



OPEN ACCESS

EDITED BY

Rushendhiran Kesavan,
University of Texas Southwestern Medical
Center, United States

REVIEWED BY

Rajesh Gunage,
Boston Children's Hospital and Harvard
Medical School, United States
Tatjana Sajic,
Centre Hospitalier Universitaire Vaudois
(CHUV), Switzerland

*CORRESPONDENCE

Jun Huang

✉ tzhuoyumin@163.com

[†]These authors share first authorship

RECEIVED 11 December 2024

ACCEPTED 07 February 2025

PUBLISHED 03 March 2025

CITATION

Li Y, Zeng R, Huang Y, Zhuo Y and Huang J
(2025) Integrating machine learning and
single-cell sequencing to identify shared
biomarkers in type 1 diabetes mellitus and
clear cell renal cell carcinoma.
Front. Oncol. 15:1543806.
doi: 10.3389/fonc.2025.1543806

COPYRIGHT

© 2025 Li, Zeng, Huang, Zhuo and Huang. This
is an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Integrating machine learning and single-cell sequencing to identify shared biomarkers in type 1 diabetes mellitus and clear cell renal cell carcinoma

Yi Li^{1†}, Rui Zeng^{2,3†}, Yuhua Huang⁴, Yumin Zhuo⁵
and Jun Huang^{4*}

¹Department of Ultrasound, The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou, China, ²Department of Pathology, School of Medicine, South China University of Technology, Guangzhou, China, ³Department of Pathology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Guangzhou, China, ⁴Department of Ultrasound, The First Affiliated Hospital of Jinan University, Guangzhou, China, ⁵Department of Urology, The First Affiliated Hospital of Jinan University, Guangzhou, China

Purpose: Type 1 diabetes mellitus (T1DM), as an autoimmune disease, can increase susceptibility to clear cell renal cell carcinoma (ccRCC) due to its proinflammatory effects. ccRCC is characterized by its subtle onset and unfavorable prognosis. Thus, the aim of this study was to highlight prevention and early detection opportunities in high-risk populations by identifying common biomarkers for T1DM and ccRCC.

Methods: Based on multiple publicly available datasets, WGCNA was applied to identify gene modules closely associated with T1DM, which were then integrated with prognostic DEGs in ccRCC. Subsequently, the LASSO and SVM algorithms were employed to identify shared hub genes between the two diseases. Additionally, clinical samples were used to validate the expression patterns of these hub genes, and scRNA-seq data were utilized to analyze the cell types expressing these genes and to explore potential mechanisms of cell communication.

Results: Overall, three hub genes (KIF21A, PIGH, and RPS6KA2) were identified as shared biomarkers for T1DM and ccRCC. Analysis of clinical samples and multiple datasets revealed that KIF21A and PIGH were significantly downregulated and that PIG was upregulated in the disease group. KIF21A and PIGH are mainly expressed in NK and T cells, PRS6KA2 is mainly expressed in endothelial and epithelial cells, and the MIF signaling pathway may be related to hub genes.

Conclusion: Our results demonstrated the pivotal roles of hub genes in T1DM and ccRCC. These genes hold promise as novel biomarkers, offering potential avenues for preventive strategies and the development of new precision treatment modalities.

KEYWORDS

type 1 diabetes mellitus, clear cell renal cell carcinoma, key genes, machine learning, single cell sequencing

1 Introduction

Diabetes mellitus (DM) affects approximately 463 million people worldwide, with type 1 diabetes mellitus (T1DM) accounting for nearly 10%. Remarkably, the prevalence of T1DM is increasing globally (1, 2). Although there is no definitive consensus on the exact pathogenic mechanism of T1DM, it is fundamentally characterized as a chronic autoimmune disorder involving the destruction of insulin-producing pancreatic islet cells (β cells) (3). By identifying autoantibodies, autoreactive CD8⁺ T cells are the primary immune cells responsible for the death of β cells, and islet-specific CD4⁺ T cells can produce cytokines that promote the activation of B cells and islet-specific CD8⁺ T cells (4, 5). Additionally, recent studies have emphasized the significance of natural killer cells (NKs) in autoimmune contexts. NKs can eliminate specific cells, generate memory cells, respond to antigens, and interact with CD8⁺ T cells, promoting immune-mediated assaults on β cells (6, 7).

T1DM increases cancer risk, and previous research has highlighted its connection to renal cell carcinoma (RCC). Coexisting T1DM in RCC patients increases the risk of recurrence and distant metastasis, contributing to a poorer prognosis (8). Clear cell renal cell carcinoma (ccRCC) constitutes up to 80% of all pathological subtypes of RCC (9), and nearly one-third of ccRCC patients present with metastases at the initial diagnosis (10). The link between RCC and diabetes has become increasingly evident over the years, both in terms of incidence and prognosis (11, 12). DM, a major consequence of metabolic syndrome, is more prevalent in ccRCC patients, and those with DM and ccRCC tend to have worse outcomes compared to ccRCC patients without DM. Diabetes is associated with higher recurrence rates, increased distant metastases, and reduced survival rates (8). Epidemiological studies also highlight that individuals with diabetes have a significantly higher risk of RCC compared to non-diabetic individuals (13). This connection is particularly prominent in conditions associated with metabolic syndrome, such as T1DM.

The tumor microenvironment (TME) in ccRCC is characterized by a substantial presence of tumor-infiltrating immune cells (14). Although a connection between T1DM and immune cell activity exists, the exact extent of its influence on the TME still needs to be understood. Therefore, it is imperative to assess the potential impact of other related diseases on the TME. Therefore, employing machine learning techniques that integrate multiple diseases can facilitate a more comprehensive analysis of meaningful biomarkers (15, 16). The primary aim of this study was to identify common biomarkers shared between T1DM and ccRCC, with the objective of establishing a more comprehensive early diagnostic model for ccRCC based on novel potential serum biomarkers. Another objective of this study was to identify biomarkers with increased sensitivity and specificity for the early screening and diagnosis of ccRCC in T1DM patients. Additionally, we hypothesized that these identified markers may play a role in the immune mechanism underlying the pathogenesis of ccRCC. In this work, we endeavored to provide new insights into the impact of the TME.

In this study, we used a variety of machine learning algorithms and weighted gene coexpression network (WGCNA) methods to

screen out possible shared biomarkers between T1DM and ccRCC. We further identified the core genes closely related to the diagnosis and prognosis of T1DM patients and ccRCC patients and verified the expression levels of these genes in clinical samples. Immune infiltration analysis revealed that these core genes were closely related to the immune response of NK cells and T cells in the immune microenvironment of ccRCC, and single-cell RNA sequencing (scRNA-seq) data confirmed this result (Figure 1). These findings reveal potential alterations in the immune microenvironment of ccRCC in the context of T1DM and provide a new target for the prevention and treatment of ccRCC.

2 Materials and methods

2.1 Transcriptome data acquisition and processing

Transcriptome RNA sequencing (RNA-seq) data of ccRCC samples were obtained from The Cancer Genome Atlas (TCGA). The RNA-seq data selected for analysis were normalized using the Fragments Per Kilobase of Transcript Per Million Mapped Reads (FPKM) method. After eliminating duplicated data entries and samples with missing information, the final TCGA-KIRC cohort comprised 29 normal samples and 400 cancer samples, of which 141 were from females and 259 were from males, aged 68.0 ± 10.64 years. All patients had accompanying clinical information.

The T1DM peripheral blood mononuclear cell (PBMC) transcriptome dataset GSE55098, the T1DM validation cohort GSE9006, and the ccRCC transcriptome validation cohort GSE53757 were obtained from the GEO database. The GPL96 and GPL97 platforms from the UTSW Medical Center were used for GSE9006. Our study used data based on the GPL97 platform, including 24 normal PBMC samples and 43 PBMC samples from T1DM patients (26 women and 17 men; aged 10.1 ± 3.8 years). The GPL97 platform from Shanghai JiaoTong University Ruijin Hospital was used for GSE53757, which included 10 normal PBMC samples and 12 PBMC samples from patients with T1DM (5 women and 7 men; aged 17.5 ± 3.68 years). The GPL570 platform from The Mayo Clinic was used for GSE53757, which consists of 72 normal samples and 72 ccRCC samples. Details of the dataset are shown in [Supplementary Table S1](#). The demographic characteristics of the TCGA-KIRC cohort involved in this study are shown in [Supplementary Table S4](#).

2.2 Weighted gene coexpression network analysis

We chose WGCNA as the analysis method for the T1DM transcriptome data because this method can identify coexpressed gene modules closely related to T1DM as a whole, revealing the pathway by which T1DM influences the development and prognosis of ccRCC. We employed the “WGCNA” package (17) to construct a weighted adjacency matrix by selecting appropriate thresholds. Subsequently, the weighted adjacency matrix was

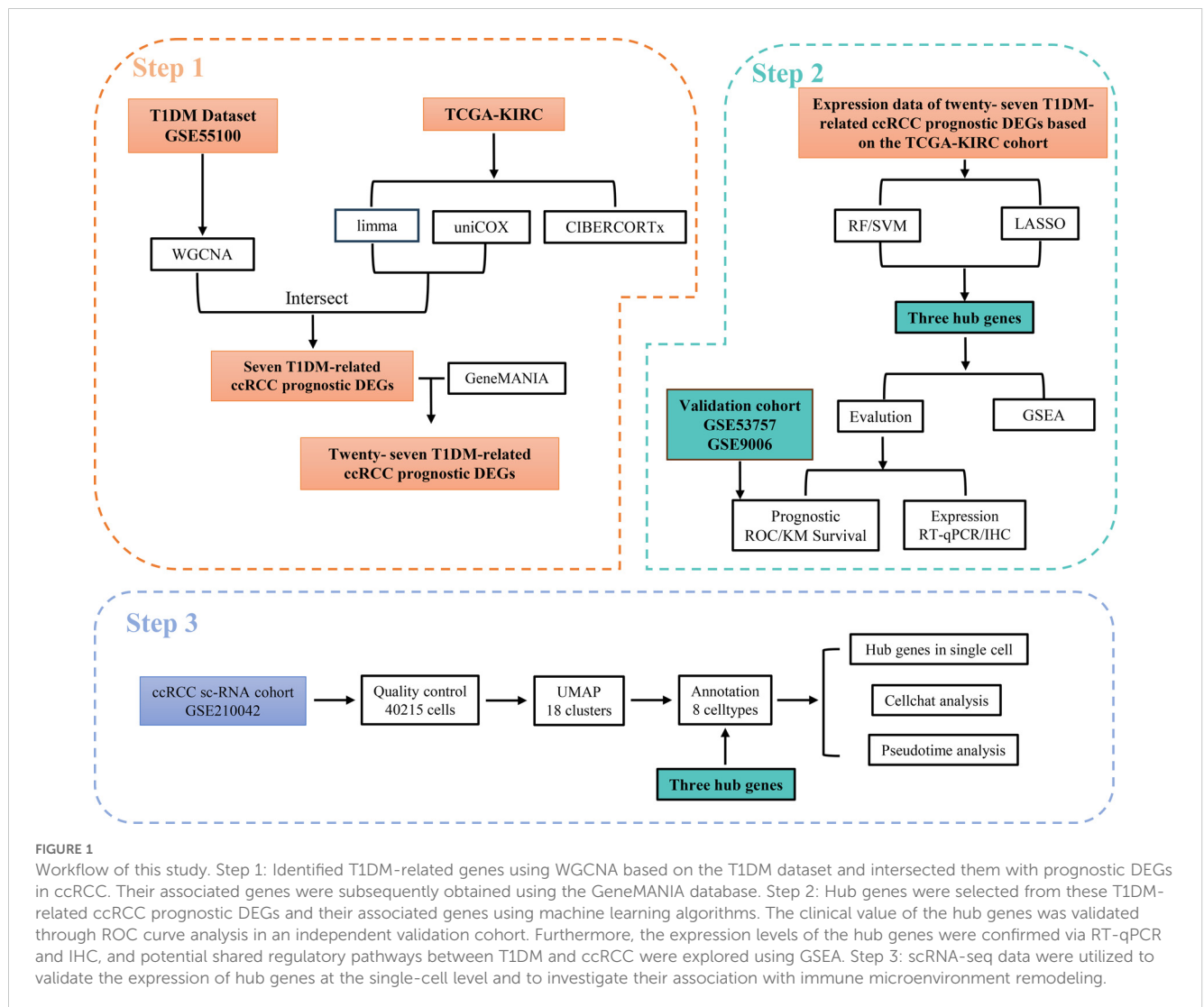


FIGURE 1

Workflow of this study. Step 1: Identified T1DM-related genes using WGCNA based on the T1DM dataset and intersected them with prognostic DEGs in ccRCC. Their associated genes were subsequently obtained using the GeneMANIA database. Step 2: Hub genes were selected from these T1DM-related ccRCC prognostic DEGs and their associated genes using machine learning algorithms. The clinical value of the hub genes was validated through ROC curve analysis in an independent validation cohort. Furthermore, the expression levels of the hub genes were confirmed via RT-qPCR and IHC, and potential shared regulatory pathways between T1DM and ccRCC were explored using GSEA. Step 3: scRNA-seq data were utilized to validate the expression of hub genes at the single-cell level and to investigate their association with immune microenvironment remodeling.

transformed into a topological overlap matrix (TOM). To identify modules with the most significant correlation with T1DM, we applied the dynamic tree-cutting algorithm to divide the network into distinct modules and extracted the genes within the modules.

2.3 Identification of differentially expressed genes and ccRCC prognostic genes

The “limma” package (18) was used to identify differentially expressed genes (DEGs) between normal and diabetic samples in the GSE55098 dataset and between normal and tumor samples in the TCGA-KIRC cohort. Genes with an adjusted p value < 0.05 and $|\log_2FC| > 1$ were considered DEGs. Then, we performed univariate Cox regression analysis using the “survival” package to identify prognostic genes (p value < 0.05). Genes meeting the criteria of both DEGs in ccRCC, specific prognostic genes for ccRCC, and present in the modules most strongly correlated with T1DM were classified as T1DM-related prognostic DEGs for subsequent analysis.

2.4 GeneMANIA-associated analysis and PPI network construction

Due to the limited number of T1DM-related prognostic DEGs, directly applying them to machine learning models may lead to unreliable results. Drawing on the approach adopted in previous studies (19–21), the GeneMANIA database (22) was used to identify genes associated with T1DM-related prognostic DEGs. The STRING database (23) and Cytoscape were used to construct a protein-protein interaction (PPI) network.

2.5 Functional enrichment analysis

To investigate the pathways shared between T1DM and ccRCC, we performed Gene Ontology (GO) term enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses using the Metascape database. These analyses were based on T1DM-related prognostic DEGs and their associated genes (24).

2.6 Hub gene screening and validation based on the machine learning algorithm

We applied three machine learning methods to identify key biomarkers associated with ccRCC prognosis from T1DM-associated ccRCC prognostic DEGs and their associated genes based on the TCGA-KIRC cohort. First, we evaluated the efficiency of the random forest (RF) and support vector machine recursive feature elimination (SVM-RFE) algorithms in the TCGA-KIRC cohort through residual and receiver operating characteristic (ROC) curve analyses. We then selected the genes identified by one of these algorithms. It is important to clarify that our study's machine learning models were trained using the presence or absence of renal cancer as the response variable, with 20 ccRCC biomarkers related to T1DM as covariates. Notably, we utilized the “randomForest” package (25) for constructing the random forest model and the “Kernlab” package (26) for constructing the SVM-RFE model. The random forest model employed default feature selection based on mean decrease impurity, while the SVM-RFE model employed recursive feature impurity removal for feature selection. We implemented K-fold cross-validation for feature extraction, ensuring that both the SVM/RF and least absolute shrinkage and selection operator (LASSO) models utilized this data splitting method for accurate and reliable feature comparisons. Subsequently, LASSO regression was performed, and we selected one of the algorithms according to the advantages and disadvantages of RF or SVM-RFE and intersection with the genes found in LASSO regression analysis as the hub genes.

2.7 Evaluation of the prognostic efficacy of the hub genes

The prognostic efficacy of the hub genes in T1DM patients and ccRCC patients was evaluated by constructing ROCs and calculating the corresponding area under the curve (AUC) using the “pROC” package. After grouping based on the cutoff values calculated by the “survival” package, Kaplan–Meier survival curves corroborated a substantial correlation between the hub genes and the survival outcomes of ccRCC patients.

2.8 Gene set enrichment analysis of the hub genes in patients with T1DM and ccRCC

GSEA is a computational method used to determine whether predefined sets of genes show statistically significant, coordinated differences between two biological states. We conducted single-gene GSEA of the hub genes in the T1DM cohort GSE55098 and TCGA-KIRC cohort using the “fgsea” package (27), with the hallmark gene set obtained from the MsigDB database (28, 29). Specifically, samples were divided into high- and low-expression groups based on the expression levels of the hub genes. Differential expression analysis was then conducted to identify DEGs between these two

groups. GSEA was subsequently applied to evaluate the association between the target genes and specific biological processes or pathways.

2.9 Assessing the role of the hub genes in the tumor microenvironment

The “Cibersort” package (30) was used to analyze the abundance ratios of 22 types of immune cells in the TCGA ccRCC cohort and to evaluate the interactions between different immune cells. Furthermore, the associations between the hub genes and specific immune cells were investigated.

2.10 Single-cell RNA sequencing analysis

We obtained ccRCC scRNA-seq data comprising seven tumor samples and two adjacent normal tissue samples from the GSE210042 dataset in the GEO database. In this study, we only used single-cell data from tumor samples. Data preprocessing was conducted using the “Seurat” package (version 4.1.0) (31) with the following criteria for the tumor samples: nFeature_RNA > 200, nCount_RNA > 1000, and fewer than 20% of mitochondrial genes. A total of 50,487 cells were obtained for subsequent analysis. The top 2000 highly variable genes were identified using the “FindVariableGenes” function, and principal component analysis was subsequently conducted. Cell types were annotated based on known renal-specific marker genes from a previous study (32) or queried in PanglaoDB (33) to assign cell types based on cluster-specific marker genes. Pseudotime analysis was performed using the “monocle3” package (34). We examined hub gene expression across all cell clusters and explored cellular crosstalk using the “CellChat” package (35).

2.11 Human renal RNA extraction and quantitative real-time polymerase chain reaction

Tumor and adjacent normal tissues were collected from 12 ccRCC patients who underwent radical nephrectomy, and RNA was extracted from these tissues. This study was approved by the Ethics Committee of the First Affiliated Hospital of Jinan University, and written informed consent was obtained from all patients and control individuals.

Total RNA was extracted using the EZ-Press RNA Purification Kit (EZbioscience, USA). cDNA was synthesized using the PrimeScript RT Kit (TaKaRa, Japan) through reverse transcription. RT–qPCR detection was performed using a CFX96 real-time PCR system (Bio-Rad, USA) following the SYBR Green method (Vazyme, China). After normalization to GAPDH expression, relative mRNA expression levels were determined. The mRNA-specific primer sequences can be found in [Supplementary Table S2](#).

2.12 Statistical analysis

Statistical analyses were conducted using R 4.1.1 and GraphPad Prism 8 (GraphPad Software, Inc.), results are based on two-group comparisons, and no multiple-group comparisons were performed. Continuous variables were described as the mean standard deviation (SD) when normally distributed or median (interquartile range, IQR) when not, and categorical variables were described as numbers (percentage). The means of medians for continuous variables in two independent groups were tested by the Wilcoxon rank-sum test. The enumeration data analyzed by using t-test. $p < 0.05$ was considered statistically significant. The error bars indicate the standard deviation in the figures.

3 Results

3.1 Identification of T1DM-related prognostic DEGs

WGCNA was used to determine an optimal soft threshold ($\beta = 9$) for constructing a scale-free network (Figure 2A) after multiple iterations. After excluding the gray module, which contained genes not involved in clustering, we identified 32 modules. Correlation analysis between these modules and T1DM highlighted the magenta module, consisting of 189 genes, as having the highest correlation with T1DM (correlation coefficient = 0.75, $P < 0.001$) (Figure 2B). A scatterplot further illustrates the correlation coefficient (0.75, $P < 0.001$) between GS and MM within the magenta module (Figure 2C).

In the TCGA-KIRC cohort, we identified 1521 DEGs, comprising 947 upregulated and 574 downregulated DEGs (Figure 2D; Supplementary Table S3). Subsequently, 1554 prognosis-related genes were identified through univariate Cox regression. The intersection of DEGs, prognostic genes, and magenta module genes revealed seven genes designated as T1DM-related prognostic DEGs (Figure 2E).

3.2 Identification of genes associated with the prognosis of patients with T1DM and enrichment analysis

A set of 20 genes associated with the T1DM-related prognostic DEGs was identified using GeneMANIA (Figure 3A). Subsequently, a PPI network was visualized through Cytoscape (Figure 3B). To gain insight into the functional implications of these genes, enrichment analysis was performed via the Metascape database. The results indicated that the following pathways may represent pathways shared between T1DM patients and ccRCC patients: the mitogen-activated protein kinase (MAPK) signaling pathway, the biosynthesis of cofactors, protein domain-specific binding,

phosphotransferase activity, and the modulation of chemical synaptic transmission (Figures 3C, D).

3.3 Selection of shared biomarkers between T1DM patients and ccRCC patients using multiple machine learning methods

We first used the RF and SVM-RFE models for gene selection. Supplementary Figure S1A shows gene selection based on the minimum error point in the RF model ($X=152$). The RF method exhibited superior machine learning capabilities, as evidenced by smaller residuals (Supplementary Figures S1B–C) and higher AUC values (Supplementary Figure S2A) than those of the SVM. Consequently, leveraging the RF method, we identified seven candidates from 27 T1DM-related prognostic differentially expressed genes (DEGs) and their associated genes. Furthermore, LASSO regression showed high AUC values in the TCGA-KIRC cohort (Supplementary Figure S2B), and identified nine genes with the lowest binomial deviance (Supplementary Figures S1D, E). The intersection of genes detected by both the RF method and LASSO regression revealed three hub genes, KIF21A, PIGH, and RPS6KA2, as illustrated in a Venn diagram (Supplementary Figure S1F).

3.4 Validation of the expression patterns of the hub genes

We validated the expression of three hub genes in both T1DM patients and ccRCC patients. Similarly to PIGH, KIF21A was expressed at low levels in both T1DM (Figure 4A) and ccRCC (Figure 4D) patients (Figures 4B, E). Conversely, the opposite trend was observed for RPS6KA2 (Figures 4C, F).

To address the issue of patient diversity, the study analyzed the demographic characteristics of the TCGA-KIRC cohort (Supplementary Table S4), which primarily includes White and Black patients. While no significant effects of these factors on gene expression were observed, however, the potential influence of ethnicity on disease mechanisms should not be ignored. Thus, to mitigate this limitation, we validated our findings using RT-qPCR on clinical samples from Asian patients, demonstrating the applicability of our results across diverse populations. This partially addresses the underrepresentation in public datasets. By RT-qPCR, we confirmed the downregulation of KIF21A (Figure 5A) and PIGH (Figure 5B) and the upregulation of RPS6KA2 (Figure 5C) in ccRCC tissues compared to normal tissues. Immunohistochemical (IHC) analysis of samples from the Human Protein Atlas (HPA) database further confirmed these trends: KIF21A (Figure 5D) and PIGH (Figure 5E) exhibited low expression, while RPS6KA2 (Figure 5F) exhibited high expression in ccRCC tissues compared to normal tissues. These findings

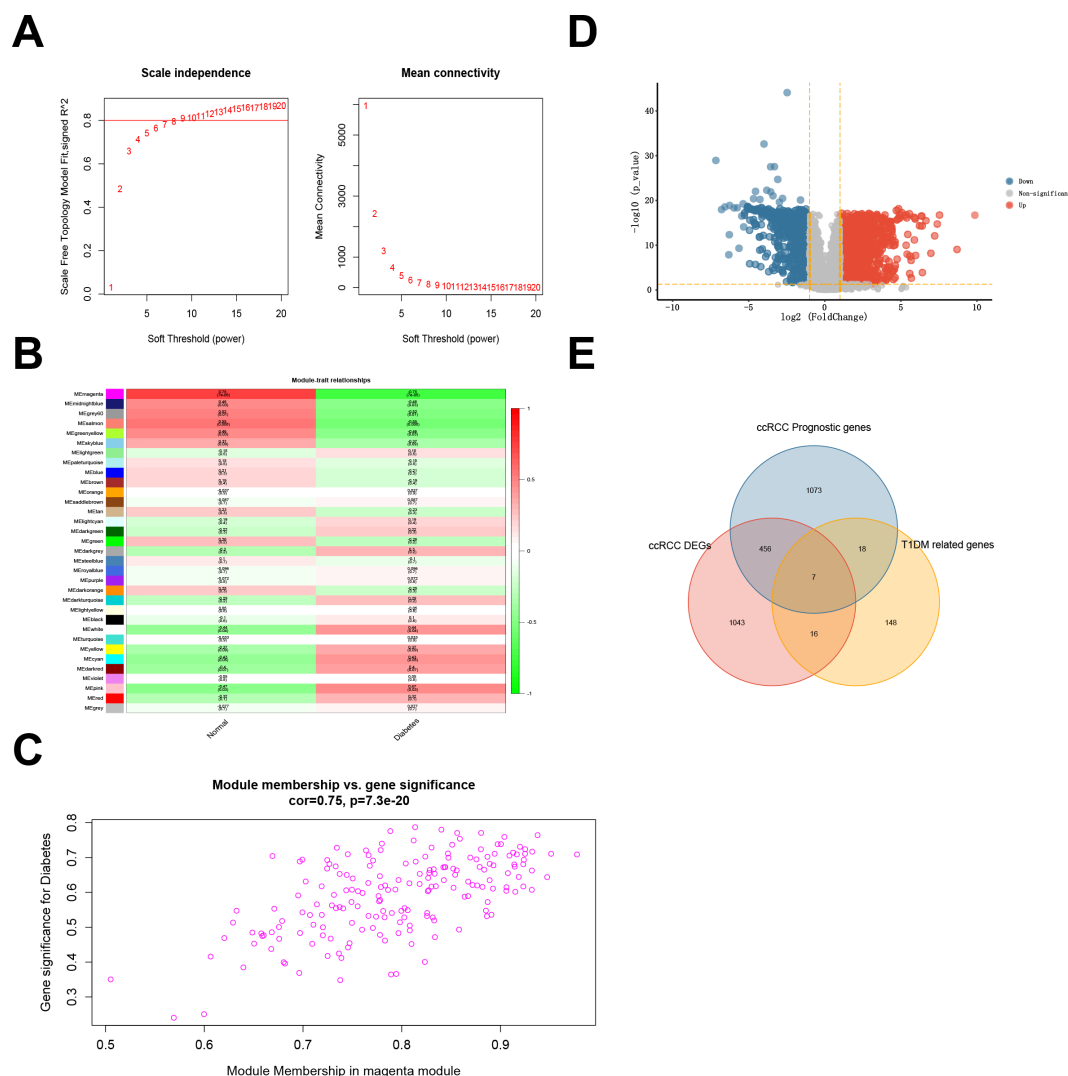


FIGURE 2

(A) Fitting index of the scale-free topology module under different soft thresholds (left) and network connectivity under different soft thresholds (right). (B) Correlations between module eigengenes and type 1 diabetes. (C) Scatterplot showing the relationship between gene significance (GS) for type 1 diabetes and module membership (MM) in the magenta module. (D) Differentially expressed genes (DEGs) in ccRCC samples compared with normal samples in the TCGA-KIRC cohort. (E) Intersection between the DEGs and prognostic genes from the TCGA-KIRC cohort and the magenta module genes from the T1DM GEO dataset.

provide strong evidence for consistent expression patterns of these three hub genes in both T1DM patients and ccRCC patients.

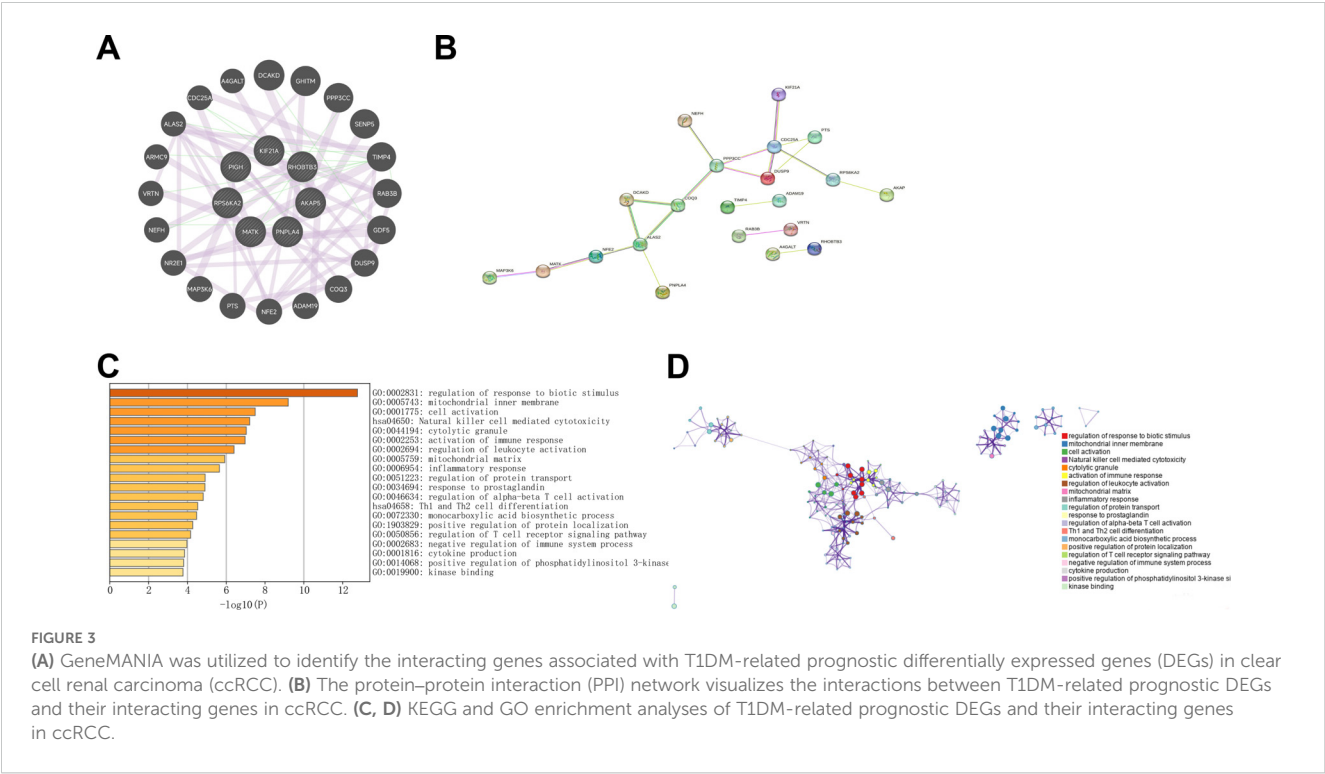
3.5 Validation of the prognostic efficacy of the hub genes

Univariate Cox regression analysis was performed to validate the prognostic significance of the hub genes in both T1DM patients and ccRCC patients. KIF21A, PIGH, and RPS6KA2 demonstrated significant prognostic performance in T1DM patients, with AUC values of 0.933, 0.858, and 0.908, respectively (Figure 6A). In ccRCC patients, the AUC values were 0.964, 0.992, and 0.826, respectively (Figure 6B). K-M survival curves confirmed the significant correlation between these genes and the survival of ccRCC patients (Figure 6C).

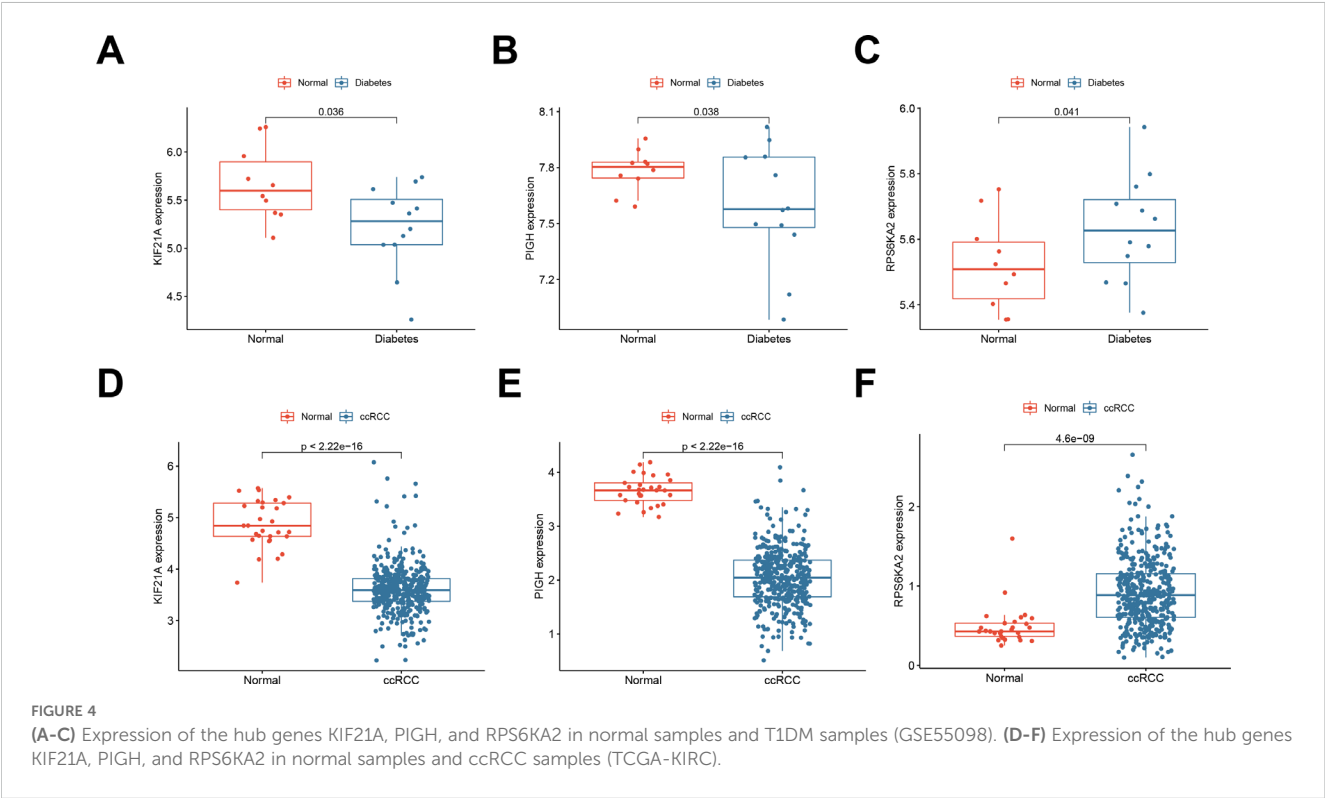
We also assessed the prognostic performance of the hub genes in independent validation datasets for both T1DM (GSE9006) and ccRCC (GSE53757) patients. In T1DM patients, the AUC values for KIF21A, PIGH, and RPS6KA2 were 0.783, 0.800, and 0.850, respectively (Figure 7A). In ccRCC patients, the AUC values were 0.972, 0.912, and 0.839, respectively (Figure 7B). Combining the results from the original and validation cohorts, it is evident that these hub genes exhibit superior prognostic efficacy in both T1DM patients and ccRCC patients.

3.6 Single-gene GSEA of the hub genes in patients with T1DM and ccRCC

We performed single-gene GSEA of the hub genes in the T1DM and ccRCC cohorts.



The results showed that the hub genes were involved in the estrogen response, epithelial-mesenchymal transition, inflammatory response, angiogenesis, KRAS signaling, and other pathways in both disease groups (Figures 8A, B). The identified pathways highlight shared mechanisms underlying T1DM and ccRCC, supporting KIF21A, PIGH, and RPS6KA2 as shared biomarkers. Such as estrogen signaling enhances insulin-like growth factor-1 receptor activity, exacerbating the effects of elevated insulin and promoting ccRCC development (36). Increased oxidative phosphorylation activates the pyruvate dehydrogenase complex, reprogramming glucose metabolism and the TCA cycle, thereby facilitating tumor growth and metastasis



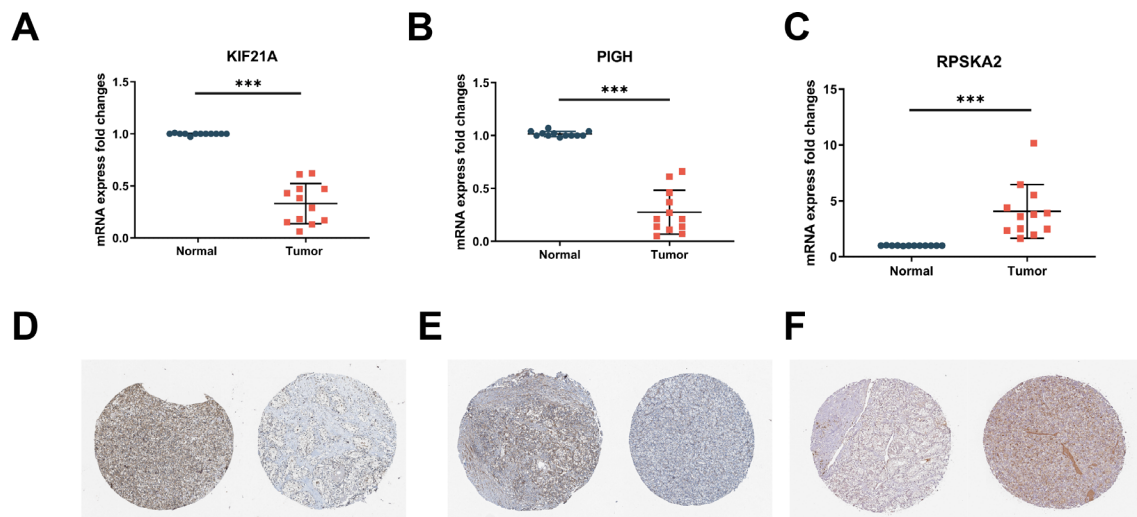


FIGURE 5 (A–C) The mRNA expression levels of KIF21A (A), PIGH (B), and RPS6KA2 (C) in clinical samples were detected by RT–qPCR. (D–F) Immunohistochemistry of KIF21A (D), PIGH (E), and RPS6KA2 (F) in normal tissues (left) and ccRCC tissues (right) from the HPA database. *** indicates $P < 0.001$.

(37). Additionally, partial complement system activation increases immune cell infiltration in the tumor microenvironment while promoting immune evasion, ultimately driving distant metastasis (38). These shared pathways underscore the interconnectedness of T1DM and ccRCC, with dysregulated mechanisms contributing to ccRCC risk and progression in T1DM patients. The roles of KIF21A, PIGH, and RPS6KA2 in these pathways reinforce their potential as diagnostic and therapeutic biomarkers.

3.7 Associations between immune cell infiltration and hub genes in ccRCC

We compared immune cell infiltration between the normal and ccRCC groups in the TCGA-KIRC cohort (Figure 9A). Most immune cells, including memory B cells, NK cells, T helper cells, dendritic cells, and monocytes, exhibited increased infiltration in the normal group. Conversely, immunosuppressive cells such as Tregs and resting immune cells were more abundant in the ccRCC group (Figure 9B).

We also examined the correlation between the hub genes and immune cells infiltrating ccRCC. KIF21A and PIGH exhibited positive correlations with activated NK cells, follicular helper T cells, and $\gamma\delta$ T cells, while they had negative correlations with eosinophils, resting NK cells, naive CD4 T cells, and Tregs, and RPS6KA2 was positively correlated with NK cells resting. It was negatively correlated with NK cells activated and M1 macrophages (Figures 9C–E). Notably, there was a significant positive association between KIF21A and PIGH with activated NK cells, which are immune cells with broad-spectrum anti-tumor effects, while RPS6KA2 was negatively correlated with its activation (Figures 9F–H). Given the low expression of KIF21A and PIGH in ccRCC and the high expression of RPS6KA2, KIF21A and PIGH

may act as tumor suppressor genes that promote NK cell activation in ccRCC, while RPS6KA2 may function as an oncogene that inhibits NK cell activation.

Finally, we analyzed the interactions between different immune cells and revealed positive correlations between activated NK cells and TFH cells, $\gamma\delta$ T cells, and resting mast cells but negative correlations between eosinophils and naive CD4 T cells (Figure 9I).

3.8 Single-cell sequencing analysis

After quality control, we analyzed the ccRCC microenvironment using scRNA-seq data, which involved clustering 40215 cells into 18 distinct clusters (Figure 10A). These 18 clusters were further categorized into eight cell populations through cell cluster annotation (Supplementary Figure S1) (Figure 10B). Pseudotime analysis revealed the pseudotime trajectory of these cell populations (Figure 10C). Our single-cell analysis showed that KIF21A was primarily localized in NK and T cells, PIGH was mainly expressed in NK, T, and endothelial cells, and RPS6KA2 targeted epithelial and endothelial cells (Figure 10D).

We also investigated interactions among these eight cell populations and noted a strong correlation between NK cells and CD8+ T cells (Figure 10E). CellChat analysis revealed interaction patterns among tumor, immune, and stromal cells in the ccRCC microenvironment. In pathways such as the SPP1, THBS, and GDF pathways, epithelial cells act as prominent senders, while NK cells and T cells act as receivers. Conversely, in the CD99 pathway, NK and T cells were senders, and epithelial cells were receivers (Figure 10F). Notably, in the macrophage migration inhibitory factor (MIF) pathway, epithelial cells were the strongest sender, indicating robust MIF pathway activity between epithelial cells, NK cells, and T cells (Figure 10G).

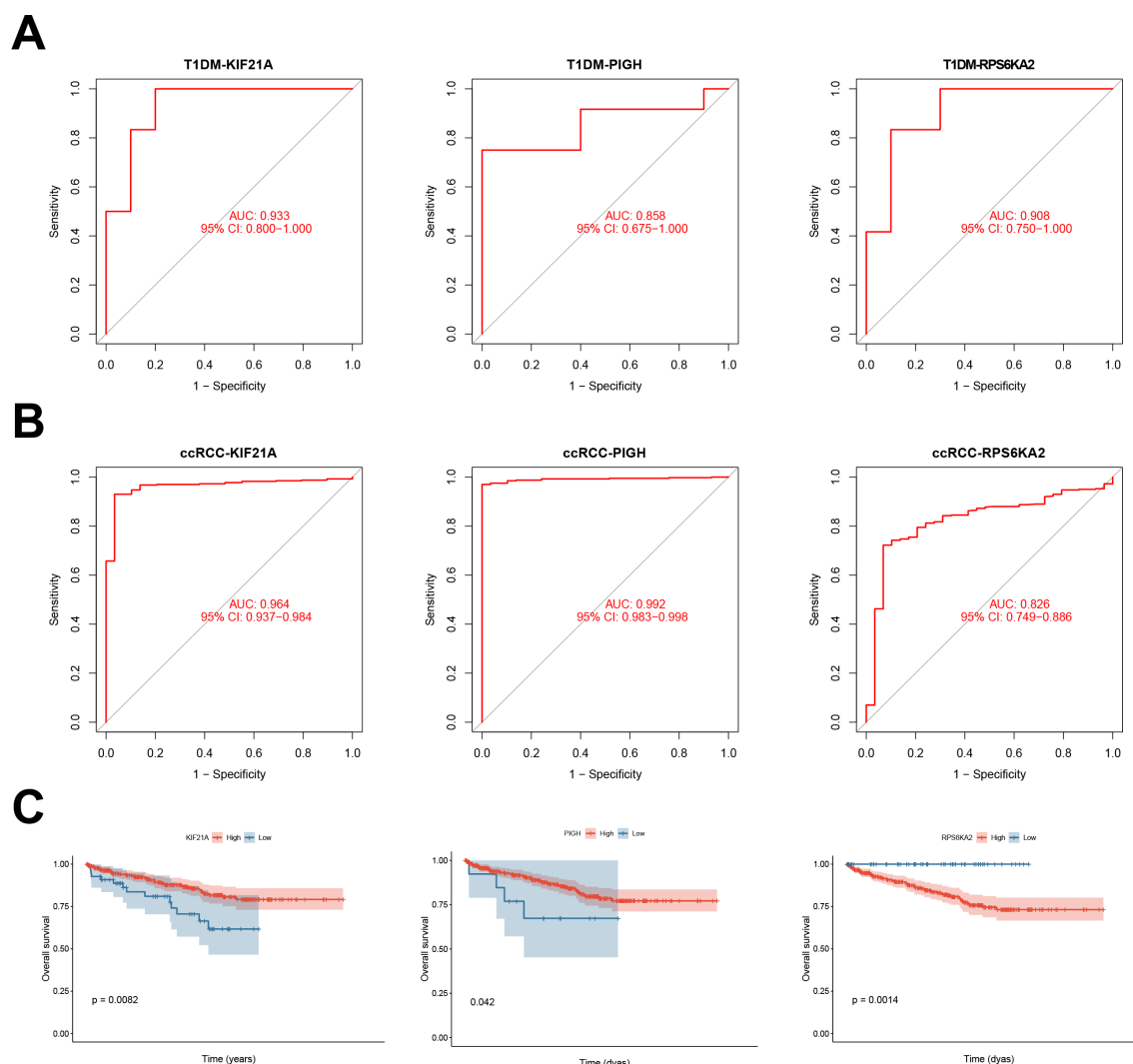


FIGURE 6

(A) ROC curves showing the prognostic performance of KIF21A, PIGH, and RPS6KA2 in patients with T1DM. (B) ROC curves depicting the prognostic performance of KIF21A, PIGH, and RPS6KA2 in ccRCC. (C) Kaplan-Meier survival curves illustrating the prognostic impact of KIF21A, PIGH, and RPS6KA2 in ccRCC patients.

4 Discussion

Currently, the precise pathological mechanism underlying ccRCC remains elusive; however, gaining insights into the mechanisms driving the initiation of ccRCC is essential for formulating effective prevention strategies. Despite its relatively low incidence in the population, however, the onset of ccRCC is insidious, and its propensity for metastasis intensifies in advanced stages. Even after surgical resection, the recurrence rate of ccRCC is approximately 20–40%. For those with metastatic ccRCC, despite the combination of chemotherapy, the prognosis remains grim (39). Therefore, early detection and prevention of ccRCC has important clinical significance, especially for high-risk groups such as diabetic patients.

T1DM is an autoimmune disorder characterized by the progressive destruction of β cells (5). The human leukocyte antigen risk alleles present islet antigens to CD4⁺ T cells,

initiating an immune response against β cells. Additionally, CD8⁺ T cells are stimulated by cytokines produced by B and T cells (40, 41). Consequently, as the primary cytotoxic effectors in specific immune responses, CD8⁺ T cells are pivotal for destroying β cells during the progression from insulinitis to T1DM (42).

Recent studies indicate that the incidence of T1DM is progressively increasing; while it predominantly develops in children and young adults, many patients are diagnosed with this condition during adulthood. In such cases, the destruction of β cells is variable, exhibiting characteristics of both T1DM and T2DM. This unique manifestation may be associated with the penetrance (43, 44) of the immune system. Notably, long-term DM is correlated with an elevated risk of numerous cancers (45). In patients with long-term DM, metabolic syndrome mainly includes obesity and high blood sugar. Obesity triggers changes in the immune microenvironment, including hyperinsulinemia, elevated insulin-like growth factor levels, and chronic inflammation, which

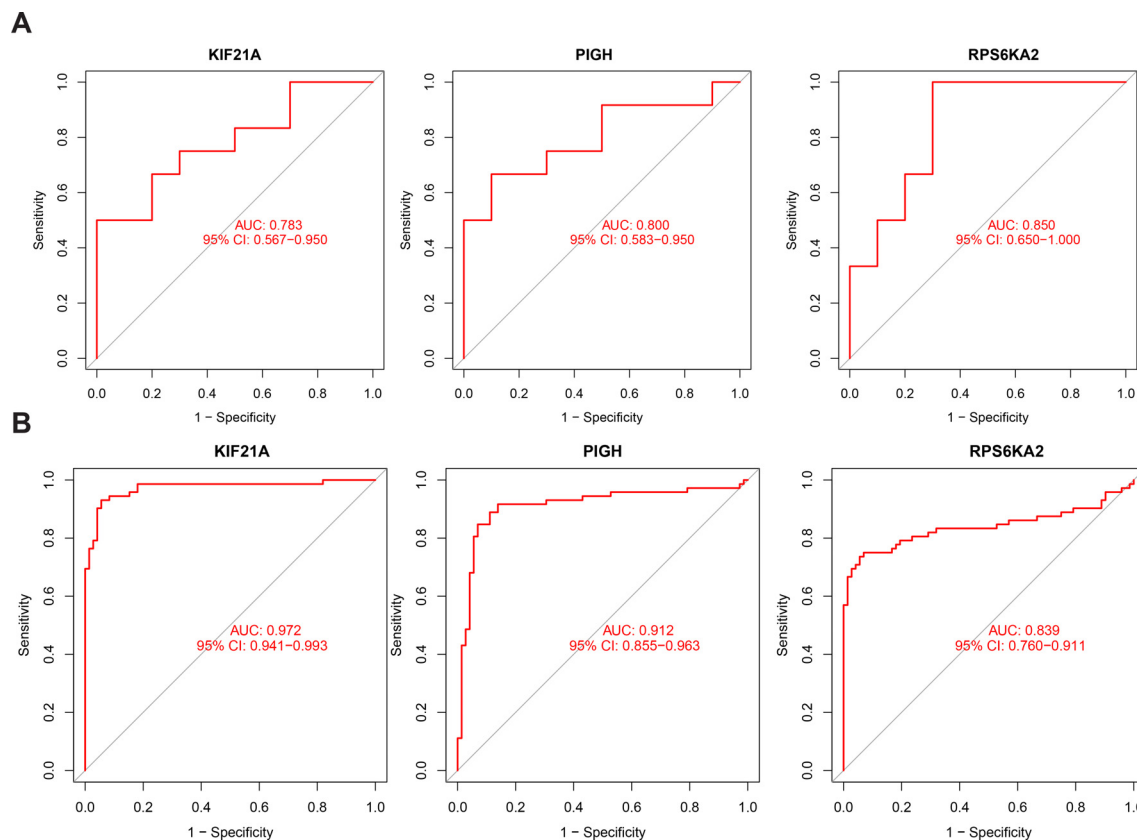


FIGURE 7

(A) ROC curve of KIF21A, PIGH, and RPS6KA2 in the validation cohort for T1DM. (B) ROC curve of KIF21A, PIGH, and RPS6KA2 in the validation cohort for ccRCC.

collectively contribute to cancer progression (46). Many cancers, including ccRCC, rely on increased carbohydrate metabolism. Cancer cells in hyperglycemic patients can absorb more glucose, leading to increased proliferation, metastasis, and poor prognosis (47), and hyperglycemia induces oxidative stress and DNA damage, promoting genomic instability and cancer initiation (48). Thus, as a chronic proinflammatory condition, DM significantly affects the tumor microenvironment, promoting cancer occurrence (49).

By integrating multiple transcriptome data and ccRCC scRNA-seq data, we used multiple machine learning methods to screen three biomarkers that are common between T1DM patients and ccRCC patients (KIF21A, PIGH, and RPS6KA2), which are linked to tumor and immune infiltration, and to predict the trajectory and prognosis of T1DM patients and ccRCC patients in an independent validation cluster. Our results revealed multiple pathways linking T1DM with ccRCC, which may be related to increased inflammation and an adverse prognosis for patients with ccRCC. For example, in the MAPK pathway, both diseases may involve several cascade protein kinases that coordinate intracellular signaling. This pathway influences lymphocyte differentiation, activation, and inhibition (50). Overactivation of the MAPK pathway could promote the transformation of renal cysts to ccRCC and enhance the proliferation of cancer cells, thereby enhancing the progression of the disease (51).

Among these biomarkers, KIF21A plays a pivotal role in microtubule assembly, and research has shown that its activity is significantly decreased in several cancers, subsequently influencing the regulation of cell polarity and migration (52). Previous investigations have shed light on the role of KIF21A in lung cancer. The expression of KIF21A decreases owing to DNA methylation in patients with lung cancer. This decrease in expression heightens the risk of distant metastasis of the cancer (53).

PIGH is an anchoring mechanism for numerous cell membrane-bound proteins and plays an integral role in synthesizing glycosyl phosphatidylinositol (54). Downregulation of PIGH expression leads to heightened chemotherapy resistance and bolsters the phenomenon of immune escape within the realm of cancer (55).

RPS6KA2, a serine/threonine protein kinase family member, is substantially overexpressed in several cancer types, including prostate, breast, and pancreatic cancers. Intriguingly, such elevated RPS6KA2 levels have been associated with fortified drug resistance and the reversal of cancer cell apoptosis typically induced by chemotherapy (56).

Consistent with the above conclusion, our study showed that ccRCC patients with low KIF21A and PIGH expression and elevated RPS6KA2 levels exhibit a poorer prognosis and overall

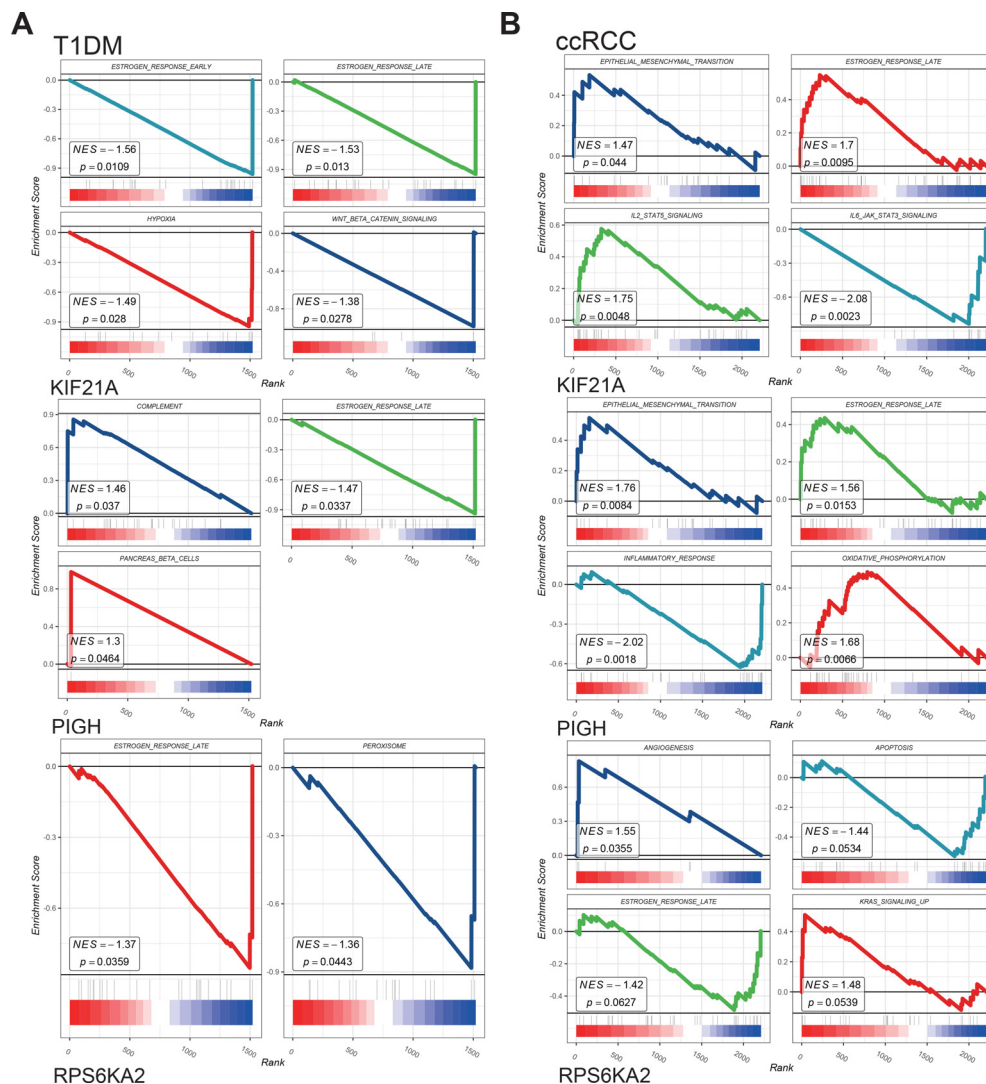
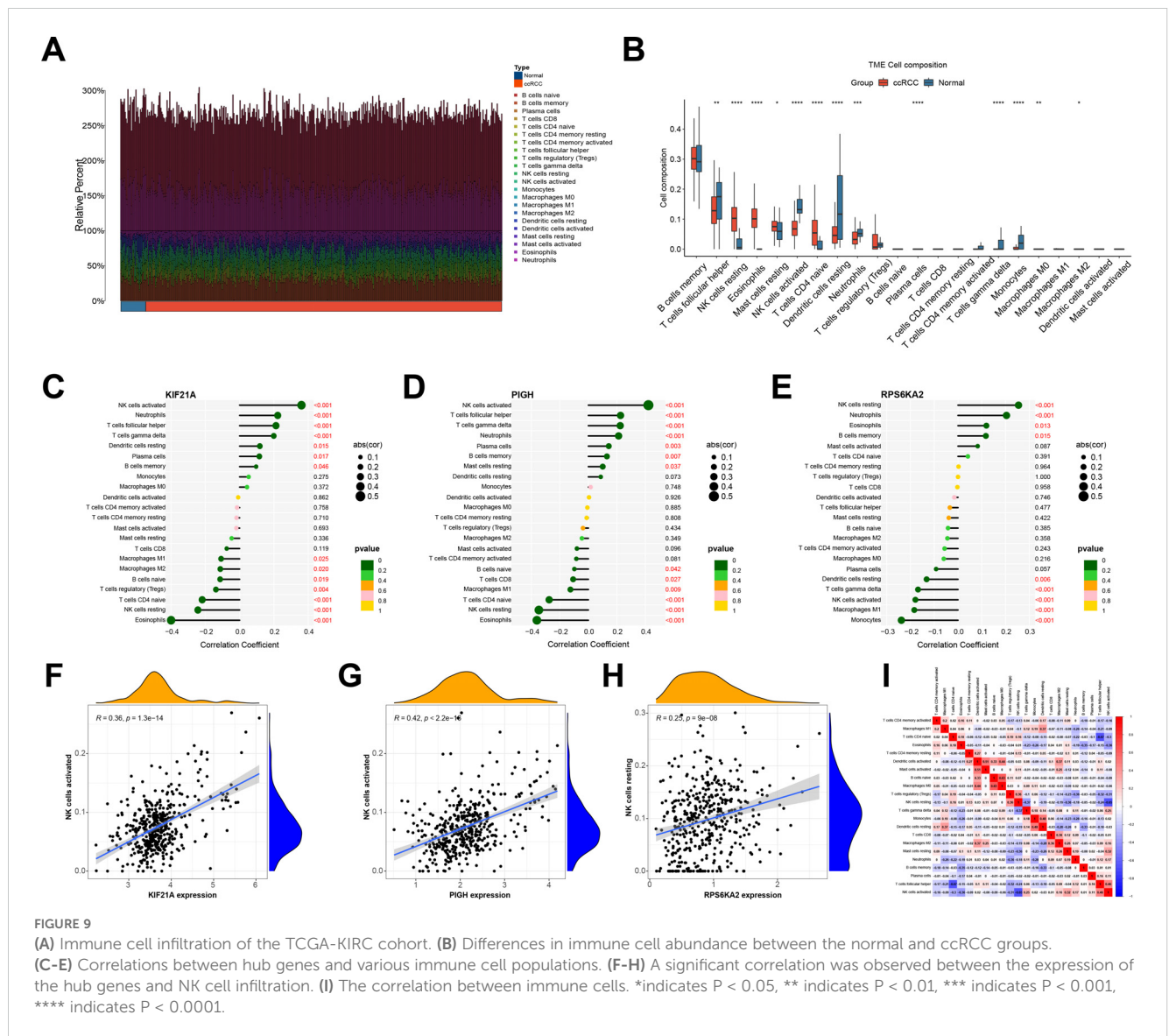


FIGURE 8
(A) GSEA analysis for hub genes in T1DM. (B) GSEA analysis for hub genes in ccRCC.

survival than their low-risk counterparts. The analysis of the scRNA-seq data revealed that KIF21A and PIGH predominantly localize to NK cells, CD4+ T cells, and CD8+ T cells, and decreased expression of these genes may reduce the activation of NK cells and reduce the number of NK cells in patients with T1DM, increasing the risk of cancer onset (6, 57). RPS6KA2 is mainly localized in endothelial and epithelial cells. Single-gene GSEA revealed that RPS6KA2 was significantly enriched in pathways associated with angiogenesis and KRAS signaling in the TCGA cohort with high RPS6KA2 expression. Therefore, RPS6KA2 may play a role in two ways: on the one hand, it may promote endothelial cell angiogenesis, and on the other hand, it may promote the progression of malignant epithelial cells by activating the KRAS signaling pathway; on the other hand, in diabetic patients, hyperglycemia and inflammation trigger endothelial dysfunction, which may eventually lead to vascular remodeling and renal damage (58). Given that RPS6KA2 is localized primarily within

endothelial cells, RPS6KA2 may play a role in transforming tumor-derived endothelial cells into cancer cells and promoting vascularization. The vascularization process is closely related to tumor proliferation and metastasis (59). Thus, RPS6KA2 may be essential for multiple vital steps in tumor development. Moreover, we used CellChat to investigate cell interactions within the ccRCC microenvironment. Our findings revealed that epithelial cells, NK cells, and T cells all play active roles in the MIF signaling pathway, indicating their potential involvement in mediating interactions among these cell types.

This study focuses on exploring potential associations between T1DM-related genes and ccRCC risk, laying the groundwork for future mechanistic research. While it remains unclear whether genes such as KIF21A, PIGH, and RPS6KA2 directly drive disease processes or are secondary to other factors, we propose that their dysregulation may result from the proinflammatory and metabolic changes associated with T1DM. This dysregulation likely impacts shared



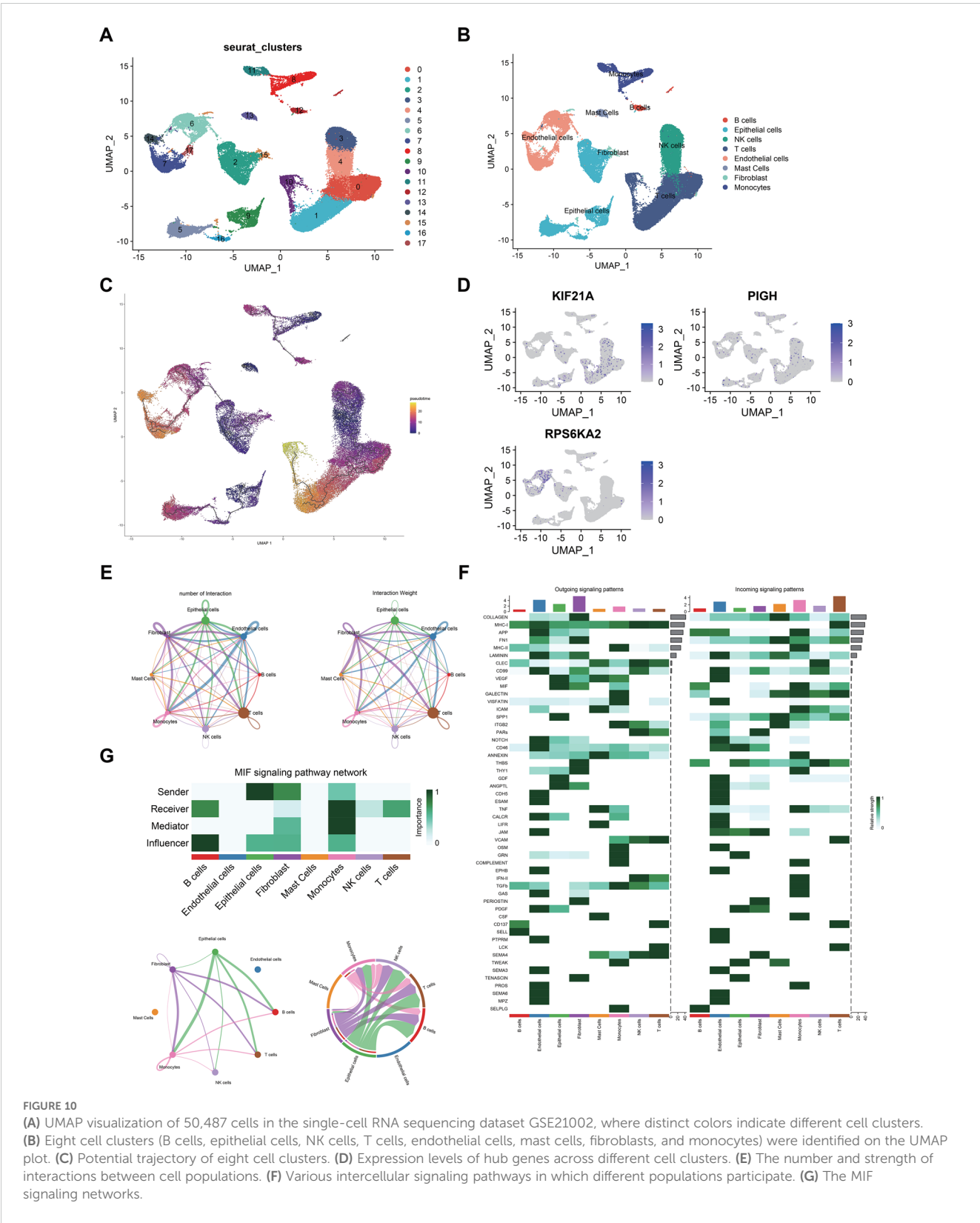
biological pathways, including estrogen response, oxidative phosphorylation, and signal transduction, thereby increasing ccRCC risk or accelerating its progression. Moreover, the proinflammatory state in T1DM may exacerbate gene dysregulation, further driving processes that promote ccRCC development and progression.

We validated the prognostic value of these markers in both the original and independent datasets, confirming their importance in ccRCC patient survival and management. Although no significant effect of race on hub gene expression was observed, the potential impact of race on disease mechanisms should not be overlooked. Thus, to mitigate this limitation, we validated our findings using RT-qPCR on clinical samples from Asian patients, demonstrating the applicability of our results across diverse populations. This partially addresses the underrepresentation in public datasets.

The strong association of these hubgenes with ccRCC prognosis and the stability of their expression across ethnic groups highlight their potential as biomarkers for early detection and personalized

patient management. We propose incorporating these biomarkers into liquid biopsy platforms, such as ctDNA or RNA-based assays, using advanced techniques like digital PCR and NGS to enable early, non-invasive detection of ccRCC, especially in high-risk populations like T1DM patients.

Due to our study's reliance on bioinformatics methods, issues with sample quality and batch effects within public databases may have introduced bias in the analysis results. Consequently, additional *in vivo* or *in vitro* experiments are essential to further validate our findings. In this study, distinct external validation sets were utilized for ccRCC and T1DM patients. However, it is imperative to substantiate the robustness of the hub genes by incorporating more external datasets in future investigations. Furthermore, our conclusions are drawn from diverse patient cohorts and lack validation within the same individual. Thus, future research should employ animal models that integrate ccRCC and T1DM to elucidate the potential relationship between these two diseases.



5 Conclusion

In this study, T1DM-related genes were identified using WGCNA, and three hub genes (KIF21A, PIGH, and RPS6KA2) were selected through feature selection in machine learning models that intersected

with ccRCC prognostic DEGs; we also explored their role and potential mechanism in the development of ccRCC. Single-cell RNA sequencing analysis has laid the groundwork for obtaining a comprehensive understanding of the functions and interactions of these biomarkers within the immune microenvironment. This work is

anticipated to offer possibilities for the prevention or early detection of ccRCC in T1DM patients and to provide novel targets for new pharmacotherapies.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repositories and accession numbers are provided in Table S1. Specifically, the T1DM peripheral blood mononuclear cell transcriptome dataset (GSE55098), the T1DM validation cohort (GSE9006), and the ccRCC transcriptome validation cohort (GSE53757) were obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). Additionally, the TCGA-KIRC cohort was obtained from The Cancer Genome Atlas (TCGA) (<https://portal.gdc.cancer.gov/>). All datasets used in this study are publicly available in these databases, they can be accessed at any time.

Ethics statement

This study was approved by the Ethics Committee of Jinan University. Both patients and controls provided written informed consent. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

YL: Conceptualization, Methodology, Writing – review & editing, Data curation, Formal analysis, Investigation, Writing – original draft. RZ: Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Software. YH: Data curation, Formal analysis, Software, Writing – review & editing, Methodology. YZ: Writing – review & editing, Project administration, Resources, Supervision, Validation. JH: Project administration, Resources, Supervision, Writing – review & editing, Conceptualization, Funding acquisition, Methodology.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

References

- Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res Clin Pract.* (2022) 183:109119. doi: 10.1016/j.diabres.2021.109119
- Forlenza GP, Rewers M. The epidemic of type 1 diabetes: what is it telling us? *Curr Opin Endocrinol Diabetes Obes.* (2011) 18:248–51. doi: 10.1097/MED.0b013e32834872ce
- Au KM, Medik Y, Ke Q, Tisch R, Wang AZ. Immune checkpoint-bioengineered beta cell vaccine reverses early-onset type 1 diabetes. *Adv Mater.* (2021) 33:e2101253. doi: 10.1002/adma.202101253
- Ohashi PS, Oehen S, Buerki K, Pircher H, Ohashi CT, Odermatt B, et al. Ablation of "tolerance" and induction of diabetes by virus infection in viral antigen transgenic mice. *Cell.* (1991) 65:305–17. doi: 10.1016/0092-8674(91)90164-T

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2025.1543806/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

(A) Gene selection using the random forest algorithm. (B) Comparison of the reverse cumulative residual distribution between the random forest and support vector machine methods. (C) Comparison of the residual values between the random forest and SVM-RFE methods; the red dots represent the root mean square of the residuals. (D) Selection of the optimal parameter (λ) for LASSO regression through cross-validation. (E) Gene selection using LASSO regression. (F) The hub genes were selected as the intersection of genes identified by the random forest and LASSO algorithms. * RF, SVM-RFE, and LASSO analyses were performed using the TCGA-KIRC cohort.

SUPPLEMENTARY FIGURE 2

(A) Comparison of the receiver operating characteristic (ROC) curves between the random forest and SVM methods in TCGA-KIRC cohort. (B) The ROC curve of LASSO method in TCGA-KIRC cohort.

SUPPLEMENTARY FIGURE 3

Bubble plot depicting marker gene expression levels in each cell type.

5. Bluestone JA, Buckner JH, Herold KC. Immunotherapy: Building a bridge to a cure for type 1 diabetes. *Science*. (2021) 373:510–6. doi: 10.1126/science.abb1654
6. Gardner G, Fraker CA. Natural killer cells as key mediators in type 1 diabetes immunopathology. *Front Immunol*. (2021) 12:722979. doi: 10.3389/fimmu.2021.722979
7. Rodacki M, Milech A, de Oliveira JE. NK cells and type 1 diabetes. *Clin Dev Immunol*. (2006) 13:101–7. doi: 10.1080/17402520600877182
8. Labochka D, Moszczuk B, Kukwa W, Szczylik C, Czarnecka AM. Mechanisms through which diabetes mellitus influences renal cell carcinoma development and treatment: A review of the literature. *Int J Mol Med*. (2016) 38:1887–94. doi: 10.3892/ijmm.2016.2776
9. Jonasch E, Gao J, Rathmell WK. Renal cell carcinoma. *Bmj*. (2014) 349:g4797. doi: 10.1136/bmj.g4797
10. Khan Y, Slattery TD, Pickering LM. Individualizing systemic therapies in first line treatment and beyond for advanced renal cell carcinoma. *Cancers (Basel)*. (2020) 12. doi: 10.3390/cancers12123750
11. Zhang Q, Chen P, Tian R, He J, Han Q, Fan L. Metabolic syndrome is an independent risk factor for fuhrman grade and TNM stage of renal clear cell carcinoma. *Int J Gen Med*. (2022) 15:143–50. doi: 10.2147/IJGM.S346972
12. Pruthi DK, Miller G, Ankerst DP, Neumair M, Capitanio U, Correa AF, et al. Diabetes, obesity, and pathological upstaging in renal cell carcinoma: results from a large multi-institutional consortium. *J Urol*. (2023) 210:750–62. doi: 10.1097/JU.0000000000003650
13. Saly DL, Eswarappa MS, Street SE, Deshpande P. Renal cell cancer and chronic kidney disease. *Adv Chronic Kidney Dis*. (2021) 28:460–468.e1. doi: 10.1053/j.ackd.2021.10.008
14. Zeng R, Li Y, He DM, Sun MZ, Huang WQ, Wang YH, et al. Potassium channel-related genes are a novel prognostic signature for the tumor microenvironment of renal clear cell carcinoma. *Front Oncol*. (2022) 12:1013324. doi: 10.3389/fonc.2022.1013324
15. Zhao Z, Ren J, Xie S, Zou L, Zhao Q, Zeng S, et al. Identification of biomarkers associated with CD8+ T cells in rheumatoid arthritis and their pan-cancer analysis. *Front Immunol*. (2022) 13:1044909. doi: 10.3389/fimmu.2022.1044909
16. Zhu E, Shu X, Xu Z, Peng Y, Xiang Y, Liu Y, et al. Screening of immune-related secretory proteins linking chronic kidney disease with calcific aortic valve disease based on comprehensive bioinformatics analysis and machine learning. *J Transl Med*. (2023) 21:359. doi: 10.1186/s12967-023-04171-x
17. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinf*. (2008) 9:559. doi: 10.1186/1471-2105-9-559
18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. (2015) 43:e47. doi: 10.1093/nar/gkv007
19. Wang N, Li Y, Zhou X, Wang X, Yang G. Comprehensive analysis identifies ARHGEF6 as a potential prognostic and immunological biomarker in lung adenocarcinoma. *Comput Biol Med*. (2023) 153:106448. doi: 10.1016/j.combiomed.2022.106448
20. Cao Y, Yang Y, Hu Q, Wei G. Identification of potential drug targets for rheumatoid arthritis from genetic insights: a Mendelian randomization study. *J Transl Med*. (2023) 21:616. doi: 10.1186/s12967-023-04474-z
21. Rajadnya R, Sharma N, Mahajan A, Ulhe A, Patil R, Hegde M, et al. Novel systems biology experimental pipeline reveals matairesinol's antimetastatic potential in prostate cancer: an integrated approach of network pharmacology, bioinformatics, and experimental validation. *Brief Bioinform*. (2024) 25. doi: 10.1093/bib/bbae466
22. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. (2010) 38:W214–20. doi: 10.1093/nar/gkq537
23. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. (2021) 49: D605–d612. doi: 10.1093/nar/gkaa1074
24. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. (2019) 10:1523. doi: 10.1038/s41467-019-09234-6
25. Ishwaran H, Kogalur UB. Consistency of random survival forests. *Stat Probab Lett*. (2010) 80:1056–64. doi: 10.1016/j.spl.2010.02.020
26. Scharl T, Grü B, Leisch F. Mixtures of regression models for time course gene expression data: evaluation of initialization and random effects. *Bioinformatics*. (2010) 26:370–7. doi: 10.1093/bioinformatics/btp686
27. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. *BioRxiv*. (2016) p:060012. doi: 10.1101/060012
28. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. (2003) 34:267–73. doi: 10.1038/ng1180
29. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U.S.A.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
30. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. (2015) 12:453–7. doi: 10.1038/nmeth.3337
31. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. (2024) 42:293–304. doi: 10.1038/s41587-023-01767-y
32. Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR, et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science*. (2018) 361:594–9. doi: 10.1126/science.aat1699
33. Franzén O, Gan LM, Björkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)*. (2019) 2019. doi: 10.1093/database/baz046
34. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. (2014) 32:381–6. doi: 10.1038/nbt.2859
35. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. (2021) 12:1088. doi: 10.1038/s41467-021-21246-9
36. Sun L, Chao F, Luo B, Ye D, Zhao J, Zhang Q, et al. Impact of estrogen on the relationship between obesity and renal cell carcinoma risk in women. *EBioMedicine*. (2018) 34:108–12. doi: 10.1016/j.ebiom.2018.07.010
37. Tian S, Wang R, Wang Y, Chen R, Lin T, Xiao X, et al. p32 regulates glycometabolism and TCA cycle to inhibit ccRCC progression via copper-induced DLAT lipoylation oligomerization. *Int J Biol Sci*. (2024) 20:516–36. doi: 10.7150/ijbs.84399
38. Li Y, Maimaiti M, Yang B, Lu Z, Zheng Q, Lin Y, et al. Comprehensive analysis of subtypes and risk model based on complement system associated genes in ccRCC. *Cell Signal*. (2023) 111:110888. doi: 10.1016/j.cellsig.2023.110888
39. Ljungberg B, Campbell SC, Choi HY, Jacqmin D, Lee JE, Weikert S, et al. The epidemiology of renal cell carcinoma. *Eur Urol*. (2011) 60:615–21. doi: 10.1016/j.eururo.2011.06.049
40. Ilonen J, Lempainen J, Veijola R. The heterogeneous pathogenesis of type 1 diabetes mellitus. *Nat Rev Endocrinol*. (2019) 15:635–50. doi: 10.1038/s41574-019-0254-y
41. Willcox A, Richardson SJ, Bone AJ, Foulis AK, Morgan NG. Analysis of islet inflammation in human type 1 diabetes. *Clin Exp Immunol*. (2009) 155:173–81. doi: 10.1111/j.1365-2249.2008.03860.x
42. Klein J, Sato A. The HLA system. First of two parts. *N Engl J Med*. (2000) 343:702–9. doi: 10.1056/NEJM200009073431006
43. Buzzetti R, Zampetti S, Maddaloni E. Adult-onset autoimmune diabetes: current knowledge and implications for management. *Nat Rev Endocrinol*. (2017) 13:674–86. doi: 10.1038/nrendo.2017.99
44. Sabbah E, Savola K, Ebeling T, Kulmala P, Vähäsalo P, Ilonen J, et al. Genetic, autoimmune, and clinical characteristics of childhood- and adult-onset type 1 diabetes. *Diabetes Care*. (2000) 23:1326–32. doi: 10.2337/diacare.23.9.1326
45. Pearson-Stuttard J, Zhou B, Kontis V, Bentham J, Gunter MJ, Ezzati M. Worldwide burden of cancer attributable to diabetes and high body-mass index: a comparative risk assessment. *Lancet Diabetes Endocrinol*. (2018) 6:e6–e15. doi: 10.1016/s2213-8587(18)30150-5
46. Iyengar NM, Hudis CA, Dannenberg AJ. Obesity and cancer: local and systemic mechanisms. *Annu Rev Med*. (2015) 66:297–309. doi: 10.1146/annurev-med-050913-022228
47. Pereira CS, Molz P, Palazzo RP, de Freitas TA, Maluf SW, Horta JA, et al. DNA damage and cytotoxicity in adult subjects with prediabetes. *Mutat Res*. (2013) 753:76–81. doi: 10.1016/j.mrgentox.2013.02.002
48. Scappaticcio L, Maiorino MI, Bellastella G, Giugliano D, Esposito K. Insights into the relationships between diabetes, prediabetes, and cancer. *Endocrine*. (2017) 56:231–9. doi: 10.1007/s12020-016-1216-y
49. Devaraj S, Dasu MR, Park SH, Jialal I. Increased levels of ligands of Toll-like receptors 2 and 4 in type 1 diabetes. *Diabetologia*. (2009) 52:1665–8. doi: 10.1007/s00125-009-1394-8
50. Pearson G, Robinson F, Beers Gibson T, Xu BE, Karandikar M, Berman K, et al. Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. *Endocr Rev*. (2001) 22:153–83. doi: 10.1210/edrv.22.2.0428
51. Wu H, He D, Biswas S, Shafiqzaman M, Zhou X, Charron J, et al. mTOR activation initiates renal cell carcinoma development by coordinating ERK and p38MAPK. *Cancer Res*. (2021) 81:3174–86. doi: 10.1158/0008-5472.CAN-20-3979
52. Weng Z, Shang Y, Yao D, Zhu J, Zhang R. Structural analyses of key features in the KANK1-KIF21A complex yield mechanistic insights into the cross-talk between microtubules and the cell cortex. *J Biol Chem*. (2018) 293:215–25. doi: 10.1074/jbc.M117.816017
53. Sun W, Iijima T, Kano J, Kobayashi H, Li D, Morishita Y, et al. Frequent aberrant methylation of the promoter region of sterile alpha motif domain 14 in pulmonary adenocarcinoma. *Cancer Sci*. (2008) 99:2177–84. doi: 10.1111/j.1349-7006.2008.00965.x

54. Tremblay-Laganière C, Kaiyrzhanov R, Maroofian R, Nguyen TTM, Salayev K, Chilton IT, et al. PIGH deficiency can be associated with severe neurodevelopmental and skeletal manifestations. *Clin Genet.* (2021) 99:313–7. doi: 10.1111/cge.13877
55. Loeff FC, Rijs K, van Egmond EHM, Zoutman WH, Qiao X, Kroes WGM, et al. Loss of the GPI-anchor in B-lymphoblastic leukemia by epigenetic downregulation of PIGH expression. *Am J Hematol.* (2019) 94:93–102. doi: 10.1002/ajh.25337
56. Milosevic N, Kühnemuth B, Mühlberg L, Ripka S, Griesmann H, Lölkes C, et al. Synthetic lethality screen identifies RPS6KA2 as modifier of epidermal growth factor receptor activity in pancreatic cancer. *Neoplasia.* (2013) 15:1354–62. doi: 10.1593/neo.131660
57. Nekoua MP, Dechaumes A, Sane F, Alidjinou EK, Moutairou K, Yessoufou A, et al. Enteroviral pathogenesis of type 1 diabetes: the role of natural killer cells. *Microorganisms.* (2020) 8. doi: 10.3390/microorganisms8070989
58. Wong WT, Wong SL, Tian XY, Huang Y. Endothelial dysfunction: the common consequence in diabetes and hypertension. *J Cardiovasc Pharmacol.* (2010) 55:300–7. doi: 10.1097/FJC.0b013e3181d7671c
59. Lin PP. Aneuploid circulating tumor-derived endothelial cell (CTEC): A novel versatile player in tumor neovascularization and cancer metastasis. *Cells.* (2020) 9. doi: 10.3390/cells9061539