Discover Oncology

RESEARCH Open Access

Identification of senescence-related genes as diagnostic biomarkers for gastric cancer using bioinformatics and machine learning

Xiaobo Li[1†], Zhenggen Piao[2†], Mengyue Lei[3], Dongyuan Xu[4], TouFeng Jin[2*] and Lan Liu[1*]

[†]Xiaobo Li and Zhenggen Piao contributed equally to this work (co-first author).

*Correspondence:
TouFeng Jin
tfjin@ybu.edu.cn
Lan Liu
lliu@ybu.edu.cn

Full list of author information is available at the end of the article

Abstract

Gastric cancer (GC) presents a significant global health challenge with a poor prognosis due to late detection. This study combines single-cell RNA sequencing and bulk transcriptomics to identify senescence-related gastric cancer genes (SGCGs) as diagnostic biomarkers for GC. Using weighted gene co-expression network analysis (WGCNA) and machine learning-based feature selection, we identified 20 core SAGs enriched in mitochondrial and cell cycle pathways. An RF-XGBoost ensemble model achieved high predictive accuracy (ROC = 0.841), with PNPT1 emerging as a key driver through SHAP analysis. Experimental validation confirmed overexpression of PNPT1 in GC cells, with its expression correlating with age-related progression. A web-based Shiny app was developed to support clinical risk stratification. These findings highlight the importance of SGCGs in GC development and offer a translational tool for early detection and personalized treatment.

Keywords Gastric cancer, Cellular senescence, Diagnostic biomarker, Machine learning, Bioinformatics

1 Background

Gastric cancer (GC) remains a significant global health challenge, ranking as the fifth most common malignancy and the fourth leading cause of cancer-related mortality worldwide [1]. The disease exhibits marked geographic variations, with the highest incidence rates observed in East Asia, Eastern Europe, and South America, where it accounts for roughly 1 in 12 cancer-related deaths [2]. Despite advances in surgical interventions, chemotherapy, targeted therapies, and immunotherapy, the five-year survival rate for GC lingers at 30–40%, largely attributable to late-stage diagnoses, tumor heterogeneity, and limited early detection strategies [3, 4]. This underscores the critical need for reliable biomarkers to enable early identification, risk stratification, and personalized therapeutic approaches.

Cellular senescence, a state of irreversible cell cycle arrest accompanied by a distinctive senescence-associated secretory phenotype (SASP), plays a dual role in cancer biology—acting as a tumor suppressor by halting proliferation while paradoxically

promoting tumorigenesis through microenvironment remodeling, immune evasion, and therapy resistance [5]. Emerging evidence links senescence-related pathways to age-associated diseases, including cardiovascular disorders, neurodegeneration, and various malignancies. In GC, senescence is increasingly regarded as a factor that contributes to cancer development, remodels the tumor microenvironment and affects the efficacy of immunotherapy [6]. Targeting senescence-related markers such as CBX4 demonstrates therapeutic potential in overcoming chemotherapy resistance and suppressing tumor progression in gastric cancer [7]. Furthermore, Senescence-associated genes (SAGs) may influence pathogenesis by modulating mitochondrial function, DNA damage responses, and inflammatory signaling, yet their specific prognostic and diagnostic utility remains underexplored [8].

The global aging population exacerbates the burden of age-related diseases, including cancer, with chronological age serving as an independent risk factor for GC. Traditional aging biomarkers, such as DNA methylation clocks, inflammatory markers, and microbiome profiles, offer insights into biological aging but fall short in capturing cancer-specific senescence dynamics [9]. High-throughput omics data, combined with artificial intelligence (AI) techniques like machine learning, provide a promising avenue for identifying robust SAG signatures [10]. AI-driven models have revolutionized age prediction using methylation and imaging datasets. In recent years, significant advancements in multi-omics data (including single-cell technologies) and systematic computational biology algorithms have broadened their potential applications—particularly in areas such as tumor biomarker prediction [11, 12]. While their application to multidimensional transcriptomic data in gastric cancer (GC) remains limited.

To bridge these gaps, this study integrates single-cell RNA sequencing (scRNA-seq) with bulk transcriptomics from GC cohorts to delineate senescence-related gastric cancer genes (SGCGs). Employing explainable machine learning, we developed a predictive model that correlates SGCGs expression with clinical outcomes, offering mechanistic insights into senescence-driven GC progression and a translational tool for risk assessment. This approach not only enhances understanding of senescence in GC but also supports precision medicine by outperforming conventional markers.

2 Material and method

2.1 Single-cell RNA sequencing analysis

Single-cell RNA sequencing data were obtained from the GEO database (accession number: GSE183904). We specifically isolated specimens' data from GSE183904 annotated with "Primary Normal" and "Primary Tumor" samples. Data preprocessing included quality control, normalization, and batch effect correction. Downstream analysis was performed using the Seurat R package. High-quality cells were retained with mitochondrial gene content below 20% and expressing between 200 and 7000 genes. The "Log Normalize" method was applied for normalization, and the top 3000 highly variable genes were identified. Dimensionality reduction was performed using Principal Component Analysis (PCA), followed by Uniform Manifold Approximation and Projection (UMAP) visualization based on the top 30 principal components. Cell clustering was conducted using the "FindClusters" function (resolution parameter = 0.8). Cell type annotation was performed based on canonical marker genes.

2.2 GEO transcriptomic microarray analysis

Bulk transcriptomic datasets (GSE26942, n = 217 and GSE27342, n = 160) were downloaded from the GEO database, including 377 gastric cancer samples and matched healthy controls. Data normalization and differential expression analysis were performed using the limma R package, with batch effect correction by algorithm. Differentially expressed genes (DEGs) were identified using thresholds of | Fold Change| > 1.5 and adjusted p-value < 0.05. Heatmaps and volcano plots were generated using the pheatmap and ggplot2 packages, respectively. A part of GSE26942 (20%) was separated out as an internal test set before model training. For robust validation of our predictive model, we conducted independent external validation cohorts sourced from two well-established gastric cancer datasets: GSE54129 (n = 132) and GSE66229 (n = 400).

2.3 Weighted gene co-expression network analysis (WGCNA)

A gene co-expression network was constructed using the WGCNA R package. Before WGCNA, we applied a variance filter (data = data [apply(data,1,sd) > 0.5,]) to remove genes with very low variability across samples. The soft threshold power was determined using the scale-free topology fit index ($R^2 > 0.9$). An adjacency matrix was constructed and transformed into a topological overlap matrix (TOM). Average linkage hierarchical clustering was employed to identify gene modules. Correlations between each module and gastric cancer phenotypes were calculated, and module genes significantly associated with GC were selected for further analysis.

WGCNA was performed on the training set expression matrix containing all genes (after filtering out those with low variance, SD < 0.5). Module detection was carried out using the automatic network construction function with a soft-threshold power of 10. Modules significantly correlated with gastric cancer (|cor| > 0.3, p < 0.05) were selected, and all genes within these modules were extracted for downstream analysis. The Venn diagram was constructed using three independent gene sets: (i) significantly differentially expressed genes (DEGs) identified by limma, (ii) genes from the gastric-cancer-associated WGCNA modules, and (iii) an external senescence-associated gene set (ASG).

2.4 Integration of senescence-associated gene datasets

Senescence-associated genes (SAGs) were systematically curated from two principal databases: CellAge [13–15] and SenMayo [16]. Gene symbols were extracted from each database (Gene symbol column for CellAge; Gene(human) column for SenMayo) and merged into a non-redundant union set using the unique() function in R (v4.2.1) to eliminate duplicates. This combined SAG list (termed ASG hereafter) was subsequently intersected with differentially expressed genes (DEGs) and WGCNA module genes to identify senescence-associated gastric cancer genes (SGCGs). No additional filtering was applied to resolve inter-database discrepancies, ensuring maximal inclusivity of senescence-related signals (Supplementary Table 1).

2.5 KEGG and GO enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using the clusterProfiler R package. Significantly enriched terms were identified with p-value < 0.05 and false discovery rate (FDR) < 0.1.

Results were visualized using the ggplot2 package, including dot plots, circle plots, and bar plots to display significantly enriched biological processes, molecular functions, and cellular components.

2.6 Machine learning feature engineering and model construction

A machine learning pipeline was established using the caret framework in R (version 4.1.3). The training set (merged GSE26942 and GSE27342 after batch correction) was first standardized using z-score normalization. An 80:20 stratified split was applied to create training and internal test sets, preserving the proportion of gastric cancer and control samples. Model performance was evaluated using repeated tenfold cross-validation (3 repeats) with the area under the receiver operating characteristic curve (AUC-ROC) as the primary optimization metric.

A systematic evaluation of 101 machine learning algorithm combinations was performed, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Logistic Regression (LR), as well as ensemble methods and additional algorithms (Lasso, PLS, LDA, Decision Tree, CatBoost, LightGBM). For each algorithm, hyperparameters were optimized exclusively on the training set using a combination of grid search (for small parameter spaces) and random search (for larger spaces) with up to 50 random combinations.

The hyperparameter search ranges for key algorithms were as follows:

XGBoost: nrounds (50–150), max_depth (3–7), eta (0.01–0.3), gamma (0–0.2), colsample_bytree (0.6–1.0), min_child_weight (1–5), subsample (0.6–1.0).
Random Forest: mtry (2–20)Lasso (glmnet): alpha fixed at 1 (Lasso), lambda automatically selected over 100 values.
CatBoost: iterations (100–500), depth (4–8), learning_rate (0.01–0.1).
LightGBM: num_leaves (15–63), max_depth (5–10), learning_rate (0.01–0.1), feature_fraction (0.6–1.0).

After tuning, the best-performing models were selected based on cross-validated AUC. An ensemble model combining RF and XGBoost (soft voting) was constructed to improve predictive robustness. Final model performance was assessed on the held-out internal test set and two independent external validation cohorts (GSE54129, GSE66229) using AUC-ROC, accuracy, sensitivity, and specificity.

2.7 SHAP interpretability analysis and feature importance ranking

The fastshap R package was employed to perform SHAP (Shapley Additive exPlanations) analysis for interpreting machine learning model predictions and addressing the "black-box" issue, which quantifying both the magnitude and directional impact (positive/negative contribution to GC prediction probability) of each feature. Global SHAP values represented the mean absolute influence across all samples, while local explanations visualized instance-specific effects. Biologically, positive SHAP values indicate the gene's elevated expression increases gastric cancer likelihood, whereas negative values suggest protective effects. Feature importance ranking was determined by mean absolute SHAP values.

2.8 Cell culture and western blot analysis

Human gastric mucosal epithelial cells (GES-1) and gastric carcinoma cells (HGC-27) were cultured under standard conditions. Total protein was extracted and quantified using the BCA method. Equal amounts of protein samples were separated by SDS-PAGE and transferred to PVDF membranes. After blocking with 5% skim milk, membranes were incubated overnight at 4 °C with primary antibodies against PNPT1 (Servicebio, GB111520-100, 1:1000 dilution) and ACTIN (Servicebio, GB11001-100, 1:5000 dilution). After TBST washing, membranes were incubated with horseradish peroxidase-conjugated secondary antibodies (1:5000 dilution) for 2 h at room temperature. Protein bands were detected using ECL chemiluminescence, and grayscale analysis was performed using ImageJ software.

2.9 Shiny-based interactive web application development

A web-based prediction platform was developed using the R Shiny framework (https://lixiaobogc.shinyapps.io/Aging_GC_Prediction_System/). The front-end interface included sliders for gene expression level input, risk prediction result display, and historical record query modules. The back-end integrated the trained machine learning model for real-time gastric cancer risk probability calculation. QR code generation technology was implemented to enable cross-device access, with compatibility for electronic medical record system integration.

2.10 Statistical analysis

All statistical analyses were performed using R software (version 4.1.3). For comparisons of continuous variables between two groups, Student's t-test (for normally distributed data) or Wilcoxon rank-sum test (for non-normal distributions) was applied, as appropriate. Statistical significance was defined as $p < 0.05$, and multiple testing corrections were performed using the false discovery rate (FDR) method (Benjamini–Hochberg procedure) where indicated. Differential expression analysis between gastric cancer and control samples was conducted using the limma package, with moderated t-statistics and empirical Bayes smoothing; resulting p-values were adjusted by FDR. In weighted gene co-expression network analysis (WGCNA), Pearson correlation coefficients between module eigengenes and clinical traits were computed, and their significance was evaluated using Student's t-distribution (corPvalueStudent). For single-cell RNA-seq data, differences in senescence module scores between tumor and normal epithelial cells were assessed with the two-sided Wilcoxon rank-sum test. Receiver operating characteristic (ROC) curve analysis was performed using the pROC package. Area under the curve (AUC) values were reported as descriptive measures of discriminative ability. When comparing the predictive performance of multiple machine-learning models, pairwise AUC differences were tested for statistical significance using the DeLong test (implemented in pROC).

Data visualization was carried out using R packages including ggplot2, pheatmap, ComplexHeatmap, and ggvenn. All statistical tests were two-tailed unless otherwise specified.

3 Results

3.1 Single-cell transcriptomics reveals elevated senescence-associated gene signatures in gastric cancer epithelium

After collecting and analyzing single-cell RNA sequencing data (GSE183904) from human gastric cancer tissues and healthy controls (Fig. 1A), a standardized processing pipeline (encompassing data acquisition, quality control assessment, mitochondrial gene filtering, and dimensionality reduction visualization) was applied, ultimately yielding 23 distinct cell clusters (Fig. 1B). Subsequently, we performed cell type annotation based on canonical marker genes, successfully identifying Mast cells, Fibroblasts, Chief cells, Macrophages, B cells, T cells, Epithelial cells, and Endothelial cells. We then conducted dimensionality reduction visualization for these fundamental cell-type populations (Fig. 1C-D).

Analysis of sample origins revealed that tumor tissues exhibited higher infiltration of immune cells and epithelial-derived tumor cells, along with fewer normal chief cells and fibroblasts (Fig. 1E, Figure S1A-D). Comparative analysis of cell type proportions revealed significantly higher canonical senescence marker CDKN1A in gastric cancer

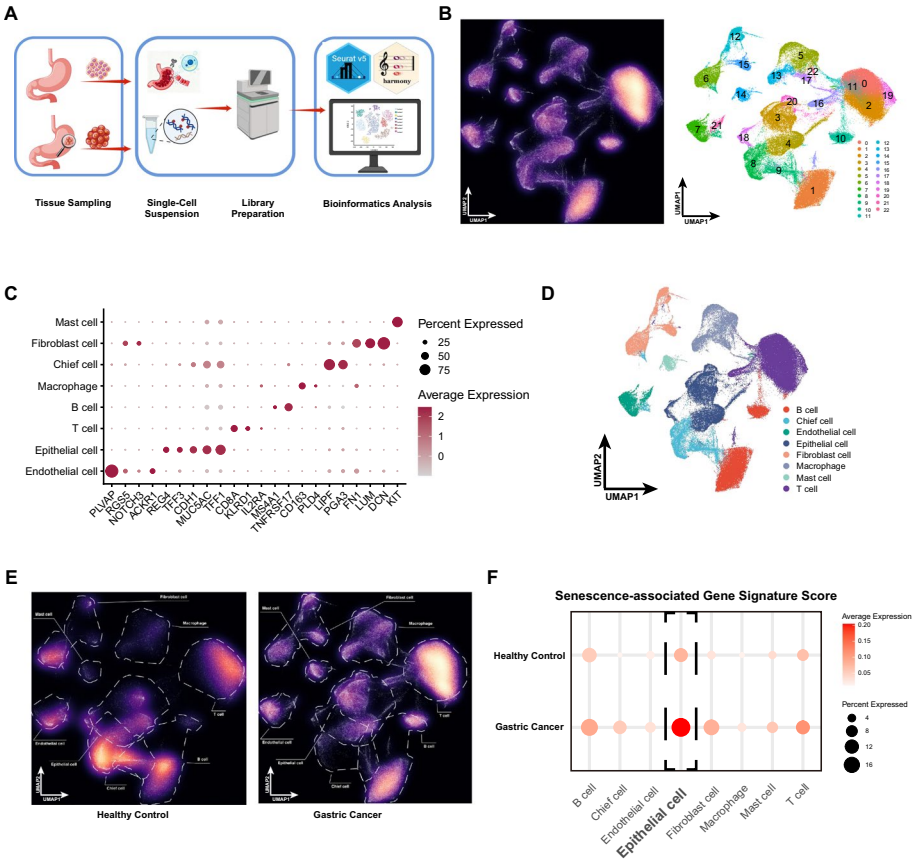

Fig. 1 Single-cell transcriptomics of senescence in gastric cancer epithelium. **A** Single-cell workflow diagram illustrating tissue sampling, single-cell suspension preparation, library construction, and bioinformatics analysis. **B** UMAP visualization showing basic clustering and density patterns in gastric cancer versus healthy control. **C** Dot plot displaying marker genes for different cell types. **D** UMAP dimensionality reduction plot with cell type annotations, clustered by expression profiles using UMAP. **E** Comparative analysis of cell type composition and density in gastric cancer versus healthy control. **F** Dot plot showing senescence-associated gene signature scores across different cell types in gastric cancer versus healthy control

epithelial cells, suggesting a potential role of cellular senescence in gastric carcinogenesis (Fig. 1F).

3.2 Transcriptomic sequencing identifies senescence-associated feature genes in gastric cancer

To further investigate gene alterations during gastric cancer progression, we performed a systematic integrative analysis of transcriptomic datasets (GSE26942 and GSE27342) with 377 gastric cancer samples and matched healthy controls. The standardized analytical pipeline included: data retrieval, batch effect correction, normalization, differential expression analysis, and gene module analysis (Fig. 2A, Figure S2A-D). Heatmap and volcano plot analyses revealed significant transcriptional differences between healthy and gastric cancer tissues (Fig. 2B). Using weighted gene co-expression network analysis (WGCNA), we identified 27 distinct gene expression modules. Among these, two

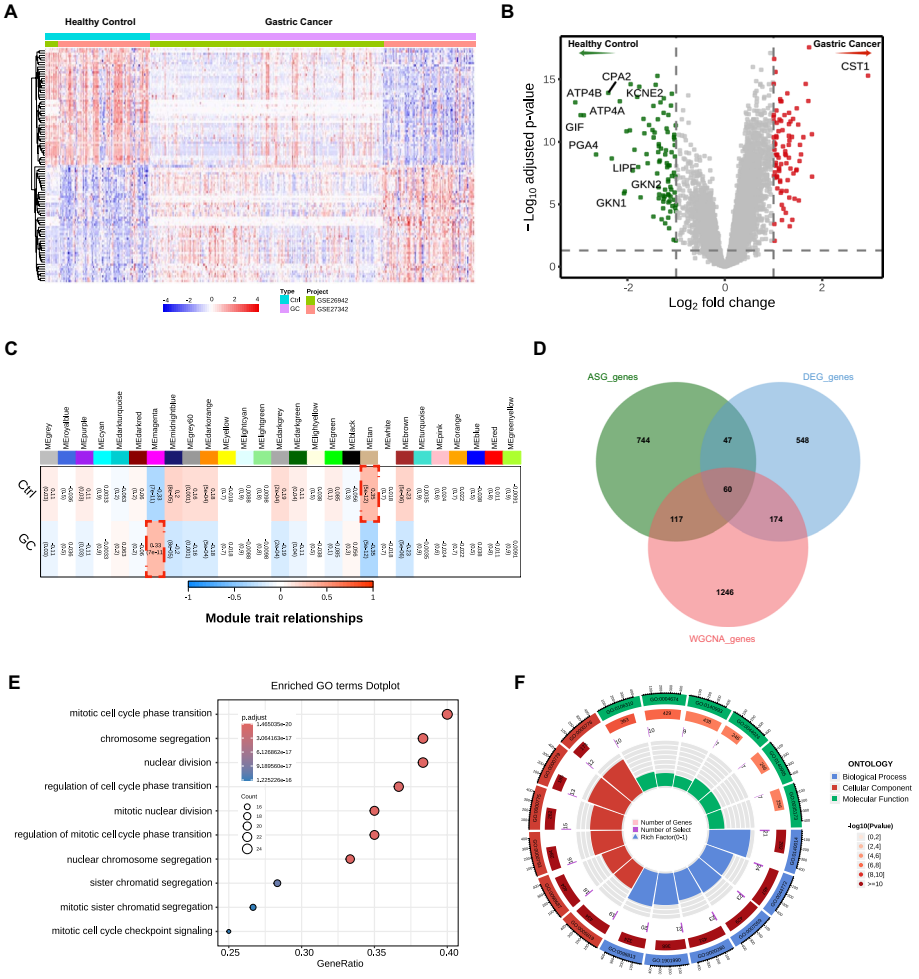

Fig. 2 Transcriptomic identification of senescence-associated genes in gastric cancer. **A** Heatmap of differentially expressed genes from integrated transcriptomic datasets. **B** Volcano plot of differentially expressed genes (DEGs) in GC samples and healthy samples. **C** WGCNA module-trait relationship heatmap highlighting modules with high correlation to phenotypic traits. **D** Venn diagram showing overlaps among senescence-associated genes (ASGs), DEGs, and WGCNA-identified genes. **E** Dot plot of enriched GO terms representing significant biological processes. **F** Circular plot of gene ontology terms categorized by biological process, molecular function, and cellular component

modules (METan and MEmagenta modules) showed strong phenotypic correlations with gastric cancer characteristics (Fig. 2C, Figure S2E-F). Through systematic curation of aging-related gene sets (ASGs) and intersection analysis with gastric cancer differentially expressed genes (DEGs) and WGCNA module genes, we further pinpointed over 60 Senescence-associated Gastric Cancer Genes (SGCGs) (Fig. 2D). Pathway enrichment analysis revealed that senescence-associated gastric cancer genes (SGCGs) were predominantly enriched in mitochondrial biological processes (Fig. 2E-F).

3.3 Machine learning-powered feature engineering pinpoints core senescence-driving oncogenes in gastric carcinogenesis

Building upon our initial set of senescence-associated gastric cancer genes (SGCGs), we employed machine learning-based feature selection to identify core driver genes. We established a machine learning pipeline comprising data standardization, train-test splitting, and batch model testing (Fig. 3A). Our analysis revealed partially overlapping feature parameters across different models (Fig. 3B-C). To address this issue, we performed

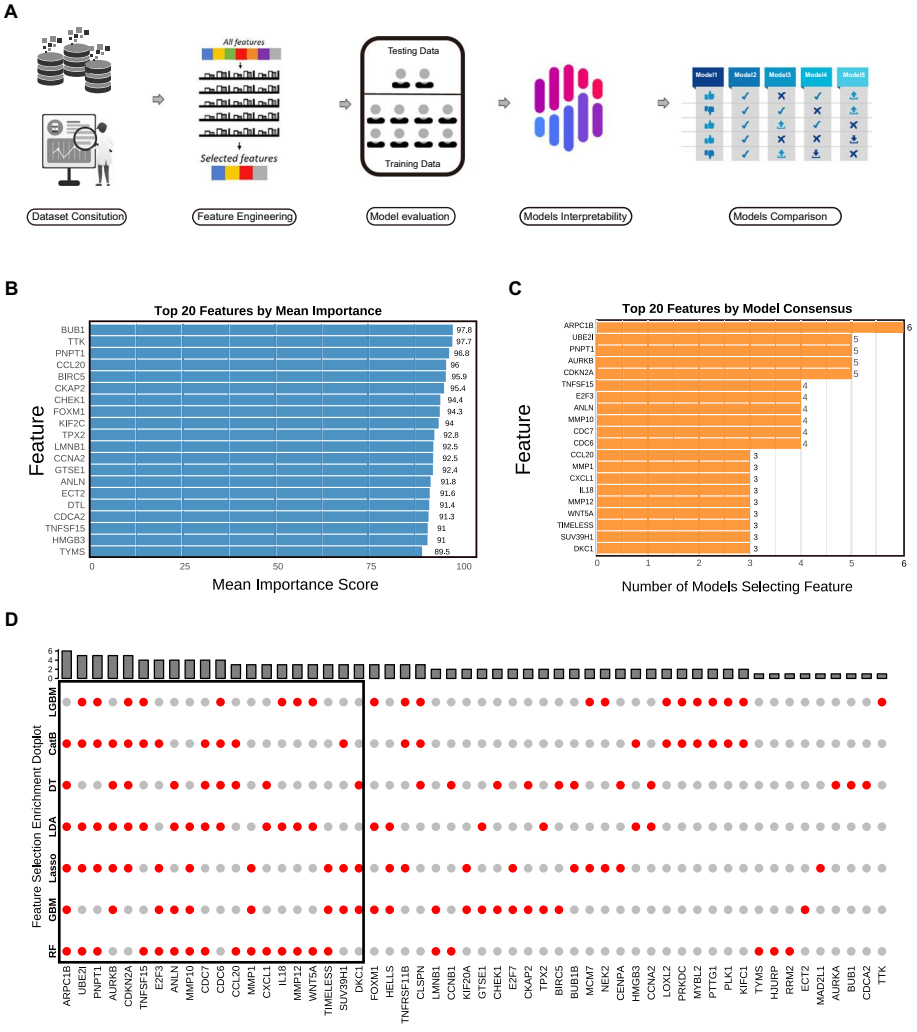

Fig. 3 Machine learning feature engineering for core senescence-driving genes. **A** Schematic of machine learning feature engineering pipeline. **B** Bar plot ranking the top 20 features by average importance score. **C** Consensus top 20 feature genes selected by multiple machine learning models. **D** Distribution and cumulative proportion of key genes across feature engineering approaches

dual importance ranking and visualized the expression patterns of these parameters within each model. Ultimately, we selected the top 20 predictive parameters as core SGCGs driver genes (Fig. 3D).

3.4 Interpretable machine learning constructs an optimal gastric cancer prediction model based on core SGCGs genes

To establish a robust predictive model, we systematically evaluated 101 machine learning algorithm combinations for constructing senescence-associated gastric cancer (SAGC) prediction models (Fig. 4A). Through comprehensive benchmarking involving classification performance tests and ROC analysis, the RF-XGBoost ensemble demonstrated superior predictive accuracy (mean ROC = 0.841, with individual genes reaching maximum ROCs of 0.936) (Fig. 4B-D). The final model incorporated all 20 top-ranking

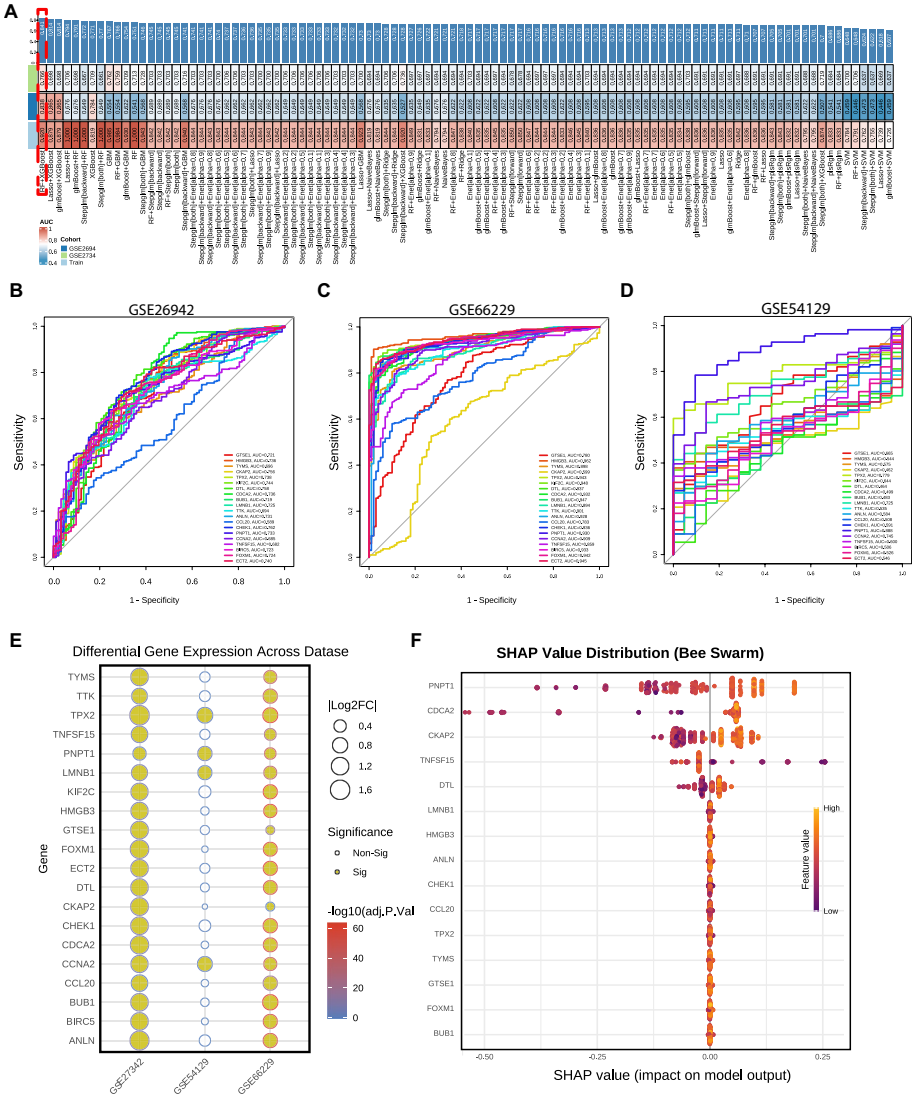

Fig. 4 Interpretable machine learning model for gastric cancer prediction. **A** Heatmap of mean ROC values across combinatorial machine learning approaches. **B** ROC curves comparing model performance in GSE26942 datasets. **C** ROC curves comparing model performance GSE66229 datasets. **D** ROC curves comparing model performance GSE54129 datasets. **E** Differential gene expression across datasets with significance annotations. **F** SHAP value distribution (beeswarm plot) showing individual gene contributions

parameters, which maintained exceptional predictive performance across both internal test sets and external validation cohorts (Fig. 4E). SHAP (SHapley Additive exPlanations) analysis further elucidated individual gene contributions, revealing PNPT1 emerged as the highest-impact risk factor (consistently positive SHAP values). SHAP analysis biologically contextualized these predictions by quantifying both effect sizes and directions. The beeswarm plot's horizontal dispersion further revealed expression-dependent effect modulation, where stronger PNPT1 overexpression correlated with disproportionately higher risk predictions (Fig. 4F).

3.5 Validation of core SGCGs and experimental confirmation of PNPT1

To substantiate the generalizability of the identified core senescence-associated gastric cancer genes (core SGCGs), we initially conducted validation studies utilizing single-cell transcriptomic datasets. These analyses demonstrated that the top 20 genes incorporated in our predictive model exhibited statistically significant overexpression in gastric epithelial tumor cells (Fig. 5A). Subsequent experimental validation efforts employed a

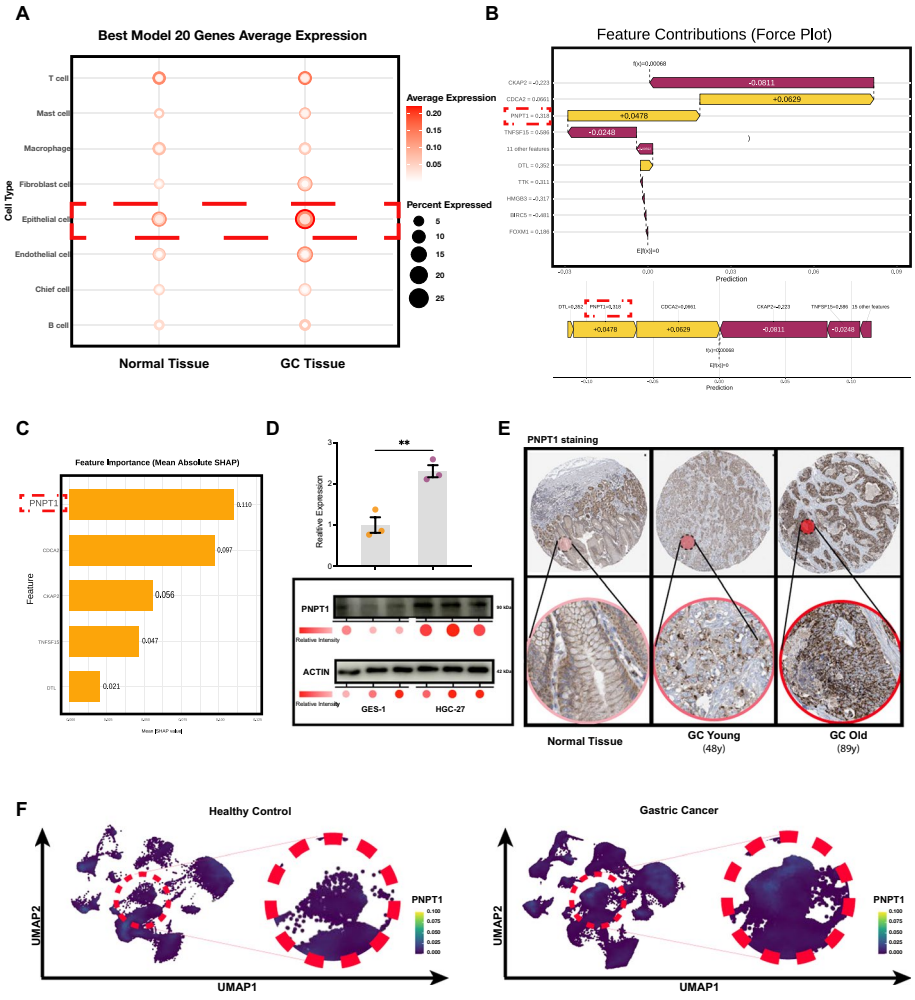

Fig. 5 Validation and experimental confirmation of core SGCGs. **A** Scatter plot comparing the average expression of the top 20 genes across normal and GC tissues by cell type. **B** Contribution force plot highlighting key predictive genes. **C** SHAP-based feature importance ranking. **D** PNPT1 detection in normal versus cancerous gastric cells. **E** Immunohistochemical comparison of PNPT1 protein expression in normal tissue versus GC cases (young: 48y; old: 89y). **F** UMAP plots overlayed with PNPT1 expression patterns

Fig. 6 Web-Based predictive system for senescence-associated gastric cancer Risk. **A** Interface of "Senescence-associated Gene-based Early Gastric Cancer Prediction System" showing gene expression input. **B** Interface of gastric cancer risk predictions. **C** Interface of historical prediction records. **D** Output interface displaying risk score, probability estimates, and system QR code

dual-criterion selection approach, evaluating both gene contribution metrics and mean absolute SHAP values, which collectively identified PNPT1 as the for further investigation (Fig. 5B-C). Comparative protein expression analyses were performed using three biologically relevant cellular models: primary gastric mucosal epithelial cells (GES-1), immortalized gastric epithelial cells, and HGC-27 gastric carcinoma cells. Western blot quantification revealed consistent PNPT1 overexpression across malignant cell populations (fold-change = 2.03, $p = 0.0056$ by Student's t-test) (Fig. 5D). Furthermore, immunohistochemical analysis of a cohort with age-related clinical parameters from the HPA database revealed that PNPT1 was not only highly expressed in gastric cancer but also exhibited an age-dependent expression trend (Fig. 5E). Single-cell analysis further verified PNPT1's clustered expression pattern in gastric cancer (Fig. 5F).

3.6 Web-based predictive system using an interpretable machine learning model

To facilitate the clinical translation of our predictive model, we developed an interactive website (https://lixiaobogc.shinyapps.io/Aging_GC_Prediction_System/), allowing users to conveniently input clinical parameters via adjustable sliders and receive instant risk stratification predictions (Fig. 6A-B). The system automatically archives historical prediction records for future reference (Fig. 6C). To optimize mobile accessibility, we implemented QR code functionality for seamless cross-device operation. This platform is designed for potential integration with electronic medical record systems to support clinical decision-making in real-world healthcare settings (Fig. 6D).

4 Discussion

In this comprehensive study, we employed an integrated multi-omics approach combining single-cell transcriptomics, bulk RNA sequencing, machine learning algorithms, experimental validation, and web-based tool development to systematically explore, screen, and validate core senescence-associated gastric cancer genes (SGCGs). The established predictive model was subsequently translated into a clinically applicable

web-based decision support tool designed to facilitate risk assessment and therapeutic decision-making in gastric cancer management.

Accurate and comprehensive evaluation and prediction tools are essential for informing treatment decisions and prognostic assessments in patients with early-stage gastric cancer (GC) [17]. Currently, there is no unified consensus on biomarkers for GC prediction, and while previous risk prediction tools for GC exist, their integration into clinical practice remains limited due to issues such as lack of validation across diverse populations and insufficient sensitivity for early detection [17–21]. The biological process of senescence in tumor initiation and progression remains incompletely understood. Recent research highlights the association between GC progression, and cellular senescence. Studies by Gu et al. have identified that compromised mitochondrial homeostasis can induce senescence and lead to the generation of antineoplastic oxidative stress, which showing senescence -induced new avenues for the treatment of GC [7, 22]. Given that age is an independent risk factor for senescence, we explored senescence gene expression patterns from this perspective to elucidate their role in GC development. Prior literature has demonstrated associations between senescence and other malignancies, such as colorectal and breast cancers, where senescence-related genes contribute to therapy resistance and metastatic potential [8, 23–27].

In the exploration of biomarkers, traditional differential expression analysis is inadequate for handling the explosion of multidimensional sequencing data [28, 29]. Integrative approaches combining bulk and single-cell transcriptomics are emerging as a transformative paradigm for biomarker discovery [11, 12], particularly in oncology research, where systematic application of advanced computational biology tools is becoming essential [30, 31].The adoption of more sophisticated feature identification methodologies has thus gained critical importance. In this study, machine learning-based feature engineering played a pivotal role by enabling the systematic selection of core SGCGs from high-dimensional datasets. Through algorithms like random forest and XGBoost, we prioritized genes based on importance scores, reducing noise and identifying predictive features that traditional methods might overlook. This approach not only enhanced model interpretability via SHAP values but also improved generalizability across datasets, as evidenced by high AUROC scores in both internal and external validations [28, 29, 32, 33].

Our feature gene set includes numerous genes related to cell cycle regulation, DNA synthesis/degradation, and genomic stability. For instance, CDCA2 (cell division cycle associated 2) has been implicated in promoting cell proliferation and tumorigenesis across various cancers, including hepatocellular carcinoma and colorectal cancer, where its overexpression correlates with poor prognosis and resistance to apoptosis by modulating pathways like PI3K/AKT [34–36]. Mitochondria are regarded as promising therapeutic targets in cancer treatment [37]. Similarly, PNPT1 (polyribonucleotide nucleotidyltransferase 1), a mitochondrial RNA-degrading enzyme, is upregulated in GC and other tumors, contributing to metabolic reprogramming and cisplatin resistance by enhancing mitochondrial oxidative phosphorylation and inhibiting apoptosis [38, 39]. GTSE1 (G2 and S-phase expressed 1), a microtubule-associated protein, facilitates cell migration and invasion in cancers such as prostate and breast, often via EMT promotion and correlation with advanced clinical stages, making it a marker for metastasis and poor survival [40, 41]. While our western blotting validation confirmed PNPT1's

overexpression in gastric cancer tissues, the functional mechanisms of PNPT1 in GC pathogenesis extend beyond the scope of this diagnostic modeling study. Nevertheless, we recognize the critical importance of elucidating its mechanistic roles to facilitate clinical translation. Future investigations should prioritize.

This study has several limitations. First, the validation cohorts could be expanded in size and diversity to further support the robustness of the core gene set across global populations. Second, experimental validations were conducted in limited cell lines, warranting broader testing in additional models, including patient-derived organoids. Therefore, more extensive clinical trials are needed to continually refine the predictive efficacy of the core gene set.

In summary, we constructed and validated a senescence feature-based GC-related gene prediction model for early GC prediction, which serves as a simple and convenient tool for early GC detection.

5 Conclusions

In conclusion, this study successfully integrates multi-omics data and explainable machine learning to identify and validate a signature of senescence-related genes (SGCGs) as robust diagnostic biomarkers for gastric cancer. We delineated the landscape of cellular senescence within the gastric tumor microenvironment at single-cell resolution and pinpointed a core set of 20 SGCGs, with PNPT1 emerging as a key driver. The developed RF-XGBoost ensemble model demonstrated high predictive accuracy, which was further substantiated by experimental validation confirming the overexpression and age-associated progression of PNPT1 in GC. The translation of these findings into a user-friendly, web-based Shiny application provides a practical tool for GC risk assessment, bridging the gap between computational discovery and clinical application. Our work not only underscores the critical role of senescence-related pathways in gastric carcinogenesis but also offers a novel, interpretable, and translatable framework for the early detection and personalized management of gastric cancer.

Abbreviations

AI	Artificial intelligence
AUC-ROC	Area under the receiver operating characteristic curve
DEGs	Differentially expressed genes
FDR	False discovery rate
GC	Gastric cancer
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
PCA	Principal component analysis
RF	Random forest
SAGs	Senescence-associated genes
SASP	Senescence-associated secretory phenotype
scRNA-seq	Single-cell RNA sequencing
SGCGs	Senescence-related gastric cancer genes
SHAP	SHapley Additive exPlanations
SVM	Support vector machine
TOM	Topological overlap matrix
UMAP	Uniform manifold approximation and projection
WGCNA	Weighted gene co-expression network analysis
XGBoost	Extreme gradient boosting

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s12672-026-04967-5.

Additional file1 (PDF 7001 KB)

Additional file2 (XLSX 20 KB)

Acknowledgements
We acknowledge the Gene Expression Omnibus (GEO) for hosting the dataset.

Author contributions
X.L. and Z.P. contributed equally to this work as co-first authors. Conceptualization: X.L., T.J., L.L.. Methodology: X.L., Z.P., M.L.. Formal analysis and investigation: X.L., Z.P., D.X.. Writing—original draft preparation: X.L. and Z.P.. Supervision: T.J., L.L.. All authors read and approved the final manuscript.

Funding
This study was supported by the Natural Science Research Foundation of Jilin Province for Sciences and Technology (Grant Numbers YDZJ202301ZYTS173 to Lan Liu and YDZJ202201ZYTS227 to Toufeng Jin) and the Project of the Education Department of Jilin Province, China (Grant Number JJKH20180910KJ to Toufeng Jin).

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. The original data supporting this research were derived from the GEO database under accession numbers GSE183904, GSE26942, GSE27342, GSE54129, and GSE66229.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors have read and approved the final version of the manuscript and consent to its publication. The manuscript has not been published previously and is not under consideration for publication elsewhere.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Pathology, The Affiliated Hospital of Yanbian University (Yanbian Hospital), Jilin, China
[2]Department of General Surgery, The Affiliated Hospital of Yanbian University (Yanbian Hospital), Jilin, China
[3]Department of Surgery, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China
[4]Center of Morphological Experiment, Medical College of Yanbian University, Jilin, China

Received: 28 November 2025 / Accepted: 6 April 2026
Published online: 14 April 2026

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Ca Cancer J Clin. 2024;74:229–63. https://doi.org/10.3322/caac.21834.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Ca Cancer J Clin. 2021;71:209–49. https://doi.org/10.3322/caac.21660.
3. Zeng Y, Jin RU. Molecular pathogenesis, targeted therapies, and future perspectives for gastric cancer. Semin Cancer Biol. 2022;86:566–82. https://doi.org/10.1016/j.semcancer.2021.12.004.
4. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. Lancet. 2020;396:635–48. https://doi.org/10.1016/S0140-6736(20)31288-5.
5. Ma L, Yu J, Fu Y, He X, Ge S, Jia R, et al. The dual role of cellular senescence in human tumor progression and therapy. Medcomm. 2024;5:e695. https://doi.org/10.1002/mco2.695.
6. Cao Z, Wang Z, Yang L, Li T, Tao X, Niu X. Reshaping the immune microenvironment and reversing immunosenescence by natural products: prospects for immunotherapy in gastric cancer. Semin Cancer Biol. 2025;110:1–16. https://doi.org/10.1016/j.semcancer.2025.02.002.
7. Gu Y, Xu T, Fang Y, Shao J, Hu T, Wu X, et al. CBX4 counteracts cellular senescence to desensitize gastric cancer cells to chemotherapy by inducing YAP1 SUMOylation. Drug Resist Updat. 2024;77:101136. https://doi.org/10.1016/j.drup.2024.101136.
8. Xuan J, Li F, Li J, Gong C, Li J, Mo Z, et al. The role of senescence genes in the treatment, prognosis, and tumor microenvironment of gastric cancer. Am J Transl Res. 2023;15:6926–38.
9. Levy JJ, Diallo AB, Saldias Montivero MK, Gabbita S, Salas LA, Christensen BC. Insights to aging prediction with AI based epigenetic clocks. Epigenomics. 2025;17:49–57. https://doi.org/10.1080/17501911.2024.2432854.
10. Sathe A, Grimes SM, Lau BT, Chen J, Suarez C, Huang RJ, et al. Single-cell genomic characterization reveals the cellular reprogramming of the gastric tumor microenvironment. Clin Cancer Res. 2020;26:2640–53. https://doi.org/10.1158/1078-0432.CCR-19-3231.
11. Li B, Lin R, Hua Y, Ma B, Chen Y. Single-cell RNA sequencing reveals TMEM71 as an immunomodulatory biomarker predicting immune checkpoint blockade response in breast cancer. Discov Oncol. 2025;16:1256. https://doi.org/10.1007/s12672-025-03068-z.

12. Chen Y, Anwar M, Wang X, Zhang B, Ma B. Integrative transcriptomic and single-cell analysis reveals IL27RA as a key immune regulator and therapeutic indicator in breast cancer. Discov Oncol. 2025;16:977. https://doi.org/10.1007/s12672-025-02811-w.
13. Avelar RA, Ortega JG, Tacutu R, Tyler EJ, Bennett D, Binetti P, et al. A multidimensional systems biology analysis of cellular senescence in aging and disease. Genome Biol. 2020;21:91. https://doi.org/10.1186/s13059-020-01990-9.
14. Tacutu R, Thornton D, Johnson E, Budovsky A, Barardo D, Craig T, et al. Human ageing genomic resources: new and updated databases. Nucleic Acids Res. 2018;46:D1083–90. https://doi.org/10.1093/nar/gkx1042.
15. de Magalhães JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics. 2009;25:875–81. https://doi.org/10.1093/bioinformatics/btp073.
16. Saul D, Kosinsky RL, Atkinson EJ, Doolittle ML, Zhang X, LeBrasseur NK, et al. A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. Nat Commun. 2022;13:4827. https://doi.org/10.1038/s41467-022-32552-1.
17. Hu B-L, Xie M-Z, Li K-Z, Li J-L, Gui Y-C, Xu J-W. Genome-wide analysis to identify a novel distant metastasis-related gene signature predicting survival in patients with gastric cancer. Biomed Pharmacother. 2019;117:109159. https://doi.org/10.1016/j.biopha.2019.109159.
18. Monahan K, Hampel H, Stoffel E, Kupfer SS. In our scope of practice: genetic risk assessment and testing for gastrointestinal cancers and polyposis in gastroenterology. Gastroenterology. 2025;169:1147–65. https://doi.org/10.1053/j.gastro.2025.06.001.
19. Kotelevets SM. Role of artificial intelligence in screening and medical imaging of precancerous gastric diseases. World J Clin Oncol. 2025;16:107993. https://doi.org/10.5306/wjco.v16.i9.107993.
20. Darvishi A, Dadashzadeh Asl A, Golshaniniya P, Alipour A, Naseri A, Hamzehzadeh S, et al. Serum markers and gastric cancer: an umbrella review. BMC Gastroenterol. 2025;25:737. https://doi.org/10.1186/s12876-025-04349-1.
21. Bharadwaj HR, Dahiya DS, Dalal P, Fuad M, Raza HA, Ibrahim M, et al. Artificial intelligence in population-level gastroenterology and hepatology: a comprehensive review of public health applications and quantitative impact. Dig Dis Sci. 2025. https://doi.org/10.1007/s10620-025-09452-7.
22. Hao X, Bu W, Lv G, Xu L, Hou D, Wang J, et al. Disrupted mitochondrial homeostasis coupled with mitotic arrest generates antineoplastic oxidative stress. Oncogene. 2022;41:427–43. https://doi.org/10.1038/s41388-021-02105-9.
23. He J, Fu Z, Zou B, Lei X, Lei L, Yang Q, et al. Identification the cellular senescence associated lncRNA LINC01579 in gastric cancer. J Cell Mol Med. 2025;29:e70360. https://doi.org/10.1111/jcmm.70360.
24. Cao L, Chen F, Xu L, Zeng J, Wang Y, Zhang S, et al. Prognostic cellular senescence-related lncRNAs patterns to predict clinical outcome and immune response in colon cancer. Front Immunol. 2024;15:1450135. https://doi.org/10.3389/fimmu.2024.1450135.
25. Hu Y, Li X, Ma L, Xu H, Zhu S, Liu A, et al. Construction and evaluation of a prognostic model for breast cancer based on aging related genes. Sci Rep. 2025;15:33756. https://doi.org/10.1038/s41598-025-98194-7.
26. Shi J, Hou Z, Fan L, Hu C, Ma N, Huang E. Development and experimental validation of a senescence-related long non-coding RNA signature for prognostic prediction and immune microenvironment characterization in gastric cancer patients. J Gastrointest Oncol. 2024;15:2413–36. https://doi.org/10.21037/jgo-24-792.
27. Dinami R, Petti E, Ostano P, Iachettini S, Rizzo A, Maresca C, et al. Multi-omics analyses integration reveals a novel TRF2-miR-181a-5p-S100A10 regulatory axis in colon cancer. Cell Mol Biol Lett. 2025;30:121. https://doi.org/10.1186/s11658-025-00813-8.
28. Phillips KA. Assessing the value of next-generation sequencing technologies: an introduction. Value Health. 2018;21:1031–2. https://doi.org/10.1016/j.jval.2018.06.012.
29. Matsuoka T, Yashiro M. Bioinformatics analysis and validation of potential markers associated with prediction and prognosis of gastric cancer. Int J Mol Sci. 2024;25:5880. https://doi.org/10.3390/ijms25115880.
30. Zhang B, Lau LY, Chen Y, Xie R. Bibliometric analysis of immune-related acute kidney injury induced by cancer immunotherapy (2000–2025). Naunyn Schmiedebergs Arch Pharmacol. 2026;399:2647–62. https://doi.org/10.1007/s00210-025-04582-1.
31. Li J, Wu Z, Pan Y, Chen Y, Chu J, Cong Y, et al. GNL3L exhibits pro-tumor activities via NF-κB pathway as a poor prognostic factor in acute myeloid leukemia. J Cancer. 2024;15:4072–80. https://doi.org/10.7150/jca.95339.
32. Song G, Wang L, Tang J, Li H, Pang S, Li Y, et al. Circulating metabolites as potential biomarkers for the early detection and prognosis surveillance of gastrointestinal cancers. Metabolomics. 2023;19:36. https://doi.org/10.1007/s11306-023-02002-0.
33. Chen Y, Wang B, Zhao Y, Shao X, Wang M, Ma F, et al. Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. Nat Commun. 2024;15:1657. https://doi.org/10.1038/s41467-024-46043-y.
34. Lin X, Zou Z, Zhong J, Wang T, Ma W, Hu T, et al. The role of CDCA2 in tumor genesis, prognosis and future treatments. Eur J Cancer. 2024;211:114308. https://doi.org/10.1016/j.ejca.2024.114308.
35. Kluger MT. Minilaparotomy under local infiltration in a high risk case. Anaesth Intensive Care. 1993;21:247.
36. Wang S, Cao K, Liao Y, Zhang W, Zheng J, Li X, et al. CDCA2 protects against oxidative stress by promoting BRCA1-NRF2 signaling in hepatocellular carcinoma. Oncogene. 2021;40:4368–83. https://doi.org/10.1038/s41388-021-01855-w.
37. Liu Y, Chen Y, Wang F, Lin J, Tan X, Chen C, et al. Caveolin-1 promotes glioma progression and maintains its mitochondrial inhibition resistance. Discover Oncol. 2023;14:161. https://doi.org/10.1007/s12672-023-00765-5.
38. Sasahara M, Kanda M, Tanaka C, Shimizu D, Umeda S, Takami H, et al. Therapeutic antibody targeting natriuretic peptide receptor 1 inhibits gastric cancer growth via BCL-2-mediated intrinsic apoptosis. Int J Cancer. 2024;154:1272–84. https://doi.org/10.1002/ijc.34831.
39. Brat DJ, Scheithauer BW, Medina-Flores R, Rosenblum MK, Burger PC. Infiltrative astrocytomas with granular cell features (granular cell astrocytomas): a study of histopathologic features, grading, and outcome. Am J Surg Pathol. 2002;26:750–7. https://doi.org/10.1097/00000478-200206000-00008.
40. Lai W, Zhu W, Li X, Han Y, Wang Y, Leng Q, et al. GTSE1 promotes prostate cancer cell proliferation via the SP1/FOXM1 signaling pathway. Lab Invest. 2021;101:554–63. https://doi.org/10.1038/s41374-020-00510-4.

41. Yang H, Toy EC, Baker B. Sexual dysfunction in the elderly patient. Prim Care Update Ob Gyns. 2000;7:269–74. https://doi.org/10.1016/s1068-607x(00)00058-5.

Publisher's Note