scientific reports

Check for updates

OPEN

Integrated bioinformatics, machine learning, and experimental validation identify a four-gene diagnostic signature for cervical cancer associated with PI3K/AKT signaling

Hailong Zhang[1,2,6], Luhong Xie[1,3,6], Yameng Liu[1,4,6], Hanlin Yang[1,2], Shaoju Min[1,2], Yurong Zhu[1,2], Jie Ren[5]✉, Xuehui Li[1,2]✉ & Yujie Tan[1,2]✉

Early and accurate diagnosis remains a major challenge in cervical cancer management. This study aimed to identify reliable diagnostic biomarkers for cervical cancer by integrating bioinformatics and machine learning approaches and to further validate their biological relevance experimentally. Transcriptomic data from the Gene Expression Omnibus and The Cancer Genome Atlas were analyzed using differential expression analysis, weighted gene co-expression network analysis, and three machine learning algorithms to identify core genes. Diagnostic performance was evaluated using receiver operating characteristic curves and a nomogram model. Functional relevance was explored by drug sensitivity analysis, ssGSEA, immune infiltration analysis, and single-cell RNA sequencing. RT-qPCR validation was performed in 10 paired cervical cancer and adjacent normal tissues, while Western blotting was performed in three paired tissue samples. In vitro validation was conducted using SiHa and HeLa cells. Four genes, CCND1, TRIP13, MYBL2, and GNB4, were identified as potential diagnostic biomarkers, and the combined model showed superior diagnostic performance compared with any single gene (AUC = 0.989). Treatment with 3-methyladenine altered the expression of these genes, suggesting their potential association with PI3K/AKT-related pathway activity. Moreover, siRNA-mediated GNB4 knockdown suppressed cervical cancer cell proliferation and reduced PI3K and AKT phosphorylation, providing preliminary evidence for the functional involvement of GNB4 in PI3K/AKT pathway activation. CCND1, TRIP13, MYBL2, and GNB4 may serve as promising diagnostic biomarkers for cervical cancer. Their dysregulation was associated with PI3K/AKT pathway activity and may reflect molecular alterations involved in cervical cancer progression. In particular, GNB4 showed potential diagnostic relevance and preliminary functional significance, suggesting that it may represent a candidate biomarker and molecular target for further investigation.

Keywords Cervical cancer, Diagnostic biomarkers, Machine learning, PI3K/AKT signaling pathway, Bioinformatics analysis

Globally, cervical cancer remains a significant threat to women's health, accounting for approximately 600,000 new diagnoses and more than 300,000 deaths each year, with the highest burden observed in developing countries[1]. Although human papillomavirus vaccination and early screening programs have contributed to a reduction in the overall incidence, most patients are still diagnosed at advanced stages because they do

[1]Center for Clinical Laboratories, The Affiliated Hospital of Guizhou Medical University, Guiyang, China. [2]School of Clinical Laboratory Science, Guizhou Medical University, Guiyang, China. [3]Department of Clinical Laboratory, Guizhou Hospital, The First Affiliated Hospital of Sun Yat-sen University, Guiyang, China. [4]School of Clinical Medicine, Guizhou Medical University, Guiyang, China. [5]Department of Obstetrics and Gynecology, The Affiliated Hospital of Guizhou Medical University, Guiyang, China. [6]These authors contributed equally: Hailong Zhang, Luhong Xie and Yameng Liu. ✉email: rj624@163.com; 18275281480@163.com; tanyujie@gmc.edu.cn

not exhibit obvious symptoms during the early stages of the disease. This delay in detection often postpones treatment and results in poor clinical outcomes[2].Moreover, female malignant tumors, including cervical cancer, breast cancer, and ovarian cancer, still face common clinical challenges, such as tumor heterogeneity, insufficient early diagnostic accuracy, resistance to conventional therapies, and the lack of robust molecular markers for individualized treatment decision-making[3,4]. For this reason, improving the sensitivity and specificity of early diagnosis has become an urgent priority in cervical cancer prevention and management. In current clinical practice, Papanicolaou smear testing and human papillomavirus screening are commonly used methods. However, these approaches have certain limitations, including relatively high false-negative rates and limited specificity, which may lead to misdiagnosis or overtreatment[5]. With the rapid development of transcriptomics and high-throughput sequencing technologies, bioinformatics-based gene expression analysis has become a powerful approach for identifying novel biomarkers for early cancer detection[6,7].

The formation and progression of tumors involve an extremely complex network of molecular interactions and a high degree of biological heterogeneity. In recent years, machine learning strategies have attracted increasing attention in biomarker discovery because they can efficiently handle high-dimensional biomedical data[8]. Compared with traditional feature selection methods that mainly rely on differential expression, univariate statistical testing, or single network-topological parameters, machine learning approaches can capture complex nonlinear relationships, reduce data dimensionality, and improve the stability and predictive performance of biomarker screening from high-throughput omics data[4,9].Support vector machine-recursive feature elimination, least absolute shrinkage and selection operator regression, and random forest algorithms have proven to be effective and show strong translational potential in studies of various cancers[10].Recent studies on female malignant tumors have further demonstrated that machine learning, single-cell analysis, and multi-omics integration can help identify disease-associated cell populations, molecular subtypes, prognostic signatures, and potential therapeutic targets, thereby providing new strategies for precision oncology[3,4,9]. Long-term infection with human papillomavirus, disruption of the cell cycle, immune evasion, and metabolic alterations have been strongly associated with the development of cervical cancer[11] Aberrant gene expression has also been reported to influence tumor biological behavior, therapeutic response, and immunological sensitivity[12]. In addition, advances in single-cell sequencing technologies have enabled detailed characterization of intratumoral heterogeneity, providing more precise cellular-level evidence for early detection and targeted therapy[13].

This study integrated multiple cervical cancer transcriptomic datasets from the Gene Expression Omnibus database and applied differential expression analysis, weighted gene co-expression network analysis, and machine learning approaches to uncover key genes with diagnostic relevance. Functional characteristics were further investigated through single-sample gene set enrichment analysis, immune infiltration assessment, drug sensitivity prediction, and single-cell validation. Experimental validation was subsequently performed in clinical tissues and cervical cancer cell lines. Notably, this study provides the first systematic evidence of elevated GNB4 expression and its potential diagnostic value in cervical cancer. Our findings offer new molecular insights and candidate biomarkers for early precision diagnosis and personalized treatment of cervical cancer.

Materials and methods

Data sources

Publicly available transcriptome datasets were obtained from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) and The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/). These included GSE63514 (52 samples comprising 24 cervical cancer tissues and 28 normal tissues; platform GPL570), GSE52903 (72 samples comprising 55 cervical cancer tissues and 17 normal tissues; platform GPL6244), GSE168652 (single-cell RNA sequencing data from cervical cancer and normal tissues generated using the 10 × Genomics platform), and TCGA (transcriptomic profiles and clinical information from 306 cervical cancer tissues and 3 normal cervical tissues). Detailed information on all public datasets used in this study is summarized in Supplementary Table S1.

Differentially expressed gene analysis

Differentially expressed gene (DEG) analysis was performed using the GSE63514 dataset. The raw expression matrix was preprocessed in R. Duplicate gene symbols were collapsed by calculating the average expression value using the avereps function in the limma package. The distribution of expression values across samples was assessed using boxplots before and after normalization. If the maximum expression value exceeded 50, log2 transformation was performed using $\log2(x+1)$.The expression matrix was then normalized using the normalizeBetweenArrays function in the limma package.As the differential expression analysis was performed within a single dataset from the same platform, no cross-platform batch-effect correction was applied at this step. Differential analysis between cervical cancer tissues and normal tissues was conducted using the limma package. Linear models were fitted using lmFit, followed by contrast analysis with makeContrasts and empirical Bayes moderation using eBayes. Genes with $|\log2\text{fold change}| > 1$ and adjusted p value < 0.05 were considered DEGs. Heatmaps and volcano plots were generated using the pheatmap and ggplot2 packages, respectively.

Weighted co-expression network analysis (WGCNA)

Weighted gene co-expression network analysis (WGCNA) was performed using the WGCNA package in R. After removing genes and samples with excessive missing values using the `goodSamplesGenes` function, genes with a mean expression value ≤ 0.5 were excluded. The top 5,000 genes with the highest variance in the GSE63514 dataset were selected for network construction. Candidate soft-thresholding powers from 1 to 20 were evaluated using the `pickSoftThreshold` function, and a soft-thresholding power of 6 was selected according to the scale-free topology criterion. An adjacency matrix was constructed and transformed into a topological overlap matrix (TOM). Gene modules were identified using the dynamic tree cut method with `deepSplit = 2`,

`pamRespectsDendro = FALSE`, and a minimum module size of 50. Similar modules were merged with a module eigengene dissimilarity threshold of 0.25. Pearson correlation analysis was then performed between module eigengenes and clinical traits, and the modules most strongly associated with cervical cancer were selected for subsequent analysis.

Functional enrichment analysis

Gene Ontology (GO) and KEGG pathway enrichment analyses were carried out using the clusterProfiler package. Enrichments with an adjusted p-value below 0.05 were deemed significant, and the results were illustrated using bubble plots.

Protein-protein interaction network construction

Protein-protein interaction (PPI) networks were constructed using the STRING database (version 11.0). Genes were converted to ENTREZ IDs and mapped to STRING identifiers. Interaction pairs with confidence scores > 0.7 were retained. The networks were visualized using the igraph and ggraph packages. Betweenness centrality was calculated to identify hub genes, with the cutoff defined as the mean + 0.1 × standard deviation.

Machine learning-based gene selection

Machine learning-based feature selection was performed using SVM-RFE, LASSO regression, and random forest (RF) algorithms based on the GSE63514 training dataset. Candidate genes obtained from the previous screening steps were used as the input gene set. For SVM-RFE, the e1071, caret, and custom msvmRFE scripts were used, and 10-fold cross-validation was performed for feature ranking and selection. The SVM-RFE model was constructed using a linear kernel. LASSO regression was performed using the glmnet package with a binomial model, alpha = 1, nlambda = 100, and 10-fold cross-validation. The optimal regularization parameter was selected according to lambda.min, and genes with non-zero coefficients were retained. RF analysis was conducted using the randomForest package. An initial RF model was constructed with ntree = 350, and the number of trees corresponding to the minimum error rate was selected to rebuild the final RF model. The mtry parameter was set according to the default setting of the randomForest package for classification analysis. Gene importance was ranked according to the importance score. Finally, the overlapping genes identified by SVM-RFE, LASSO, and RF were defined as key candidate diagnostic genes, and their diagnostic performance was further evaluated in the independent validation dataset GSE52903.

Co-expression and functional association analysis of diagnostic genes

To explore the potential relationships among the four diagnostic genes, correlation analysis was performed using the normalized and batch-corrected expression data from GSE63514 and GSE52903. Pearson correlation coefficients were calculated among CCND1, GNB4, TRIP13, and MYBL2, and the results were visualized as a correlation matrix. In addition, functional association networks of the four diagnostic genes were explored using the GeneMANIA database. The network was used to evaluate potential co-expression, physical interaction, co-localization, pathway, and shared protein domain relationships among the candidate genes and their associated molecules.

Diagnostic model construction and ROC analysis

Receiver operating characteristic (ROC) curves were generated using the pROC package to evaluate diagnostic performance. The area under the curve (AUC) was calculated. Individual disease risk prediction was performed using nomograms constructed with the rms package.Decision curve analysis (DCA) was performed using the dcurves package in R to evaluate the net benefit of the four-gene diagnostic model across different threshold probabilities in both the GSE63514 training cohort and the GSE52903 validation cohort.

Drug sensitivity prediction

Drug sensitivity was analyzed using the oncoPredict software based on data from the Genomics of Drug Sensitivity in Cancer (GDSC) database. Comparisons of IC50 values between the high- and low-expression groups were performed. Spearman correlation analysis was also conducted to assess the relationships between gene expression and drug response.

Immune infiltration analysis

The CIBERSORT algorithm was applied to estimate immune cell infiltration based on the LM22 gene signature. The parameters were set as follows: permutations = 1000 and quantile normalization = TRUE. Only samples with $p < 0.05$ were included in the analysis. The proportions of immune cells were visualized using stacked bar charts and boxplots. The Wilcoxon rank-sum test was used to compare differences between groups, while Spearman correlation analysis was applied to examine the associations between key genes (CCND1, TRIP13, MYBL2, and GNB4) and the extent of immune cell infiltration in tumor tissues. Cell types with zero variance were excluded. Correlation results were visualized using heatmaps and linkET network plots.

ssGSEA pathway enrichment analysis

Single-sample gene set enrichment analysis (ssGSEA) was performed using the GSVA package based on the MSigDB Hallmark gene set (h.all.v7.5.1.symbols.gmt). The parameters were set as kcdf = "Gaussian" and absRanking = TRUE. Pathway activity scores were correlated with key gene expression using Spearman correlation analysis and visualized with boxplots and heatmaps.

Single-cell RNA sequencing analysis

Single-cell RNA sequencing data were processed using the Seurat package, encompassing quality control, normalization, dimensionality reduction via t-SNE, and clustering. Cell types were annotated based on established marker genes, and the expression patterns of key genes were examined across the identified cell populations.To further evaluate epithelial cell heterogeneity, canonical epithelial markers were first examined in tumor and normal groups. Because very few normal epithelial marker-positive cells were identified in this dataset, tumor-derived epithelial cells were extracted and subjected to secondary clustering. Marker genes for each tumor epithelial subcluster were identified using the FindAllMarkers function in Seurat, and the expression patterns of CCND1, TRIP13, MYBL2, and GNB4 across tumor epithelial subclusters were visualized using DotPlot and FeaturePlot analyses.Cell-cell communication analysis was performed using CellChat. Tumor-derived epithelial cells, lymphocytes, macrophages, and smooth muscle cells were included, whereas cell populations with very low cell numbers were excluded from the analysis. The CellChat human ligand-receptor database was used to infer potential intercellular communication networks. Interaction number, interaction strength, epithelial cell-derived ligand-receptor interactions, incoming and outgoing signaling patterns, and representative pathway-specific ligand-receptor contributions were visualized to characterize communication patterns within the cervical cancer tumor microenvironment.

Cell culture

Human cervical cancer cell lines (SiHa and HeLa) were obtained from Procell Life Science & Technology Co., Ltd., and normal cervical epithelial cells (HCerEpic) were purchased from the Beijing Baina Chuanglian Biotechnology Institute. Cells were cultured in DMEM supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin at 37 °C in a humidified incubator with 5% CO_2. The culture medium was replaced or cells were passaged every two days.

siRNA transfection

Small interfering RNAs (siRNAs) targeting GNB4 and the corresponding negative control (si-NC) were purchased from CHENEN BIOLOGY, Guiyang, China. SiHa and HeLa cells were seeded in 6-well plates and cultured until they reached 60–70% confluence. Transient transfection was performed using Lipofectamine 3000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Briefly, siRNAs and transfection reagent were separately diluted in Opti-MEM medium and then mixed gently for incubation at room temperature for 15 min. The complexes were subsequently added to the cells and incubated for 6 h, after which the medium was replaced with fresh complete culture medium. After 48 h of transfection, cells were collected for subsequent Western blotting analysis to evaluate knockdown efficiency. The siRNAs with the highest silencing efficiency were selected for further experiments.Primer sequences are listed in Supplementary Table S2.

Cell proliferation assay

Cell proliferation was assessed using the Cell Counting Kit-8 (CCK-8) assay (Meilunbio, Guiyang, China). After transfection with GNB4 siRNA or si-NC for 24 h, SiHa and HeLa cells were harvested and seeded into 96-well plates at a density of 2×10^3-5×10^3 cells/well in 100 µL of complete medium. At the indicated time points (0, 24, 48, and 72 h), 10 µL of CCK-8 reagent was added to each well, followed by incubation at 37 °C for 1–2 h. The absorbance was then measured at 450 nm using a microplate reader. Cell proliferation curves were plotted based on the optical density values.

Clinical tissue samples

Cervical cancer tissues and matched adjacent normal tissues were collected from patients at the Affiliated Hospital of Guizhou Medical University (Approval No. 2023-231-01). Written informed consent was obtained from all participants prior to sample collection. Three tumor tissues and three matched adjacent tissues (approximately 2 cm from the tumor margins) were obtained. All samples were independently evaluated by two experienced pathologists. Patients who had received chemotherapy, radiotherapy, or hormone therapy prior to surgery were excluded.

RT-qPCR analysis

Total RNA was isolated using TRIzol reagent, and cDNA synthesis was performed with the PrimeScript RT Kit (Takara). Quantitative PCR using SYBR Green was carried out under the following cycling conditions: 95 °C for 30 s, then 40 cycles of 95 °C for 5 s and 60 °C for 30 s. GAPDH was used as the internal reference, and relative expression levels were determined using the $2^{-\Delta\Delta Ct}$ method. Primer sequences are listed in Supplementary Table S2.

Western blot analysis

Human cervical cancer tissues and adjacent normal tissues were minced and thoroughly ground in liquid nitrogen, followed by total protein extraction using tissue lysis buffer supplemented with protease inhibitors. HCerEpic, SiHa, and HeLa cells were collected by trypsinization and lysed in RIPA buffer supplemented with protease inhibitors. Samples were incubated on ice for 30 min, followed by sonication and centrifugation at 12,000 × g for 15 min at 4 °C. Supernatants were collected, and protein concentrations were measured using a BCA assay kit (KeyGEN, China). Equal amounts of protein were separated by SDS-PAGE and transferred onto PVDF membranes. Membranes were blocked with 5% non-fat milk for 2 h and then incubated overnight at 4 °C with the corresponding primary antibodies listed in Table S3. Following three washes with TBST, membranes were incubated with HRP-conjugated secondary antibodies at room temperature for 1 h. Protein bands were detected using enhanced chemiluminescence (ECL) reagents and quantified with ImageJ software.

During densitometric analysis, local background signals were subtracted from each band before normalization. The relative expression levels of total proteins were normalized to GAPDH, whereas phosphorylated proteins were normalized to their corresponding total proteins. GAPDH was used as the internal control. The PI3K/AKT pathway inhibitor 3-methyladenine (3-MA, IM0190) was purchased from Beijing Solarbio Science & Technology Co., Ltd. Detailed antibody information is provided in Table S3.

Statistical analysis

All statistical analyses were performed using R software (version 4.3.1) and GraphPad Prism 9.0. The normality of continuous variables was assessed using the Shapiro–Wilk test. For normally distributed data, comparisons between two groups were performed using Student's t-test, whereas non-normally distributed data were analyzed using the Wilcoxon rank-sum test. For paired clinical tissue samples, the paired Student's t-test or Wilcoxon matched-pairs signed-rank test was used according to data distribution. Comparisons among three or more groups were performed using one-way ANOVA followed by Tukey's post hoc test for normally distributed data, or the Kruskal–Wallis test followed by Dunn's post hoc test for non-normally distributed data. Correlations were assessed using Pearson correlation analysis for normally distributed data or Spearman correlation analysis for non-normally distributed data. Receiver operating characteristic (ROC) analysis was conducted using the pROC package. Data are presented as mean ± standard deviation (SD), and a two-sided p value < 0.05 was considered statistically significant.

Results

Identification of DEGs

The raw expression data were first normalized, and the distribution of gene expression levels before and after normalization is shown in Fig. 1A and B. After normalization, the distributions across samples were more consistent compared with the pre-normalization values, indicating that the data quality was adequate for subsequent analyses. The limma package was used to analyze the normalized expression matrix, with screening criteria set as $|\log_2 \text{fold change}| > 1$ and $p < 0.05$. In total, 2,669 differentially expressed genes (DEGs) were identified, including 1,391 upregulated and 1,278 downregulated genes. The heatmap of DEGs (Fig. 1C) revealed distinct expression patterns between tumor and normal tissues, while the volcano plot (Fig. 1D) illustrated the overall distribution of significantly dysregulated genes. Notably, some of the upregulated genes included CDKN2A, SYCP2, and ELM2, whereas SCNN1B and CRISP3 were strongly downregulated in cervical cancer samples. Gene set enrichment analysis (GSEA) was conducted on the entire expression matrix to investigate the biological processes associated with the observed gene expression changes. Pathways related to cell cycle regulation, DNA replication, and histone acetylation-related cancer were enriched in tumor tissues. In contrast, pathways associated with arachidonic acid metabolism, histidine metabolism, phenylalanine metabolism, primary bile acid biosynthesis, and sulfur metabolism were significantly downregulated (Fig. 1E and F). These results suggest that these pathways may play a role in the development of cervical cancer. Overall, this analysis identified a cluster of key DEGs and their associated biological pathways, providing a solid foundation for constructing a co-expression network and selecting major candidate genes for further study.

WGCNA and functional enrichment analysis

To identify co-expression gene modules systematically associated with cervical cancer development, the expression matrix of the GSE63514 dataset was preprocessed, and all genes with a mean FPKM value > 0.5 were retained to reduce noise from low-abundance transcripts. Weighted gene co-expression network analysis (WGCNA) was then applied to construct a gene-based co-expression network related to cervical cancer.

Hierarchical clustering was initially performed to assess sample similarity and detect potential outliers. The clustering pattern was uniform, and no apparent outliers were observed (Fig. 2A), indicating that the dataset was suitable for network construction.

Soft-threshold power selection was subsequently conducted to determine the optimal network parameter. A soft-threshold power of 6 produced a scale-free topology fit index greater than 0.85 and an acceptable mean connectivity (Fig. 2B). Therefore, a power of 6 was employed for subsequent analyses. This parameter was used to hierarchically cluster genes and to identify co-expression modules using the dynamic tree cut algorithm. A total of 17 independent modules were identified, each represented by a distinct color (Fig. 2C). The modules were then correlated with clinical traits (normal versus cervical cancer), and module-trait relationship heatmaps were generated (Fig. 2D). The green module showed a strong positive correlation with cervical cancer ($r = 0.81$, $p = 3 \times 10^{-13}$), whereas the cyan module demonstrated a strong negative correlation with tumor status ($r = -0.66$, $p = 2 \times 10^{-19}$), suggesting that these modules may play important regulatory roles in cervical cancer development. The significantly correlated modules were intersected with the differentially expressed genes, resulting in 776 candidate genes (Fig. 2E). These genes exhibited substantial changes in expression and high network connectivity, indicating significant biological relevance. Functional enrichment analysis indicated that the overlapping genes were primarily involved in cancer-related pathways, including cell cycle regulation, cellular senescence, human papillomavirus infection, and the p53 signaling pathway, according to KEGG analysis (Fig. 2F). Gene Ontology analysis revealed that these genes mainly participate in organelle fission, chromosome segregation, nuclear division, and microtubule binding, all of which are processes closely associated with mitosis (Fig. 2G). Overall, WGCNA successfully identified key gene modules and candidate genes most strongly linked to the cervical cancer phenotype, providing a solid foundation for subsequent hub gene screening and mechanistic investigations.

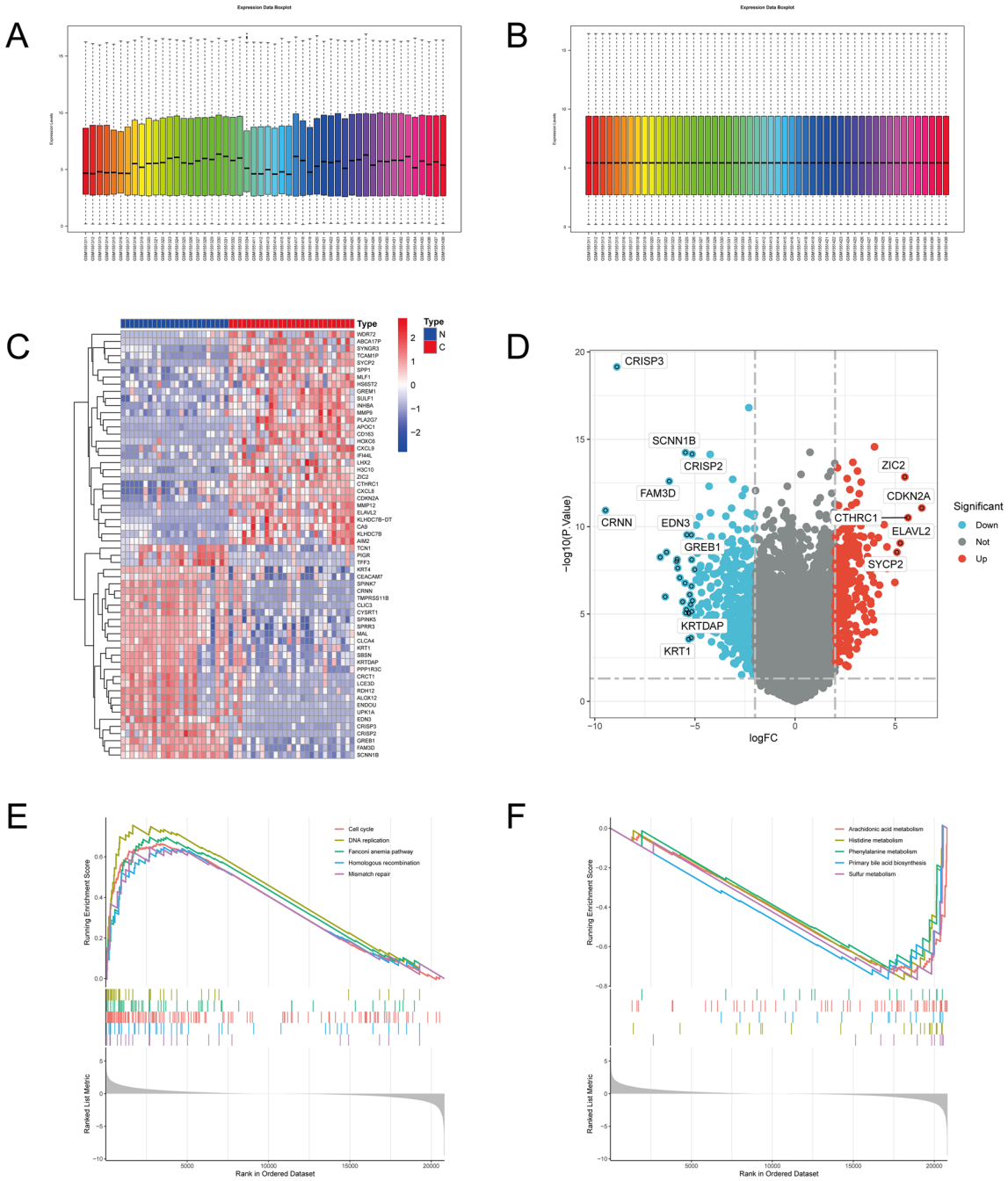

Fig. 1. Identification and functional enrichment analysis of differentially expressed genes. (**A**, **B**) Distribution of gene expression levels before and after normalization. (**C**) Heatmap showing differential expression patterns between tumor and normal tissues. (**D**) Volcano plot displaying significantly upregulated and downregulated genes. (**E**, **F**) GSEA results illustrating significantly enriched pathways associated with cervical cancer.

Identification of hub diagnostic genes

The next step in examining protein-protein interactions among the intersecting genes and identifying potential functional hub genes involved constructing a protein-protein interaction (PPI) network based on the STRING database. The intersecting genes were uploaded to STRING, and interactions with a confidence score ≥ 0.7 were retained. High-confidence interaction data were obtained using the STRINGdb R package. The resulting PPI network contained 153 nodes and 529 edges, forming a relatively tight structure, which suggests a high degree of functional coordination among these genes (Fig. 3A). To identify key regulatory nodes within the network, the betweenness centrality of all genes was calculated. Genes with values exceeding the mean + 0.1 standard deviation were selected as hub gene candidates, resulting in 50 potential hubs. A sub-network of these genes displayed closer and more interconnected patterns (Fig. 3B), indicating that they play a crucial role in maintaining molecular network stability and signal transmission related to cervical cancer. To further refine the candidate

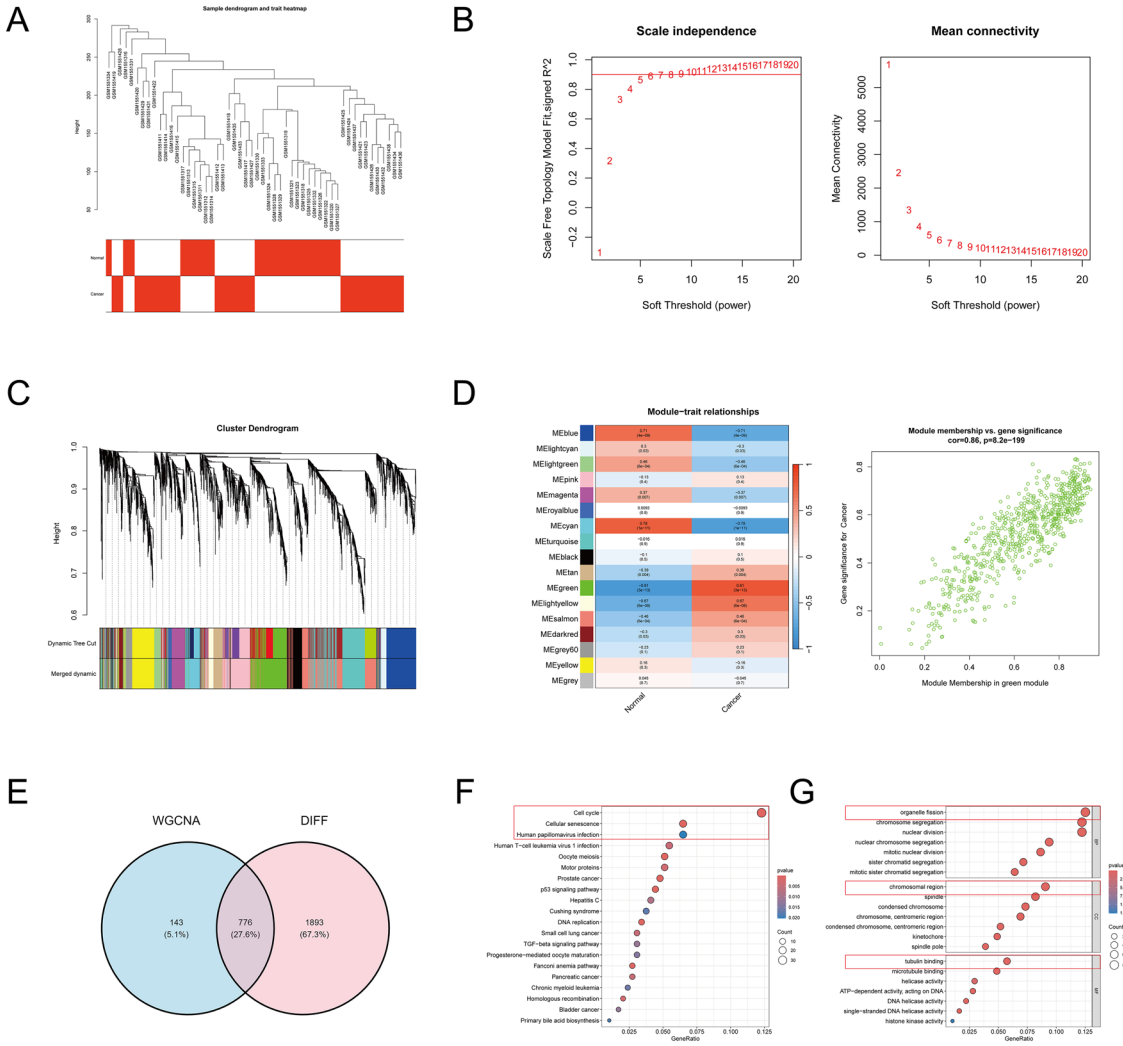

Fig. 2. Identification of key gene modules and functional enrichment analysis based on WGCNA. (**A**) Sample clustering dendrogram showing overall sample similarity and the absence of obvious outliers. (**B**) Soft-threshold selection analysis, showing scale-free topology fit index (left) and mean connectivity (right), indicating an optimal power of 6. (**C**) Gene dendrogram and module assignment identified by dynamic tree cut, with different colors representing distinct modules. (**D**) Module-trait relationship heatmap showing correlations between modules and cervical cancer status, with the green module exhibiting a strong positive association. (**E**) Venn diagram showing the overlap between DEGs and genes from key WGCNA modules. (**F**) KEGG pathway enrichment analysis of intersecting genes. (**G**) Gene Ontology enrichment analysis of intersecting genes.

genes and enhance selection rigor, three machine learning methods with distinct feature-selection processes were applied: support vector machine-recursive feature elimination (SVM-RFE), least absolute shrinkage and selection operator (LASSO) regression, and random forest analysis. The highest cross-validation accuracy in SVM-RFE was achieved using 10 features (Fig. 3C). LASSO regression identified nine genes with non-zero coefficients at the optimal penalty parameter λ (Fig. 3D). Concurrently, random forest analysis selected the top 20 features based on mean decrease in Gini index (Fig. 3E). By integrating the results of these three algorithms, four common genes CCND1, TRIP13, MYBL2, and GNB4 were consistently identified (Fig. 3F). Their repeated selection across multiple methods indicates strong stability and reliability. Together, these findings suggest that CCND1, TRIP13, MYBL2, and GNB4 are promising diagnostic biomarkers and potential molecular classifiers for cervical cancer.

External validation of diagnostic genes

The expression profiles and diagnostic significance of the four key genes (CCND1, TRIP13, MYBL2, and GNB4) were further validated using the independent GSE52903 cohort. To explore the potential relationships among these genes, correlation analysis was performed using the normalized and batch-corrected expression data from GSE63514 and GSE52903. MYBL2 was positively correlated with TRIP13, whereas CCND1 showed weak or negative correlations with the other genes (Fig. 4A). GeneMANIA analysis further suggested that the four genes

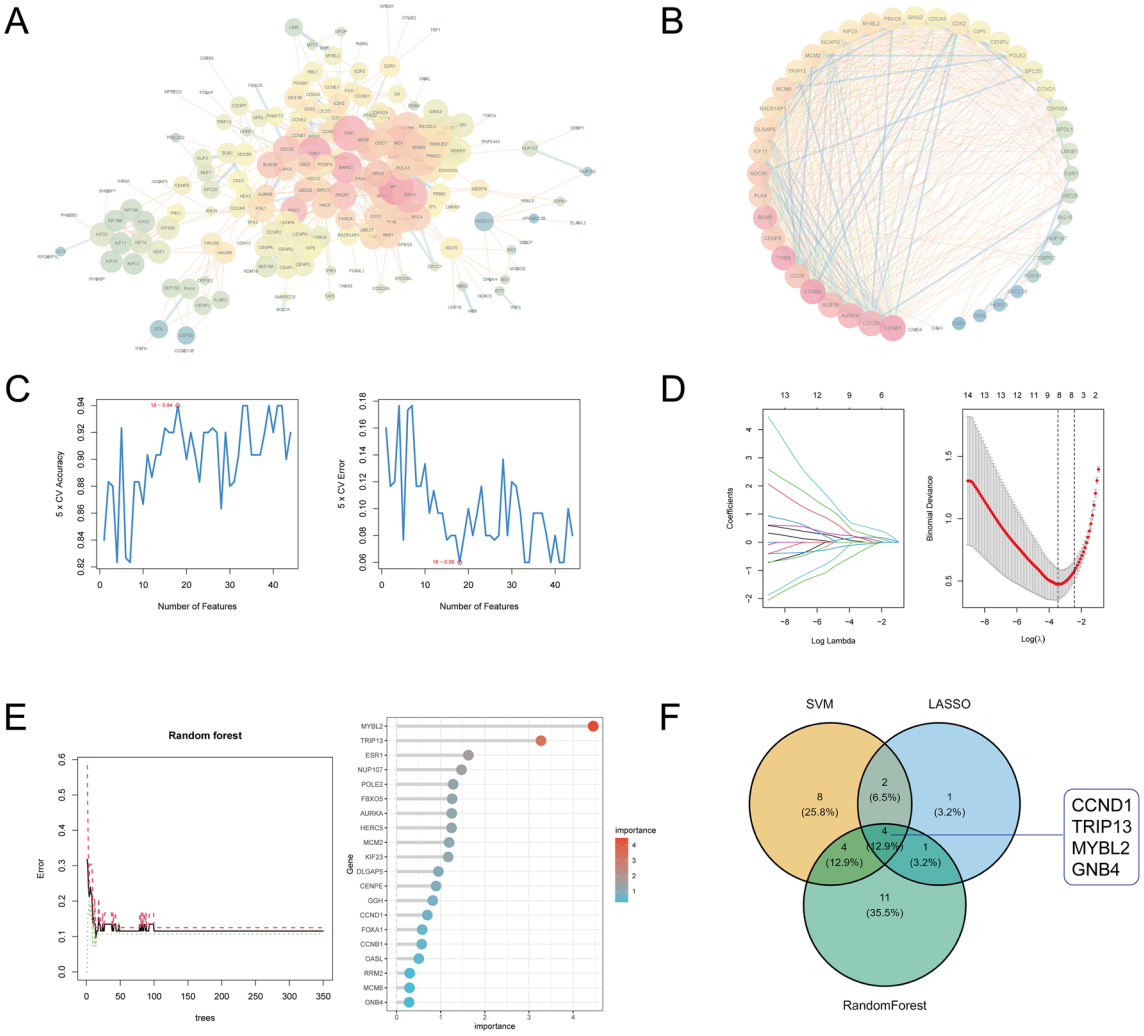

Fig. 3. Identification of potential hub diagnostic genes based on PPI network and machine learning algorithms. (**A**) Comprehensive PPI network constructed using the STRING database. Nodes represent proteins, and edges indicate protein-protein interactions. Different colors denote distinct network clusters. (**B**) Core PPI subnetwork consisting of 50 hub genes identified based on betweenness centrality. Node color intensity reflects centrality, and edges represent high-confidence interactions. (**C**) Feature selection process using the SVM-RFE algorithm to determine the optimal number of features. (**D**) Feature selection and optimal gene identification using LASSO regression. (**E**) Gene importance ranking determined by random forest analysis. (**F**) Venn diagram showing overlapping genes identified by the three machine learning algorithms, yielding four key candidates.

were directly or indirectly associated with cell-cycle- and proliferation-related genes, including CDK4, CDK6, RB1, CCNB1, E2F4, and CDKN1A (Supplementary Fig. S1).

Differential expression analysis showed that TRIP13, MYBL2, and GNB4 were upregulated, whereas CCND1 was downregulated in cervical cancer tissues (Fig. 4B). ROC analysis showed that MYBL2 and TRIP13 had strong diagnostic performance, with AUC values of 0.987 and 0.958, respectively, while CCND1 and GNB4 showed moderate diagnostic performance, with AUC values of 0.768 and 0.732, respectively (Fig. 4C). The four-gene nomogram model achieved an AUC of 0.989, indicating better diagnostic performance than any individual gene (Fig. 4D,E). To further assess the diagnostic performance of the four-gene model, additional ROC analyses were conducted in the GSE63514 training cohort and TCGA-CESC dataset. The nomogram based on CCND1, TRIP13, MYBL2, and GNB4 showed excellent discrimination in GSE63514 (Supplementary Fig. S2A–C). TCGA-CESC was used as an exploratory validation cohort, and a four-gene diagnostic score was calculated based on gene expression patterns. ROC analysis indicated good discriminatory ability (Supplementary Fig. S2D). However, because only three normal samples were available in TCGA-CESC, these results should be interpreted with caution.

Furthermore, decision curve analysis showed that the four-gene model had a higher net benefit than the treat-all and treat-none strategies in both the GSE63514 training cohort and the GSE52903 validation cohort

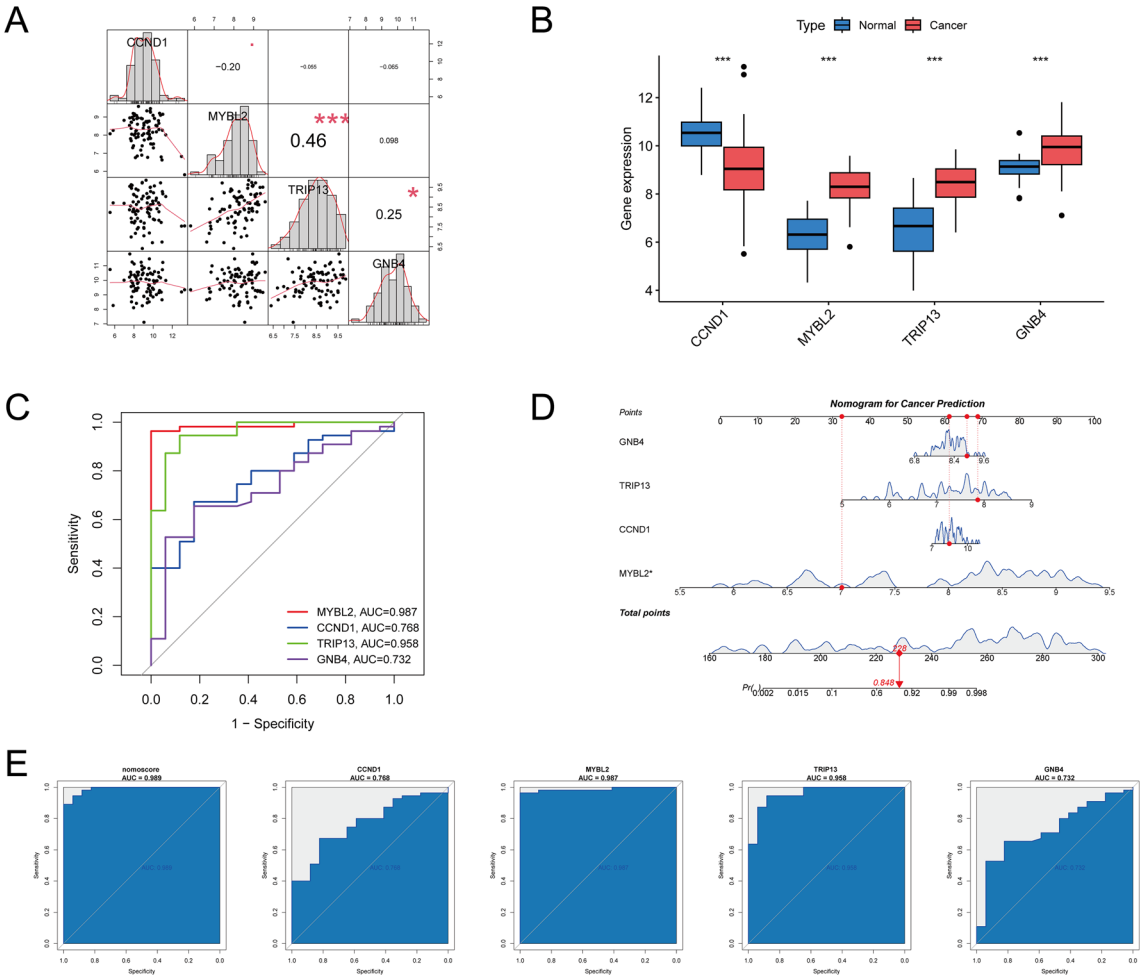

Fig. 4. Diagnostic performance of key genes in cervical cancer. (**A**) Pearson correlation analysis among CCND1, TRIP13, MYBL2, and GNB4. (**B**) Differential expression of the four key genes between normal and cervical cancer tissues. (**C**) ROC curves evaluating the diagnostic performance of individual genes. (**D**) Nomogram model predicting cervical cancer risk based on the four-gene signature. (**E**) Comparison of ROC curves and AUC values between single-gene and combined diagnostic models.

(Supplementary Fig. S3). These results suggest that the four-gene model may serve as a potential diagnostic signature for cervical cancer.

Drug sensitivity and pathway analysis

The associations between the expression levels of CCND1, TRIP13, MYBL2, and GNB4 and the half-maximal inhibitory concentration (IC50) values of various anticancer drugs were investigated to explore the potential roles of these key genes in therapeutic response. A significant positive correlation was observed between CCND1 expression and the IC50 of the TAF1 inhibitor TAF1_5496 ($R = 0.31$, $p = 0.000288$) (Fig. 5A), indicating that samples with lower CCND1 expression may be more sensitive to this inhibitor. In contrast, TRIP13 expression was significantly and inversely correlated with the IC50 of the Wnt pathway inhibitor WIKI4_1940 (R = -0.52, $p = 1.34 \times 10^{-9}$) (Fig. 5B), suggesting that tumors with higher TRIP13 expression are more sensitive to Wnt pathway inhibition. Similarly, MYBL2 expression showed a negative correlation with the IC50 of the proteasome inhibitor bortezomib (R = -0.39, $p = 1.69 \times 10^{-6}$) (Fig. 5C), and GNB4 expression was negatively associated with the IC50 of the estrogen receptor degrader fulvestrant (R = -0.40, $p = 1.75 \times 10^{-9}$) (Fig. 5D). These results suggest that the expression levels of the key genes may be associated with predicted sensitivity to specific anticancer agents. However, given that these correlations were weak to moderate and were derived from computational prediction, these findings should be interpreted as exploratory and require further validation through in vitro drug treatment experiments.

To further investigate the molecular mechanisms underlying these correlations, single-sample gene set enrichment analysis (ssGSEA) was performed to assess differences in pathway activities between tumor and normal tissues. As shown in Fig. 5E, multiple cancer-related signaling pathways were significantly dysregulated in cervical cancer tissues, with the PI3K/AKT pathway notably upregulated ($p < 0.01$). Correlation analysis further demonstrated that the expression levels of TRIP13, MYBL2, and GNB4 were positively associated with

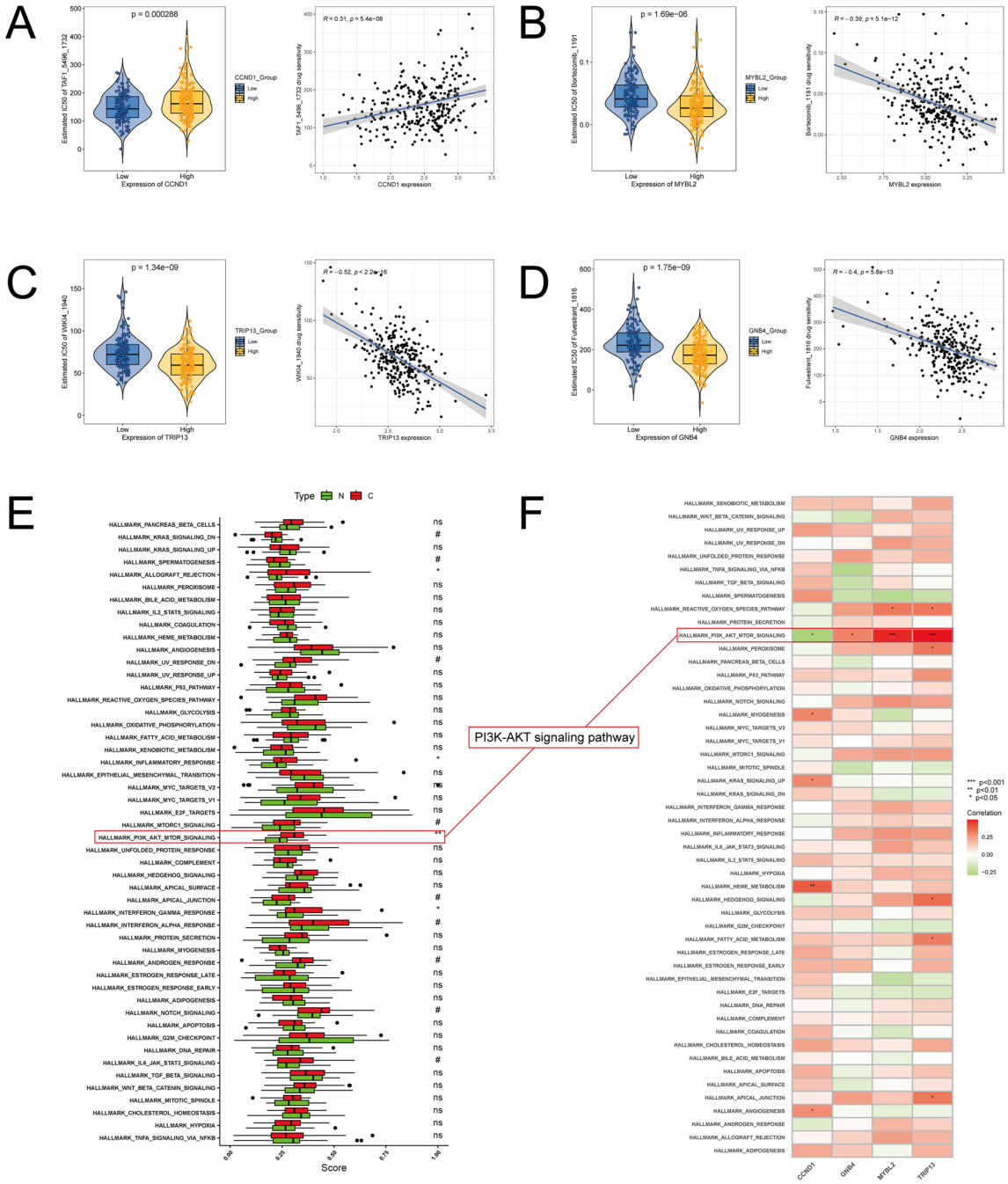

Fig. 5. ssGSEA-based functional and drug sensitivity analysis. (**A**) Positive correlation between CCND1 expression and sensitivity to the TAF1 inhibitor. (**B**) Association between TRIP13 expression and sensitivity to the Wnt pathway inhibitor. (**C**) Correlation between MYBL2 expression and sensitivity to the proteasome inhibitor bortezomib. (**D**) Correlation between GNB4 expression and sensitivity to the estrogen receptor degrader fulvestrant. (**E**) Boxplots showing differences in pathway enrichment scores between cervical cancer and normal tissues. (**F**) Heatmap illustrating correlations between key gene expression and pathway enrichment scores, highlighting the PI3K/AKT pathway.

the enrichment scores of the PI3K/AKT pathway ($p < 0.01$) (Fig. 5F), suggesting that these genes may contribute to cervical cancer progression by regulating cell proliferation, survival, and apoptosis through this pathway.

Immune infiltration analysis

To investigate the potential roles of key biomarkers in the cervical cancer immune microenvironment, the CIBERSORT algorithm was used to estimate the relative proportions of 22 immune cell types in tumor and adjacent normal tissues. Differences in immune cell composition between cervical cancer and normal samples are shown in Fig. 6A. Differential analysis revealed that activated memory CD4$^+$ T cells, follicular helper T

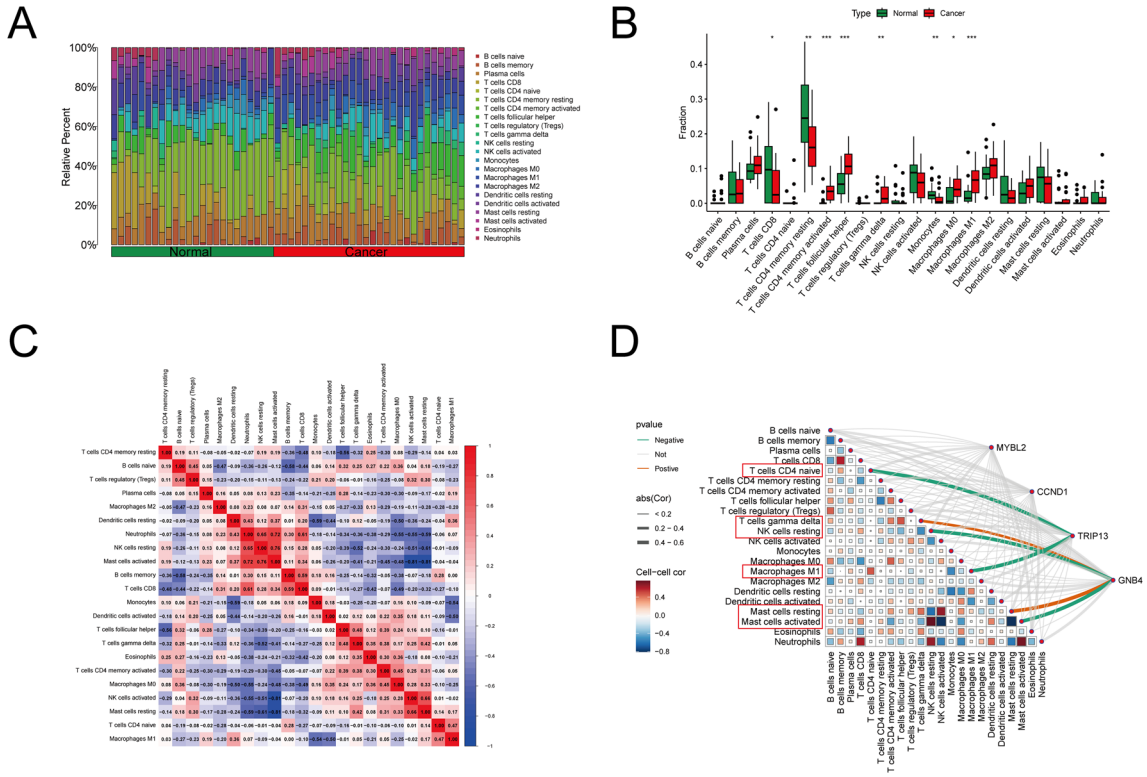

Fig. 6. Immune infiltration analysis. (**A**) Stacked bar plot showing the proportions of immune cells in cervical cancer and normal tissues. (**B**) Comparison of immune cell infiltration between tumor and normal samples. (**C**) Correlation heatmap of immune cell subsets. (**D**) Network visualization showing correlations between key genes and immune cell infiltration.

cells, γδ T cells, M0 macrophages, and M1 macrophages were significantly elevated ($p < 0.05$), whereas resting memory $CD4^+$ T cells, $CD8^+$ T cells, and monocytes were significantly reduced ($p < 0.05$) in tumor tissues (Fig. 6B), suggesting that an imbalance in the immune microenvironment may be closely associated with cervical cancer progression. Correlation analysis among immune cell subsets revealed statistically significant associations (Fig. 6C). Notably, activated mast cells exhibited a strong positive correlation with resting natural killer cells, suggesting that these two cell types may function cooperatively in tumor immunoregulation. Furthermore, the relationships between key genes (CCND1, TRIP13, MYBL2, and GNB4) and immune cell infiltration were evaluated (Fig. 6D). TRIP13 expression was negatively correlated with M1 macrophages and activated $CD4^+$ T cells ($p < 0.05$). In contrast, GNB4 expression was positively correlated with activated mast cells and γδ T cells, and negatively correlated with resting natural killer cells ($p < 0.05$). Overall, these findings suggest that the identified key genes may modulate the immune microenvironment in cervical cancer by regulating the infiltration and activity of specific immune cell subpopulations, thereby contributing to tumor progression. This evidence further supports their potential as therapeutic targets in immunotherapy.

Single-cell validation

To validate the cellular heterogeneity and expression patterns of the identified diagnostic genes in cervical cancer, single-cell RNA sequencing data from GSE168652 were systematically analyzed. Visualization using t-distributed stochastic neighbor embedding (t-SNE) showed distinct distribution patterns between tumor and normal tissues, suggesting that cellular heterogeneity may play a crucial role in cervical cancer development (Fig. 7A). Subsequent clustering analysis identified 12 distinct cell subpopulations based on transcription patterns (right panel of Fig. 7A). Cell types were annotated using known marker genes, including CDH1, CD163, FCGR2A, COL1A2, CDKN2A, and CD27. The main cell populations comprised epithelial cells, endothelial cells, smooth muscle cells, fibroblasts, lymphocytes, and macrophages (Fig. 7B–D). The expression profiles of the most significant diagnostic genes GNB4, TRIP13, MYBL2, and CCND1 were subsequently visualized using single-cell expression distributions (Fig. 7E). It is noteworthy that the three key genes, GNB4, TRIP13, and MYBL2, were highly enriched in epithelial cell clusters, suggesting their involvement in the pathological proliferation and differentiation of cervical epithelial cells. In contrast, CCND1 was mostly found in fibroblasts and partially in epithelial cells, which may indicate its involvement in tumor-stroma interactions and regulation of the microenvironment.

To further improve the single-cell analysis, epithelial marker expression was first examined in tumor and normal groups. Epithelial marker-positive cells were mainly distributed in the tumor group, whereas very few were detected in the normal group (Supplementary Fig. S4A). Therefore, tumor-derived epithelial cells were

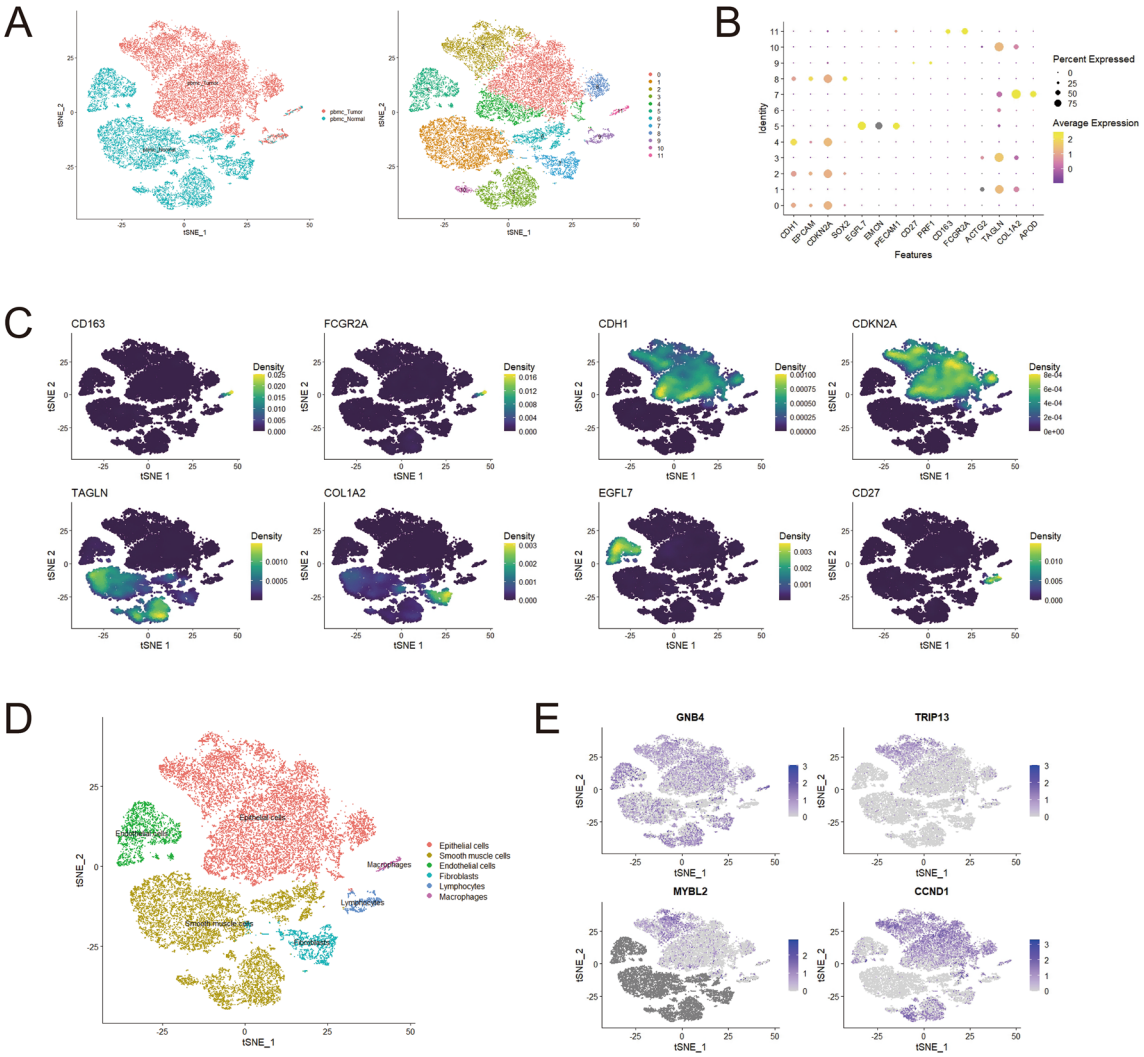

Fig. 7. Single-cell transcriptome analysis. (**A**) t-SNE visualization showing cell distribution and clustering patterns in tumor and normal tissues. (**B**) Bubble plot illustrating cell cluster annotation based on representative marker genes. (**C**) t-SNE density plots of typical marker gene expression. (**D**) Cell type annotation map of identified clusters. (**E**) t-SNE plots showing single-cell expression of CCND1, GNB4, TRIP13, and MYBL2.

further extracted for subcluster analysis. Six tumor epithelial subclusters were identified, showing distinct marker gene expression patterns (Supplementary Fig. S4B,C,F). CCND1, TRIP13, MYBL2, and GNB4 displayed heterogeneous and subcluster-specific expression patterns in tumor epithelial cells (Supplementary Fig. S4D,E). CellChat analysis was further performed to explore cell-cell communication within the cervical cancer tumor microenvironment. Active interactions were observed among epithelial cells, lymphocytes, macrophages, and smooth muscle cells, particularly involving epithelial cells and macrophages (Supplementary Fig. S5A–C). Epithelial cell-derived ligand-receptor interactions and incoming/outgoing signaling patterns suggested the involvement of immune- and microenvironment-related pathways, including SPP1, CXCL, MIF, CCL, TGFβ, MHC-I, and MHC-II (Supplementary Fig. S5D–F). Contribution analysis further identified representative ligand-receptor pairs, such as SPP1-CD44, CXCL16-CXCR6, MIF-($CD74^+CXCR4$), and MIF-($CD74^+CD44$) (Supplementary Fig. S5G–I).

Collectively, the single-cell transcriptomic analysis confirmed the presence of heterogeneous expression patterns of the key diagnostic genes at the cellular level, indicating that these genes may contribute to cervical cancer development through cell type-specific regulatory mechanisms. The additional epithelial subcluster and CellChat analyses further suggest that malignant epithelial cells may interact with immune and stromal-related components in the cervical cancer microenvironment. These findings provide valuable cellular-level insights for future mechanistic studies.

Experimental validation of gene expression

In order to confirm the differential expression profiles of the vital diagnostic genes (CCND1, GNB4, TRIP13, and MYBL2) identified by bioinformatics analysis, clinical specimens and cell lines were used to perform

Western blotting and RT-qPCR. Western blotting was performed using three paired cervical cancer and adjacent normal tissue samples, while RT-qPCR validation was expanded to 10 paired clinical tissue samples to improve the reliability of expression validation. In paired clinical samples, CCND1 expression was lower in cervical cancer tissues than in adjacent normal tissues, whereas TRIP13, MYBL2, and GNB4 were upregulated in tumor tissues (Fig. 8A,C–F; $p < 0.01$ or $p < 0.001$). The expanded RT-qPCR results from 10 paired tissue samples further supported the differential expression patterns of these four genes in cervical cancer.Further evidence from cervical cancer cell lines (SiHa and HeLa) and normal cervical epithelial cells (HCerEpic) indicated that CCND1 expression in both cancer cell lines was significantly decreased compared to normal cells (Fig. 8B,G; $p < 0.01$ or $p < 0.001$). Conversely, TRIP13 and MYBL2 expression levels were markedly higher in SiHa and HeLa cells (Fig. 8B,H–I; $p < 0.01$ or $p < 0.001$). GNB4 expression was also significantly increased in HeLa cells (Fig. 8B,J; $p < 0.001$). These experimental results are consistent with observations from the bioinformatics analysis, further confirming that CCND1, GNB4, TRIP13, and MYBL2 are differentially expressed in cervical cancer tissues and cell lines. Taken together, these findings provide additional experimental evidence supporting the potential diagnostic relevance of these four genes in cervical cancer.

Effects of PI3K/AKT inhibition on key genes

To examine the regulatory relationship between the key genes and the PI3K/AKT signaling pathway, the cervical cancer cell lines SiHa and HeLa were treated with the pathway inhibitor 3-methyladenine (3-MA, 50 μM) for 24 h.The concentration and treatment duration were selected based on previous studies using 3-MA as a

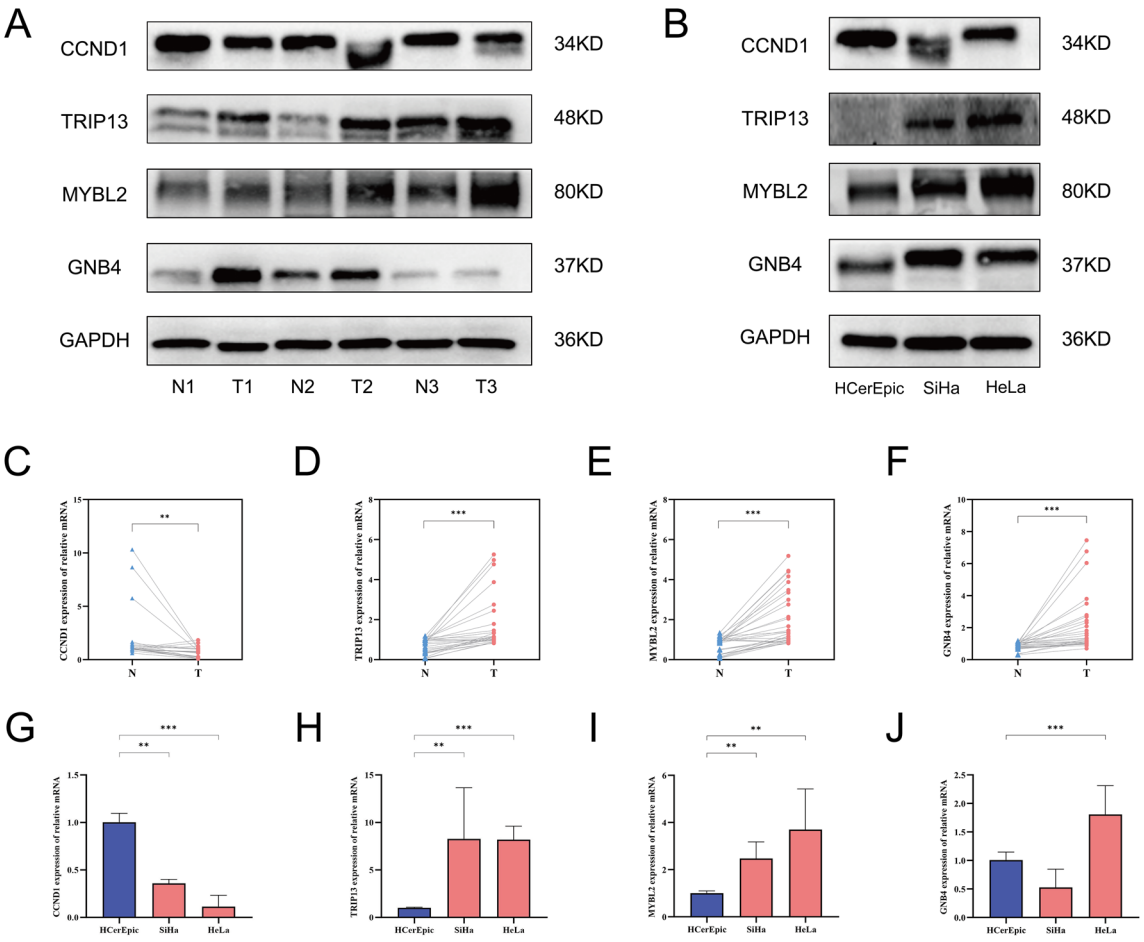

Fig. 8. Experimental validation of key gene expression in cervical cancer. (**A**) Protein expression levels of CCND1, GNB4, TRIP13, and MYBL2 in cervical cancer tissues (C) and adjacent normal tissues (P) detected by Western blotting. (**B**) Protein expression levels of CCND1, GNB4, TRIP13, and MYBL2 in cervical cancer cell lines (SiHa and HeLa) and normal cervical epithelial cells (HCerEpic). (**C**–**F**) RT-qPCR analysis of CCND1, TRIP13, MYBL2, and GNB4 mRNA expression in 10 paired cervical cancer tissues and adjacent normal tissues. (**G**–**J**) mRNA expression levels of CCND1, TRIP13, MYBL2, and GNB4 in cervical cancer cell lines and normal cervical epithelial cells. For clinical tissue samples, statistical significance was determined using the Wilcoxon matched-pairs signed-rank test. For cell experiments, data are presented as mean ± SD from three independent experiments, and statistical analysis was performed using the Kruskal-Wallis test followed by Dunn's post hoc test. T, tumor tissue; N, adjacent normal tissue. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; ns, not significant.

PI3K/autophagy inhibitor under similar in vitro conditions[14,15]. Subsequently, the expression levels of CCND1, TRIP13, MYBL2, and GNB4 were assessed using Western blotting and RT-qPCR. In HeLa cells, CCND1 protein expression significantly increased upon 3-MA treatment, whereas the protein levels of TRIP13, MYBL2, and GNB4 were markedly reduced. These changes were further confirmed using densitometric analysis (Fig. 9A,B). In SiHa cells, a similar trend was observed: CCND1 was upregulated, while TRIP13, MYBL2, and GNB4 were downregulated following 3-MA treatment (Fig. 9C,D). These results suggest that inhibition of PI3K/autophagy-

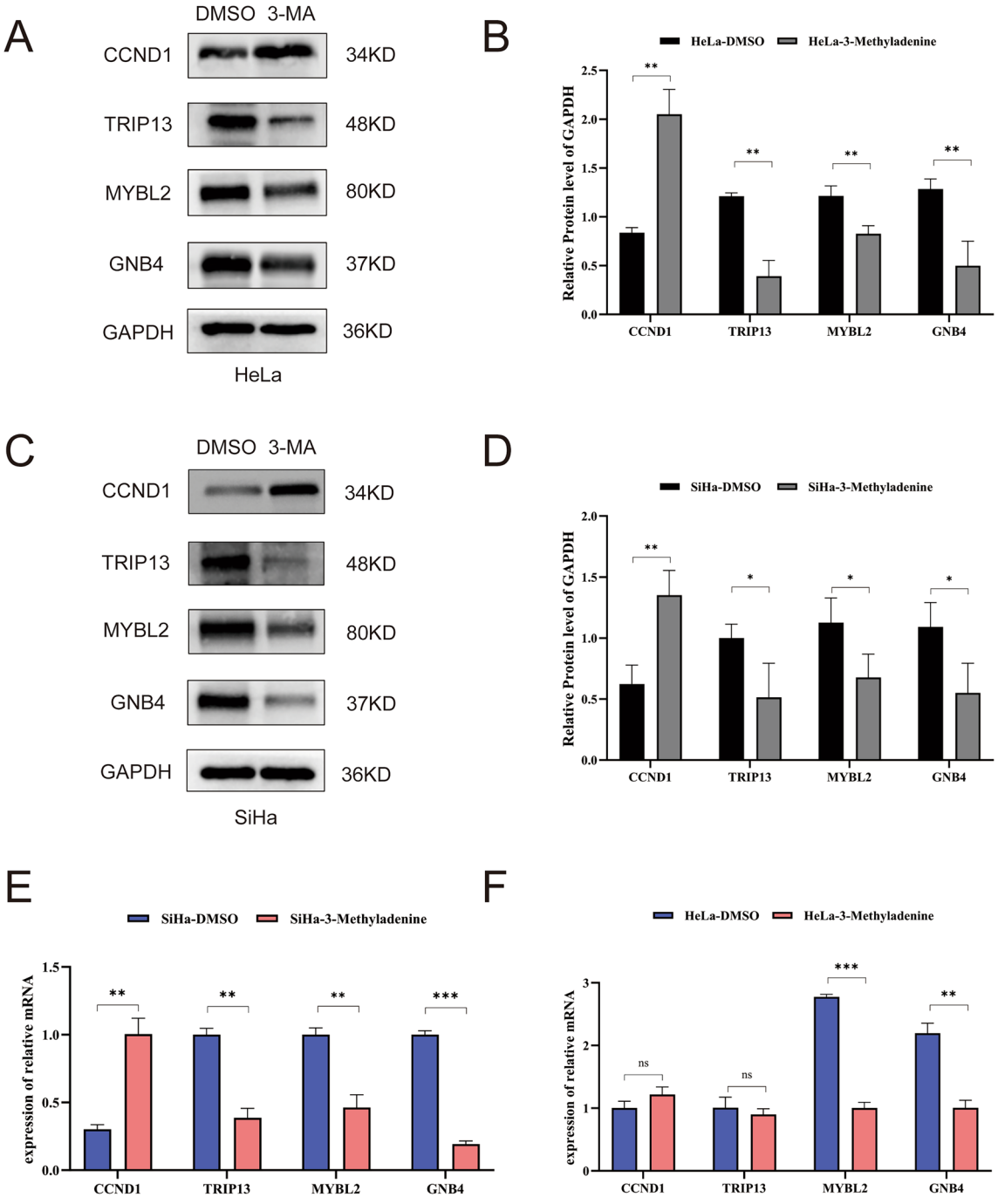

Fig. 9. Effects of PI3K/AKT pathway inhibition on key gene expression in SiHa and HeLa cells. (**A**) Protein expression levels of CCND1, TRIP13, MYBL2, and GNB4 in HeLa cells following 3-MA treatment. (**B**) Quantitative densitometric analysis of protein bands in HeLa cells. (**C**) Protein expression levels of CCND1, TRIP13, MYBL2, and GNB4 in SiHa cells after 3-MA treatment. (**D**) Quantitative densitometric analysis of protein bands in SiHa cells. (**E**) RT-qPCR analysis of mRNA expression levels of key genes in SiHa cells. (**F**) RT-qPCR analysis of mRNA expression levels of key genes in HeLa cells. Data are presented as mean ± SD from three independent experiments. Statistical significance was determined using the Mann-Whitney U test for two-group comparisons. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; ns, not significant.

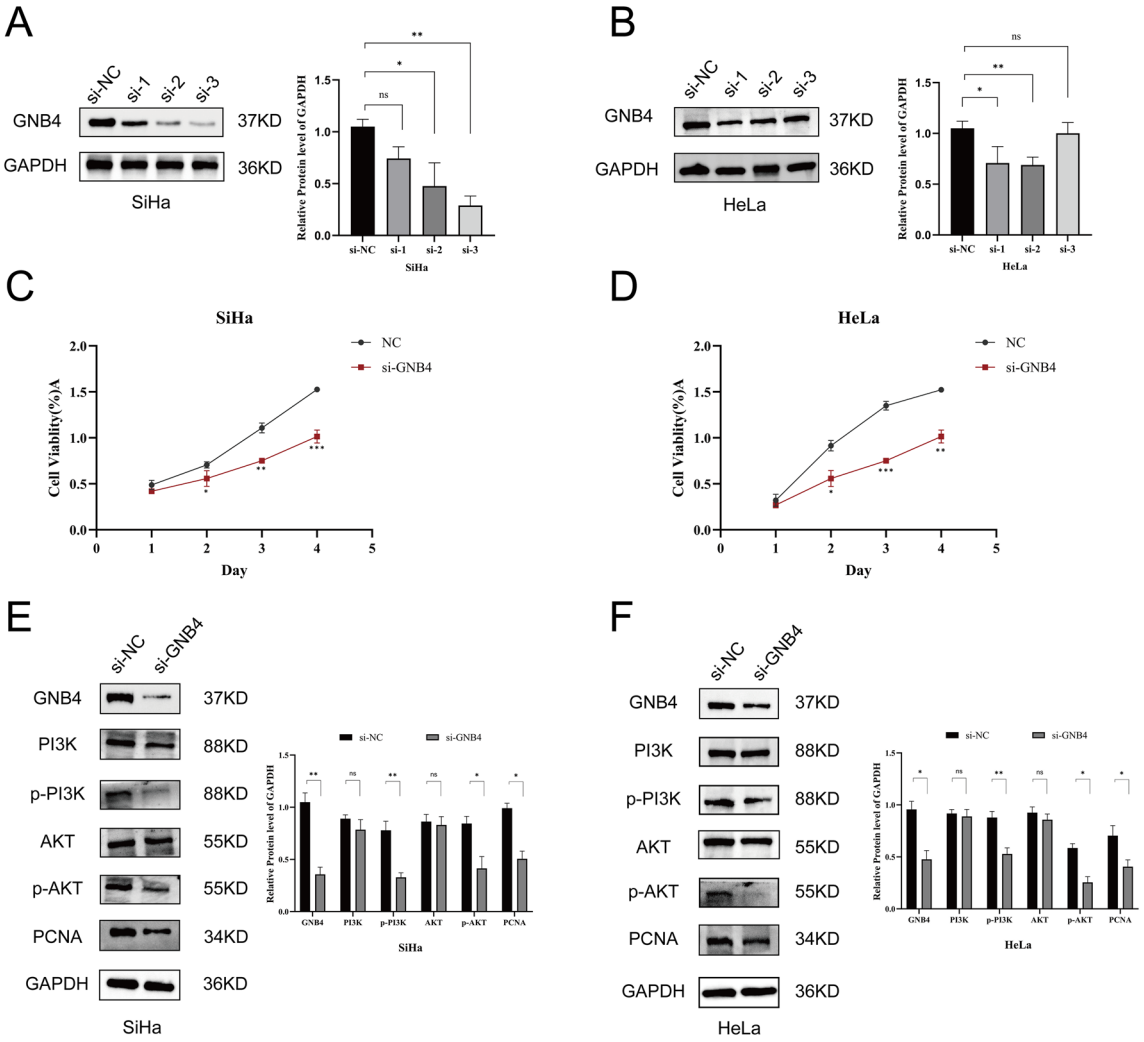

Fig. 10. Effects of GNB4 knockdown on cervical cancer cell proliferation and PI3K/AKT pathway activation. (**A**, **B**) Western blot analysis and densitometric quantification of GNB4 protein expression in SiHa and HeLa cells after transfection with three different GNB4 siRNAs. (**C**, **D**) CCK-8 assays showing the effects of GNB4 knockdown on the proliferative activity of SiHa and HeLa cells. (**E**, **F**) Western blot analysis and densitometric quantification of GNB4, PI3K, p-PI3K, AKT, p-AKT, and PCNA protein expression in SiHa and HeLa cells following GNB4 knockdown. Data are presented as mean ± SD from three independent experiments. For comparisons among multiple groups, statistical significance was determined using the Kruskal-Wallis test followed by Dunn's post hoc test. For two-group comparisons, the Mann-Whitney U test was used. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; ns, not significant.

related signaling may be associated with altered expression of the four key genes. RT-qPCR analysis also confirmed these observations at the transcriptional level. CCND1 mRNA expression was markedly increased, whereas TRIP13, MYBL2, and GNB4 mRNA levels were substantially decreased in SiHa cells after 3-MA treatment (Fig. 9E). Similar expression patterns were observed in HeLa cells (Fig. 9F). Taken together, these findings indicate that the expression of CCND1, TRIP13, MYBL2, and GNB4 may be associated with PI3K/AKT pathway activity in cervical cancer cells. However, these results do not establish direct regulatory relationships between the PI3K/AKT pathway and each individual gene. Further gain- or loss-of-function experiments are needed to clarify whether CCND1, TRIP13, and MYBL2 directly participate in PI3K/AKT-mediated cervical cancer progression.

GNB4 knockdown suppresses cell proliferation and PI3K/AKT activation

To investigate the biological role of GNB4 in cervical cancer cells, three different siRNAs targeting GNB4 were transfected into SiHa and HeLa cells, and knockdown efficiency was assessed by Western blotting. In SiHa cells, all three siRNAs reduced GNB4 protein expression to different extents, with si-GNB4-3 showing the highest silencing efficiency. In HeLa cells, both si-GNB4-1 and si-GNB4-2 effectively decreased GNB4 protein expression, among which si-GNB4-2 exhibited more stable knockdown efficiency. These siRNAs were therefore selected for subsequent experiments (Fig. 10A,B).The CCK-8 assay demonstrated that GNB4 knockdown

significantly inhibited the proliferative activity of both SiHa and HeLa cells compared with the NC group, and this inhibitory effect became more pronounced over time, indicating that GNB4 is required for the maintenance of cervical cancer cell proliferation (Fig. 10C,D).To further explore the underlying mechanism, the expression of PI3K/AKT pathway-related proteins was examined by Western blotting. As shown in Fig. 10E,F, silencing of GNB4 markedly decreased the protein levels of p-PI3K, p-AKT, and PCNA in both SiHa and HeLa cells, whereas the total levels of PI3K and AKT remained largely unchanged. These results suggest that GNB4 knockdown suppresses cervical cancer cell proliferation, at least in part, by inhibiting PI3K-AKT pathway activation rather than altering total PI3K or AKT expression.

Discussion

One of the most common cancers of the female reproductive system, with a complex pathogenesis and significant heterogeneity, cervical cancer remains a major cause of morbidity and reduced quality of life in women[16]. Although overall survival rates have improved over the last several decades, low rates of early detection, poor responses to therapy, and high risks of recurrence remain important clinical challenges[17]. For this reason, there is a need to identify reliable molecular biomarkers that can facilitate early detection and appropriate treatment of cervical cancer. With the advent of high-throughput sequencing and microarray technologies, bioinformatics-based tools have proven extremely useful in the discovery of cancer biomarkers[18]. In this study, we systematically evaluated multiple cervical cancer datasets (GSE63514 and GSE52903) and integrated differential expression analysis, WGCNA, and machine learning methods. Using this multi-layered approach, the four most valuable diagnostic biomarkers CCND1, GNB4, TRIP13, and MYBL2 were identified and validated. These biomarkers performed consistently across independent datasets, indicating their potential clinical relevance for both the diagnosis and therapeutic targeting of cervical cancer.

CCND1 has been identified as one of the predominant diagnostic genes in the present study. Unlike its typical oncogenic role in most solid tumors, which usually leads to overexpression and induction of cell growth[19,20], CCND1 was significantly downregulated in cervical cancer. Previous studies have shown that specific microenvironmental factors or viral infections, including HPV infection, can inhibit the expression and function of CCND1 through upstream regulatory mechanisms, such as inhibition of the pRb pathway by E6/E7, which causes cell cycle arrest[21]. In addition, CCND1 can undergo alternative splicing to generate different cyclin D1 isoforms, including cyclin D1a and cyclin D1b, which differ in their C-terminal regions and may exhibit distinct biological behaviors[22,23]. Cyclin D1 is also regulated by post-translational modifications, particularly phosphorylation-dependent ubiquitination, which affects its stability, degradation, and potentially its electrophoretic mobility during Western blot analysis[24]. Therefore, the atypical CCND1 band migration observed in a limited number of samples may be related to isoform variation or post-translational modification. This particular expression pattern may suggest different mechanisms of cell cycle regulation in cervical cancer, which require further investigation. The TRIP13 and MYBL2 proteins have emerged as key molecules in cancer research[25,26]. Both genes were highly upregulated in cervical cancer samples and cell lines and exhibited strong diagnostic potential. TRIP13 has been shown in previous studies to promote chromosomal instability and chemoresistance, thereby enhancing tumor growth and contributing to treatment failure[27]. MYBL2 is closely associated with tumor proliferation and metastatic potential and serves as a poor prognostic indicator in many cancers[28]. Additional insights into their roles in drug resistance and immune evasion could provide novel molecular targets for personalized therapy. It is noteworthy that this study is the first to report the high expression of GNB4 in cervical cancer. As a member of the G protein beta subunit family, GNB4 participates in signaling pathways that regulate cell proliferation and migration[29]. However, its role in cancers, particularly cervical cancer, remains poorly understood. We propose that GNB4 may serve as a novel biomarker and a potential therapeutic target.

The potential relationships among CCND1, GNB4, TRIP13, and MYBL2 may partly explain why these genes formed an effective diagnostic combination. Although the four genes did not all show strong pairwise correlations, they represented complementary biological processes. CCND1 is closely related to cell-cycle regulation, TRIP13 is associated with chromosomal instability and cell division, MYBL2 participates in proliferation-related transcriptional regulation, and GNB4 is involved in G-protein-mediated signal transduction. GeneMANIA analysis further suggested that these genes may be functionally connected through cell-cycle-related networks involving CDK4, CDK6, RB1, CCNB1, and CDKN1A. Therefore, the four-gene model may capture both shared and complementary molecular alterations in cervical cancer, which may contribute to its diagnostic performance. However, these predicted functional associations require further experimental validation.

The tumor microenvironment plays a critical role in cancer development and therapy responses. Growing evidence suggests a significant influence of immune cell infiltration and tumor-immune interactions on clinical outcomes[30]. In our analysis of the immune infiltrate, we demonstrated that several key biomarkers were significantly correlated with the expression of $CD8^+$ T cells, NK cells, and macrophages, suggesting that they may contribute to remodeling the immune microenvironment. Heterogeneous single-cell transcriptomic profiles were also observed across different cell populations. Epithelial cells were the primary sites of GNB4, TRIP13, and MYBL2 expression, indicating their potential involvement in malignant epithelial transformation, whereas CCND1 expression in fibroblasts may suggest a possible role in stromal regulation. Further tumor epithelial subcluster analysis revealed six distinct malignant epithelial subclusters, and CCND1, TRIP13, MYBL2, and GNB4 showed heterogeneous and subcluster-specific expression patterns across these subclusters. These findings suggest that the four candidate genes may reflect different transcriptional states within malignant epithelial cells rather than uniform expression across all tumor cells. In addition, CellChat analysis indicated active communication among epithelial cells, lymphocytes, macrophages, and smooth muscle cells, with epithelial cells and macrophages showing relatively strong interaction activities. Several immune- and microenvironment-related pathways, including SPP1, CXCL, MIF, CCL, TGFβ, MHC-I, and MHC-II, were identified, suggesting

that malignant epithelial cells may interact with immune and stromal-related components in the cervical cancer microenvironment. These results further support the cellular heterogeneity and microenvironmental relevance of the identified diagnostic genes, although further experimental validation is required to clarify their precise regulatory roles. Regarding translational research, the multi-gene diagnostic model developed in this study demonstrated potential clinical utility. The combination model was more consistent and accurate than individual markers, which may help reduce diagnostic bias caused by tumor heterogeneity, an important consideration for early screening and risk stratification. At present, several biomarkers are commonly used or investigated in cervical cancer screening, diagnosis, and disease monitoring, including SCC-Ag, HPV E6/E7 mRNA, and p16/Ki-67. SCC-Ag is mainly used for disease monitoring, therapeutic response evaluation, and recurrence prediction in patients with cervical squamous cell carcinoma; however, its sensitivity for early diagnosis is limited[31]. HPV E6/E7 mRNA testing reflects transcriptionally active HPV infection and has higher specificity than HPV DNA testing in identifying clinically relevant lesions[32]. p16/Ki-67 dual staining is useful for triaging HPV-positive or cytologically abnormal women and may reduce unnecessary colposcopy referral while maintaining good detection performance for high-grade cervical lesions[33,34]. Nevertheless, these markers mainly reflect viral oncogene activity or cell-cycle dysregulation and may not fully capture the transcriptomic heterogeneity of cervical cancer.Compared with these conventional or clinically investigated markers, the four-gene model constructed in the present study integrates host-derived transcriptomic alterations associated with cell-cycle regulation, signal transduction, tumor proliferation, and immune-related features. Therefore, this model may provide complementary molecular information rather than replacing existing screening strategies. In particular, it may help improve risk stratification when used together with HPV-based testing, cytology, or p16/Ki-67 dual staining.In clinical practice, this model may potentially be implemented using platforms such as RT-qPCR, digital PCR, or multiplex fluorescence assays, which offer advantages in simplicity, reproducibility, and turnaround time. These approaches are less costly and more feasible for mass screening and primary care settings than high-throughput sequencing. Furthermore, multi-gene evaluation mitigates the impact of biological variability and technical error, thereby enhancing clinical validity. However, because the present model has not yet been directly compared with SCC-Ag, HPV E6/E7 mRNA, or p16/Ki-67 in a prospective clinical cohort, its clinical superiority cannot be concluded at this stage. Future studies should evaluate the diagnostic performance and cost-effectiveness of this model in combination with existing clinical markers.

Although the four-gene diagnostic model showed high AUC values in both the GSE63514 training cohort and the GSE52903 validation cohort, the potential risk of overfitting should be considered. Decision curve analysis further suggested that the model provided a higher net benefit than the treat-all and treat-none strategies across a wide range of threshold probabilities, supporting its potential clinical usefulness. In addition, TCGA-CESC was incorporated as an exploratory supplementary validation resource. However, only three normal cervical tissue samples were available in TCGA-CESC; therefore, the TCGA-CESC-based ROC result should be interpreted cautiously and cannot replace validation in an independent cohort with sufficient normal controls. Overall, the four-gene model may represent a complementary molecular tool for cervical cancer diagnosis and risk assessment, but further validation in larger, multicenter, prospective clinical cohorts is required before clinical application.

Drug sensitivity analysis further suggested that the expression levels of the key genes may be associated with predicted responsiveness to specific anticancer agents. In particular, higher TRIP13 and MYBL2 expression levels were correlated with lower predicted IC50 values of Wnt pathway inhibitors and bortezomib, respectively. However, these associations were based on computational prediction using oncoPredict and showed weak-to-moderate correlation strengths. Therefore, these findings should be regarded as preliminary clues for potential therapeutic sensitivity rather than direct evidence for treatment stratification. Further in vitro drug treatment experiments and clinical validation are required to confirm their predictive value.

Additionally, ssGSEA analysis revealed that the PI3K/AKT signaling pathway was highly activated in cervical cancer. This pathway is known to regulate proliferation, invasion, metabolic reprogramming, and drug resistance[35]. In the present study, the expression levels of TRIP13, MYBL2, and GNB4 were positively associated with PI3K/AKT pathway enrichment scores, suggesting that these genes may be related to PI3K/AKT pathway activity in cervical cancer. Furthermore, 3-MA treatment altered the expression of CCND1, TRIP13, MYBL2, and GNB4 in both SiHa and HeLa cells, providing preliminary cellular evidence supporting a potential association between these key genes and PI3K/AKT-related signaling. However, these findings should be interpreted cautiously, as they do not demonstrate direct regulatory relationships between the PI3K/AKT pathway and each individual gene.Although the four-gene diagnostic model included CCND1, TRIP13, MYBL2, and GNB4, the functional experiments in this study mainly focused on GNB4. This design was based on the relatively limited evidence regarding the role of GNB4 in cervical cancer, together with its elevated expression and potential diagnostic relevance observed in our bioinformatics and experimental analyses. GNB4 knockdown significantly suppressed cervical cancer cell proliferation and reduced p-PI3K and p-AKT levels, providing preliminary evidence that GNB4 may be involved in PI3K/AKT pathway activation. In contrast, CCND1, TRIP13, and MYBL2 were mainly supported by expression validation, pathway association analysis, and 3-MA inhibitor experiments in the present study. Therefore, their functional roles and direct regulatory relationships with the PI3K/AKT pathway remain to be further investigated using gain- and loss-of-function experiments. In particular, although CCND1 is a well-known cell-cycle regulator, its downregulated expression pattern and specific biological role in cervical cancer require additional mechanistic studies.

Despite the comprehensive analyses and experimental validations performed in this study, several limitations should be noted. First, the clinical sample size was relatively small and requires validation in larger, multicenter cohorts. Second, the mechanistic investigations were primarily based on bioinformatics predictions and preliminary experiments. Further studies, including functional assays, animal models, and prospective clinical trials, are necessary to elucidate the biological roles and therapeutic potential of these genes. In summary, this

study integrated bioinformatics and machine learning methods to identify and validate CCND1, GNB4, TRIP13, and MYBL2 as putative biomarkers for cervical cancer. These genes were found to be associated with immune regulation, drug responsiveness, and tumor progression. The multi-gene diagnostic model demonstrated high accuracy and reliability, suggesting its potential for practical implementation in early diagnosis and risk assessment. More extensive future studies, including clinical data, treatment outcomes, and cost-effectiveness analyses, are warranted to evaluate its real-world clinical applicability and to facilitate the clinical implementation of these biomarkers.

Data availability

The public datasets analysed during the current study are available from the Gene Expression Omnibus database, including GSE63514, GSE52903 and GSE168652. The TCGA-CESC transcriptomic and clinical data analysed in this study are available from the Genomic Data Commons Data Portal. Drug sensitivity-related data used for prediction were derived from the Genomics of Drug Sensitivity in Cancer database and processed using the oncoPredict R package. The experimental data generated during the current study, including RT-qPCR data, Western blot quantification data and membrane strip images corresponding to the Western blot figures, are included in this article and its Supplementary Information files. Additional data supporting the findings of this study are available from the corresponding author upon reasonable request.

Received: 16 April 2026; Accepted: 9 June 2026
Published online: 23 June 2026

References

1. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74** (3), 229–263 (2024).
2. Mayadev, J. S. et al. Global challenges of radiotherapy for the treatment of locally advanced cervical cancer. *Int. J. Gynecol. Cancer*. **32** (3), 436–445 (2022).
3. Yuan, Q. et al. BPHL promotes TNBC stemness by resolving R-loops via POLR2A lactylation inhibition and BARD1-mediated ubiquitination. *Cancer Lett.* **645**, 218388 (2026).
4. Yu, G. et al. A machine learning-driven framework integrating cell death and senescence signatures for multi-target drug design and immunotherapy optimization in ovarian cancer. *NPJ Precis. Oncol.* (2026).
5. Philp, L. et al. Pap tests in the diagnosis of cervical cancer: Help or hinder? *Gynecol. Oncol.* **150** (1), 61–66 (2018).
6. Jiang, P. et al. Big data in basic and translational cancer research. *Nat. Rev. Cancer*. **22** (11), 625–639 (2022).
7. Huang, J. et al. Bioinformatics tools and resources for cancer and application. *Chin. Med. J. (Engl)*. **137** (17), 2052–2064 (2024).
8. Fu, X. et al. Application of machine learning for high-throughput tumor marker screening. *Life Sci.* **348**, 122634 (2024).
9. Yuan, Q. et al. Machine learning-based identification of kbhb-affected tumor cell subsets as prognostic and therapeutic targets in breast cancer. *J. Transl Med.* **24** (1), 71 (2025).
10. Sakkal, M. & Hajal, A. A. Machine learning predictions of tumor progression: How reliable are we? *Comput. Biol. Med.* **191**, 110156 (2025).
11. Burd, E. M. Human papillomavirus and cervical cancer. *Clin. Microbiol. Rev.* **16** (1), 1–17 (2003).
12. Zitvogel, L. et al. Mechanism of action of conventional and targeted anticancer therapies: reinstating immunosurveillance. *Immunity* **39** (1), 74–88 (2013).
13. Lei, Y. et al. Applications of single-cell sequencing in cancer research: progress and perspectives. *J. Hematol. Oncol.* **14** (1), 91 (2021).
14. Cao, X. et al. Autophagy inhibition enhances apigenin-induced apoptosis in human breast cancer cells. *Chin. J. Cancer Res.* **25** (2), 212–222 (2013).
15. Wu, Y. T. et al. Dual role of 3-methyladenine in modulation of autophagy via different temporal patterns of inhibition on class I and III phosphoinositide 3-kinase. *J. Biol. Chem.* **285** (14), 10850–10861 (2010).
16. Caruso, G. et al. Cervical cancer: a new era. *Int. J. Gynecol. Cancer*. **34** (12), 1946–1970 (2024).
17. Podwika, S. E. & Duska, L. R. Top advances of the year: Cervical cancer. *Cancer* **129** (5), 657–663 (2023).
18. Győrffy, B. Integrated analysis of public datasets for the discovery and validation of survival-associated genes in solid tumors. *Innov. (Camb)*. **5** (3), 100625 (2024).
19. Wang, J. et al. Influence of cyclin D1 splicing variants expression on breast cancer chemoresistance via CDK4/CyclinD1-pRB-E2F1 pathway. *J. Cell. Mol. Med.* **27** (7), 991–1005 (2023).
20. Jiang, D. et al. Prognostic significance of CCND1 amplification/overexpression in smoking patients with esophageal squamous cell carcinoma. *Cancer Genet.* **278-279**, 1–8 (2023).
21. Scarth, J. A. et al. The human papillomavirus oncoproteins: a review of the host pathways targeted on the road to transformation. *J. Gen. Virol.* **102** (3). (2021).
22. Lu, F. et al. An alternatively spliced cyclin D1 isoform, cyclin D1b, is a nuclear oncogene. *Cancer Res.* **63** (21), 7056–7061 (2003).
23. Wang, J. et al. Aberrant Cyclin D1 splicing in cancer: from molecular mechanism to therapeutic modulation. *Cell. Death Dis.* **14** (4), 244 (2023).
24. Lin, D. I. et al. Phosphorylation-dependent ubiquitination of cyclin D1 by the SCF(FBX4-alphaB crystallin) complex. *Mol. Cell.* **24** (3), 355–366 (2006).
25. Chen, C. et al. Role of TRIP13 in human cancer development. *Mol. Biol. Rep.* **51** (1), 1088 (2024).
26. Bayley, R., Ward, C. & Garcia, P. MYBL2 amplification in breast cancer: Molecular mechanisms and therapeutic potential. *Biochim. Biophys. Acta Rev. Cancer*. **1874** (2), 188407 (2020).
27. Jacob Bunu, S. et al. TRIP13 - a potential drug target in cancer pharmacotherapy. *Bioorg. Chem.* **151**, 107650 (2024).
28. Huang, Y. et al. MYBL2 promotes proliferation of clear cell renal cell carcinoma by regulating TOP2A and activating AKT/mTOR signaling pathway. *Faseb J.* **39** (2), e70330 (2025).
29. Hamby, M. E. et al. Inflammatory mediators alter the astrocyte transcriptome and calcium signaling elicited by multiple G-protein-coupled receptors. *J. Neurosci.* **32** (42), 14489–14510 (2012).
30. Rui, R., Zhou, L. & He, S. Cancer immunotherapies: advances and bottlenecks. *Front. Immunol.* **14**, 1212476 (2023).
31. Fu, J. et al. The role of squamous cell carcinoma antigen (SCC Ag) in outcome prediction after concurrent chemoradiotherapy and treatment decisions for patients with cervical cancer. *Radiat. Oncol.* **14** (1), 146 (2019).
32. Zhu, Y. et al. Performance of p16/Ki67 immunostaining, HPV E6/E7 mRNA testing, and HPV DNA assay to detect high-grade cervical dysplasia in women with ASCUS. *BMC Cancer*. **19** (1), 271 (2019).
33. Wentzensen, N. et al. Clinical Evaluation of Human Papillomavirus Screening With p16/Ki-67 Dual Stain Triage in a Large Organized Cervical Cancer Screening Program. *JAMA Intern. Med.* **179** (7), 881–888 (2019).

34. Ouh, Y. T. et al. Enhancing Cervical Cancer Screening: Review of p16/Ki-67 Dual Staining as a Promising Triage Strategy. *Diagnostics (Basel)*. **14** (4), 451 (2024).
35. Yu, L., Wei, J. & Liu, P. Attacking the PI3K/AKT signaling pathway for targeted therapeutic treatment in human cancer. *Semin Cancer Biol.* **85**, 69–94 (2022).

Acknowledgements
We sincerely thank all contributors who made this study possible. We are grateful to the researchers and institutions for generating and sharing the GEO and TCGA datasets used in this work. We also thank the staff of Guizhou Medical University Affiliated Hospital for their assistance in sample collection and clinical coordination. In addition, we appreciate the technical support and valuable suggestions provided by our laboratory colleagues. Their contributions greatly facilitated the completion of this study.

Author contributions
H.Z., L.X., and Y.L. contributed equally to this work. Y.T., X.L., and J.R. conceived and designed the study. H.Z., L.X., Y.L., H.Y., S.M., and Y.Z. performed the experiments, collected the data, and conducted the formal analysis. H.Z., L.X., and Y.L. drafted the manuscript. J.R., X.L., and Y.T. supervised the study and critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Funding
This study was supported by the National Natural Science Foundation of China (Grant No. NSFC82360588), the Science and Technology Fund Project of Guizhou Health Commission (Grant No. gzwkj2024-184), the Key Laboratory for Chronic Disease Biomarkers of Guizhou Medical University (Grant No. 2024fy004), the Science and Technology Foundation of Health Commission of Guizhou Province (Grant No. gzwkj2026-073), the General Program of Guizhou Provincial Department of Science and Technology [Grant No. Qiankehe Basic Research MS(2026)843], and the Guizhou Provincial Postgraduate Research Fund Project (Grant No. 2024YJSKYJJ308).

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval
This study was performed in accordance with the principles of the Declaration of Helsinki and was approved by the Ethics Committee of the Affiliated Hospital of Guizhou Medical University (Approval No. 2023-231-01).

Consent to participate
Written informed consent was obtained from all individual participants included in this study.

Consent to publish
The authors affirm that human research participants provided informed consent for publication of the data included in this study.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-57689-7.

Correspondence and requests for materials should be addressed to J.R., X.L. or Y.T.