



Research article

CENPE is a diagnostic and prognostic biomarker for cervical cancer

Peiqiang Peng^{a,1}, Jingying Zheng^{b,1}, Kang He^a, Kai Wang^a, Longyun Wang^a,
Xufei Zheng^a, Hao Wu^a, Zhenning Yang^a, Shuang Zhang^{a,*}, Lijing Zhao^{a,**}

^a Department of Medical Rehabilitation, School of Nursing, Jilin University, 965 Xinjiang Street, Chaoyang District, Changchun, 130021, China

^b Department of Gynecology and Obstetrics, The Second Hospital of Jilin University, No.4026, Yatai Street, Changchun City, 130000, Jilin Province, China

ARTICLE INFO

Keywords:

CENPE
CESC
Diagnostic
Prognostic
Biomarker

ABSTRACT

Cervical squamous cell carcinoma (CESC) is a common cancer in women. Despite advancements in early diagnosis through high-risk human papillomavirus (HPV) screening, challenges remain in predicting and treating the disease. Hence, the identification of novel biomarkers for prognosis and therapeutic targets is crucial.

CENPE, a microtubule-end directed motor protein that accumulates during the G2 phase, is recognized for its involvement in promoting cancer growth and progression. However, its specific role in CESC remains unclear. This research investigated the expression of CENPE in CESC utilizing data from The Cancer Genome Atlas (TCGA), which was further validated through gene expression profiles, the Human Protein Atlas (HPA), and clinical data. The study utilized Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Set Enrichment Analysis (GSEA), and immune infiltration analysis to elucidate the role of CENPE in CESC. Additionally, Protein-protein interaction (PPI) networks and competing endogenous RNA (CeRNA) networks involving CENPE and its differentially expressed genes were established. Furthermore, Kaplan-Meier survival analysis was conducted to evaluate the impact of CENPE on patient prognosis.

Our study revealed an upregulation of CENPE expression in cervical cancer tissues, which promotes the progression of CESC through IL-6-mediated PI3K-Akt and MAPK signaling pathways. The significant associations with ACNG3, LY6H, and SLC6A7 suggest that CENPE may play a role in tumor growth and metastasis, potentially involving the nervous system. Moreover, the correlations with ARIH1, KDM1A, KDM5B, and NSD3 indicate that CENPE could be a promising target for drug development. Our analysis of the ROC curve demonstrated a high diagnostic accuracy of CENPE in CESC (AUC: 0.997, CI: 0.990–1.000). Subgroup analysis highlighted substantial effects in patients under 50 years old, those with a height under 160 cm, individuals in peri- and post-menopausal stages, and patients in clinical stages 1 and 4. Additionally, COX regression analysis indicated that older age, lower BMI, and higher CENPE expression are associated with decreased 1-year, 3-year, and 5-year survival rates.

In conclusion, CENPE emerges as a crucial factor in the initiation and advancement of cervical cancer, showing potential as a novel target for therapeutic interventions.

* Corresponding author.

** Corresponding author.

E-mail addresses: Zhangs@jlu.edu.cn (S. Zhang), Zhao_lj@jlu.edu.cn (L. Zhao).

¹ These authors have equally contribution for this work.

Abbreviations:

ACC	adrenocortical carcinoma
AOD	average optical density
AUC	Area Under the Curve
AML	acute myeloid leukemia
BLCA	bladder urothelial carcinoma
BLBC	basal-like breast cancer
BRCA	breast invasive carcinoma
CENP	centromere protein
CENPE	Centromeric protein E
CESC	Cervical squamous cell carcinoma
CeRNA	competing endogenous RNA
CHOL	cholangiocarcinoma
CI	Confidence Interval
COAD	colon cancer
DSS	disease-specific survival
ESCA	esophageal carcinoma
FIGO	Federation of Gynecology and Obstetrics
GBM	glioblastoma multiforme
GSEA	Gene Set Enrichment Analysis
HNSC	head and neck squamous cell carcinoma
HPA	Human Protein Atlas
HPV	human papillomavirus
KEGG	Kyoto Encyclopedia of Genes and Genomes
KICH	chromophobe tumor of the kidney
KIRC	kidney renal clear cell carcinoma
KIRP	kidney renal papillary cell carcinoma
LGG	low-grade glioma
LIHC	liver hepatocellular carcinoma
LUAD	lung adenocarcinoma
LUSC	lung squamous cell carcinoma
OS	overall survival
OV	ovarian cancer
PAAD	pancreatic adenocarcinoma
PCPG	pheochromocytoma and paraganglioma
PFI	progression-free interval
PPI	Protein-protein interaction
PRAD	prostate adenocarcinoma
READ	rectum adenocarcinoma
RTKs	receptor tyrosine kinases
SAC	spindle assembly checkpoint
SKCM	skin cutaneous melanoma
siRNA	small interfering RNA
ssGSEA	single-sample gene set enrichment analysis
TCGA	The Cancer Genome Atlas
THCA	thyroid carcinoma
THYM	thymoma
UCEC	endometrial adenocarcinoma
UCS	uterine carcinosarcoma

1. Introduction

CESC is a prevalent malignant tumor globally [1], with over 500,000 new cases and more than 274,000 deaths reported annually [2]. The survival rates for CESC are inversely related to the International Federation of Gynecology and Obstetrics (FIGO) stage at the time of diagnosis [3]. While HPV vaccines have shown significant efficacy in reducing cervical cancer incidence [4,5], HPV infection alone does not explain all cases of CESC [6]. Current prognosis prediction relies heavily on tumor-node-metastasis (TNM) staging, which may not fully encompass the biological diversity of CESC [7]. This underscores the critical necessity for identifying new biomarkers that can serve as both therapeutic targets and prognostic indicators.

Centromeric protein E (CENPE), a member of the centromere protein (CENP) family [8], functions as a microtubule plus-end directed motor protein that is upregulated during the G2 phase [9,10]. It plays a crucial role in the spindle assembly checkpoint (SAC), chromosome alignment, and microtubule capture at kinetochores during mitosis [11]. CENPE is responsible for ensuring the proper attachment of chromosomes to spindle microtubules and regulating spindle checkpoint signaling to prevent chromosome missegregation, thereby preserving the accuracy and stability of cell division [12,13]. Dysregulation of CENPE can result in chromosomal instability, a significant contributor to the development of various cancers [14,15].

Research has shown that rapidly dividing cells often display increased levels of CENPE expression, which is upregulated in various solid tumors [16,17]. Examples of such tumors include esophageal adenocarcinoma [18], clear cell renal cell carcinoma [19], colorectal cancer [20], and lung adenocarcinoma [8]. Moreover, heightened CENPE expression has been linked to unfavorable outcomes in breast cancer [21] and has the potential to facilitate the advancement of ovarian cancer [22]. Studies on prostate cancer have demonstrated that the genetic removal or pharmacological inhibition of CENPE can lead to a significant reduction in tumor growth [23]. Furthermore, the suppression of CENPE in human neuroblastoma cell lines through the use of small interfering RNA (siRNA) has been found to impede cell proliferation [24]. These observations suggest that CENPE may have a role in the development and progression of cancer.

Despite the existing insights, the precise function of CENPE in CESC has not been thoroughly investigated. This study seeks to fill this gap by investigating the role of CENPE in CESC using bioinformatics methods and evaluating its influence on patient survival outcomes. The objective is to construct a prognostic model that overcomes the current challenges in the diagnosis, treatment, and prognosis prediction of CESC.

2. Materials and methods

2.1. Design of the study

CENPE is linked to various tumor progression, but its role in CESC is unreported. This study examines CENPE's influence on CESC.

Firstly, this study demonstrates that CENPE expression is significantly elevated in tumors, particularly in CESC: 1) Pan-cancer paired and unpaired analyses reveal a notable increase in CENPE across multiple tumor types, including CESC. 2) Further validation from various sources, including the GEO database, the HPA database, and collected clinical samples, confirms the elevation of CENPE in CESC.

Secondly, a more in-depth exploration of the biological functions and pathways influenced by CENPE: 1) We conduct single-gene analyses to identify molecular sets significantly associated with CENPE variation in CESC (differential gene sets). 2) The differential gene sets undergo GO, KEGG, and GSEA analyses to determine the biological functions and pathways affected by CENPE that are related to tumor occurrence and development.

Thirdly, identify key molecules in the differential gene sets: 1) Use PPI networks to find core and strongly correlated molecules affecting CENPE function in CESC and their relevance to tumor processes. 2) Build the CeRNA regulatory network of these molecules to confirm their regulatory interactions.

Fourthly, evaluate CENPE's impact on the tumor immune microenvironment: 1) Perform correlation and stratification analyses based on CENPE expression, followed by ssGSEA, to show its significant influence. 2) Analyze correlations with common immune targets to validate its effect on tumor immune processes.

Fifthly, confirm that CENPE significantly impacts CESC patients and can be targeted for diagnosis and treatment: 1) Analyze CENPE expression across disease stages to establish its harmful overexpression in CESC patients. 2) Illustrate these effects through survival curve analyses. 3) Identify patient subgroups most affected by CENPE and those with better treatment outcomes. 4) Use TCGA and GEO databases for validation, showing CENPE's high sensitivity and specificity as a diagnostic marker.

Sixthly, Perform univariate and multivariate regression analyses to identify factors influencing CESC patients: 1) Assess independent effects and CENPE's influence on these factors. 2) Verify if CENPE jointly affects outcomes and validate further.

Seventhly, Create a clinical prediction model based on regression results: 1) Develop a survival prediction model by integrating various factors. 2) Assess its accuracy.

2.2. Differential expression analysis of CENPE

The RNA-seq differential expression data of CENPE in pan-cancer were acquired from UCSC XENA (<https://xenabrowser.net/datapages/>) and processed uniformly in TPM (Transcripts per Million Reads) format [25]. For the TCGA UCEC project, both paired and unpaired RNA sequencing data for CENPE were initially provided in Level 3 HTSeq-FPKM (Fragments per Kilobase per Million Reads) format, which were subsequently converted to TPM format through log2 transformation. Validation datasets GSE7410 and GSE9705 were retrieved using the GEOquery package (version 2.68.0) [26]. Data normalization was carried out utilizing the limma package (version 3.56.2), with only the probe exhibiting the highest signal value retained in cases where multiple probes corresponded to the same molecule. All statistical analyses and visualizations were executed using R (version 4.3.3).

2.3. Analysis of immunohistochemical staining

Immunohistochemical staining images of CENPE in CESC and normal tissue sections were obtained from the Human Protein Atlas (HPA) database (<https://www.proteinatlas.org/>). The sections were stained using standardized antibodies and experimental

protocols.

2.4. Single-gene differential analysis of CENPE in CESC

Single-gene differential analysis was performed on RNA-seq data obtained from the CESC project of TCGA (<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>) utilizing the DESeq2 package (version 1.42.1) [27]. Volcano plots were created based on the results of the single-gene differential analysis using ggplot2 (version 3.5.0), with criteria set at $|\log_2(\text{FC})| > 1$ and $p.\text{adj} < 0.05$.

2.5. Functional enrichment analysis of CENPE in CESC

Functional enrichment analyses, including GO, KEGG, and GSEA, were performed on the results from single-gene differential analysis using the clusterProfiler package (version 4.10.1) [28]. Gene ID conversion was handled with the org.Hs.eg.db package (version 3.17.0). Z scores were calculated using the GOpilot package (version 1.0.2) to assess the relevance of CENPE to the enrichment pathways. GSEA utilized the curated gene set c2.cgp.v7.2.symbols.gmt [29], with significant enrichment defined by criteria of FDR < 0.25 and $p.\text{adjust} < 0.05$. Visualization of the analysis results was conducted using ggplot2 (version 3.5.0).

2.6. Correlation analysis of CENPE

Single-gene correlation analysis of expression profile data in TPM format was conducted using the STAT package (version 0.1.0). Differentially expressed genes were analyzed for PPI using the STRING database [30], with network analysis performed using Cytoscape software. HUB genes were identified with the MCODE plugin, and correlation analysis was carried out between these HUB genes and CENPE. Results were visualized using heatmaps and chord diagrams with ggplot2 (version 3.5.0).

HUB genes that showed significant correlation with CENPE ($p < 0.01$) were used to construct a lncRNA-miRNA-mRNA network based on CLIP-seq data from the ENCORI database [31]. Visualization of the network was adjusted using Cytoscape.

2.7. Analysis of immunoinfiltration in CENPE

The relative infiltration levels of 24 immune cell types were evaluated using the GSVA software (version 1.51.17) through single-sample gene set enrichment analysis (ssGSEA) and Spearman correlation analysis [32]. The markers for the analysis were obtained from immunoprofiling studies [33]. Subsequently, the samples were categorized into groups based on CENPE down-regulation and up-regulation. The GSVA software (version 1.51.17) was employed to compute the enrichment percentage of various immune cell infiltrations within each subgroup. Finally, the association between immune cells that exhibited statistically significant infiltration ($p < 0.001$) and CENPE was depicted using chord diagrams and heatmaps.

2.8. Clinical information, survival analysis, and receiver operating characteristic (ROC) curves

Clinical data sourced from the TCGA database was utilized for analysis. Box plots were generated using the boxplot package in R. Survival analysis of CESC patients was conducted using the survival package (version 3.6.4), and Kaplan-Meier survival curves for overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI) were created using the survminer package (version 0.4.9). The prognostic accuracy of CENPE was evaluated through ROC analysis using the pROC package (version 1.18.5), with validation carried out in the GSE7410 and GSE9750 databases. Subsequently, subgroup analysis was performed on variables such as age, height, and clinical stage in CESC patients, and overall survival curves were plotted.

2.9. Regression analysis and survival prediction

Univariate and multivariate Cox regression analyses were performed utilizing the survival package (version 3.6.4) to evaluate various clinicopathological factors and CENPE expression. The most significant p-value was selected as the threshold for CENPE expression. Subsequently, a diagnostic model was developed based on the Cox regression analysis results using the rms package (version 6.8.0) and survival package (version 3.6.4). Calibration plots were then generated to evaluate the accuracy of the diagnostic model.

2.10. Immunohistochemical staining

Immunohistochemical staining was conducted in triplicate for each group. The samples were fixed in 10 % formalin at room temperature for 24 h, then sectioned to a thickness of 3 μm and embedded in paraffin. Following deparaffinization, the sections underwent antigen retrieval using EDTA buffer in a microwave. To prevent nonspecific binding, the sections were treated with 5 % bovine serum albumin for 20 min at room temperature. The rabbit anti-App antibody (1:200, AF6084, Affinity) was left overnight at 4 °C. Subsequently, the secondary antibody, goat anti-rabbit IgG conjugated with horseradish peroxidase (GB23204; 1:200; Servicebio), was applied for 30 min at 37 °C. Staining was visualized using 3,3'-diaminobenzidine (Boster Biological Technology, Inc.), and the sections were counterstained with 0.1 % hematoxylin (Boster Biological Technology, Inc.) for 2 min. Images were captured with a light microscope at $\times 200$ or $\times 400$ magnification. The quantification of positive cell density was performed using Image-Pro Plus 6.0

(Media Cybernetics, Inc.), and the results were reported as average optical density (AOD) values.

2.11. Statistical analysis

Data were reported as mean \pm standard deviation (SD). Statistical analyses were performed utilizing SPSS 26.0 and R version 4.3.3. Student's t-test was employed to evaluate variances in CENPE expression between CESC tumor tissues and adjacent tissues. One-way ANOVA was utilized for comparisons across multiple groups. The correlation between CENPE expression and clinical data of CESC patients was examined using the Mann-Whitney *U* test. Statistical significance was defined as a p-value of <0.05 .

3. Results

3.1. Expression of CENPE in Pan-Cancer and Cervical Cancer

Using the XENA database for pan-cancer non-paired analysis, it was observed that CENPE exhibits overexpression in various cancer types when compared to normal tissues. These cancer types include adrenocortical carcinoma (ACC), bladder urothelial carcinoma (BLCA), CESC, endometrial adenocarcinoma (UCEC), breast invasive carcinoma (BRCA), cholangiocarcinoma (CHOL), colon cancer (COAD), basal-like breast cancer (BLBC), esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), chromophobe tumor of the kidney (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), acute myeloid leukemia (AML), low-grade glioma (LGG), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), ovarian cancer (OV), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), skin cutaneous melanoma (SKCM), thyroid carcinoma (THCA), thymoma (THYM), and uterine carcinosarcoma (UCS) (Fig. 1A).

In paired samples, elevated levels of CENPE were detected in the following types of cancer in comparison to adjacent tissues: BLCA, BRCA, CHOL, COAD, ESCA, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, READ, STAD, and UCEC (Fig. 1B).

In unpaired samples, the expression of CENPE was consistently elevated in tumor tissues in comparison to normal tissues (Fig. 1C). The validation datasets GSE9750 and GSE9701 also supported the increased levels of CENPE in tumor tissues (Fig. 1D and E). Additionally, CENPE exhibited high expression in tumor tissues irrespective of the lymph node metastasis status (Fig. 1E).

Analysis of the HPA database indicated that the CENPE protein was not detectable in normal tissues but exhibited high expression in tumor tissues (Fig. 1F). Immunohistochemical staining showed a slight elevation in CENPE levels in precancerous lesions, with a notable upregulation in both high-grade and low-grade CESC (Fig. 1G and H).

3.2. Single-gene differential analysis and functional enrichment analysis of CENPE

Through a single-gene differential analysis of CENPE, it was determined that 612 genes satisfied the threshold criteria of $|\log_2(\text{FC})| > 1$ and $p.\text{adj} < 0.05$. Within this set, 519 genes exhibited high expression levels, while 93 genes showed low expression levels (refer to Fig. 2A). Following this analysis, further investigations were carried out through GO, KEGG, and GSEA on these identified genes.

The GO analysis unveiled that the differentially expressed genes are implicated in various biological processes. These processes include the immune system (specifically the humoral immune response), tumor invasion and metastasis, tumor microenvironment, cell polarity and morphology (such as extracellular structure organization, apical part of the cell, and apical plasma membrane), tumor metabolism, degradation, growth, and organic acid transport (including negative regulation of hydrolase activity, response to nutrient levels, and organic acid transport), tumor signaling transduction (involving response to tumor necrosis factor and receptor ligand activity), enzyme activity (comprising endopeptidase activity, enzyme inhibitor activity, serine-type endopeptidase activity, and serine-type endopeptidase inhibitor activity), cytokines (such as cytokine activity and growth factor activity), electrophysiology, matrix components, and hormones (including voltage-gated ion channel activity, collagen-containing extracellular matrix, and hormone activity) as illustrated in Fig. 2B.

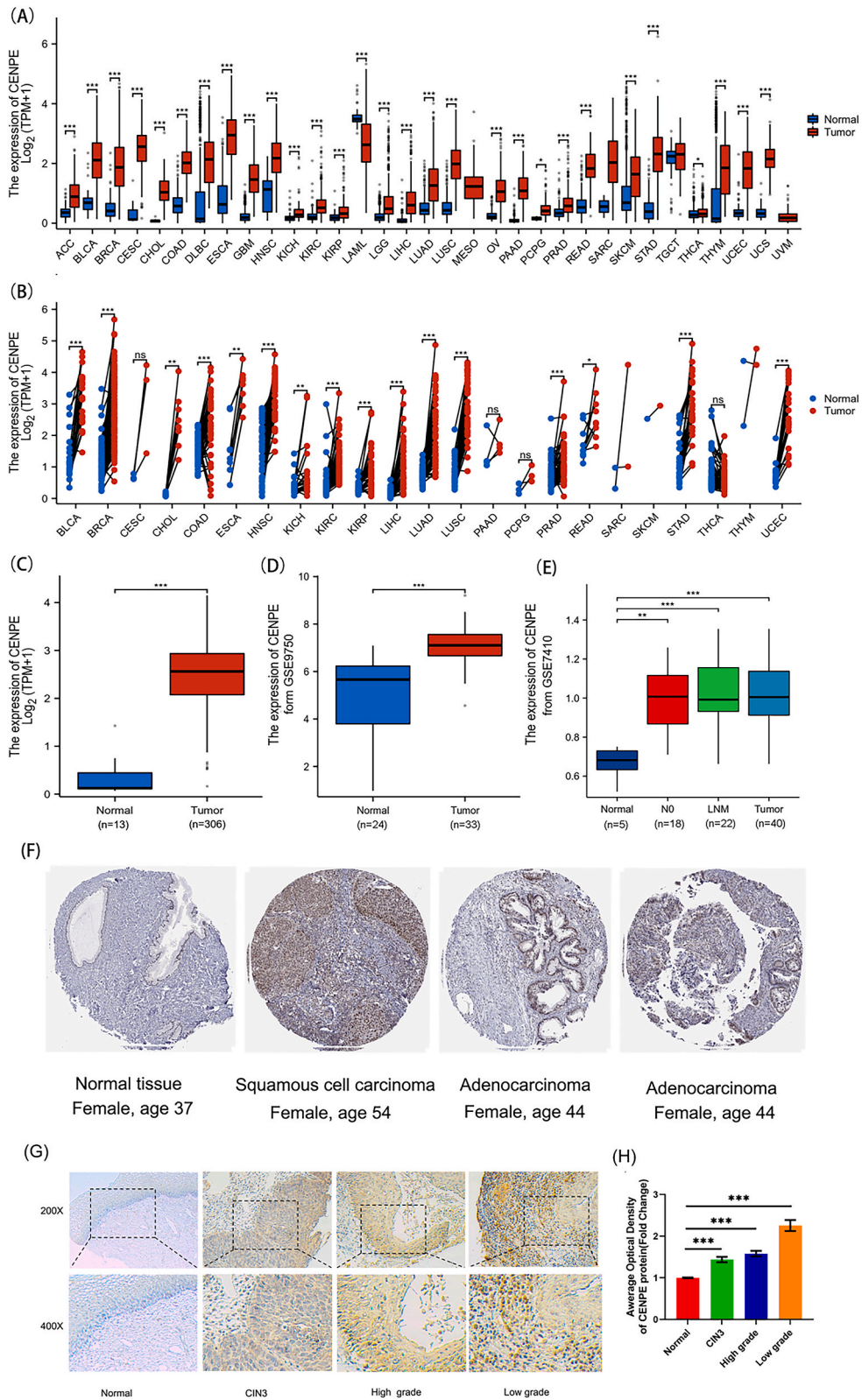
KEGG analysis reveals that the pathways primarily impact signaling and regulation (such as Neuroactive ligand-receptor interaction, Cytokine-cytokine receptor interaction, Cholinergic synapse), host defense mechanisms and immune regulation (including Complement and coagulation cascades, Viral protein interaction with cytokine and cytokine receptor, *Staphylococcus aureus* infection, Primary immunodeficiency), as well as metabolism and secretion (specifically Bile secretion) (Fig. 2C and D).

GSEA analysis reveals that the identified genes are predominantly associated with various pathways including the PI3K-Akt signaling pathway, MAPK signaling pathway, cancer pathways, growth factors, second messenger signaling transduction, and receptor tyrosine kinase signaling (Fig. 2E and F).

3.3. Correlation analysis of CENPE

A PPI network was constructed for 612 differentially expressed genes (Fig. 3A). Genes located closer to the center of the interaction network graph exhibit a higher number of connections with other genes. Subsequently, utilizing the MCODE plugin, 9 hub genes were identified: MT3, FAIM2, TCEAL6, SLC6A7, SYNPR, VSTM2B, CALY, LY6H, and CACNG3 (Fig. 3B). A co-expression heatmap was then generated for these genes along with CENPE, revealing significant correlations among CACNG3, LY6H, and SLC6A7 (Fig. 3C).

Subsequently, the three genes were utilized to establish a lncRNA-miRNA-mRNA network involving CENPE, utilizing CLIP-seq data sourced from the ENCORI database. LY6H was omitted from the analysis due to the absence of CLIP-seq data (Fig. 3D).



(caption on next page)

Fig. 1. The Expression of CENPE in Pan-Cancer and Cervical Cancer. (A) Differential analysis of CENPE expression in 33 tumors based on the data in the XENA Database. (B) Pan-cancer analysis of paired samples based on TCGA database. (C) Differential analysis of CENPE expression in unpaired samples. (D) Differential analysis of CENPE expression based on GSE9760 data. (E) Differential analysis of CENPE expression based on GSE7410 data. (F) CENPE protein expression based on the HPA database. (G, H) CENPE immunohistochemical expression and data statistics based on clinical samples.

3.4. Immunoinfiltration analysis of CENPE

Based on the levels of CENPE expression, the data on expression profiles were categorized into high and low expression groups. Subsequently, an assessment was conducted to determine the variations in levels of immune cell infiltration between these groups (refer to Fig. 4A). In the high CENPE expression group, Tcm, T-helper cells, and Th2 cells exhibited elevated levels of immune infiltration. Conversely, in the low expression group, B cells, cytotoxic cells, dendritic cells (DC), eosinophils, immature DC (iDC), mast cells, NK CD56 bright cells, neutrophils, plasmacytoid DC (pDC), T cells, TFH, and Th17 cells demonstrated heightened levels of immune infiltration. Following this, ssGSEA was employed for immune infiltration analysis to investigate the influence of CENPE on the tumor microenvironment. Spearman correlation analysis indicated a positive correlation between CENPE and T-helper cells, Th2 cells, Tcm, and Tgd, while a negative correlation was observed with macrophages, T cells, neutrophils, eosinophils, Th17 cells, iDC, mast cells, TFH, cytotoxic cells, DC, NK CD56 bright cells, B cells, and pDC (refer to Fig. 4B), consistent with the findings in Fig. 4A. Subsequently, the correlation between CENPE and the 11 significantly associated immune cells was visually represented using heatmaps and chord diagrams (refer to Fig. 4C and D).

Finally, an examination of the correlation between CENPE and various prevalent immune therapy targets indicated notable connections with ARIH1, KDM1A, KDM5B, and NSD3 (Fig. 4E–H).

3.5. Clinical information, survival and ROC curves

Upon examining the correlation between CENPE expression and clinical data, it was observed that CENPE expression is only marginally affected by menopausal status and obesity (Fig. 5A and B). However, it is significantly linked to clinical staging, histological grading, tumor size, distant metastasis, lymph node involvement, and high expression in outcomes for CESC (Fig. 5C–H).

ROC curve analysis indicated that the expression of CENPE exhibits a high level of diagnostic accuracy for CESC, with an Area Under the Curve (AUC) of 0.957 and a Confidence Interval (CI) of 0.900–1.000 (Fig. 5I). This finding was further supported by similar results observed in the validation datasets GSE9740 (Fig. 5G–AUC = 0.891, CI = 0.811–0.972) and GSE7401 (Fig. 5K–AUC = 0.970, CI = 0.924–1.000).

Samples were stratified into high and low expression groups according to the CENPE expression profile data. Subsequently, an analysis of OS, DSS, and PFI was conducted across various patient groups with CESC. The findings indicated reduced survival durations for patients belonging to the high CENPE expression group (Fig. 5L–N).

Subgroup analysis, based on clinical information, indicated shorter survival times in patients aged 50 years or younger (Fig. 5O), those in early and perimenopausal status (Fig. 5P), individuals with a height of 160 cm or less (Fig. 5Q), patients in clinical stages I and IV (Fig. 5R), and those in pathological stages T1 and T4 (Fig. 5S) exhibiting high CENPE expression.

3.6. Regression analysis and survival prediction

Univariate and multivariate Cox regression analyses were performed to ascertain autonomous prognostic determinants for patients with CESC, taking into account a range of clinical-pathological factors and levels of CENPE expression. The findings indicated that CENPE expression levels, age, BMI, clinical stage, and primary therapy outcome were notable independent prognostic factors (Table 1).

Based on these factors, predictive models were formulated to calculate the 1-, 3-, and 5-year survival probabilities for patients with CESC (Fig. 6A). The calibration plots for these models were constructed, revealing a concordance index (C-index) of 0.78, indicating a moderate level of accuracy in predictive capacity (Fig. 6B). Furthermore, the calibration plots demonstrated close alignment.

4. Discussion

Cervical cancer ranks as the fourth most prevalent malignancy among women globally. Despite notable progress in screening and treatment, a considerable risk of mortality persists [34]. Consequently, current research efforts are directed towards enhancing diagnostic and therapeutic approaches to diminish mortality rates and enhance the quality of life for patients. The utilization of high-throughput technologies and bioinformatics has facilitated the acquisition of omics data for investigating cancer alterations and identifying associated biomarkers. This methodology has emerged as a crucial research tool for disease diagnosis and clinical prognosis assessment by elucidating modifications in the structure or functionality of systems, organs, tissues, cells, and subcellular components.

Utilizing data from the TCGA and GEO databases, the study identified heightened transcriptional expression levels of CENPE in CESC compared to normal tissues. Additionally, protein expression levels in CESC tissues were found to be elevated based on the HPA database. These findings align with the research conducted by El-Arabey et al. [35,36], suggesting a correlation between CENPE and tumorigenesis as well as disease progression.

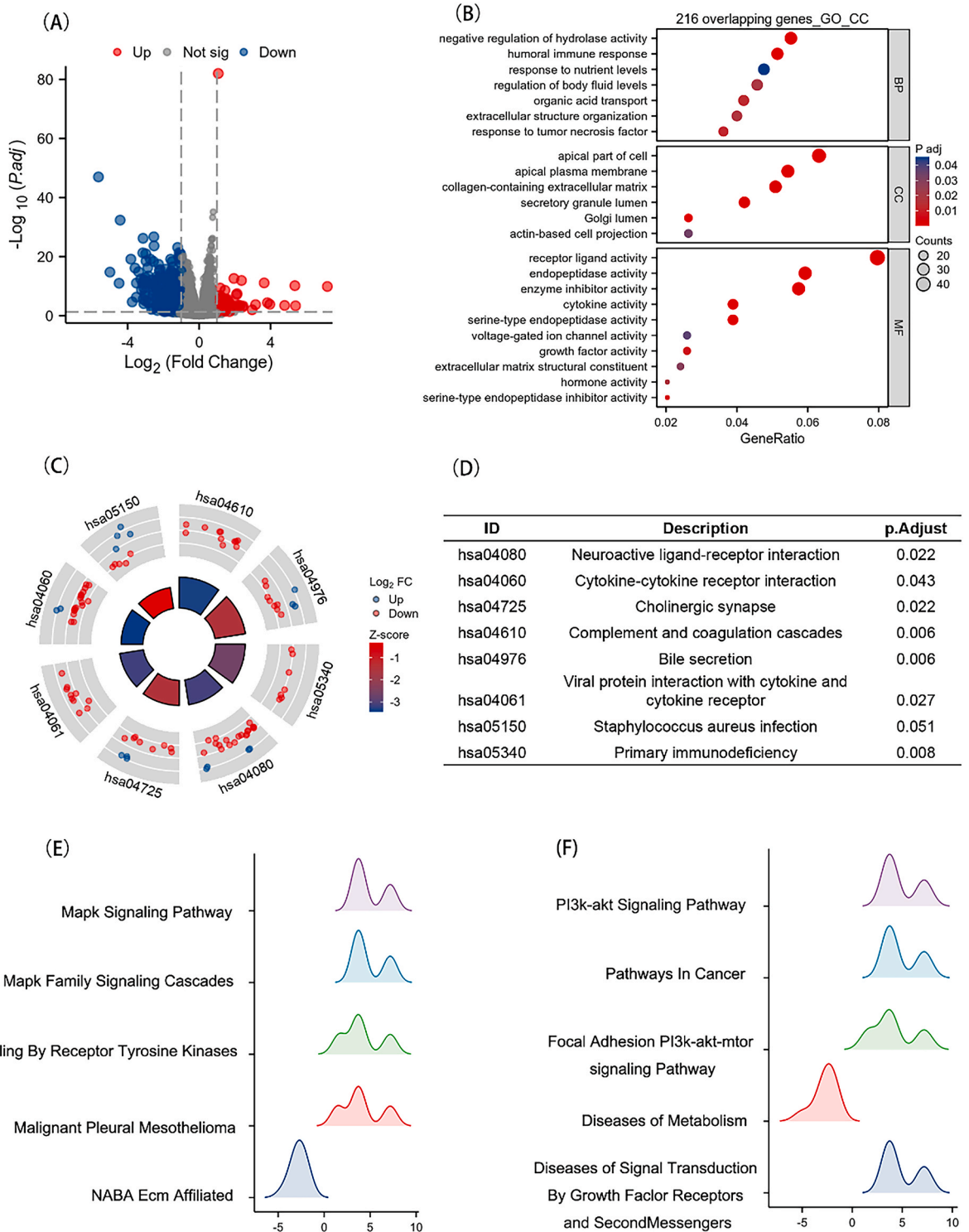


Fig. 2. Single-gene differential analysis and Functional enrichment analysis of CENPE. (A) Volcano plot of single gene differential analysis of CENPE. (B) Gene Ontology (GO) analysis results. (C) Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis results. (D) KEGG analysis category names corresponding to KEGG identifiers. (E,F) Gene set enrichment analysis (GSEA) results. When the horizontal coordinate is positive, CENPE expression correlates positively with this pathway; when negative, it correlates inversely.

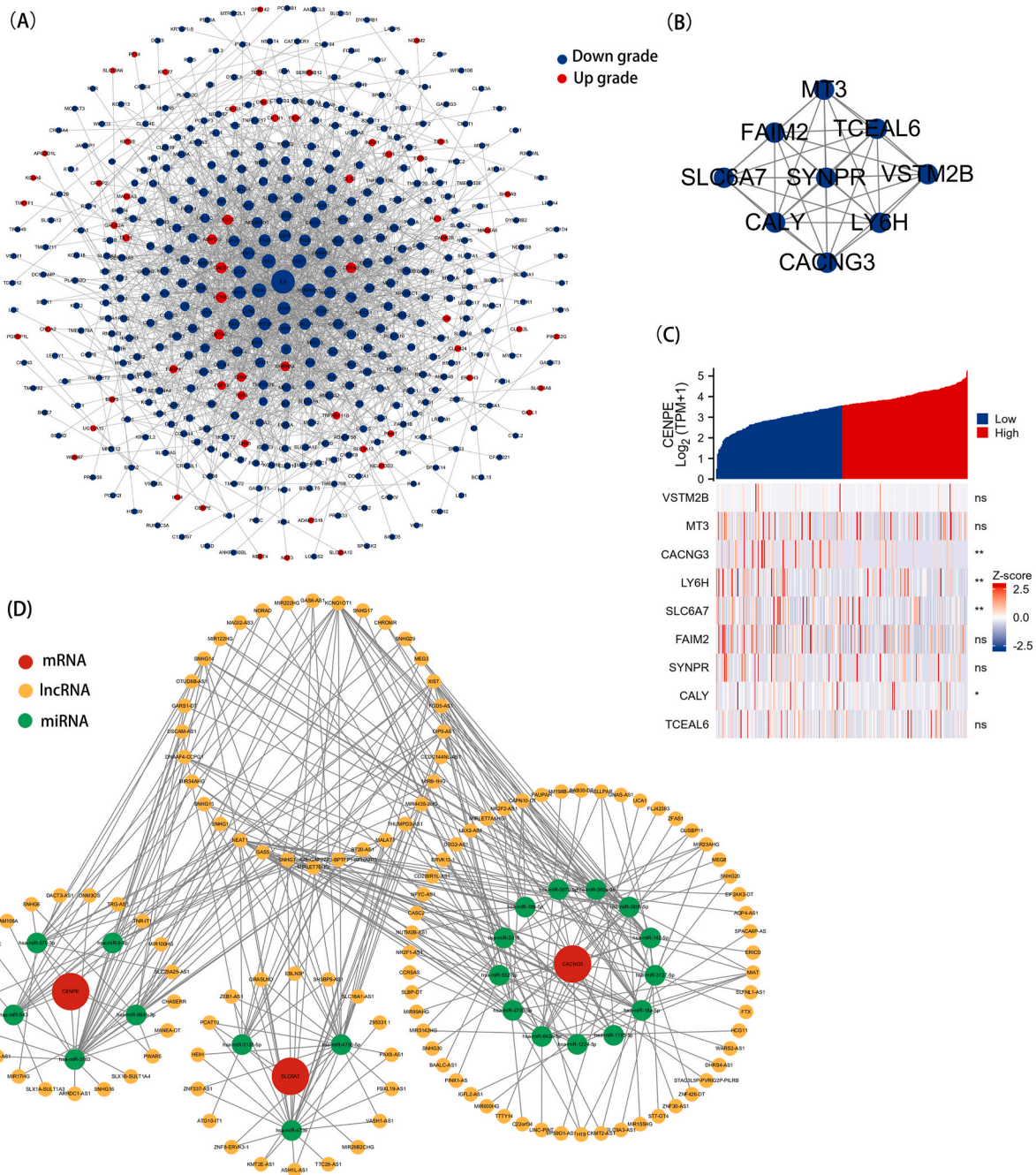


Fig. 3. Correlation analysis of CENPE. (A) Protein–protein interaction network (PPI) of differentially expressed genes in single gene differential analysis. (B) PPI network map of the HUB gene. (C) Single gene co-expression heatmap of the HUB gene and CENPE. (D) The lncRNA-miRNA-mRNA network based on CLIP-seq data for CENPE, CACNG3, and SLC6A7.

To elucidate the potential molecular mechanisms underlying the promotion of tumor development by CENPE, an initial single-gene differential analysis was conducted, followed by GO, KEGG, and GSEA on the differentially expressed genes. The GO analysis revealed that genes related to CENPE are implicated in various biological processes, including the immune system, tumor invasion and metastasis, tumor microenvironment, cell polarity and morphology, tumor metabolism, degradation, growth, organic acid transport, tumor signaling transduction, enzyme activity, cytokines, electrophysiology, extracellular matrix, and hormonal regulation.

The close association between CENPE and CESC tumors is supported by additional evidence. Notably, the differential genes of CENPE are linked to serine-type endopeptidase activity. Serine proteases, a subset of the protease family, predominantly target proteins containing specific serine residues, thereby modulating cellular signaling and protein degradation [37,38]. This implies that

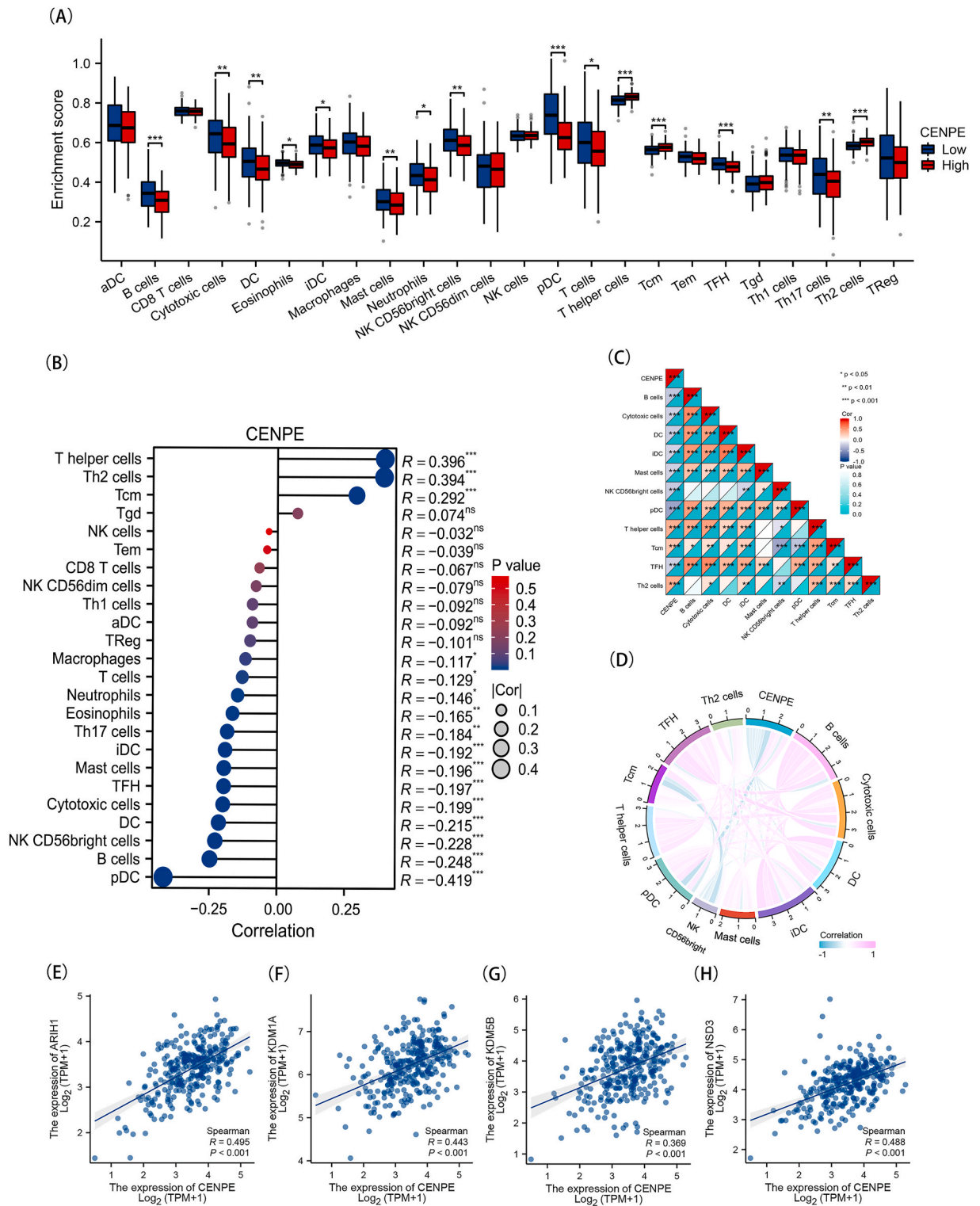
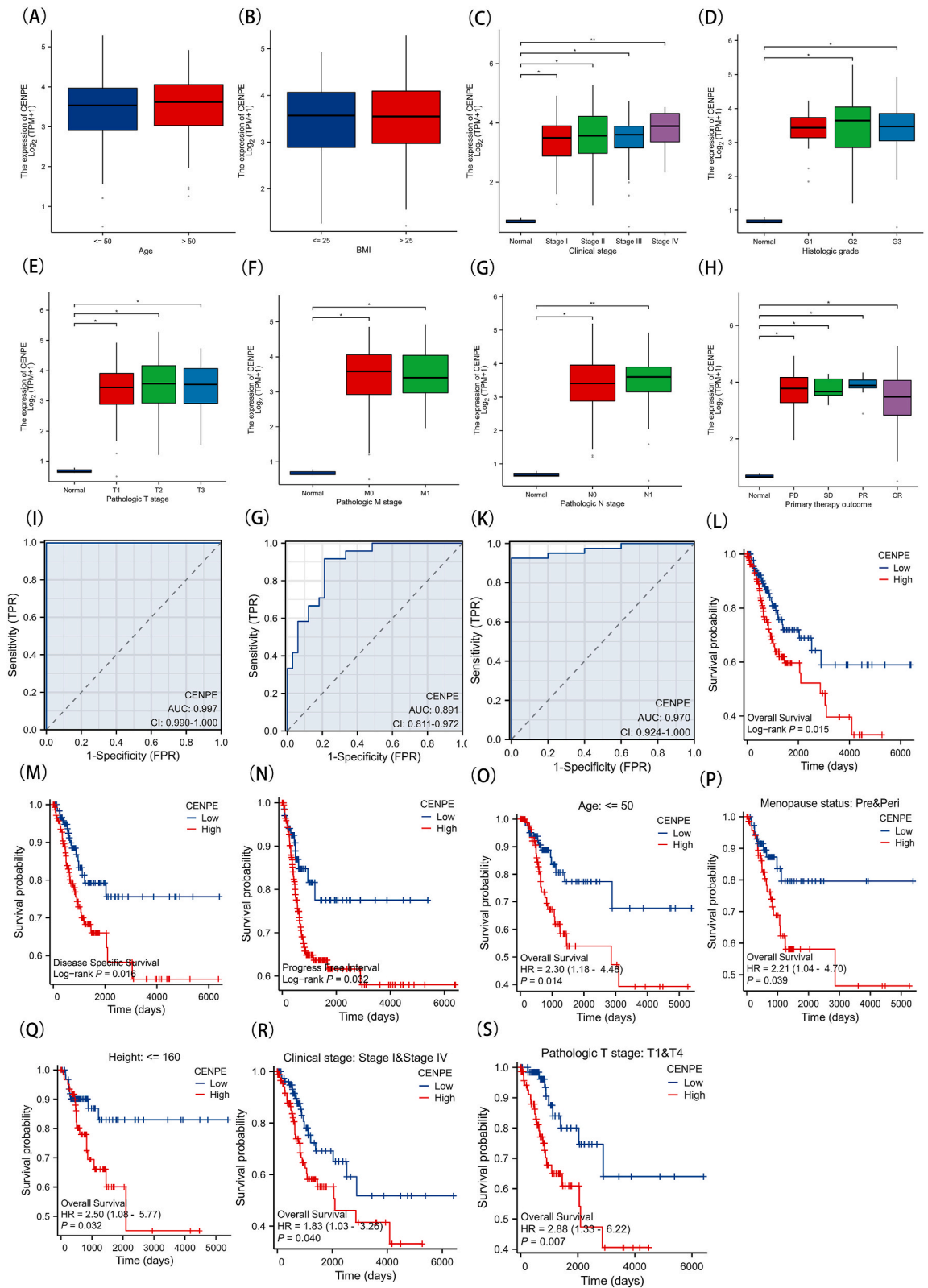


Fig. 4. Immunoinfiltration analysis of CENPE. (A) Immune cell infiltration levels across varying CENPE expression levels. (B) Correlation between CENPE and infiltration levels of 24 immune cells. (C) Correlation heatmap of CENPE with 11 significantly associated immune cells. (D) Chord diagram of CENPE with 11 significantly associated immune cells. (E–H) The correlation between CENPE expression and ARIH1, KDM1A, KDM5B, and NSD3. Significance identifier: ns (no significance), $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.



(caption on next page)

Fig. 5. Clinical information, survival and ROC curves. (A,B) The expression of CENPE among different age and BMI groups in CESC population. (C–H) The expression levels of CENPE among populations with different clinical and pathological factors. (I–K) The AUC curve of CENPE for diagnosing CESC in TCGA dataset, GSE9740, and GSE7401. (L–N) The Kaplan-Meier survival curves illustrating the association between CENPE expression and overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI). (O–S) Kaplan-Meier survival curves showing CENPE expression’s association with OS in subgroups based on clinical and pathological factors.

Table 1
The Cox regression analysis results of CENPE and various clinical pathological factors.

Characteristics	Total(N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95 % CI)	P value	Hazard ratio (95 % CI)	P value
Group	306				
low	156	Reference		Reference	
High	150	1.790 (1.110–2.887)	0.017	4.752 (1.359–16.625)	0.015
Age	306	1.017 (0.999–1.034)	0.062	1.035 (0.993–1.078)	0.106
BMI	260	0.953 (0.911–0.996)	0.034	0.956 (0.881–1.038)	0.285
Clinical stage	299				
Stage I	162	Reference		Reference	
Stage II	71	0.784 (0.398–1.543)	0.481	0.334 (0.061–1.825)	0.206
Stage IV	22	4.376 (2.354–8.136)	< 0.001	1.470 (0.135–15.982)	0.752
Stage III	44	1.482 (0.754–2.915)	0.254	1.053 (0.291–3.813)	0.937
Histologic grade	273				
G1	19	Reference			
G2	135	1.284 (0.394–4.186)	0.678		
G3	119	1.110 (0.331–3.720)	0.866		
Histological type	306				
Adenocarcinoma	48	Reference			
Squamous cell carcinoma	253	1.078 (0.551–2.109)	0.825		
Adenosquamous	5	1.858 (0.235–14.704)	0.557		
Menopause status	233				
Pre	126	Reference			
Post	82	1.160 (0.668–2.014)	0.597		
Peri	25	0.645 (0.225–1.848)	0.414		
Primary therapy outcome	161				
CR	143	Reference		Reference	
PD	8	16.872 (6.693–42.533)	< 0.001	27.296 (6.918–107.699)	< 0.001
SD	5	0.000 (0.000 - Inf)	0.997	0.000 (0.000 - Inf)	0.998
PR	5	6.548 (1.447–29.630)	0.015	10.147 (1.766–58.307)	0.009
Treatment type	305				
Pharmaceutical	147	Reference			
Radiation Therapy	158	1.059 (0.666–1.682)	0.809		
Abortion	306				
No	213	Reference			
Yes	93	1.148 (0.705–1.869)	0.579		

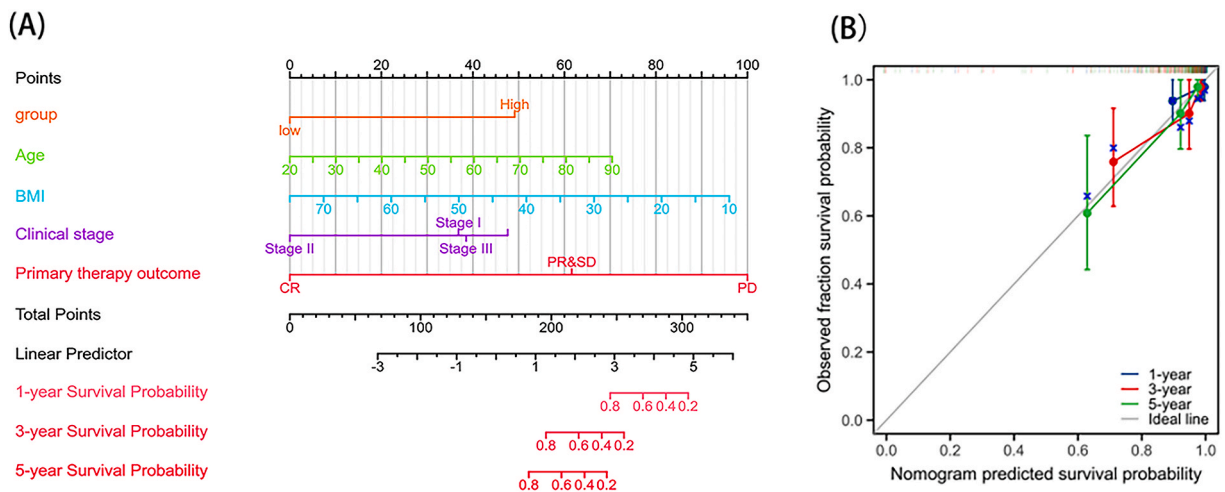


Fig. 6. A survival model for CESC patients using CENPE expression levels. (A) Nomogram model. (B) Calibration graph.

CENPE might enhance the proliferation of CESC tumors by impeding protein degradation through the inhibition of serine protease activity. Analysis using the KEGG database indicates that these pathways primarily impact signaling and regulation, host defense mechanisms, immune regulation, as well as metabolism and secretion processes. Furthermore, the GSEA demonstrates that the differentially expressed genes are predominantly associated with the PI3K-Akt signaling pathway and the MAPK signaling pathway. Consistent with these findings, Sui et al. [39] propose a correlation between the progression of CESC and the PI3K-Akt signaling pathway, while Katopodis et al. [40] suggest an association between female cancers and the MAPK signaling pathway. It is noteworthy that the GSEA analysis reveals the activation of receptor tyrosine kinases (RTKs). RTKs represent a class of receptor proteins abundantly present on the cell membrane, crucial for cellular signal transduction. Upon ligand binding, typically a growth factor, an RTK undergoes a conformational change, activating its intrinsic tyrosine kinase activity. Consequently, tyrosine residues are phosphorylated, leading to the activation of downstream signaling pathways such as the PI3K-Akt and MAPK signaling pathways [41]. These signaling cascades govern cell growth, proliferation, differentiation, and survival, thereby facilitating the initiation and progression of cancer.

To enhance the understanding of CENPE's involvement in CESC initiation and progression, a PPI network was constructed for the differentially expressed genes. IL-6 emerged as a common factor linked to multiple genes [42]. Consequently, it is postulated that CENPE exerts a significant influence on CESC by modulating IL-6. Furthermore, IL-6 has the ability to stimulate tumor cell proliferation and invasion by activating signaling pathways like PI3K-Akt and MAPK, thereby promoting cancer advancement. This observation is consistent with the outcomes of the present study. Subsequently, key genes were identified within this gene set, and a detailed analysis of CENPE revealed notable connections with CACNG3, LY6H, and SLC6A7. LY6H [43], SLC6A7 [44], and CACNG3 [45] are all associated with the nervous system. The study's findings suggest that CENPE might have a critical role in controlling tumor growth, metastasis, and responses to treatment through interactions with the nervous system. Extensive experimental and clinical data support the influence of the nervous system on the tumor microenvironment, immune modulation, and behavior of tumor cells [46–49]. Particularly, Marco Arese and colleagues noted that all peripheral cancer types interact with neuronal structures in the late stages, particularly bladder, prostate, pancreatic, colorectal, lung, head and neck cancers, and cholangiocarcinoma [50]. The nervous system may influence cancer through direct communication with cancer cells via neural signals, releasing neurotransmitters and neurotrophic factors that affect growth and metastasis. It may also indirectly alter immune cell functions in the tumor microenvironment and impact cell interactions, thereby influencing tumor progression [51]. This approach provides a targeted neuro-related pathway treatment strategy for CESC and also offers evidence for the direction of non-pharmacological interventions in regulating the nervous system for therapeutic purposes. Nevertheless, the mechanisms of nervous system function in CESC are not fully understood, necessitating further research to clarify these findings.

Additionally, common long non-coding RNAs (lncRNAs) were identified among CENPE, SLC6A7, and CACNG3, indicating potential regulatory interactions among them through these lncRNAs. This further underscores the relationship between CENPE and SLC6A7, as well as CACNG3.

Given the correlation between CENPE and IL-6, we analyzed the association between CENPE expression and 24 immune cell types in CESC patients to further explore its impact on the tumor microenvironment. CENPE significantly influences various immune cells, including T helper cells, Th2 cells, Tcm, iDC, mast cells, TFH, cytotoxic cells, DC, B cells, pDC, and NK CD56bright cells, highlighting its role in shaping the tumor microenvironment. Targeting CENPE may positively influence the tumor microenvironment in CESC and enhance the outcomes of immunotherapy. Specifically, targeting CENPE could improve immunotherapy response rates by increasing immune cell infiltration and promoting anti-tumor responses. Furthermore, combining CENPE targeting with immune checkpoint inhibitors may amplify treatment effects by modulating the tumor microenvironment.

Moreover, CENPE is notably correlated with ARIH1, KDM1A, KDM5B, and NSD3, which are common immunotherapy targets [52–55]. This suggests that CENPE could be a potential target for drug development due to its association with these key therapeutic targets.

Our study also explored CENPE's implications for CESC diagnosis and treatment. Clinical data show high CENPE expression across different tumor stages, significantly affecting patient survival and demonstrating high sensitivity and specificity as a therapeutic target. Notably, in the validation dataset GSE7401, which includes early-stage cervical cancer samples, CENPE exhibited excellent diagnostic performance (AUC: 0.970, CI: 0.924–1.000). This underscores its potential for accurate early-stage diagnosis, enhancing its clinical value. Although CENPE shows high sensitivity and specificity, it cannot diagnose all cases. A thorough assessment with other molecular biomarkers is necessary to reduce the risk of false positives or negatives, even if the probability is low. Subgroup analyses revealed that CENPE significantly influences survival in patients aged <50, with height <160 cm, peri- and post-menopausal status, and clinical stages I and IV, indicating its sensitivity in these populations. Age, height, gender, and hormonal status may influence CENPE activity in tumor cells. Patients under 50 often have a higher capacity for cell division, potentially increasing CENPE activity. Individuals shorter than 160 cm may experience variations in CENPE expression due to genetic, nutritional, or endocrine factors. Furthermore, hormonal differences in perimenopausal and postmenopausal individuals may also affect CENPE expression.

To evaluate CENPE's role in predicting patient prognosis, we performed COX regression analysis using clinical data and created nomogram plots based on the results. Our analysis revealed that patient survival is influenced by factors such as clinical pathological staging and clinical outcomes. Notably, older age, lower BMI, and higher CENPE expression are associated with decreased survival rates at 1, 3, and 5 years. Age-related physiological decline may reduce treatment responsiveness and increase complication risks, while a lower BMI indicates poorer nutrition and diminished disease-fighting capability, raising mortality risk. The calibration plots demonstrate a strong correlation between actual and predicted OS values at 1, 3, and 5 years, validating the accuracy of our prognostic predictions.

In conclusion, CENPE is highly expressed in CESC and significantly impacts patient survival. It demonstrates excellent diagnostic

sensitivity and specificity and serves as a reliable prognostic biomarker with strong predictive efficacy.

CRediT authorship contribution statement

Peiqiang Peng: Writing – original draft. **Jingying Zheng:** Writing – review & editing. **Kang He:** Conceptualization. **Kai Wang:** Data curation. **Longyun Wang:** Formal analysis. **Xufei Zheng:** Investigation. **Hao Wu:** Methodology. **Zhenning Yang:** Software. **Shuang Zhang:** Supervision. **Lijing Zhao:** Project administration, Funding acquisition.

Ethical approval and consent to participate

This study have been approved by the jilin university ethics committee and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

Availability of data and material

All the data are publicly available, and the raw data can be accessed from the GEO database at <https://www.ncbi.nlm.nih.gov/geo/>.
XENA at <https://xenabrowser.net/datapages/>.

Consent for publication

All authors consent to the publication of this manuscript. We confirm that the manuscript has not been published or submitted for publication elsewhere. We have read and approved the final version of the manuscript and agree with its publication.

Financial Support

This study was supported by the Jilin Provincial Department of Science and Technology project (grant number: 20210204200YY).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Zhao lijing reports article publishing charges and statistical analysis were provided by Jilin Provincial Department of Science and Technology project. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Thanks to Jilin University for providing experimental platform support.

References

- [1] R.L. Siegel, et al., *Cancer statistics, 2021*. *CA cancer J clin* 71 (1) (2021) 7–33.
- [2] M. Arbyn, et al., Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis, *Lancet Global Health* 8 (2) (2020) e191–e203.
- [3] P.A. Cohen, et al., Cervical cancer, *Lancet* 393 (10167) (2019) 169–182.
- [4] L.E. Markowitz, et al., Prevalence of HPV after Introduction of the vaccination Program in the United States, *Pediatrics* 137 (3) (2016) e20151968.
- [5] B. Sehnal, et al., Current data on the efficacy of prophylactic HPV vaccination in the primary prevention of cervical lesions, *Ceska Gynekol.* 87 (2) (2022) 124–130.
- [6] I. Kalliala, et al., Incidence and mortality from cervical cancer and other malignancies after treatment of cervical intraepithelial neoplasia: a systematic review and meta-analysis of the literature, *Ann. Oncol.* 31 (2) (2020) 213–227.
- [7] L. Li, et al., Construction of a four-mRNA prognostic signature with its ceRNA network in CESC, *Sci. Rep.* 12 (1) (2022) 10691.
- [8] L. Shan, et al., CENPE promotes lung adenocarcinoma proliferation and is directly regulated by FOXM1, *Int. J. Oncol.* 55 (1) (2019) 257–266.
- [9] J.R. Testa, et al., Chromosomal localization of the genes encoding the kinetochore proteins CENPE and CENPF to human chromosomes 4q24–q25 and 1q32–q41, respectively, by fluorescence in situ hybridization, *Genomics* 23 (3) (1994) 691–693.
- [10] Y.H. Yang, Y.L. Wei, Z.Y. She, Kinesin-7 CENP-E in tumorigenesis: Chromosome instability, spindle assembly checkpoint, and applications, *Front. Mol. Biosci.* 11 (2024) 1366113.
- [11] K.W. Yu, et al., Mechanisms of kinesin-7 CENP-E in kinetochore-microtubule capture and chromosome alignment during cell division, *Biol. Cell.* 111 (6) (2019) 143–160.
- [12] D.A. Thrower, M.A. Jordan, L. Wilson, Modulation of CENP-E organization at kinetochores by spindle microtubule attachment, *Cell Motil Cytoskeleton* 35 (2) (1996) 121–133.
- [13] X. Liu, et al., Mitotic motor CENP-E cooperates with PRC1 in temporal control of central spindle assembly, *J. Mol. Cell Biol.* 12 (8) (2020) 654–665.
- [14] A. Tighe, V.L. Johnson, S.S. Taylor, *Truncating APC mutations have dominant effects on proliferation, spindle checkpoint control, survival and chromosome stability*. *J Cell Sci* 117 (Pt 26) (2004) 6339–6353.
- [15] D.H. Bach, W. Zhang, A.K. Sood, Chromosomal instability in tumor initiation and development, *Cancer Res.* 79 (16) (2019) 3995–4002.
- [16] E. Hitti, et al., Systematic analysis of AU-rich element expression in cancer reveals common functional clusters regulated by key RNA-binding proteins, *Cancer Res.* 76 (14) (2016) 4068–4080.
- [17] Y. Yuan, et al., Comprehensive pan-cancer analysis identifies centromere-associated protein E as a novel prognostic and immunological biomarker in human tumors, *Biochim. Biophys. Acta Gen. Subj.* 1867 (6) (2023) 130346.

- [18] X. Zhu, et al., CENPE expression is associated with its DNA methylation status in esophageal adenocarcinoma and independently predicts unfavorable overall survival, *PLoS One* 14 (2) (2019) e0207341.
- [19] Q. Wang, et al., CENPA promotes clear cell renal cell carcinoma progression and metastasis via Wnt/ β -catenin signaling pathway, *J. Transl. Med.* 19 (1) (2021) 417.
- [20] Z. Zhang, X. Zhang, Identification of m6A-related biomarkers Associated with Prognosis of colorectal cancer, *Med. Sci. Mon. Int. Med. J. Exp. Clin. Res.* 27 (2021) e932370.
- [21] T.F. Li, et al., Overexpression of kinesin superfamily members as prognostic biomarkers of breast cancer, *Cancer Cell Int.* 20 (2020) 123.
- [22] P. Guruvaiiah, et al., ATAD2 is a driver and a therapeutic target in ovarian cancer that functions by upregulating CENPE, *Cell Death Dis.* 14 (7) (2023) 456.
- [23] Y. Liang, et al., LSD1-Mediated epigenetic reprogramming drives CENPE Expression and prostate cancer progression, *Cancer Res.* 77 (20) (2017) 5479–5490.
- [24] N.J. Balamuth, et al., Serial transcriptome analysis and cross-species integration identifies centromere-associated protein E as a novel neuroblastoma target, *Cancer Res.* 70 (7) (2010) 2749–2758.
- [25] J. Vivian, et al., Toil enables reproducible, open source, big biomedical data analyses, *Nat. Biotechnol.* 35 (4) (2017) 314–316.
- [26] S. Davis, P.S. Meltzer, GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor, *Bioinformatics* 23 (14) (2007) 1846–1847.
- [27] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.* 15 (12) (2014) 550.
- [28] G. Yu, et al., clusterProfiler: an R package for comparing biological themes among gene clusters, *OMICS* 16 (5) (2012) 284–287.
- [29] A. Subramanian, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.* 102 (43) (2005) 15545–15550.
- [30] L.J. Jensen, et al., STRING 8—a global view on proteins and their functional interactions in 630 organisms, *Nucleic Acids Res.* 37 (Database issue) (2009). D412–6.
- [31] J.H. Li, et al., starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data, *Nucleic Acids Res.* 42 (Database issue) (2014). D92–7.
- [32] S. Hänzelmann, R. Castelo, J. Guinney, GSEA: gene set variation analysis for microarray and RNA-seq data, *BMC Bioinformatics* 14 (2013) 7.
- [33] G. Bindea, et al., Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer, *Immunity* 39 (4) (2013) 782–795.
- [34] W. Small Jr., et al., Cervical cancer: a global health crisis, *Cancer* 123 (13) (2017) 2404–2412.
- [35] A.A. El-Arabey, S.A. Salama, A.R. Abd-Allah, CENP-E as a Target for Cancer Therapy: where Are We Now? *Life Sci.*, vol. 208, 2018, pp. 192–200.
- [36] G. legiani, et al., CENPE inhibition Leads to mitotic Catastrophe and DNA Damage in medulloblastoma cells, *Cancers* 13 (5) (2021).
- [37] N.D. Rawlings, A.J. Barrett, A. Bateman, MEROPS: the database of proteolytic enzymes, their substrates and inhibitors, *Nucleic Acids Res.* 40 (Database issue) (2012). D343–50.
- [38] R. Tagirasa, E. Yoo, Role of serine proteases at the tumor-stroma interface, *Front. Immunol.* 13 (2022) 832418.
- [39] X. Sui, et al., hsa-mir-133a-2 promotes the proliferation and invasion of cervical cancer cells by targeting the LAMB3-mediated PI3K/ATK pathway, *Cancer Med.* 12 (5) (2023) 5874–5888.
- [40] P. Katopodis, et al., p38 β - MAPK11 and its role in female cancers, *J. Ovarian Res.* 14 (1) (2021) 84.
- [41] R. Trenker, N. Jura, Receptor tyrosine kinase activation: from the ligand perspective, *Curr. Opin. Cell Biol.* 63 (2020) 174–185.
- [42] S. Kaur, et al., A panoramic review of IL-6: structure, pathophysiological roles and inhibitors, *Bioorg. Med. Chem.* 28 (5) (2020) 115327.
- [43] M. Wu, et al., Unbalanced regulation of α 7 nAChRs by Ly6h and NACHO contributes to neurotoxicity in alzheimer's disease, *J. Neurosci.* 41 (41) (2021) 8461–8474.
- [44] K.M. Reid, et al., MED27, SLC6A7, and MPPE1 Variants in a complex neurodevelopmental Disorder with severe dystonia, *Mov. Disord.* 37 (10) (2022) 2139–2146.
- [45] E. Shan, et al., Integrated profiling identifies CACNG3 as a prognostic biomarker for patients with glioma, *BMC Cancer* 23 (1) (2023) 846.
- [46] A.H. Zahalka, P.S. Frenette, Nerves in cancer, *Nat. Rev. Cancer* 20 (3) (2020) 143–157.
- [47] J.L. Saloman, et al., Can stopping nerves, Stop Cancer? *Trends Neurosci* 39 (12) (2016) 880–889.
- [48] H.S. Venkatesh, et al., Neuronal activity promotes glioma growth through neuroligin-3 secretion, *Cell* 161 (4) (2015) 803–816.
- [49] L. Wang, et al., Crosstalk between the nervous system and tumor microenvironment: functional aspects and potential therapeutic strategies, *Cancer Lett.* 594 (2024) 216986.
- [50] M. Arese, et al., Tumor progression: the neuronal input, *Ann. Transl. Med.* 6 (5) (2018) 89.
- [51] F. Winkler, et al., Cancer neuroscience: State of the field, emerging directions, *Cell* 186 (8) (2023) 1689–1707.
- [52] G. Yuan, et al., Elevated NSD3 histone methylation activity drives squamous cell lung cancer, *Nature* 590 (7846) (2021) 504–508.
- [53] Y. Fang, G. Liao, B. Yu, LSD1/KDM1A inhibitors in clinical trials: advances and prospects, *J. Hematol. Oncol.* 12 (1) (2019) 129.
- [54] Y. Wu, et al., ARIH1 signaling promotes anti-tumor immunity by targeting PD-L1 for proteasomal degradation, *Nat. Commun.* 12 (1) (2021) 2346.
- [55] S.M. Zhang, et al., KDM5B promotes immune evasion by recruiting SETDB1 to silence retroelements, *Nature* 598 (7882) (2021) 682–687.