

Single-Cell–Derived Malignant Epithelial Programs Define Prognostic Risk and Therapeutic Vulnerability in Ovarian Cancer

Jingwen Si^{1†}, Hanbo Li^{1†}, Han Zhang^{2†}, Yan Shen^{1✉}

1. Department of Pathology, Tianjin Central Hospital of Gynecology Obstetrics, Tianjin, China.
2. Tianjin Chest Hospital, Tianjin University, Tianjin, China.

† These authors contributed equally to this work and share first authorship.

✉ Corresponding author: Yan Shen (Email: serina_shen@163.com).

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <https://ivyspring.com/terms> for full terms and conditions.

Received: 2026.04.02; Accepted: 2026.06.11; Published: 2026.07.03

Abstract

Background: Ovarian cancer (OV) is characterized by pronounced intratumoral heterogeneity and unfavorable clinical outcomes, yet clinically applicable biomarkers for risk stratification remain limited. Although bulk transcriptomic models have been widely proposed, they often overlook the cellular origins of prognostic signals. Advances in single-cell RNA sequencing (scRNA-seq) provide an opportunity to resolve malignant epithelial diversity and identify biologically meaningful predictors.

Methods: Single-cell transcriptomic datasets were integrated to construct a comprehensive epithelial landscape of OV. Copy number variation profiles inferred by inferCNV were used to distinguish malignant epithelial cells, followed by subpopulation identification and trajectory inference using Slingshot and Monocle3. Subcluster-specific gene signatures were projected onto bulk RNA-seq cohorts, and prognostic genes were screened using Cox regression and LASSO modeling to establish a multigene risk score. The model was validated across multiple independent cohorts. Multi-omics analyses, including pathway enrichment, mutational profiling, tumor mutation burden, immune features, and drug sensitivity prediction, were performed to explore biological and clinical relevance. Functional assays were conducted to validate the role of the key gene PSMB1.

Results: Malignant epithelial cells exhibited substantial transcriptional heterogeneity and distinct evolutionary trajectories, revealing subpopulations associated with differential clinical outcomes. A nine-gene risk model derived from these subpopulations demonstrated robust and consistent prognostic performance across multiple datasets. The risk score was closely associated with tumor progression pathways, immune suppression, genomic instability, and reduced therapeutic sensitivity. Notably, PSMB1 was identified as a critical oncogenic factor, and its knockdown significantly inhibited proliferation and colony formation in OV cells.

Conclusions: This study establishes a single-cell–informed prognostic framework that links malignant epithelial heterogeneity to clinical outcomes and therapeutic response in OV. The proposed model provides biologically interpretable risk stratification and highlights PSMB1 as a potential therapeutic target, offering new insights for precision oncology.

Keywords: OV, scRNA-seq, Epithelial cells, inferCNV, immunotherapy, PSMB1.

1. Introduction

OV is a highly fatal malignancy of the female reproductive system, noted for its significant heterogeneity and subtle early symptoms [1]. As a result, the majority of patients are diagnosed at advanced stages [2]. Despite advancements in surgery,

chemotherapy, targeted therapy, and immunotherapy, the overall prognosis remains poor, with a low five-year survival rate [3]. Particularly under homogenized treatment strategies, the need for precise risk stratification and individualized therapeutic

guidance has become a pressing clinical challenge [4, 5].

Intratumoral heterogeneity in OV encompasses molecular and cellular variations, not just histological differences. Traditional bulk RNA sequencing technologies, which capture averaged signals from mixed populations, are limited in their ability to resolve the specific roles of distinct cellular subsets during tumor initiation and progression. scRNA-seq has recently transformed cancer research by allowing detailed analysis of intratumoral cellular heterogeneity with exceptional resolution [6]. Unlike bulk transcriptomic approaches, which obscure the contributions of individual cell types, scRNA-seq allows researchers to distinguish and characterize distinct cellular populations within the tumor microenvironment [7, 8]. In OV, where epithelial cells constitute a major component of the tumor mass, this technology has unveiled remarkable transcriptional diversity among epithelial subpopulations. Emerging evidence indicates that these subsets differ not only in gene expression patterns, but also in functional attributes such as proliferative potential, differentiation status, metabolic activity, and immune modulatory capacity [9, 10]. Notably, some epithelial clusters exhibit stem-like phenotypes, enhanced survival capabilities, or resistance to cytotoxic insults, suggesting their potential role as key drivers of tumor progression, metastasis, and therapeutic failure. Integrating single-cell data into prognostic models and therapeutic decisions is essential to accurately reflect the biological complexity of OV. Recent single-cell and multi-omics studies have further shown that malignant tumor-cell subpopulations can reflect distinct biological programs and prognostic features, supporting the value of linking cellular heterogeneity to patient-level risk stratification [11].

While several prognostic models for OV have been proposed based on bulk transcriptomic data, most lack insights into the biological origins of their constituent genes, failing to account for the heterogeneity at the cellular level [7, 12, 13]. This limits both their robustness and interpretability. This study seeks to combine scRNA-seq and bulk RNA-seq data to pinpoint crucial genes from malignant epithelial subpopulations, and develop a prognostic scoring system informed by biological data using machine learning methods. Furthermore, we systematically evaluate the association of this signature with immune contexture, mutational landscape, pathway activity, and immunotherapy responsiveness across multiple omics layers, thereby enhancing its clinical applicability and translational potential.

We developed a detailed epithelial atlas of OV by combining single-cell and bulk transcriptomic data,

revealing subpopulations with different levels of malignancy. Subcluster-specific marker genes were subsequently used to derive a prognostic model with clear cellular origins through machine learning algorithms, and its predictive performance was validated across multiple independent cohorts. Multi-omics analyses further revealed that the model signature was closely associated with signaling pathway activity, tumor immune microenvironment, and therapeutic response. In addition, functional assays confirmed that the core model gene, PSMB1, plays a tumor-promoting role, offering mechanistic insights and a potential target for therapeutic intervention.

2. Methods

2.1 Summary of Public Databases Employed in the Study

The single-cell transcriptomic datasets analyzed in this study were curated from the GEO database, specifically datasets GSE154600 and GSE184880. GSE154600 includes five primary OV samples, while GSE184880 provides single-cell RNA profiles for five normal ovarian tissues and seven OV tumor specimens, enabling comparisons between malignant and non-malignant states. Bulk RNA sequencing data were obtained from both TCGA and GEO repositories. The TCGA-OV cohort was employed as the primary training set for model construction. Three independent datasets—GSE13876, GSE140082, and GSE26712—were selected from GEO to serve as external validation sets, enabling robust assessment of model performance across platforms. Before model validation and signature projection, bulk expression matrices from different cohorts were log₂-transformed when necessary, matched by gene symbols, and further corrected for cross-dataset batch effects using the ComBat function in the sva R package, with cohort or platform information used as the batch variable. Clinical metadata and somatic mutation annotations (MAF files) for TCGA-OV were obtained from the Genomic Data Commons (GDC) portal using the TCGAbiolinks R package [14]. To examine the model's cross-cancer applicability, pan-cancer expression profiles and corresponding clinical annotations were downloaded from the UCSC Xena platform. TIDE scores, used to predict immunotherapy responses, were obtained from the official TIDE web interface (<http://tide.dfci.harvard.edu/>).

2.2 Single-Cell Processing Pipeline and Cross-Sample Harmonization

scRNA-seq data were processed following a standard pipeline using the Seurat framework [15, 16].

Prior to downstream analyses, rigorous quality control was applied based on the number of transcripts, the number of expressed genes, and the proportion of mitochondrial transcripts per cell. Cells that did not meet the defined thresholds were removed; filtering criteria are provided in Supplementary Table 1. After quality filtering, gene expression values were normalized, and genes with high variability across cells were retained for further analysis. The dataset was then scaled, and potential confounders such as mitochondrial gene expression were accounted for during preprocessing. To mitigate technical variability across samples, the Harmony algorithm was employed for batch effect correction. This integration approach allowed cells from different datasets to align within a shared low-dimensional representation. Clustering was performed in the principal component space, and cellular heterogeneity was visualized using a two-dimensional embedding. Cell identities were annotated based on the expression of canonical lineage markers, supplemented by manual verification using reference databases and published literature. Visualization of cell clusters and gene expression patterns was enhanced using the SCP package [17], ensuring high-resolution graphical outputs.

2.3 Mapping Intercellular Communication Networks between Normal and Tumor Tissues

To uncover intercellular communication differences between OV tissues and normal ovarian samples, we constructed signaling interaction networks using single-cell transcriptomic data. Cell-cell communication was deduced using known ligand-receptor interactions, with interaction strengths calculated separately for tumor and normal conditions. Enrichment analysis was conducted to identify signaling pathways exhibiting differential activity between the two tissue types. Subsequent analyses focused on epithelial cells, investigating their role as signaling initiators. We analyzed pathways that showed significant upregulation or downregulation in tumor-derived epithelial populations relative to normal counterparts. The analysis leveraged a high-confidence interaction database and accounted for both expression levels and interaction likelihood. Final outputs were summarized at the signaling pathway level, highlighting key communication patterns reshaped by the tumor microenvironment.

2.4 Genomic Instability Estimation via inferCNV in Epithelial Populations

We utilized the inferCNV package [18] in R to analyze large-scale copy number variations (CNVs) from single-cell transcriptomic data, aiming to study

genomic instability in epithelial cells. Epithelial cells derived from normal ovarian tissues were designated as the reference group. By comparing gene expression intensities along chromosomal positions, relative CNV patterns were generated to reflect potential genomic gains or losses. A dual-criteria approach was employed to identify malignant epithelial cells. Cells were deemed malignant if their average CNV value exceeded 0.2 or their Pearson correlation coefficient was greater than 0.2 compared to the top 5% of cells with the highest CNV signals. This approach enhances the robustness of malignant cell detection under conditions of intratumoral heterogeneity and provides insight into chromosomal alterations characteristic of tumorigenic epithelial subsets.

2.5 Pseudotime Trajectory Analysis Using Slingshot and Monocle3

To elucidate the potential dynamic transitions of malignant epithelial cells during tumor progression, we performed pseudotime trajectory reconstruction using two complementary algorithms: Slingshot [19] and Monocle3 [20]. Slingshot was first employed to infer lineage trajectories based on predefined clustering and low-dimensional embeddings. This approach fits continuous lineage structures across clusters and assigns pseudotime values to individual cells according to their spatial progression along the inferred paths. To validate the robustness of the trajectory and explore cellular transitions within a different computational framework, Monocle3 was applied to the same set of malignant cells. This method constructs a principal graph that models the developmental landscape and orders cells along trajectories by learning the global topology of transcriptional states. Both tools enabled the reconstruction of developmental hierarchies and provided a foundation for downstream branch-specific enrichment analysis.

2.6 Association Analysis Between Malignant Cell Subpopulations and Patient Prognosis

To investigate the potential impact of distinct malignant epithelial subpopulations on overall prognosis, we first identified cluster-specific marker genes based on the tumor cells previously inferred by inferCNV. Representative gene sets for each malignant cluster were identified through differential expression analysis. The gene signatures were then applied to the bulk RNA-seq data from the TCGA-OV cohort. For each patient, enrichment scores corresponding to each cluster-specific gene set were calculated using a gene set scoring algorithm. Patients were categorized into high- and low-score groups according to the median enrichment score. Kaplan-Meier survival curves and

log-rank tests were used to assess the prognostic value of each malignant subpopulation. This method facilitated the identification of clusters linked to adverse outcomes, offering potential insights for discovering therapeutic targets.

2.7 Translating Single-Cell Insights into a Multigene Risk Score

We examined the prognostic significance of specific malignant epithelial subpopulations in OV, concentrating on clusters 0, 1, and 3, previously linked to patient survival. Marker genes for each cluster were identified using Seurat's FindAllMarkers function, applying thresholds of $|\log_2FC| > 0.25$, adjusted p-value < 0.05 , and $\text{min.pct} > 0.25$. These marker genes were then mapped to the bulk RNA-seq expression data from the TCGA-OV cohort. A univariate Cox regression analysis was conducted to identify genes significantly associated with patient survival ($p < 0.05$). To construct a robust prognostic model, we applied the Least Absolute Shrinkage and Selection Operator (LASSO) regression using the glmnet R package. The optimal penalty parameter lambda was determined through tenfold cross-validation. Genes selected by LASSO were further refined using stepwise regression with the Akaike Information Criterion (stepAIC) to develop the final multigene Cox regression model.

The risk score for each patient was then calculated using the following formula:

$$\text{Risk score} = \sum (\text{Coef}_i \times \text{Exp}_i)$$

where Coef_i is the regression coefficient and Exp_i is the expression level of gene i .

Patients were categorized into high- and low-risk groups according to the median risk score. Kaplan-Meier survival analysis and log-rank tests were then performed to evaluate the prognostic performance of the model.

2.8 Gene Set Enrichment Analysis Highlights Molecular Distinctions Between Risk Groups

We conducted Gene Set Variation Analysis (GSVA) on all samples to evaluate hallmark pathway enrichment levels [21], utilizing gene sets from the MSigDB H collection to investigate functional differences between distinct risk groups. Patients were categorized into high- and low-risk groups according to the median risk score. Genes with differential expression were identified using criteria of $|\log_2 \text{fold change}| > 1$ and $\text{FDR} < 0.05$. GO Biological Process and KEGG pathway enrichment analyses were performed to determine functional patterns specific to each group. Heatmaps and bubble plots were utilized to visualize the enriched biological functions and

signaling pathways.

2.9 Integrated Evaluation of Risk Model and Tumor Mutation Burden

We conducted a detailed analysis of the mutational landscape using the TCGA-OV cohort to explore the relationship between risk scores and somatic mutation features. MAF files were processed and visualized using the Maftools R package [22] to analyze the global mutation profile, covering mutation frequency, variant classifications, types, and SNV classes. Subsequently, we extracted mutation data for the prognostic model-associated genes across all patients and illustrated their distribution using waterfall plots to assess the mutational frequency and potential biological relevance of each gene. Patients were categorized into high- and low-risk groups based on the median risk score, and their mutational landscapes were separately plotted to compare mutational spectra. TMB, defined as nonsynonymous mutations per megabase, was compared across risk subgroups. A correlation analysis was performed to assess the relationship between TMB and the risk score. Patients were stratified by TMB and risk score, and Kaplan-Meier survival curves with log-rank tests evaluated the prognostic impact of their combined effect. This multi-layered analysis provides additional support for the interpretability and clinical relevance of the proposed prognostic model.

2.10 Pan-Cancer Landscape of Risk Score Correlation with Tumor Phenotypic Programs

We conducted a pan-cancer analysis using TCGA transcriptomic data from 33 solid tumors, including OV, to thoroughly assess the relationship between the constructed risk score and tumor phenotype-related pathways across different cancer types. The risk score for each sample was calculated as previously described. We utilized hallmark gene sets from the MSigDB database to quantify key tumor biological processes, specifically focusing on angiogenesis and epithelial-mesenchymal transition (EMT). GSVA was used to calculate pathway enrichment scores for each sample. Pearson correlation analysis was conducted to evaluate the associations between risk scores and pathway activities across various tumor types. The results were visualized using scatter plots, with Pearson coefficients and p-values annotated to highlight tumor-type-specific associations.

2.11 Cell Lines and Cell Culture

The human OV cell lines SKOV3 and ES-2 were obtained from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China). All cell lines were

authenticated by short tandem repeat (STR) profiling and confirmed to be free of mycoplasma contamination prior to experimentation. SKOV3 and ES-2 cells were cultured in RPMI-1640 medium (Gibco, USA) supplemented with 10% fetal bovine serum (FBS) (Gibco, USA), 100 U/mL penicillin, and 100 µg/mL streptomycin, and maintained in a humidified incubator at 37 °C with 5% CO₂. Cells were passaged at 70–80% confluence using 0.25% trypsin-EDTA and were used within 10 passages for all experiments to ensure stability.

2.12 Cell Proliferation Assay (CCK-8)

To assess the impact of PSMB1 on OV cell proliferation, the Cell Counting Kit-8 (CCK-8, Beyotime, China) assay was conducted. SKOV3 and ES-2 cells transfected with siRNA were seeded into 96-well plates at a density of 5×10^3 cells per well, with three technical replicates per group. From Day 1 to Day 7 post-transfection, 10 µL of CCK-8 reagent was added daily to each well, followed by a 2-hour incubation. Absorbance at 450 nm was measured using a microplate reader. Each experiment was performed in triplicate, and the mean OD values were used to generate cell growth curves over the 7-day period.

2.13 Colony Formation Assay

To further investigate the effect of PSMB1 on the proliferative and survival capacities of OV cells, a colony formation assay was performed. SKOV3 and ES-2 cells transfected with siRNA were seeded into 6-well plates at a density of 500 cells per well. Both control and siRNA groups were included, with three replicates each. Cells were cultured for approximately 10–14 days until visible colonies formed. The plates were gently washed with PBS, fixed in 4% paraformaldehyde for 15 minutes, and stained with 0.1% crystal violet for 30 minutes. After removing excess dye, images were captured, and colonies larger than 50 µm were counted using ImageJ software. Colony formation rates were calculated and statistically analyzed.

3. Results

3.1 Single-cell Transcriptomic Profiling of OV Samples, Including Data Preprocessing, Batch Effect Mitigation, and Cell Type Classification

We initiated our analysis by performing stringent quality control on the scRNA-seq data derived from OV samples. As illustrated in Figure 1A, the unfiltered dataset included cells with low gene counts, abnormal transcript levels, or high mitochondrial content. Applying filtering criteria centralized and

standardized the distribution of nFeature_RNA, nCount_RNA, and percent.mt across samples (Figure 1B), improving data reliability for subsequent analyses. We utilized the Harmony algorithm to correct potential batch effects among samples. Prior to adjustment, cells from different samples formed distinct clusters in the low-dimensional space (Figure 1C). After integration, cells were well mixed across batches, demonstrating successful elimination of technical variation (Figure 1D). Given the confounding influence of cell cycle genes on dimensionality reduction, we regressed out these effects using Seurat's normalization pipeline. Before correction, cells at different phases exhibited biased distributions along principal components (Figure 1E). Post-regression, the impact of cell cycle was substantially mitigated, yielding a more uniform representation (Figure 1F).

Following preprocessing and integration, a total of 97,875 high-quality cells were retained for further analysis. Unsupervised clustering and UMAP projection revealed 36 distinct cellular subpopulations (Figure 1G), highlighting the extensive heterogeneity within the tumor ecosystem. Using canonical markers and reference-based classification, we identified these clusters as major cell lineages: T cells, B cells, epithelial cells, macrophages, dendritic cells, fibroblasts, and endothelial cells (Figure 1H). The spatial mapping of tissue origin indicated that the majority of cells were derived from tumor tissues rather than adjacent normal regions (Figure 1I), suggesting that the tumor microenvironment dominates the cellular composition landscape. We analyzed the expression profiles of key marker genes, such as CD3D, NKG7, EPCAM, and MKI67, which exhibited unique distribution patterns among the identified cell types (Figure 1J), supporting the accuracy of the classification. Finally, we measured the abundance of each cell type across different patients and tissue sources (Figure 1K). T cells exhibited the highest overall abundance and showed marked variability among patients. Tumor tissues were enriched with T cells, B cells, and epithelial cells, while fibroblasts and endothelial cells were comparatively sparse. These compositional shifts underscore the complex and heterogeneous immune landscape of OV.

3.2 Cellular Communication Heterogeneity Reveals Immune Microenvironment Reprogramming in Ovarian Cancer

We systematically compared signaling pathway activity between ovarian tumor and normal tissues to identify differences in cell-cell communication (Figure 2A). The tumor samples exhibited enhanced signaling via SPP1, FN1, MHC-I, TIGIT, and COMPLEMENT

pathways, which are associated with immunosuppression, extracellular matrix remodeling, and antigen presentation. In contrast, CD45, LAMININ, VCAM, and VEGF pathways were more

prominent in normal tissues, indicating their involvement in immune surveillance and basement membrane maintenance.

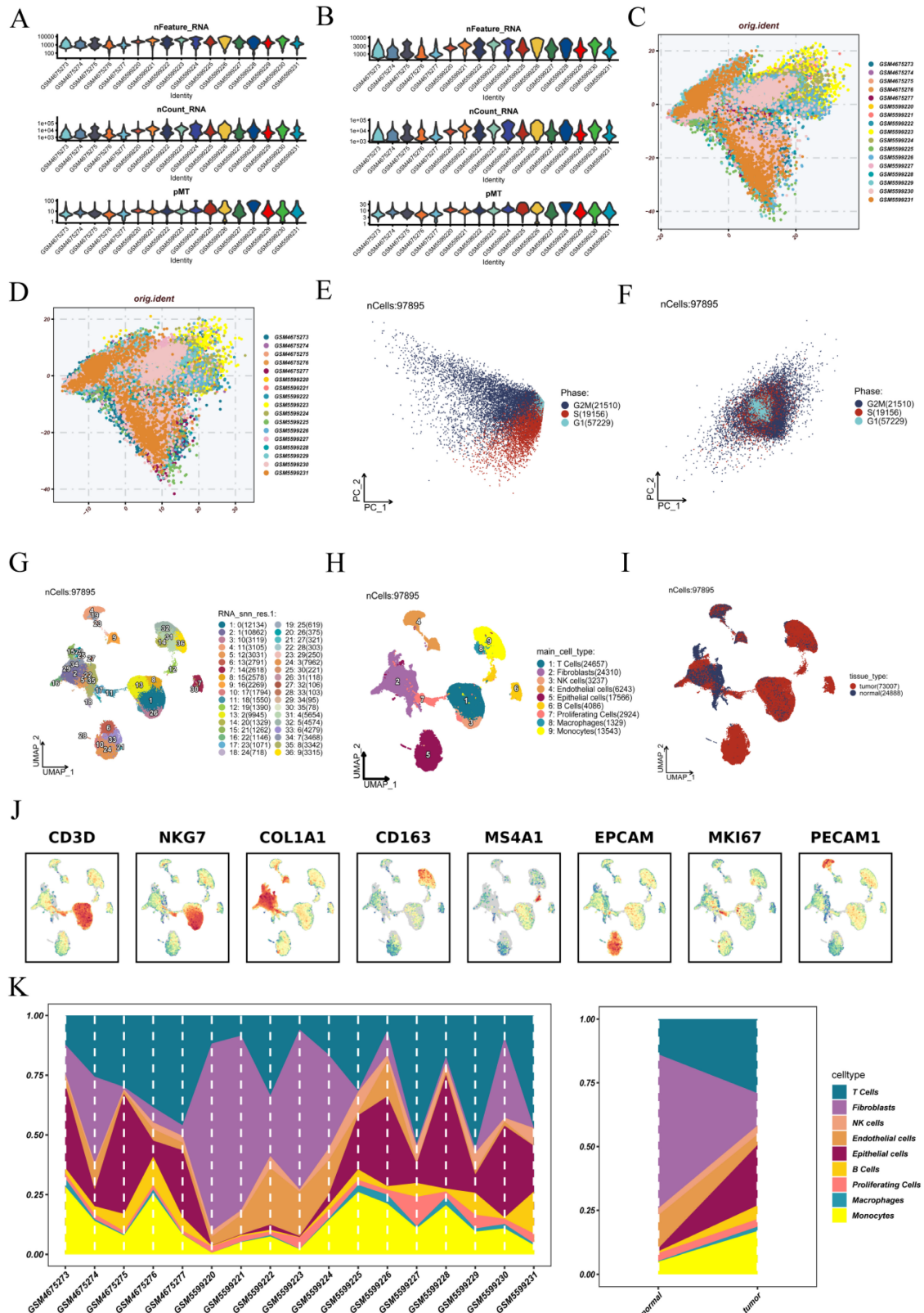


Figure 1. Single-cell RNA sequencing data undergoes quality control, batch correction, and cell-type annotation. (A–B) Distribution of gene counts (nFeature_RNA), transcript counts (nCount_RNA), and mitochondrial gene percentages (percent.mt) across all samples before and after quality control. (C–D) Visualization

Focusing on epithelial cells as signal senders, we observed that multiple outgoing interactions were markedly elevated in tumor tissues (Figure 2B). The interactions MDK–NCL, APP–CD74, MIF–(CD74+CD44), MIF–(CD74+CXCR4), and HLA-E–CD94–NKG2C are mainly associated with immune regulation and antigen recognition. Conversely, pathways related to cell adhesion—such as COL4A1–CD44, COL4A2–CD44, and CD99–CD99—were notably downregulated in tumors (Figure 2C), suggesting impaired epithelial adhesion and a potential shift toward enhanced migratory or invasive behavior.

3.3 Dissecting Epithelial Cell Heterogeneity Through CNV-Based Malignancy Profiling

To further investigate the heterogeneity and malignant characteristics of epithelial cells, we performed reclustering analysis on the epithelial population. UMAP-based dimensionality reduction revealed 20 transcriptionally distinct subclusters, indicating substantial intragroup heterogeneity (Figure 3A). Tissue origin annotation showed that the majority of epithelial cells were derived from tumor samples, with a higher abundance compared to those from normal tissues (Figure 3B). We utilized normal tissue-derived epithelial cells as a reference to apply inferCNV for inferring large-scale chromosomal alterations at single-cell resolution. The results revealed widespread copy number gains and losses in tumor-derived cells, suggesting the presence of genomic instability (Figure 3C). Cells exhibiting either a mean CNV value greater than 0.2 or a Pearson correlation exceeding 0.2 with the top 5% high-CNV cells were classified as malignant epithelial cells (Figure 3D). Further reclustering of these malignant cells revealed additional transcriptional diversity within the tumor epithelial compartment (Figure 3E). Functional enrichment analysis demonstrated that specific subpopulations exhibited distinct pathway preferences. In particular, subcluster 0 was strongly associated with metabolic and stress-related pathways, including GLYCOLYSIS, MTORC1_SIGNALING, and REACTIVE_OXYGEN_SPECIES_PATHWAY, suggesting a role in maintaining tumor viability and potential therapy resistance (Figure 3F).

3.4 Developmental Trajectories Reveal Divergent Evolution of Malignant Epithelial Subpopulations

To delineate the developmental trajectories of

malignant epithelial cells, we employed the Slingshot algorithm for pseudotime inference. As shown in Figure 4A, two distinct lineages were identified, with Cluster 1 serving as the common root state, while Clusters 0 and 2 appeared at the terminal ends of the respective trajectories, suggesting divergent differentiation fates. This branching pattern implies that malignant cells may originate from a shared progenitor-like state and evolve along separate paths under tumor microenvironmental pressures. Figure 4B presents the pseudotime distribution of cells across the inferred trajectories, highlighting a gradual and continuous transition in cellular states.

Functional enrichment analysis comparing the two lineages is shown in Figure 4C. Genes enriched along Lineage 1 were primarily associated with the androgen receptor signaling pathway, cellular response to amino acid starvation, negative regulation of autophagy, and intracellular steroid hormone receptor signaling, suggesting an adaptive response to hormonal and nutrient signals. In contrast, Lineage 2 showed enrichment in pathways associated with cell migration, immune modulation, and extracellular matrix remodeling, suggestive of a more aggressive and potentially invasive phenotype.

To validate the robustness of the inferred trajectories, we further applied Monocle3 for pseudotime reconstruction. As shown in Figure 4D, cells were colored by their original clusters, while Figure 4E illustrates a continuous trajectory along the pseudotime axis, confirming the developmental transitions observed with Slingshot.

3.5 Association Analysis Between Malignant Cell Subpopulations and Patient Prognosis

We calculated enrichment scores for each patient in the TCGA-OV cohort using marker genes specific to each malignant epithelial subcluster, as illustrated in Supplement Figure 1. Patients were divided into high- and low-score groups based on the median score, followed by Kaplan–Meier survival analysis. The findings indicated a significant association between specific subclusters and patient prognosis. High enrichment in cluster 0 ($p = 0.012$) and cluster 2 ($p = 0.082$) correlated with poorer survival, while clusters 1 ($p = 0.02$), 3 ($p = 0.038$), and 4 ($p = 0.16$) were linked to better outcomes. Considering both statistical significance and survival trends, we selected marker genes from clusters 0, 1, and 3—those most strongly associated with prognosis—for subsequent model construction.

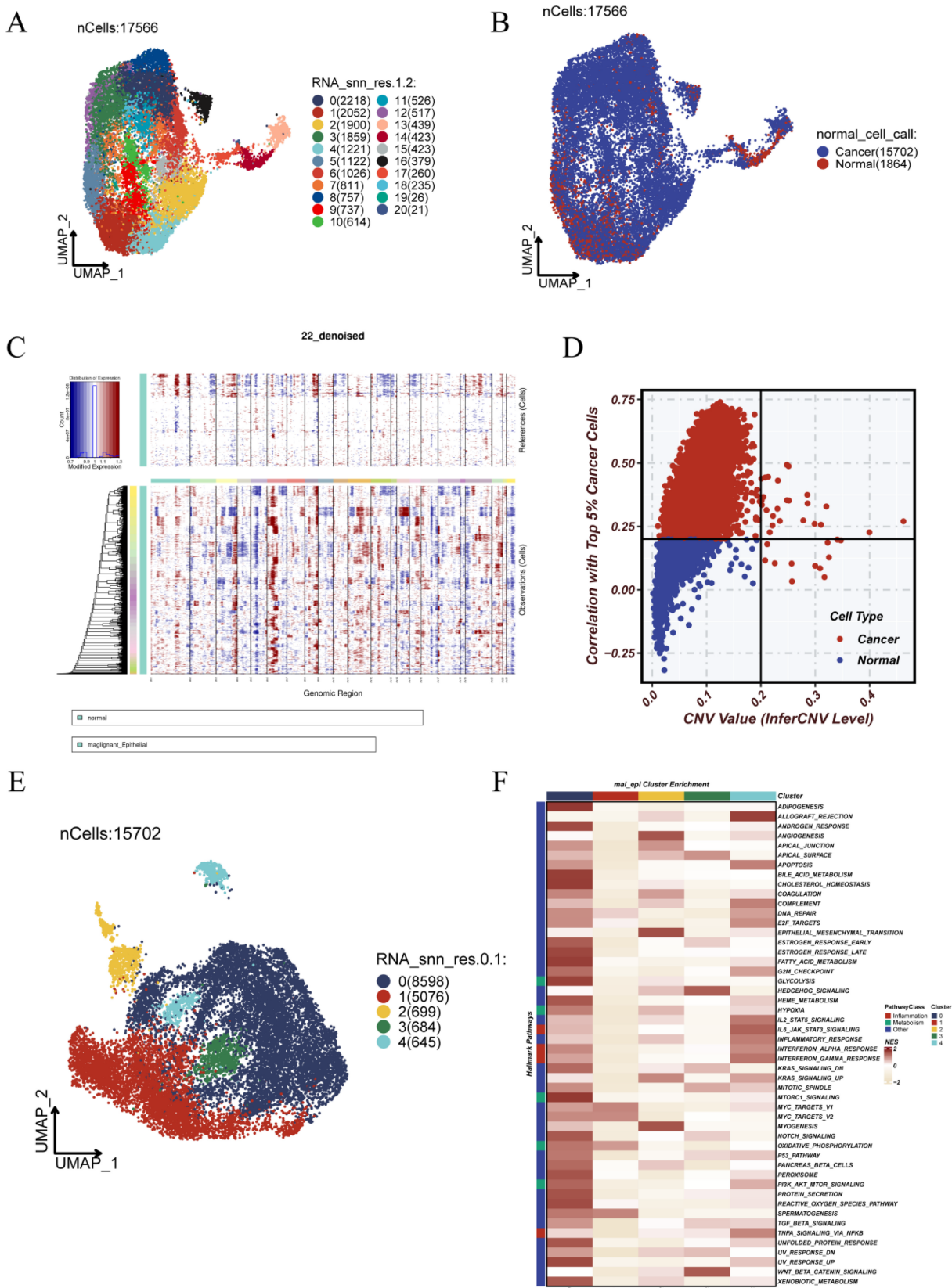


Figure 3. Dissecting Epithelial Cell Heterogeneity Through CNV-Based Malignancy Profiling. (A) UMAP-based reclustering of epithelial cells identified 20 heterogeneous subpopulations. (B) Epithelial cells were classified by tissue origin, with tumor-derived cells marked in red and normal tissue cells in blue. (C) Single-cell copy number variations (CNVs) were deduced using inferCNV. The heatmap displays genomic CNV patterns, with the upper panel representing reference (normal) cells and the lower panel showing predicted tumor cells. (D) Malignant epithelial cells were detected by analyzing CNV values and their association with the top 5% of cells exhibiting high CNV. Cells with CNV values > 0.2 or correlation > 0.2 were classified as malignant (red). (E) Reclustering of malignant epithelial cells revealed distinct substructures and intratumoral heterogeneity. (F) Heatmap of pathway enrichment across malignant subclusters, highlighting their functional divergence.

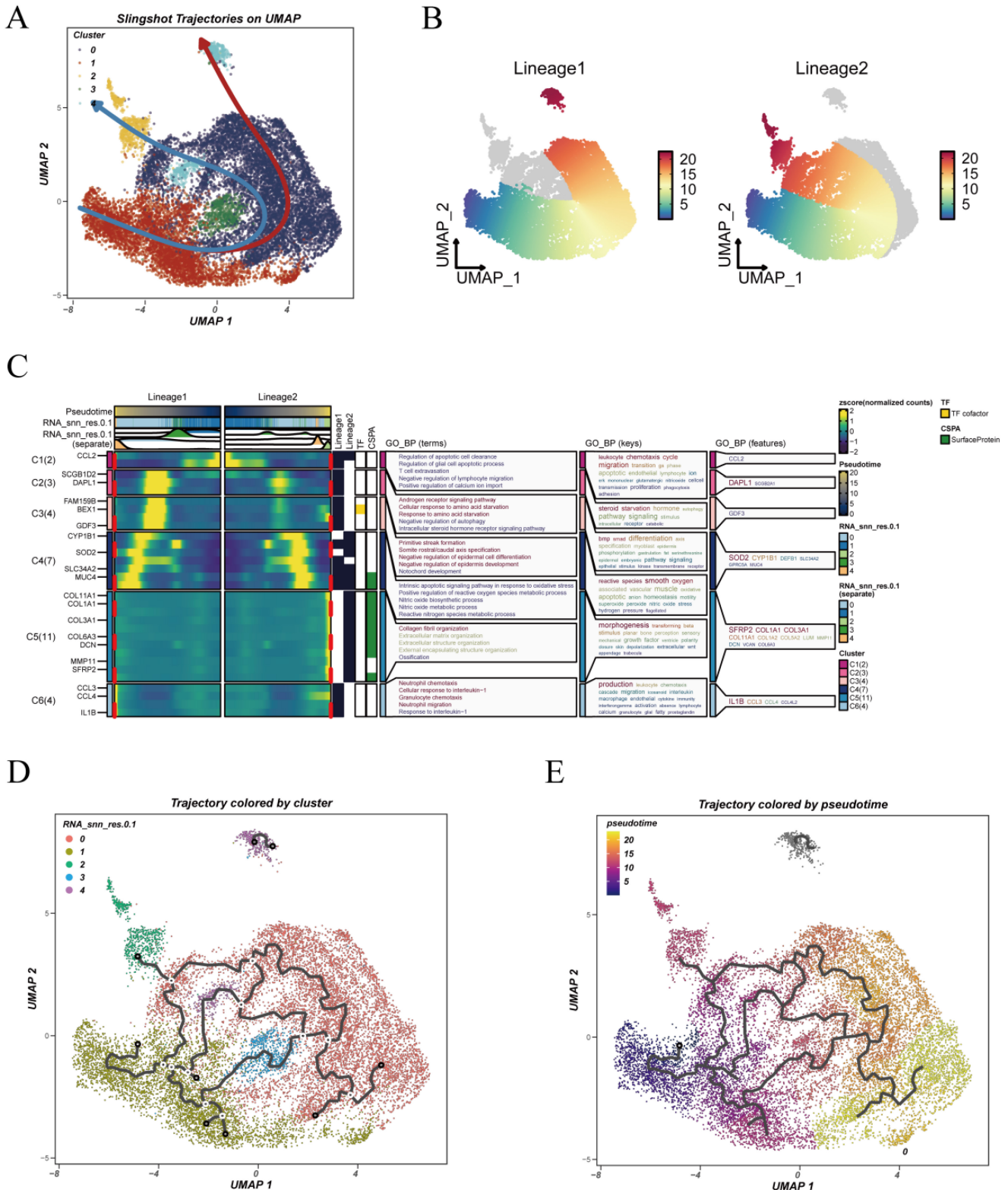


Figure 4. Pseudotime trajectory analysis reveals differentiation paths of malignant epithelial cells. (A) UMAP projection analyzed using the Slingshot algorithm reveals two primary developmental trajectories: Lineage 1 and Lineage 2. (B) Pseudotime progression of cells along each trajectory, with color gradients representing transitions from early to late pseudotime states. (C) GO enrichment analysis of subclusters along the two lineages, highlighting functional heterogeneity across biological processes, molecular functions, and cellular components. (D) Pseudotime trajectory constructed using Monocle3, with cells colored by their original clustering identities, illustrating the correspondence between trajectory and cellular states. (E) Monocle3-derived pseudotime projection colored by pseudotime values, demonstrating the directionality and continuity of the differentiation process.

3.6 From Single Cells to Signatures: Building a Prognostic Model for OV

To develop a prognostic signature based on malignant epithelial subpopulations, we first identified cluster-specific marker genes from clusters 0, 1, and 3—each showing significant correlation with patient survival in Supplement Figure 1. Figure 5A illustrates the use of univariate Cox regression to assess the prognostic significance of these genes, highlighting risk factors in red and protective factors in blue. Subsequently, LASSO regression was applied to reduce feature redundancy, with ten-fold cross-validation determining the optimal penalty parameter. We selected the minimum value of the tuning parameter ($\lambda_{\min} = 0.0228$) to construct the model (Figure 5B–C).

The resulting multivariate Cox regression model incorporated nine genes with the following formula:

$$\text{Risk Score} = (-0.0038 \times \text{ASRGL1}) + (-0.0008 \times \text{CXCR4}) + (0.0025 \times \text{PPP1R15A}) + (0.0033 \times \text{PSMB1}) + (0.0057 \times \text{RAB20}) + (0.0011 \times \text{RAC1}) + (0.0002 \times \text{RPL12}) + (-0.0046 \times \text{SDF2L1}) + (-0.0125 \times \text{SNRPD1}).$$

The model coefficients are displayed in Figure 5D, highlighting the relative contribution of each gene. Positive coefficients (e.g., RAB20 and PSMB1) indicate risk-enhancing effects, whereas negative coefficients (e.g., SNRPD1 and SDF2L1) suggest protective roles. Figure 5E displays a heatmap of model gene expression patterns in patients, categorized into high- and low-risk groups according to the median risk score.

To evaluate the stability and generalizability of the constructed prognostic model across different OV cohorts, we conducted validation analyses in four independent datasets: TCGA-OV, GSE13876, GSE140082, and GSE26712. Kaplan-Meier survival analysis in the TCGA-OV cohort showed that patients in the high-risk group had significantly shorter overall survival than those in the low-risk group ($p < 0.0001$). The AUC values at 1, 3, and 5 years were 0.628, 0.683, and 0.707, respectively, indicating moderate but consistent predictive performance. In the GSE13876 cohort, the model also showed significant prognostic stratification ($p < 0.0001$), with AUC values of 0.629, 0.677, and 0.686 at 1, 3, and 5 years, respectively. In the GSE140082 cohort, despite the relatively shorter follow-up time, a significant survival difference was observed ($p = 0.0068$), with AUC values of 0.670 and 0.676 at 1 and 3 years. In the GSE26712 cohort, the model achieved statistically significant survival separation ($p = 0.02$), with AUC values of 0.577, 0.639, and 0.685. These results support the reproducibility of

the risk score across independent datasets, while indicating moderate rather than strong discriminative ability.

3.7 Gene Set Enrichment Analysis Highlights Molecular Distinctions Between Risk Groups

Building upon the previously established prognostic model, we next investigated the functional differences between risk subgroups using GSVA and enrichment analyses. GSVA results revealed increased activity in pathways linked to malignant progression, such as the P53 signaling pathway, apical surface organization, and Notch signaling, in the high-risk group. The low-risk group showed an enrichment of pathways associated with immune homeostasis and metabolic regulation (Figure 7A), indicating distinct functional states between the groups. Gene Ontology enrichment analysis indicated that the high-risk group was mainly linked to the negative regulation of hydrolase activity and proteolysis (Figure 7B). The low-risk group demonstrated notable enrichment in nucleosome assembly and protein–DNA complex assembly processes, which are integral to chromatin remodeling and maintaining nuclear structure (Figure 7C).

At the KEGG pathway level, the high-risk group showed increased activity in immune and signaling pathways, such as the calcium signaling pathway and cytokine–cytokine receptor interaction (Figure 7D). The low-risk group showed enrichment in DNA replication and homologous recombination pathways, indicating improved genome maintenance capacity (Figure 7E). In summary, the two risk subgroups exhibit substantial divergence in their molecular and functional profiles, reinforcing the biological interpretability and robustness of the constructed risk stratification system.

3.8 Integrated Evaluation of Risk Model and Tumor Mutation Burden

We investigated the mutational characteristics associated with the risk score in the TCGA-OV cohort (Figure 8A). Among the variant classifications, missense mutations were the most prevalent. Single nucleotide polymorphisms (SNPs) accounted for the majority of variant types, with C>T being the most common single nucleotide variant (SNV) class. The genes TP53, TTN, MUC16, and CSMD3 were among the most commonly mutated, with TP53 exhibiting the highest mutation frequency. The nine genes incorporated into the prognostic model displayed relatively low mutation frequencies, with the highest alteration rate not exceeding 2.88% (Figure 8B).

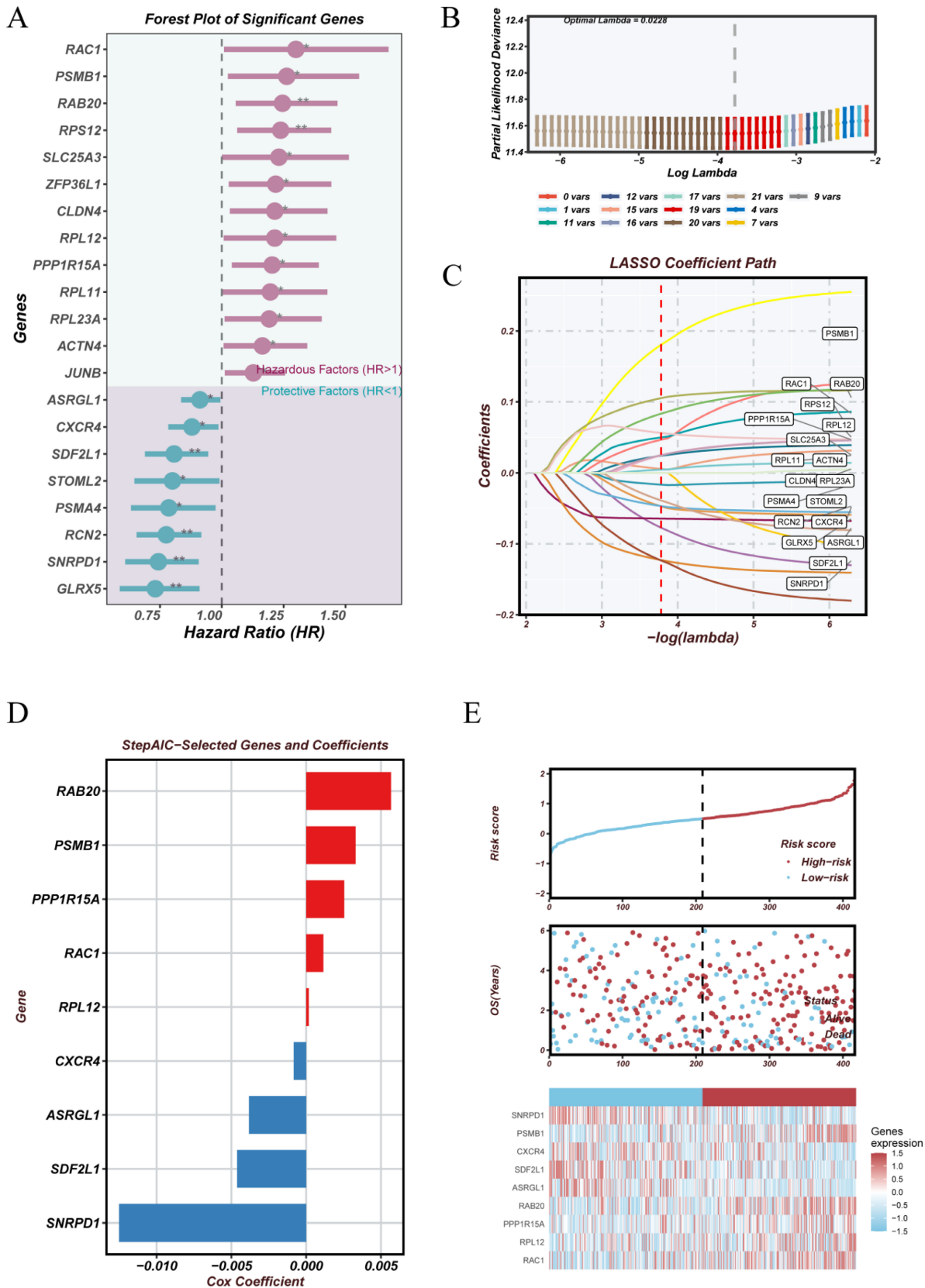


Figure 5. Construction of a malignant epithelial cell-derived prognostic risk model. (A)Conducted univariate Cox regression analysis on marker genes identified from clusters 0, 1, and 3. Genes significantly linked to prognosis are displayed, with high-risk factors (HR > 1) highlighted in red and protective factors (HR < 1) in blue. (B) Utilizing ten-fold cross-validation to select tuning parameters in LASSO regression. The optimal lambda value is selected to minimize partial likelihood deviance. (C) LASSO coefficient paths of selected genes across varying penalty parameters, illustrating the shrinkage behavior of model coefficients. (D) Final gene selection using the stepwise AIC algorithm applied to the LASSO results. Bar plot displays the Cox regression coefficients of the retained genes. (E) Risk score distribution and patient stratification. The top section displays risk scores ranked from low to high, the middle section illustrates the distribution of survival status, and the bottom section presents a heatmap of model gene expression profiles across high- and low-risk groups.

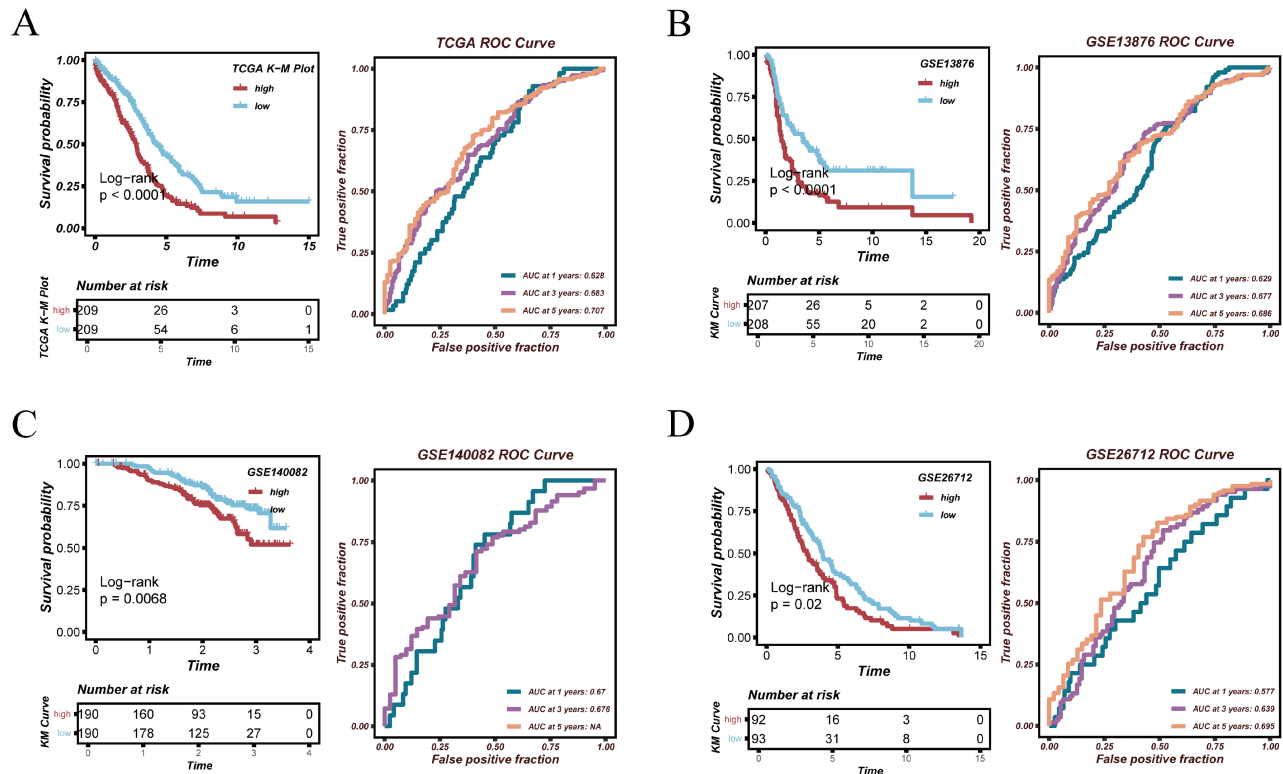


Figure 6. Validation of Prognostic Performance Across Multiple Cohorts. (A) Kaplan–Meier survival analysis and time-dependent ROC evaluation were conducted on the TCGA-OV cohort. (B) Survival analysis and ROC curves in the GSE13876 validation cohort. (C) Prognostic validation in the GSE140082 cohort was conducted using Kaplan–Meier analysis and receiver operating characteristic (ROC) evaluation. (D) External validation in the GSE26712 dataset, showing survival difference and predictive performance. Patients were categorized into high- and low-risk groups according to the median risk score. AUC values for 1-, 3-, and 5-year survival were calculated to assess the predictive accuracy of the model.

The analysis of somatic mutation profiles showed distinct patterns between high- and low-risk subgroups, with TP53 mutations being prominent in both (Figure 8C). The analysis revealed that TMB levels were significantly lower in the high-risk group than in the low-risk group (Figure 8D). A notable inverse relationship was identified between TMB and the risk score ($R = -0.19$, $p = 0.0013$; Figure 8E). Notably, survival analysis integrating both TMB and risk stratification demonstrated that patients with low risk and high TMB had the most favorable outcomes, whereas those with high risk and low TMB had the worst prognosis (Figure 8F). These findings highlight the potential of integrating TMB with the risk model to enhance prognostic accuracy.

3.9 Elevated Risk Scores Coupled with Enhanced Angiogenesis and EMT: Evidence from 33 Cancers

We conducted a pan-cancer correlation analysis to explore the biological significance of the model-derived risk score in relation to hallmark tumor phenotypes, focusing on angiogenesis and epithelial-mesenchymal transition (EMT). Figure 9A illustrates a positive correlation between the risk score and angiogenesis activity in various cancer types, such as OV, LUAD, BRCA, and PAAD. This indicates that

patients in the high-risk group might have increased neovascularization potential, which could facilitate tumor progression and metastasis. Figure 9B also shows a strong positive correlation between the risk score and EMT pathway activation in various cancers, including OV, PRAD, KIRC, and LUAD. The results suggest that individuals at high risk are more prone to exhibit aggressive phenotypes with enhanced metastatic potential. Overall, the results highlight that the identified risk signature is not only prognostically relevant but also reflects key malignant characteristics of tumor biology.

3.10 Integrating Immune Signatures and Drug Sensitivity Prediction to Assess Clinical Utility

Expanding on established risk stratification, we examined variations in immunotherapeutic response and drug sensitivity between high- and low-risk groups using immune checkpoint expression analysis, TIDE prediction, and chemotherapeutic sensitivity profiling. Figure 10A illustrates that immune checkpoint-related genes, including HAVCR2, LAIR1, TNFRSF14, TNFRSF25, and TNFSF9, are significantly upregulated in the high-risk group, suggesting an immunosuppressive or exhausted immune microenvironment.

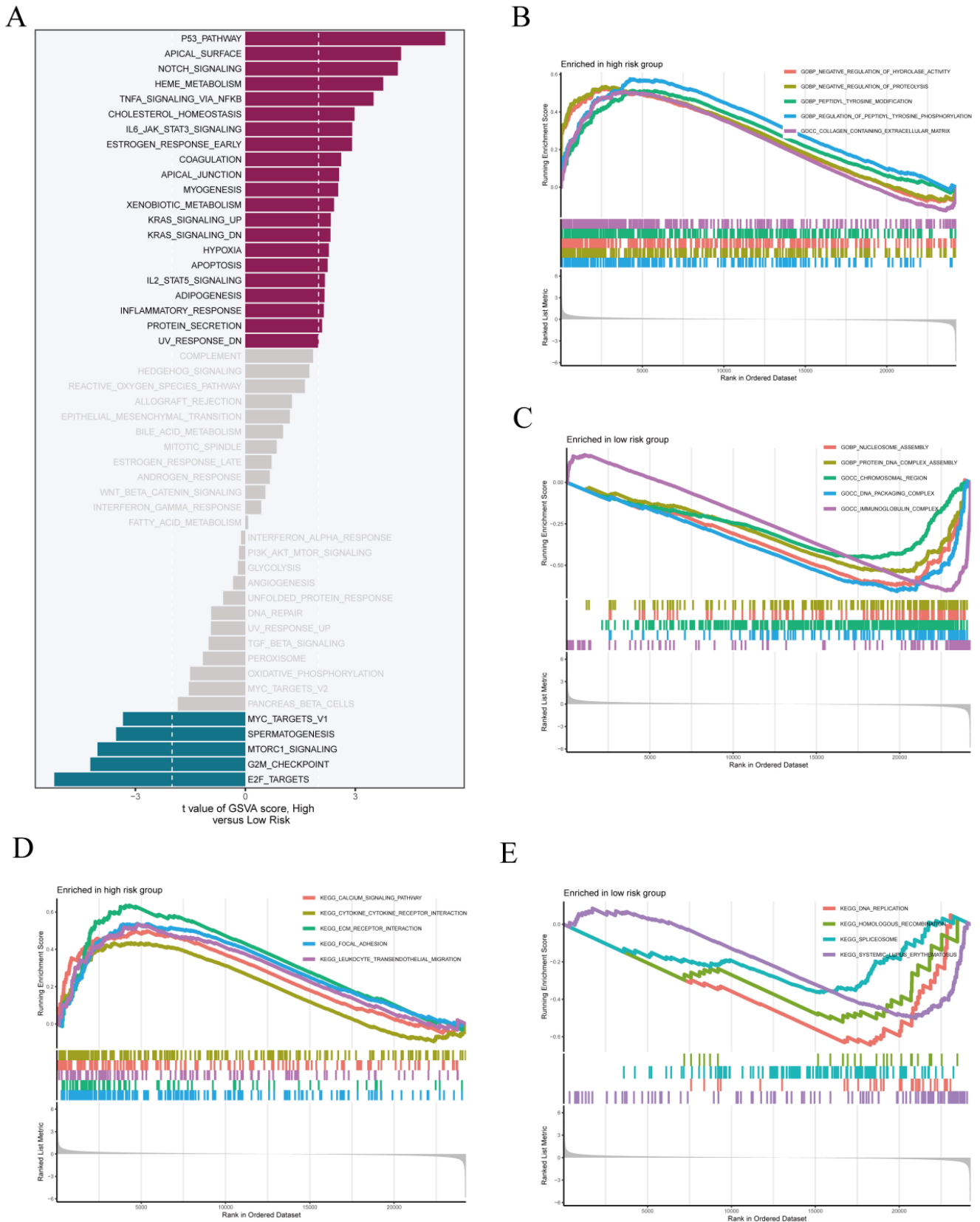


Figure 7. Functional enrichment analysis comparing high-risk and low-risk groups. (A) Gene Set Variation Analysis (GSVA) was performed to assess pathway activation differences between high- and low-risk groups, highlighting positively enriched pathways in red and negatively enriched ones in blue. Gene Ontology (GO) enrichment analysis identified key biological processes that were upregulated in both the (B) high-risk group and the (C) low-risk group. KEGG pathway analysis highlighted the significantly enriched pathways in both the high-risk (D) and low-risk (E) groups.

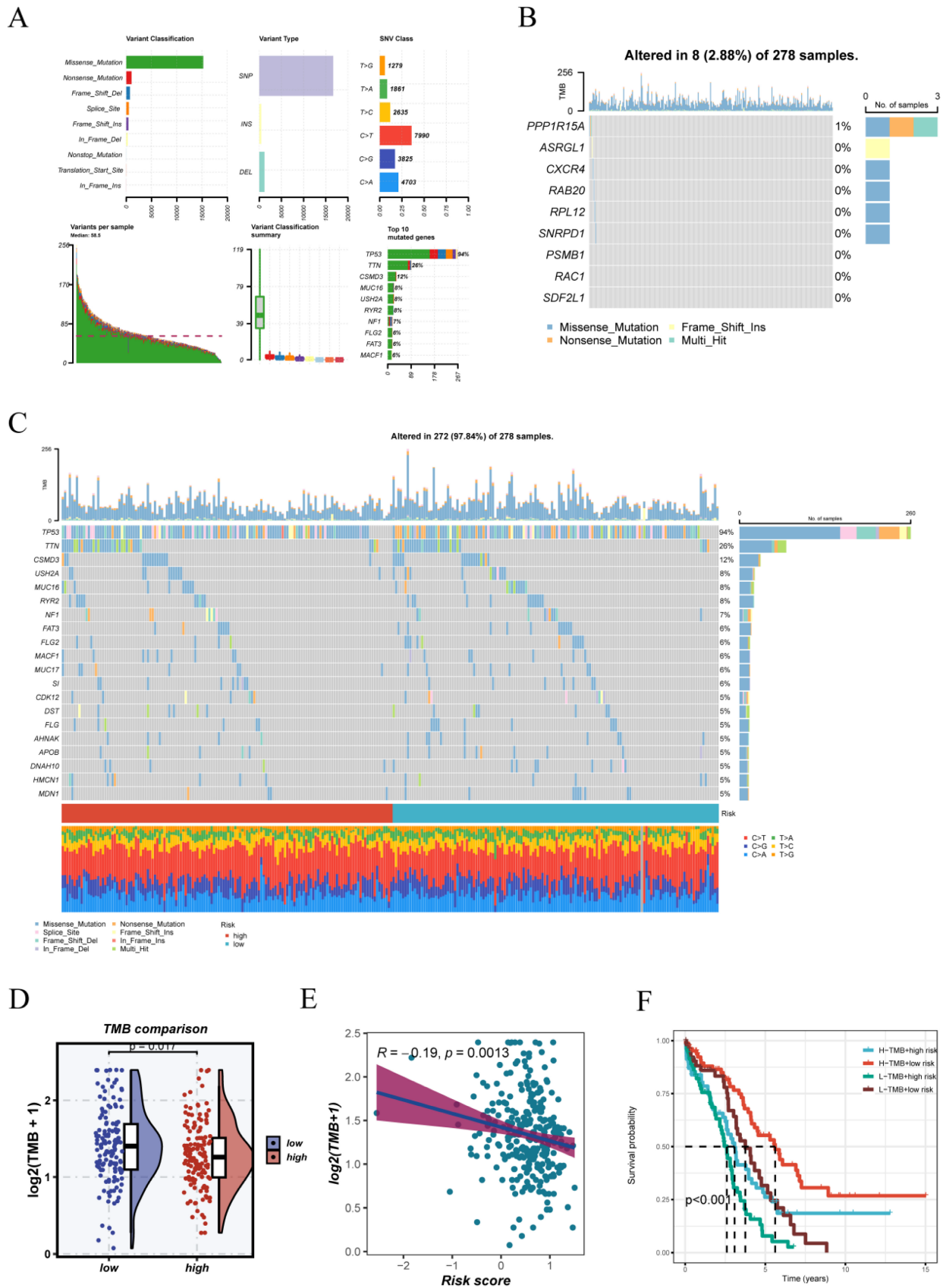
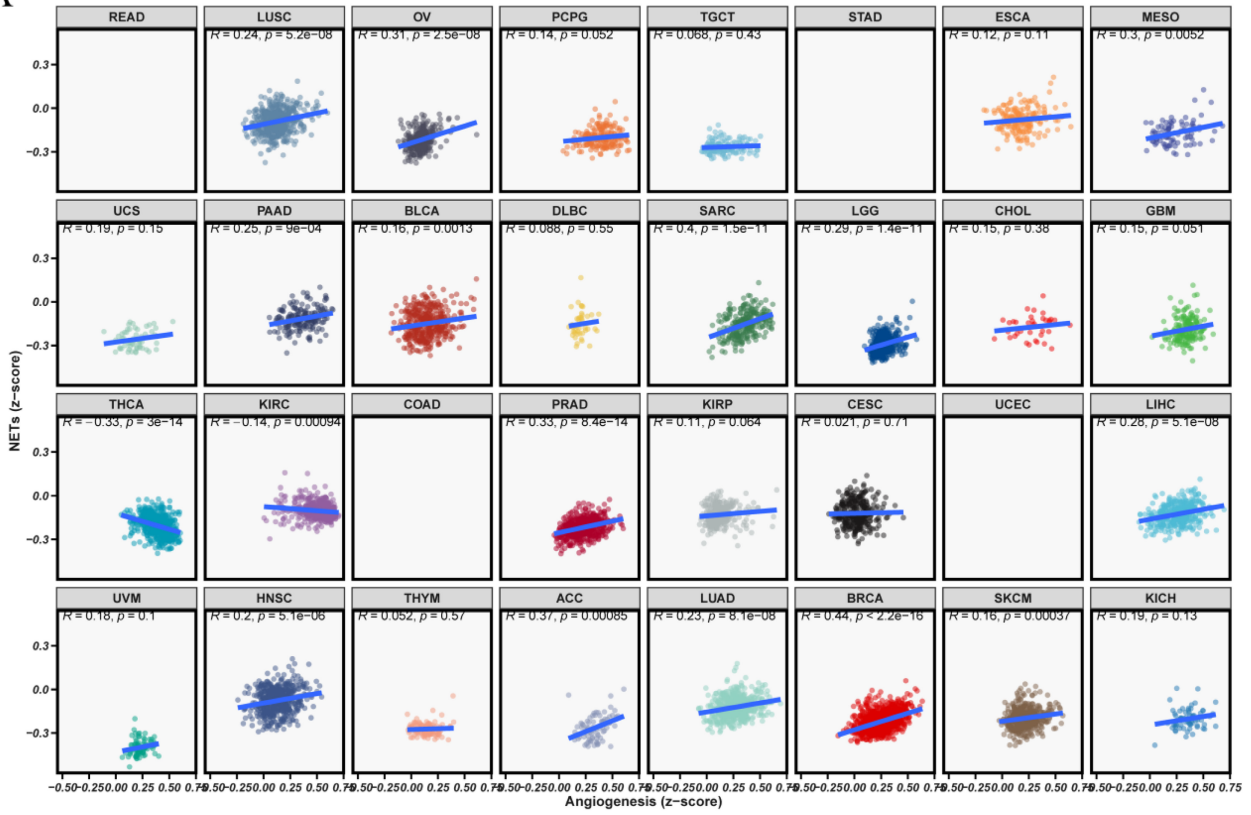


Figure 8. Somatic mutation and TMB analysis in TCGA-OV cohort. (A) Summary of somatic mutation profiles in TCGA-OV, including mutation classification, variant types, SNV classes, and top mutated genes. (B) Mutation status of the nine genes included in the prognostic model, with overall mutation frequencies. (C) Oncoplot illustrating the mutation landscape stratified by risk groups; mutation types and clinical annotations are shown below. (D) Tumor mutation burden (TMB) was compared between high- and low-risk groups, revealing that the high-risk group had lower TMB levels. (E) Correlation analysis revealed a significant negative relationship between risk score and TMB. (F) Combined survival analysis of risk score and TMB level.

A



B

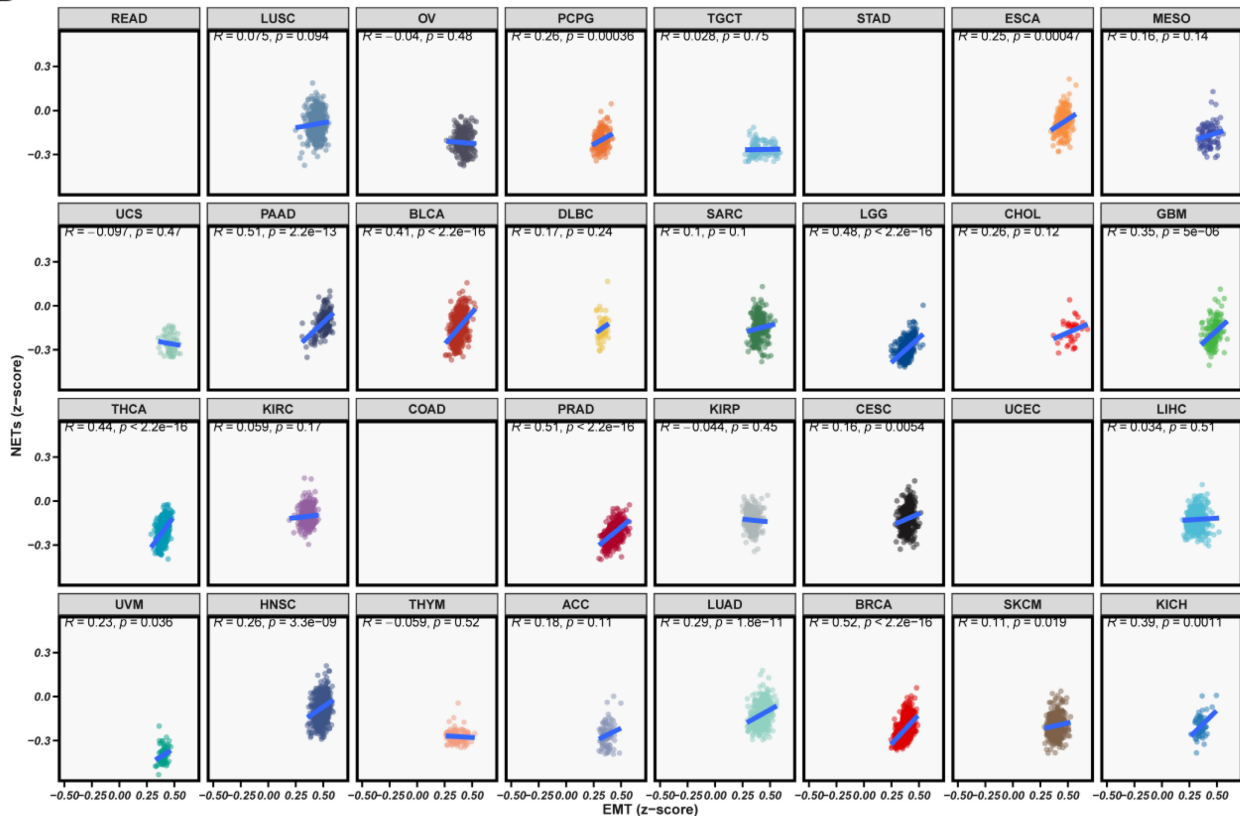


Figure 9. Association between risk score and tumor phenotype-related pathways in pan-cancer. (A) Scatter plots illustrating the relationship between risk scores and angiogenesis pathway activity across TCGA pan-cancer cohorts. (B) Scatter plots depicting the correlation between risk score and epithelial-mesenchymal transition (EMT) signature. Each dot represents an individual tumor sample, color-coded by cancer type. Each subplot displays Pearson correlation coefficients and corresponding p-values.

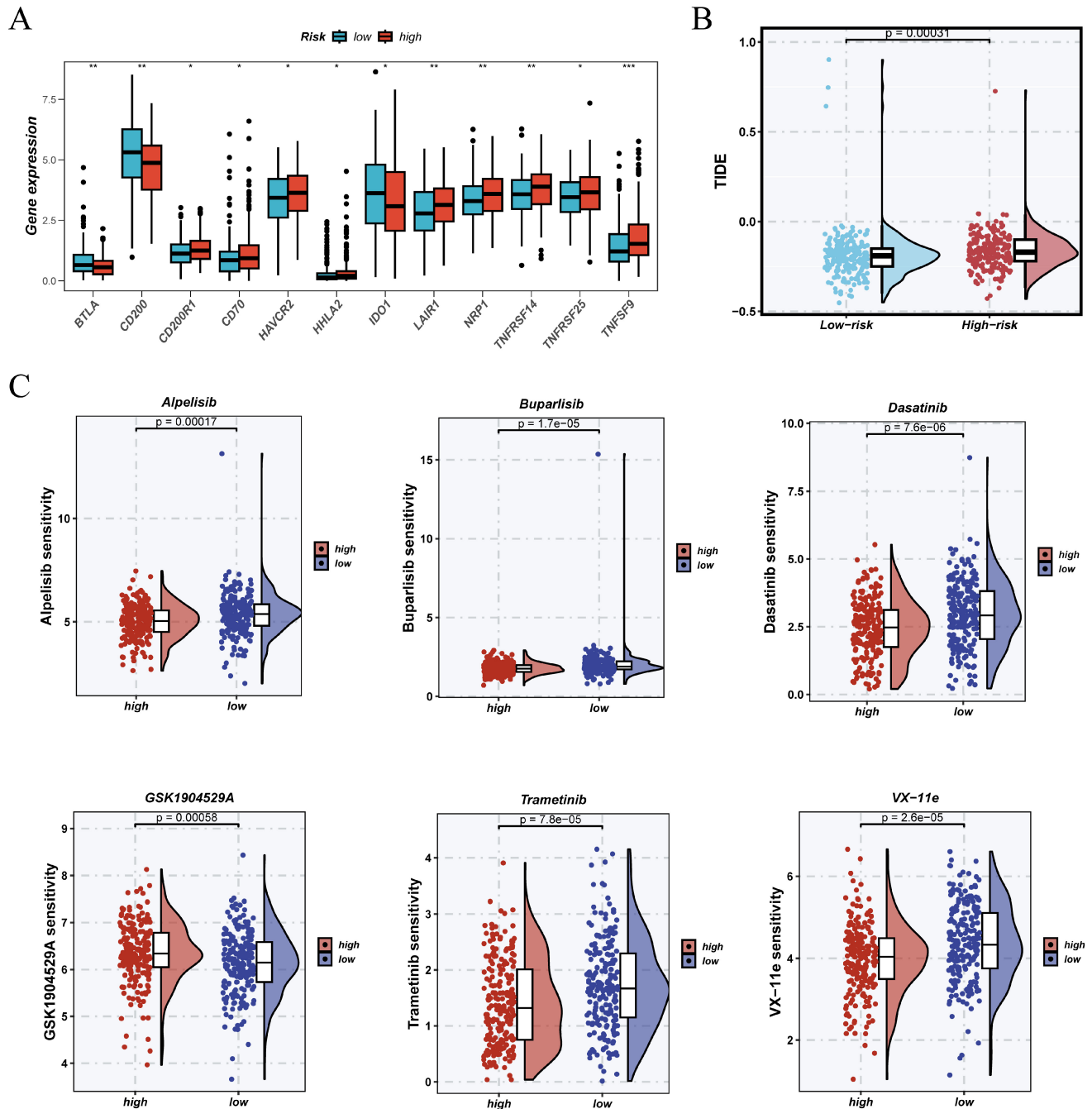


Figure 10. Potential of the risk score for predicting immunotherapy response and drug sensitivity. (A) Differential expression of immune checkpoint-related genes was observed between high- and low-risk groups. (B) TIDE analysis revealed that the high-risk group had a greater potential for immune evasion, indicating a possibly poorer response to immunotherapy. (C) The OncoPredict algorithm identified notable variations in drug sensitivity between the two groups.

TIDE analysis (Figure 10B) indicated that high-risk patients generally had elevated TIDE scores, implying a greater chance of immune evasion and diminished response to immune checkpoint blockade therapies. The OncoPredict algorithm's drug sensitivity profiling (Figure 10C) revealed that high-risk patients had increased IC50 values for several agents, such as Alpelisib, Buparlisib, Dasatinib, GSK1904529A, Trametinib, and VX-11e, suggesting reduced predicted sensitivity to these treatments.

These findings suggest that high-risk individuals may exhibit an immune-suppressive phenotype and diminished therapeutic response, underscoring the necessity for alternative or combination treatment strategies.

3.11 Expression Profile and Functional Validation of Tumor-Promoting Gene PSMB1

During model development, PSMB1 emerged as a key gene of interest due to its consistently high

hazard ratio in univariate Cox regression analysis and its relatively large coefficient in the final risk model, indicating its substantial contribution to risk stratification. Therefore, we selected PSMB1 for further experimental investigation. Analysis of GTEx and TCGA datasets showed significantly higher PSMB1 expression in tumor tissues than in normal controls (Figure 11A). Pan-cancer survival analysis further demonstrated that high PSMB1 expression was associated with worse prognosis across multiple cancer types, underscoring its potential as a prognostic biomarker (Figure 11B). Figure 11C presents a Sankey diagram using TCGA data, showing the relationships between PSMB1 expression and clinical variables, including age, TNM stage, and survival status.

To explore its functional role, siRNA-mediated knockdown of PSMB1 was established in SKOV3 and ES-2 OV cell lines, with qRT-PCR confirming effective gene silencing (Figure 11D–E). CCK-8 assays demonstrated that PSMB1 silencing markedly inhibited cell proliferation (Figure 11F). Colony formation assays demonstrated a significant decrease in tumorigenic potential following PSMB1 knockdown (Figure 11G), indicating that PSMB1 functions as a tumor-promoting gene in OV cells.

4. Discussion

OV is among the deadliest female reproductive system malignancies, primarily due to its subtle onset and the absence of effective early screening methods [22]. As a result, most patients are diagnosed at advanced stages [23]. Despite improvements in surgical techniques, chemotherapy, and the introduction of targeted and immunotherapeutic options, overall survival rates remain unsatisfactory, with five-year survival hovering below 30% [24]. In this context, the development of reliable molecular biomarkers and risk stratification tools is urgently needed to support personalized treatment approaches.

This study developed a single-cell transcriptomic atlas of epithelial cells in OV, identifying subpopulations with different malignant potentials. We then extracted subcluster-specific marker genes and applied machine learning algorithms to develop a biologically grounded, nine-gene risk score system. The model showed moderate but consistent prognostic performance across independent cohorts and was associated with multi-omics features, including mutation burden, immune microenvironment traits, and therapeutic responsiveness, indicating its potential translational relevance. This observation is consistent with recent evidence that malignant cell clusters identified by single-cell or multi-omics analysis may provide

clinically relevant information for prognostic interpretation and therapeutic exploration [25].

Among the genes incorporated in the model, ASRGL1 has been implicated in cellular metabolism, and although upregulated in some solid tumors, its expression in this study was associated with favorable prognosis, potentially reflecting context-dependent roles in specific epithelial subsets [26, 27]. CXCR4, a chemokine receptor broadly involved in metastasis and immune evasion, exhibited a negative coefficient here, suggesting tissue-specific heterogeneity in its prognostic impact [28]. PPP1R15A, known for its role in stress response and translational control, was associated with poor outcomes, possibly due to its protective effects on tumor cells under adverse conditions [28–30]. PSMB1, the focus of our functional validation, encodes a proteasome subunit involved in cell cycle progression and apoptosis inhibition [31]. Its involvement in tumor progression across various cancers has been suggested, and we validated its oncogenic role in OV using qRT-PCR, CCK-8, and colony formation assays.

RAB20, a member of the Ras GTPase family, is linked to vesicular transport and cell migration, and may promote invasive behavior in tumor cells [32, 33]. RAC1, a well-characterized Rho GTPase, drives cytoskeletal remodeling and has been validated as a pro-tumorigenic factor in various cancers [34]. RPL12 encodes a ribosomal protein and, although underexplored, may support tumor cell proliferation through enhanced protein synthesis, as suggested by its positive association with poor prognosis [35]. SDF2L1, an endoplasmic reticulum-associated protein involved in stress response and protein folding, emerged as a protective factor, possibly due to its role in maintaining cellular homeostasis [36, 37]. SNRPD1, a component of the spliceosomal complex, has been associated with tumor cell proliferation and unfavorable outcomes in several cancers [38, 39]. However, in the present OV-derived risk model, SNRPD1 showed a negative coefficient, suggesting that its prognostic effect may be context-dependent and may reflect the specific transcriptional state captured by this malignant epithelial cell-derived signature.

From a translational perspective, our risk score model offers several advantages. It is biologically interpretable due to its derivation from well-defined malignant epithelial subpopulations. It demonstrated consistent predictive performance in both TCGA and multiple external validation cohorts. Furthermore, its associations with clinical features, immune infiltration, mutation profiles, and drug sensitivity highlight its potential utility in guiding personalized therapy.

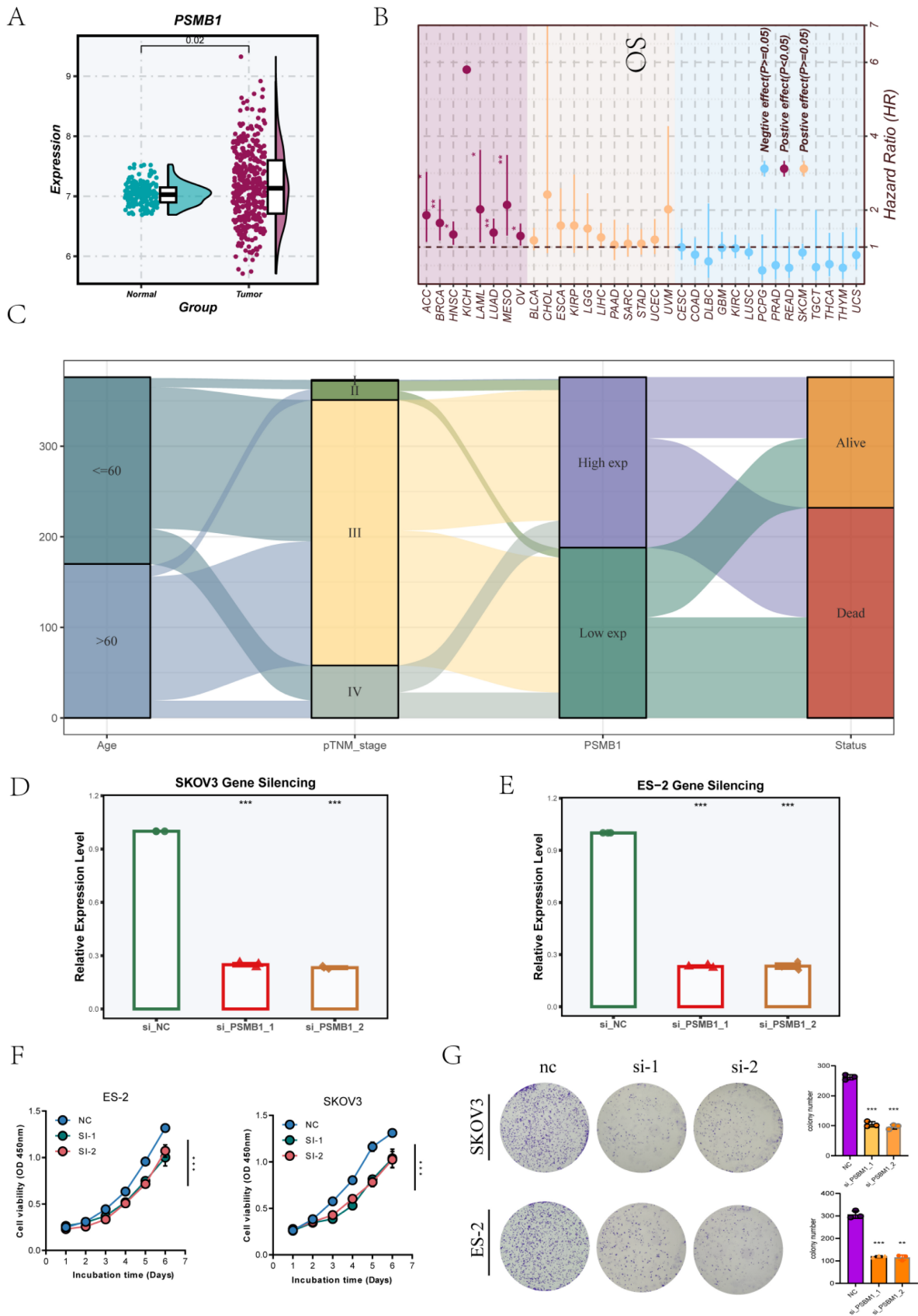


Figure 11. Expression Profile and Functional Validation of Tumor-Promoting Gene PSMB1. (A) Analysis of GTEx and TCGA datasets showed significant overexpression of PSMB1 in tumor tissues. (B) Pan-cancer survival analysis revealed that elevated PSMB1 expression correlates with poor prognosis in various cancer types. (C) The Sankey diagram based on TCGA data depicted the associations among PSMB1 expression, age, TNM stage, and survival status. (D-E) siRNA knockdown models of PSMB1 were effectively created in SKOV3 and ES-2 OV cell lines, with gene silencing efficiency verified by qRT-PCR. (F) CCK-8 assays demonstrated that PSMB1 knockdown suppressed cell proliferation. (G) Colony formation assays further validated the tumor-promoting role of PSMB1.

Nonetheless, this study has limitations. Although the model was validated in publicly available cohorts, prospective clinical validation in larger and more diverse populations is necessary to confirm its generalizability. Moreover, the accuracy of cell subpopulation identification and batch correction may influence downstream gene selection and model stability. Although PSMB1 knockdown was shown to inhibit OV cell proliferation and colony formation, the current experimental validation was limited to *in vitro* growth-related assays. Further studies incorporating migration, invasion, apoptosis, rescue experiments, protein-level validation, and *in vivo* tumor models are warranted to more comprehensively define the functional role and therapeutic relevance of PSMB1 in OV. Future studies integrating spatial transcriptomics, proteomics, and metabolomics could provide deeper insights into the biology of key subpopulations and uncover novel therapeutic targets. In addition, TIDE-based immunotherapy prediction and OncoPredict-based drug sensitivity analysis should be interpreted as exploratory computational estimates rather than direct evidence of clinical treatment response. These findings require further validation in ovarian cancer cohorts with documented immunotherapy outcomes, pharmacological response data, and experimental models.

In summary, this study developed a robust epithelial cell-derived prognostic model for OV by integrating single-cell and bulk transcriptomic data. The model provides reproducible risk stratification and also sheds light on the biological mechanisms and therapeutic implications of tumor heterogeneity, offering a potentially useful framework for precision oncology in OV.

Supplementary Material

Supplementary figures and tables.
<https://www.jcancer.org/v17p1274s1.pdf>

Acknowledgments

The authors of this article would like to thank the online databases such as TCGA and GEO for providing the data.

Funding

This study was supported by the Tianjin Municipal Health and Medical Science and Technology Project (TJWJ2024ZD2008).

Data Availability Statement

All data used in this study are publicly available. The single-cell RNA sequencing datasets were obtained from the Gene Expression Omnibus (GEO) database under accession numbers GSE154600 and

GSE184880. Bulk RNA-seq datasets used for external validation were also retrieved from GEO, including GSE13876, GSE140082, and GSE26712.

The TCGA-OV cohort was downloaded from The Cancer Genome Atlas (TCGA) via the UCSC Xena platform (<https://xena.ucsc.edu/>). Somatic mutation data (MAF files) were obtained from the Genomic Data Commons (GDC) portal (<https://portal.gdc.cancer.gov/>).

TIDE (Tumor Immune Dysfunction and Exclusion) scores were obtained from the TIDE web server (<http://tide.dfci.harvard.edu/>) for immunotherapy response prediction.

All relevant data are publicly accessible from the above-mentioned repositories. Further information can be obtained from the corresponding author upon reasonable request.

Author Contributions

Jingwen Si and Han Zhang conceptualized the study and performed data analysis and manuscript drafting. Hanbo Li contributed to data analysis and experimental validation. Yan Shen critically reviewed and approved the final version of the manuscript for submission. All authors have read and approved the final manuscript and consent to its publication.

Competing Interests

The authors have declared that no competing interest exists.

References

- Battistini C, Cavallaro U. Patient-Derived *In Vitro* Models of Ovarian Cancer: Powerful Tools to Explore the Biology of the Disease and Develop Personalized Treatments. *Cancers (Basel)*. 2023; 15.
- Xiao Y, Bi M, Guo H, Li M. Multi-omics approaches for biomarker discovery in early ovarian cancer diagnosis. *EBioMedicine*. 2022; 79: 104001.
- Cubillos-Ruiz JR, Fiering S, Conejo-Garcia JR. Nanomolecular targeting of dendritic cells for ovarian cancer therapy. *Future Oncol*. 2009; 5: 1189-92.
- Tang W, Li J, Chang X, Jia L, Tang Q, Wang Y, et al. Construction of a novel prognostic-predicting model correlated to ovarian cancer. *Biosci Rep*. 2020; 40.
- Qin Y, Sheng Y, Ren M, Hou Z, Xiao L, Chen R. Identification of necroptosis-related gene signatures for predicting the prognosis of ovarian cancer. *Sci Rep*. 2024; 14: 11133.
- Zhu J, Fan Y, Xiong Y, Wang W, Chen J, Xia Y, et al. Delineating the dynamic evolution from preneoplasia to invasive lung adenocarcinoma by integrating single-cell RNA sequencing and spatial transcriptomics. *Exp Mol Med*. 2022; 54: 2060-76.
- Shao W, Lin Z, Xiahou Z, Zhao F, Xu J, Liu X, et al. Single-cell RNA sequencing reveals that MYBL2 in malignant epithelial cells is involved in the development and progression of ovarian cancer. *Front Immunol*. 2024; 15: 1438198.
- Winterhoff BJ, Maile M, Mitra AK, Sebe A, Bazzaro M, Geller MA, et al. Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells. *Gynecol Oncol*. 2017; 144: 598-606.
- Carey KM, Young CD, Clark AJ, Dammer EB, Singh R, Lillard JW, Jr. Subtype-specific analysis of gene co-expression networks and immune cell profiling reveals high grade serous ovarian cancer subtype linkage to variable immune microenvironment. *J Ovarian Res*. 2024; 17: 240.
- Zhao F, Jiang X, Li Y, Huang T, Xiahou Z, Nie W, et al. Characterizing tumor biology and immune microenvironment in high-grade serous ovarian cancer via single-cell RNA sequencing: insights for targeted and personalized immunotherapy strategies. *Front Immunol*. 2024; 15: 1500153.
- Wang G, Shi C, He L, Li Y, Song W, Chen Z, et al. Identification of the tumor metastasis-related tumor subgroups overexpressed NENF in triple-negative breast cancer by single-cell transcriptomics. *Cancer Cell Int*. 2024; 24: 319.

12. Wang T, Tian L, Wei B, Li J, Zhang C, Long R, et al. Effect of fibroblast heterogeneity on prognosis and drug resistance in high-grade serous ovarian cancer. *Sci Rep.* 2024; 14: 26617.
13. Ding H, Hu B, Guo R. Comprehensive analysis of single cell and bulk data develops a promising prognostic signature for improving immunotherapy responses in ovarian cancer. *PLoS One.* 2024; 19: e0298125.
14. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 2016; 44: e71.
15. Zhang P, Wang D, Zhou G, Jiang S, Zhang G, Zhang L, et al. Novel post-translational modification learning signature reveals B4GALT2 as an immune exclusion regulator in lung adenocarcinoma. *J Immunother Cancer.* 2025; 13.
16. Li P, Zhang H, Sun L, Wu X. Toward precision oncology in LUAD: a prognostic model using single-cell sequencing and WGCNA based on a disulfidptosis relative gene signature. *Front Immunol.* 2025; 16: 1581915.
17. Shi M, Zhang H, Ma L, Wang X, Sun D, Feng Z. Innovative prognostic modeling in ESCC: leveraging scRNA-seq and bulk-RNA for dendritic cell heterogeneity analysis. *Front Immunol.* 2024; 15: 1352454.
18. Chen K, Wang Y, Hou Y, Wang Q, Long D, Liu X, et al. Single cell RNA-seq reveals the CCL5/SDC1 receptor-ligand interaction between T cells and tumor cells in pancreatic cancer. *Cancer Lett.* 2022; 545: 215834.
19. Zhang H, Dai J, Mu Q, Zhao X, Lin Z, Wang K, et al. Macrophage heterogeneity and oncogenic mechanisms in lung adenocarcinoma: insights from scRNA-seq analysis and predictive modeling. *Front Immunol.* 2024; 15: 1491872.
20. Zhao AY, Unterman A, Abu Hussein NS, Sharma P, Nikola F, Flint J, et al. Single-Cell Analysis Reveals Novel Immune Perturbations in Fibrotic Hypersensitivity Pneumonitis. *Am J Respir Crit Care Med.* 2024; 210: 1252-66.
21. Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013; 14: 7.
22. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018; 28: 1747-56.
23. Liberto JM, Chen SY, Shih IM, Wang TH, Wang TL, Pisanic TR, 2nd. Current and Emerging Methods for Ovarian Cancer Screening and Diagnostics: A Comprehensive Review. *Cancers (Basel).* 2022; 14.
24. Tavares V, Marques IS, Melo IG, Assis J, Pereira D, Medeiros R. Paradigm Shift: A Comprehensive Review of Ovarian Cancer Management in an Era of Advancements. *Int J Mol Sci.* 2024; 25.
25. Wang G, Cao J, Lu C, Cao Y, Wang S, Chen Z, et al. Multiomics profiling identifies the poor prognostic role of a tumor cluster with GNA15 overexpression in triple-negative breast cancer. *Front Immunol.* 2025; 16: 1659183.
26. Garcia-Montojo M, Fathi S, Rastegar C, Simula ER, Doucet-O'Hare T, Cheng YHH, et al. TDP-43 proteinopathy in ALS is triggered by loss of ASRGL1 and associated with HML-2 expression. *Nat Commun.* 2024; 15: 4163.
27. Zhao H, Gao Y, Miao J, Chen S, Li J, Li Z, et al. Single-cell RNA-seq highlights a specific carcinoembryonic cluster in ovarian cancer. *Cell Death Dis.* 2021; 12: 1082.
28. Li J, Jiang K, Qiu X, Li M, Hao Q, Wei L, et al. Overexpression of CXCR4 is significantly associated with cisplatin-based chemotherapy resistance and can be a prognostic factor in epithelial ovarian cancer. *BMB Rep.* 2014; 47: 33-8.
29. Ding ZB, Chen Y, Zheng YR, Wang YY, Deng WD, Zheng JH, et al. Inhibition of PPP1R15A alleviates osteoporosis via suppressing RANKL-induced osteoclastogenesis. *Acta Pharmacol Sin.* 2024; 45: 790-802.
30. Magg V, Manetto A, Kopp K, Wu CC, Naghizadeh M, Lindner D, et al. Turnover of PPP1R15A mRNA encoding GADD34 controls responsiveness and adaptation to cellular stress. *Cell Rep.* 2024; 43: 114069.
31. Li L, Bai Y, Zhou Y, Jiang Y, Tong W, Li G, et al. PSMB1 Inhibits the Replication of Porcine Reproductive and Respiratory Syndrome Virus by Recruiting NBR1 To Degrade Nonstructural Protein 12 by Autophagy. *J Virol.* 2023; 97: e0166022.
32. Tan X, Yuan G, Wang Y, Zou Y, Luo S, Han H, et al. RAB20 Promotes Proliferation via G2/M Phase through the Chk1/cdc25c/cdc2-cyclinB1 Pathway in Penile Squamous Cell Carcinoma. *Cancers (Basel).* 2022; 14.
33. Peng Z, Duan M, Zhao K, Tang Y, Liang F. RAB20 deficiency promotes the development of silicosis via NLRP3 inflammasome. *Front Immunol.* 2022; 13: 967299.
34. De P, Aske JC, Dey N. RAC1 Takes the Lead in Solid Tumors. *Cells.* 2019; 8.
35. Chen Y, Hu J, Zhao P, Fang J, Kuang Y, Liu Z, et al. Rpl12 is a conserved ribophagy receptor. *Nat Cell Biol.* 2025; 27: 477-92.
36. Gong W, Martin TA, Sanders AJ, Jiang A, Sun P, Jiang WG. Location, function and role of stromal cell-derived factors and possible implications in cancer (Review). *Int J Mol Med.* 2021; 47: 435-43.
37. Sasako T, Ohsugi M, Kubota N, Itoh S, Okazaki Y, Terai A, et al. Hepatic Sdf2l1 controls feeding-induced ER stress and regulates metabolism. *Nat Commun.* 2019; 10: 947.
38. Dai X, Yu L, Chen X, Zhang J. SNRPD1 confers diagnostic and therapeutic values on breast cancers through cell cycle regulation. *Cancer Cell Int.* 2021; 21: 229.
39. Dai X, Cai L, Zhang Z, Li J. SNRPD1 conveys prognostic value on breast cancer survival and is required for anthracycline sensitivity. *BMC Cancer.* 2023; 23: 376.