Discover Oncology

Research

PLOD3 as a novel oncogene in prognostic and immune infiltration risk model based on multi-machine learning in cervical cancer

Lingling Qiu[1] · Xiuchai Qiu[2] · Xiaoyi Yang[3]

Received: 12 December 2024 / Accepted: 3 March 2025
Published online: 11 March 2025

Abstract
Cervical carcinoma (CC) remains a significant global health issue despite advancements in screening and treatment. To improve prognostic accuracy and therapeutic strategies, we developed a multi-machine learning prognostic model based on metabolic-associated genes. This study integrated genomic, transcriptomic, and spatial data from multiple databases to identify key metabolic genes with a causal relationship to CC. We identified 112 key metabolic genes, which were used to construct and validate a prognostic model through various machine learning algorithms. GO and KEGG enrichment analysis revealed the MAPK cascade plays a crucial role in metabolic processes. To pinpoint key metabolic genes, we constructed WGCNA and extracted 337 key genes. Supervised principal component analysis and random survival forests were incorporated into the final model, which showed strong predictive ability in classifying patients. Furthermore, the model demonstrated notable variations in immune cell infiltration among risk categories, which shown regulatory T cells may be involved in immune suppression, and natural killer cells might have a limited effect in tumor clearance. Spatial transcriptomics and single-cell analyses further validated the model, uncovering tumor heterogeneity and distinct intercellular communication patterns associated with different risk levels. The functional experiment results indicated that down expression of PLOD3 could suppress the proliferation of CC cell. In this study, offer a precision medicine methods for predicting patient outcomes as well as fresh insights into the metabolic foundations, which may contribute to the prognosis and immunotherapy of CC. Additionally, we discovered PLOD3 to be a novel oncogene in CC. These findings imply that this model may be applied to assess prognostic risk and identify potential therapeutic targets for CC patients.

Keywords Cervical carcinoma · Metabolic · Spatial transcriptomics · Immune · PLOD3

1 Introduction

Cervical carcinoma (CC) continues to be a major gynecological cancer, even with the recent reduction in cases and deaths due to better HPV screening [1, 2]. This disease remains a global health issue, affecting women of all ages. According to the World Health Organization (WHO), approximately 311,365 deaths and 569,847 new cases of cervical

Lingling Qiu, Xiuchai Qiu and Xiaoyi Yang have contributed equally to this work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02031-2.

✉ Lingling Qiu, linglingqiu2021@163.com | [1]Department of Reproductive Medicine, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, China. [2]Department of Public Health, Jinjiang City Hospital Jinnan Branch, Fujian Province, China. [3]Department of Oncology, Shanghai General Hospital, , Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China.

cancer are reported annually [3]. Despite progress, early-stage or localized CC is often diagnosed late, which limits the effectiveness of surgery. Although some patients benefit from surgery in early-stage cases, the majority are diagnosed at more advanced stages [4]. Traditional treatments like chemotherapy and radiation have shown limited efficacy in significantly extending survival for advanced CC patients [5]. Therefore, it is crucial to improve CC screening methods and develop innovative therapeutic strategies that integrate tumor metabolism, immune responses, and predictive modeling to enhance clinical outcomes.

Tumor cells are known to alter their metabolic pathways to acquire the nutrients necessary for their growth and persistence [6]. This metabolic reprogramming can weaken the immune system's ability to fight tumors, leading to further cancer development and reduced effectiveness of immunotherapy [6]. Tumor formation and progression are significantly influenced by the interplay between immune cells and cellular metabolism [7]. Immune cells, such as macrophages and T cells, are essential for recognizing and attacking cancer cells. However, the metabolic reprogramming of tumor cells can create a hostile environment that impairs the effectiveness of these immune responses [8]. Tumor cells often increase their uptake of glucose and other nutrients to support rapid growth and survival, which can deprive immune cells of the energy they need to function effectively [9–11]. Additionally, the metabolic byproducts of tumor cells, such as lactate, can create an acidic microenvironment that further inhibits immune cell activity [12]. Understanding these metabolic alterations and their impact on immune cell function is essential for developing new therapeutic strategies that can enhance the immune response and improve outcomes for cervical cancer patients [13]. Research has consistently associated poor outcomes in cancer patients with disrupted metabolic activities involving nucleotides and carbohydrates [14]. However, the detailed interactions between these metabolic changes and immune responses in cervical carcinoma (CC) remain unclear. Thus, there is a pressing need for practical and comprehensive models to predict the prognosis of CC patients accurately.

Traditionally, prognostic models used Cox regression. This method, however, is based on the assumption of linear relationships, which restricts its ability to represent the intricate, multidimensional, and nonlinear interactions among various prognostic factors. Advances in genomic profiling and big data analytics now offer significant opportunities for creating sophisticated risk models, thus improving patient care and prognosis assessment. Machine learning (ML) has become increasingly popular in medical research due to its capacity to process complex and extensive datasets and deliver highly precise predictions [15]. This makes ML exceptionally effective for analyzing medical data. Studies have demonstrated that ML surpasses traditional statistical methods in predicting survival rates, particularly when managing numerous variables [16].

In this research, we developed a prognostic model for cervical cancer centered on metabolic genes linked to patient outcomes. Additionally, we explored the model's relationship with the tumor microenvironment and immune response in terms of immune cell distribution, cell communication and pseudotime analysis, which help to determine its clinical relevance in predicting cervical cancer prognosis. Furthermore, the model's performance had been validated based on single cell data and spatial transcriptomes. In conclusion, this approach aids in understanding the interplay between tumor immunity and metabolism. Our findings offer novel insights into forecasting cervical cancer prognosis, addressing the need for precision medicine approaches that can predict patient outcomes more accurately and uncover the metabolic underpinnings of CC progression. Furthermore, we identified PLOD3 as a novel oncogene in CC, which highlights the potential for our model to assess prognostic risk and identify therapeutic targets. These findings contribute to advancing immunotherapy and prognosis methodologies in CC patients.

2 Method and materials

2.1 Data collection

To develop a metabolism-based prognostic model for cervical cancer to enhance precision medicine in clinical settings, this study utilized multiple datasets. Two GWAS datasets, 7PQTL and 9PQTL, were acquired to investigate genetic variants associated with metabolic traits relevant to cervical cancer. Additionally, five cervical cancer-related GWAS datasets were downloaded from the IEU (MRC Integrative Epidemiology Unit) database, including ukb-b-8777 (sample size: 462,933), ieu-b-4876 (sample size: 199,086), ebi-a-GCST90018817 (sample size: 239,158), ebi-a-GCST90018597 (sample size: 61,581), and bbj-a-98 (sample size: 90,336). These datasets provided insights into genetic predispositions and associations relevant to cervical cancer development and progression. The TCGA-CESC

cervical cancer dataset, obtained from the UCSC Xena platform, included gene expression sequencing data (Count, FPKM, and normalized TPM values) for 293 patients, along with clinical data such as age, gender, survival time, and status. Additionally, one RNAseq dataset (GSE52903, Homo sapiens, GPL6244, 55 samples) and one single-cell dataset (GSE208653, Homo sapiens, 3 samples) were downloaded from the GEO database, focusing on squamous cervical carcinoma patients. A spatial transcriptomics sample from a cervical cancer patient (Dataset ID: STDS0000113) was obtained from the STOmicsDB database to capture spatially resolved gene expression patterns. Finally, 23,497 metabolism-related coding genes identified from the GeneCards database were integrated into the analysis to explore metabolic pathways and their implications in cervical cancer. This comprehensive dataset collection facilitated the development of a robust metabolism-based prognostic model for cervical cancer.

2.2 Mendelian randomization analysis to identify potential causal metabolic genes and functional enrichment

Genetic variants are used as instrumental variables (IVs) in Mendelian Randomization (MR) analysis, which looks into causal links based on three main assumptions: The genetic instruments have three characteristics: (i) they are linked to the exposure; (ii) they are unaffected by any confounding factors; and (iii) they only influence the result as a result of the exposure. Strongly associated and independent variants ($p < 5 \times 10^{-8}$, linkage disequilibrium [LD] $r^2 < 0.001$, window size = 10,000 kb) unrelated by condition to cervical cancer were selected as IVs. As our primary analysis strategy, we used the random effects inverse variance weighted (IVW) method from the "TwoSampleMR" package, taking *p*-values less than 0.05 as evidence of possible causal links. This method identified proteins and their corresponding coding genes that exhibit a positive causal relationship with cervical cancer occurrence in both the 7PQTL and 9PQTL GWAS datasets, based on data from ukb-b-8777, ieu-b-4876, ebi-a-GCST90018817, ebi-a-GCST90018597, and bbj-a-98. Subsequent functional enrichment analysis of these genes was conducted using the Reactome pathway database and Metascape database (www.metascape.org/). A *p*-value of less than 0.05 was considered statistically significant for enrichment.

2.3 WGCNA analysis

Using the "WGCNA" package in R, we built a gene co-expression network based on candidate genes and the TCGA-CESC dataset. The following steps were included in the process: First, a similarity matrix was created from each transcript's expression levels using the matched Pearson correlation values. The formula amn = | cMn | β was then used to convert this similarity matrix into an adjacency matrix, where cMn stands for the Pearson correlation between paired genes and amn = the relation between gene pairs. The adjacency matrix underwent additional conversion to a topological overlap matrix when β = 3. Genes with comparable expression patterns were finally grouped into distinct modules using a bottom-up approach, with a minimum module size cutoff of 10. For additional examination, core genes were defined as those having a *p*-value of 0.05 or below.

2.4 One-stop prognostic feature gene screening and model construction using multiple machine learning approaches

To identify genes significantly correlated with cervical cancer prognosis, we conducted a univariate Cox regression analysis. Subsequently, we developed a robust predictive model by integrating ten machine learning algorithms: Ridge Regression, Cox Partial Least Squares Regression (plsRcox), CoxBoost, Stepwise Cox, Random Survival Forests (RSF), Gradient Boosting Machine (GBM), Survival Support Vector Machine (Survival-SVM), Supervised Principal Components (SuperPC), Elastic Net (Enet), and Lasso.

By combining these algorithms, we generated 116 machine learning models, emphasizing those with dimension reduction and variable selection capabilities, such as RSF, LASSO, CoxBoost, and Stepwise Cox. These models were applied to the TCGA-CESC dataset to select key genes and construct prognostic models. The most effective model, determined by the highest C-index in the validation cohort GSE52903, was designated as the Metabolic Score (MS). Patients were stratified into high-risk and low-risk groups based on the MS median value. Expression variations of

significant genes between these groups were visualized, and the prognostic value of MS was assessed through ROC curve analysis and survival analysis.

2.5 Immune infiltration analysis

Using the "estimate" R package, we performed differential analyses on tumor immune scores, stromal scores, and tumor purity in order to investigate the discriminative power of the metabolic score (MS) risk score in tumor immunity. Next, single-sample gene set enrichment analysis (ssGSEA) was carried out. ssGSEA is a method used to assess the activity of gene sets (such as pathways or functional sets) in individual samples by quantifying the enrichment of gene sets through the cumulative distribution function of genes within a sample, thereby revealing activities associated with specific biological processes. We conducted ssGSEA for each patient to obtain scores across 28 types of immune cells and visualized the results using heatmaps and differential box plots.

2.6 Single-cell transcriptomic analysis for the validation of key genes

We used the R "Seurat" package to process and analyze data from single-cell RNA sequencing (scRNA-seq). scRNA-seq data was normalized using the "LogNormalize" method by utilizing the "NormalizeData" function in Seurat. Then, Seurat objects were created from the normalized data. For quality control (QC), we determined the proportion of ribosomal or mitochondrial genes and eliminated cells with low quality. Cells containing more than 20% ribosomal RNA and those with gene counts between 200 and 3000 were eliminated. The top 3000 genes were determined to be the most variable features using Seurat's FindVariableFeatures function. This information was then used to normalize the scRNA-seq data for each cell. In addition, we used the Seurat objects as the basis for the ScaleData and RunPCA routines to get the principal component (PC) count. Uniform Manifold Approximation and Projection (UMAP) was utilized to achieve additional dimensionality reduction. Lastly, we used the Idents and DimPlot tools to annotate and illustrate the primary cell types or subtypes, drawing on previously established literature. We then used the CellChat software, which has a database of ligand-receptor interactions for both people and mice, to perform an analysis of cell communication. First, we evaluated the primary signaling inputs and outputs between all kinds of cell clusters using CellChat and CellChatDB.human. Then, we used the netVisual_circle function to show the cell communication network from the target cell cluster to various clusters. Furthermore, we used the R program Monocle to investigate the connection between model genes and cell pseudotime trajectories. The selection of highly variable genes was done using the following standards: an empirical value for dispersion ≥ 1 times the dispersion fit and an average expression ≥ 0.1. The DDRTree approach was used to reduce dimensionality. The "plot_pseudotime_heatmap" function was then used to visualize the dynamic expression of model genes across various TME cell types in HGG pseudotime trajectories.

2.7 Spatial transcriptomics analysis

Spatial transcriptomic data were processed using the Seurat package, employing the 'SC Transform' function for normalization. Subsequently, the expression matrix was subjected to machine learning model scoring to compute risk scores for each spot. Within Seurat, the 'SpatialFeaturePlot' function was utilized to visualize these spot risk scores, illustrating the distribution of risk across different spatial locations.

2.8 Cell culture and transfection

The human CC cell line Hela were provided by the National Collection Authenticated Cell Cultures (Shanghai, China). Hela cells are the first aneuploid epithelioid cell line derived from human tissue and maintained in continuous culture.

Hela cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco, Grand Island, NY, USA) supplemented with 10% fetal bovine serum (FBS, Invitrogen, San Diego, CA, USA) and 1% penicillin–streptomycin solution to prevent contamination. Cells were maintained at 37 °C in a humidified incubator under 5% CO_2 to ensure optimal

growth conditions. The coding sequence of the human PLOD3 gene was amplified by PCR and cloned into the pEZ-M03 vector, a mammalian expression vector containing a CMV promoter for high-level expression. The construct was confirmed by Sanger sequencing to ensure correct insertion and sequence fidelity. For transfection, Hela cells were seeded in 6-well plates and transfected at 70–80% confluency using Lipofectamine 3000 reagent (Catalog No. L3000015, Invitrogen, China) following the manufacturer's protocol. Empty pEZ-M03 vector was used as a negative control to validate the specificity of PLOD3 overexpression effects. These detailed procedures ensure reproducibility and clarity for further experimental replication.

2.9 Colony formation assay

In order to assess how PLOD3 influences the development of cervical cancer cell colonies, $1 \times 10^{\wedge}$ cells were planted into each well of a 6-well plate. The cells were meticulously reconstituted and incubated at 37 °C with 5% CO_2 in a medium enhanced with 10% fetal bovine serum (FBS). After 7 days of incubation, the formation of colonies was assessed, with clusters of ≥ 30 cells being counted as a single colony. The assay was repeated three times to ensure the statistical validity of the results.

2.10 Statistical analysis

R software (version 4.3.1) was used for all statistical analyses. When comparing differences between two groups, the Wilcoxon rank-sum test was utilized, and when comparing differences between more than two groups, the Kruskal–Wallis test was applied. Correlation analyses were conducted using Spearman's correlation approach. *p*-values below 0.05 were regarded as statistically significant. The aforementioned methodology section has a detailed description of the various specific statistics that were involved.

3 Results

3.1 Identification of genes with a positive causal relationship to cervical cancer incidence

To identify genes that have a positive causal relationship with the incidence of cervical cancer, we conducted a two-sample Mendelian randomization analysis. Initially, we collected two proteomic GWAS datasets, 7pqtl and 9pqtl, and five cervical cancer GWAS datasets from the IDE database: ukb-b-8777, ieu-b-4876, ebi-a-GCST90018817, ebi-a-GCST90018597, and bbj-a-98. We performed Mendelian randomization analyses using the proteomic GWAS data in conjunction with the cervical cancer GWAS data. After retrieving and merging exposure and outcome data, we employed the Inverse Variance Weighted (IVW) method to identify instrumental variables (IVs). Reliable results were obtained from both GWAS datasets; specifically, 880 SNP loci were selected as IVs in the 7pqtl dataset (Fig. 1a), and 500 SNP loci were identified as IVs in the 9pqtl dataset (Fig. 1b). Through the connections established by the IVs, a total of 3656 proteins were identified that have a positive causal relationship with the incidence of cervical cancer from the two proteomic GWAS datasets.

3.2 Metabolism-related positive causal relationship gene enrichment analysis

To explore the positive impact of metabolic activity in cervical cancer patients on their prognosis, we identified 3,099 metabolism-related genes from 3656 proteins that have a positive causal relationship with cervical cancer (Fig. 2a). These 3099 genes were used as candidates for subsequent analysis. Gene Ontology (GO) enrichment analysis revealed significant enrichments in cellular components (CC), molecular functions (MF), and biological processes (BP). Notably, the most significant enrichment in BP was observed in the positive regulation of the MAPK cascade, which plays a crucial role in metabolic processes such as regulating carbohydrate, lipid, and protein metabolism. The external side of the plasma membrane was most significantly enriched in CC, critical for regulating material ingress and egress as well as signal transduction, thus profoundly influencing cellular metabolic activities. In MF,

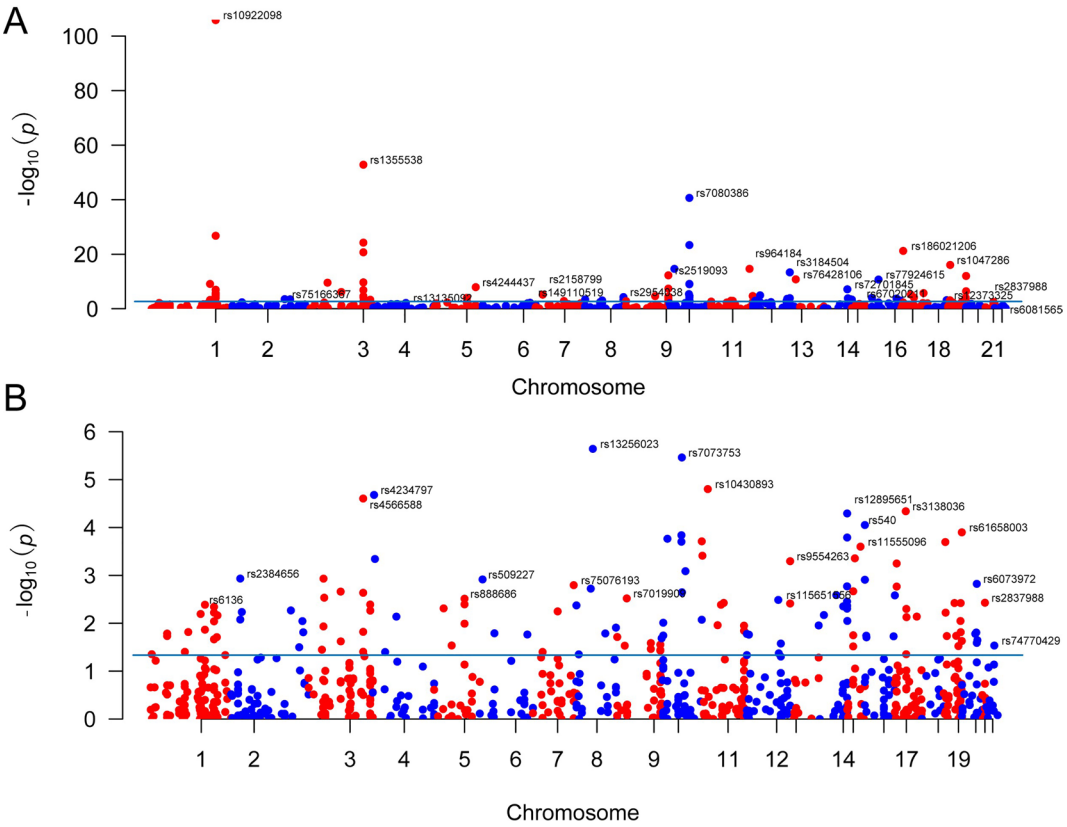

Fig. 1 Mendelian randomization analysis. **a** Manhattan plot showing the significance of associations between all instrumental variables (IVs) from the 7pqtl proteomic dataset and cervical cancer incidence. **b** Manhattan plot displaying the significance of associations between all IVs from the 9pqtl proteomic dataset and cervical cancer incidence

receptor ligand activity was most significantly enriched, serving as a mode of interaction between cells and their external environment. By regulating receptor-ligand binding, cells can appropriately respond to external stimuli, thereby modulating their function and metabolic status (Fig. 2b). In the subsequent KEGG enrichment analysis, we displayed the top 30 enriched pathways, with the Cytokine-cytokine receptor interaction pathway containing the most candidate genes and showing the highest significance (Fig. 2c). Additionally, Reactome pathway enrichment analysis revealed that 10 reactome-related pathways were significantly enriched with candidate genes, and the interaction relationships among these 10 pathways are illustrated in Fig. 2d.

3.3 Identification of key metabolic genes and their co-expression modules in cervical cancer

To identify key metabolic genes associated with cervical cancer for potential prognostic biomarkers, we conducted a weighted gene co-expression network analysis (WGCNA) using the TCGA-CESC dataset. A soft threshold of 4 was applied to construct a scale-free network (Fig. 3a), followed by the generation of a topological matrix and hierarchical clustering to identify gene modules (Fig. 3b). Modules with at least 50 genes were defined, and similar modules were merged based on module eigengenes. This process resulted in 13 distinct modules, with unassigned genes grouped in the grey module. Among the 13 modules, two showed statistically significant differences in eigengenes, excluding the grey module (Fig. 3c). Analysis of Module Membership (MM) and Gene Significance (GS) revealed a positive correlation in these two modules, indicating that highly prognostic genes are central to these modules. From these, we identified 337 key genes for further investigation (Fig. 3d).

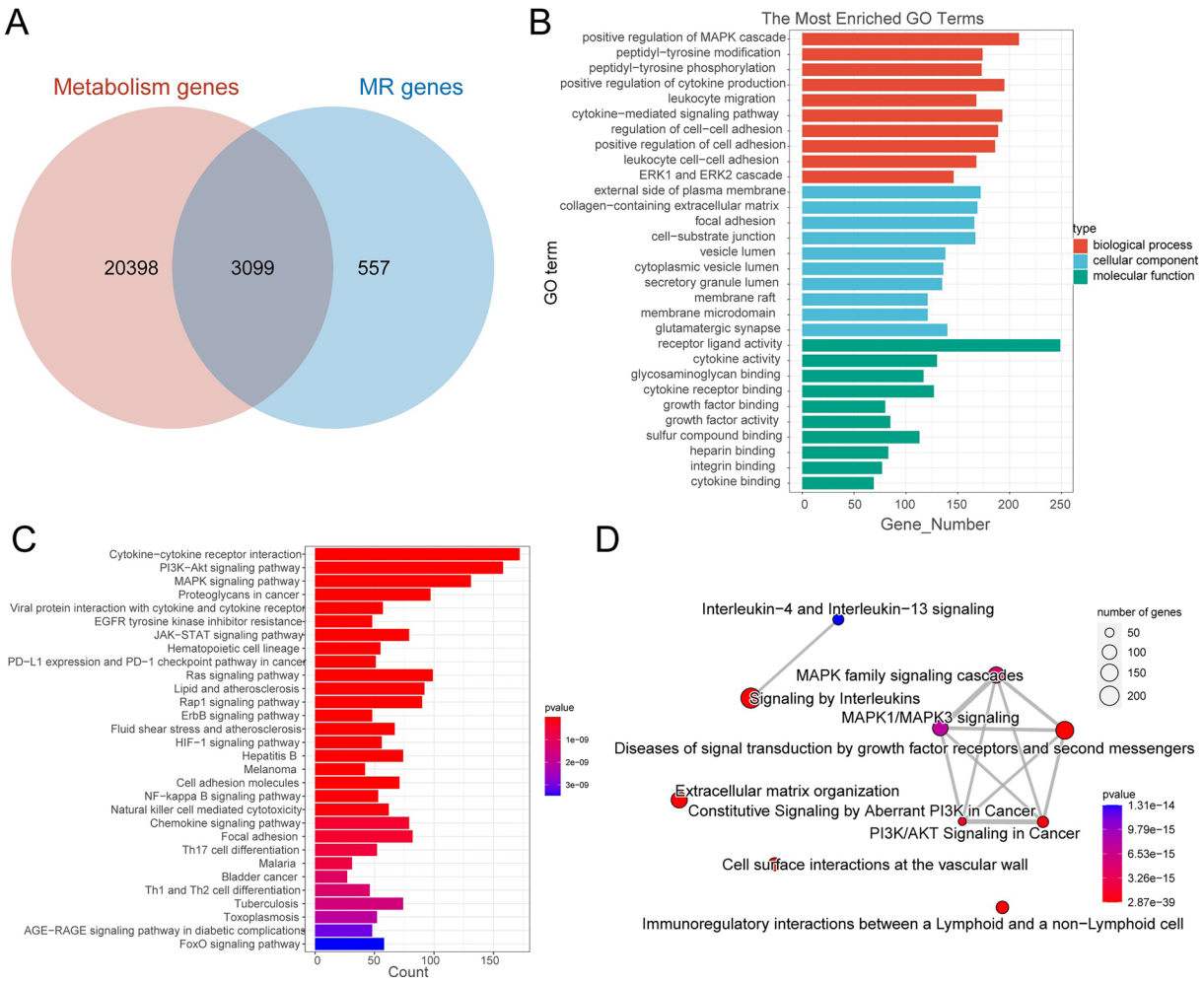

Fig. 2 Functional enrichment of metabolism-related genes with positive causal relationships to cervical cancer. **a** Venn diagram showing the intersection of metabolism-related genes with genes having a positive causal relationship with cervical cancer incidence. **b** Bar chart of Gene Ontology (GO) enrichment analysis. **c** Bar chart of Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. **d** Network diagram of Reactome pathway enrichment analysis

3.4 Development of a cervical cancer prognostic model using multiple machine learning approaches

The identification of key metabolism-related genes is crucial for advancing the development of metabolic biomarkers for predicting cervical cancer. To identify these key genes, we conducted a univariate Cox regression analysis within the set of key genes to further select features associated with prognosis. We identified 112 prognostic-related key genes with significant univariate Cox analysis results ($p < 0.05$) from the TCGA-CESC dataset for model construction. Using the TCGA-CESC dataset as the training set and GSE52903 as the validation set, we then concentrated on creating the best prognostic model by feeding these 112 prognostic-related important genes into several machine learning models. The combination of Random Survival Forests (RSF) and Supervised Principal Component Analysis (SuperPC), as seen in Fig. 4a, was chosen as the final model because it produced the highest C-index (0.698) in the validation set out of 116 machine learning techniques.

3.5 Development of a prognostic model for cervical cancer patients using a combination of lasso and RSF machine learning

Based on a comparison of C-indexes across 112 machine learning model combinations, we selected the optimal combination of the Random Survival Forest (RSF) and Supervised Principal Component Analysis (SuperPC) algorithms to

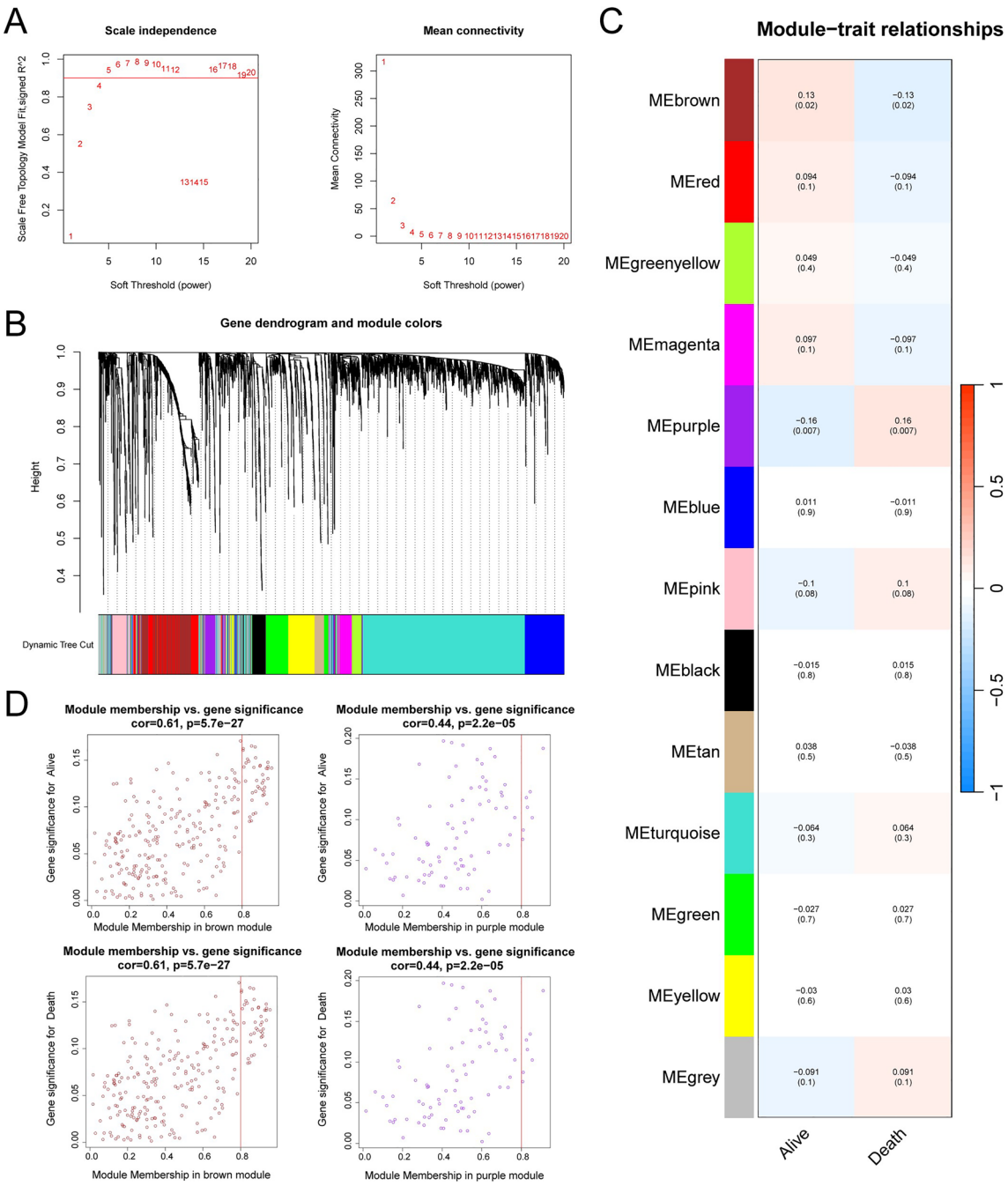

Fig. 3 Identification of key genes via WGCNA. **a** Determination of the optimal soft threshold for WGCNA, set at soft power = 4. **b** WGCNA clustering dendrogram. **c** Thirteen modules identified by WGCNA, each row representing genes within a module, with two columns showing different prognostic outcomes for cervical cancer. Each cell contains the corresponding correlation and *p*-value, colored according to a correlation color scale. **d** Scatter plots of Gene Module Membership (MM) and Gene Significance (GS) illustrating the relationship between genes and modules, and between genes and traits

construct a metabolism-related prognostic model for cervical cancer patients. The RSF algorithm further refined the selection of key genes, which were then utilized to develop the SuperPC prognostic model. Ultimately, 22 model genes were selected to establish the optimal model. These 22 key genes were used to stratify subgroups based on the median expression levels, and prognostic analyses were conducted for each subgroup. The Kaplan–Meier (K–M) survival curves for each subgroup, as shown in Figures S1A–L, demonstrated significant prognostic differences. Based on the expression of these 22 model genes, a risk score (Metabolic Score, MS) was computed for each patient using the SuperPC model.

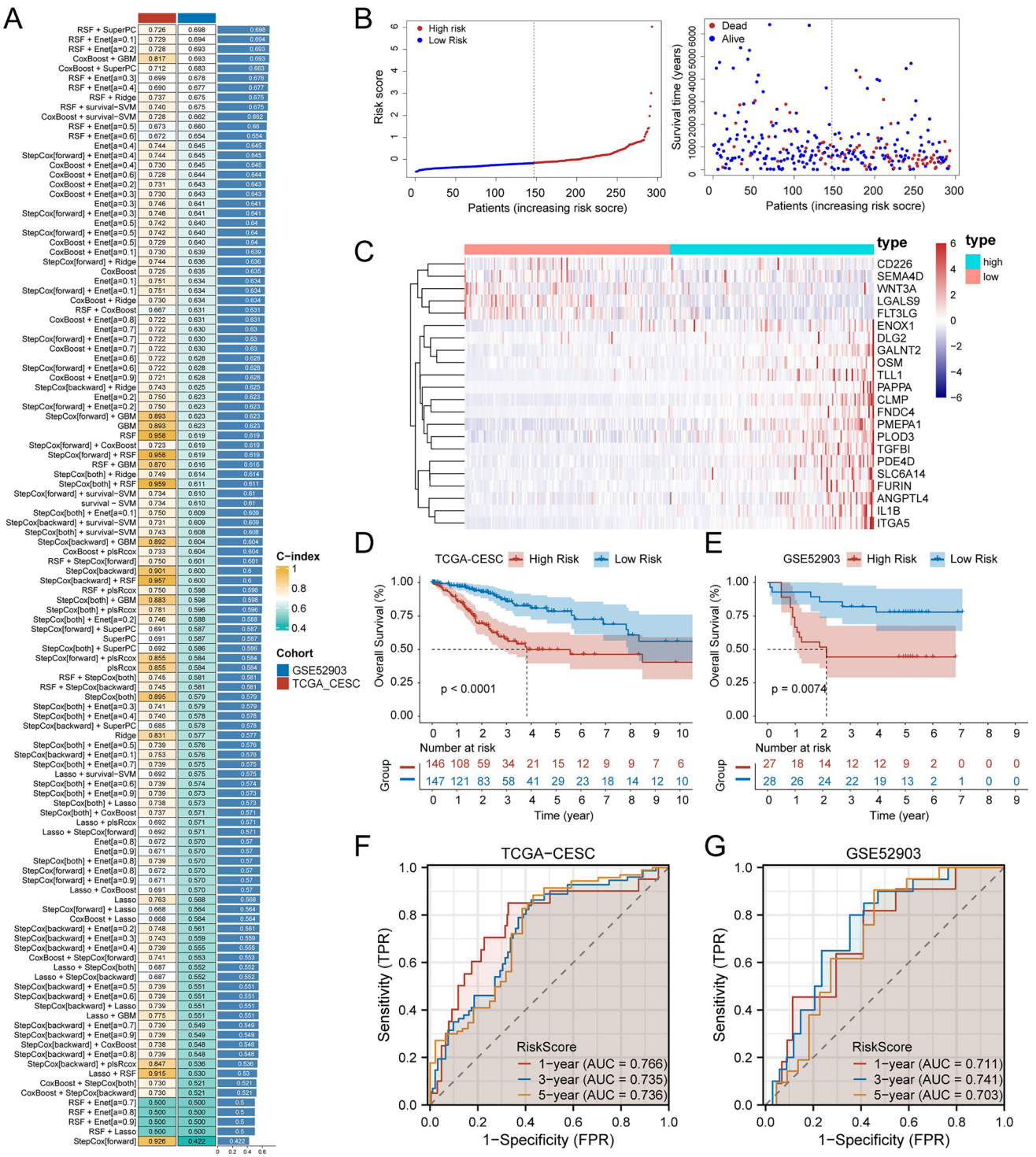

Fig. 4 Selection of key genes and prognostic models. **a** Heatmap of the C-index for 112 machine learning model combinations based on key genes in the training dataset TCGA-CESC and the validation dataset GSE52903. **b** Distribution of risk scores and survival status for cervical cancer patients derived from the SuperPC model. **c** Heatmap of gene expression for features identified by the SuperPC model

Based on the median MS, patients were classified as high or low metabolic risk groups. A distribution chart of patient death statuses showed that the high-risk group had a higher mortality rate (Fig. 4b). Figure 4c displays the risk factor expression heatmaps developed between the various risk groups, indicating that the majority of genes were expressed at higher levels in the high-risk group. Figure 4d shows that the K–M curves indicate a substantial difference in prognoses between patients in the high-risk and low-risk groups ($p < 0.0001$). This conclusion was confirmed in the GSE52903 validation dataset ($p = 0.0074$) (Fig. 4e). Subsequently, time-ROC curves for prognosis were plotted based on patient risk scores and survival statuses. In the TCGA-CESC dataset, the AUC values for MS in predicting patient outcomes were consistently above 0.73 (Fig. 4f); in the validation dataset GSE52903, the results were reliably confirmed with AUC values above 0.7 (Fig. 4g), indicating that MS is an effective prognostic model for cervical cancer patients.

3.6 Differences in immune infiltration among cervical cancer patients of varying risk levels

Tumor immune activity plays a crucial role in the development and progression of cervical cancer. We assessed the degree of tumor immune infiltration in cervical cancer patients using ESTIMATE analysis, which provides an overall score for immune cells, stromal cells, and tumor purity. We evaluated the disparities in scores between various risk groups using Wilcoxon tests. According to the data, the high-risk group had a larger concentration of stromal cells (Fig. 5b) and a much lower immune cell infiltration (Fig. 5a) than the low-risk group. Additionally, there was a trend toward lower overall immunological ESTIMATE scores (Fig. 5d) and higher tumor purity in the high-risk group (Fig. 5c), but there was no statistically significant difference in either measure for either tumor purity or overall ESTIMATE scores between the two groups.

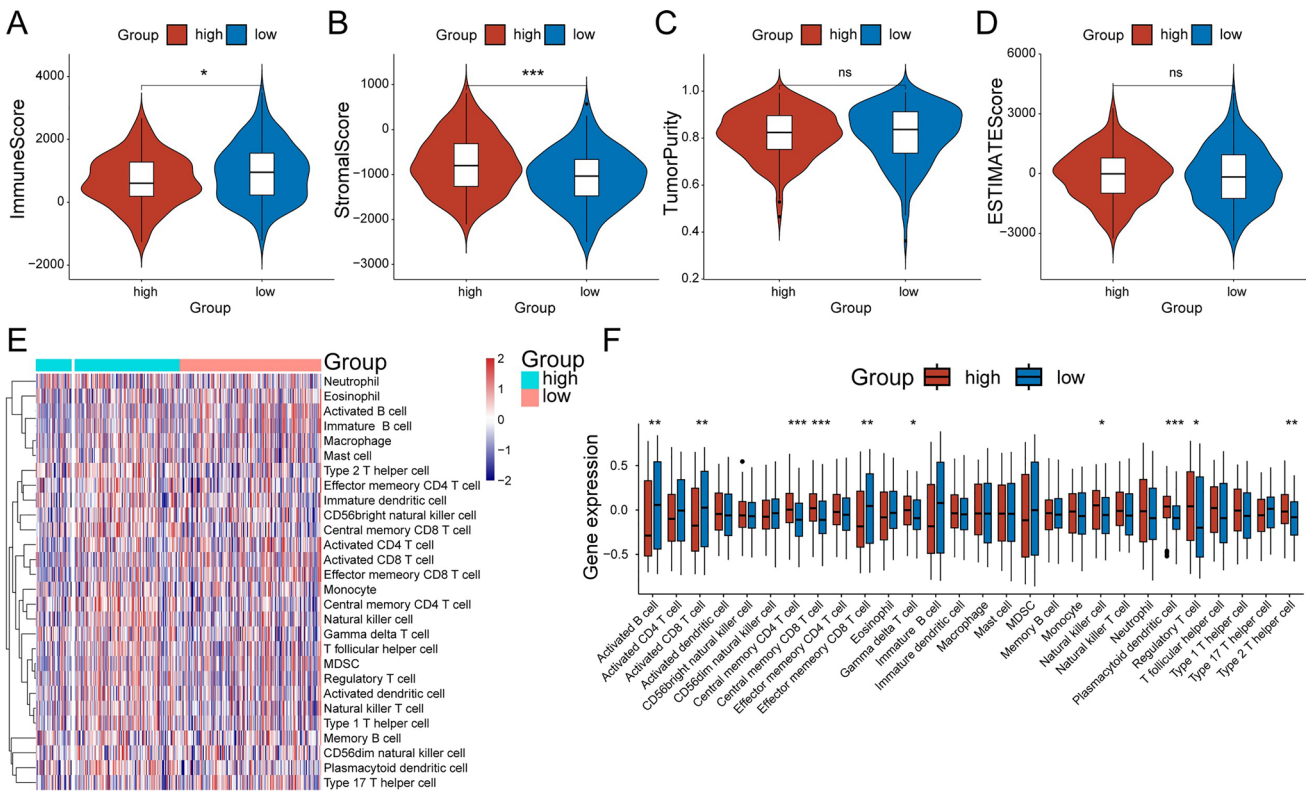

Fig. 5 Analysis of immune cell distribution in high- and low-risk groups of cervical cancer patients. **a** Differences in immune scores between high- and low-risk groups. **b** Differences in stromal scores between high- and low-risk groups. **c** Differences in tumor purity scores between high- and low-risk groups. **d** Differences in ESTIMATE immune status scores between high- and low-risk groups. **e** Heatmap showing differences in the abundance of 28 immune cells between high- and low-risk groups based on ssGSEA immune infiltration analysis. **f** Box plot showing differences in the abundance of 28 immune cells between high- and low-risk groups based on ssGSEA immune infiltration analysis. *Indicates $p < 0.05$, **indicates $p < 0.01$, ****indicates $p < 0.001$. **d** Kaplan–Meier (KM) curves for patients in the TCGA-CESC training set stratified into high- and low-risk groups, showing significantly worse prognosis in the high-risk group. **e** Kaplan–Meier (KM) curves for patients in the TCGA-CESC training set, emphasizing the pronounced prognostic difference between high- and low-risk groups. **f** Time-dependent Receiver Operating Characteristic (ROC) curves demonstrating the prognostic prediction efficacy of the Metabolic Score (MS) in the TCGA-CESC dataset. **g** Time-dependent ROC curves demonstrating the prognostic prediction efficacy of MS in the GSE52903 dataset

To gain more detailed insights into the differences in immune cell abundance, we also employed the ssGSEA method to assess the abundance of 28 types of immune cells (Fig. 5e). Ten distinct immune cell types showed significant differences in abundance between the high- and low-risk groups. The low-risk group had higher concentrations of effector memory CD8 + T cells, active B cells, and CD8 + T cells, all of which are essential for immunological responses. Active B cells are involved in antibody production, while active CD8 + T cells and effector memory CD8 + T cells participate in combating pathogen infections and clearing tumor cells, potentially reflecting a normal or healthier immune response that could help control the progression of cervical cancer or enhance responsiveness to treatment. Conversely, central memory CD4 + T cells, central memory CD8 + T cells, γδ T cells, natural killer cells, plasmacytoid dendritic cells, regulatory T cells, and type 2 helper T cells showed significant differences in abundance, being more prevalent in high-risk cervical cancer patients. These cell types may indicate an abnormal state of the immune system or phenomena of immune escape, associated with the development and worsening of cervical cancer. For example, regulatory T cells may be involved in immune suppression, and natural killer cells might have a limited effect in tumor clearance (Fig. 5f).

3.7 Spatial transcriptomic visualization of tumor tissue spots and their prognostic impact

To investigate whether the Metabolic Score (MS) can differentiate the prognostic impact of various tumor tissue regions at a spatial level, we conducted an in-depth exploration using spatial transcriptomics. Through the characterization of individual spots within the spatial transcriptome, we identified regions marked in Fig. 6a as keratin pearls, a structure commonly identified in pathology. These regions exhibited lower total RNA counts (Figure S2A) and fewer gene numbers (Figure S2B), yet they had higher MS risk scores (Fig. 6b). The majority of high-risk spots were also located in these areas (Fig. 6c). This suggests that MS may also be capable of identifying tumor-specific regions that impact patient prognosis, potentially aiding in the development of precision treatment strategies.

In additional, we observed that spots with different attributes tend to cluster at specific visible locations, suggesting significant heterogeneity in the cellular characteristics within these clustered areas. To further analyze this, we utilized single-cell sequencing data from two cervical squamous cell carcinoma patients available in the GSE208653 dataset.

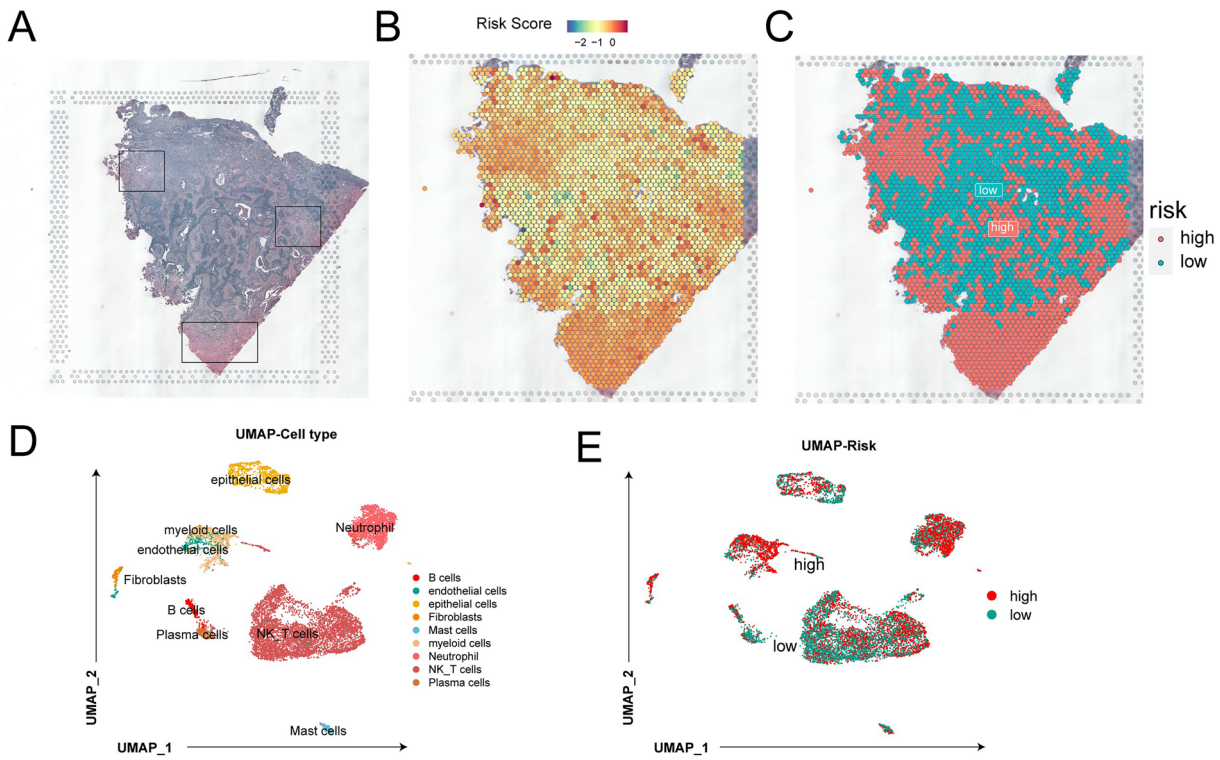

Fig. 6 Spatial transcriptomics and single-cell analysis. **a** Cervical cancer patient tissue section with spatial orientation. **b** Feature plot showing the distribution of Metabolic Score (MS) across different spots in the spatial transcriptome. **c** Risk attributes of different spots identified based on MS in the spatial transcriptome. **d** UMAP plot illustrating nine distinct cell types identified in the single-cell dataset. **e** UMAP plot showing different cell risk attributes in the single-cell dataset

After excluding low-quality cells through a single-cell quality control process (Figures S2C–F), we identified a total of 10,920 cells. Normalization processes were applied to mitigate batch effects between samples (Figure S1G). The cells were then subjected to unsupervised clustering, resulting in the identification of 22 distinct cell clusters (Figure S2H).The primary cell types, such as B cells, endothelial cells, epithelial cells, fibroblasts, macrophages, myeloid cells, neutrophils, NK T cells, and plasma cells, were identified through manual annotation of the cells based on the expression of marker genes (Figure S2I) (Fig. 6d). Using machine learning techniques, we assessed each cell's MS score and divided them into high-risk and low-risk categories (Fig. 6e). To gain more detailed insights into the differences in immune cell abundance, we also employed the ssGSEA method to assess the abundance of immune cells (Figure S3).

3.8 Single-cell data reveals heterogeneity among cells with different risk attributes

In the high-risk cell clusters, extensive communication was observed among various cells (Figure S4A), especially between epithelial cells and fibroblasts, myeloid cells, and others (Figure S4B). Pseudotime analysis distinguished the cells into five distinct states (Figure S4C, D), revealing that cells in state 5 are typically found early in the timeline and gradually transition to state 3 over time. The MS model gene pseudotime heatmap showed that high expression of 17 genes dynamically changed over time within the high-risk group cell cluster, indicating significant expression changes across different pseudotime periods (Figure S4E).

Compared to the high-risk cell clusters, cell communication in the low-risk clusters was generally weaker (Fig. 7a), with epithelial cells showing interactions with myeloid cells, endothelial cells, neutrophils, and mast cells, among others, but at overall lower intensities (Fig. 7b). Pseudotime analysis allowed for the differentiation of cells into six distinct states (Fig. 7c, d), suggesting that cells in state 1, found early in the timeline, gradually evolve into state 7, and eventually differentiate into states 3, 5, and 6. This indicates that low-risk group cells exhibit a broader range of temporal states. The heatmap of MS model gene expression changes over pseudotime in the low-risk group also revealed dynamic changes, with 14 genes showing varied levels of high expression over different time periods (Fig. 7e). The extent and duration of high expression differed from those observed in high-risk group cells, highlighting differences in gene expression dynamics between the groups.

3.9 PLOD3 as a novel oncogene in cervical cancer

To elucidate the functional role of PLOD3 in cervical cancer, we employed both clone formation assays and Ki67 immunofluorescence staining. The clone formation assay (Fig. 8a), Ki67 immunofluorescence staining (Fig. 8b) and CCK8 results revealed that silencing PLOD3 expression markedly reduced the clonogenic capacity of cervical cancer cells.

4 Discussion

Our study successfully constructed and validated a multi-machine learning prognostic model based on metabolic genes for CC. The identification of key metabolic genes that are associated with cervical cancer will help to further develop prognostic biomarkers for cervical cancer. More importantly, our model leveraged the extensive capabilities of machine learning to analyze complex datasets, surpassing traditional statistical methods in predicting survival outcomes of CC. By incorporating metabolic genes associated with prognosis, our model provides a more comprehensive understanding of the factors influencing cervical cancer progression and patient survival. The validation of our model using single-cell and spatial transcriptomics further underscores its robustness and applicability.

Our findings highlight the significant role of metabolic pathways in shaping the tumor microenvironment and influencing immune cell activity [17]. The ESTIMATE analysis demonstrated that high-risk cervical cancer patients have significantly lower immune cell infiltration compared to low-risk patients. This reduced infiltration might indicate a compromised immune surveillance in high-risk patients, potentially facilitating tumor growth and metastasis [18]. On the other hand, the higher content of stromal cells in the high-risk group could contribute to a more supportive microenvironment for tumor progression, as stromal cells are known to interact with cancer cells, promoting their survival and dissemination [19].

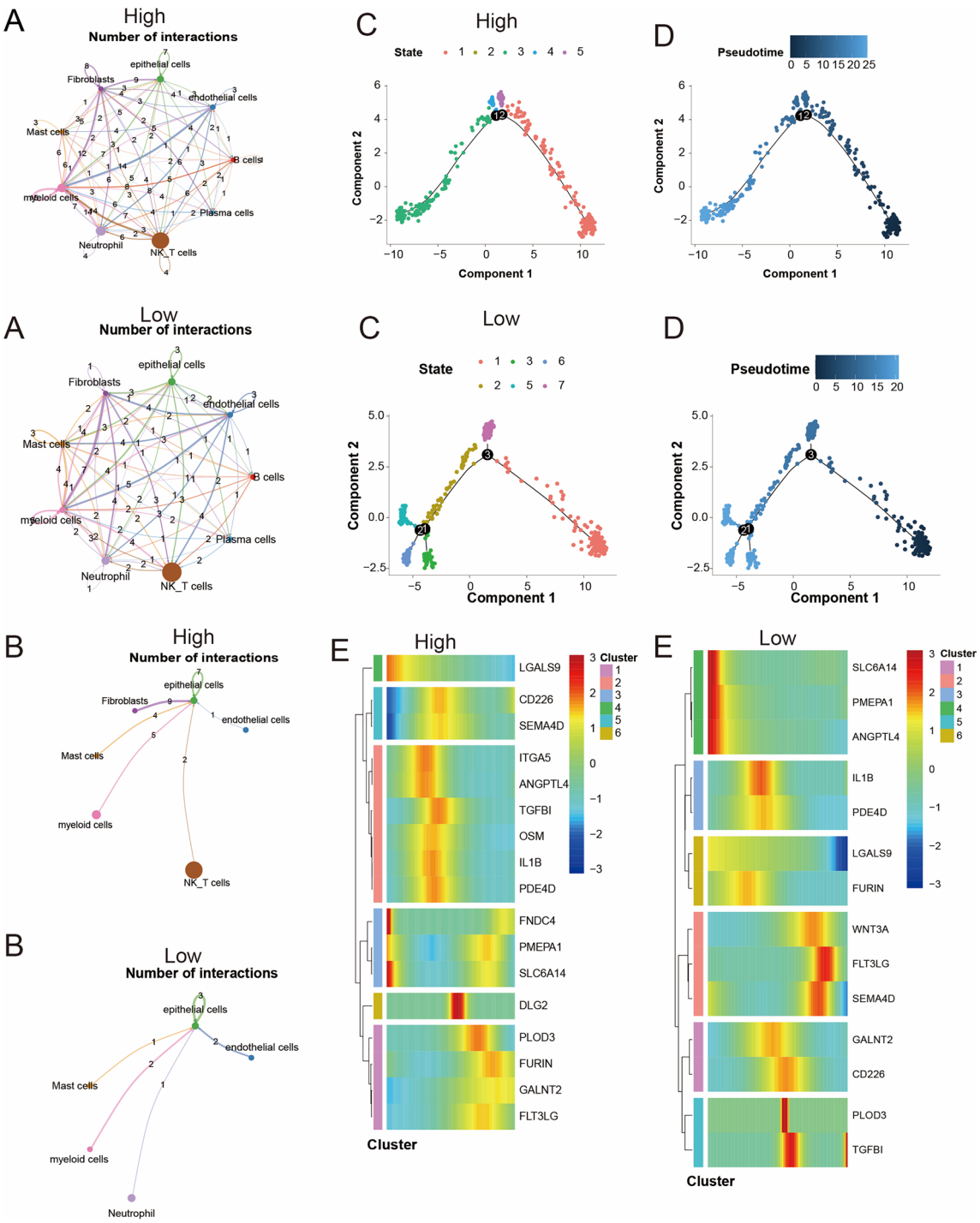

Fig. 7 Cell communication and pseudotime analysis in low-risk cell clusters. **a** Network depicting cell communication within the low-risk group. **b** Communication patterns between epithelial cells and other cell types in the low-risk group. **c** Identification of five distinct pseudotime states in low-risk group cells, organized by pseudotime events. **d** Pseudotime trajectory map for cells in the low-risk group. **e** Heatmap showing the expression changes of MS model genes over pseudotime in the low-risk group

Furthermore, active B cells, active CD8 + T cells, and effector memory CD8 + T cells were found to be more abundant in the low-risk group. These cells play critical roles in adaptive immunity, with active B cells producing antibodies and CD8 + T cells targeting and eliminating infected and tumor cells [20]. Their higher abundance in low-risk patients suggests a more robust and effective immune response, which could be instrumental in controlling tumor growth

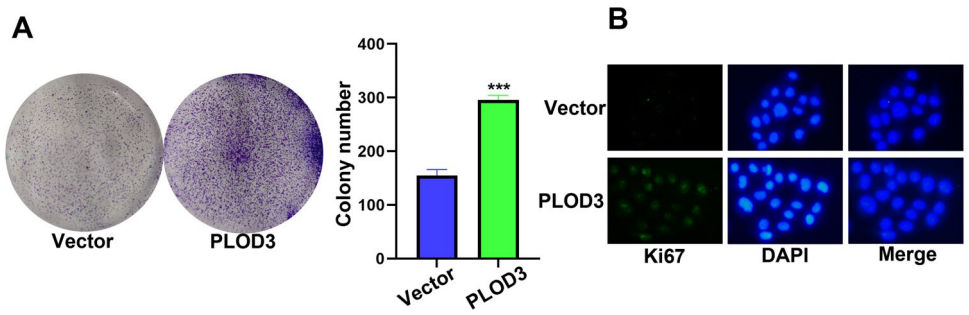

Fig. 8 Analysis of PLOD3 expression and functional role in cervical cancer. **a** Clone formation assay demonstrating a significant reduction in the clonogenic capacity of cervical cancer cells following PLOD3 knockdown. **b** Ki67 immunofluorescence staining illustrating decreased proliferative activity in cervical cancer cells with silenced PLOD3 expression. **c** CCK8 results revealed that silencing PLOD3 expression markedly reduced the clonogenic capacity of cervical cancer cells

and enhancing the efficacy of cancer treatments. In contrast, high-risk patients exhibited higher levels of central memory CD4 + T cells, γδ T cells, natural killer cells, regulatory T cells. This pattern points towards a more dysfunctional or suppressed immune state. For instance, regulatory T cells are known for their role in suppressing immune responses, which can aid in tumor immune escape [21]. Similarly, the increased presence of natural killer cells and other immune cell types in high-risk patients may reflect an attempted but ineffective immune response, as these cells might be functionally impaired or unable to overcome the immunosuppressive tumor microenvironment [22]. Understanding these immune differences provides valuable insights for developing targeted therapies. For low-risk patients, enhancing the existing robust immune response might be a viable strategy, potentially through immune checkpoint inhibitors or vaccines that boost the activity of CD8 + T cells and B cells. For high-risk patients, strategies to modify the tumor microenvironment, such as reducing stromal support or reversing immune suppression, might be necessary to make the tumors more susceptible to immune-mediated destruction.

In addition, our findings highlight the utility of spatial transcriptomics in delineating tumor heterogeneity and its prognostic impact. The ability to map high-risk metabolic scores to specific tumor regions such as keratin pearls can inform precision medicine strategies. By identifying and targeting these high-risk areas, treatment can be tailored to address the most aggressive parts of the tumor, potentially improving patient outcomes. Additionally, our analysis of single-cell data reveals significant heterogeneity among cells with varying risk attributes in cervical cancer, providing insights into cellular communication and gene expression dynamics across different risk groups.

In high-risk cell clusters, extensive intercellular communication was observed, particularly between epithelial cells and fibroblasts, myeloid cells, and other cell types. This extensive communication network suggests a highly interactive tumor microenvironment [23], which may facilitate tumor progression and immune evasion. Compared to high-risk clusters, cell communication in low-risk clusters was generally weaker. This reduced intercellular communication might reflect a less aggressive tumor microenvironment [24], potentially leading to better prognosis and response to treatment. In addition, pseudotime analysis further distinguished the high-risk cells into five distinct states, revealing a dynamic progression of cellular states. The strong intercellular interactions and temporal progression of cellular states indicate a highly adaptive tumor microenvironment [25], which may contribute to treatment resistance and poor prognosis. Thus, targeting the key genes and pathways identified in high-risk clusters might disrupt these interactions and improve therapeutic outcomes. In contrast, the weaker cell communication and broader range of temporal states in low-risk clusters suggest a less aggressive tumor behavior, which may be more amenable to treatment.

Despite the promising results, there are limitations to our study that warrant further investigation. The complexity of metabolic interactions and their impact on immune responses require more detailed exploration to fully elucidate the underlying mechanisms. Additionally, while our model showed high accuracy in predicting patient outcomes, its clinical utility needs to be further validated in larger, independent cohorts to ensure its generalizability.

In conclusion, our study provides a novel and effective prognostic model for cervical carcinoma, emphasizing the critical role of metabolic pathways in tumor progression and immune response. The integration of machine learning and advanced genomic techniques offers a powerful approach to improve prognosis prediction and guide therapeutic strategies. Future research should continue to explore the intricate interactions between metabolism and immunity to develop more targeted and effective treatments for cervical cancer patients.

Acknowledgements We express our gratitude to the TCGA and GEO databases for providing access to the data.

Author contributions QLL conceptualised and designed the study. QX conducted the experiments. YX analysed the data and prepared the initial manuscript draft. All authors reviewed and approved the final manuscript.

Funding No fund.

Data availability The datasets used or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

Gene list LGALS9; SLC6A14; CD226; PMEPA1; SEMA4D; ANGPTL4; ITGA5; IL1B; ANGPTL4; TGFBI; PDE4D; OSM; LGALS9; IL1B; PDE4D; FURIN; FNDC4; WNT3A; PMEPA1; FLT3LG; SLC6A14; SEMA4D; DLG2; PLOD3; GALNT2; FURIN; CD226; GALNT2; PLOD3; FLT3LG; TGFBI21.

References

1. Chibwesha CJ, Stringer JSA. Cervical cancer as a global concern: contributions of the dual epidemics of HPV and HIV. JAMA. 2019;322(16):1558–60.
2. Choi S, Ismail A, Pappas-Gogos G, Boussios S. HPV and cervical cancer: a review of epidemiology and screening uptake in the UK. Pathogens. 2023;12(2):298.
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
4. Monk BJ, Tan DSP, Hernández Chagüi JD, Takyar J, Paskow MJ, Nunes AT, Pujade-Lauraine E. Proportions and incidence of locally advanced cervical cancer: a global systematic literature review. Int J Gynecol Cancer. 2022;32(12):1531–9.
5. Burmeister CA, Khan SF, Schäfer G, Mbatani N, Adams T, Moodley J, Prince S. Cervical cancer therapies: current challenges and future perspectives. Tumour Virus Res. 2022;13: 200238.
6. Jang M, Kim SS, Lee J. Cancer cell metabolism: implications for therapeutic targets. Exp Mol Med. 2013;45(10): e45.
7. Zhang X, Song W, Gao Y, Zhang Y, Zhao Y, Hao S, Ni T. The role of tumor metabolic reprogramming in tumor immunity. Int J Mol Sci. 2023;24(24):17422.
8. Allison KE, Coomber BL, Bridle BW. Metabolic reprogramming in the tumour microenvironment: a hallmark shared by cancer cells and T lymphocytes. Immunology. 2017;152(2):175–84.
9. Kedia-Mehta N, Finlay DK. Competition for nutrients and its role in controlling immune responses. Nat Commun. 2019;10(1):2123.
10. Newsholme P. Cellular and metabolic mechanisms of nutrient actions in immune function. Nutr Diabetes. 2021;11(1):22.
11. Newsholme P. Cellular and metabolic mechanisms of nutrient actions in immune function. Eur J Clin Nutr. 2021;75(9):1328–31.
12. Wang ZH, Peng WB, Zhang P, Yang XP, Zhou Q. Lactate in the tumour microenvironment: from immune modulation to therapy. EBioMedicine. 2021;73: 103627.
13. Dai G, Ou J, Wu B. A predictive study of metabolism reprogramming in cervical carcinoma. Ann Transl Med. 2022;10(7):414.
14. Schmidt DR, Patel R, Kirsch DG, Lewis CA, Vander Heiden MG, Locasale JW. Metabolomics in cancer research and emerging applications in clinical oncology. CA Cancer J Clin. 2021;71(4):333–58.
15. Habehh H, Gohel S. Machine learning in healthcare. Curr Genomics. 2021;22(4):291–300.
16. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina (Kaunas). 2020;56(9):455.
17. Cassim S, Pouyssegur J. Tumor microenvironment: a metabolic player that shapes the immune response. Int J Mol Sci. 2019;21(1):157.
18. Janssen LME, Ramsay EE, Logsdon CD, Overwijk WW. The immune system in cancer metastasis: friend or foe? J Immunother Cancer. 2017;5(1):79.
19. Monteran L, Zait Y, Erez N. It's all about the base: stromal cells are central orchestrators of metastasis. Trends Cancer. 2024;10(3):208–29.
20. Raskov H, Orhan A, Christensen JP, Gögenur I. Cytotoxic CD8+ T cells in cancer and cancer immunotherapy. Br J Cancer. 2021;124(2):359–67.
21. Facciabene A, Motz GT, Coukos G. T-regulatory cells: key players in tumor immune escape and angiogenesis. Cancer Res. 2012;72(9):2162–71.

22. Dean I, Lee CYC, Tuong ZK, Li Z, Tibbitt CA, Willis C, Gaspal F, Kennedy BC, Matei-Rascu V, Fiancette R, Nordenvall C, Lindforss U, Baker SM, Stockmann C, Sexl V, Hammond SA, Dovedi SJ, Mjösberg J, Hepworth MR, Carlesso G, Clatworthy MR, Withers DR. Rapid functional impairment of natural killer cells following tumor entry limits anti-tumor immunity. Nat Commun. 2024;15(1):683.
23. Dominiak A, Chełstowska B, Olejarz W, Nowicka G. Communication in the cancer microenvironment as a target for therapeutic interventions. Cancers (Basel). 2020;12(5):1232.
24. Son B, Lee S, Youn H, Kim E, Kim W, Youn B. The role of tumor microenvironment in therapeutic resistance. Oncotarget. 2017;8(3):3933–45.
25. Barkley D, Moncada R, Pour M, Liberman DA, Dryg I, Werba G, Wang W, Baron M, Rao A, Xia B, França GS, Weil A, Delair DF, Hajdu C, Lund AW, Osman I, Yanai I. Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. Nat Genet. 2022;54(8):1192–201.