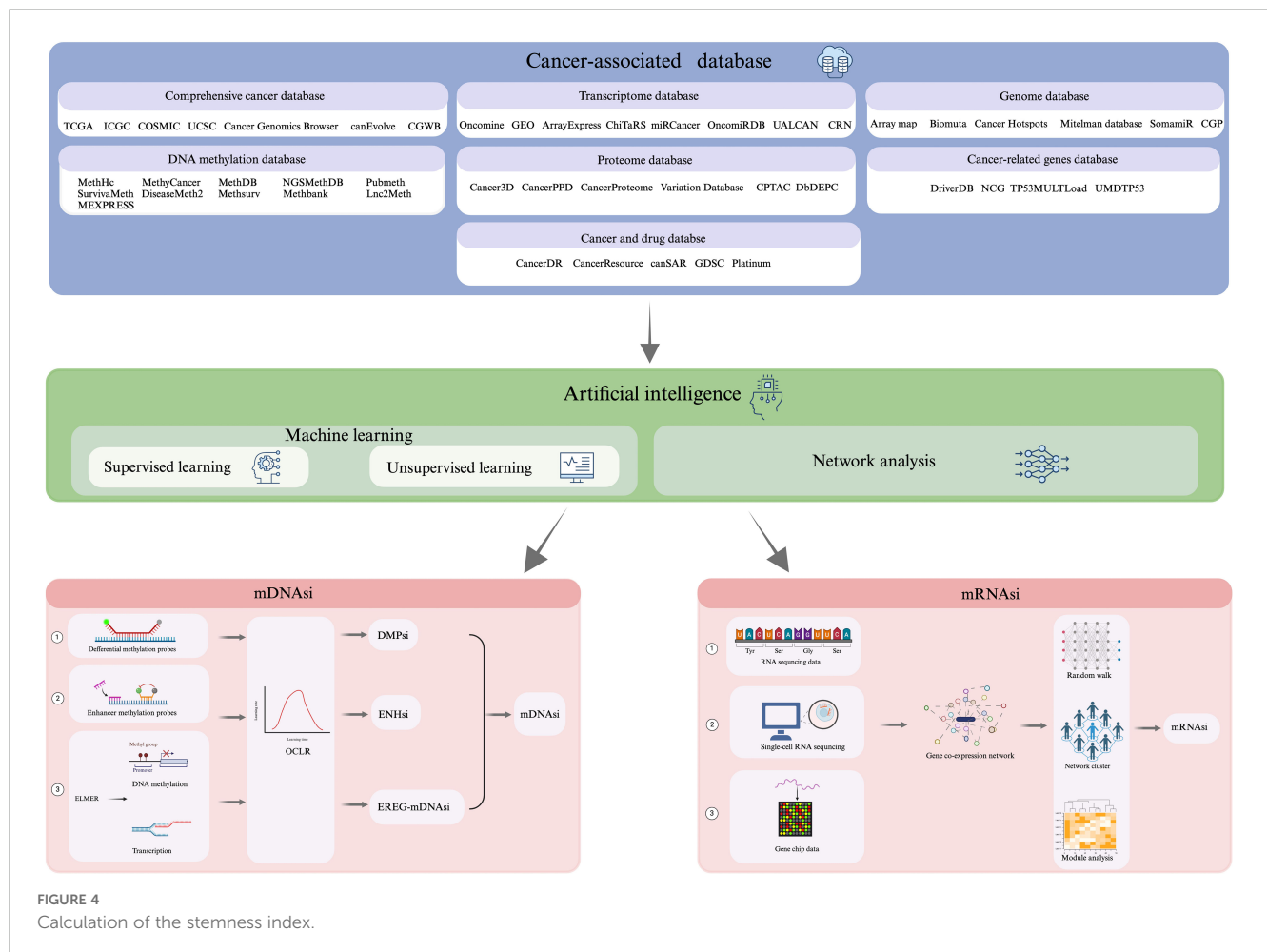
## 4 Artificial intelligence driven stemness techniques in stemness analysis

### 4.1 Calculation of DNA methylation-based stemness index

The process workflow for calculation of DNA Methylation-based Stemness Index has been illustrated in Figure 4. DNA methylation is the addition of methyl groups to cytosine residues in the DNA molecule, which regulates gene expression and cell differentiation, and is widespread in epigenetic modifications in eukaryotes, including CpG island methylation and non-CpG island methylation (52). During cell differentiation, stem cells usually have low levels of DNA methylation, whereas non-stem cells have high levels of DNA methylation (53). Malta (48) defined the mDNasi using OCLR by combining: supervised classification between ESC/iPSC and their progenies, iPSCs and their progeny. Roadmap in the ChromHMM software as well as ELMER (Enhancer Linking by







Methylation/Expression Relationships, ELMER) define mDNAsi on the basis of known methylation patterns of stemness and non-stemness genes using statistical models or machine learning algorithms to compute stemness indices including Euclidean distance, Pearson correlation coefficient and methylation clustering. The stemness index (SI) calculated from DNA methylation data involves the identification of specific CpG sites that exhibit differential methylation patterns between stem cells and non-stem cells. The differences in methylation levels at these CpG sites serve as a basis for quantifying the stemness of a particular cell sample or a tumor. The process typically involves identification of differential CpG Sites; calculation of Stemness Index; and quantification of Stemness (54). However, DNA methylation stemness features cannot be deciphered through a range of probes, however, different methylated regions are used as inputs, whereby, three methods are being employed depending on the input features (14, 48): (i) DMPsi (differentially methylated probes-based stemness index) uses differentially methylated probe regions (containing many filtering conditions) as input to the OCLR algorithm to construct predictive models; (ii) ENHsi (enhancer-based stemness index) uses methylated probes of enhancer regions as input to the OCLR algorithm to construct predictive models; and

(iii) EREG-mDNAsi uses the ELMER package to reconstruct gene regulatory networks from DNA methylation and transcriptome expression data, and use the identified features as inputs to the OCLR algorithm to construct predictive models, which can generate methylated probes and genes as output. The DNA methylation-based approach for calculating stemness index consists of two steps (55): (i) Firstly, “stemness genes” specifically expressed in stem cells are identified based on Gene Set Enrichment Analysis (GSEA), Principal Component Analysis (PCA), and Machine Learning, and (ii) Secondly, the DNA methylation level of each CpG site in the promoter region is determined using microarray or sequencing-based methods to calculate the methylation score for each gene. Once the methylation levels are scored for each gene, the DNA methylation-based SI was calculated using the “EpiScore” algorithm (56), which identifies CpG sites that are specifically methylated in stem cells and also differentially methylated in cancer cells. Subsequently, the SI scores are used to determine the average methylation level of CpG sites within the stem cell-specific CpG set. For instance, Liu (57) et al. characterized copy number alteration and genome-wide DNA methylation of meningioma subtypes using random forests and constructed a meningioma progression score (MPscore) using the stemness index.

## 4.2 Calculation of mRNA expression-based stemness index

The process for calculation of mRNAsi has been illustrated in the Figure 4. The mRNAsi was calculated based on gene expression patterns, functional annotations, gene networks, and gene differential expression data while using statistical models or machine learning algorithms (58). The most commonly used data are single-cell RNA sequencing data, gene chip data, or RNA sequencing data (44). By using gene co-expression networks to analyze co-expression patterns between genes, functional annotation information, and gene or protein interaction networks, stemness indices are calculated using random walk algorithms, network clustering, and modular analysis (36). Pertinently, Malta (48) validated mRNAsi by applying it to an external dataset consisting of stem and somatic differentiated cells and scored the molecular subtypes of breast cancers and gliomas with higher SI values for all stem cell samples than for differentiated cell samples. Likewise, Tan (59) obtained potential molecular subtypes of GBM patients from the GlioVis dataset by using the Consensus Cluster Plus (CC) R package based on unsupervised clustering analysis into gene.clusters.C1 (C1), gene.clusters.C2 (C2) and gene.clusters.C3 (C3), followed by the analysis of the TME variants, immune cell infiltration, and stemness indices for the three subtypes. Similarly, Sun (60) used a non-negative matrix decomposition algorithm to efficiently reduce the dimensionality of the integrated dataset (an effective dimensionality reduction method widely used to differentiate molecular patterns in high-dimensional genomic data), classified the expression of anoikic-related genes into Cluster 1 and Cluster 2, and analyzed the differences between the two clusters in terms of TME, stemness indices, and clinical traits and constructed the risk-scoring model to evaluate the relationship between risk scores of glioblastoma and pan-cancers and the TME, stemness, clinical traits, and response to immunotherapy. Da (45) first used the hclust function to cluster the samples and remove the outliers, followed by the use of the soft Power = sft\$powerEstimate command to select the optimal soft threshold to ensure that the interactions between lncRNAs conformed to the scale-free

distribution to the greatest extent possible and constructed the neighbor-joining matrix by calculating the topological overlap matrix (TO). Afterwards, hierarchical clustering using (1-TO) was used as the distance metric, selecting key modules by identifying them through the dynamic shear tree algorithm, defining the lncRNAs in the key modules as stemness index-associated lncRNAs, and finally constructing stemness index-associated lncRNA markers for predicting prognosis in breast cancer patients. Li (61) et al. used a one-class logistic regression machine learning algorithm (OCLR) to extract the transcriptomic and epigenetic feature sets derived from untransformed pluripotent stem cells and their differentiated progeny to calculate the mRNAsi values (mRNAsi ranges from 0 to 1, the closer the mRNAsi is to 1, the stronger the stem cells' features are), and found that the distribution of immune cells differed significantly between high and low mRNAsi lung cancer subtypes. Additionally, Table 1 summarizes the characteristics of four commonly used stemness indices—mRNAsi, mDNAsi, DMPsi, and ENHsi—highlighting their underlying molecular basis, data sources, strengths, limitations, and typical applications. mRNAsi is based on gene expression data and reflects transcriptional activity, while mDNAsi utilizes genome-wide DNA methylation patterns to capture epigenetic regulation. DMPsi refines mDNAsi by focusing on differentially methylated positions, thereby improving specificity in stemness evaluation. ENHsi further emphasizes enhancer-associated methylation, offering insights into non-coding regulatory mechanisms. Each index provides a distinct perspective on tumor stemness, complementing one another in cancer classification, prognosis prediction, and the study of epigenetic plasticity.

## 5 Clinical applications of AI-based stemness indices in cancer

The Stemness Index (SI) has gained significant importance in cancer research as a valuable metric for quantifying the extent of stem cell-like characteristics within a tumor cell population (62). By examining gene expression patterns and identifying genes that are

TABLE 1 Comparative analysis of common stemness indic.

| Stemness index | Basis                               | Data source                  | Advantages  | Limitations  | Applications   |
|----------------|-------------------------------------|------------------------------|---|--|--|
| mRNAsi         | Gene expression (mRNA)              | TCGA, GEO (bulk RNA-seq)     | Widely used; captures transcriptional activity; applicable across cancers | May miss epigenetic regulation; sensitive to tumor heterogeneity                 | Identifying stem-like phenotypes; correlating with prognosis and immune infiltration |
| mDNAsi         | DNA methylation patterns            | TCGA, GEO (450K/850K arrays) | Reflects epigenetic regulation; more stable than RNA-based indices        | Limited to methylation sites; may not fully capture transcriptional dynamics     | Cancer classification; predicting therapy resistance and epigenetic reprogramming    |
| DMPsi          | Differentially methylated positions | TCGA, GEO                    | Focuses on informative CpG sites; increased specificity                   | Dependent on high-quality DMP identification; limited cross-cohort comparability | Distinguishing tumor subtypes with differential stemness                             |
| ENHsi          | Enhancer-associated DNA methylation | ENCODE, TCGA                 | Targets regulatory regions; reflects enhancer activity driving stemness   | Complex analysis pipeline; enhancer regions less annotated in some cancers       | Understanding non-coding regulatory influence on stemness and cancer progression     |

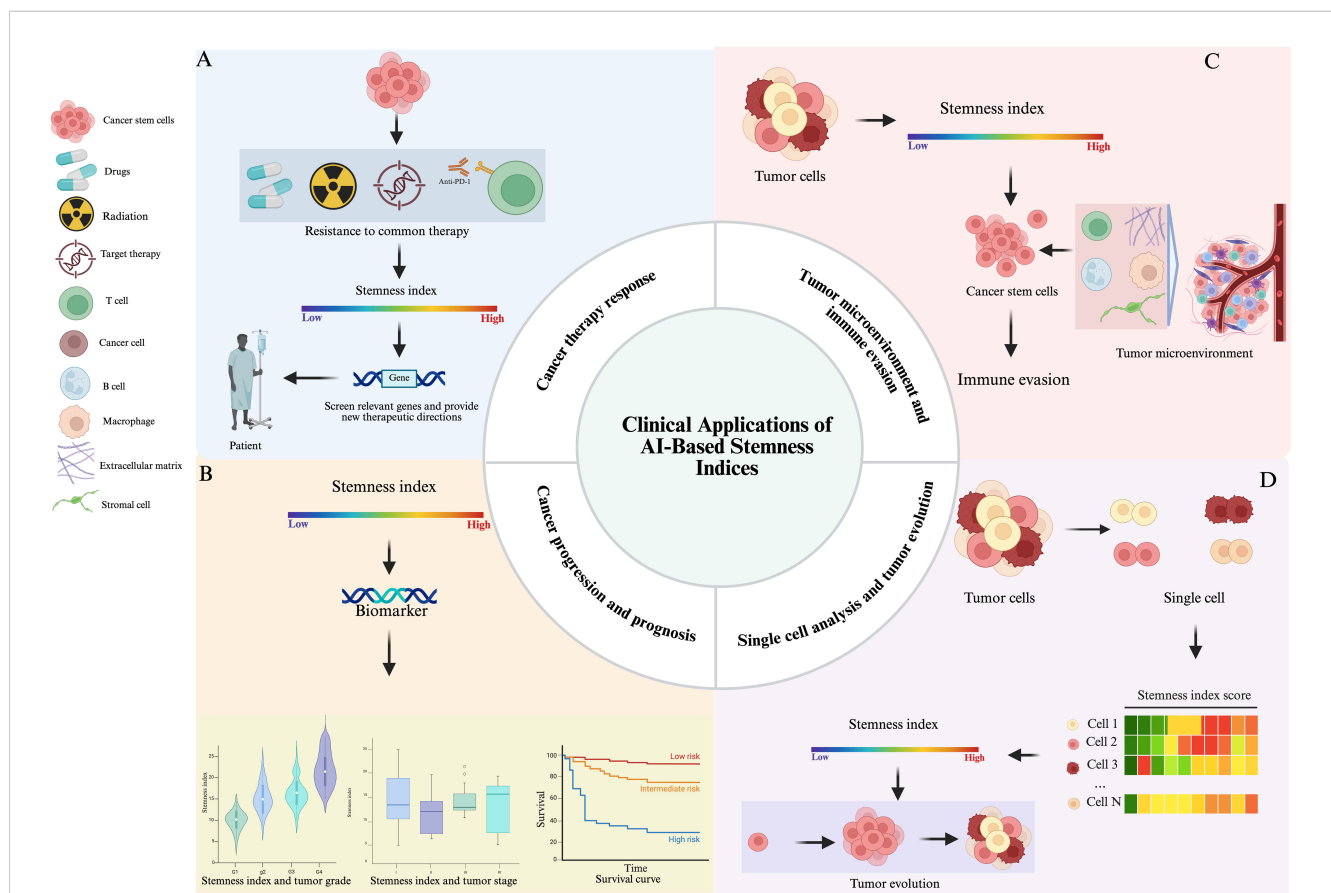
commonly expressed in stem cells, the stem cell index can be used to predict tumor aggressiveness, patient prognosis, and response to therapy (63), providing new targets and strategies for cancer management (50).

## 5.1 Stemness index in cancer therapy response

Currently, common treatment modalities for cancer include traditional radiotherapy, targeted therapy, and immunotherapy (42). The stem cell index identifies patients with tumors that have high levels of stem cells, reduces the number of patients who develop resistance to traditional chemotherapy and radiation, provides targeted therapies against CSCs, and improves treatment response rates (Figure 5A). CSCs are thought to be responsible for tumorigenesis, progression, and recurrence, and targeting these cells may improve overall therapeutic outcomes in cancer patients (42). In this regard, Guo (64) et al., screened 16 genes related to stem cell characteristics of IGC and 43 genes of DGC using mRNAsi. They preliminarily analyzed the relationship between the clinical features of gastric adenocarcinomas and the mRNAsi scores, and found that the tumor samples had higher stemness

indices than the normal samples, whereby, there was a significant difference between intestinal-type and diffuse-type gastric carcinomas; and that the stemness-properties-related genes were related to the cell cycle, and they could be used as a therapeutic target for inhibiting the stem cells of gastric cancer.

Furthermore, using the mRNAsi from The Cancer Genome Atlas (TCGA) to assess and correct tumor purity, along with the exploration of gene modules and key genes through weighted gene co-expression network analysis (WGCNA), has shown that grade III and IV tumors have higher mRNAsi and corrected mRNAsi scores than grade I and II tumors (65). This research verified the expression of 13 key genes between advanced platinum-resistant and sensitive SOC samples in two Gene Expression Omnibus (GEO) datasets, which showed CDC20 to be a potential platinum-sensitive indicator in advanced SOC (65). Moreover, the CTSF gene as a risk factor for resistance to a variety of tested drugs through drug susceptibility analysis, extensive resistance in the CTSF gene may be a potential reason for affecting disease outcome in patients with basal breast cancer, and selumetinib, SB590885, PLX4720, and Dabrafenib may be potential therapeutic or adjuvant therapeutic agents (66). Additionally, Shi (67) et al. screened 380 tumor stemness and immune (TSI)-related genes. Using a machine learning method, they constructed a five-gene TSI-





specific signature (TSISig) comprising CPS1, CCR2, NT5E, ANLN, and ABCC2. This process involved integrating the tumor stemness index (based on mRNA expression, mRNAsi), immune score, mRNA expression profiles, and clinical data from the TCGA database. TSISig demonstrated robust prognostic predictive ability and served as an effective indicator for tumor recurrence and response to radiotherapy and immunotherapy in LUAD patients. Moreover, the stem cell index can also be used to identify potential therapeutic targets for CSCs by analyzing the gene expression patterns of stem cell-like tumor cells and identifying signaling pathways and proteins that are critical for the maintenance of stemness and tumorigenicity for the development of new drugs and therapies against CSCs. The mRNA expression-based stemness index (mRNAsi), which can represent degrees of dedifferentiation of HCC samples, was calculated by Feng (68) et al. to predict the drug response of sorafenib therapy and prognosis. Unsupervised cluster analysis was conducted to distinguish mRNAsi-based subgroups, and gene/gene set functional enrichment analysis was employed to identify key sorafenib resistance-related pathways. By analyzing the core regulatory genes of the PPAR signaling pathway, they identified four candidate target genes, (i) retinoid X receptor beta (RXRB), (ii) nuclear receptor subfamily 1 group H member 3 (NR1H3), (iii) cytochrome P450 family 8 subfamily B member 1 (CYP8B1), and (iv) stearyl-CoA desaturase (SCD), as a signature to distinguish the response of sorafenib. They proposed and validated that the RXRB and NR1H3 could directly regulate NR1H3 and SCD, respectively. The results endorsed the combined use of SCD inhibitors and sorafenib as a promising therapeutic approach (68).

Finally, the stemness index has shown promise in identifying cancer patients who may benefit from targeted therapy or immunotherapy (42). In this regard, Wang (69) et al. retrieved gene expression data of 60 patients with gastrointestinal mesenchymal stromal tumor GIST from the Array Express database, applied CIBERSORT to calculate the immune infiltration level, used ssGSEA and ESTIMATE to calculate the cancer stemness index and the tissue purity, and implemented the connectivity map (CMAP) database to screen target drugs based on GIST's CSC-like properties to screen targeted drugs. Consequently, the results suggested that there were differences in immune infiltration levels between metastatic and non-metastatic GIST groups and that low levels of T-cell infiltration were associated with high tumor purity and tumor stemness index, with correlation coefficients of -0.87 and -0.61 ( $p < 0.001$ ), respectively (69). In addition, the cancer stemness index was positively correlated with cell purity ( $p < 0.001$ ) and was higher in the metastatic group than in the non-metastatic group ( $p = 0.0017$ ). Through the pharmacological mechanism of topoisomerase inhibitors, six molecular complexes may serve as the targets for GIST treatment (69). Studies for non-small cell lung cancer found that patients with a high stemness index responded better to immune checkpoint inhibitor therapy (PD-1/PD-L1 blockade) (70).

## 5.2 The stemness index serves as a potential biomarker for cancer progression and prognosis

Currently, the main factors affecting the prognosis of tumor patients include tumor grade, treatment modality, and independent prognostic influences. The stemness index has been investigated as a potential biomarker for predicting prognosis in patients with various types of cancers and identification of distant metastases in patients with high-risk cancers (57).

In a study of 355 breast cancer patients, the stemness index was found to be an independent predictor of distant metastasis-free survival and overall survival, suggesting that the stemness index can be used as a prognostic biomarker in breast cancer (71). Guo (72) analyzed gastric adenocarcinoma STAD cases in The Cancer Genome Atlas (TCGA) based on mRNAsi. mRNAsi analysis was performed on STAD by differential expression, survival analysis, clinical stage, and gender. Weighted gene co-expression network analysis (WGCNA) was used to identify useful modules and key genes, and enrichment analysis was carried out to annotate the functions and pathways of key genes. Finally, the expression levels of key genes in all the cancers were validated using the Gene Expression Omnibus (GEO) database in STAD, and the protein-protein interaction network was used to determine the relationship between the key genes. The results showed a decrease in mRNAsi scores with increasing tumor stage and T-stage, and a higher overall survival in highly grouped patients (72). Lyu (73) found that stemness index based on corrected mRNA expression was up-regulated in renal clear cell carcinoma (KIRC) tissues compared to non-tumor tissues and increased with tumor stage and grade. Similarly, EZH2 expression was associated with tumor-infiltrating immune cells, and epigallocatechin-3-gallate (EGCG) was determined to be the most potent inhibitor of EZH2. Notably, the percentage of FoxP3+ Treg cells in the peripheral blood mononuclear cells of ccRCC patients was significantly lower when cultured in spheroids pretreated with sunitinib, thus, Zhao (74) calculated mRNAsi from more than 500 lung adenocarcinoma patients from TCGA database based on a one-class logistic regression machine learning algorithm for pluripotent stem cells and their post-differentiation mRNA expression. mRNAsi-related key genes were identified by weighted correlation network analysis, and the results suggested that the mRNAsi was significantly higher in LUAD compared to normal lung tissues, whereby, patients with advanced LUAD demonstrated higher mRNAsi and poorer overall survival (OS). EZH2 was identified as a CSC marker and prognostic factor in KIRC patients. Huang (66) found that basal-like breast cancer carries the highest mRNAsi among all four breast cancer subtypes, and 385 mRNAsi-related genes were positively correlated with high mRNAsi values of basal breast cancer. High mRNAsi is closely associated with active cell cycle, DNA replication and metabolic reprogramming in basal-like breast cancer. Among them, TRIM59, SEPT3, RAD51AP1 and EXO1 can be used as

independent protective factors, and CTSF and ABHD14B are used as risk factors, and the establishment of a prognostic model containing mRNAsi-related genes can effectively predict the survival in patients diagnosed with basal type breast cancer subtypes. Tan (75) classified gliomas into low-grade gliomas and glioblastomas based on mRNAsi-related genes by consensus clustering of TCGA dataset, and developed prognostic features related to stemness subtypes, which could effectively predict the prognosis in glioma patients. Tang (76) used single-sample gene set enrichment analysis (GSEA) to calculate the relative activities of the metabolic pathways in pancreatic ductal adenocarcinoma (PDAC) samples, and found that the overall survival (OS) of patients with high mRNAsi values was significantly lower than the patients with low mRNAsi values ( $P = 0.003$ ). Moreover, weighted gene co-expression network analysis (WGCNA) revealed eight independent gene modules significantly associated with mRNAsi and 12 metabolic pathways, and two PDAC subgroups were identified based on unsupervised clustering of the key genes in each module, which demonstrated that PDAC samples with high mRNAsi values exhibited aberrant activation of multiple metabolic pathways, and the patients exhibited poor prognosis.

Overall, the AI-based Stem Cell Index stands as a powerful and evolving tool in cancer research with profound implications for understanding tumor grading, staging, and predicting prognosis (Figure 5B). The ongoing use, refinement, and integration of the AI-based Stem Cell Index in cancer research promises to revolutionize cancer treatment by offering more precise diagnostics, tailored treatments, and a deeper understanding of the intricate mechanisms governing tumor development and progression, thus, significantly improving patient outcomes and quality of life in cancer affected individuals.

### 5.3 Stemness index in tumor microenvironment and immune evasion

In the TME, a high stemness index is often associated with aggressive tumor behavior, therapy resistance, and poor prognosis. CSCs, which typically have high stemness scores, interact dynamically with components of the TME encompassing immune cells, stromal cells, and extracellular matrix (ECM) to maintain their stem-like state (77, 78) (Figure 5C). These interactions promote immune evasion by inducing immunosuppressive signaling pathways, modulating antigen presentation, and recruiting regulatory T cells and myeloid-derived suppressor cells. The TME, thus, becomes a sanctuary for CSCs, shielding them from immune surveillance and enhancing their survival (79, 80). Understanding the relationship between stemness and immune evasion offers valuable insights for therapeutic strategies aimed at disrupting CSC niches, reactivating anti-tumor immunity, and improving the efficacy of immunotherapies in cancers with high stemness signatures. Additionally, integrating stemness indices into prognostic models may aid in patient stratification and personalized treatment design (81).

### 5.4 Stemness index in single cell analysis and tumor evolution

In the context of single-cell analysis, the stemness index is particularly valuable for identifying subpopulations within heterogeneous tumor tissues. Figure 5D illustrates how the stemness index facilitates the identification of tumor subpopulations in heterogeneous tissues during single-cell analysis. By applying transcriptomic profiling at single-cell resolution, researchers can calculate stemness scores for individual tumor cells, uncovering gradients of differentiation and identifying (CSCs). These CSCs are often associated with therapy resistance, metastasis, and poor prognosis (82). Understanding the stemness index in single-cell data provides insights into tumor evolution, as cancer progresses through branching paths of clonal expansion, differentiation, and selection. Tumor cells with high stemness indices may serve as founders of new subclones, driving tumor heterogeneity and adaptive evolution under treatment pressure. Moreover, the spatial and temporal dynamics of stemness across a tumor can reveal how specific microenvironmental niches support CSC maintenance (50). Incorporating stemness indices into single-cell and spatial transcriptomics datasets helps reconstruct tumor lineage trajectories and evolutionary hierarchies (83, 84). Ultimately, this approach can inform therapeutic strategies by targeting CSC populations and their supporting environments, potentially improving long-term treatment outcomes by disrupting the cellular plasticity that fuels tumor progression and recurrence.

## 6 Cancer specific insights into applicability of the AI-driven methods in various cancer types

AI-driven stemness indices have been widely applied across various cancers to reveal prognostic patterns, immune landscapes, and therapeutic vulnerabilities, as shown in Table 2.

### 6.1 Lung cancer – lung adenocarcinoma

In LUAD, the AI-derived mRNA stemness index (mRNAsi) is significantly higher in tumors than in normal lung tissue and increases with tumor stage. Patients with high mRNAsi had notably worse overall survival. A study combining OCLR-based mRNAsi with immune profiling identified a set of 144 “immune-stemness” genes. Hub genes including IL-6, FPR2, RLN3 were linked to poor prognosis and correlated with immune checkpoints and tumor mutational burden (TMB) (85).

### 6.2 Colorectal cancer

AI-based analysis stratified CRC patients into high- and low-mRNAsi groups. High-mRNAsi was associated with poorer overall

TABLE 2 Cancer specific insights into applicability of the AI-driven methods in various cancer type.

| Cancer type        | Application of stemness index   | Key findings   | AI/ML methods used                     | Notable biomarkers/pathways              |
|--------------------|---|--|--|--|
| NSCLC (Lung)       | Classification into stemness subtypes; prognosis and immunotherapy prediction | High-stemness linked to poor survival and immune exclusion               | LASSO, XGBoost, Random Forest          | ARTN, CD8+ T-cell exclusion, EMT         |
| HCC (Liver)        | Prognostic modeling; immune profiling; methylation integration                | High-stemness associated with low immune scores and worse prognosis      | OCLR, LASSO, Multi-omics models        | Wnt signaling, methylation patterns      |
| CRC (Colon/Rectum) | Risk stratification and chemotherapy sensitivity                              | High-stemness tumors resistant to standard therapies                     | LASSO Cox, SVM                         | EMT, angiogenesis, tumor mutation burden |
| BLCA (Bladder)     | Subtype identification and immune evasion prediction                          | High-stemness subtypes show immune suppression and worse outcome         | Boruta, SVM, Consensus Clustering      | TNFAIP6, immune checkpoints              |
| PDAC (Pancreas)    | Spatial distribution of CSCs and immune interaction analysis                  | AI revealed immune-suppressive niches near stem-like cells               | Pathology AI + spatial transcriptomics | CD133+, CD8+ spatial mapping             |
| GC (Gastric)       | Drug sensitivity prediction and subtype classification                        | High-stemness linked to chemotherapy resistance                          | GSVA, XGBoost                          | Wnt/ $\beta$ -catenin, MAPK signaling    |
| Pan-Cancer         | Cross-cancer stemness comparison and biomarker discovery                      | Universal links between high stemness, immune evasion, and mutation load | Deep Learning, Multi-omics AI          | Shared CSC markers, immune escape genes  |

survival in stage IV CRC, increased TMB, and altered immune infiltration patterns. Prognostic stemness signature: Weighted gene co-expression network analysis (WGCNA) and LASSO-Cox regression identified a three-gene prognostic signature (PARBP, KNSTRN, KIF2C). This signature was validated via tissue immunofluorescence and incorporated into a nomogram outperforming TNM staging. High-stemness CRC tumors showed lower immune/stromal scores and reduced infiltration by macrophages, but higher CD8+ and T follicular helper cells, suggesting specific immune microenvironment remodeling (86).

### 6.3 Breast cancer

WGCNA of breast cancer transcriptomes linked mRNAsi to hub cell cycle genes (CDC20, PLK1, BUB1/BUB1B, NCAPG, KIF20A). These genes were overexpressed in advanced tumor stages and are promising therapeutic targets. Breast cancer stem cells (BCSCs) often display CD44+/CD24- phenotypes, undergo epithelial-mesenchymal transition (EMT), and are driven by Notch, HER2, and NF- $\kappa$ B signaling. Their metabolism supports self-renewal and therapy resistance (87).

### 6.4 Bladder, pancreatic, and gastric cancer

In bladder Cancer, AI-based clustering helps identify stemness-driven subtypes that are associated with poor prognosis and immune evasion. For example, TNFAIP6 was discovered as a critical gene in high-stemness tumors (88). AI combined with spatial pathology can reveal how cancer stem cells are arranged within the tumor and how they interact with the immune system (e.g., CD133+ cells co-located with immune suppression zones) in

Pancreatic Ductal Adenocarcinoma (PDAC) (89). In Gastric Cancer (GC), stemness indices are used to stratify patients and predict drug sensitivity. Moreover, certain pathways (like Wnt signaling) are often enriched in high-stemness subtypes in GC (90).

## 6.5 Glioblastoma

While direct mRNAsi studies are limited, deep learning-based radiogenomic pipelines have been used to automatically segment tumors and predict survival in GBM. A 3D CNN radiomic signature yielded a C-index of 0.67 (vs. 0.64 for traditional methods), with significant patient stratification (91). Wnt/ $\beta$ -catenin stemness signaling: Preclinical work highlights Wnt pathway activation in glioblastoma stem cells (GSCs), with elevated  $\beta$ -catenin, TCF/LEF1, LGR5, and c-Myc—suggesting a possible basis for integrating AI-based signaling pathway quantification (92).

## 7 Challenges and limitations

Despite the promising advances in applying artificial intelligence (AI) to develop stemness indices in cancer, several challenges and limitations remain, which need to be addressed before these tools can be fully integrated into clinical practice.

One major challenge is the biological complexity of cancer stemness. Cancer stem cells (CSCs) are highly heterogeneous, existing in dynamic states that vary not only between tumor types but also within a single tumor. Capturing this heterogeneity through AI models is difficult, particularly when using bulk transcriptomic data that averages signals from diverse cell populations. This can obscure important variations, limiting the accuracy of stemness indices (93). Another significant limitation is

the lack of standardized frameworks for defining and measuring stemness. Different studies use varying gene sets, computational methods, and cutoffs to calculate stemness scores, leading to inconsistent results that are hard to compare or reproduce. This heterogeneity complicates efforts to validate stemness indices across cohorts and cancer types (94). Data-related issues also pose challenges. Heterogeneity across datasets, including differences in sequencing platforms, sample processing, and patient demographics, can introduce biases. Many AI models are trained on retrospective public datasets such as TCGA, which may not represent the diversity of clinical populations, affecting the generalizability of findings (95).

Another critical limitation is the lack of prospective clinical validation. Most studies rely on retrospective analyses, and few have demonstrated how AI-based stemness indices perform in predicting patient outcomes or guiding therapy decisions in real-time clinical settings. Moreover, overfitting remains a concern, especially with complex machine learning models applied to relatively small datasets. This can lead to overly optimistic performance metrics that do not hold up in independent validation. Finally, integrating AI-derived stemness indices with existing clinical workflows requires models to be interpretable and explainable, yet many AI methods remain “black boxes,” which limits clinician trust and adoption. Addressing these challenges through rigorous methodological standardization, large-scale prospective studies, and explainable AI will be essential for translating stemness indices into practical cancer care tools (96).

While stemness indices such as mRNasi and mDNasi have demonstrated promising applications in cancer research, several studies have reported contradictory or limited findings. For instance, the predictive value of stemness scores varies across tumor types; in some cancers, high stemness correlates with poor prognosis, while in others, no significant association is observed (97). Additionally, discrepancies arise when comparing indices derived from different data platforms (e.g., TCGA vs. GEO), often due to batch effects and data normalization inconsistencies. Some AI-based models show reduced reproducibility when applied to external validation cohorts, highlighting concerns about overfitting and lack of generalizability (98). Moreover, stemness scores sometimes fail to reflect the functional heterogeneity of cancer stem cells (CSCs) within the tumor microenvironment. These inconsistencies underscore the need for standardized methodologies, cross-cohort validation, and integration of multi-omics data to improve the reliability of stemness-based metrics in cancer biology.

## 8 Future directions

The application of artificial intelligence (AI) to quantify cancer stemness is a rapidly evolving field with substantial promise, but several key areas warrant further development to maximize clinical impact. One important future direction is the integration of spatial transcriptomics and pathology-based AI. By combining gene expression data with spatial localization of cells within the tumor

microenvironment (TME), researchers can gain a more nuanced understanding of how cancer stem cells (CSCs) interact with immune and stromal cells, potentially revealing new therapeutic targets. Another promising avenue is the advancement of single-cell stemness models. Current AI-based stemness indices primarily rely on bulk tumor data, which can obscure heterogeneity. Single-cell technologies, coupled with machine learning, will allow more precise identification of CSC subpopulations and their dynamic states, improving prognostic accuracy and therapy stratification. Furthermore, the transition from retrospective computational analyses to prospective clinical trials is essential. Validating AI-derived stemness scores in real-world patient cohorts will help establish their utility in guiding treatment decisions and predicting outcomes. Developing pan-cancer deep learning frameworks that integrate multi-omics data across tumor types can uncover universal and cancer-specific stemness signatures, enhancing personalized medicine. Finally, synergizing AI with mechanistic biology is crucial for drug discovery. Understanding the molecular underpinnings of AI-identified stemness features will facilitate development of novel therapies targeting CSCs, addressing therapy resistance and relapse. Overall, these directions emphasize a multidisciplinary approach to fully harness AI's potential in stemness research and precision oncology.

Collectively, these challenges highlight the need to refine stemness modeling—potentially through transfer learning, data augmentation, multi-omics integration, and single-cell approaches—to enhance its reliability and clinical applicability, particularly in the context of rare tumors.

## 9 Conclusions

The stemness index, constructed on the basis of artificial intelligence, is a measure of the extent of stem cell-like features in cancer cells that has recently emerged as a promising biomarker for identifying different cancer subtypes, thus, aiding in prognostication and therapeutic decision-making in breast, colorectal, lung, hepatocellular carcinoma, and glioblastoma. However, its clinical utility encounters various challenges and limitations that are required to be addressed for enhanced applicability in cancer management. Most importantly, lack of standardized methods for calculating the Stemness Index is one of the major challenges, leading towards variability and potential biases in results across studies and datasets. Therefore, establishing standardized protocols is crucial for ensuring consistency and comparability of the results, thus obtained. Additionally, limited availability of high-quality datasets, especially for rare or less common cancer subtypes is a significant constraint. Since training and validation of the Stemness Index models require diverse and comprehensive datasets, the scarcity of such datasets for less prevalent cancers hinders the accuracy and generalizability of the Index. While promising, the clinical utility and reliability of the cancer stemness indices need rigorous validation through large-scale prospective studies. Validating its effectiveness in predicting prognosis, treatment response, and guiding therapeutic decisions in diverse patient



populations is essential for its implementation in a clinical setting. Various factors, including TME, patient demographics, and comorbidities, could influence the applicability of the Stemness Index. Therefore, it is critically important to apprehend and account for the confounder to ensure the accuracy and reliability of the stemness index in reflecting accurate stemness characteristics in various cancer subtypes. Nonetheless, additional research studies are highly necessitated to overcome existing challenges and address the aforementioned limitations, and to establish the clinical relevance and utility of the stemness index as a putative biomarker for various cancer subtypes, at large.

## Author contributions

LL: Writing – original draft, Visualization. QP: Writing – original draft. JQ: Conceptualization, Writing – original draft, Writing – review & editing. YC: Writing – review & editing. JL: Writing – review & editing, Supervision, Conceptualization. YL: Validation, Writing – review & editing, Supervision. JX: Visualization, Writing – review & editing. RD: Validation, Writing – review & editing. TY: Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Youth Fund, China (Grant No.

82003138); the Sichuan Science and Technology Program for International Cooperation (Grant No. 2024YFHZ0331); the Medical Science and Technology Development Project of Clinical Medicine in Southwest Medical University (Grant No. 2024LCYXZX24).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Clevers H. The intestinal crypt, a prototype stem cell compartment. *Cell*. (2013) 154:274–84. doi: 10.1016/j.cell.2013.07.004
- Doulatov S, Notta F, Laurenti E, Dick JE. Hematopoiesis: a human perspective. *Cell Stem Cell*. (2012) 10:120–36. doi: 10.1016/j.stem.2012.01.006
- Desai A, Yan Y, Gerson SL. Concise reviews: cancer stem cell targeted therapies: toward clinical success. *Stem Cells Transl Med*. (2019) 8:75–81. doi: 10.1002/sctm.18-0123
- Mayani A, Chávez-González A, Vázquez-Santillan K, Contreras J, Guzman ML. Cancer stem cells: biology and therapeutic implications. *Arch Med Res*. (2022) 53:770–84. doi: 10.1016/j.arcmed.2022.11.012
- Battle E, Clevers H. Cancer stem cells revisited. *Nat Med*. (2017) 23:1124–34. doi: 10.1038/nm.4409
- Baker M. Cancer stem cells tracked. *Nature*. (2012) 488:13–4. doi: 10.1038/488013a
- Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*. (1994) 367:645–8. doi: 10.1038/367645a0
- Vohra LM, Mooghal M, Khan W, Sohail H, Zeeshan S. Comprehensive targeted treatment options available for breast cancer stem cells; a literature review of the last 10 years' developments. *JPMA. J Pak Med Assoc*. (2023) 73:S47–55. doi: 10.47391/JPMA.AKUS-08
- Gimple RC, Yang K, Halbert ME, Agnihotri S, Rich JN. Brain cancer stem cells: resilience through adaptive plasticity and hierarchical heterogeneity. *Nat Rev Cancer*. (2022) 22:497–514. doi: 10.1038/s41568-022-00486-x
- Prince ME, Sivanandan R, Kaczorowski A, Wolf GT, Kaplan MJ, Dalerba P, et al. Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *Proc Natl Acad Sci U. S. A*. (2007) 104:973–8. doi: 10.1073/pnas.0610117104
- Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, Guba M, et al. Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. *Cell Stem Cell*. (2007) 1:313–23. doi: 10.1016/j.stem.2007.06.002
- Eramo A, Lotti F, Sette G, Pilozzi E, Biffoni M, Di Virgilio A, et al. Identification and expansion of the tumorigenic lung cancer stem cell population. *Cell Death Differ*. (2008) 15:504–14. doi: 10.1038/sj.cdd.4402283
- Fang JS, Deng YW, Li MC, Chen FH, Wang YJ, Lu M, et al. Isolation and identification of brain tumor stem cells from human brain neuroepithelial tumors. *Zhonghua Yi Xue Za Zhi*. (2007) 87:298–303. doi: 10.3760/j.issn:0376-2491.2007.05.003
- Liu X, Li WJ, Puzanov I, Goodrich DW, Chatta G, Tang DG. Prostate cancer as a dedifferentiated organ: androgen receptor, cancer stem cells, and cancer stemness. *Essays Biochem*. (2022) 66:291–303. doi: 10.1042/EBC20220003
- Wan Q, Ren X, Tang J, Ma K, Deng Y-P. Cross talk between tumor stemness and microenvironment for prognosis and immunotherapy of uveal melanoma. *J Cancer Res Clin Oncol*. (2023) 149:11951–68. doi: 10.1007/s00432-023-05061-x
- Medema JP. Cancer stem cells: the challenges ahead. *Nat Cell Biol*. (2013) 15:338–44. doi: 10.1038/ncb2717
- Gevaert O, Villalobos V, Sikic BI, Plevritis SK. Identification of ovarian cancer driver genes by using module network integration of multi-omics data. *Interface Focus*. (2013) 3:20130013. doi: 10.1098/rsfs.2013.0013
- Gingold J, Zhou R, Lemischka IR, Lee D-F. Modeling cancer with pluripotent stem cells. *Trends Cancer*. (2016) 2:485–94. doi: 10.1016/j.trecan.2016.07.007
- Chen Y-M, Hsiao T-H, Lin C-H, Fann YC. Unlocking precision medicine: clinical applications of integrating health records, genetics, and immunology through artificial intelligence. *J BioMed Sci*. (2025) 32:16. doi: 10.1186/s12929-024-01110-w
- Hope A, Verduin M, Dilling TJ, Choudhury A, Fijten R, Wee L, et al. Artificial intelligence applications to improve the treatment of locally advanced non-small cell lung cancers. *Cancers*. (2021) 13:2382. doi: 10.3390/cancers13102382



21. Zhang Yuhang, You Peimeng, Liu Ruizhe, Lu Yingwei, Li Jinjing, Lei Yu, et al. Artificial intelligence in clinical trials of lung cancer: current and future prospects. *Intell Oncol.* (2025) 1:34–51. doi: 10.1016/j.intonc.2024.11.003
22. Kumar VE, Nambiar R, De Souza C, Nguyen A, Chien J, Lam KS, et al. Targeting epigenetic modifiers of tumor plasticity and cancer stem cell behavior. *Cells.* (2022) 11:1403. doi: 10.3390/cells11091403
23. Hallis SP, Kim JM, Kwak M-K. Emerging role of NRF2 signaling in cancer stem cell phenotype. *Mol Cells.* (2023) 46:153–64. doi: 10.14348/molcells.2023.2196
24. Pattabiraman DR, Weinberg RA. Tackling the cancer stem cells - what challenges do they pose? *Nat Rev Drug Discov.* (2014) 13:497–512. doi: 10.1038/nrd4253
25. Jones CL, Inguva A, Jordan CT. Targeting energy metabolism in cancer stem cells: progress and challenges in leukemia and solid tumors. *Cell Stem Cell.* (2021) 28:378–93. doi: 10.1016/j.stem.2021.02.013
26. Landeros N, Castillo I, Pérez-Castro R. Preclinical and clinical trials of new treatment strategies targeting cancer stem cells in subtypes of breast cancer. *Cells.* (2023) 12:720. doi: 10.3390/cells12050720
27. D'Antonio L, Fieni C, Ciummo SL, Vespa S, Lotti L, Sorrentino C, et al. Inactivation of interleukin-30 in colon cancer stem cells via CRISPR/Cas9 genome editing inhibits their oncogenicity and improves host survival. *J Immunother Cancer.* (2023) 11:e006056. doi: 10.1136/jitc-2022-006056
28. Leon G, MacDonagh L, Finn SP, Cuffe S, Barr MP. Cancer stem cells in drug resistant lung cancer: targeting cell surface markers and signaling pathways. *Pharmacol Ther.* (2016) 158:71–90. doi: 10.1016/j.pharmthera.2015.12.001
29. Thankamony AP, Saxena K, Murali R, Jolly MK, Nair R. Cancer stem cell plasticity - A deadly deal. *Front Mol Biosci.* (2020) 7:79. doi: 10.3389/fmolb.2020.00079
30. Nassar D, Blanpain C. Cancer stem cells: basic concepts and therapeutic implications. *Annu Rev Pathol.* (2016) 11:47–76. doi: 10.1146/annurev-pathol-012615-044438
31. Huang T, Song X, Xu D, Tiek D, Goenka A, Wu B, et al. Stem cell programs in cancer initiation, progression, and therapy resistance. *Theranostics.* (2020) 10:8721–43. doi: 10.7150/thno.41648
32. Zhou T, Zhang LY, He JZ, Miao ZM, Li YY, Zhang YM, et al. Review: Mechanisms and potential treatment of radioresistance in non-small cell lung cancer. *Front Immunol.* (2023) 14:1133899. doi: 10.3389/fimmu.2023.1133899
33. Shibue T, & Weinberg, R. A. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat Rev Clin Oncol.* (2017) 14:611–29. doi: 10.1038/nrclinonc.2017.44
34. Biddle A, Mackenzie IC. Cancer stem cells and EMT in carcinoma. *Cancer Metastasis Rev.* (2012) 31:285–93. doi: 10.1007/s10555-012-9345-0
35. Visvader JE, Lindeman GJ. Cancer stem cells: current status and evolving complexities. *Cell Stem Cell.* (2012) 10:717–28. doi: 10.1016/j.stem.2012.05.007
36. Fedyanin M, Anna P, Elizaveta P, Sergei T. Role of stem cells in colorectal cancer progression and prognostic and predictive characteristics of stem cell markers in colorectal cancer. *Curr Stem Cell Res Ther.* (2017) 12:19–30. doi: 10.2174/1574888X11666160905092938
37. Pardee AB, Li CJ. Two controls of cell proliferation underlie cancer relapse. *J Cell Physiol.* (2018) 233:8437–40. doi: 10.1002/jcp.26597
38. Clarke MF, Dick JE, Dirks PB, Eaves CJ, Jamieson CH, Jones DL, et al. Cancer stem cells—perspectives on current status and future directions: AACR Workshop on cancer stem cells. *Cancer Res.* (2006) 66:9339–44. doi: 10.1158/0008-5472.CAN-06-3126
39. Li R, Wu R, Zhao L, Wu M, Yang L, Zou H. P-glycoprotein antibody functionalized carbon nanotube overcomes the multidrug resistance of human leukemia cells. *ACS Nano.* (2010) 4:1399–408. doi: 10.1021/nn9011225
40. Hermann PC, Bhaskar S, Cioffi M, Heeschen C. Cancer stem cells in solid tumors. *Semin Cancer Biol.* (2010) 20:77–84. doi: 10.1016/j.semcancer.2010.03.004
41. Liu DH, Wen GM, Song CL, Xu ZJ, Ren F, Zhao ZY, et al. The role of the stemness index-associated signature in the analysis of the tumorigenesis of liver cancer patients of different races. *Am J Cancer Res.* (2023) 13:802–17.
42. Zheng H, Xie J, Song K, Yang J, Xiao H, Zhang J, et al. StemSC: a cross-dataset human stemness index for single-cell samples. *Stem Cell Res Ther.* (2022) 13:115. doi: 10.1186/s13287-022-02803-5
43. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U. S. A.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
44. Zhao Z, Mu H, Feng S, Liu Y, Zou J, Zhu Y. Identification of biomarkers associated with hepatocellular carcinoma stem cell characteristics based on Co-expression network analysis of transcriptome data and stemness index. *Crit Rev Eukaryotic Gene Express.* (2022) 32:47–60. doi: 10.1615/CritRevEukaryotGeneExpr.2021039692
45. Qian D, Qian C, Ye B, Xu M, Wu D, Li J, et al. Development and validation of a novel stemness-index-related long noncoding RNA signature for breast cancer based on weighted gene Co-expression network analysis. *Front Genet.* (2022) 13:760514. doi: 10.3389/fgene.2022.760514
46. Zhang Y, Liu D, Li F, Zhao Z, Liu X, Gao D, et al. Identification of biomarkers for acute leukemia via machine learning-based stemness index. *Gene.* (2021) 804:145903. doi: 10.1016/j.gene.2021.145903
47. Greener JG, Kandathil SM, Moffat L, & Jones, D. T. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol.* (2022) 23:40–55. doi: 10.1038/s41580-021-00407-0
48. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell.* (2018) 173:338–354.e15. doi: 10.1016/j.cell.2018.03.034
49. Lian H, Han YP, Zhang YC, Zhao Y, Yan S, Li QF, et al. Integrative analysis of gene expression and DNA methylation through one-class logistic regression machine learning identifies stemness features in medulloblastoma. *Mol Oncol.* (2019) 13:2227–45. doi: 10.1002/1878-0261.12557
50. Zhang Y, Zhang X, Huang X, Tang X, Zhang M, Li Z, et al. Tumor stemness score to estimate epithelial-to-mesenchymal transition (EMT) and cancer stem cells (CSCs) characterization and to predict the prognosis and immunotherapy response in bladder urothelial carcinoma. *Stem Cell Res Ther.* (2023) 14:15. doi: 10.1186/s13287-023-03239-1
51. Pinto JP, Kalathur RK, Oliveira DV, Barata T, Machado RS, Machado S, et al. StemChecker: a web-based tool to discover and explore stemness signatures in gene sets. *Nucleic Acids Res.* (2015) 43:W72–77. doi: 10.1093/nar/gkv529
52. Ning B, Li W, Zhao W, Wang R. Targeting epigenetic regulations in cancer. *Acta Biochim Biophys Sin.* (2016) 48:97–109. doi: 10.1093/abbs/gmv116
53. Nator KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, et al. Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell.* (2012) 10:620–34. doi: 10.1016/j.stem.2012.02.013
54. Zhang Q, Sun S, Xie Q, Wang X, Qian J, Yao J, et al. FAM81A identified as a stemness-related gene by screening DNA methylation sites based on machine learning-accessed stemness in pancreatic cancer. *Epigenomics.* (2022) 14:569–88. doi: 10.2217/epi-2022-0098
55. Jian J, Yuan C, Ji C, Hao H, Lu F. DNA methylation-based subtypes of acute myeloid leukemia with distinct prognosis and clinical features. *Clin Exp Med.* (2023) 23:2639–49. doi: 10.1007/s10238-022-00980-4
56. Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, et al. A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Mol Oncol.* (2015) 9:555–68. doi: 10.1016/j.molonc.2014.10.012
57. Liu F, Qian J, Ma C. MPscore: A novel predictive and prognostic scoring for progressive meningioma. *Cancers.* (2021) 13:1113. doi: 10.3390/cancers13051113
58. Zhang B, He Y, Ma G, Zhang L, Qi P, Han D, et al. Identification of stemness index-related long noncoding RNA SNHG12 in human bladder cancer based on WGCNA. *Mol Cell Probes.* (2022) 66:101867. doi: 10.1016/j.mcp.2022.101867
59. Tan C, Wei Y, Ding X, Han C, Sun Z, Wang C. Cell senescence-associated genes predict the Malignant characteristics of glioblastoma. *Cancer Cell Int.* (2022) 22:411. doi: 10.1186/s12935-022-02834-1
60. Sun Z, Zhao Y, Wei Y, Ding X, Tan C, Wang C. Identification and validation of an aneuploidy-associated gene signature to predict clinical character, stemness, IDH mutation, and immune infiltration in glioblastoma. *Front Immunol.* (2022) 13:939523. doi: 10.3389/fimmu.2022.939523
61. Li N, Li Y, Zheng P, Zhan X. Cancer stemness-based prognostic immune-related gene signatures in lung adenocarcinoma and lung squamous cell carcinoma. *Front Endocrinol.* (2021) 12:755805. doi: 10.3389/fendo.2021.755805
62. Feng YD, Du J, Chen HL, Shen Y, Jia YC, Zhang PY, et al. Characterization of stem cell landscape and assessing the stemness degree to aid clinical therapeutics in hematologic Malignancies. *Sci Rep.* (2024) 14:23743. doi: 10.1038/s41598-024-74806-6
63. Wang Z, Wang Y, Yang T, Xing H, Wang Y, Gao L, et al. Machine learning revealed stemness features and a novel stemness-based classification with appealing implications in discriminating the prognosis, immunotherapy and temozolomide responses of 906 glioblastoma patients. *Briefings Bioinf.* (2021) 22:1–20. doi: 10.1093/bib/bbab032
64. Guo R, Chu A, Gong Y. Identification of cancer stem cell-related biomarkers in intestinal-type and diffuse-type gastric cancer by stemness index and weighted correlation network analysis. *J Transl Med.* (2020) 18:418. doi: 10.1186/s12967-020-02587-3
65. Sun X, Liu Q, Huang J, Diao G, Liang Z. Transcriptome-based stemness indices analysis reveals platinum-based chemo-therapeutic response indicators in advanced-stage serous ovarian cancer. *Bioengineered.* (2021) 12:3753–71. doi: 10.1080/21655979.2021.1939514
66. Huang K, Wu Y, Xie Y, Huang L, Liu H. Analyzing mRNA-related genes identifies novel prognostic markers and potential drug combination for patients with basal breast cancer. *Dis Markers.* (2021) 2021:4731349. doi: 10.1155/2021/4731349
67. Shi H, Han L, Zhao J, Wang K, Xu M, Shi J, et al. Tumor stemness and immune infiltration synergistically predict response of radiotherapy or immunotherapy and relapse in lung adenocarcinoma. *Cancer Med.* (2021) 10:8944–60. doi: 10.1002/cam4.4377
68. Feng T, Wu T, Zhang Y, Zhou L, Liu S, Li L, et al. Stemness analysis uncovers that the peroxisome proliferator-activated receptor signaling pathway can mediate fatty acid homeostasis in sorafenib-resistant hepatocellular carcinoma cells. *Front Oncol.* (2022) 12:912694. doi: 10.3389/fonc.2022.912694

69. Wang J, Ren H, Wu W, Zeng Q, Chen J, Han J, et al. Immune infiltration, cancer stemness, and targeted therapy in gastrointestinal stromal tumor. *Front Immunol.* (2021) 12:691713. doi: 10.3389/fimmu.2021.691713
70. Zheng S, Zheng D, Dong C, Jiang J, Xie J, Sun Y, et al. Development of a novel prognostic signature of long non-coding RNAs in lung adenocarcinoma. *J Cancer Res Clin Oncol.* (2017) 143:1649–57. doi: 10.1007/s00432-017-2411-9
71. Pei J, Wang Y, Li Y. Identification of key genes controlling breast cancer stem cell characteristics via stemness indices analysis. *J Transl Med.* (2020) 18:74. doi: 10.1186/s12967-020-02260-9
72. Guo SH, Ma L, Chen J. Identification of prognostic markers and potential therapeutic targets in gastric adenocarcinoma by machine learning based on mRNAsi index. *J Oncol.* (2022) 2022:8926127. doi: 10.1155/2022/8926127
73. Lyu C, Wang L, Stadlbauer B, Noessner E, Buchner A, Pohla H, et al. Identification of EZH2 as cancer stem cell marker in clear cell renal cell carcinoma and the anti-tumor effect of epigallocatechin-3-gallate (EGCG). *Cancers.* (2022) 14:4200. doi: 10.3390/cancers14174200
74. Zhao M, Chen Z, Zheng Y, Liang J, Hu Z, Bian Y, et al. Identification of cancer stem cell-related biomarkers in lung adenocarcinoma by stemness index and weighted correlation network analysis. *J Cancer Res Clin Oncol.* (2020) 146:1463–72. doi: 10.1007/s00432-020-03194-x
75. Tan J, Zhu H, Tang G, Liu H, Wanggou S, Cao Y, et al. Molecular subtypes based on the stemness index predict prognosis in glioma patients. *Front Genet.* (2021) 12:616507. doi: 10.3389/fgene.2021.616507
76. Tang R, Liu X, Wang W, Hua J, Xu J, Liang C, et al. Identification of the roles of a stemness index based on mRNA expression in the prognosis and metabolic reprogramming of pancreatic ductal adenocarcinoma. *Front Oncol.* (2021) 11:643465. doi: 10.3389/fonc.2021.643465
77. Luo Y, Xu W-B, Ma B, Wang Y. Novel stemness-related gene signature predicting prognosis and indicating a different immune microenvironment in HNSCC. *Front Genet.* (2022) 13:822115. doi: 10.3389/fgene.2022.822115
78. Li Y-R, Fang Y, Lyu Z, Zhu Y, Yang L. Exploring the dynamic interplay between cancer stem cells and the tumor microenvironment: implications for novel therapeutic strategies. *J Transl Med.* (2023) 21:686. doi: 10.1186/s12967-023-04575-9
79. Wu B, Shi X, Jiang M, Liu H. Cross-talk between cancer stem cells and immune cells: potential therapeutic targets in the tumor immune microenvironment. *Mol Cancer.* (2023) 22:38. doi: 10.1186/s12943-023-01748-4
80. Das S, Khan TH, Sarkar D. Comprehensive review on the effect of stem cells in cancer progression. *Curr Tissue Microenviron.Rep.* (2024) 5:39–59. doi: 10.1007/s43152-024-00053-6
81. Fu S, Tan Z, Shi H, Chen J, Zhang Y, Guo C, et al. Development of a stemness-related prognostic index to provide therapeutic strategies for bladder cancer. *NPJ Precis. Oncol.* (2024) 8:14. doi: 10.1038/s41698-024-00510-3
82. Zhang Z, Wang ZX, Chen YX, Wu HX, Yin L, Zhao Q, et al. Integrated analysis of single-cell and bulk RNA sequencing data reveals a pan-cancer stemness signature predicting immunotherapy response. *Genome Med.* (2022) 14:45. doi: 10.1186/s13073-022-01050-w
83. Sarkar H, Lee E, Lopez-Darwin SL, Kang Y. Deciphering normal and cancer stem cell niches by spatial transcriptomics: opportunities and challenges. *Genes Dev.* (2025) 39:64–85. doi: 10.1101/gad.351956.124
84. Nam AS, Chaligne R, Landau DA. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nat Rev Genet.* (2021) 22:3–18. doi: 10.1038/s41576-020-0265-5
85. Chen R, Liu Y, Xie J. Construction of a pathomics model for predicting mRNAsi in lung adenocarcinoma and exploration of biological mechanism. *Heliyon.* (2024) 10:e37100. doi: 10.1016/j.heliyon.2024.e37100
86. Wei R, Quan J, Li S, Liu H, Guan X, Jiang Z, et al. Integrative analysis of biomarkers through machine learning identifies stemness features in colorectal cancer. *Front Cell Dev Biol.* (2021) 9:724860. doi: 10.3389/fcell.2021.724860
87. Wu Z-H, Zhang Y-J, Jia C-L. Cancer stem cell characteristics by network analysis of transcriptome data stemness indices in breast carcinoma. *J Oncol.* (2020) 2020:8841622. doi: 10.1155/2020/8841622
88. Qiu H, Deng X, Zha J, Wu L, Liu H, Lu Y, et al. Machine learning-based characterization of stemness features and construction of a stemness subtype classifier for bladder cancer. *BMC Cancer.* (2025) 25:717. doi: 10.1186/s12885-025-14109-9
89. Zhou T, Man Q, Li X, Xie Y, Hou X, Wang H, et al. Artificial intelligence-based comprehensive analysis of immune-stemness-tumor budding profile to predict survival of patients with pancreatic adenocarcinoma. *Cancer Biol Med.* (2023) 20:196–217. doi: 10.20892/j.issn.2095-3941.2022.0569
90. Xiang R, Song W, Ren J, Wu J, Fu T. Identification of stem cell-related subtypes and risk scoring for gastric cancer based on stem genomic profiling. *Stem Cell Res Ther.* (2021) 12:563. doi: 10.1186/s13287-021-02633-x
91. Liu J, Jiang S, Wu Y, Zou R, Bao Y, Wang N, et al. Deep learning-based radiomics and machine learning for prognostic assessment in IDH-wildtype glioblastoma after maximal safe surgical resection: a multicenter study. *Int J Surg (lond. Engl.).* (2025) 111(7):4576–85. doi: 10.1097/JIS9.0000000000002488
92. Guan R, Zhang X, Guo M. Glioblastoma stem cells and wnt signaling pathway: molecular mechanisms and therapeutic targets. *Chin Neurosurg J.* (2020) 6:25. doi: 10.1186/s41016-020-00207-z
93. Rich JN. Cancer stem cells: understanding tumor hierarchy and heterogeneity. *Med (Baltimore).* (2016) 95:S2–7. doi: 10.1097/MD.0000000000004764
94. Liu M, Zhou R, Zou W, Yang Z, Li Q, Chen Z, et al. Machine learning-identified stemness features and constructed stemness-related subtype with prognosis, chemotherapy, and immunotherapy responses for non-small cell lung cancer patients. *Stem Cell Res Ther.* (2023) 14:238. doi: 10.1186/s13287-023-03406-4
95. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: a call for open science. *Patterns (n. Y. N.Y.).* (2021) 2:100347. doi: 10.1016/j.patter.2021.100347
96. Aliferis C, Simon G. Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI. In: Simon GJ, Aliferis C, editors. *Artificial intelligence and machine learning in health care and medical sciences: best practices and pitfalls.* Springer, Cham (CH) (2024).
97. Yuan H, Yu Q, Pang J, Chen Y, Sheng M, Tang W. The value of the stemness index in ovarian cancer prognosis. *Genes (Basel).* (2022) 13:993. doi: 10.3390/genes13060993
98. Sartori F, Codicè F, Caranzano I, Rollo C, Birolo G, Fariselli P. A comprehensive review of deep learning applications with multi-omics data in cancer research. *Genes (Basel).* (2025) 16:648. doi: 10.3390/genes16060648