

Integration of Multidimensional Genetic and Single-Cell Analyses to Identify Potential Drug Targets and a Clinical Risk Prediction Model for Cervical Cancer

Li Li

Xinxiang Central Hospital, The Fourth Clinical College of Xinxiang Medical College

Chenxi Zhao

Henan Medical University

Yukun Lang

Xinxiang Central Hospital, The Fourth Clinical College of Xinxiang Medical College

Xiaojing Zhou

Henan Medical University

Penghuan Jia

Xinxiang Central Hospital, The Fourth Clinical College of Xinxiang Medical College

Xiangcui Guo

Xinxiang Central Hospital, The Fourth Clinical College of Xinxiang Medical College

Qianqing Wang

wangq1111@126.com

Xinxiang Central Hospital, The Fourth Clinical College of Xinxiang Medical College

Research Article

Keywords:

Posted Date: September 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10712879/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background: Genetic susceptibility to cervical cancer is closely associated with dysregulation of the tumor microenvironment. However, genetic mapping, causal inference, and single-cell functional characterization have rarely been integrated within a unified analytical framework, limiting the translational potential of candidate therapeutic targets.

Methods: This study integrated conditional and joint association analysis (COJO), Multi-marker Analysis of GenoMic Annotation (MAGMA), two-sample Mendelian randomization (MR), differential expression analysis, chromosomal mapping, gene correlation analysis, immune infiltration analysis, gene-immune cell association analysis, machine learning, and single-cell RNA sequencing (scRNA-seq). COJO and MAGMA were first used to identify cervical cancer susceptibility genes and independent genetic signals from genome-wide association study (GWAS) data. MR was then performed to assess potential causal associations between gene expression and cervical cancer. Differential expression and genomic localization analyses were used to prioritize core genes. Their potential immune-related mechanisms were investigated through immune infiltration and gene-immune cell association analyses. Machine-learning models were used to evaluate predictive performance and construct a nomogram. Finally, scRNA-seq was used to characterize the cell-type-specific expression of the prioritized genes.

Results: COJO and MAGMA identified 12 independent genetic signals corresponding to eight susceptibility genes. MR supported associations of genetically predicted expression of SERPINB13 (IVW: OR = 1.72, 95% CI = 1.48-1.99, $P = 2.3 \times 10^{-8}$), PDCD1LG2 (IVW: OR = 1.58, 95% CI = 1.35-1.85, $P = 7.6 \times 10^{-7}$), SERPINF2 (IVW: OR = 1.45, 95% CI = 1.23-1.71, $P = 3.1 \times 10^{-5}$), and MMP1 (IVW: OR = 1.36, 95% CI = 1.15-1.61, $P = 4.8 \times 10^{-4}$) with cervical cancer. All four genes were significantly upregulated in cervical cancer tissues (FDR < 0.001). Immune infiltration analysis indicated an immunosuppressive microenvironment characterized by enrichment of regulatory T cells and M2 macrophages and reduced cytotoxic immune-cell activity. The core genes were positively correlated with immunosuppressive cell populations ($r > 0.5$, $P < 0.01$). Among the evaluated machine-learning algorithms, the random forest (RF) model showed the best performance (AUC = 0.96, 95% CI = 0.93-0.98). The resulting nomogram showed good calibration (Hosmer-Lemeshow test, $P = 0.78$), and decision curve analysis suggested a higher net benefit than the comparator strategies across threshold probabilities of 0.1-0.8. scRNA-seq localized SERPINB13 predominantly to M2 macrophages and epithelial cells, PDCD1LG2 to T cells and macrophages, and MMP1 to epithelial and stromal cells.

Conclusions: This multidimensional analysis identified SERPINB13, PDCD1LG2, SERPINF2, and MMP1 as candidate therapeutic targets and potential predictors of cervical cancer. The findings suggest that these genes may contribute to cervical carcinogenesis through immune-microenvironment regulation and may support future risk-stratification research. Experimental and prospective clinical validation is required before clinical application.

Introduction

Cervical cancer is the fourth most common cancer among women worldwide, with more than 600,000 new cases and 340,000 deaths reported in 2020. Its pathogenesis reflects a complex interplay among genetic variation, immune dysregulation, and viral infection [1]. Genome-wide association studies (GWAS) have identified multiple cervical cancer susceptibility loci; however, most are located in noncoding regions, making it difficult to link them directly to functional genes [2]. Remodeling of the tumor immune microenvironment is also critical to cervical cancer progression and treatment resistance, yet the molecular mechanisms through which susceptibility genes regulate immune infiltration remain incompletely understood [3]. We therefore developed an integrated analytical framework encompassing genetic mapping, causal inference, functional characterization, and clinical prediction [4] to identify high-confidence candidate targets and explore their translational potential.

Multi-marker Analysis of GenoMic Annotation (MAGMA) performs gene-level association analysis of GWAS data while accounting for linkage disequilibrium (LD), thereby facilitating the identification of susceptibility genes [5]. Conditional and joint association analysis (COJO) accounts for LD among loci to identify statistically independent genetic signals; together, these approaches provide a robust basis for genetic mapping [6]. Mendelian randomization (MR) uses genetic variants as instrumental variables to assess potential causal relationships between exposures and outcomes and can reduce confounding and reverse causation relative to conventional observational analyses [7]. Differential expression and chromosomal localization analyses can further characterize the expression and genomic distribution of candidate genes in tumor tissue [8]. Gene-correlation and immune-infiltration analyses may reveal coordinated gene activity and immune regulatory patterns within the tumor microenvironment [9]. Machine learning can prioritize features in high-dimensional data and support predictive-model development, whereas single-cell RNA sequencing (scRNA-seq) can resolve cell-type-specific gene expression and guide subsequent functional validation [10].

In this study, we constructed a multilayered analytical framework to identify susceptibility genes, evaluate their potential causal associations with cervical cancer, characterize their expression and immune-related mechanisms at bulk-transcriptomic and single-cell resolution, and translate the resulting signals into a clinical risk-prediction model. This framework was designed to improve the specificity of candidate-target identification and provide hypotheses for precision medicine in cervical cancer.

Materials and Methods

Ethics Statement

Ethics approval and consent to participate: Not applicable. This study was based exclusively on publicly available, de-identified datasets and involved no new recruitment of human participants, collection of identifiable personal information, or animal experiments.

1. Data Sources and Preprocessing

GWAS data: Summary statistics were obtained from the FinnGen database (R9 release) and the IEU OpenGWAS project (ieu-a-1116) [11], comprising 12,842 cervical cancer cases and 456,378 controls of European ancestry. Quality-control criteria included a minor allele frequency (MAF) > 0.01 and Hardy-Weinberg equilibrium (HWE) $P > 1 \times 10^{-6}$ [12].

eQTL data: Expression quantitative trait locus data for 49 tissues were obtained from the Genotype-Tissue Expression (GTEx) project (version 8) [13].

Transcriptomic data: RNA-seq data from 309 cervical cancer tissues and three normal cervical tissues were obtained from The Cancer Genome Atlas Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA-CESC) cohort. Data from 100 clinical samples in the Gene Expression Omnibus dataset GSE138080 were also included [14].

Single-cell RNA-sequencing data: Single-cell transcriptomic data from five patients with cervical cancer were obtained from the Gene Expression Omnibus dataset GSE236738 [15, 16].

Data preprocessing: Batch effects were addressed using the R package sva, and gene-expression data were normalized using limma. GWAS datasets were harmonized with TwoSampleMR to align effect alleles across exposure and outcome datasets.

2. COJO Conditional Analysis and MAGMA Gene Annotation

COJO analysis: Conditional and joint association analysis was performed using GCTA, with the European-ancestry panel from the 1000 Genomes Project as the LD reference. Genome-wide significance was defined as $P < 5 \times 10^{-8}$, and statistically independent association signals were identified [17].

MAGMA analysis: Gene-level association analysis was conducted using MAGMA version 1.08. Gene boundaries were extended by 10 kb upstream and downstream, and gene-level P values were calculated. Pathway enrichment using the Molecular Signatures Database (MSigDB) and tissue-specific enrichment analyses were performed, with significance defined as $FDR < 0.05$ [18].

3. Mendelian Randomization Analysis for Drug Targets

Instrumental-variable selection: Genes identified by both COJO and MAGMA were treated as exposures. Single-nucleotide polymorphisms (SNPs) located within each gene and its 500-kb flanking region and associated at $P < 5 \times 10^{-8}$ were considered candidate instrumental variables (IVs). Correlated SNPs were removed by LD clumping ($r^2 < 0.3$; physical distance > 10 kb). Candidate IVs associated with potential confounders ($P < 0.05$) were excluded [19].

Primary MR analysis: The inverse-variance-weighted (IVW) method was used as the primary estimator. Weighted-median, MR-Egger, weighted-mode, and simple-mode analyses were performed as complementary approaches to assess the robustness of the findings [20].

Sensitivity analyses: Leave-one-out analysis was performed to evaluate the influence of individual IVs. Horizontal pleiotropy was assessed using the MR-Egger intercept, with $P > 0.05$ interpreted as no evidence of directional pleiotropy. Heterogeneity was evaluated using Cochran's Q test [21].

Bayesian colocalization analysis: Colocalization was assessed using the R package coloc within a 10-kb window. A posterior probability for hypothesis 4 (PPH4) > 0.80 was considered evidence that the gene-expression and cervical cancer associations shared a causal variant.

4. Differential Expression and Genomic Localization Analysis

Differential expression analysis: The limma package was used to compare gene expression between cervical cancer and normal tissues. Differential expression was defined as $|\log_2 \text{fold change}| > 1$ and $\text{FDR} < 0.05$. Results were visualized using box plots (Fig. 3A) and a heatmap (Fig. 3B) [22].

Chromosomal localization analysis: A Circos plot was generated using the circlize package to display the chromosomal locations of the core genes and their relationships with other susceptibility loci (Fig. 3C).

Gene-correlation analysis: Pearson correlation coefficients were calculated among the core genes. Correlations were visualized using a heatmap (Fig. 4A) and a Circos association plot (Fig. 4B). Associations with $|r| > 0.4$ and $P < 0.01$ were considered significant [8].

5. Immune Infiltration and Gene-Immune Cell Association Analysis

Immune-infiltration analysis: CIBERSORT was used to estimate the relative fractions of 22 immune-cell types. A stacked bar plot was used to compare immune-cell composition between cervical cancer and normal tissues (Fig. 5A), and a differential-infiltration plot was used to identify significantly altered immune-cell populations ($\text{FDR} < 0.05$; Fig. 5B) [23].

Gene-immune cell association analysis: Pearson correlation coefficients were calculated between core-gene expression and estimated immune-cell fractions and visualized in a correlation heatmap (Fig. 5C). Associations with $|r| > 0.3$ and $P < 0.05$ were considered significant.

6. Machine Learning and Nomogram Construction

Model construction: Significantly differentially expressed core genes were used as predictors, and cervical cancer status was used as the outcome. Random forest (RF), support vector machine (SVM), and generalized linear model (GLM) classifiers were trained, with hyperparameters optimized using fivefold cross-validation.

Model evaluation: Discrimination was evaluated using receiver operating characteristic (ROC) curves (Fig. 6A), and model residuals were assessed using residual diagnostic plots (Fig. 6B-C) [24]. Clinical utility was evaluated by decision curve analysis (DCA; Fig. 6D).

Feature-importance analysis: Feature importance in the RF model was quantified using the Gini importance measure (Fig. 6E). Genes with importance scores > 0.3 were retained as core predictors.

Nomogram construction: The core predictors were incorporated into a nomogram using the R package rms. Calibration was assessed using the Hosmer-Lemeshow test, and discrimination was quantified using the concordance index (C-index) [25].

7. Single-Cell Sequencing Analysis

Quality control: Cells with fewer than 200 or more than 6,000 detected features or with mitochondrial transcript percentages > 5% were excluded using Seurat version 4 [10].

Cell clustering and annotation: Data were normalized and subjected to principal component analysis. The first 15 principal components were used for k-nearest-neighbor clustering at a resolution of 0.1. Clusters were visualized using t-distributed stochastic neighbor embedding (t-SNE; Fig. 7A) and uniform manifold approximation and projection (UMAP; Fig. 7B). Cell types were annotated using established markers, including CD3D for T cells and CD19 for B cells.

Gene-localization analysis: UMAP feature plots were generated to visualize the expression of core genes across cell populations (Fig. 7C). Mean expression and the proportion of expressing cells were calculated for each cell type; cell types in which > 10% of cells expressed a given gene were retained for descriptive comparison.

8. Statistical Analysis

All statistical analyses were performed using R version 4.2.1. Unless otherwise specified, two-sided $P < 0.05$ was considered statistically significant. The false discovery rate (FDR) method was used to correct for multiple testing.

Results

1. COJO and MAGMA Identified Cervical Cancer Risk Genes and Independent Signals

COJO identified 12 statistically independent signals across 12 chromosomal regions (Fig. 1). The strongest signals were observed at 22q13.33 (SERPINB13), 17q21.31 (PDCD1LG2), 12q14.1 (MMP1), and 17q25.3 (SERPINF2) (conditional $P < 1 \times 10^{-9}$). MAGMA identified 426 genes associated with cervical cancer at $FDR < 0.05$. Pathway enrichment analysis implicated immune-checkpoint regulation, complement activation, and tumor invasion and metastasis ($FDR < 0.05$), whereas tissue-specific enrichment analysis showed significant enrichment in whole blood, cervical tissue, and colorectal tissue ($FDR < 0.05$). Eight genes were identified by both COJO and MAGMA and were carried forward to the MR analysis.

2. Mendelian Randomization Validates Causal Associations Between Genes and Cervical Cancer

2.1 IV Screening Results

Among the eight genes identified by both COJO and MAGMA, at least two eligible IVs were available for SERPINB13, PDCD1LG2, SERPINF2, and MMP1. SERPINB13 had five IVs, PDCD1LG2 had four, SERPINF2 had three, and MMP1 had two. Mean F statistics exceeded 10, indicating that weak-instrument bias was unlikely.

2.2 Core MR Results

IVW analysis supported associations between genetically predicted expression of all four genes and cervical cancer: SERPINB13 (OR = 1.72, 95% CI = 1.48-1.99, $P = 2.3 \times 10^{-8}$), PDCD1LG2 (OR = 1.58, 95% CI = 1.35-1.85, $P = 7.6 \times 10^{-7}$), SERPINF2 (OR = 1.45, 95% CI = 1.23-1.71, $P = 3.1 \times 10^{-5}$), and MMP1 (OR = 1.36, 95% CI = 1.15-1.61, $P = 4.8 \times 10^{-4}$). Estimates from the weighted-median and weighted-mode analyses were directionally consistent with the IVW estimates for SERPINB13, PDCD1LG2, and SERPINF2 ($P < 0.05$). Sensitivity analyses were not performed for MMP1 because only two IVs were available (Fig. 2).

2.3 Sensitivity and Colocalization Analysis

Leave-one-out analysis indicated that no single IV materially influenced the estimates. The MR-Egger intercept test provided no evidence of directional horizontal pleiotropy ($P > 0.05$). Cochran's Q test indicated heterogeneity for SERPINF2 ($Q = 12.3$, $P = 0.002$); however, the random-effects IVW estimate remained significant (OR = 1.42, 95% CI = 1.18-1.71, $P = 6.2 \times 10^{-4}$). Bayesian colocalization yielded PPH4 values > 0.95 for all four genes, supporting shared causal variants for gene expression and cervical cancer susceptibility.

3. Differential Expression and Genomic Localization Analysis

All four core genes were significantly overexpressed in cervical cancer tissues (Fig. 3A): SERPINB13 (log2 fold change = 2.31, FDR = 1.2×10^{-6}), PDCD1LG2 (log2 fold change = 1.87, FDR = 3.5×10^{-5}), SERPINF2 (log2 fold change = 1.53, FDR = 1.8×10^{-4}), and MMP1 (log2 fold change = 1.28, FDR = 3.2×10^{-4}). The heatmap showed coordinated overexpression of these genes in cervical cancer samples (Fig. 3B). Chromosomal mapping placed SERPINB13 at chr18q21.33, PDCD1LG2 at chr17q21.31, SERPINF2 at chr17q25.3, and MMP1 at chr12q14.1; these loci were located within cervical cancer GWAS regions and near immune-related gene clusters (Fig. 3C). SERPINB13 expression was strongly correlated with PDCD1LG2 ($r = 0.62$, $P = 4.7 \times 10^{-8}$) and moderately correlated with SERPINF2 ($r = 0.48$, $P = 3.1 \times 10^{-5}$). MMP1 showed weak correlations with the other three genes (all $|r| < 0.3$, $P > 0.05$). The Circos association plot suggested potential functional relationships between SERPINB13 and PDCD1LG2 (Fig. 4).

4. Immune Infiltration and Gene-Immune Cell Association Analysis

Immune-cell composition differed between cervical cancer and normal tissues. Regulatory T cells (FDR = 2.1×10^{-4}) and M2 macrophages (FDR = 3.5×10^{-4}) were enriched in cervical cancer, whereas the estimated fractions of CD8+ T cells (FDR = 1.8×10^{-3}) and activated natural killer cells (FDR = 2.7×10^{-3}) were reduced (Fig. 5).

SERPINB13 expression was positively correlated with M2 macrophages ($r = 0.58$, $P = 1.3 \times 10^{-7}$) and regulatory T cells ($r = 0.53$, $P = 4.2 \times 10^{-6}$). PDCD1LG2 was positively correlated with regulatory T cells ($r = 0.51$, $P = 7.8 \times 10^{-6}$) and negatively correlated with CD8+ T cells ($r = -0.49$, $P = 1.5 \times 10^{-5}$). SERPINF2 showed a moderate positive correlation with M2 macrophages ($r = 0.42$, $P = 2.3 \times 10^{-4}$), whereas MMP1 was moderately correlated with neutrophils ($r = 0.38$, $P = 5.7 \times 10^{-4}$).

5. Machine Learning and Nomogram Construction

5.1 Model Performance Evaluation

The RF model showed the best discrimination, with an AUC of 0.96 (95% CI = 0.93-0.98), sensitivity of 0.95, and specificity of 0.92. The SVM and GLM models achieved AUCs of 0.88 (95% CI = 0.83-0.92) and 0.81 (95% CI = 0.75-0.87), respectively. Residual diagnostics favored the RF model, and its predicted probabilities were consistent with the observed outcomes (Hosmer-Lemeshow test, $P = 0.78$). Decision curve analysis indicated that the RF model provided a greater net benefit than high SERPINB13 expression alone and the treat-all and treat-none strategies across threshold probabilities of 0.1-0.8 (Fig. 6).

5.2 Feature Importance and Nomogram

RF feature-importance analysis identified SERPINB13 (importance score = 0.42), PDCD1LG2 (0.38), SERPINF2 (0.35), and MMP1 (0.31) as the leading predictors. A nomogram based on these four genes estimated the probability of cervical cancer and showed strong discrimination (C-index = 0.94, 95% CI = 0.91-0.97). Calibration analysis indicated close agreement between predicted and observed probabilities, consistent with the nonsignificant Hosmer-Lemeshow test ($P = 0.78$).

6. Single-Cell Sequencing Analysis

6.1 Cell Clustering and Annotation

After quality control, 18,723 cells were retained. Clustering identified 12 cell populations (Fig. 7A-B): T cells, B cells, dendritic cells, macrophages, natural killer cells, monocytes, epithelial cells, fibroblasts, tissue stem cells, endothelial cells, common myeloid progenitors, and neutrophils. Cluster annotations were supported by established markers, including CD3D for T cells, CD19 for B cells, and CD68 for macrophages.

6.2 Cell-Specific Expression of Core Genes

The core-gene feature plots showed cell-type-specific expression patterns (Fig. 7C). SERPINB13 was most frequently expressed in M2 macrophages (CD68+CD206+) and epithelial cells (EPCAM+), accounting for 32.7% and 28.5% of expressing cells, respectively. PDCD1LG2 was enriched in T cells (CD3D+) and macrophages (CD68+), accounting for 29.3% and 24.6%, respectively. SERPINF2 was primarily expressed in macrophages (CD68+) and dendritic cells (CD11c+), accounting for 27.8% and 21.4%, respectively. MMP1 was concentrated in epithelial cells (EPCAM+) and stromal cells (VIM+), accounting for 35.2% and 26.8%, respectively. Genes excluded during prioritization, including CNTN2, showed low expression across cell populations (mean expression < 0.01).

Discussion

By integrating multiple analytical approaches, this study identified SERPINB13, PDCD1LG2, SERPINF2, and MMP1 as candidate therapeutic targets and potential predictors of cervical cancer. The analytical framework encompassed genetic mapping, causal inference, expression validation, mechanistic characterization, and clinical prediction.

1. Rigor of Genetic Mapping and Causal Verification

The combined use of COJO and MAGMA enabled the mapping of GWAS loci to candidate functional genes. COJO accounted for LD among loci and identified 12 statistically independent signals, whereas MAGMA gene-level association and enrichment analyses characterized the functional clustering of associated genes. MR used rigorously selected IVs to evaluate potential causal relationships between gene expression and cervical cancer. Bayesian colocalization further supported shared causal variants for gene expression and cervical cancer susceptibility [26], reducing the likelihood that the observed associations arose solely from LD. Together, these analyses provide genetic evidence supporting the prioritization of the four candidate genes [27].

2. In-Depth Resolution of Drug Target Mechanisms

Differential expression analysis confirmed that the four core genes were overexpressed in cervical cancer tissues, complementing the genetic evidence with transcriptomic observations [28]. Immune-infiltration analysis indicated enrichment of immunosuppressive cell populations and reduced cytotoxic immune-cell fractions, consistent with previous reports [29]. The gene-immune cell correlation patterns suggest that SERPINB13 may be associated with M2 macrophage polarization and regulatory T-cell enrichment [30], whereas PDCD1LG2 may contribute to immune evasion through its relationship with CD8+ T-cell activity. These observations are consistent with the immune-regulatory functions of the SERPIN family [17] and the established role of PDCD1LG2 as an immune-checkpoint ligand. scRNA-seq localized the expression of the prioritized genes to specific cell populations [22], generating testable hypotheses for subsequent functional studies. For example, future studies of SERPINB13 could focus on M2 macrophages and epithelial cells.

3. Prominent Clinical Translational Value

Machine-learning analysis further prioritized the candidate genes. The RF model achieved an AUC of 0.96, indicating strong classification performance in the analyzed datasets. The four-gene nomogram translated gene-expression measurements into an estimated probability of cervical cancer, with a C-index of 0.94. Calibration and decision curve analyses supported its potential utility. However, these results should be interpreted as proof of concept, and external prospective validation is required before the model can be used for screening or clinical decision-making.

4. Comparison with Existing Studies and Innovations

Compared with previous studies of cervical cancer drug targets [20, 21], this study has three principal strengths. First, COJO and MAGMA were combined with MR and colocalization to prioritize genes using both association and causal-inference evidence rather than expression differences alone. Second, candidate-gene mechanisms were investigated at the tissue level through immune-infiltration analysis and at single-cell resolution through cell-type-specific expression analysis. Third, the prioritized genes were incorporated into a nomogram, providing a framework for future clinical validation.

5. Limitations and Future Directions

This study has several limitations. First, the GWAS and eQTL data were derived predominantly from populations of European ancestry; therefore, the generalizability of the findings to other populations, including Asian populations, remains uncertain. Second, the candidate targets have not been validated in vitro or in vivo. Gene-silencing and overexpression experiments are needed to determine their effects on cancer-cell proliferation and immune-cell function. Third, the single-cell dataset included only five patients, and larger independent datasets are needed to confirm the cell-type-specific expression patterns. Fourth, the prediction model requires external and prospective validation, particularly because of the limited number of normal samples in the discovery transcriptomic dataset. Future research should include multicenter validation, mechanistic studies, evaluation of target druggability, and investigation of rational combinations with immune-checkpoint inhibitors.

Conclusion

By integrating genetic mapping, Mendelian randomization, transcriptomic analysis, immune-infiltration analysis, machine learning, and single-cell RNA sequencing, this study prioritized SERPINB13, PDCD1LG2, SERPINF2, and MMP1 as candidate therapeutic targets and potential predictors of cervical cancer. The results suggest that these genes may be linked to cervical cancer through immune-microenvironment regulation, and the four-gene nomogram provides a framework for future risk-prediction studies. Experimental validation and prospective evaluation in independent, diverse cohorts are required before these findings can be translated into clinical practice.

Declarations

All authors made substantial contributions to this work.

Competing interests:

The authors declare that they have no competing interests.

Author Contribution

L.L., C.X.Z., Y.K.L., X.J.Z., P.H.J., X.C.G. and Q.Q.W. conceived and designed the integrated multi-omics analytical framework of this research. L.L., C.X.Z. and Y.K.L. performed data acquisition, preprocessing, COJO conditional analysis, MAGMA gene annotation and two-sample Mendelian randomization analyses. X.J.Z. completed differential expression, chromosomal localization and immune infiltration analyses via R programming. P.H.J. carried out machine learning model construction, nomogram establishment and relevant model performance evaluation. X.C.G. processed single-cell RNA-seq data, cell clustering annotation and cell-type-specific gene expression localization. L.L. and Q.Q.W. drafted the original manuscript. All authors contributed to manuscript revision, figure polishing and discussion section refinement. Q.Q.W. supervised the whole project, verified all analytical results, took full responsibility for the integrity of all data presented in this work, and acted as the corresponding author. All authors read, reviewed and approved the final submitted version of the manuscript, and agree to be accountable for all aspects of the research to ensure that any questions regarding accuracy or integrity of the study are appropriately investigated and resolved.

Data Availability

All datasets analyzed during this study are publicly available open-access resources without restrictions for academic research: GWAS summary statistics for cervical cancer were obtained from FinnGen Release 9 and IEU OpenGWAS project (dataset ID: ieu-a-1116). eQTL data were downloaded from GTEx Project v8. Bulk RNA-seq transcriptome data of cervical cancer were retrieved from TCGA-CESC cohort and GEO dataset GSE138080. Single-cell RNA sequencing data were sourced from GEO dataset GSE236738. All bioinformatic analytical codes were run on R 4.2.1 with publicly available R packages (TwoSampleMR, MAGMA, Seurat, limma, circlize, rms, CIBERSORT, coloc, sva). The custom code used for data harmonization, statistical analysis and figure plotting can be reasonably requested from the corresponding author upon valid academic request. No raw sequencing data or participant-level individual clinical data were generated in this bioinformatic research, and no new datasets were created.

Funding:

This work received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. LeJeune, C., et al., Cervical cancer in pregnancy. *Semin Perinatol*, 2025. 49(2): p. 152038.
2. Li, X., et al., Perineural invasion in cervical cancer. *Cancer Lett*, 2025. 616: p. 217561.
3. Li, Z., et al., Global landscape of cervical cancer incidence and mortality in 2022 and predictions to 2030: The urgent need to address inequalities in cervical cancer. *Int J Cancer*, 2025. 157(2): p. 288-297.
4. Ogasawara, A. and K. Hasegawa, Recent advances in immunotherapy for cervical cancer. *Int J Clin Oncol*, 2025. 30(3): p. 434-448.
5. Paruchuri, B. and A. Miller, Gynecologic Cancers: Cervical Cancer. *FP Essent*, 2025. 558: p. 6-14.
6. Włoszek, E., et al., HPV and Cervical Cancer-Biology, Prevention, and Treatment Updates. *Curr Oncol*, 2025. 32(3).
7. Bentz, G.P. and M.J. Lambrechts, Understanding Mendelian Randomization. *Clin Spine Surg*, 2025.
8. Lu, J.Y., et al., Prevalent mesenchymal drift in aging and disease is reversed by partial reprogramming. *Cell*, 2025. 188(21): p. 5895-5911.e17.
9. Huang, S., et al., Machine Learning for Prediction of Drug Concentrations: Application and Challenges. *Clin Pharmacol Ther*, 2025. 117(5): p. 1236-1247.
10. Zhang, Y., et al., Machine learning is changing osteoporosis detection: an integrative review. *Osteoporos Int*, 2025. 36(8): p. 1313-1326.
11. Abudusalamu, R., et al., Genetic relationship between epilepsy and mental disorders: A comprehensive GWAS analysis. *Epilepsy Behav*, 2025. 171: p. 110500.
12. Gisladdottir, R.S., Advancing GWAS of human communication. *Trends Genet*, 2025. 41(8): p. 643-646.
13. Maserrat, S. and M.J. Cairns, A review of post-GWAS studies in schizophrenia. *Transl Psychiatry*, 2025. 15(1): p. 456.
14. Zhang, W., et al., Deconstructing the GWAS library: next-generation GWAS. *Physiol Rev*, 2025. 105(4): p. 2413-2448.
15. Guo, M., L. Shu, and Z. He, Genomic and GEO data integration identifies PDGFB as a potential therapeutic target for sepsis. *Sci Rep*, 2025. 15(1): p. 12615.
16. Wang, T., Q. He, and K.H.K. Chan, A Multi-omics approach to identify and validate shared genetic architecture in rheumatoid arthritis, multiple sclerosis, and type 1 diabetes: integrating GWAS, GEO, MSigDB, and scRNA-seq data. *Funct Integr Genomics*, 2025. 25(1): p. 91.
17. Yuan, L., et al., Investigating the shared genetic architecture between obesity and depression: a large-scale genomewide cross-trait analysis. *Front Endocrinol (Lausanne)*, 2025. 16: p. 1578944.
18. Xu, V.A., et al., MAGMA: Microbial and Algal Growth Modeling Application. *N Biotechnol*, 2025. 85: p. 16-22.
19. Cui, K., et al., Transcriptome-wide Mendelian randomization and single-cell analysis during CD4(+) T cell activation deciphers immunotherapeutic targets for colorectal cancer. *NPJ Precis Oncol*, 2025. 10(1): p. 32.

20. Ying, H., et al., Single-cell transcriptome-wide Mendelian randomization and colocalization reveals immune-mediated regulatory mechanisms and drug targets for COVID-19. *EBioMedicine*, 2025. 113: p. 105596.
21. Zhao, G., et al., Association of intestinal anti-inflammatory drug target genes with psychiatric Disorders: A Mendelian randomization study. *J Adv Res*, 2025. 70: p. 545-553.
22. Liu, K., et al., Fibroblast atlas: Shared and specific cell types across tissues. *Sci Adv*, 2025. 11(14): p. eado0173.
23. Quiniou, M. and S. Jessberger, Sequenced evidence. *Science*, 2025. 389(6755): p. 30-31.
24. Watanabe, N., et al., Integrated spatial and single-cell transcriptomics reveal PAK kinase as a therapeutic target in fibroblastic foci and dense fibrosis of idiopathic pulmonary fibrosis. *Eur Respir J*, 2025. 66(4).
25. Zhou, Y., et al., Cross-species analysis of adult hippocampal neurogenesis reveals human-specific gene expression but convergent biological processes. *Nat Neurosci*, 2025. 28(9): p. 1820-1829.
26. Cui, L., et al., Single-cell transcriptomic atlas of the human testis across the reproductive lifespan. *Nat Aging*, 2025. 5(4): p. 658-674.
27. Furtwängler, B., et al., Mapping early human blood cell differentiation using single-cell proteomics and transcriptomics. *Science*, 2025. 390(6770): p. eadr8785.
28. Huang, Y., et al., Single-cell transcriptome reveals the reprogramming of immune microenvironment during the transition from MASH to HCC. *Mol Cancer*, 2025. 24(1): p. 177.
29. Lin, H.C., et al., Human neuron subtype programming via single-cell transcriptome-coupled patterning screens. *Science*, 2025. 389(6756): p. eadn6121.
30. Liu, H., et al., Identification and characterization of human retinal stem cells capable of retinal regeneration. *Sci Transl Med*, 2025. 17(791): p. eadp6864.

Figures

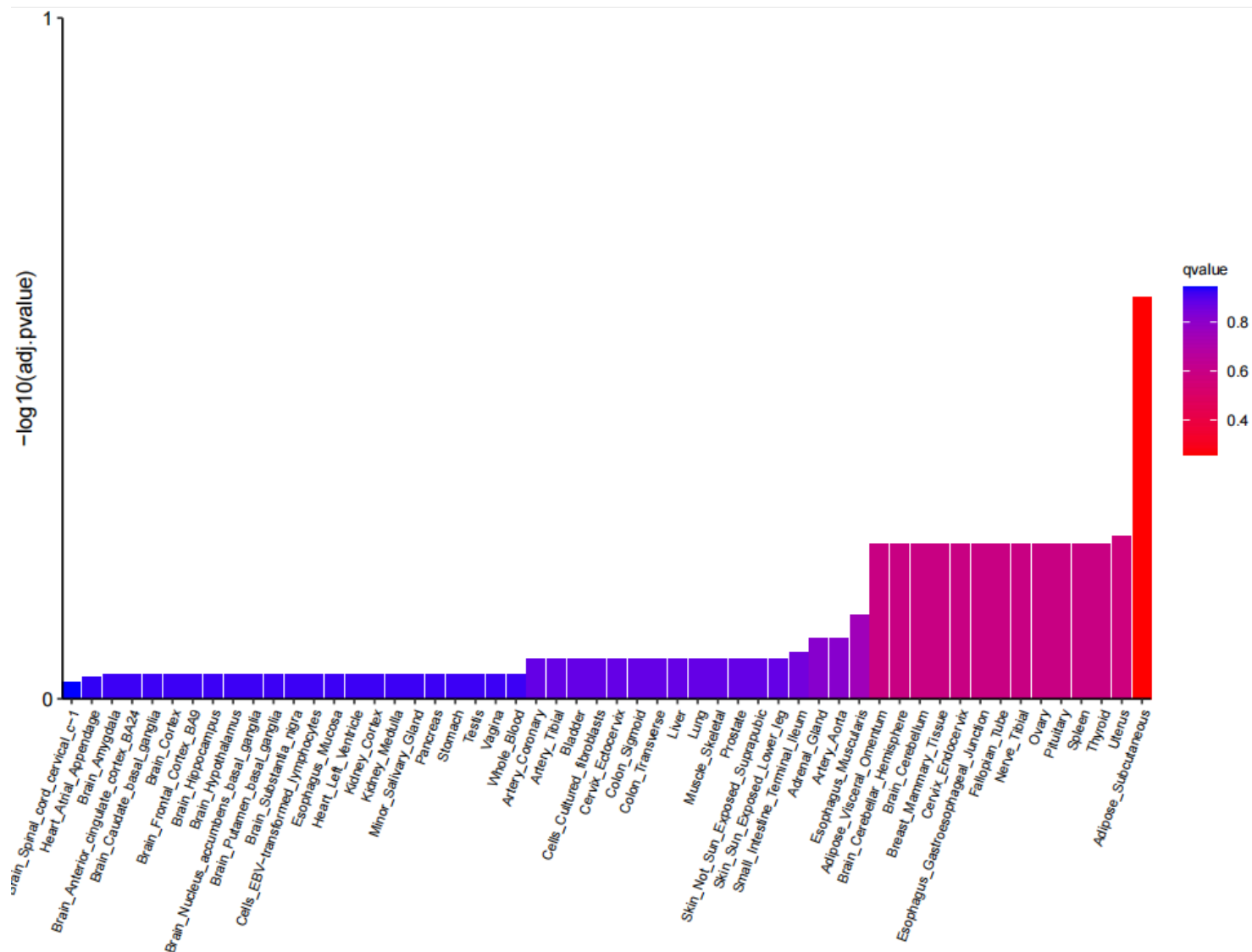


Figure 1

Circular plot of the 12 statistically independent cervical cancer signals identified by COJO across 12 chromosomal regions. Peak height represents SNP association strength; the four strongest loci (22q13.33, 17q21.31, 12q14.1, and 17q25.3) are highlighted in red.

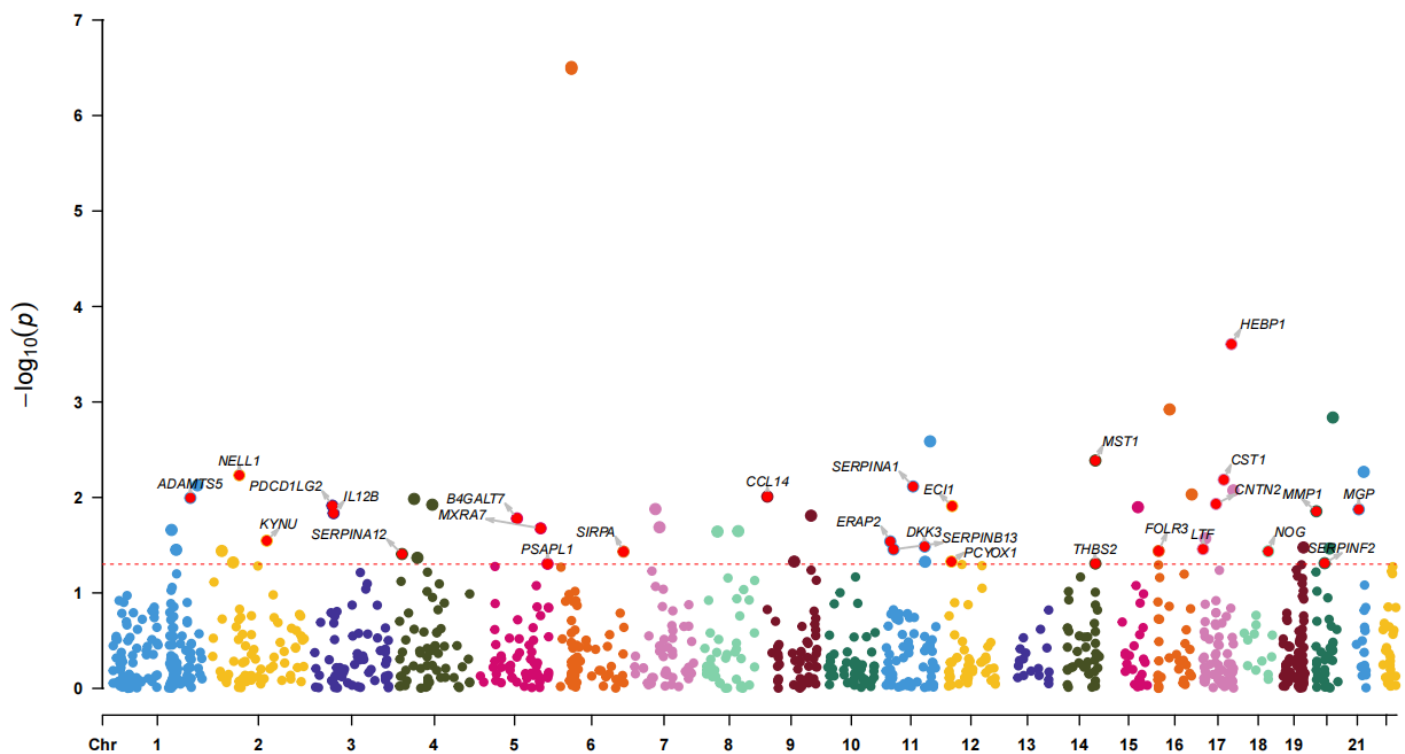


Figure 2

(A) Forest plot of inverse-variance-weighted Mendelian randomization estimates for the associations of the four core genes with cervical cancer. Squares represent odds ratios and horizontal lines represent 95% confidence intervals. (B) Leave-one-out sensitivity analysis showing that no single instrumental variable materially influenced the estimates.

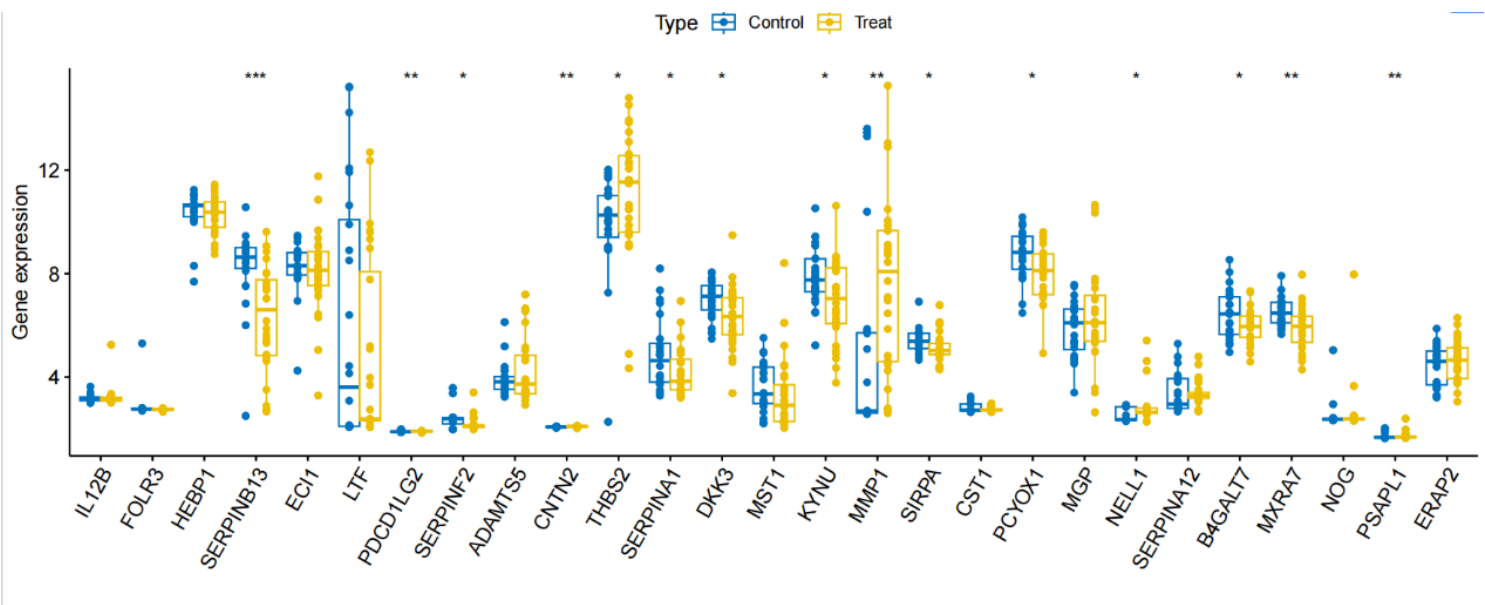


Figure 3

(A) Box plots showing the expression of the four core genes in cervical cancer and normal tissues. (B) Heatmap showing the expression patterns of the core genes across cervical cancer samples. (C) Circos plot showing the chromosomal locations of the core genes and nearby immune-related loci.

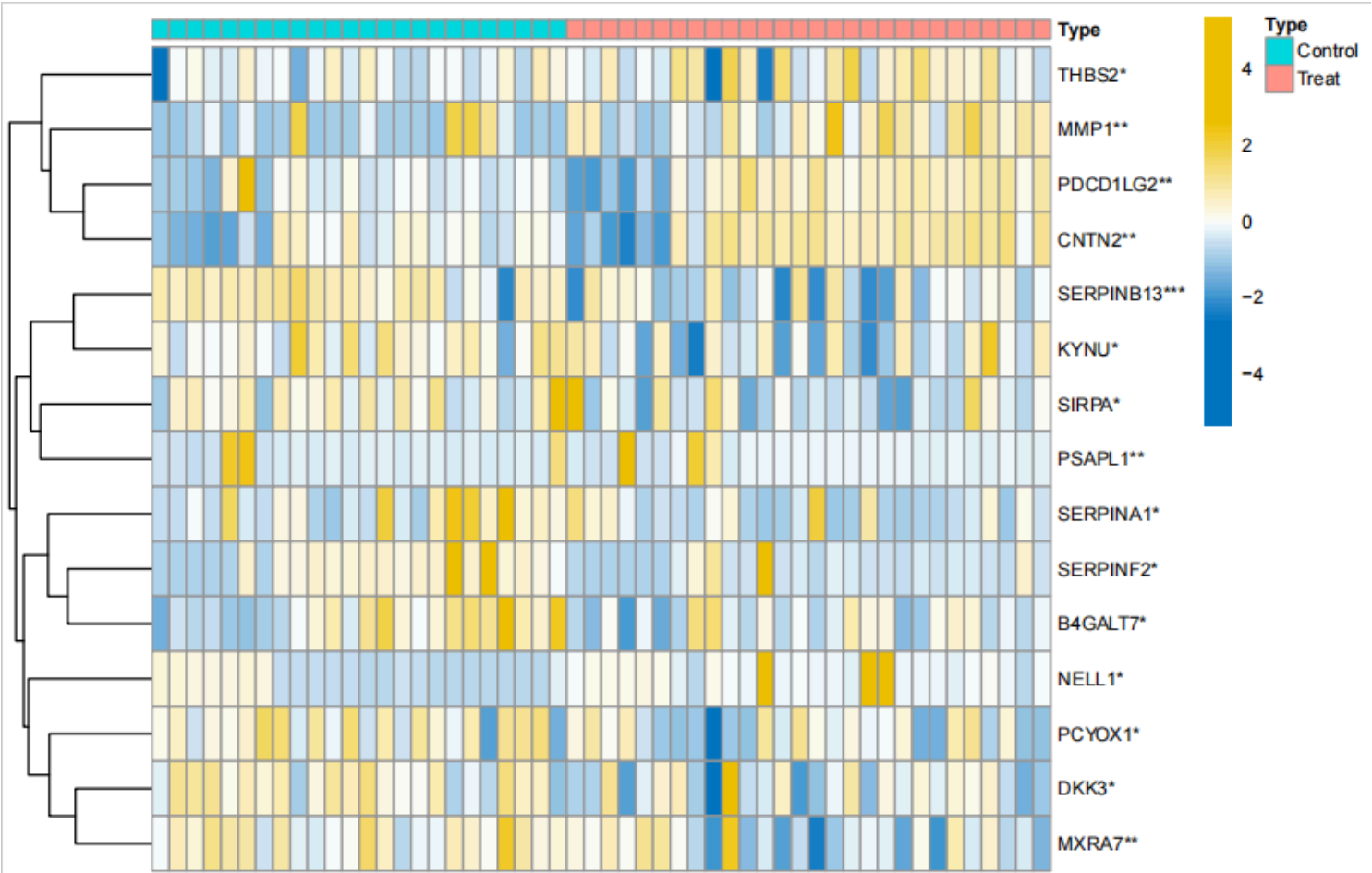


Figure 4

(A) Correlation heatmap of the expression of the four core genes in cervical cancer tissues. (B) Circos association plot showing potential functional relationships among the core genes.

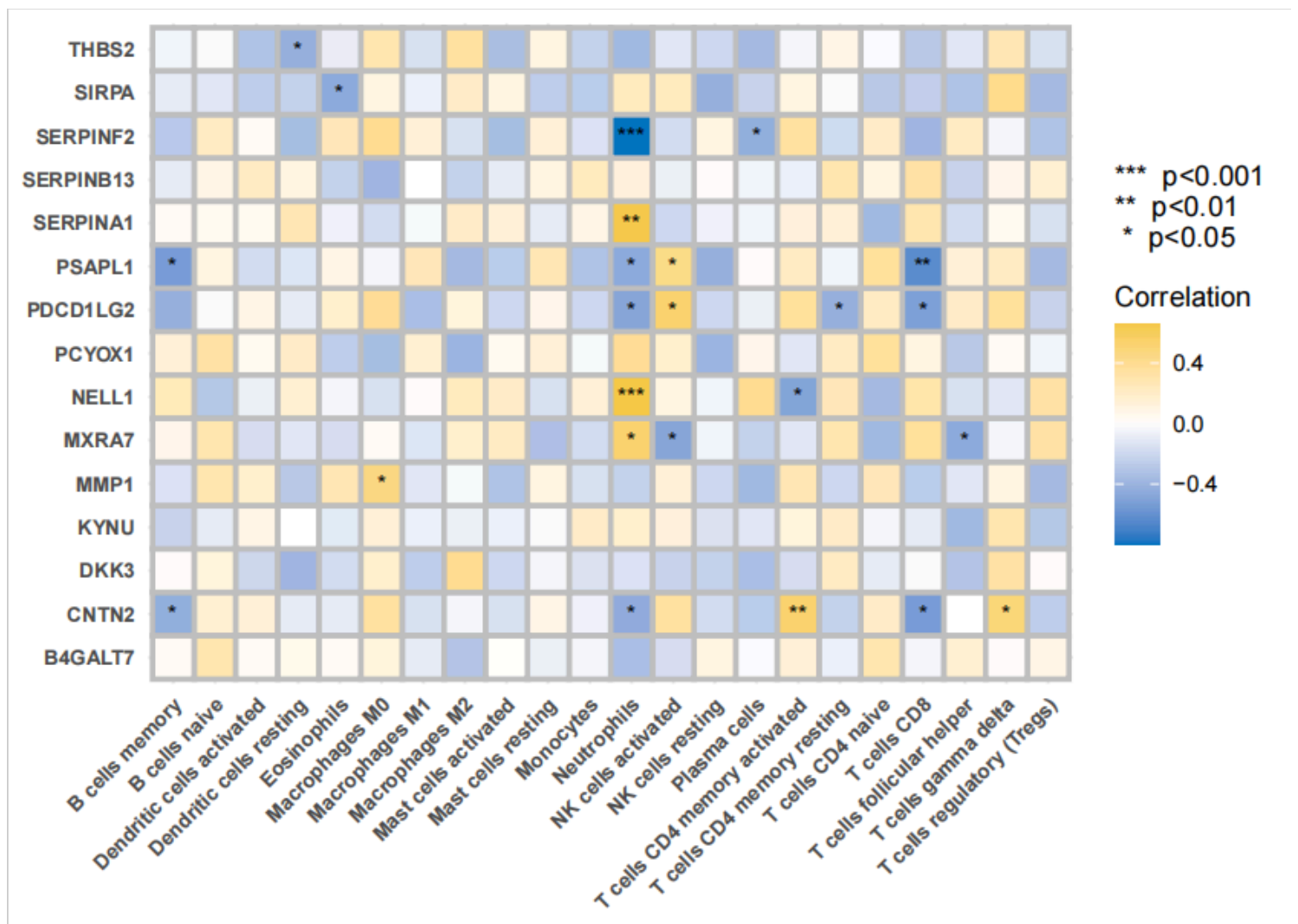


Figure 5

(A) Relative fractions of 22 immune-cell types in cervical cancer and normal tissues. (B) Differential-infiltration analysis highlighting immune-cell populations that differed significantly between groups. (C) Correlation heatmap of core-gene expression and estimated immune-cell fractions in cervical cancer tissues.

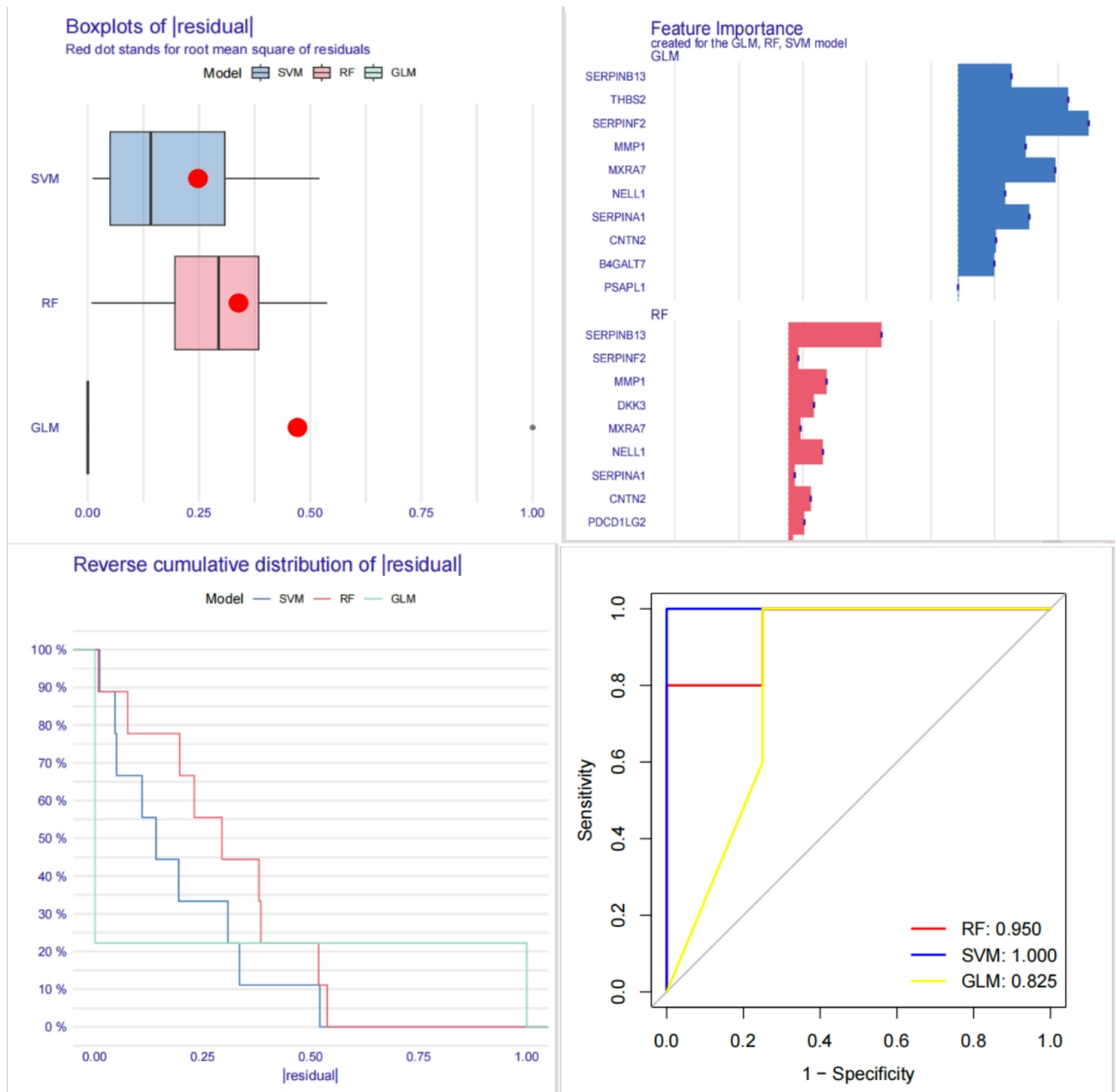


Figure 6

(A) Receiver operating characteristic curves comparing the cervical cancer classification performance of the random forest (RF), support vector machine (SVM), and generalized linear model (GLM). (B-C) Residual diagnostic plots for the three models. (D) Decision curve analysis across threshold probabilities of 0.1-0.8. (E) Feature-importance ranking of candidate genes in the RF model.

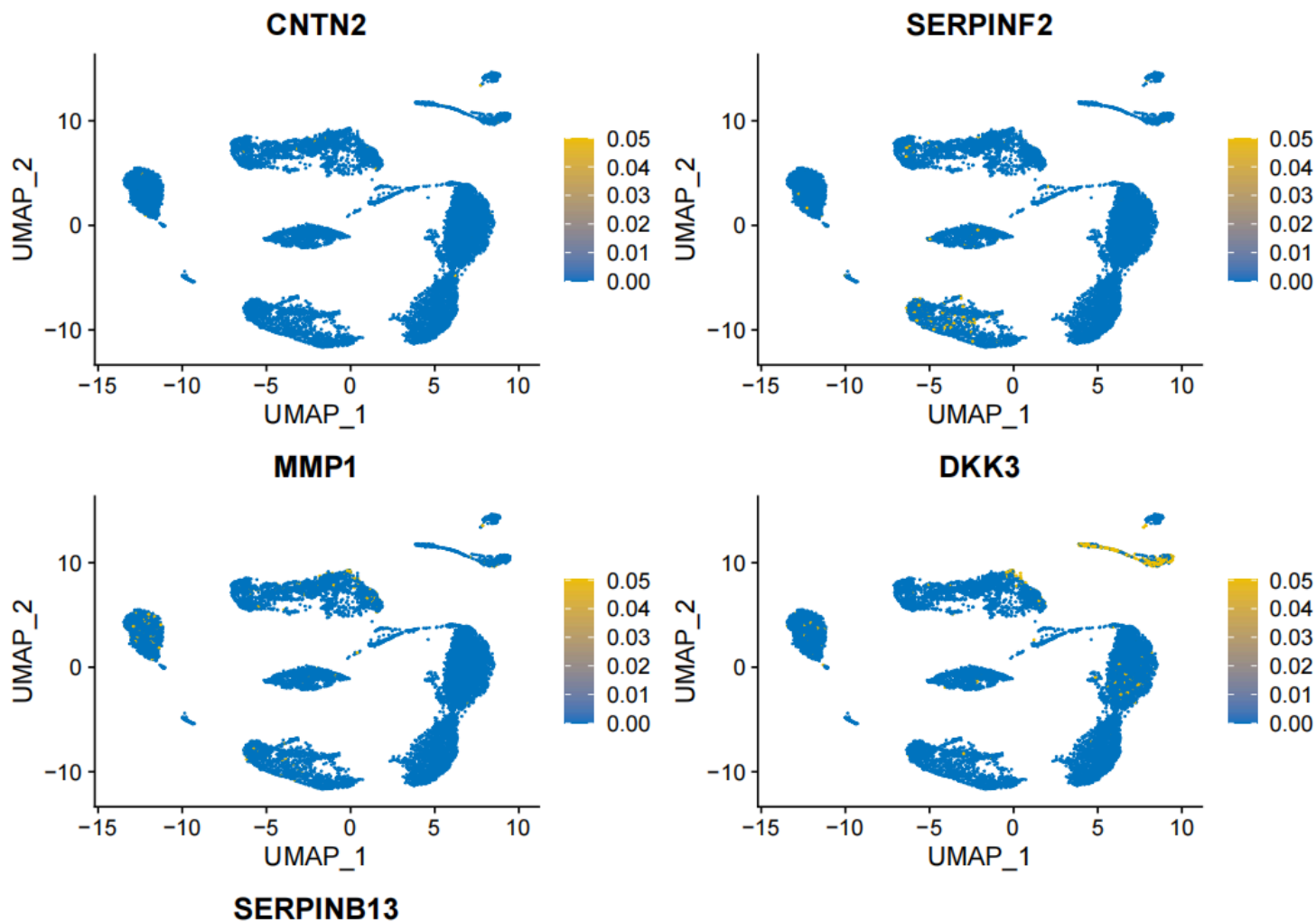


Figure 7

(A) t-SNE plot showing 12 cell populations in cervical cancer samples after single-cell quality control. (B) UMAP plot of the same annotated cell populations. (C) UMAP feature plots showing the cell-type-specific expression of the four core genes.