Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrated bulk and single-cell transcriptomics identifies matrix- stiffness-associated prognostic and diagnostic signatures in clear-cell renal cell carcinoma

Yongkang Xiao
Longhui County People's Hospital (Longhui Hospital, Jishou University)
Xiuqi Luo
Longhui County People's Hospital (Longhui Hospital, Jishou University)
Fei Hu
Longhui County People's Hospital (Longhui Hospital, Jishou University)
Chen Luo
Longhui County People's Hospital (Longhui Hospital, Jishou University)
Feiyun Wu
Longhui County People's Hospital (Longhui Hospital, Jishou University)
Hui Chen
Longhui County People's Hospital (Longhui Hospital, Jishou University)
Chenghui Li
Longhui County People's Hospital (Longhui Hospital, Jishou University)
Renyu Gong
Longhui County People's Hospital (Longhui Hospital, Jishou University)
Chengyan Guo
178739412990163.com

Longhui County People's Hospital (Longhui Hospital, Jishou University)

Research Article

Keywords: clear-cell renal cell carcinoma, matrix stiffness, prognostic model, diagnostic classifier, single-cell RNA sequencing

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11662614/v1

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

Additional Declarations:
No competing interests reported.

Tables 1 to 3 are available in the Supplementary Files section.

Integrated bulk and single-cell transcriptomics identifies matrix-stiffness-associated prognostic and diagnostic signatures in clear-cell renal cell carcinoma

Yongkang Xiao[1,†], Xiuqi Luo[1,†], Fei Hu[1], Chen Luo[1], Feiyun Wu[1], Hui Chen[1], Chenghui Li[1], Renyu Gong[1], Chengyan Guo[1,#]

[1]Department of Urology, Longhui County People's Hospital (Longhui Hospital Affiliated to Jishou University), Shaoyang 422200, Hunan, China

[†]These authors contributed equally to this work.

[#]Correspondence address. Tel: 17873941299; E-mail: 17873941299@163.com

Abstract

Background: Matrix stiffness is a biophysical feature of the tumour microenvironment linked to malignant behaviour, immune regulation and treatment response, but its transcriptional correlates in clear-cell renal cell carcinoma (ccRCC) remain insufficiently characterised. We investigated whether a previously reported matrix-stiffness-associated transcriptional set could support parallel prognostic stratification and tumour-normal diagnostic classification in ccRCC, and resolved representative genes at single-cell resolution.

Methods: TCGA-KIRC tumour-normal differentially expressed genes were intersected with a previously reported 1,417-gene matrix-stiffness-associated transcriptional set, yielding 203 overlapping genes for GO and KEGG enrichment analysis using clusterProfiler. Univariable Cox regression identified 96 overall-survival-associated candidates, which were used as input for STRING-based protein-protein interaction analysis. The 30 genes with the highest network node degree entered 117 survival-learning workflows. TCGA-KIRC served as the model development cohort, whereas E-MTAB-1980, CPTAC and GSE167573 contributed to cross-cohort workflow selection and benchmarking rather than serving as untouched independent validation cohorts. In parallel, WGCNA tumour-normal modules were intersected with the prognosis-associated candidate set, and the resulting genes were evaluated in 103 diagnostic-classification workflows. Model attribution, expression and outcome evidence were integrated to prioritise COL4A4 from the prognostic signature and ADA from the diagnostic classifier. Clinical, microenvironmental and genomic associations were characterised, followed by cell-type localisation, pseudotime analysis and virtual perturbation in a 102,368-cell ccRCC atlas.

Results: Among 1,353 differentially expressed genes, 203 overlapped with the matrix-stiffness-associated transcriptional set and 96 were associated with overall survival at nominal $P<0.05$. Five survival-learning workflows shared the highest mean C-index (0.747) across the three cross-cohort workflow-selection datasets. Lasso plus SuperPC yielded an eleven-gene prognostic model with cohort-specific C-indices of 0.627, 0.690, 0.739 and 0.811 in TCGA-KIRC, E-MTAB-1980, CPTAC and GSE167573, respectively. Because the three external cohorts contributed to workflow selection, their performance estimates are benchmarking results rather than independent validation estimates. High-risk patients had shorter survival in all four cohorts, and the score remained associated with outcome after adjustment for age and pathologic N and M categories. COL4A4 combined high model attribution, a strong inverse association with risk (rho=-0.87), tumour downregulation and favourable gene-level survival. The parallel WGCNA diagnostic branch reduced 722 tumour-normal module genes to 15 candidates and generated an ADA, CCL5, HAVCR2, LGALS1 and LGALS9 classifier with a mean AUC of 0.895. CCL5 had the highest classifier SHAP contribution, whereas ADA showed a significant adverse survival association ($P=0.007$) and was selected for cell-resolved investigation. In the single-cell atlas, COL4A4 showed its strongest class-level transcript signal in B cells and declined along pseudotime, whereas ADA was concentrated in immune compartments, particularly T and NK cells, and followed a non-linear rise-and-fall trajectory. Virtual COL4A4 perturbation produced B-cell-dominant responses, while virtual ADA loss and gain of function converged on NK cells.

Conclusions: In ccRCC, the transferred matrix-stiffness-associated gene set supported an overall-survival model and a separate tumour-normal classifier. COL4A4 emerged from the prognostic analysis, whereas ADA was selected for cell-resolved follow-up from the diagnostic analysis. Both candidates warrant evaluation in independent cohorts and focused experimental systems.

Keywords: clear-cell renal cell carcinoma; matrix stiffness; prognostic model; diagnostic classifier; single-cell RNA sequencing

1.Introduction

Clear-cell renal cell carcinoma (ccRCC) is characterised by extensive genomic, metabolic, spatial and immune heterogeneity, which limits uniform risk assessment and contributes to variable benefit from contemporary systemic therapy[1-9]. Proteogenomic and spatial studies have further shown that aggressive behaviour emerges from coordinated tumour-cell and microenvironmental states rather than from a single molecular lesion[8-11]. Molecular models therefore require both reproducible evaluation across cohorts and a biological structure that can be interpreted at gene and cell levels.

The ECM is an active regulator of tumour behaviour. Its biochemical composition, collagen architecture, viscoelasticity and stiffness influence receptor signalling, metabolism, invasion, immune-cell access and therapeutic response[12-21]. In ccRCC, increased spatial coupling of integrin and collagen-IV-associated signals has been observed in immunoresistant regions, while stiffness-responsive PIEZO1-Ca2+/calpain/YAP signalling promotes malignant phenotypes [22-23]. Basement-membrane-related signatures also carry prognostic and immune information in ccRCC[24-25], and recent experimental evidence links COL4A4 loss to JAK2/STAT3 activation, tumour progression and microenvironmental reprogramming[26].

Recent ccRCC studies increasingly integrate bulk, single-cell and spatial transcriptomics to resolve malignant epithelial, metabolic, vascular, fibroblast and immune programmes[10-11,27-39]. Most transcriptomic classifiers are built around a single clinical task. A survival model estimates future outcome among patients with cancer, whereas a tumour-normal classifier distinguishes malignant from non-malignant tissue. Analysing both from a common candidate space makes it possible to compare the signals each task captures, but their performance and biological meaning must be judged against their separate endpoints.

We applied the 1,417-gene matrix-stiffness-associated transcriptional set to ccRCC in two analytical branches. For survival modelling, 117 workflows were compared and an eleven-gene prognostic signature was derived; model attribution, expression and outcome evidence were then used to prioritise COL4A4. In parallel, WGCNA and 103 tumour-normal classification workflows produced a five-gene diagnostic classifier, from which ADA was selected for cell-resolved follow-up. We examined both genes in the same 102,368-cell ccRCC atlas using cell-type localisation, pseudotime and virtual perturbation.

2. Methods

2.1 Data sources and preprocessing

Transcriptomic expression profiles, corresponding clinical information, and somatic mutation data of kidney renal clear cell carcinoma (KIRC) were downloaded from The Cancer Genome Atlas (TCGA) database. Additional public datasets were used for prognostic workflow selection and cross-cohort benchmarking, tumour-normal expression comparison, and diagnostic-classifier benchmarking. These datasets included GSE167573, GSE213324, GSE251905, and GSE289907 from the Gene Expression Omnibus, E-MTAB-1980 (101 samples) from ArrayExpress, and the CPTAC ccRCC cohort (94 samples) from the CPTAC Data Portal. Matrix-stiffness-associated genes were based on a previously reported transcriptional set assembled from MSigDB and literature-derived sources; after gene-symbol harmonisation and duplicate removal, 1,417 genes were retained for the present analyses. Single-cell transcriptomic data for KIRC were obtained from GSE207493.

For bulk transcriptomic data, TCGA-KIRC expression profiles were normalised, and samples with the required clinical follow-up information were retained for downstream survival analysis. For GEO cohorts, probes were annotated and converted to official gene symbols according to the corresponding platform files. When multiple probes mapped to a single gene, the average expression value was used to represent gene expression. Single-cell RNA-seq data were processed using the Seurat workflow, including quality-control filtering, normalisation, batch-effect correction, dimensionality reduction, and clustering.

2.2 Identification of matrix stiffness-associated differentially expressed genes

Differential expression analysis between KIRC tumour and normal tissues was performed using the limma package, with $|\log2FC| > 1.5$ and adjusted $P < 0.05$ as the screening thresholds. The resulting differentially expressed genes were intersected with the 1,417-gene matrix-stiffness-associated transcriptional set to identify matrix-stiffness-associated differentially expressed genes. Volcano plots, heatmaps, and Venn diagrams were generated to visualise the expression distribution and intersection results.

2.3 Functional enrichment and OS-associated candidate screening

Gene Ontology (GO) biological-process and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed on the 203 overlapping genes using the clusterProfiler R package. Enriched terms and pathways were visualised using bar plots. Separately, univariable Cox proportional-hazards regression was applied to these 203 genes to identify genes associated with overall

survival (OS). Genes with nominal P<0.05 (n=96) were retained as OS-associated candidates for subsequent STRING-based protein-protein interaction (PPI) analysis.

2.4 Construction of the ensemble machine-learning prognostic model

A protein-protein interaction (PPI) network was constructed from the 96 OS-associated genes using the STRING database and visualised using Cytoscape[40]. Genes were ranked according to network node degree, and the 30 highest-degree genes were retained as candidate features for prognostic-model learning.

To establish a robust prognostic prediction model for KIRC, an integrated machine-learning framework was implemented using the Mime R package[41]. In total, 117 algorithm combinations were generated by integrating ten commonly used survival algorithms, including Lasso, Ridge, Elastic Net (Enet), CoxBoost, stepwise Cox regression (StepCox), partial least-squares Cox regression (plsRcox), supervised principal component analysis (SuperPC), generalised boosted regression modelling (GBM), random survival forest (RSF), and survival support vector machine (survival-SVM). Lasso and SuperPC were implemented according to their established formulations[42-43].

All model fitting, hyperparameter optimisation and feature selection were conducted in the TCGA-KIRC development cohort under a unified 10-fold cross-validation framework. For regularised Cox models (Lasso, Ridge and Enet), the optimal regularisation parameter lambda was determined by cross-validation, and the Elastic Net mixing parameter alpha was screened across a predefined grid. The C-index was used as the principal evaluation metric for the 117 model combinations. E-MTAB-1980, CPTAC and GSE167573 were used to compare candidate workflows and select the final workflow according to the mean C-index across these three cohorts. Because these cohorts contributed to model selection, their reported C-indices represent cross-cohort selection-set benchmarking rather than independent external validation. Individual risk scores were then calculated using the selected model, and patients were stratified into high- and low-risk groups using the cohort-specific median risk score.

2.5 Evaluation of the prognostic model and clinicopathological associations

In the TCGA development cohort, risk-score distributions, survival status, and model-gene expression patterns were evaluated. Kaplan-Meier survival curves were used to compare OS between the high- and low-risk groups, and time-dependent receiver operating characteristic (ROC) curves were used to evaluate predictive performance at different time points. The same risk-score formula was applied to E-MTAB-1980,

CPTAC, and GSE167573; because expression-value distributions differ across platforms, risk stratification in each cohort used the within-cohort median risk score.

Differences in risk scores were compared across clinicopathological groups, including overall pathologic stage and pathologic T, N, and M categories, using the Wilcoxon rank-sum test or Kruskal-Wallis test as appropriate. Risk score and clinical variables were further incorporated into univariable and multivariable Cox regression analyses to evaluate whether the risk score remained independently associated with OS in patients with KIRC. A nomogram integrating the risk score and clinical variables was constructed using the rms package, and its predictive performance was evaluated using calibration curves and decision-curve analysis.

2.6 SHAP analysis and cross-cohort expression analysis of model genes

To improve model interpretability, SHapley Additive exPlanations (SHAP) analysis was used to quantify the contribution of each model gene to the prediction output. SHAP feature-importance, beeswarm, dependence, waterfall, and force plots were generated. SHAP values were used for model interpretation and, together with gene-risk correlations, tumour-normal expression, and gene-level OS analyses, for prioritisation of COL4A4.

Public pan-cancer expression data were used to evaluate expression patterns of model genes across tumour types, with particular emphasis on KIRC. Tumour-normal expression of the eleven prognostic-model genes was assessed in TCGA-KIRC and additionally evaluated in processed series-matrix data from GSE289907, GSE213324, and GSE251905. Expression values were retained on each dataset's processed scale. Where probe annotation was required, probes were mapped to gene symbols and the probe with the highest mean expression was retained when multiple probes mapped to the same gene. Normal and tumour groups were compared separately within each dataset using two-sided Wilcoxon rank-sum tests; cross-platform expression values were not pooled.

2.7 Somatic mutation, immune, and drug-sensitivity analyses

Somatic mutation data from TCGA-KIRC were analysed using the maftools package to generate mutation landscapes for the high- and low-risk groups. ESTIMATE was used to calculate ImmuneScore, StromalScore, ESTIMATEScore, and inferred tumour purity[44]. The TIP framework was used to evaluate activity across the cancer-immunity cycle[45], and TIDE was used to estimate immune-dysfunction and exclusion features[46]. Drug-sensitivity prediction used transcriptome-based pharmacogenomic modelling with GDSC-derived

response models[47], and DrugReflector analysis was used to compare transcriptional connectivity patterns between risk groups[48].

2.8 WGCNA and tumour-normal diagnostic classification

Weighted gene co-expression network analysis (WGCNA) was performed using the WGCNA R package (version 1.73)[49] to identify gene co-expression modules associated with tumour status. Soft-thresholding powers (β) were evaluated to approximate a scale-free network topology. A weighted adjacency matrix was then constructed and transformed into a topological overlap matrix (TOM), and TOM-based dissimilarity was used for hierarchical clustering. Gene modules were identified using the dynamic tree-cutting algorithm, and module eigengenes were correlated with tumour-normal status. Modules showing the strongest negative and positive correlations with tumour status were selected for downstream analysis. Genes from the selected modules were intersected with the 96 OS-associated matrix-stiffness genes to define candidates for subsequent diagnostic-classification modelling.

A comprehensive diagnostic machine-learning panel was then applied to the 15 candidates. The framework included regularisation methods (Lasso, Ridge, and ElasticNet), generalised linear-model approaches (Stepglm, glmBoost, and plsRglm), ensemble-learning methods (Random Forest and XGBoost), pattern-recognition models (SVM, LDA, and Naive Bayes), and gradient-boosting approaches. The workflow incorporated feature selection and predictive model fitting under stratified 10-fold cross-validation and generated 103 diagnostic-classification workflows. Model performance was evaluated by AUC across TCGA-KIRC, GSE289907, GSE213324, and GSE251905. The retained glmBoost plus random-forest classifier contained ADA, CCL5, HAVCR2, LGALS1, and LGALS9. SHAP was used to interpret feature contributions, and each classifier gene was further evaluated using tumour-normal expression, single-gene ROC analysis, and median-dichotomised OS analysis. Because these cohorts contributed to classifier comparison and selection, they were treated as comparative benchmarking cohorts rather than untouched independent clinical validation cohorts.

2.9 Single-cell transcriptomic analysis

Single-cell sequencing data were processed using the standard Seurat workflow[50], including quality-control filtering, normalisation, highly variable gene selection, principal component analysis, UMAP dimensionality reduction, batch-effect correction, and clustering, with integration choices considered in relation to atlas-level integration benchmarking[51]. Cell types were annotated using scMayoMap, and the expression distributions of focal genes were assessed across cell populations. After quality control and doublet removal, 102,368 cells were retained for downstream analysis.

2.10 Single-cell pseudotime analysis

For cell populations enriched for focal-gene expression, Monocle was applied to order cells along inferred pseudotime trajectories and to describe gene-expression changes across transcriptional states.

2.11 In silico gene knockout analysis using scTenifoldKnk

In silico gene knockout analysis was performed using the scTenifoldKnk R package[52] to evaluate the potential regulatory impact of focal genes in KIRC-related cell populations. The raw count matrix of the selected cell population was extracted from the Seurat object and used as input. For each target gene, scTenifoldKnk was applied to simulate gene knockout at the single-cell transcriptomic level. Briefly, the algorithm constructed gene regulatory networks before and after virtual knockout, followed by manifold alignment to estimate transcriptomic perturbation induced by target-gene deletion. Genes showing significant regulatory shifts after knockout were identified according to the perturbation score and adjusted statistical significance. The top dysregulated genes were retained for downstream functional interpretation, thereby helping to infer the potential regulatory roles of key matrix-stiffness-associated genes in KIRC.

2.12 Virtual overexpression analysis

To evaluate the potential regulatory effects of the focal genes, virtual overexpression analysis was performed using a scTenifoldOE workflow based on the scTenifoldNet network-perturbation framework[53]. The raw counts matrix from single-cell RNA-seq data was used as input, and major cell types were randomly downsampled before analysis. Highly variable genes were selected and combined with each target gene to generate the input gene set. Virtual overexpression was performed separately for COL4A4 and ADA, with the

overexpression factor set to 10. Gene regulatory networks were constructed under wild-type and virtual-overexpression conditions, followed by manifold alignment to estimate transcriptional perturbations induced by target-gene overexpression. Genes with adjusted $P<0.05$ were considered significantly affected, and the top affected genes for each target were used for GO enrichment analysis.

2.13 Statistical analysis

All tests were two-sided unless defined otherwise by the software. Continuous associations were assessed using Spearman correlation. Kaplan-Meier groups were compared using log-rank tests, and hazard ratios with 95% confidence intervals were estimated using Cox models. Multiple-testing correction used the Benjamini-Hochberg method where recorded. Nominal $P<0.05$ defined the univariable Cox screen, whereas $FDR<0.05$ defined adjusted enrichment significance. Patients or cohorts, rather than individual cells, were treated as the units of independent clinical replication.

2.14 Ethics and data use

This study reused de-identified data from public repositories and did not recruit participants or collect new specimens. No additional institutional ethics approval or individual informed consent was required for this secondary analysis of publicly available de-identified datasets.

3: Results

3.1 Matrix-stiffness-associated transcriptomic screening defines the shared discovery space in ccRCC

Tumour-normal differential-expression analysis in TCGA-KIRC identified 1,353 genes at $|\log2 \text{ fold change}| > 1.5$ and adjusted $P<0.05$, including 663 upregulated and 690 downregulated genes. Intersection with a previously reported 1,417-gene matrix-stiffness-associated transcriptional set yielded 203 candidates (Figure 1A-C). Univariable Cox regression further identified 96 genes associated with overall survival at the prespecified nominal $P<0.05$ threshold (Figure 1D). The 203-gene overlap was used for functional enrichment, whereas the 96-gene OS-associated subset was carried forward into PPI-based prognostic feature selection and the WGCNA-derived diagnostic branch.

GO biological-process enrichment analysis of the 203 overlapping genes highlighted extracellular matrix organisation, extracellular structure organisation, external encapsulating structure organisation, cell-substrate

and cell-matrix adhesion, ERK1/ERK2 cascade-related processes, and chemotaxis and taxis processes (Figure 1E). KEGG pathway enrichment of the same 203 genes highlighted integrin signalling, ECM-receptor interaction, PI3K-Akt signalling, focal adhesion, cytoskeleton in muscle cells, and regulation of the actin cytoskeleton. Additional enriched pathways included protein digestion and absorption, AGE-RAGE signalling in diabetic complications, complement and coagulation cascades, and chemokine signalling (Figure 1F). A high-confidence STRING-derived protein-protein interaction network was constructed using the 96 OS-associated genes as input (Figure 1G), and the 30 highest-degree genes were retained for subsequent survival modelling (Figure 1H). Consensus clustering of the same 96-gene expression space separated 514 TCGA-KIRC cases into groups of 284 and 230 patients with distinct expression patterns and overall survival ($P<0.001$; Supplementary Figure S1). These analyses distinguish the broader 203-gene set used for functional characterisation from the 96-gene OS-associated subset used for downstream prognostic and diagnostic analyses.

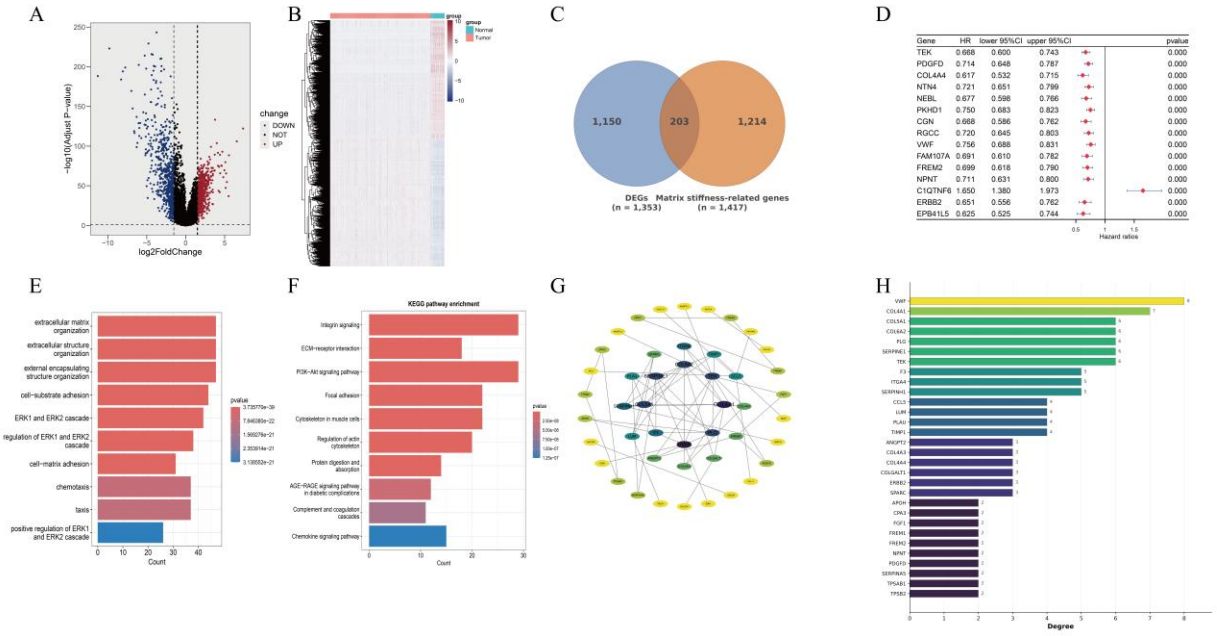


Figure 1. Matrix-stiffness-associated transcriptomic screening and functional enrichment analysis in ccRCC. A, Volcano plot showing differentially expressed genes between TCGA-KIRC tumour and normal tissues. B, Heatmap showing the expression patterns of differentially expressed genes. C, Venn diagram showing the overlap between 1,353 differentially expressed genes and 1,417 matrix-stiffness-associated genes, yielding 203

overlapping candidates. D, Forest plot of selected genes from the univariable Cox analysis; 96 genes in total were associated with OS at nominal P<0.05. E, Gene Ontology biological-process enrichment analysis of the 203 overlapping genes. F, KEGG pathway enrichment analysis of the same 203 genes. G, STRING-derived protein-protein interaction network constructed using the 96 OS-associated genes as input. H, Node-degree ranking of the 30 highest-degree genes retained for subsequent prognostic-model construction.

3.2 Multi-algorithm survival modelling identifies an eleven-gene matrix-stiffness-associated prognostic signature

We compared 117 survival-learning workflows, using TCGA-KIRC as the development cohort and E-MTAB-1980, CPTAC and GSE167573 as cross-cohort workflow-selection cohorts. Five workflows shared the highest mean C-index (0.747) across these three selection cohorts (Figure 2A). Lasso plus SuperPC was retained because it combined co-leading performance with a compact eleven-gene representation.

The model comprised COL4A4, TIMP1, FGF1, F3, PLAU, COLGALT1, COL5A1, FREM2, CPA3, COL4A3 and PDGFD and was defined as: Risk score = -2.056304 x COL4A4 + 2.054580 x TIMP1 - 1.394876 x FGF1 + 1.603805 x F3 + 1.293289 x PLAU + 1.254472 x COLGALT1 + 1.772723 x COL5A1 - 2.333829 x FREM2 - 1.741431 x CPA3 - 1.185251 x COL4A3 - 2.726067 x PDGFD.

The C-indices were 0.627 in TCGA-KIRC, 0.690 in E-MTAB-1980, 0.739 in CPTAC and 0.811 in GSE167573 (Figure 2B). The 1-, 3- and 5-year AUCs were 0.607, 0.562 and 0.615 in TCGA-KIRC; 0.649, 0.668 and 0.612 in E-MTAB-1980; 0.778, 0.611 and 0.651 in CPTAC; and 0.792, 0.911 and 0.707 in GSE167573, respectively (Figure 2C,E).

Using the median score within each cohort, high-risk patients had shorter overall survival in TCGA-KIRC (HR=2.15, 95% CI 1.60-2.90, P<0.001), E-MTAB-1980 (HR=2.51, 95% CI 1.11-5.69, P=0.035), CPTAC (HR=3.91, 95% CI 1.62-9.48, P=0.004) and GSE167573 (HR=5.47, 95% CI 1.56-19.18, P=0.015; Figure 2D). Benchmarking against 25 published signatures showed favourable relative performance across multiple cohort-time combinations (Supplementary Figure S2).

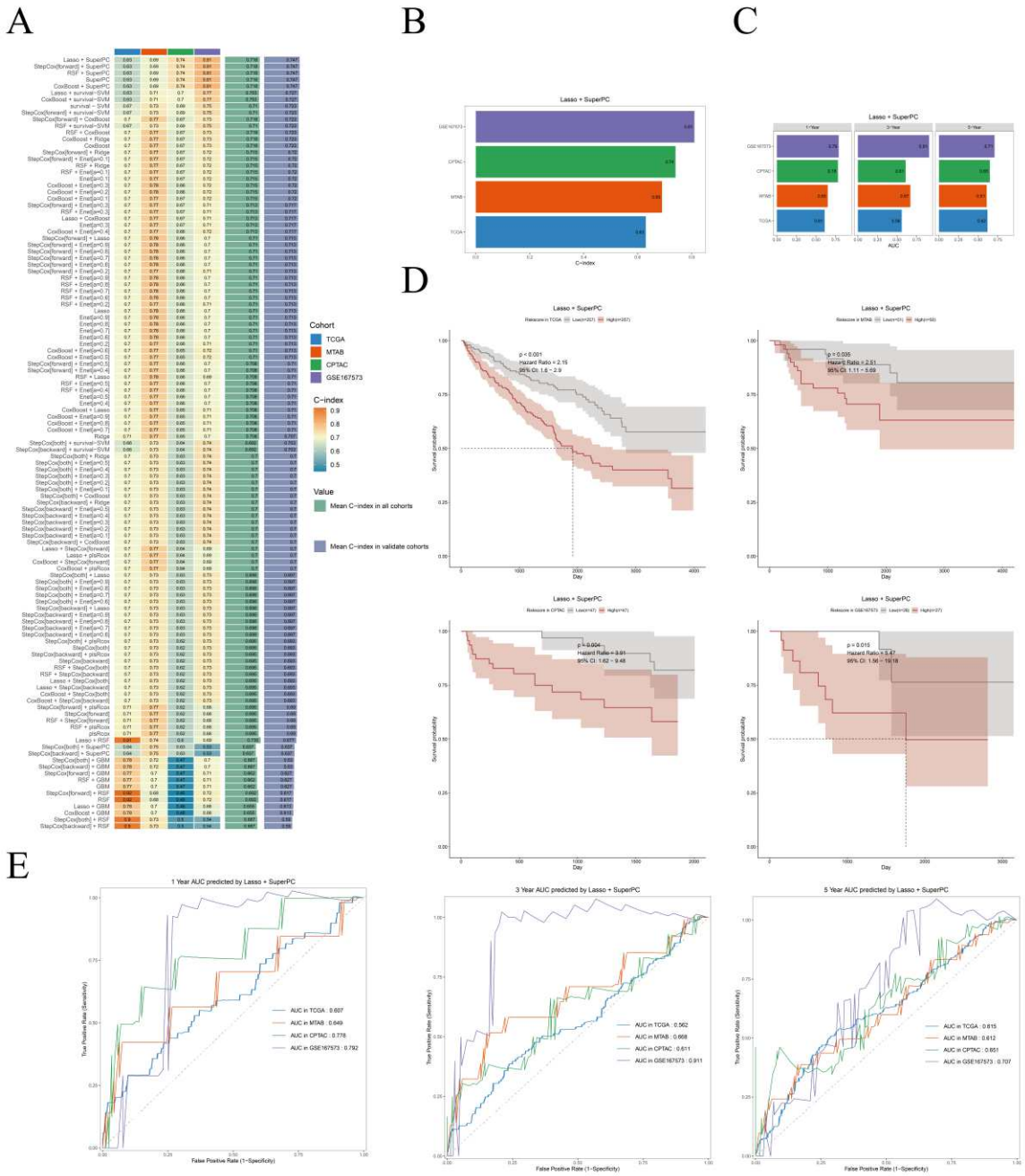


Figure 2. Construction and cross-cohort benchmarking of the eleven-gene prognostic model. A, Comparison of C-indices across 117 survival-learning workflows using the three cross-cohort selection datasets. B, Cohort-specific C-indices of the retained Lasso plus SuperPC model. C, Summary of 1-, 3- and 5-year time-dependent AUCs. D, Kaplan-Meier overall-survival curves for high- and low-risk groups defined by cohort-specific median scores. E, Time-dependent ROC curves.

3.3 Integrated model and outcome evidence prioritises COL4A4 within the prognostic signature

SHAP analysis quantified the contribution of individual genes within the fitted model. COL4A3 had the highest mean absolute SHAP value (0.369), followed by COL4A4 (0.310) and PDGFD (0.280), identifying these genes as the leading contributors to model predictions (Figure 3A-E).

Spearman analysis in 514 TCGA-KIRC samples provided an additional view of gene-risk relationships. FREM2 showed the strongest inverse correlation with the composite score (rho=-0.88), closely followed by COL4A4 (rho=-0.87); COL4A3 also showed a strong inverse association (rho=-0.73). TIMP1 showed the strongest positive correlation (rho=0.76; Figure 3F).

Gene-level survival and tumour-normal expression further distinguished COL4A4. Higher COL4A4 expression was associated with longer survival ($P<0.001$), whereas median-split survival differences were not significant for COL4A3 ($P=0.114$) or F3 ($P=0.0913$). (Supplementary Figure S3) Pan-cancer analysis demonstrated marked tumour-type dependence (Supplementary Figure S4). Thus, COL4A4 was selected through convergence of high model attribution, one of the strongest inverse risk associations, tumour downregulation, favourable survival and subsequent cell-resolved evidence, rather than by a single rank. The model coefficients and interpretive metrics are summarised in Table 2.

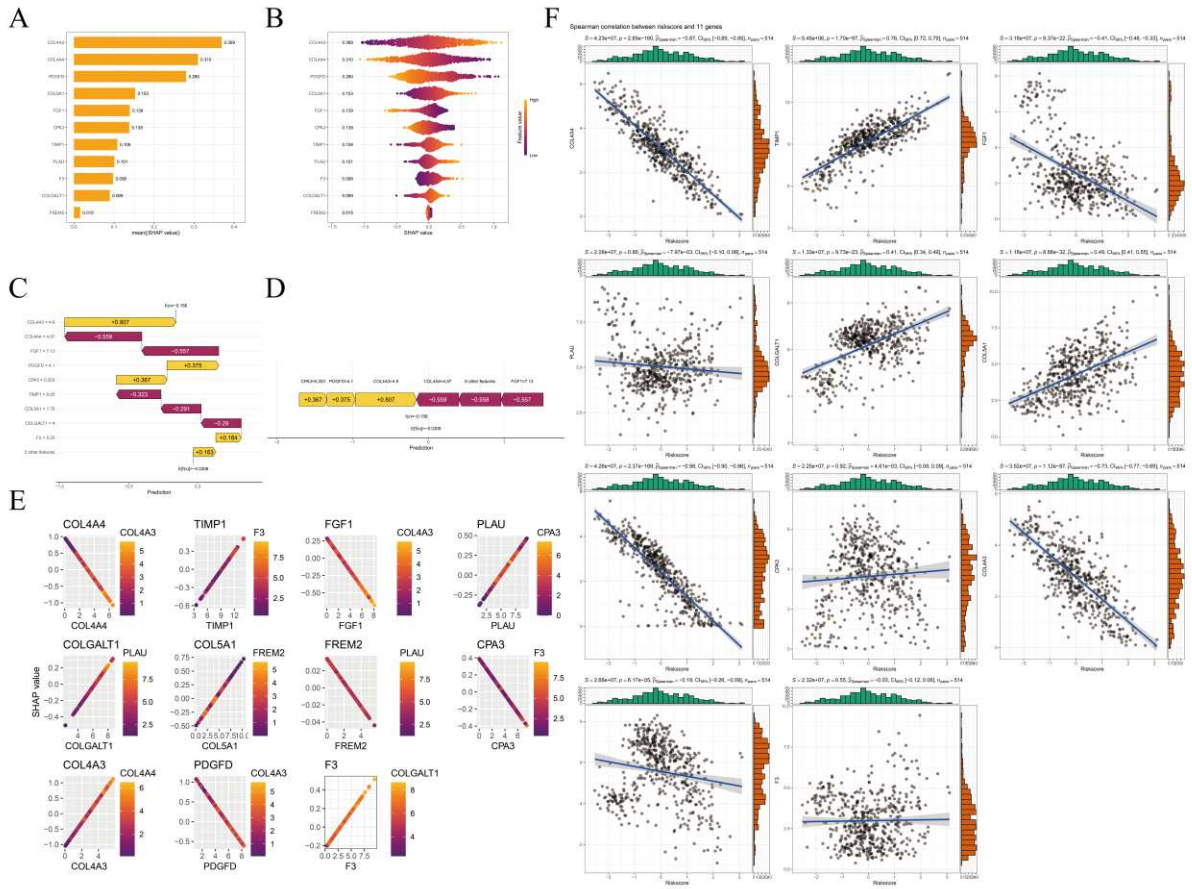


Figure 3. Model interpretation and prioritisation of COL4A4. A, Global feature importance ranked by mean absolute SHAP value. B, SHAP summary plot. C, D, Representative patient-level SHAP explanations. E, SHAP dependence plots. F, Spearman correlations between the eleven model genes and the composite risk score in 514 TCGA-KIRC cases.

3.4 The matrix-stiffness-associated prognostic signature retains independent clinical information

The risk score was significantly associated with overall survival in univariable analysis and remained associated after adjustment for age and pathologic N and M categories (HR=1.523, 95% CI 1.243-1.867, P<0.001; Figure 4A,B).

A nomogram integrating age, pathologic N, pathologic M and the molecular score showed agreement between predicted and observed survival probabilities, while decision-curve analysis indicated potential net benefit for the nomogram and for models including the molecular score (Figure 4C-E).

Risk scores increased in advanced T categories, N1 disease, M1 disease and stage III/IV tumours, linking the molecular signature to clinicopathological progression (Supplementary Figure S5).

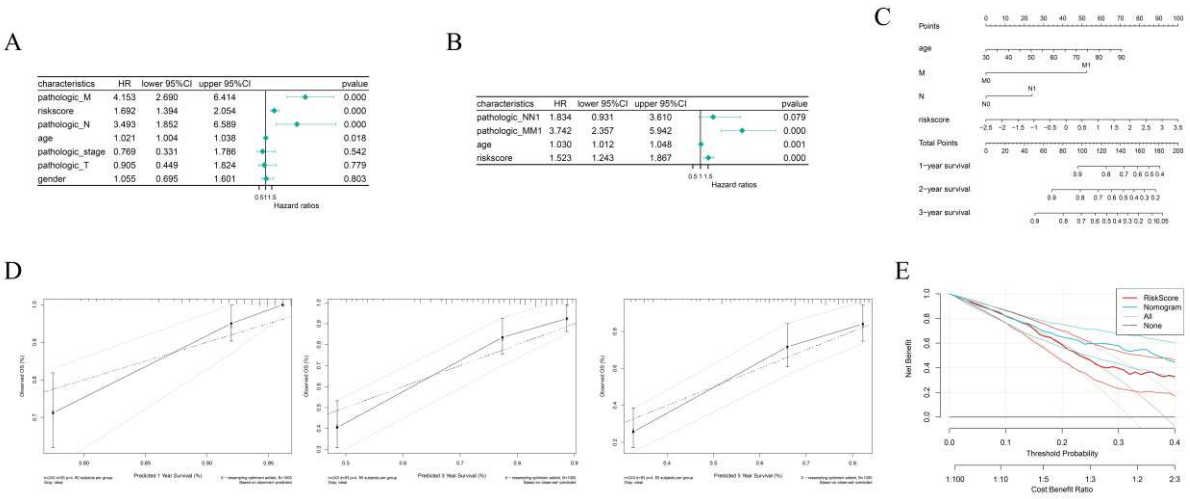


Figure 4. Clinical evaluation of the matrix-stiffness-associated prognostic signature. A, Univariable Cox analysis. B, Multivariable Cox model including age, pathologic N, pathologic M and risk score. C, Nomogram integrating the retained clinical variables and molecular score. D, Calibration plots for 1-, 3- and 5-year overall survival. E, Decision-curve analysis.

3.5 The prognostic signature is associated with immune-stromal remodelling

High-risk tumours had higher ImmuneScore, StromalScore and ESTIMATEScore and lower inferred tumour purity (all P<0.001; Figure 5A-D). Immune-cell deconvolution showed coordinated differences across multiple populations, including increases in activated CD4 memory T cells, follicular helper T cells, regulatory T cells and M0 macrophages and reductions in monocytes, M1 macrophages, resting dendritic cells and resting mast cells in the evaluated analyses (Figure 5E-G).

Extended immune analyses showed differences in multiple steps of the cancer-immunity cycle. High-risk tumours also had higher overall TIDE, dysfunction and exclusion scores, accompanied by lower MSI and CD274 scores and altered CD8-related features (Supplementary Figure S6). VHL, PBRM1, TTN, SETD2 and BAP1 were among the most frequently altered genes across risk groups. Exploratory pharmacogenomic

analyses identified 94 compounds with higher predicted IC50 and 40 with lower predicted IC50 in high-risk tumours (Supplementary Figure S7).

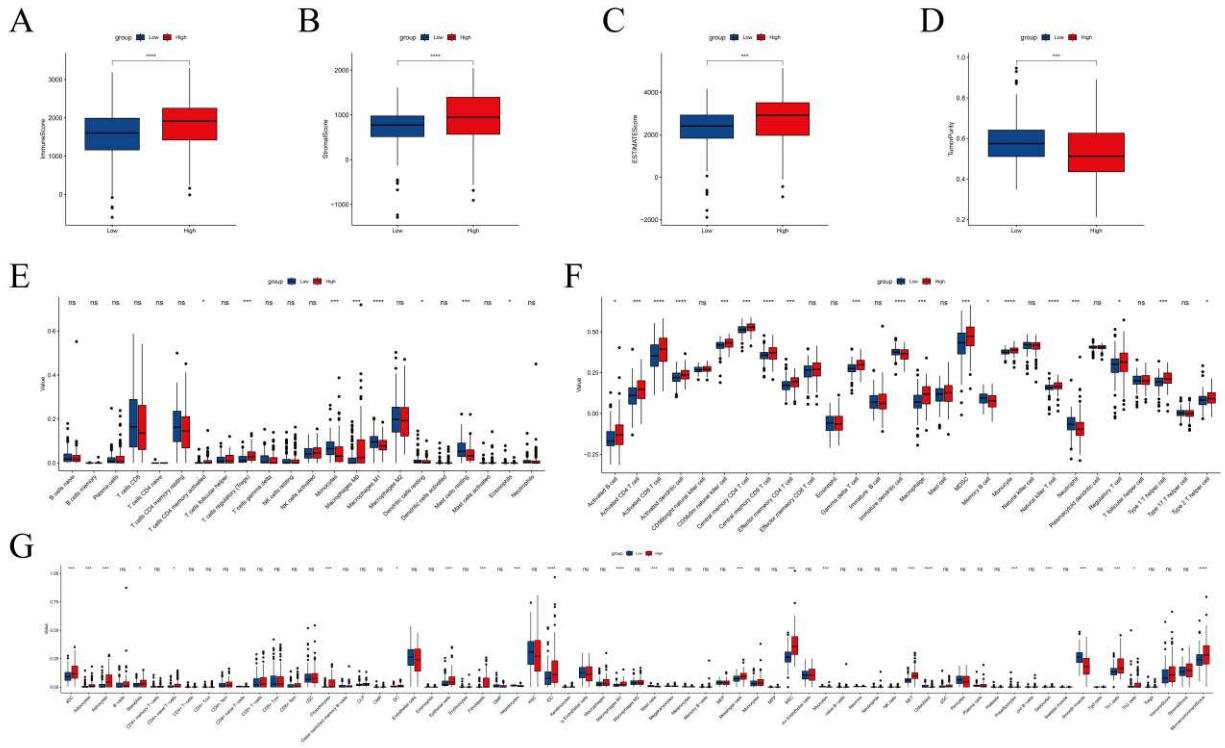


Figure 5. Immune and stromal associations of the matrix-stiffness-associated prognostic signature. A-D, ImmuneScore, StromalScore, ESTIMATEScore and inferred tumour purity in low- and high-risk groups. E-G, Immune-cell composition inferred by three complementary deconvolution analyses.

3.6 WGCNA and machine learning define a five-gene tumour-normal diagnostic classifier

WGCNA was applied to tumour and normal kidney transcriptomes to derive a diagnostic branch from the shared matrix-stiffness-associated discovery space. The turquoise and red modules represented the two extreme tumour-associated modules, corresponding to the strongest negative and strongest positive correlations with tumour status. Their 722-gene union intersected with the 96 prognosis-associated matrix-stiffness genes to yield 15 candidates (Supplementary Figure S8).

All 15 candidates entered 103 tumour-normal diagnostic-classification workflows. The retained glmBoost plus random-forest classifier comprised ADA, CCL5, HAVCR2, LGALS1 and LGALS9 and achieved a mean AUC of 0.895 across TCGA-KIRC, GSE289907, GSE213324 and GSE251905 (Figure 6A). All five genes

showed reproducible tumour-normal expression differences across the four evaluated datasets (Figure 6B-E), and cohort-specific single-gene ROC analyses supported their discriminatory information (Supplementary Figure S9).

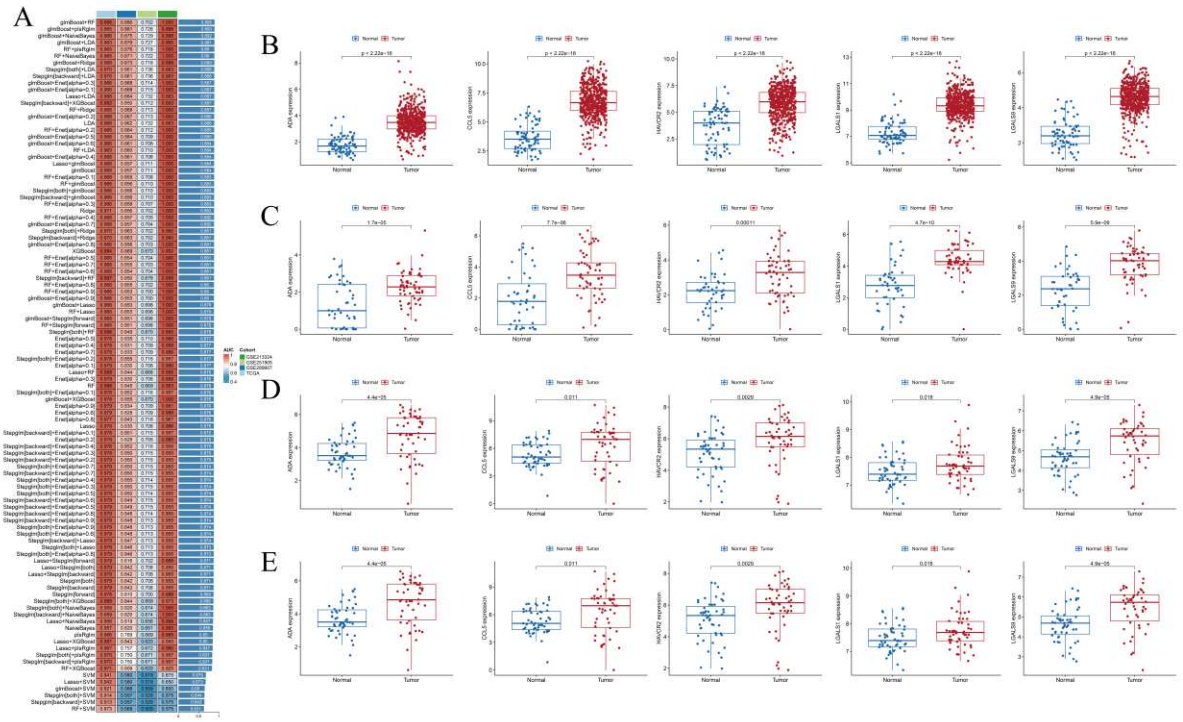


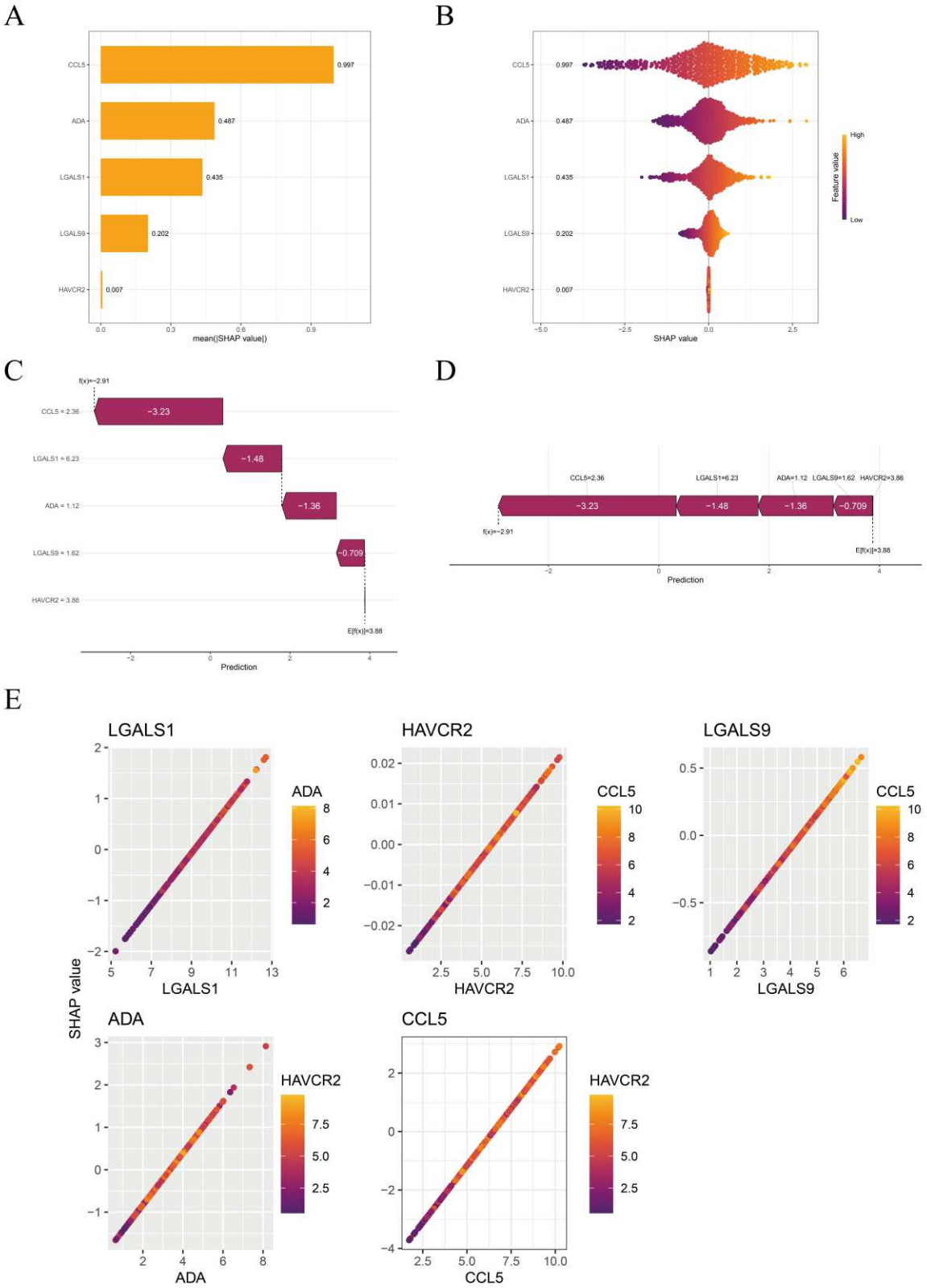
Figure 6. Construction and expression-level evaluation of the five-gene tumour-normal diagnostic classifier. A, AUC comparison across 103 diagnostic-classification workflows and four tumour-normal cohorts. B-E, Tumour-normal expression of ADA, CCL5, HAVCR2, LGALS1 and LGALS9 in TCGA-KIRC and the three external kidney cohorts.

3.7 Model interpretation and outcome evidence prioritise ADA for cell-resolved investigation

SHAP interpretation ranked CCL5 as the leading contributor to diagnostic-classifier prediction, followed by ADA and LGALS1 (Figure 7A,B). Representative patient-level explanations and dependence plots illustrated how the five features contributed to individual predictions (Figure 7C-E). CCL5 was therefore retained as the dominant classifier feature rather than being replaced by the gene with the strongest survival association.

Gene-level overall-survival analysis separated diagnostic contribution from outcome association. Median-split survival P values were 0.007 for ADA, 0.003 for HAVCR2, 0.051 for LGALS1, 0.189 for CCL5 and 0.281 for LGALS9 (Supplementary Figure S10). Higher ADA expression was associated with poorer survival, whereas the HAVCR2 association was in the favourable direction.

ADA was selected for deeper cell-resolved analysis because it combined diagnostic-classifier membership, reproducible tumour-normal discrimination, a significant adverse survival association and an interpretable immune-cell distribution. This selection criterion is distinct from classifier feature importance and does not alter the leading SHAP position of CCL5. The evidence supporting selection of ADA for deeper cell-resolved analysis is summarised in Table 3.


A
CCL5
ADA
LGALS1
LGALS9
HAVCR2
0.997
0.487
0.435
0.202
0.007
mean(|SHAP value|)
B
SHAP value
Feature value
High
Low
C
f(x)=-2.91
CCL5 = 2.36
LGALS1 = 6.23
ADA = 1.12
LGALS9 = 1.62
HAVCR2 = 3.86
-3.23
-1.48
-1.36
-0.709
E[f(x)]=3.88
Prediction
D
CCL5=2.36
LGALS1=6.23
ADA=1.12
LGALS9=1.62
HAVCR2=3.86
Prediction
E
LGALS1
HAVCR2
LGALS9
ADA
CCL5
SHAP value

370

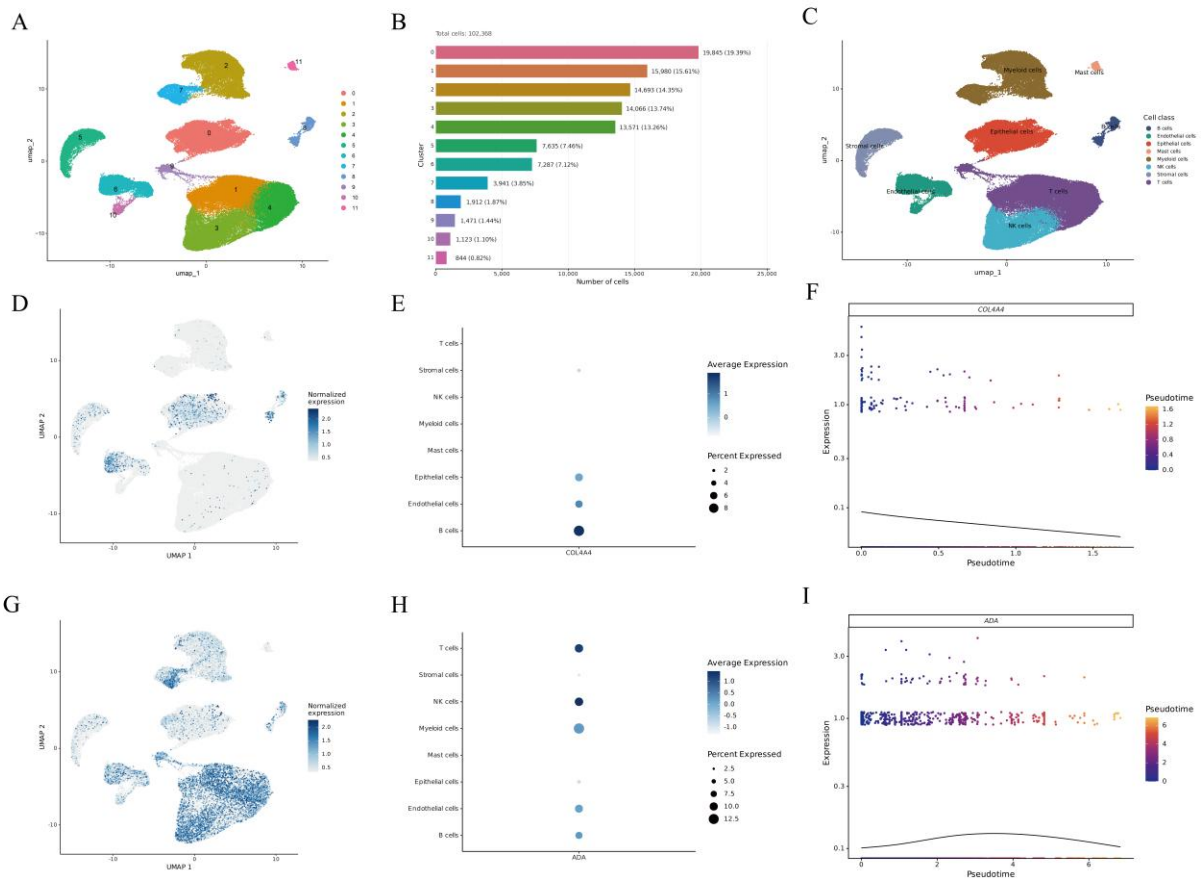
Figure 7. SHAP interpretation of the five-gene tumour-normal diagnostic classifier. A, Global feature importance ranked by mean absolute SHAP value. B, SHAP summary plot. C, D, Representative patient-level SHAP explanations. E, SHAP dependence plots for selected feature pairs.

3.8 COL4A4 and ADA occupy distinct cellular states in the ccRCC ecosystem

We next compared the focal genes from the prognostic signature and diagnostic classifier in GSE207493. After quality control and doublet removal, the atlas comprised 102,368 cells. Unsupervised clustering identified 12 clusters that were consolidated into B, endothelial, epithelial, mast, myeloid, NK, stromal and T-cell classes (Figure 8A-C).

COL4A4 showed its strongest class-level transcript signal and largest expressing-cell fraction in B cells, followed by endothelial and epithelial cells (Figure 8D,E). Its fitted expression trend declined progressively along the inferred pseudotime trajectory (Figure 8F), indicating attenuation of the COL4A4-associated transcriptional state across the modelled transition.

ADA displayed a broader immune-centred distribution, with prominent signal in T and NK cells and detectable expression across additional immune compartments (Figure 8G,H). Unlike the monotonic COL4A4 pattern, ADA followed a non-linear trajectory: fitted expression increased during early-to-intermediate pseudotime and subsequently declined (Figure 8I).



3.9 Virtual perturbation predicts divergent COL4A4- and ADA-associated network responses

Virtual COL4A4 knockout identified B cells as the dominant responding population. IGF1 and CD24 showed the strongest positive responses, whereas LRIG2, GDAP2, SHISA8, EYA3, CYB5D2, SLC25A33, AIM2 and VPREB3 were among the leading negative responses (Figure 9A,B). Virtual overexpression again produced a B-cell-dominant response, with BCL2A1, COTL1, CD3G, CCR7, GZMK, MS4A1, SNN, ID2, LTB and KIF2A among the leading positive genes (Figure 9D,E). The COL4A4 GO rankings did not reach FDR<0.05 and were interpreted descriptively (Figure 9C,F).

Both directions of virtual ADA perturbation identified NK cells as the dominant responding population (Figure 9G,J). Virtual knockout emphasised heat-response and protein-refolding genes and yielded significant ranked terms related to protein folding, heat response and chromosome segregation (Figure 9H,I). Virtual

overexpression highlighted immune and metabolic response genes (Figure 9K); its GO rankings did not reach FDR<0.05 and were therefore interpreted descriptively (Figure 9L).

4. Discussion

The two modelling branches yielded related but non-identical biological signals. The survival signature was weighted toward matrix and basement-membrane-associated genes, and COL4A4 was selected for follow-up because model attribution, risk correlation, tumour expression and gene-level survival pointed in a consistent direction. In the diagnostic classifier, CCL5 contributed most strongly to prediction; ADA was taken forward because its tumour-normal signal coincided with an adverse overall-survival association and a distinct immune-cell distribution. The cell-resolved analyses therefore examine two candidates selected through different evidence paths, not two components of a single established mechanism.

The 203-gene overlap was enriched for extracellular-matrix organisation, cell-substrate adhesion and ERK-related processes, while KEGG analysis highlighted integrin signalling, ECM-receptor interaction, focal adhesion and PI3K-Akt signalling. This profile fits the broader view that the tumour matrisome regulates cell behaviour and immune access[12-21]. There is also ccRCC-specific support for this connection: spatial coupling between integrin and collagen-IV signals has been reported in immunoresistant regions[22], and PIEZO1-dependent stiffness signalling has been linked to renal tumour progression[23]. An important qualification is the origin of the input set: the 1,417-gene collection was assembled in an HCC-related context and transferred to ccRCC[62]. The enrichment results therefore identify matrix-associated transcriptional programmes shared with that source set; they do not show that stiffness was directly measured in the ccRCC samples or that mechanical forces caused the observed clinical associations.

COL4A4 was not the highest-ranked feature on every metric. COL4A3 had a larger mean absolute SHAP value (0.369 versus 0.310), while FREM2 showed a slightly stronger inverse correlation with the composite risk score (rho=-0.88 versus -0.87). COL4A4 was prioritised because several independent readouts converged: it retained a high model attribution, was strongly inversely associated with risk, was downregulated in tumour tissue and showed a favourable gene-level survival association. This pattern is compatible with earlier ccRCC

studies in which basement-membrane-related signatures carried prognostic information[24-25]. It also provides a clinical-transcriptomic context for experimental work linking COL4A4 loss and promoter hypermethylation to JAK2/STAT3 activation and microenvironmental reprogramming[26]. The current analysis does not reproduce that mechanism; it adds evidence from model interpretation and single-cell localisation, while protein deposition, methylation status and pathway activity remain untested. Other members of the prognostic signature, including COLGALT1, TIMP1 and CPA3, also have reported links to ccRCC biology[54-56].

ADA requires a similarly specific interpretation. CCL5 was the leading SHAP feature in the five-gene classifier, so ADA should not be described as the classifier's dominant diagnostic marker. Its value for follow-up came from the combination of classifier membership, a significant adverse overall-survival association and expression across immune compartments. ADA catalyses the conversion of adenosine to inosine and therefore sits within purine metabolism, which can influence immune-cell activity. A previous pan-cancer analysis reported ADA-related expression and immune associations in KIRC[60]. Our data are consistent with an immune-metabolic role, but transcript abundance alone cannot establish enzyme activity, local adenosine concentrations or the functional effect of ADA-expressing cells on NK- and T-cell responses.

The high-risk group also showed greater inferred immune and stromal content, together with changes in the composition of several immune populations. Higher TIDE, dysfunction and exclusion scores occurred alongside evidence of immune-cycle activity. This combination is compatible with immune recruitment that does not necessarily translate into effective tumour control. Single-cell and spatial studies in ccRCC have described coordinated interactions among tumour, stromal, myeloid, endothelial and lymphoid compartments[10,17,22,39,61]; the present risk-associated pattern may reflect a similar multicellular context. It should not, however, be read as evidence that the eleven-gene score directly measures immune function.

The single-cell results help localise the two candidates, although they do not settle their cellular origin or function. COL4A4 had its strongest class-level transcript signal in B cells, and virtual COL4A4 perturbation also produced B-cell-dominant network responses. These observations do not establish B cells as the source of

deposited type IV collagen. The signal could arise from a restricted B-cell subset, and RNA abundance does not demonstrate translation, assembly or secretion of collagen-IV protein. Subcluster-level analysis and spatial RNA/protein colocalisation are needed before assigning a matrix-producing role to these cells. ADA was prominent in T and NK cells; both virtual knockout and overexpression produced NK-cell-dominant responses. This convergence identifies NK-cell networks as responsive in the model, but it does not mean that loss and gain of ADA have equivalent biological effects. scTenifold outputs depend on the sampled network and modelling assumptions. Graded ADA manipulation in purified or co-cultured NK/T cells, accompanied by ADA activity, adenosine/inosine measurements and functional immune assays, would help determine whether the predicted response is reproducible experimentally. COL4A4 likewise requires spatial and protein-level validation.

Several limitations affect interpretation. The analysis was retrospective and combined cohorts generated on different platforms. E-MTAB-1980, CPTAC and GSE167573 contributed to selection of the 117 prognostic workflows; their C-indices are therefore selection-set benchmarks, not estimates from untouched independent validation cohorts. The tumour-normal datasets also contributed to diagnostic-workflow comparison, so the five-gene classifier has not yet been evaluated as a locked model in an independent clinical cohort. A fixed model, preprocessing pipeline and decision threshold should be tested in datasets not used during model development or selection. In addition, cohort-specific median cut-offs do not provide a universal clinical threshold. The transferred gene set may not fully represent kidney-specific matrix biology, and deconvolution, drug-sensitivity estimation, pseudotime and virtual perturbation remain computational analyses. Finally, single-cell transcript measurements cannot substitute for protein localisation or enzyme assays. These issues define the next validation steps: independent cohort testing, direct assessment of tumour matrix properties, spatial COL4A4 analysis and functional ADA studies in relevant immune-cell systems.

Several limitations should be considered. First, this was a retrospective analysis using cohorts generated on different platforms. E-MTAB-1980, CPTAC and GSE167573 contributed to selection of the 117 prognostic workflows; therefore, their reported C-indices are selection-set benchmarks, not independent external-validation estimates, and may be optimistic. The tumour-normal cohorts likewise supported diagnostic-

workflow comparison rather than evaluation of a locked classifier in an untouched clinical cohort. Independent validation will require a prespecified model, fixed preprocessing and cut-offs, and cohorts not used for model development or selection. Second, cohort-specific median risk thresholds do not define a universal clinical cut-off. Third, the 1,417-gene matrix-stiffness-associated set was transferred from an HCC study based on MSigDB and literature-derived pathways; it is neither kidney-specific nor a direct mechanical measurement[62]. Finally, deconvolution, drug-sensitivity estimates, pseudotime and virtual perturbation are computational inferences, while single-cell RNA sequencing measures transcripts rather than collagen-IV deposition or ADA enzymatic activity. Prospective studies should combine spatial RNA/protein localisation, direct measurement of renal-tumour stiffness, matrix-controlled COL4A4 perturbation, assessment of JAK2/STAT3 and PIEZO1/YAP signalling, and graded ADA manipulation in purified or co-cultured NK/T-cell systems.

The study connects a matrix-associated transcriptional programme with two complementary prediction tasks and generates distinct cell-level hypotheses for COL4A4 and ADA. The immediate priority is to evaluate the models in untouched cohorts and test each candidate in the cellular context indicated by the single-cell analyses.

Data availability statement

The transcriptome datasets analysed during the current study are derived from multiple publicly accessible repositories: The Cancer Genome Atlas (TCGA), the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO), ArrayExpress and CPTAC Data Portal. The specific datasets and their corresponding direct access links are as follows: TCGA-KIRC: Kidney Renal Clear Cell Carcinoma project of TCGA, available at https://portal.gdc.cancer.gov/projects/TCGA-KIRC. GSE167573: available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE167573. GSE213324: available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE213324. GSE251905: available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE251905. GSE289907: available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE289907. E-MTAB-1980: available at https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1980. CPTAC ccRCC cohort: available at

https://cptac-data-portal.georgetown.edu/. GSE207493 (single-cell transcriptomic data): available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207493.

Ethics, Consent to Participate, and Consent to Publish declarations

not applicable.

Authors' contributions

Yongkang Xiao and Xiuqi Luo contributed equally to this work. Yongkang Xiao was responsible for data management, statistical analysis, and drafting of the initial manuscript. Xiuqi Luo contributed to data curation and figure preparation. Fei Hu, Chen Luo, Feiyun Wu, Hui Chen, Chenghui Li and Renyu Gong participated in data collection and literature sorting. Chengyan Guo provided overall supervision, guidance and manuscript revision. All authors have read and approved the final version of the manuscript.

Funding

No funding was received for this study.

Competing interests

All authors declare that they have no competing interests.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

519

References

[1].CHOUEIRI TK, POWLES T, PELTOLA K, et al. Belzutifan versus everolimus for advanced renal-cell carcinoma [J]. N Engl J Med, 2024, 391: 710-721.

[2].SALIBY RM, LABAKI C, JAMMIHAL TR, et al. Impact of renal cell carcinoma molecular subtypes on immunotherapy and targeted therapy outcomes [J]. Cancer Cell, 2024, 42: 732-735.

[3].KINGET L, NAULAERTS S, GOVAERTS J, et al. A spatial architecture-embedding HLA signature to predict clinical response to immunotherapy in renal cell carcinoma [J]. Nat Med, 2024, 30: 1667-1679.

[4].JAMMIHAL T, SALIBY RM, LABAKI C, et al. Immunogenomic determinants of exceptional response to immune checkpoint inhibition in renal cell carcinoma [J]. Nat Cancer, 2025, 6: 372-384.

[5].RINI BI, PLIMACK ER, STUS V, et al. Pembrolizumab plus axitinib versus sunitinib for advanced clear cell renal cell carcinoma: 5-year survival and biomarker analyses of the phase 3 KEYNOTE-426 trial [J]. Nat Med, 2025, 31: 3475-3484.

[6].VUONG L, CORNISH AE, PFEIL J, et al. Hypoxia shapes both therapeutic response and resistance in metastatic clear cell renal cell carcinoma [J]. Cancer Cell, 2026, 44: 1477-1495.e15.

[7].FERNANDES R, JABBARIZADEH B, RAJEH A, et al. Fecal microbiota transplantation plus immunotherapy in metastatic renal cell carcinoma: the phase 1 PERFORM trial [J]. Nat Med, 2026, 32: 1325-1336.

[8].QU Y, FENG J, WU X, et al. A proteogenomic analysis of clear cell renal cell carcinoma in a Chinese population [J]. Nat Commun, 2022, 13: 2052.

[9].LI Y, LIH TM, DHANASEKARAN SM, et al. Histopathologic and proteogenomic heterogeneity reveals features of clear cell renal cell carcinoma aggressiveness [J]. Cancer Cell, 2023, 41: 139-163.e17.

[10].MEI S, ALCHAHIN AM, TSEA I, et al. Single-cell analysis of immune and stroma cell remodeling in clear cell renal cell carcinoma primary tumors and bone metastatic lesions [J]. Genome Med, 2024, 16: 1.

[11].YANG G, CHENG J, XU J, et al. Metabolic heterogeneity in clear cell renal cell carcinoma revealed by single-cell RNA sequencing and spatial transcriptomics [J]. J Transl Med, 2024, 22: 210.

[12].PALLY D, ENGLUND JI, NABA A. Understanding and targeting the tumour matrisome [J]. Nat Rev Clin Oncol, Published online September 21, 2026. doi:10.1038/s41571-026-01203-5.

[13].SLEEBOOM JJF, VAN TIENDEREN GS, SCHENKE-LAYLAND K, et al. The extracellular matrix as a hallmark of cancer and metastasis: from biomechanics to therapeutic targets [J]. Sci Transl Med, 2024, 16: eadg3840.

[14].PRAKASH J, SHAKED Y. The interplay between extracellular matrix remodeling and cancer therapeutics [J]. Cancer Discov, 2024, 14: 1375-1388.

[15].BORST R, MEYAARD L, PASCOAL RAMOS MI. Understanding the matrix: collagen modifications in tumours and their implications for immunotherapy [J]. J Transl Med, 2024, 22: 382.

[16].CAMBRIA E, COUGHLIN MF, FLORYAN MA, et al. Linking cell mechanical memory and cancer metastasis [J]. Nat Rev Cancer, 2024, 24: 216-228.

[17].ZHANG J, LI J, HOU Y, et al. Osr2 functions as a biomechanical checkpoint to aggravate CD8+ T-cell exhaustion in tumour [J]. Cell, 2024, 187: 3409-3426.e24.

[18].FAN W, ADEBOWALE K, VÁNCZA L, et al. Matrix viscoelasticity promotes liver cancer progression in the pre-cirrhotic liver [J]. Nature, 2024, 626: 635-642.

[19].CHITTY JL, COX TR. The extracellular matrix in cancer: from understanding to targeting [J]. Trends Cancer, 2025, 11: 839-849.

[20].LESAVAGE BL, ZHANG D, HUERTA-LÓPEZ C, et al. Engineered matrices reveal stiffness-mediated chemoresistance in patient-derived pancreatic cancer organoids [J]. Nat Mater, 2024, 23: 1138-1149.

[21].ARPINATI L, CARRADORI G, SCHERZ-SHOUVAL R. CAF-induced physical constraints controlling T-cell state and localization in solid tumours [J]. Nat Rev Cancer, 2024, 24: 676-693.

[22].SOUPIR AC, HAYES MT, PEAK TC, et al. Increased spatial coupling of integrin and collagen IV in the immunoresistant clear-cell renal-cell carcinoma tumor microenvironment [J]. Genome Biol, 2024, 25: 308.

[23].ZHU B, LI F, YU J, et al. PIEZO1 mediates matrix stiffness-induced tumor progression in kidney renal clear cell carcinoma by activating the Ca2+/calpain/YAP pathway [J]. Biochim Biophys Acta Mol Cell Res, 2025, 1872: 119871.

[24].ZHOU T, CHEN W, WU Z, et al. A newly defined basement membrane-related gene signature for the prognosis of clear-cell renal cell carcinoma [J]. Front Genet, 2022, 13: 994208.

[25].XIA B, WANG J, ZHANG D, et al. Integration of basement membrane-related genes in a risk signature for prognosis in clear cell renal cell carcinoma [J]. Sci Rep, 2024, 14: 3893.

[26].ZHOU H, TIAN X, YANG Z, et al. Epigenetic inactivation of COL4A4 orchestrates JAK2/STAT3-driven progression and microenvironment reprogramming in renal cell carcinoma [J]. Int Immunopharmacol, 2026, 188: 117305.

[27].LIU S, LI G, YIN X, et al. Comprehensive investigation of malignant epithelial cell-related genes in clear cell renal cell carcinoma: development of a prognostic signature and exploration of tumor microenvironment interactions [J]. J Transl Med, 2024, 22: 607.

[28].FU T, ZHANG X, LIU Y, et al. Exploring the prognostic value of tumour-associated genes in clear cell renal cell carcinoma through single-cell RNA sequencing insights [J]. J Cell Mol Med, 2024, 28: e70297.

[29].CHEN J, WU X, HAN G, et al. Single-cell transcriptomic insights into ccRCC: a stemness gene signature for prognosis and treatment response prediction [J]. Discov Oncol, 2025, 16: 1862.

[30].DAI H, LU R, ZHANG M, et al. Integrating single-cell multi-omics and machine learning to reveal triaptosis heterogeneity in clear cell renal cell carcinoma [J]. Hum Genomics, 2026, 20: 93.

[31].CHEN JW, LI JY, FENG HQ, et al. FKBP10 promotes M2 polarization of macrophage via MEK/ERK/CXCL8 axis and facilitates tumor progression in clear cell renal cell carcinoma [J]. Int J Biol Sci, 2026, 22: 1807-1833.

[32].LI Y, CHEN Y, WANG L, et al. Integrating bulk and single-cell RNA-seq reveals glycolysis-associated macrophages and related tumour subgroup signatures to predict prognosis and therapy in clear cell renal cell carcinoma [J]. Hum Mutat, 2026, 2026: 3125551.

[33].WANG Y, YUE B, NI H, et al. POSTN-positive cancer-associated fibroblast-CCL3-positive macrophage crosstalk defines the immune-excluded tumour microenvironment in clear cell renal cell carcinoma [J]. Transl Oncol, 2026, 65: 102682.

[34].XU K, LI D, QIAN J, et al. Single-cell disulfidptosis regulator patterns guide intercellular communication of the tumour microenvironment that contributes to kidney renal clear cell carcinoma progression and immunotherapy [J]. Front Immunol, 2024, 15: 1288240.

[35].ZVIRBLYTE J, NAINYS J, JUZENAS S, et al. Single-cell transcriptional profiling of clear cell renal cell carcinoma reveals a tumor-associated endothelial tip cell phenotype [J]. Commun Biol, 2024, 7: 780.

[36].SONG X, ZHU Y, GENG W, et al. Spatial and single-cell transcriptomics reveal cellular heterogeneity and a novel cancer-promoting regulatory T-cell subset in human clear-cell renal cell carcinoma [J]. J Immunother Cancer, 2025, 13: e010183.

[37].KALOGIROU C, KREBS M, KUNZ AS, et al. Spatial transcriptomic profiling of metastatic renal cell carcinoma identifies chemokine-driven macrophage and CD8+ T-cell interactions predictive of immunotherapy response [J]. J Immunother Cancer, 2025, 13: e012991.

[38].MA J, HUANG Y, CHEN J, et al. FAP+ fibroblasts orchestrate tumour-microenvironment remodelling in renal cell carcinoma with tumour thrombus [J]. Nat Commun, 2025, 16: 9387.

[39].YAO W, QI J, ZHENG X, et al. Integrated single-cell and spatial transcriptomics reveals a multicellular program contributing to immunotherapy resistance and adverse clinical outcomes in clear cell renal cell carcinoma [J]. NPJ Precis Oncol, Published online July 21, 2026. doi:10.1038/s41698-026-01615-7.

[40].SHANNON P, MARKIEL A, OZIER O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks [J]. Genome Res, 2003, 13: 2498-2504.

[41].LIU H, ZHANG W, ZHANG Y, et al. Mime: a flexible machine-learning framework to construct and visualize models for clinical characteristics prediction and feature selection [J]. Comput Struct Biotechnol J, 2024, 23: 2798-2810.

[42].TIBSHIRANI R. The lasso method for variable selection in the Cox model [J]. Stat Med, 1997, 16: 385-395.

[43].BAIR E, TIBSHIRANI R. Semi-supervised methods to predict patient survival from gene expression data [J]. PLoS Biol, 2004, 2: e108.

[44].YOSHIHARA K, SHAHMORADGOLI M, MARTÍNEZ E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data [J]. Nat Commun, 2013, 4: 2612.

[45].XU L, DENG C, PANG B, et al. TIP: a web server for resolving tumor immunophenotype profiling [J]. Cancer Res, 2018, 78: 6575-6580.

[46].JIANG P, GU S, PAN D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response [J]. Nat Med, 2018, 24: 1550-1558.

[47].GEELEHER P, COX N, HUANG RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels [J]. PLoS One, 2014, 9: e107468.

[48].SUBRAMANIAN A, NARAYAN R, CORSELLO SM, et al. A next generation Connectivity Map: L1000 platform and the first 1,000,000 profiles [J]. Cell, 2017, 171: 1437-1452.e17.

[49].LANGFELDER P, HORVATH S. WGCNA: an R package for weighted correlation network analysis [J]. BMC Bioinformatics, 2008, 9: 559.

[50].STUART T, BUTLER A, HOFFMAN P, et al. Comprehensive integration of single-cell data [J]. Cell, 2019, 177: 1888-1902.e21.

[51].LUECKEN MD, BÜTTNER M, CHAICHOOMPU K, et al. Benchmarking atlas-level data integration in single-cell genomics [J]. Nat Methods, 2022, 19: 41-50.

[52].OSORIO D, ZHONG Y, LI G, et al. scTenifoldKnk: an efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation [J]. Patterns (N Y), 2022, 3: 100434.

[53].OSORIO D, ZHONG Y, LI G, et al. scTenifoldNet: a machine learning workflow for constructing and comparing transcriptome-wide gene regulatory networks from single-cell data [J]. Patterns (N Y), 2020, 1: 100139.

[54].GUO Y, WANG B, DING G, et al. Comprehensive analysis of COLGALT1 in tumor microenvironment regulation and prognosis of clear cell renal cell carcinoma [J]. Clin Exp Med, 2026, 26: 127.

[55].LI Q, WEI K, ZHANG X, et al. TIMP1 shapes an immunosuppressive microenvironment by regulating anoikis to promote the progression of clear cell renal cell carcinoma [J]. Aging (Albany NY), 2023, 15: 8908-8929.

[56].SONG X, JIAO J, QIN J, et al. Single-cell transcriptomics reveals the heterogeneity and function of mast cells in human ccRCC [J]. Front Immunol, 2025, 15: 1494025.

[57].XIE Z, FENG C, HONG Y, et al. Identification of key genes CCL5, PLG, LOX and C3 in clear cell renal cell carcinoma through integrated bioinformatics analysis [J]. Front Mol Biosci, 2025, 12: 1587196.

[58].ZOU X, LIU X, WANG H, et al. Characterization of cuproptosis signature in clear cell renal cell carcinoma by single cell and spatial transcriptome analysis [J]. Discov Oncol, 2024, 15: 300.

[59].TAN X, CHEN Y, JIANG C, et al. Multi-omics analysis identified LGALS1 (Galectin-1) as a novel therapeutic target for clear cell renal cell carcinoma via targeting epithelial or endothelial cells [J]. Discov Oncol, 2025, 16: 1894.

[60].GAO ZW, YANG L, LIU C, et al. Distinct roles of adenosine deaminase isoenzymes ADA1 and ADA2: a pan-cancer analysis [J]. Front Immunol, 2022, 13: 903461.

[61].KRISHNA C, DINATALE RG, KUO F, et al. Single-cell sequencing links multiregional immune landscapes and tissue-resident T cells in ccRCC to tumor topology and therapy efficacy [J]. Cancer Cell, 2021, 39: 662-677.e6.

[62].LI H, ZHANG J, SHI Y, et al. Identification of matrix stiffness-related molecular subtypes in HCC via integrating multi-omics analysis and machine learning algorithms [J]. J Transl Med, 2025, 23: 716.

Figures

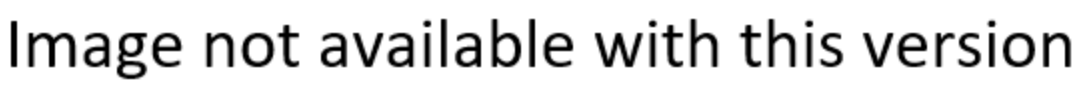


Figure 1

Figure 9: Legend not included with this version.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure.docx
- Table1.xlsx
- Table2.xlsx
- Table3.xlsx