

Cross-Omics Identification of Core Epithelial Genes Associated with Folate Metabolism in Cervical Cancer and Prediction of an IL1RN/VEGFA Diagnostic Model and Candidate Drugs

Zhengchao Yan

Nanxishan Hospital of Guangxi Zhuang Autonomous Region

Gan Gao

Liuzhou Hospital

Sijuan Tang

The Second Affiliated Hospital of Guilin Medical University

Zhenyu Yang

Nanxishan Hospital of Guangxi Zhuang Autonomous Region

Jie Xin

Nanxishan Hospital of Guangxi Zhuang Autonomous Region

Hui Lin

Nanxishan Hospital of Guangxi Zhuang Autonomous Region

Liuzhu Pan

Nanxishan Hospital of Guangxi Zhuang Autonomous Region

Shiyu Chen

Nanxishan Hospital of Guangxi Zhuang Autonomous Region

Shunwen Teng

Nanxishan Hospital of Guangxi Zhuang Autonomous Region

Liwu Zhang

Nanxishan Hospital of Guangxi Zhuang Autonomous Region

Jianlin Chen

c.j18019000@163.com

Nanxishan Hospital of Guangxi Zhuang Autonomous Region

Research Article

Keywords: cervical cancer, folate metabolism, IL1RN, VEGFA, diagnostic model, molecular docking

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11116144/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

This study integrated single-cell transcriptomics, bulk transcriptomics, GWAS, spatial transcriptomics, and drug-gene data to identify core epithelial genes associated with folate metabolism in cervical cancer. scPagwas revealed the strongest association between epithelial cells and genetic risk for the disease; hdWGCNA and WGCNA, combined with differential expression and folate metabolism gene sets, identified IL1RN and VEGFA. In the CESC cohort, IL1RN expression was downregulated and VEGFA was upregulated; external validation confirmed this trend, and the AUC of the two-gene diagnostic model was 0.771 (95% CI: 0.685–0.848). Enrichment analysis indicated that both genes are involved in DNA repair, the cell cycle, the VEGF/mTOR pathway, and the folate one-carbon pathway, and are associated with immune infiltration. Single-cell localization analysis showed that both IL1RN and VEGFA are expressed in epithelial cells; pseudo-time-series, virtual knockout, and spatial transcriptomics analyses revealed their dynamic expression, regulatory networks, and spatial heterogeneity. Molecular docking analysis demonstrated that beclomethasone can simultaneously target both IL1RN and VEGFA. In summary, IL1RN/VEGFA may serve as potential diagnostic biomarkers for cervical cancer, and beclomethasone is a candidate therapeutic agent.

1.Introduction

Cervical cancer is the fourth most common malignancy among women worldwide and one of the leading causes of cancer-related deaths in women [1]. Although persistent human papillomavirus (HPV) infection has been identified as the primary causative factor for cervical cancer, only a fraction of infected individuals ultimately progress to invasive cancer, suggesting that factors such as the host's genetic background, metabolic status, and tumor microenvironment play important regulatory roles in disease progression [2–4]. In recent years, genome-wide association studies have identified multiple susceptibility loci for cervical cancer in the HLA region and at loci such as PAX8 and CLPTM1L; however, since these studies are largely based on bulk tissue samples, they struggle to reveal the expression patterns and functional consequences of genetic risk signals within specific cellular subpopulations [5].

Folic acid, as a key nutrient in one-carbon metabolism, plays an irreplaceable role in DNA synthesis, repair, and methylation regulation [6]. Epidemiological studies have shown that folate deficiency or metabolic disorders are significantly associated with an increased risk of cervical cancer and cervical intraepithelial neoplasia [7]. Polymorphisms in genes encoding key enzymes in the folate metabolic pathway—such as methylenetetrahydrofolate reductase and methionine synthase reductase—have been shown to be associated with susceptibility to cervical cancer; however, study findings across different populations are not entirely consistent [8, 9]. At the molecular level, folate deficiency may contribute to the development and progression of cervical cancer through multiple pathways, including interference with DNA methylation patterns, effects on DNA synthesis and repair, and potentiation of oncogenic viral activity [10, 11]. However, there are still few reports on the co-expression network characteristics of folate metabolism-related genes in cervical cancer epithelial cells and their integrated analysis with genetic risk for the disease.

The interleukin-1 receptor antagonist IL1RN and vascular endothelial growth factor A (VEGFA) are tumor-associated genes that have attracted widespread attention in recent years. The IL-1RA protein encoded by IL1RN is a natural antagonist of the IL-1 signaling pathway [12]. Low IL1RN expression is associated with the severity of various malignancies, and IL1RN exerts anti-angiogenic effects by inhibiting the induction of IL-1-mediated angiogenic factors, including VEGF [13, 14]. VEGFA, on the other hand, is a central regulator of tumor angiogenesis; it drives endothelial cell proliferation, migration, and survival through the VEGF/VEGFR2 signaling axis, playing a key role in the progression, invasion, and metastasis of cervical cancer [15, 16]. Notably, the VEGF/VEGFR pathway is not only a central driver in the formation of the immunosuppressive microenvironment in gynecological malignancies but also limits the efficacy of monotherapy with immune checkpoint inhibitors through various mechanisms, such as impeding T-cell infiltration and expanding immunosuppressive cell populations [17]. Although preliminary insights into the individual functions of IL1RN and VEGFA in cervical cancer exist, their synergistic relationships at the co-expression network level in cervical epithelial cells, as well as their diagnostic value and potential for targeted intervention, have not yet been systematically investigated.

At the methodological level, the integrated analysis of single-cell RNA sequencing and aggregated GWAS data provides a new paradigm for elucidating cell-type-specific genetic risks in complex diseases [18]. The scPagwas algorithm, which maps GWAS risk loci to single-cell expression networks to calculate disease-related polygenic risk scores for each cell, has been successfully applied to identify trait-associated cell subpopulations in various cancer types [19]. High-dimensional weighted gene co-expression network analysis (hdWGCNA) enables the construction of cell-type-specific co-expression clusters at single-cell resolution, providing a powerful tool for elucidating the transcriptional regulatory mechanisms of key cell subsets [20]. Meanwhile, spatial transcriptomics technologies can resolve spatial heterogeneity in gene expression while preserving tissue spatial information, offering new approaches for validating the spatial localization of key genes in the tumor microenvironment [21]. Furthermore, drug-gene interaction analysis and molecular docking techniques based on databases such as DSigDB provide efficient computational strategies for rapidly screening multi-target drug candidates from approved medications [22]. However, there remains a gap in research regarding the systematic integration and application of the aforementioned multi-omics methods for the screening of core genes related to folate metabolism in cervical cancer, the construction of diagnostic models, and drug prediction.

Based on the research background described above, this study aims to integrate single-cell transcriptomic, bulk transcriptomic, aggregated GWAS statistics, spatial transcriptomics, and drug-gene interaction data. Using scPagwas, we map GWAS genetic risks to the single-cell level to identify cell subpopulations enriched for genetic risk; we utilize hdWGCNA to construct an epithelial cell-specific co-expression network; and we screen for core signature genes by combining WGCNA, differential expression analysis, and gene sets related to folate metabolism. Based on these findings, we will construct a cervical cancer diagnostic model based on these signature genes and validate it externally. We will further systematically analyze the biological functions of these signature genes through functional enrichment, immune infiltration, pseudo-time-series trajectories, virtual knockout, and spatial

transcriptomics analyses. Finally, we will screen for potential targeted drugs through drug enrichment analysis and molecular docking, with the aim of providing new theoretical foundations and candidate targets for the early diagnosis and precision treatment of cervical cancer.

2 Materials and Methods

2.1 Data Sources and Preprocessing

This study integrated multi-omics data, including single-cell transcriptomic data from cervical cancer, bulk transcriptomic data, aggregated GWAS statistics, spatial transcriptomic data, and drug-gene interaction data. The single-cell transcriptomic data were obtained from the ArrayExpress database (accession number: E-MTAB-12305) and included 3 samples of normal cervical tissue (Normal, N) and 4 samples of cervical cancer tissue (Tumor, T). After rigorous quality control, 70,946 high-quality cells were retained for downstream analysis. For bulk transcriptomic data, the combined datasets GSE39001 and GSE63514 were selected as the training set; GSE52903 served as an independent external validation set to verify the expression trends of feature genes and the performance of the diagnostic model. The aggregated GWAS data (ieu-b-4876) were used to characterize genetic risk associated with cervical cancer. The folate metabolism-related gene set (FMRG) was obtained from Lin et al [23]. A complete list of FMRGs is provided in Supplementary Table S1. Spatial transcriptomic data were obtained from the 10× Genomics Visium platform to analyze the spatial expression heterogeneity of key genes at the tissue level. Drug-gene interaction analysis was based on the DSigDB database, and the 3D structures of target proteins were obtained from the RCSB PDB database. The 3D structure of the small molecule beclomethasone was downloaded from the PubChem database.

Bulk transcriptomic data underwent background correction, log transformation, and quantile normalization prior to analysis to remove low-abundance genes and correct for batch effects. Unless otherwise specified, all analyses were performed in the R 4.5.3 environment; specific R packages and algorithms are detailed in the corresponding Methods section.

2.2 Single-Cell Data Clustering and Annotation

Single-cell transcriptomic data were obtained from the ArrayExpress database and included 3 samples of normal cervical tissue and 4 samples of cervical cancer tissue. Data analysis was performed using the Seurat package in R. The quality control criteria were $nFeature_RNA > 50$ and a mitochondrial gene proportion $< 15\%$; a total of 70,946 high-quality cells were retained for downstream analysis. After standardizing the data using LogNormalize, 1,500 highly variable genes were selected, followed by PCA-based dimensionality reduction and correction for batch effects using Harmony based on sample origin. Cell clustering was performed based on the Harmony dimension-reduction results (resolution 0.6), and visualization was conducted using the t-SNE algorithm. The cell clusters were annotated using the SingleR package in conjunction with multiple reference datasets. Differences in cell type composition between the tumor group and the normal control group were further compared, and cell type-specific marker genes were identified, followed by analysis of intergroup differential expression.

2.3 Integrated Analysis of Single-Cell Transcriptomics and GWAS

The scPagwas algorithm was used to integrate aggregated GWAS statistics (ieu-b-4876) with single-cell transcriptomic data; the analysis was based solely on cells from the tumor group. This algorithm maps GWAS risk loci to the single-cell expression network to calculate a disease-related polygenic risk score (TRS) for each cell, which is then aggregated by cell type to assess the contribution of different cell subpopulations to the genetic risk of the disease. Bootstrap tests were used to evaluate the significance of associations between each cell subpopulation and GWAS traits, with the strength of genetic association measured by the $-\log_{10}$ P-value. Furthermore, the Pearson correlation coefficient (PCC) between gene expression and the disease-related polygenic risk score is calculated. Significantly correlated genes are selected based on the criteria $|PCC| > 0.1$ and $\text{adj_pvalue} < 0.05$, and the top 20 genes by PCC are used for visualization.

2.4 Construction of an Epithelial Cell-Specific Co-expression Network Based on hdWGCNA

Based on single-cell data from the tumor cohort, an epithelial cell-specific gene co-expression network was constructed using the hdWGCNA algorithm. First, genes expressed in at least 5% of cells were selected, grouped by cell type and sample type, and metacells were constructed based on the Harmony dimension reduction results ($k = 25$, $\text{max_shared} = 10$) based on the Harmony dimension reduction results and standardized them. We extracted the epithelial cell expression matrix and used TestSoftPowers to evaluate the network fit at different soft thresholds, selecting a soft threshold of 9 to construct the co-expression network. Gene co-expression modules were identified using the dynamic pruning tree algorithm. We calculated the module-specific genes (hMEs), assessed intra-module gene connectivity using ModuleConnectivity, and screened for core genes in each module based on the kME value ($n_hubs = 300$). Module scores for cells were calculated using ModuleExprScore; the spatial distribution of hMEs was visualized using t-SNE feature maps; and the expression profiles of characteristic genes in each module across different cell types, as well as the correlations between modules and cell types, were presented using dot plots and violin plots.

2.5 Construction of a WGCNA-Based Co-expression Network and Identification of Core Modules

Based on the expression matrix from the merged training set, genes with a standard deviation less than 0.5 were removed. Missing values were checked for and removed using $goodSamp \leq s_{Ge} \neq s$, and outliers were eliminated via sample-level clustering (threshold height 20,000, $\text{min Size} = 10$). A clinical trait matrix comprising the Control and CESC groups was constructed. The $\pi ckSoftThreshold$ function was used to evaluate soft thresholds ranging from 1 to 20, and the optimal soft threshold was selected based on the scale-free topological fit index and average connectivity. Construct the adjacency matrix and TOM matrix to perform gene-level clustering. Identify gene co-expression modules using the

dynamic pruning tree algorithm ($\text{minModuleSize} = 60$, $\text{deepSplit} = 2$), and merge similar modules using $\text{mer} \geq \text{CloseModules}$ ($\text{cutHeight} = 0.25$). Calculate the correlation coefficients between module eigenvectors and clinical traits, and plot a module-trait association heatmap. Calculate the gene significance (GS) for each gene using “Control” as the trait, average the values by module, plot a module significance map, and evaluate the strength of association between each module and the clinical phenotype. Output the gene lists for each module for subsequent analysis.

2.6 Differential Expression Analysis, Diagnostic Model Construction, and Validation

Differential expression analysis was performed on the expression matrix obtained by merging the training datasets GSE39001 and GSE63514. Differentially expressed genes were screened using a corrected P-value < 0.05 as the criterion, and the intersection of folate metabolism-related core genes and epithelial cell Hub genes was obtained. The expression levels of the intersecting genes were compared between the control and CESC groups, and their expression trends were verified using the independent external validation dataset GSE52903. Single-gene ROC curve analysis was performed, and the AUC was calculated to evaluate the ability of individual genes to distinguish CESC from control samples. Furthermore, a CESC risk prediction scorecard was constructed based on the characteristic genes; the sum of scores for each gene was used to determine an individual’s probability of developing CESC, and calibration curves were used to assess the consistency between the model-predicted probability and the actual observed probability. For the combined gene model, a multivariable logistic regression model was constructed using the `rms::lrm` function, with CESC status as the outcome and the intersecting genes as predictors. A nomogram was generated from this model using `nomogram` with `fun = plogis` and predicted probabilities at 0.001, 0.01, 0.1, 0.3, 0.5, 0.7, 0.9, and 0.99. The predicted probability for each sample was calculated as `plogis(predict(lrmModel, newdata = rt))`. A comprehensive ROC curve was then plotted based on the gene combination model, and the AUC with its 95% confidence interval was calculated by bootstrap. Internal calibration was evaluated using `calibrate` with the bootstrap method and 1,000 resamples. Finally, the diagnostic performance of the model was validated in the independent external dataset GSE52903.

2.7 Functional Enrichment Analysis

GSEA and GSVA analyses were performed on disease group samples from the merged dataset (GSE39001 and GSE63514), using the KEGG pathway gene set file `c2.cp.kegg.Hs.symbols.gmt` as a reference [24]. For IL1RN and VEGFA, samples were divided into high-expression and low-expression groups based on their median expression levels, respectively. For GSEA, genes were ranked using the mean difference between the high- and low-expression groups as $\log_{2}FC$, with `c2.cp.kegg.Hs.symbols.gmt` as `TERM2GENE` to identify significantly enriched pathways ($p < 0.05$), and the top 5 most significantly enriched pathways in the high-expression group ($NES > 0$) and low-expression group ($NES < 0$) were presented separately. GSVA employed the ssGSEA algorithm to calculate pathway scores based on the same KEGG gene set, `c2.cp.kegg.Hs.symbols.gmt`. After normalization, the differences in

scores between the high- and low-expression groups were compared, and significantly different pathways ($p < 0.05$) were identified using a t-test and presented in bar charts.

2.8 Immune Infiltration Analysis

The CIBERSORT algorithm (perm = 1000) was used to perform deconvolution of the merged expression matrix based on the LM22 signature gene matrix, yielding the relative infiltration proportions of 22 immune cell subsets; samples with $P < 0.05$ were retained for subsequent analysis [25]. Samples were divided into a control group and a CESC group based on sample origin. Stacked bar charts of relative immune cell abundance were plotted to compare differences in immune cell composition between the two groups; box-and-whisker plots were used to display the distribution of each immune cell subset across the two groups, and significance tests for intergroup differences were performed. For samples in the CESC group, the Spearman correlation coefficients between immune cell subsets were calculated, and a correlation heatmap was plotted to analyze the regulatory relationships among immune cells. Furthermore, the expression levels of IL1RN and VEGFA were extracted and subjected to Spearman correlation analysis with the relative infiltration proportions of each immune cell type, respectively. A P-value < 0.05 was considered statistically significant, and lollipop plots were generated to display the correlation coefficients and significance levels.

2.9 Single-Cell Pseudo-Temporal Trajectory Analysis

Epithelial cells and fibroblasts were extracted from the Seurat objects, and up to 5,000 cells were randomly sampled from each cell type to control computational load. A Monocle cell dataset (CellDataSet) was constructed based on the RNA assay-normalized expression matrix, and cell type annotations were added. Standardization and dispersion estimation were performed using *estimateSizeFac* $\rightarrow$ *rs* and *estimateDispersions*, with differentially expressed genes (sig.markers\$gene) used as the sorting filter (setOrderingFilter). Dimension reduction was performed using the DDRTree method (max_components = 2), followed by cell ordering (orderCells). Pseudo-temporal trajectories, cell type trajectories, and pseudo-temporal distribution violin plots were plotted to compare the pseudo-temporal characteristics of epithelial cells and fibroblasts and to further analyze the dynamic expression patterns of the two genes during cell state transitions.

2.10 Virtual Knockout Analysis

Based on single-cell data from the tumor cohort, the RNA assay count matrix was extracted, 5,000 highly variable genes were selected, and the target genes were included in the expression matrix. Use the scTenifoldKnk algorithm to perform virtual knockouts for IL1RN and VEGFA, respectively [26], with the following parameter settings: qc_mtThreshold = 0.1, qc_minLSize = 500, nc_nNet = 10, and nc_nCells = 500. Screen for significantly differentially regulated genes using a p.adj threshold of < 0.05 and output the results. For each target gene, the genes were sorted by fold change (FC), and the top 20 genes were used to generate a bar chart.

2.11 Spatial Transcriptomics Analysis

Spatial transcriptomics data for human cervical cancer obtained using the 10× Genomics Visium platform. The expression matrix and image data were read using *Read10X* and *Read10X_Image*, respectively, to create a Seurat object (*assay* = “*Spatial*”). The distribution of *nCount_spatial* was examined using *VlnPlot* and *SpatialFeaturePlot*. Standardization was performed using *SCTran* or *m*, dimensionality reduction using *RunPCA*, clustering using $F \in dNeighb$ or *s* and $F \in dClusters$, and visualization of the reduced-dimensionality data using *RunUMAP*. Use $F \in dAllMarkers$ to identify marker genes for each cluster. Apply the *Sing* algorithm, using *ref_Human* and *HumanPrimaryCellAtlas_hpa* as reference datasets, to annotate clusters with cell types, plot UMAP and spatial distribution maps of cell types, and visualize the spatial expression patterns of model genes on tissue sections using *SpatialFeaturePlot*.

2.12 Drug Screening and Molecular Docking

Drug enrichment analysis was performed based on the DSigDB database to construct drug-gene regulatory networks. Using IL1RN and VEGFA as target genes, candidate drugs that significantly targeted both genes were screened, with a significance threshold of $p.adjust < 0.05$. Molecular docking analysis was performed on the candidate drug beclomethasone, with the target protein structures being IL1RN (PDB ID: 1ILR) and VEGFA (PDB ID: 2VPF). AutoDock Vina 1.2.3 was used for molecular docking to calculate binding energy and analyze binding modes such as hydrogen bonds, hydrophobic interactions, π - π stacking, and salt bridges between the drug and target proteins. Finally, the docking results were visualized using PyMol 2.5.5 and Ligplot v2.2.9.

2.13 Statistical Analysis

All statistical analyses were performed using R 4.5.3. Quantitative data were expressed as mean $\pm$ standard deviation or median (interquartile range) based on data distribution; comparisons between groups were performed using the independent samples t-test or the Wilcoxon signed-rank test. Categorical data were analyzed using the chi-square test or Fisher’s exact test. Correlation analysis was performed using Pearson or Spearman correlation. Survival analysis and diagnostic performance evaluation were conducted using ROC curves and AUC. P-values from multiple comparisons were corrected using the Benjamini-Hochberg method. A P-value < 0.05 or a corrected P-value < 0.05 was considered statistically significant.

3. Results

3.1 Identification of Cellular Heterogeneity and Key Genes Based on Integrated Analysis of Single-Cell Transcriptomics and GWAS

First, rigorous quality control was performed on samples from the cervical cancer case group and the control group; after quality control filtering, high-quality cells were retained for subsequent analysis. Dimension reduction was determined using principal component analysis (PCA) and the JackStraw test, followed by unsupervised clustering of cells using t-SNE. Ultimately, eight major cell subpopulations were successfully identified and annotated, including epithelial cells, fibroblasts (Fibroblasts), CD8 + T

cells, endothelial cells, macrophages, neutrophils, NK cells, and hematopoietic stem cells (HSCs), with epithelial cells and fibroblasts constituting the major components of the tumor microenvironment (Fig. 1A–E). Analysis of cell proportions revealed distinct dynamic changes in each cell subset between tumor and normal tissues (Fig. 1F). Disease-related polygenic risk scores (TRS) were significantly enriched in specific cell populations (Fig. 1G). Bootstrap analysis (Bootstrap results) revealed that epithelial cells exhibited the strongest genetic association with the trait represented by ieu-b-4876, with significance far exceeding that of other cell populations, suggesting that epithelial cells are the core cell type responsible for the genetic risk of this disease (Fig. 1H). Furthermore, by calculating the phenotype-correlated coefficient (PCC) for genes, we identified PERP, KRT17, SERPINE1, RAC1, and others as the top 20 key genes most strongly associated with genetic risk for the disease (Top 20 Genes with Highest PCC) (Fig. 1I).

3.2 Construction of an Epithelial Cell-Specific Co-expression Network Based on hdWGCNA and Integrated Analysis with Genetic Risk

To elucidate the transcriptional regulatory mechanisms of epithelial cells in disease progression, this study utilized the hdWGCNA algorithm to construct a gene co-expression network at the single-cell level. When the soft threshold was set to 9, the network fit was optimized while maintaining reasonable connectivity, consistent with the characteristics of a scale-free network (Fig. 2A). After constructing the co-expression network based on this optimal threshold, a total of 22 color-coded gene co-expression modules were identified using the Dynamic Tree Cut algorithm, named Epithelial cells-M1 through M22 (Fig. 2B). Correlation analysis between module eigengenes (hMEs) and cell types revealed that the M13 module exhibited high expression levels in epithelial cells (Fig. 2C). The t-SNE plot illustrated the cellular spatial distribution of the module eigengenes (Fig. 2D). Gene significance analysis within each module showed that hub genes in different modules all had high significance scores (Fig. 2E). We performed an integrated analysis of the core genes of the M13 module identified by hdWGCNA and the key genes identified by the previous scPagwas analysis. The results showed a significant overlap between the hdWGCNA module gene set and the scPagwas gene set, with a total of 57 overlapping genes identified (Fig. 2F).

3.3 WGCNA Analysis and Selection of Core Genes

Based on the merged transcriptomic expression matrix, a co-expression network was constructed using WGCNA. By analyzing the scale-free topology model fit (R^2) and average connectivity at different soft thresholds, the optimal soft threshold (Power) was determined to be 6 (Fig. 3A). Using the dynamic pruning tree algorithm, a total of 11 gene co-expression modules were identified (Fig. 3B). To assess the association of each module with clinical phenotypes, correlation coefficients were calculated between the module eigenvectors and the control group (Control) and the CESC group. The results showed that the blue module (MEblue) exhibited the strongest significant positive correlation with CESC ($r = 0.65$, $P = 1e-23$) (Fig. 3C). Gene Significance (GS) analysis within the modules indicated that the average GS value of the blue module was significantly higher than that of the other modules, suggesting that this module

is the core module driving CESC progression (Fig. 3D). A Venn diagram was constructed to identify the intersection of genes from the core module in WGCNA with differentially expressed genes (DEGs) and the folate metabolism-related gene set (FMRG) (Fig. 3E).

3.4 Screening of Key Feature Genes, Construction of Diagnostic Models, and External Validation

By identifying the intersection of core folate metabolism genes and epithelial cell hub genes, we identified two signature genes (IL1RN and VEGFA) (Fig. 4A). Analysis based on the combined GSE39001 and GSE63514 datasets (training set) showed that, compared with the control group, IL1RN expression was significantly downregulated in the CESC group, while VEGFA expression was significantly upregulated (Fig. 4B); Single-gene ROC analysis showed that the AUCs for distinguishing CESC from the control group using IL1RN and VEGFA were 0.644 and 0.727, respectively (Fig. 4C). Validation using the independent external dataset GSE52903 showed that the expression trends of IL1RN and VEGFA in CESC samples were consistent with those in the training set (Fig. 4D). Calibration curves demonstrated that the model's predicted CESC risk probabilities closely matched the observed probabilities (Fig. 4E), and a diagnostic model for CESC risk was constructed using these two genes (Fig. 4F). Comprehensive ROC analysis indicates that this two-gene combination model possesses strong diagnostic performance (AUC = 0.771, 95% CI: 0.685–0.848) (Fig. 4

G).

3.5 Gene Function Enrichment Analysis

GSEA and GSVA pathway enrichment analyses were performed based on the combined dataset (GSE39001 and GSE63514). For IL1RN, the high-expression group was significantly enriched in the “Epithelial cell signaling in Helicobacter pylori infection,” “TCA cycle,” and lipid metabolism pathways such as linoleic acid and arachidonic acid (Fig. 5A). In the IL1RN low-expression group, DNA repair and cell cycle-related pathways—including DNA replication, mismatch repair, base excision repair, and N-glycan biosynthesis—were significantly enriched (Fig. 5B). In the high VEGFA expression group, pathways related to tumor proliferation and progression were significantly enriched, including the cell cycle, DNA replication, the VEGF signaling pathway, the mTOR signaling pathway, and pathways in cancer; they were also enriched in the “One carbon pool by folate” pathway (Fig. 5C). In contrast, the VEGFA low-expression group showed enrichment primarily in metabolic and structural pathways, such as ribosomes, cell adhesion molecules (CAMs), and fatty acid metabolism (Fig. 5D). The results of further GSVA analysis were highly consistent with those of GSEA (Fig. 5E, F): scores for tumor-related cancer pathways and the mTOR signaling pathway were significantly elevated in the high VEGFA expression group, while features related to folate metabolism also showed significant differences in specific expression subgroups of IL1RN and VEGFA.

3.6 Correlation between Hub Biomarkers and Infiltrating Immune Cells

The CIBERSORT algorithm was used to assess the infiltration proportions of 22 immune cell subsets. A stacked bar chart of relative immune cell abundances showed significant differences in immune cell composition between the CESC group and the control group (Fig. 6A). A heatmap analysis of correlations among immune cells revealed complex regulatory relationships within the immune network (Fig. 6B). Compared with the control group, the CESC group exhibited reduced expression of resting CD4 + memory T cells, while expression of activated CD4 + memory T cells and M0 macrophages was elevated (Fig. 6C). IL1RN expression was significantly positively correlated with neutrophils and CD8 + T cells, and significantly negatively correlated with follicular helper T cells. VEGFA expression was significantly positively correlated with plasma cells; it was significantly negatively correlated with resting CD4 memory T cells, resting mast cells, and others (Fig. 6D, E).

3.7 Expression of IL1RN and VEGFA in the CESC Tumor Microenvironment

Through t-SNE-based dimensionality reduction and clustering, as well as SingleR annotation, a total of 8 major cell subpopulations were identified (Fig. 7A). Results from bubble plot and feature map analyses showed that IL1RN and VEGFA exhibited distinct cell-type-specific distribution patterns in the tumor microenvironment (Figs. 7B and C). Further quantitative analysis using violin plots confirmed that IL1RN exhibited the highest expression levels in neutrophils; in contrast, VEGFA was widely expressed in epithelial cells (Fig. 7D).

3.8 Single-Cell Trajectory Analysis

The Monocle algorithm was used to perform single-cell pseudotime trajectory analysis on two cell subpopulations: epithelial cells and fibroblasts. The results showed that the trajectories of epithelial cells and fibroblasts exhibited complex bifurcated structures, with black nodes marking the differentiation origins or branching points of cell states (Fig. 8A, B). Pseudotime distribution violin plots revealed that fibroblasts exhibited significantly higher pseudotime values, whereas epithelial cells were primarily concentrated in the early stages of the trajectories (Fig. 8D). Gene expression time series curves and heatmap analyses revealed that the expression levels of IL1RN and VEGFA exhibited significant nonlinear patterns during the early, branching, and late stages of the trajectories, indicating a high degree of stage specificity during cell state transitions (Fig. 8C, E). Gene expression trajectory feature maps visually revealed the spatial enrichment of these two genes along the evolutionary pathway from epithelial cells to fibroblasts (Fig. 8F).

3.9 Virtual Knockout and Spatial Transcriptomics Analysis

Using the scTenifoldKnn algorithm, virtual knockout analysis of characteristic genes was performed based on the tumor single-cell expression matrix. The results showed that following virtual knockout of IL1RN, the top 20 genes significantly differentially regulated in the network included RPSA, SLURP2, LY6D, and others (Fig. 9A); following virtual knockout of VEGFA, genes such as LY6D, SLURP2, and DSG3 exhibited extremely high fold changes (FC), suggesting that these key genes may play important hub roles in maintaining intracellular homeostasis or in tumor-related signaling networks (Fig. 9B). Using spatial transcriptomic data from human cervical cancer, tissue sections were clustered and cell annotated. UMAP and spatial distribution maps revealed that epithelial cells exhibited specific areas of spatial clustering within the tissue sections (Fig. 9C). Subsequently, the expression levels of IL1RN and VEGFA were mapped onto the spatial sections (Fig. 9D). The results showed that these two genes exhibited distinct spatially heterogeneous expression patterns within the tissue, with their high-expression regions exhibiting spatial overlap or boundary distribution characteristics relative to areas of epithelial cell enrichment.

3.10 Drug Screening and Molecular Docking

Drug enrichment analysis showed that beclomethasone significantly targets both IL1RN and VEGFA ($p = 1.18 \times 10^{-5}$, enrichment fold = 288.93; Fig. 10A and Table S2). Molecular docking further confirmed its binding potential, with binding energies of -6.344 kcal/mol for IL1RN (PDB: 1ILR) and -6.871 kcal/mol for VEGFA (PDB: 2VPF) (Fig. 10B and Table S3). The detailed enrichment scores and docking binding energies are provided in Supplementary Table S2 and Table S3, respectively. Binding mode analysis indicates that beclomethasone stabilizes its binding to both targets primarily through hydrogen bonds (e.g., ASP-104, ARG-102, ASP-34, GLN-37) and hydrophobic interactions (e.g., GLN-149, GLU-64); neither π - π stacking nor salt bridges were detected. In summary, beclomethasone demonstrates reliable binding capacity to simultaneously target IL1RN and VEGFA, providing a structural basis for its potential in multi-target synergistic therapy. This docking analysis serves only as a structural screening step and does not constitute biological validation of functional interactions.

4. Discussion

In this study, the scPagwas algorithm was used to project aggregated GWAS data for cervical cancer (ieu-b-4876) onto a single-cell expression network, revealing that the genetic association between epithelial cells and cervical cancer was significantly stronger than that observed in other cell subpopulations. These results are consistent with the findings from the application of scPagwas across various cancer types, indicating that this algorithm can map polygenic risk to cell-type-specific pathway activity scores, thereby effectively identifying trait-associated cell subsets [19, 27]. From the perspective of the pathogenesis of cervical cancer, HPV infection primarily targets cervical epithelial cells; its oncoproteins E5 and E6 impair MHC class I antigen presentation and interfere with interferon regulator 3 signaling, respectively, thereby promoting immune evasion [28, 29]. Therefore, epithelial cells may serve as the intersection between host genetic susceptibility and viral carcinogenesis; their higher genetic risk

scores not only have a sound biological basis but also suggest that epithelial cells play a central role in the development of cervical cancer.

Notably, folate metabolism disorders may represent another key molecular dimension through which genetic susceptibility and viral carcinogenicity act synergistically in epithelial cells [30]. As a central nutrient in one-carbon metabolism, folate has no direct physiological activity on its own; it must undergo metabolic conversion to participate in DNA synthesis, repair, and methylation maintenance [31, 32]. Multiple studies suggest that folate deficiency or metabolic disorders may increase the risk of cervical cancer and cervical intraepithelial neoplasia [33]. In epithelial cells, folate deficiency can lead to abnormally elevated expression of DNA methyltransferase 1 (DNMT1), which in turn causes abnormally high methylation in the promoter regions of tumor suppressor genes [34]. Wang et al. found in 238 samples of cervical lesions that serum folate levels decreased progressively as the severity of cervical lesions increased, while DNMT1 protein and mRNA expression showed a progressive upward trend [35]; In vitro experiments demonstrated that folate intervention can inhibit cervical cancer cell viability in a dose-dependent manner and reverse the abnormal expression of DNMT1 [36]. This suggests that folate metabolism disorders are not merely a state of nutritional deficiency but may directly participate in the malignant transformation of epithelial cells through epigenetic reprogramming.

The effects of folate metabolism on epithelial cells are also evident in its synergistic relationship with HPV infection. A cross-sectional study by Flatley et al. confirmed that both folate status and abnormal DNA methylation are associated with HPV infection and the progression of cervical lesions [35]. Chen et al. further found in a cohort of 310 participants that folate deficiency was positively correlated with high methylation and low protein expression of the FHIT gene in both CIN and cervical squamous cell carcinoma, and that this interactive effect increased with the severity of the lesion [30]; in vitro folate intervention was able to reverse the high methylation status of the FHIT gene. As a commonly abnormally methylated tumor suppressor gene in cervical cancer, inactivation of FHIT may impair the ability of epithelial cells to monitor genomic damage caused by HPV infection, thereby facilitating persistent viral infection and clonal expansion [37]. At the same time, genetic polymorphisms in MTHFR and MTRR—key enzymes in the folate metabolism pathway—also interact with HPV infection in the development of CIN; the prevalence of high-risk HPV infection, the mutation rate of MTHFR, and the methylation rate of PAX1 gradually increase as the severity of cervical lesions worsens [38]. Taken together, this evidence points to the following inference: folate metabolism disorders, by affecting the methylation homeostasis of epithelial cells, may act in concert with HPV oncoproteins to disrupt the cell's epigenetic balance.

To further elucidate the regulatory architecture within epithelial cells, this study employed hdWGCNA to construct an epithelial cell-specific co-expression network. Under a soft threshold of 9, a total of 22 gene co-expression modules were identified, among which the M13 module exhibited the highest expression level in epithelial cells. hdWGCNA has been successfully applied to elucidate co-expression programs in epithelial cells across various cancers; for example, it identified the glycolysis-related M3 module and screened for the core hub gene STOML2 in ovarian cancer [39]. In this study, we identified the

intersection of genes from the M13 module and key genes from scPagwas, yielding 57 overlapping genes. The advantage of this integrated strategy lies in the fact that hdWGCNA module genes reflect co-expression regulatory structures in epithelial cells, while scPagwas genes reflect those significantly associated with genetic risk for disease; the intersection of the two is more likely to enrich genes that simultaneously possess regulatory hub status and genetic causal contributions.

Among the 57 overlapping genes, after multiple rounds of screening and expression validation, two signature genes—IL1RN and VEGFA—were ultimately identified. In the CESC cohort, IL1RN expression was significantly downregulated, while VEGFA expression was significantly upregulated; this trend was confirmed in the external validation set GSE52903. IL1RN encodes IL-1RA, a natural antagonist of the IL-1 signaling pathway [40]. Previous studies have shown that IL1RN gene polymorphisms are associated with increased susceptibility to cervical squamous cell carcinoma; however, the direction of IL1RN expression in cervical cancer tissues remains controversial across different studies. This discrepancy may be related to heterogeneity in sample sources, detection platforms, disease staging, and HPV infection status, or it may reflect the dual functions of IL1RN in different tumor microenvironment contexts. Notably, low IL1RN expression has been shown to correlate with disease severity in various malignancies, and IL1RN exerts anti-angiogenic effects by inhibiting the induction of IL-1-mediated angiogenic factors (including VEGF) [41–43]. This provides a mechanistic explanation for the functional association between IL1RN downregulation and VEGFA upregulation.

Building on this, the potential regulatory link between folate metabolism and the IL1RN-VEGFA axis identified in this study warrants further consideration. Abnormal DNMT1 expression and methylation dysregulation caused by folate deficiency may affect the methylation status of the IL1RN promoter region, thereby contributing to the downregulation of its expression [44]; simultaneously, folate can also directly upregulate VEGF expression through the FRA-mediated STAT3 signaling pathway [45]. Thus, folate metabolism disorders may serve as a bridge linking epigenetic regulation to the remodeling of the angiogenic microenvironment, while the imbalance in IL1RN and VEGFA expression may represent a downstream effect of this chain in epithelial cells [46]. The beneficial effects of folate supplementation on metabolic and inflammatory markers in patients with CIN2/3—including a significant reduction in high-sensitivity C-reactive protein and an increase in total antioxidant capacity—also indirectly support a functional link between folate metabolic status and the local inflammatory microenvironment of the cervix [47]. In summary, the folate metabolic pathway and the IL1RN-VEGFA axis identified in this study may jointly participate in the epigenetic-microenvironmental remodeling process of epithelial cells under genetic susceptibility and HPV-induced carcinogenic stress. This integrated framework provides a testable hypothesis for understanding how genetic risk in cervical cancer epithelial cells is translated into a neoplastic phenotype via metabolic-epigenetic mechanisms, and suggests that folate metabolic status may serve as a potential auxiliary target for cervical cancer risk stratification and intervention.

Analysis of immune infiltration further revealed that these two genes are closely associated with the tumor immune microenvironment. IL1RN expression was positively correlated with the infiltration of neutrophils and CD8 + T cells. Single-cell localization analysis showed that IL1RN was most highly

expressed in neutrophils, while VEGFA was widely expressed in epithelial cells. These findings are consistent with recent research on tumor-associated neutrophils (TANs). A study on lung squamous cell carcinoma confirmed that tumor-associated neutrophils upregulate IL1RN expression, suggesting that these cells may exert immunosuppressive effects through IL1RN expression [48]. Furthermore, previous studies have shown that VEGFA-positive tumor-associated neutrophils are associated with the formation of an immunosuppressive microenvironment in cervical cancer and with resistance to PD-1 immunotherapy; Eliminating VEGFA-positive tumor-associated neutrophils can significantly enhance the antitumor efficacy of PD-1 inhibitors [49]. This study observed high expression of IL1RN in neutrophils and a functional association between IL1RN and VEGFA, suggesting that IL1RN may participate in the remodeling of the cervical cancer immune microenvironment by regulating the pro-angiogenic phenotype of neutrophils. This hypothesis requires further experimental validation in subsequent studies.

Functional enrichment analysis revealed that the VEGFA-high expression group was significantly enriched in the cell cycle, DNA replication, VEGF signaling pathway, mTOR signaling pathway, and cancer pathways, as well as in the “One-carbon pool by folate” pathway. This finding extends the function of VEGFA beyond traditional angiogenesis to the fields of folate metabolism and the one-carbon cycle. Folic acid, a key nutrient in one-carbon metabolism, is significantly associated with an increased risk of CIN3+, and this risk rises further when folic acid deficiency is combined with high levels of DNA methylation [50]. Folic acid status and abnormal DNA methylation have been shown to be associated with HPV infection and the pathogenesis of cervical cancer; folic acid directly participates in the regulation of DNA methylation through one-carbon metabolism [51]. This study identified an association between VEGFA and the folate one-carbon pool pathway, suggesting that VEGFA may contribute to cervical cancer progression associated with folate metabolism dysregulation by influencing the metabolic reprogramming of tumor cells, thereby providing new insights into the molecular link between folate metabolism and angiogenesis in cervical cancer. Furthermore, VEGFA has been shown to activate the PI3K/Akt and its downstream mTOR signaling pathways, promoting the growth and invasion of cervical cancer cells [52], which is consistent with the GSEA results of this study.

The two-gene diagnostic model based on IL1RN and VEGFA had an AUC of 0.771 (95% CI: 0.685–0.848), and the calibration curve demonstrated good agreement between predicted and observed probabilities. Although this AUC value is lower than that of models based on large gene combinations, this model is based on only two genes and has been validated using an external independent dataset, indicating high feasibility for clinical translation. Single-gene ROC analysis showed that VEGFA (AUC = 0.727) had superior diagnostic performance compared to IL1RN (AUC = 0.644), which is consistent with extensive literature reporting VEGFA as a key driver gene in cervical cancer [52]. Spatial transcriptomic analysis revealed that IL1RN and VEGFA exhibit distinct spatially heterogeneous expression patterns in tissue sections; their highly expressed regions show spatial overlap with or are distributed along the boundaries of areas enriched with epithelial cells, indicating that the microenvironment shapes the malignant epithelial program and extracellular matrix-associated immunological niches. scTenifoldKnk virtual knockout analysis showed that following the knockout of IL1RN or VEGFA, genes such as LY6D,

SLURP2, and DSG3 exhibited extremely high differential regulation fold changes within the network. LY6D, as a marker of epithelial cell differentiation, shows a regulatory association with IL1RN and VEGFA, suggesting that these two genes may be involved in maintaining the differentiated state of epithelial cells. SLURP2 and DSG3, as genes associated with the epidermal differentiation complex, exhibit responses that indicate the IL1RN/VEGFA regulatory network may be involved in processes related to epithelial–mesenchymal transition [53–55].

Drug enrichment analysis based on the DSigDB database revealed that beclomethasone significantly targets both IL1RN and VEGFA ($p = 1.18 \times 10^{-5}$, enrichment fold of 288.93). Molecular docking revealed that beclomethasone exhibits reliable binding affinity to both IL1RN (binding energy – 6.344 kcal/mol) and VEGFA (binding energy – 6.871 kcal/mol). Beclomethasone is a glucocorticoid receptor agonist widely used as an anti-inflammatory drug in the treatment of asthma and allergic rhinitis. Preliminary reports have indicated its antitumor activity, and steroidal drugs (including beclomethasone dipropionate) have been shown to possess topoisomerase II inhibitory potential [56, 57]. From a mechanistic perspective, glucocorticoids, including beclomethasone, can suppress pro-inflammatory signaling through glucocorticoid receptor-mediated transrepression, which interferes with the activity of transcription factors such as NF- κ B [58]. Given that IL1RN is a natural antagonist of the IL-1 signaling pathway and that IL-1 can induce VEGF expression, beclomethasone may inhibit the IL-1/VEGF axis through glucocorticoid receptor-mediated anti-inflammatory signaling, thereby exerting a synergistic regulatory effect on both targets. However, it should be emphasized that the molecular docking analysis in this study represents only a structural screening step; the functional regulatory effects of beclomethasone on IL1RN and VEGFA still require validation through cellular and animal experiments.

This study has several limitations. First, the single-cell data were derived from a limited sample size in the ArrayExpress database (3 normal and 4 tumor tissue samples), which may not fully reflect the heterogeneity of cervical cancer. Second, the sample size and population composition of the GWAS aggregate statistic “ieu-b-4876” may limit the generalizability of the genetic risk score. Third, the scPAGWAS analysis was based solely on cells from the tumor group and did not fully utilize the genetic risk information from normal controls. Fourth, the diagnostic model was constructed using a retrospective dataset and lacks validation in a prospective clinical cohort. Fifth, the targeted prediction for beclomethasone is based solely on computational analysis and lacks experimental validation. Sixth, the set of folate metabolism-related genes was derived from published literature and may be subject to selection bias.

5. Conclusion

In summary, this study suggests that IL1RN and VEGFA may serve as potential core epithelial cell signature genes associated with folate metabolism in cervical cancer, based on an integrated multi-omics analysis. Both genes contribute to cervical cancer progression by regulating DNA repair, the cell cycle, the VEGF/mTOR signaling pathway, and the folate one-carbon pool pathway, and are closely associated with the remodeling of the tumor immune microenvironment. The two-gene diagnostic model

demonstrates good diagnostic performance and consistency in external validation; beclomethasone, as a candidate drug that simultaneously targets IL1RN and VEGFA, warrants further investigation. Future research should focus on validation with a larger sample size, experimental studies on the functional mechanisms of IL1RN/VEGFA, and preclinical evaluation of beclomethasone's anti-cervical cancer activity.

Declarations

Ethical approval and Accordance Statement

Not applicable. This study used publicly available, de-identified data and did not involve human participants requiring informed consent. Availability of data and materials

Consent to publish

Not applicable. This manuscript does not contain any individual person's data in any form.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare that they have no competing interests. Contributions

Acknowledgements

Not applicable.

Data availability

The single-cell transcriptomic data analyzed in this study are publicly available in the ArrayExpress database under accession number E-MTAB-12305 (<https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-12305>). The bulk transcriptomic datasets are available in the Gene Expression Omnibus (GEO) under accession numbers GSE39001, GSE63514, and GSE52903 (<https://www.ncbi.nlm.nih.gov/geo/>). The aggregated GWAS summary statistics are available from the IEU OpenGWAS database under dataset ID ieu-b-4876 (<https://gwas.mrcieu.ac.uk/datasets/ieu-b-4876/>). The folate metabolism-related gene set was obtained from Lin et al., and the complete gene list is provided in Supplementary Table S1. The spatial transcriptomic data analyzed in this study were obtained from the 10x Genomics public dataset "Human Cervical Cancer (FFPE)" (Visium platform), which is publicly available at <https://www.10xgenomics.com/datasets/human-cervical-cancer-1-standard>. Drug-gene interaction data were obtained from the DSigDB database (<http://tanlab.ucdenver.edu/DSigDB/>). The 3D protein

structures were obtained from the RCSB PDB database (<https://www.rcsb.org/>), and the 3D structure of beclomethasone was obtained from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>). All other data generated or analyzed during this study are included in this published article and its supplementary information files. Further information and analysis code are available from the corresponding author upon reasonable request.

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*. 2024;74:229–63.
2. Schiffman M, et al. Carcinogenic human papillomavirus infection. *Nat Rev Dis Primers*. 2016;2:16086.
3. Singh A, et al. Role of human papillomavirus (HPV) variants and host genetic susceptibility in cervical carcinogenesis. *Arch Microbiol*. 2026;208:198.
4. Gore M, Kabekkodu SP, Chakrabarty S. Exploring the metabolic alterations in cervical cancer induced by HPV oncoproteins: From mechanisms to therapeutic targets. *Biochim Biophys Acta Rev Cancer*. 2025;1880:189292.
5. Bowden SJ, et al. Genetic variation in cervical preinvasive and invasive disease: a genome-wide association study. *Lancet Oncol*. 2021;22:548–57.
6. Stover PJ. One-carbon metabolism-genome interactions in folate-associated pathologies. *J Nutr*. 2009;139:2402–5.
7. Piyathilake CJ, et al. Lower risk of cervical intraepithelial neoplasia in women with high plasma folate and sufficient vitamin B12 in the post-folic acid fortification era. *Cancer Prev Res (Phila)*. 2009;2:658–64.
8. Gong JM, Shen Y, Shan WW, He YX. The association between MTHFR polymorphism and cervical cancer. *Sci Rep*. 2018;8:7244.
9. Badiga S, et al. A lower degree of PBMC L1 methylation in women with lower folate status may explain the MTHFR C677T polymorphism associated higher risk of CIN in the US post folic acid fortification era. *PLoS ONE*. 2014;9:e110093.
10. Crider KS, Yang TP, Berry RJ, Bailey LB. Folate and DNA methylation: a review of molecular mechanisms and the evidence for folate's role. *Adv Nutr*. 2012;3:21–38.
11. Sobral AF et al. Unveiling the Therapeutic Potential of Folate-Dependent One-Carbon Metabolism in Cancer and Neurodegeneration. *Int J Mol Sci* 25(2024).
12. Dinarello CA. Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. *Blood*. 2011;117:3720–32.
13. Ding Y, et al. Low expression of interleukin-1 receptor antagonist correlates with poor prognosis via promoting proliferation and migration and inhibiting apoptosis in oral squamous cell carcinoma. *Cytokine*. 2024;179:156595.

14. Konishi N, et al. Interleukin-1 receptor antagonist inhibits the expression of vascular endothelial growth factor in colorectal carcinoma. *Oncology*. 2005;68:138–45.
15. Chuai Y et al., *Vascular endothelial growth factor (VEGF) targeting therapy for persistent, recurrent, or metastatic cervical cancer*. *Cochrane Database Syst Rev* 3, Cd013348 (2021).
16. Nandi A, Ghosh A, Sil SK. The interplay between angiogenesis and cervical cancer: a comprehensive review. *Expert Rev Anticancer Ther*. 2025;25:633–42.
17. Wang Y, et al. VEGF/VEGFR targeting-induced vascular normalization: a key strategy to reverse cold tumor phenotype and potentiate immunotherapy in gynecologic cancers. *Int Immunopharmacol*. 2026;182:116821.
18. Zhang MJ, et al. Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat Genet*. 2022;54:1572–80.
19. Deng C et al. Pan-cancer analysis of single-cell profiles with polygenic signals reveals genetic influences on tumor immunity across cancers. *Commun Biol* 9(2026).
20. Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep Methods*. 2023;3:100498.
21. Sharma A, et al. From tissue architecture to clinical insights: Spatial transcriptomics in solid tumor studies. *Semin Oncol*. 2025;52:152389.
22. Yoo M, et al. DSigDB: drug signatures database for gene set analysis. *Bioinformatics*. 2015;31:3069–71.
23. Lin J et al. Integrated multi-omics analysis reveals folate metabolism-related genes as prognostic markers and therapeutic targets in clear cell renal cell carcinoma. *BMC Cancer* (2026).
24. Liberzon A, et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27:1739–40.
25. Newman AM, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12:453–7.
26. Osorio D, et al. scTenifoldKnn: An efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation. *Patterns (N Y)*. 2022;3:100434.
27. Ma Y, et al. Polygenic regression uncovers trait-relevant cellular contexts through pathway activation transformation of single-cell RNA sequencing data. *Cell Genom*. 2023;3:100383.
28. Miyauchi S, et al. Human papillomavirus E5 suppresses immunity via inhibition of the immunoproteasome and STING pathway. *Cell Rep*. 2023;42:112508.
29. Ronco LV, Karpova AY, Vidal M, Howley PM. Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity. *Genes Dev*. 1998;12:2061–72.
30. Bai LX, et al. Folate deficiency and FHIT hypermethylation and HPV 16 infection promote cervical cancerization. *Asian Pac J Cancer Prev*. 2014;15:9313–7.
31. Wang W, et al. Associations of RBC and Serum Folate Concentrations with Cervical Intraepithelial Neoplasia and High-Risk Human Papillomavirus Genotypes in Female Chinese Adults. *J Nutr*. 2022;152:466–74.

32. Pathak S, Bhatla N, Singh N. Cervical cancer pathogenesis is associated with one-carbon metabolism. *Mol Cell Biochem.* 2012;369:1–7.
33. Chen R, Ding C, Wang D. Association between folate level and cervical intraepithelial neoplasia risk: a systematic review and meta-analysis. *Eur J Cancer Prev.* 2025;34:185–92.
34. Ghoshal K, et al. A folate- and methyl-deficient diet alters the expression of DNA methyltransferases and methyl CpG binding proteins involved in epigenetic gene silencing in livers of F344 rats. *J Nutr.* 2006;136:1522–7.
35. Wang JT, et al. Folate deficiency and aberrant expression of DNA methyltransferase 1 were associated with cervical cancerization. *Curr Pharm Des.* 2014;20:1639–46.
36. Ding L, Ma J, Zhou Q, Wang J. [Effect of folate on the proliferation of human cervical cancer cell and relationship with HPV16]. *Wei Sheng Yan Jiu.* 2013;42:748–53.
37. Neyaz MK, et al. Effect of aberrant promoter methylation of FHIT and RASSF1A genes on susceptibility to cervical cancer in a North Indian population. *Biomarkers.* 2008;13:597–606.
38. Zhou XY, et al. Interaction of MTHFR polymorphism with PAX1 methylation in cervical cancer. *Open Life Sci.* 2025;20:20221052.
39. Liu L, et al. Single-cell and spatial analyses identify a glycolysis-high state driven by STOML2 in epithelial ovarian cancer. *Exp Cell Res.* 2026;461:115066.
40. Lennard AC. Interleukin-1 Receptor Antagonist. *Crit Rev Immunol.* 2017;37:531–59.
41. Wu S, Hu G, Chen J, Xie G. Interleukin 1 β and interleukin 1 receptor antagonist gene polymorphisms and cervical cancer: a meta-analysis. *Int J Gynecol Cancer.* 2014;24:984–90.
42. Shiiba M, et al. Interleukin-1 receptor antagonist (IL1RN) is associated with suppression of early carcinogenic events in human oral malignancies. *Int J Oncol.* 2015;46:1978–84.
43. Elaraj DM, et al. The role of interleukin 1 in growth and metastasis of human cancer xenografts. *Clin Cancer Res.* 2006;12:1088–96.
44. Valcárcel G, et al. Modulating immune cell fate and inflammation through CRISPR-mediated DNA methylation editing. *Sci Adv.* 2025;11:eadt1644.
45. Hansen MF, et al. Folic acid mediates activation of the pro-oncogene STAT3 via the Folate Receptor alpha. *Cell Signal.* 2015;27:1356–68.
46. Navarro-Rodríguez PM et al. MTHFR–folate axis as a modulator of the epigenetic landscape in autoimmune diseases (Review). *Int J Mol Med* 57(2026).
47. Sabihi S, Vahedpoor Z, Saraf-Bank S, Nourian M. Effects of folate supplementation on recurrence and metabolic status of cervical intraepithelial neoplasia grade 2/3 in overweight and obese women: a randomized double-blind placebo-controlled trial. *Eur J Clin Nutr.* 2022;76:666–70.
48. Hao B, et al. Single-cell RNA sequencing analysis revealed cellular and molecular immune profiles in lung squamous cell carcinoma. *Transl Oncol.* 2023;27:101568.
49. Yao S, et al. Eliminating VEGFA+ tumor-associated neutrophils by antibody-drug conjugates boosts antitumor immunity and potentiates PD-1 immunotherapy in preclinical models of cervical cancer.

Cell Death Dis. 2025;16:115.

50. Agudelo MC, et al. Folate deficiency modifies the risk of CIN3 + associated with DNA methylation levels: a nested case-control study from the ASCUS-COL trial. *Eur J Nutr.* 2024;63:563–72.
51. Flatley JE, et al. Folate status and aberrant DNA methylation are associated with HPV infection and cervical pathogenesis. *Cancer Epidemiol Biomarkers Prev.* 2009;18:2782–9.
52. Chen B, Zhang C, Dong P, Guo Y, Mu N. Molecular regulation of cervical cancer growth and invasion by VEGFa. *Tumour Biol.* 2014;35:11587–93.
53. Andersson N, et al. Lymphocyte antigen 6 superfamily member D is a marker of urothelial and squamous differentiation: implications for risk stratification of bladder cancer. *Biomark Res.* 2020;8:51.
54. Arredondo J, Chernyavsky AI, Jolkovsky DL, Webber RJ, Grando SA. SLURP-2: A novel cholinergic signaling peptide in human mucocutaneous epithelium. *J Cell Physiol.* 2006;208:238–45.
55. Wang T, et al. DSG3 promotes bladder cancer growth and metastasis via AKT/GSK3 β / β -catenin pathway. *J Transl Med.* 2025;23:729.
56. Cianchetti S, et al. The beclomethasone anti-inflammatory effect occurs in cell/mediator-dependent manner and is additively enhanced by formoterol: NF κ B, p38, PKA analysis. *Life Sci.* 2018;203:27–38.
57. Al-Karmalawy AA, et al. Exploring the Topoisomerase II Inhibitory Potential of Steroidal Drugs as a Recommended Mechanism of Action for Their Anticancer Activity: In Silico and In Vitro Assessments. *Drug Dev Res.* 2025;86:e70152.
58. Nakao S, et al. Dexamethasone inhibits interleukin-1 β -induced corneal neovascularization: role of nuclear factor-kappaB-activated stromal cells in inflammatory angiogenesis. *Am J Pathol.* 2007;171:1058–65.

Figures

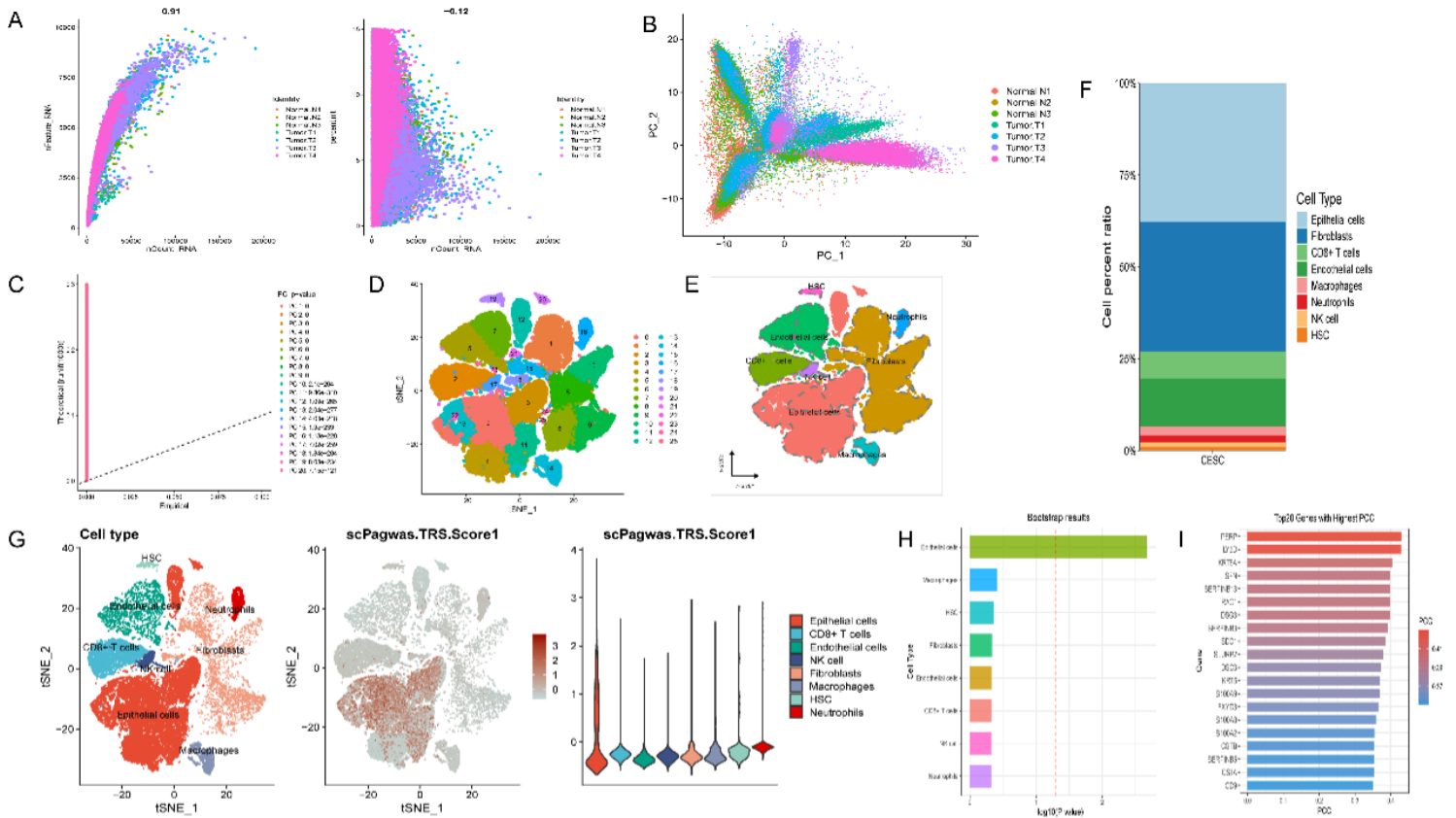


Figure 1

Construction of a single-cell transcriptome profile for cervical cancer and analysis of genetic associations with GWAS traits.

(A) Scatter plot for cell quality control. (B) Principal component analysis (PCA) plot for sample origin grouping, showing the distribution of the Normal and Tumor groups in the reduced-dimension space. (C) JackStraw test scatter plot used to assess the significance of the top 20 principal components (PCs) in the PCA; the dashed line indicates the threshold for statistical significance. (D) Initial clustering plot based on t-SNE dimensionality reduction, identifying a total of 26 cell subpopulations (Clusters 0–25). (E) t-SNE cell type atlas based on SingleR and annotated reference datasets, identifying a total of 7 major cell subpopulations. (F) Stacked bar chart showing the cellular proportions of the 7 major cell subpopulations in the disease group. (G) Visualization of scPagwas analysis results. (H) Significance of associations between each cell subpopulation and GWAS traits. The red dashed line represents the threshold for statistical significance ($P = 0.05$). (I) Bar chart of the top 20 key genes ranked by their correlation coefficient (PCC) with genetic risk for the disease; the depth of color represents the magnitude of the PCC value.

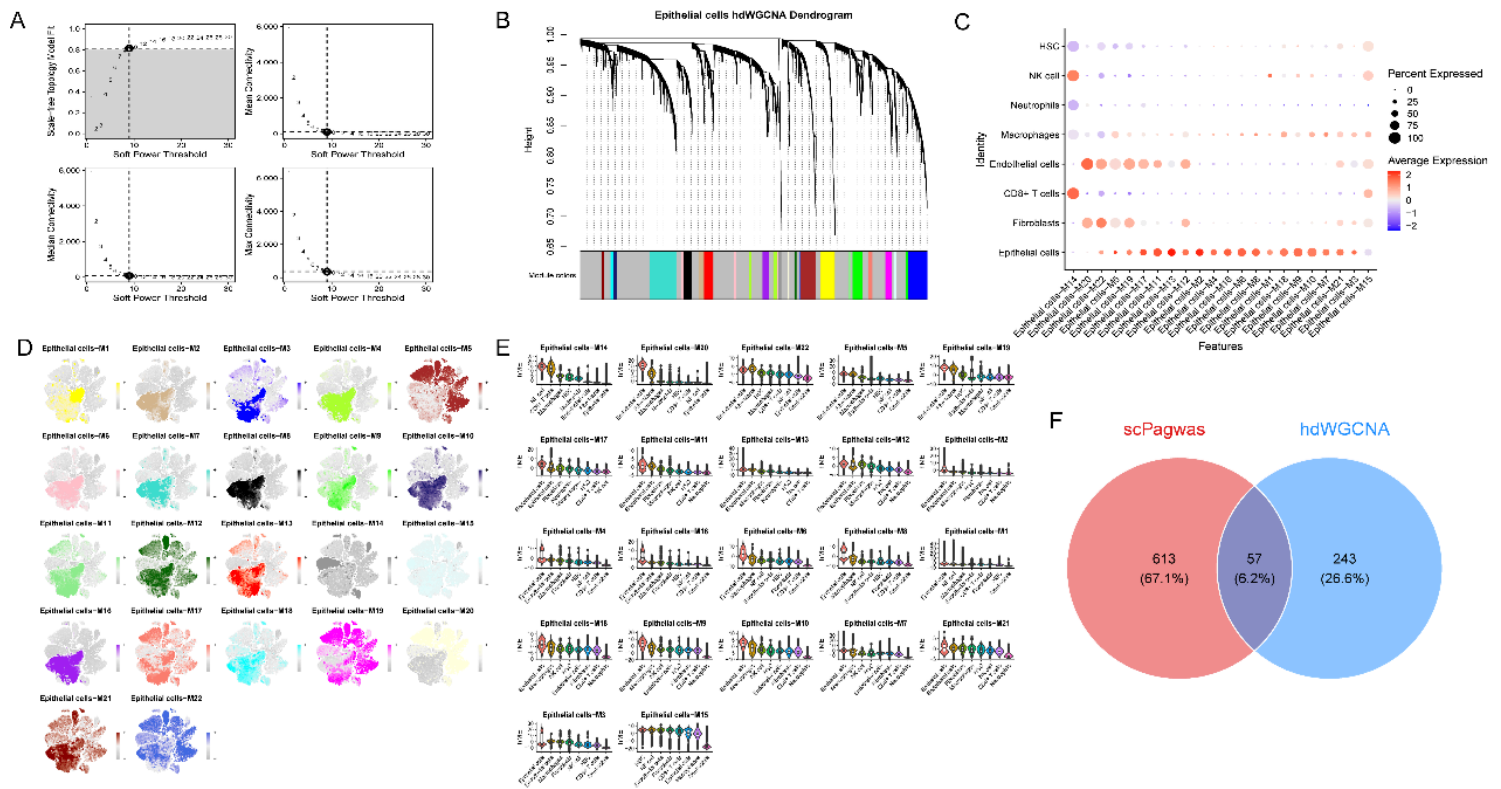


Figure 2

Construction of the epithelial cell-specific gene co-expression network.

(A) Soft threshold (Soft Power) selection plot. The red dashed line indicates the optimal soft threshold (Power = 9). (B) Hierarchical clustering dendrogram of the epithelial cell gene co-expression network. (C) Shows the expression percentage and average expression level of characteristic genes (hMEs) in each module across different cell types. (D) Feature plots of characteristic genes (hMEs) in each gene co-expression module in the t-SNE dimensionality-reduction space, illustrating the spatial distribution characteristics of module scores in the cell space. (E) Violin plots of gene significance within each module, showing the strength of association between hub genes and their corresponding phenotypes or traits. (F) Intersection of key genes identified by scPagwas and genes in the hdWGCNA module.

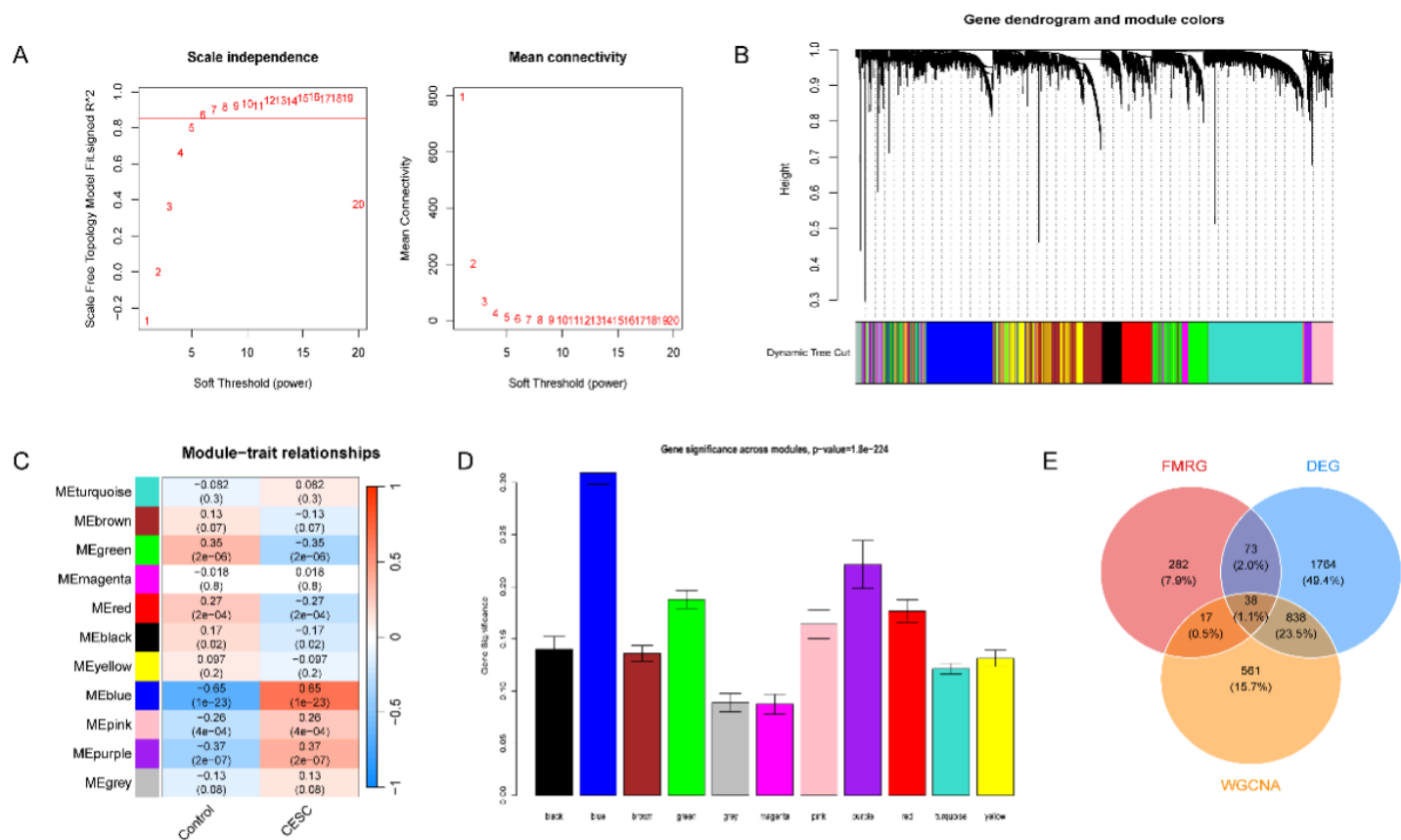


Figure 3

WGCNA analysis identifies core modules and candidate genes associated with CESC.

(A) Soft-threshold screening. The red dashed line indicates the optimal soft threshold (power = 6) that yields a scale-free network. (B) Gene clustering tree and module partitioning. (C) Heatmap of module-trait relationships. (D) Bar chart showing the distribution of gene significance within modules. (E) Intersection between the FMRG gene set, DEGs, and the WGCNA core module gene set.

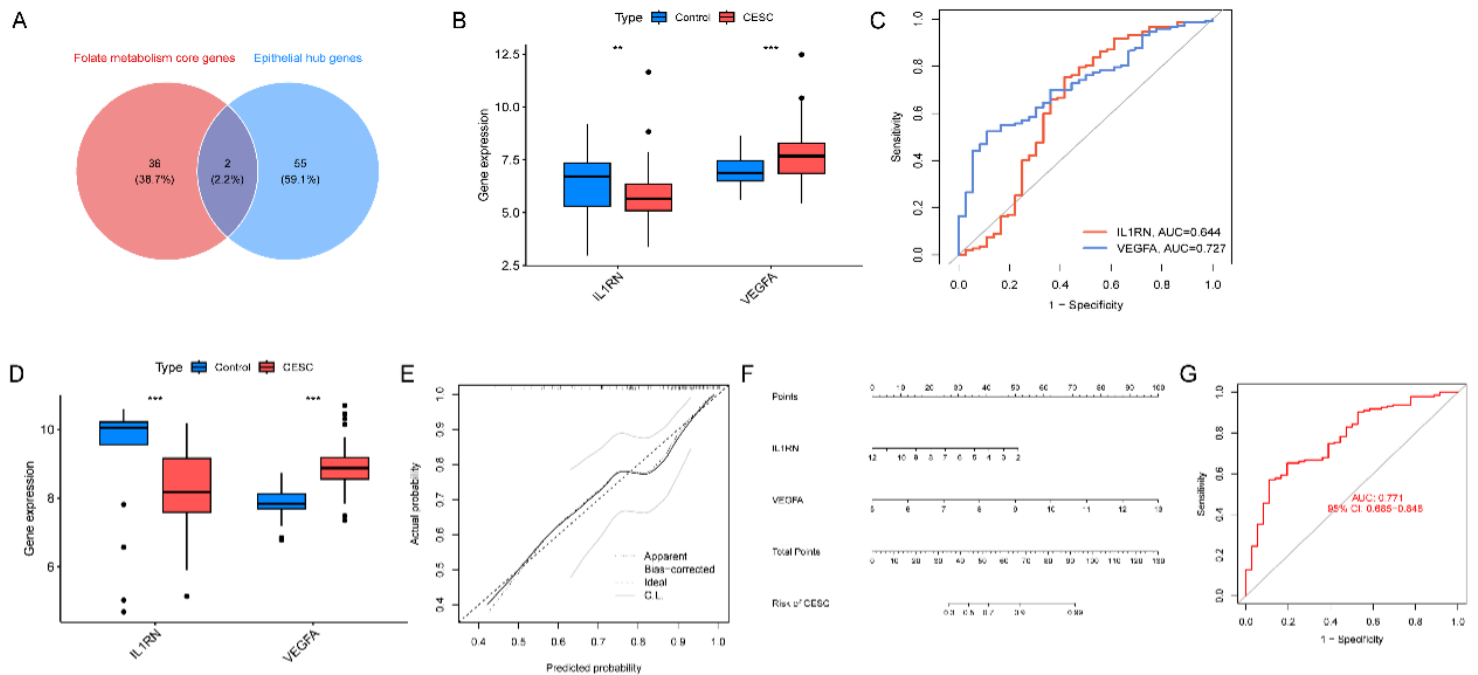


Figure 4

CESC diagnostic model constructed based on the intersection of core folate metabolism genes and epithelial cell hub genes, and its validation.

(A) A Venn diagram illustrating the intersection between core folate metabolism genes and epithelial cell hub genes. (B) Box plots showing the differential expression of IL1RN and VEGFA between the control group and the CESC group in the training set. (C) Single-gene receiver operating characteristic (ROC) curves for IL1RN and VEGFA in diagnosing CESC in the training set. (D) Box plots showing the differential expression of IL1RN and VEGFA in the external independent validation set (GSE52903). (E) Calibration curve for the nomogram model. (F) CESC risk prediction nomogram based on the two feature genes IL1RN and VEGFA; the individual's probability of developing CESC can be determined by calculating the sum of the scores for each gene. (G) ROC curve for the nomogram-based combined diagnostic model, with an AUC of 0.771 (95% CI: 0.685–0.848), indicating good diagnostic performance.

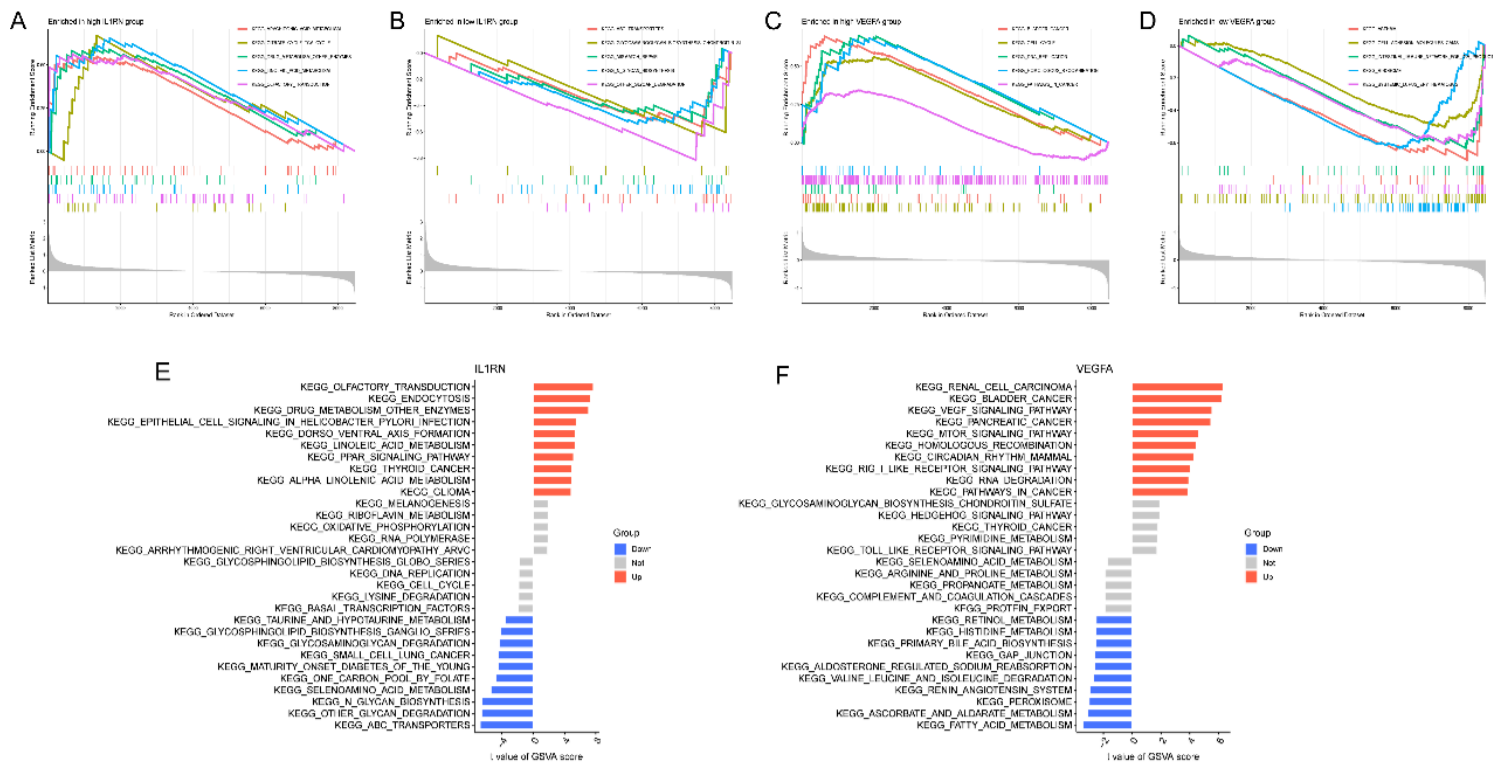


Figure 5

GSEA and GSVA enrichment analyses.

(A–B) GSEA enrichment plots of KEGG pathways for the high- and low-expression groups of IL1RN. (C–D) GSEA enrichment plots of KEGG pathways for the high- and low-expression groups of VEGFA. (E–F) Bar charts comparing differences in KEGG pathways between the high- and low-expression groups of IL1RN and VEGFA based on GSVA scores.

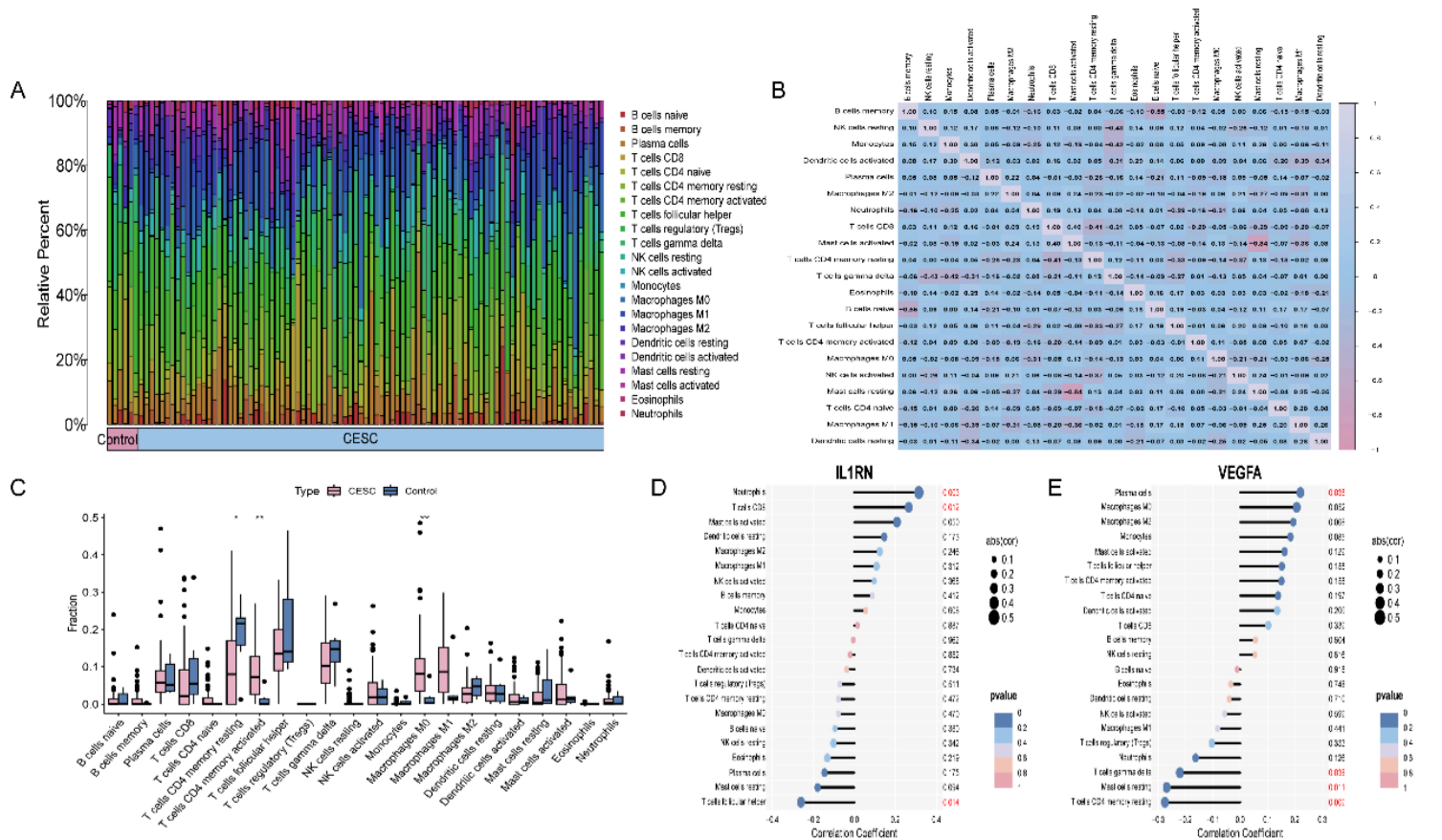


Figure 6

Analysis of the immune landscape associated with CESC using the CIBERSORT algorithm on the integrated dataset. (A) Relative infiltration proportions of 22 immune cell subsets in the control group (Control) and the cervical cancer group (CESC), as assessed by the CIBERSORT algorithm. (B) Heatmap of Spearman's correlation analysis among immune cell subsets; color intensity represents the magnitude of the correlation coefficient (red indicates a positive correlation, blue indicates a negative correlation). (C) Differences in the infiltration proportions of each immune cell subset between the CESC group and the Control group; pink represents the CESC group, and blue represents the Control group. (D) Spearman correlation lollipop plot showing the relationship between IL1RN expression and the infiltration proportions of each immune cell subset. (E) Spearman correlation lollipop plot showing the relationship between VEGFA expression and the infiltration proportions of each immune cell subset.

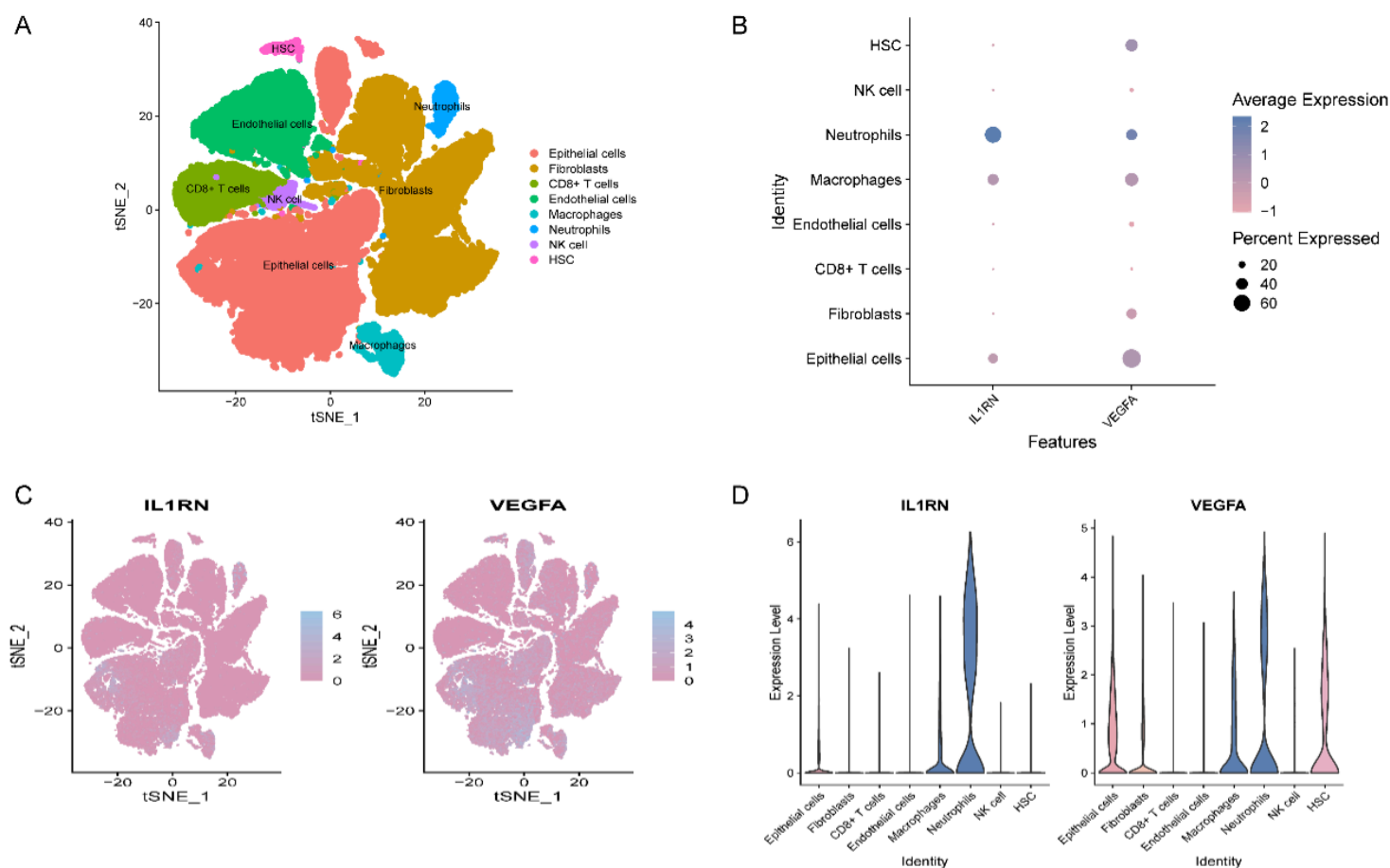


Figure 7

Single-cell transcriptomics reveals cell-specific expression patterns of key signature genes in CECs.

(A) t-SNE dimension-reduction clustering plot. (B) Dot plot showing the expression of IL1RN and VEGFA in different cell subsets. Dot size represents the proportion of cells expressing the gene, and color intensity represents the average expression level. (C) t-SNE feature plot showing the expression distribution of IL1RN and VEGFA in the dimension-reduced space. (D) Violin plot showing the distribution characteristics of IL1RN and VEGFA expression levels across cell subsets.

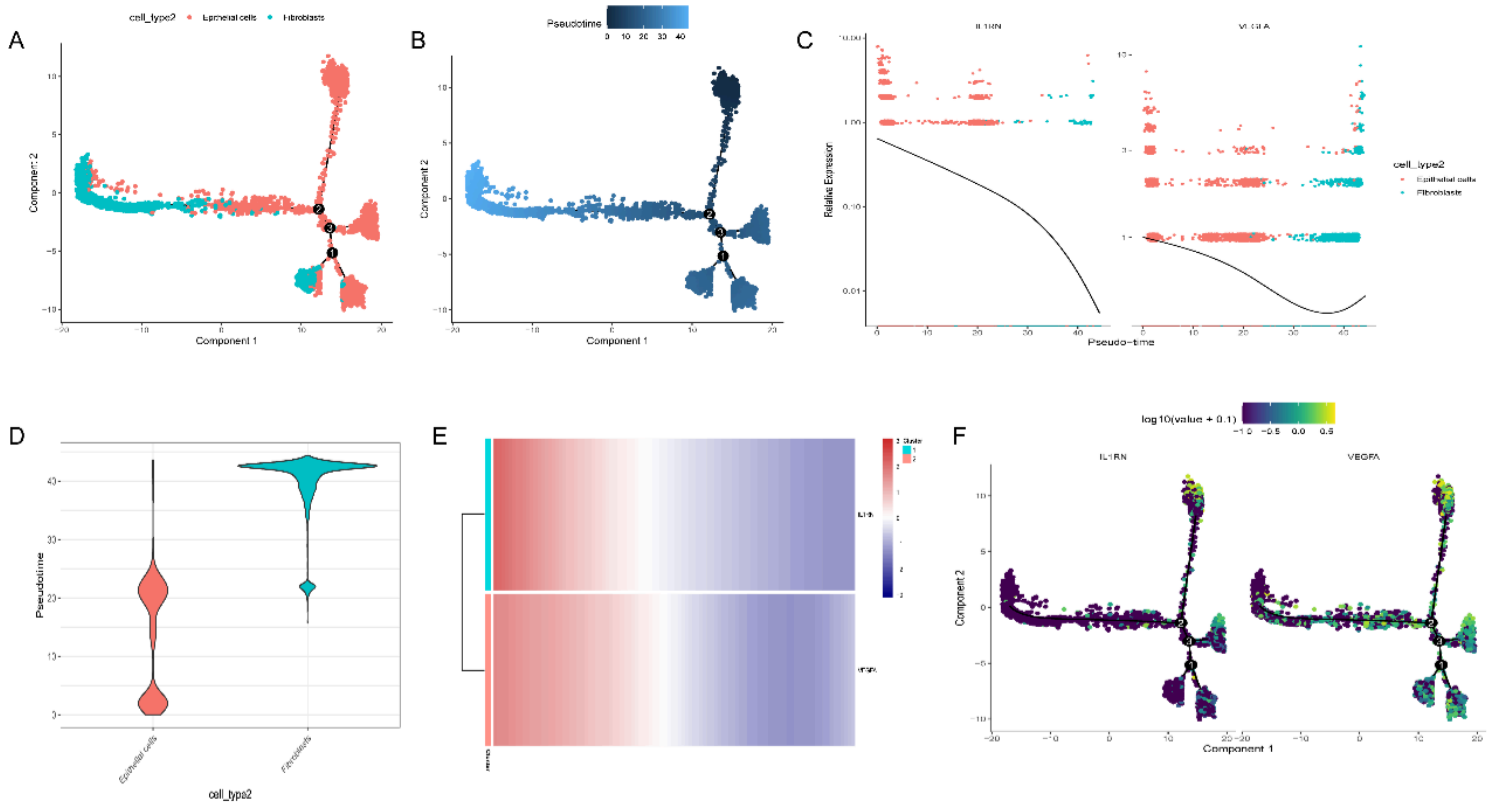


Figure 8

Construction of pseudotime trajectories for epithelial cells and fibroblasts, and the dynamic expression of the signature genes IL1RN and VEGFA.

(A) Cell differentiation trajectory map constructed using the Monocle algorithm, colored by cell type (red represents epithelial cells; blue-green represents fibroblasts), showing the spatial distribution of the two cell types along the trajectories. (B) Cell trajectories colored by pseudotime, with shades ranging from dark blue to light blue indicating a gradual increase in pseudotime values; black dots mark key branching nodes determining cell fate. (C) Fitted curves showing the relative expression levels of the key signature genes IL1RN and VEGFA as a function of pseudotime, with colors corresponding to different cell types, revealing nonlinear dynamic changes in gene expression over developmental time. (D) Violin plots showing the distribution of pseudotime values for epithelial cells and fibroblasts. (E) Heatmaps of the expression of key signature genes IL1RN and VEGFA as a function of pseudotime. (F) Feature plots of IL1RN and VEGFA along pseudotime trajectories.

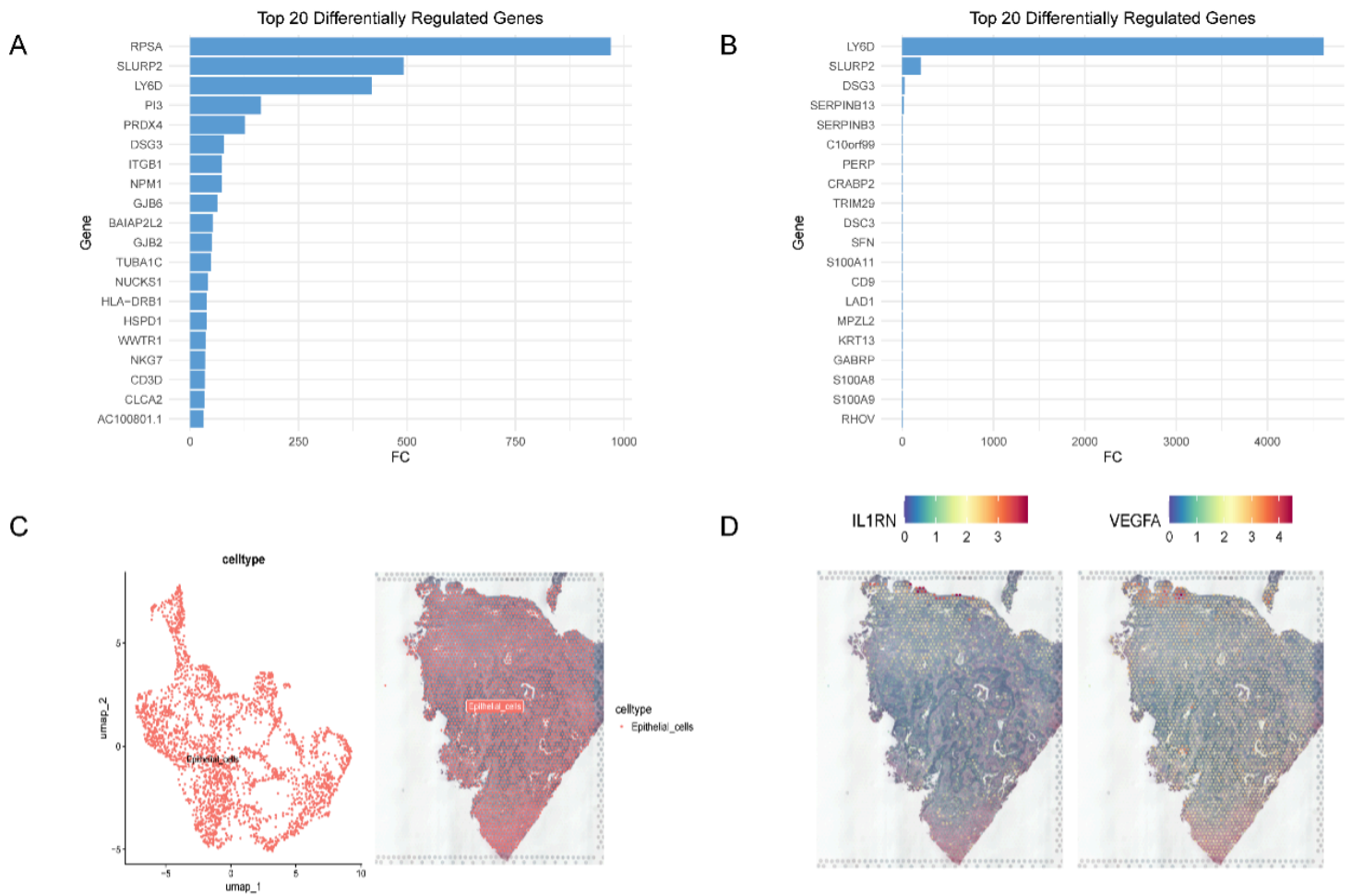


Figure 9

Regulatory Networks and Spatial Localization of Key Signature Genes

(A–B) Bar charts showing the top 20 genes ranked by fold change (FC) after virtual gene knockout in tumor single-cell data using the scTenifoldKnn algorithm. (C) The left panel shows a UMAP-based dimensionality-reduction clustering map, and the right panel shows cell type annotations on tissue sections, where red dots represent spatial clusters of epithelial cells within the tissue. (D) Spatial expression patterns of IL1RN and VEGF in cervical cancer tissue sections.

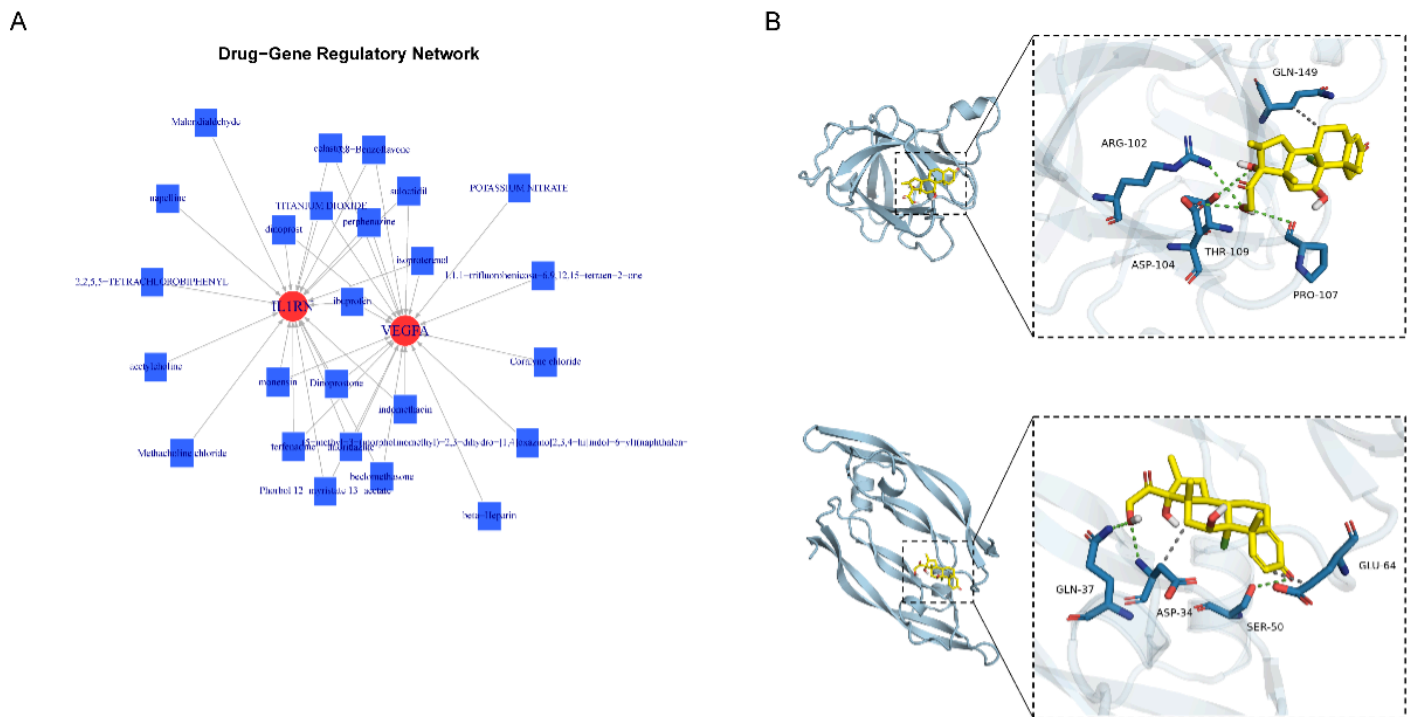


Figure 10

Analysis of the Interaction Patterns Between Drug-Gene Regulatory Networks and Molecular Docking

(A) A drug-gene regulatory network diagram constructed based on the DSigDB database. Red circular nodes represent intersecting signature genes (IL1RN and VEGFA), while blue rectangular nodes represent significantly enriched candidate drugs ($p.adjust < 0.05$); the connecting lines indicate interaction relationships between drugs and genes. (B) Molecular docking binding mode diagrams of beclomethasone with target proteins. The top figure shows the 3D binding mode of beclomethasone with IL1RN (PDB ID: 1ILR), and the bottom figure shows the 3D binding mode of beclomethasone with VEGFA (PDB ID: 2VPF). The yellow stick model represents the small molecule, the blue cartoon model represents the protein backbone, the green dashed lines indicate hydrogen bonds, and the gray dashed lines indicate hydrophobic interactions.