

Integrated Transcriptomic Analysis Identifies a Six-Gene Prognostic Signature, RNA Editing Derived Neoantigens, and miRNA Regulatory Networks in Renal Cell Carcinoma

Zohreh Mehmandoostli^{1,2}, PhD;
Mahmood Dehghani Ashkezari^{1,2}, PhD;
Seyed Morteza Seifati^{1,2}, PhD;
Gholam Ali Kardar^{3,4}, PhD

¹Department of Biology, Ashk.C., Islamic Azad University, Ashkezar, Iran;

²Medical Biotechnology Research Center, Ashk.C., Islamic Azad University, Ashkezar, Iran;

³Immunology, Asthma and Allergy Research Institute (IAARI), Tehran University of Medical Sciences, Tehran, Iran;

⁴Department of Medical Biotechnology, School of Advanced Technologies in Medicine, Tehran University of Medical Sciences, Tehran, Iran

Correspondence:

Gholam Ali Kardar, PhD;
Immunology, Asthma and Allergy Research Institute, Children Medical Center, No. 62, Dr Gharib St, Keshavarz Blvd, Postal code: 14197-33154, Tehran, Iran

Tel: +98 21 61472147

Fax: +98 21 66428995

Email: gakardar@gmail.com

Received: 21 September 2025

Revised: 21 November 2025

Accepted: 20 December 2025

What's Known

- Renal cell carcinoma is a heterogeneous malignancy with variable outcomes and limited biomarkers for prognosis or immunotherapy response.
- Frequent mutations are found in *VHL*, *SETD2*, *PBRM1*, *BAP1*, *FLCN*, and *TP53*; however, their integrated effects on tumor biology, RNA editing, and non-coding RNA regulation remain unclear.
- The synergistic effects remain unclear, and there is notable batch-related variability across different cohorts.

What's New

- A six-gene prognostic signature (*BAP1*, *SETD2*, *TP53*, *PBRM1*, *FLCN*, *VHL*) stratifies patients into high- and low-risk groups (AUC>0.70 at 1, 3, 5 years).
- This study integrates transcriptomic profiling, RNA editing analysis, and miRNA-mRNA network construction, leading to the identification of recurrent RNA editing events. Both broadly shared and population-specific predicted neoantigens were identified, revealing links between chromatin remodeling, metabolic pathways, and immune regulation.

Abstract

Background: Renal cell carcinoma (RCC) is a heterogeneous malignancy with variable outcomes and limited biomarkers for prognosis or immunotherapy response. Although mutations in *VHL*, *SETD2*, *PBRM1*, *BAP1*, *FLCN*, and *TP53* are frequent, their integrated effects on tumor biology, RNA editing, and non-coding RNA regulation remain unclear. This study aims to integrate genetic, epigenetic, and immune features to provide mechanistic insights into RCC progression and support precision immuno-oncology and vaccine development.

Methods: Transcriptomic data from TCGA-KIRC and four Gene Expression Omnibus (GEO) cohorts were normalized, batch-corrected, and integrated. Differentially expressed genes were analyzed using LASSO-Cox regression, Kaplan–Meier survival, ROC curves, and nomogram modeling. RNA editing events from REDIportal were annotated for nonsynonymous substitutions and assessed for human leukocyte antigen (HLA) class I binding. Parallel miRNA-mRNA analyses identified regulatory interactions. Functional enrichment and immune LASSO deconvolution were used to explore pathways and tumor microenvironment features.

Results: *VHL*, *SETD2*, and *BAP1* were downregulated, while *FLCN* showed heterogeneous upregulation. A six-gene prognostic signature (*BAP1*, *SETD2*, *TP53*, *PBRM1*, *FLCN*, *VHL*) stratified patients into high- and low-risk groups with AUCs>0.70 at 1, 3, and 5 years. RNA editing revealed 35 recurrent events, including 25 nonsynonymous substitutions in *TP53*, *BAP1*, and *SETD2*. Predicted neoantigens included both broadly presented and population-specific epitopes. Deregulated miRNAs highlighted post-transcriptional regulation influencing progression and immune evasion. Functional enrichment analysis implicated chromatin remodeling, metabolism, and immune regulation, while immune profiling linked *BAP1* mutations with reduced NK/Treg infiltration and identified chromatin genes associated with endothelial and immune activity.

Conclusion: This integrative study identifies a six-gene signature, recurrent RNA editing-derived neoantigens, and miRNA networks in RCC. By connecting genomic, epigenetic, and immune features, it provides mechanistic insights into RCC progression and supports precision immuno-oncology and vaccine development.

Please cite this article as: Mehmandoostli Z, Dehghani Ashkezari M, Seifati SM, Kardar GA. Integrated Transcriptomic Analysis Identifies a Six-Gene Prognostic Signature, RNA Editing Derived Neoantigens, and miRNA Regulatory Networks in Renal Cell Carcinoma. Iran J Med Sci. 2026;51(5):316-339. doi: 10.30476/ijms.2026.108742.4376.

Keywords • Carcinoma, renal cell • RNA editing • Biomarkers
• Prognosis • Antigens

Introduction

Renal cell carcinoma (RCC) represents 2–3% of adult malignancies, with clear cell RCC (ccRCC) accounting for approximately 75% of cases.¹ Despite advances in targeted therapies and immunotherapy, RCC remains clinically challenging due to molecular heterogeneity, late diagnosis, and limited predictive biomarkers.² Tumor suppressor gene mutations and alterations in oncogenes play key roles in RCC initiation, progression, and therapeutic resistance. However, the complex interplay between gene expression, mutations, and RNA-level modifications remains incompletely understood.³

Several recurrently altered genes in RCC, including von Hippel–Lindau (*VHL*), polybromo 1 (*PBRM1*), SET domain containing 2 (*SETD2*), BRCA1-associated protein 1 (*BAP1*), folliculin (*FLCN*), tumor protein p53 (*TP53*), and phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (*PIK3CA*), contribute significantly to disease pathogenesis.^{3–5} Inactivation of *VHL* is regarded as an initiating event in RCC, leading to constitutive activation of hypoxia-inducible factors (HIFs) and downstream angiogenic pathways.⁶ Mutations in *PBRM1*, *SETD2*, and *BAP1*, located on chromosome 3p, further shape RCC progression, often in a mutually exclusive manner, reflecting distinct molecular subtypes and clinical outcomes.^{7, 8} Alterations in *TP53* and *PIK3CA*, although less frequent, have been implicated in tumor progression and therapeutic resistance.⁹ Meanwhile, *FLCN*, a gene associated with Birt-Hogg-Dubé (BHD) syndrome, is emerging as an additional regulator of metabolic signaling in renal tumors.¹⁰ While these genetic alterations are well-documented, the role of RNA editing, particularly in modifying the expression and function of these key RCC genes, has not been comprehensively studied. RNA editing, especially adenosine-to-inosine (A-to-I) editing, introduces transcriptomic diversity beyond fixed DNA mutations, generating nonsynonymous changes, influencing splicing, and potentially altering immune recognition.¹¹ Editing in cancer-relevant genes such as *TP53*, *BAP1*, and *SETD2* may impact protein function or produce immunogenic neoantigens.¹² Unlike fixed mutations, RNA editing is dynamic, tissue-specific, and population-dependent, potentially influencing tumor evolution and immune surveillance.¹³ Advances in sequencing and computational prediction now enable systematic identification of RNA editing–derived neoantigens and their human leukocyte antigen

(HLA)-binding potential,¹⁴ opening avenues for personalized immunotherapy.^{15, 16} In this study, we performed an integrative analysis of transcriptomic datasets, focusing on seven RCC-associated genes (*VHL*, *SETD2*, *PBRM1*, *BAP1*, *FLCN*, *TP53*, and *PIK3CA*). By combining differential expression, mutational profiling, and RNA editing mapping, we identified canonical genetic drivers alongside RNA editing–derived alterations with potential immunogenic relevance, evaluating both shared and population-specific neoantigens. This study aims to provide a comprehensive understanding of how integrated genetic alterations, RNA editing, and miRNA regulation contribute to RCC progression and immune evasion, while identifying potential targets for personalized immunotherapy.

Materials and Methods

Data Acquisition

Transcriptomic expression data for seven key genes implicated in RCC, including *VHL*, *SETD2*, *PBRM1*, *BAP1*, *FLCN*, *TP53*, and *PIK3CA*—both tumor and matched normal kidney tissues—were obtained from The Cancer Genome Atlas (TCGA-KIRC) (NCBI, USA) “<https://portal.gdc.cancer.gov/projects/TCGA-KIRC>”, which includes 541 primary tumor samples and 72 solid tissue normal samples collected across multiple institutions in the USA. Additionally, four Gene Expression Omnibus (GEO) datasets (NIH, USA) were used: GSE11151 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11151>) (five normal, 62 tumor), submitted in 2008 by Erasmus Medical Centre (Netherlands), last updated on July 31, 2025, utilizing the Affymetrix Human Genome U133 Plus 2.0 Array (GPL570) platform; GSE46699 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46699>) (63 normal, 67 tumor), submitted in 2014 by Mayo Clinic (USA), last updated on March 25, 2019, utilizing the Affymetrix Human Genome U133 Plus 2.0 Array (GPL570) platform; GSE53757 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53757>) (72 normal, 72 tumor), submitted in 2014 by Mayo Clinic (USA), last updated on March 25, 2019, utilizing the Affymetrix Human Genome U133 Plus 2.0 Array (GPL570) platform; and GSE6344 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6344>) (10 normal, 10 tumor), submitted in 2006 by Mayo Clinic (USA), last updated on August 10, 2018, utilizing the Affymetrix Human Genome U133 Array (GPL96/GPL97) platform.

The comprehensive methodology, outlining the key analytical steps and their interrelation, is visualized in figure 1.

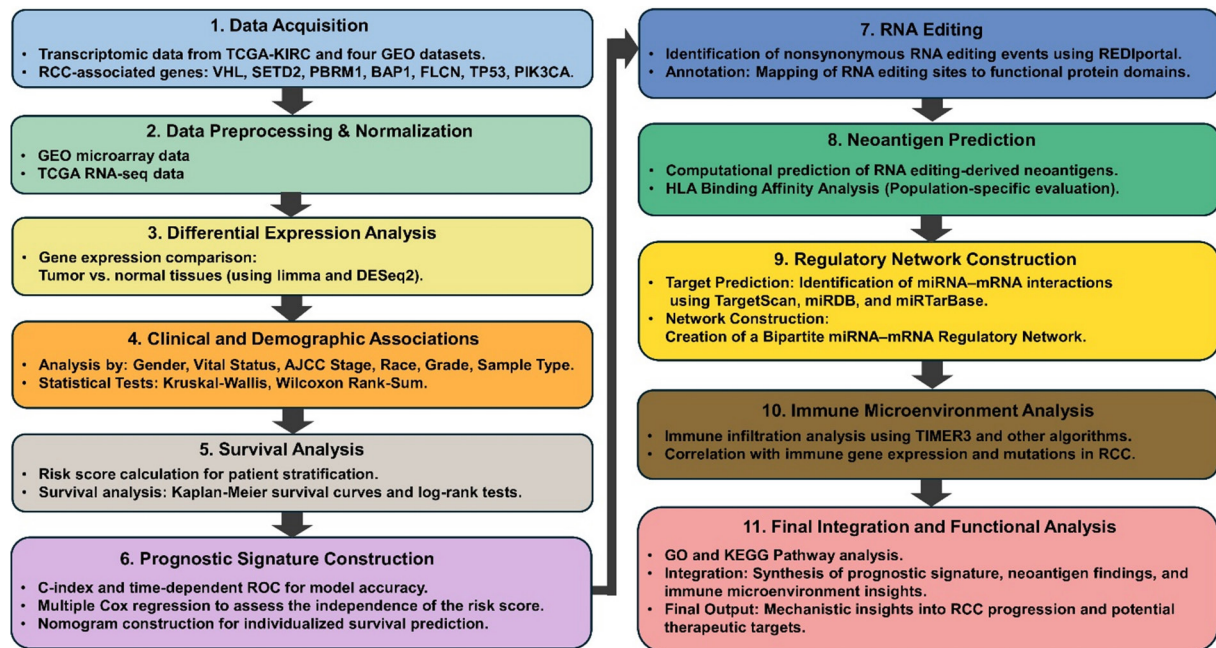


Figure 1: This flowchart illustrates the integrated transcriptomic analysis workflow for investigating Renal Cell Carcinoma (RCC) through eleven sequential methodology steps. The process encompasses data acquisition, preprocessing, differential expression analysis, clinical associations, survival analysis, prognostic signature construction, RNA editing, neoantigen prediction, regulatory network construction, immune microenvironment analysis, and final functional integration. RCC: Renal cell carcinoma; TCGA: The cancer genome atlas; KIRC: Kidney renal clear cell carcinoma; GEO: Gene expression omnibus; VHL: Von Hippel-Lindau; SETD2: SET domain containing 2; PBRM1: Polybromo 1; BAP1: BRCA1 associated protein 1; FLCN: Folliculin; TP53: Tumor protein P53; PIK3CA: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; AJCC: American joint committee on cancer; ROC: Receiver operating characteristic; HLA: Human leukocyte antigen; miRNA: MicroRNA; mRNA: Messenger RNA; TIMER3: Tumor immune estimation resource 3; GO: Gene ontology; KEGG: Kyoto encyclopedia of genes and genomes

Data Processing and Normalization

GEO microarray data were processed and normalized using R/Bioconductor (R Foundation for Statistical Computing, Austria). TCGA RNA-seq fragments per kilobase million (FPKM) data were downloaded via TCGAbiolinks (<https://gdc.cancer.gov/content/tcga-biolinks>). Clinical metadata for 613 TCGA patients, including sex, race, stage, grade, and survival, were retrieved.

Differential Expression Analysis

Tumor vs. normal expression was analyzed using limma (R package, USA) (GEO) and DESeq2 (TCGA) (Bioconductor, USA). An adjusted P value <0.05 was considered significant.

Clinical and Demographic Association

Gene expression was stratified by sex, vital status, AJCC stage, grade, and race (White, Black, American, Asian). Kruskal-Wallis and Wilcoxon tests were used for comparisons. Kaplan-Meier analysis and log-rank tests assessed prognostic significance.

Kaplan–Meier Survival Analysis

Patients were divided into high and low expression groups per gene. Survival curves were plotted using Survminer and ggplot2 (R package,

USA), with log-rank tests reporting significance.

Cross-Platform Validation

Expression patterns were validated across multiple GEO datasets, demonstrating consistent deregulation across independent cohorts.

Gene Expression Visualization and Clustering

Clustered heatmaps of seven RCC-associated genes were generated across integrated TCGA and GEO datasets. Data were log₂-transformed, quantile-normalized, and batch-corrected using ComBat (R packages/Tools, USA). Heatmaps were visualized via pheatmap with hierarchical clustering and row scaling.

Univariate and Multivariable Cox Proportional Hazards Regression

We initially performed univariate Cox proportional hazards regression for each of the seven RCC-associated genes (BAP1, FLCN, PBRM1, PIK3CA, SETD2, TP53, VHL) using the survival package in R. Genes significantly associated with overall survival (P<0.05) were subsequently included in a multivariable Cox regression model to evaluate their independent prognostic value. Hazard ratios (HRs), 95% confidence intervals (CIs), and P values were reported.

Construction of a Multi-Gene Risk Score Model

A prognostic risk score was calculated based on the linear combination of gene expression values weighted by their multivariable cox regression coefficients. Patients were dichotomized into high-risk and low-risk groups according to the median risk score. Kaplan–Meier survival curves and log-rank tests were used to assess survival differences between groups.

LASSO-Cox Regression for Optimal Gene Selection

To reduce dimensionality and avoid overfitting, we applied the least absolute shrinkage and selection operator (LASSO)-Cox regression using the glmnet (R packages, USA). Cross-validation was performed to identify the optimal penalty parameter (λ). Genes with non-zero coefficients were retained to construct an optimized prognostic signature. Risk scores based on this LASSO-selected gene panel were computed as described above.

Time-Dependent ROC and Concordance Index (C-index)

The predictive accuracy of the prognostic models was assessed using time-dependent receiver operating characteristic (ROC) analysis with the timeROC (R packages, USA). The concordance index (C-index) was also computed to evaluate model discrimination power.

Multivariable Cox Model with Clinical Covariates

To test the independence of the prognostic risk score, a multivariable Cox model was constructed including the LASSO-derived risk score and major clinical covariates (age, sex). Hazard ratios and statistical significance were reported.

Validation and Nomogram Construction

Patients were randomly divided into training and test cohorts to assess model robustness. A prognostic nomogram combining the risk score and clinical factors was developed to predict individualized survival probabilities.

Identification and Filtering of RNA Editing Sites in Cancer-Related Genes

RNA editing sites were extracted from REDportal v2.0 (USA) “<http://rediportal.cloud.ba.infn.it/atlas/>” and processed in Python (pandas) (Python Software Foundation, USA) / (Libraries, USA). Analysis focused on cancer-relevant genes (VHL, SETD2, PBRM1, BAP1, FLCN, TP53, PIK3CA). Sites were filtered for target genes using regex and selected for nonsynonymous

SNVs causing amino acid changes.

Functional Mapping of RNA Editing Sites

Nonsynonymous RNA editing events were mapped to protein domains of TP53, SETD2, and BAP1 using UniProt “<https://www.uniprot.org/>” coordinates. A Python pipeline determined whether mutations occurred within functional domains.

Clinical Annotation of RNA Editing-Derived Mutations

Identified amino acid changes were queried in ClinVar “<https://www.ncbi.nlm.nih.gov/clinvar/>” to assess clinical significance and disease associations. UniProt domain annotations contextualized mutations, and COSMIC “<https://cancer.sanger.ac.uk/cosmic>” was checked for reported somatic mutations in RCC. Only exact amino acid matches were considered.

Computational Prediction of RNA Editing-Derived Neoantigens Across Populations

RNA editing–derived amino acid substitutions from RCC patients were analyzed across five populations (Iranian, European, American, Asian, and African). Representative editing events were integrated with population-specific HLA class I alleles obtained from frequency databases and prior immunogenetic studies. Reference protein sequences for candidate genes (e.g., TP53, SETD2, BAP1) were retrieved from UniProt/Ensembl <https://www.uniprot.org/database/DB-0023> “<https://www.ensembl.org/>”, and only isoforms containing edited residues were used.

Each nonsynonymous edit was standardized (e.g., p.R337G), mapped to the protein sequence, and used to generate overlapping 8–11mer peptides containing the edited residue. Both wild-type (WT) and edited (ED) counterparts were produced. Peptide–HLA binding affinities were predicted using MHCflurry v2.1.5 <https://github.com/openvax/mhcflurry>, and only alleles supported by the tool were included. Candidate neoantigens were defined by:

- ED peptide IC₅₀ ≤ 150 nM (strong binder)
- WT peptide IC₅₀ ≥ 500 nM (weak binder)
- Significant Δ IC₅₀ (WT–ED) improvement.

For each population, we quantified the number of candidate neoantigens, binding affinity ranges, peptide length distribution, and HLA allele contributions. Cross-population comparisons distinguished public epitopes (shared across ≥2 populations) from private epitopes (population-specific). Results were visualized using log-scaled scatterplots of ED vs WT IC₅₀ values.

Mutation and Gene Expression Analysis

Seven key RCC genes (*BAP1*, *FLCN*, *PBRM1*, *PIK3CA*, *SETD2*, *TP53*, *VHL*) were analyzed for expression and mutation profiles. Gene expression data were normalized for downstream analyses, and somatic mutation data were processed into Mutation Annotation Format (MAF) using maftools “<https://github.com/PoisonAlien/maftools>”. Mutation frequencies and co-occurrence patterns were visualized with Oncoprints.

Prediction of miRNA–mRNA Interactions and Network Construction

Putative miRNA–mRNA interactions for the seven genes were retrieved from TargetScan “<https://www.targetscan.org/>”, miRDB “<https://mirdb.org/>”, and miRTarBase https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2025 (Databases, USA). Interactions with context++ scores <-0.2 (TargetScan) were retained, and miRNAs targeting ≥3 genes were defined as hub miRNAs. A bipartite miRNA–mRNA network was constructed and visualized using NetworkX, Matplotlib (Python Software Foundation, USA)/(Libraries, USA), and Cytoscape (Cytoscape Consortium, USA), with node size reflecting connectivity.

Functional and Pathway Enrichment Analysis

Gene Ontology (GO) enrichment of co-expressed genes was performed using clusterProfiler with the org.Hs.eg.db database (R packages, USA), focusing on Biological Process (BP). Significance thresholds were $P \leq 0.05$ and FDR-adjusted $P \leq 0.05$, corrected by the Benjamini-Hochberg method.

Immune Gene Expression and Immune Cell Infiltration Analysis

Associations between gene expression and immune infiltration in TCGA-KIRC were assessed using TIMER3 (USA) with multiple algorithms (ABIS, CIBERSORT, XCell, MCPOUNTER, QUANTISEQ, EPIC, CONSENSUS_TME). Correlations were calculated with Spearman's ρ ,

and results were visualized with heatmaps and bubble plots (*ggplot2* in R).

Immune Mutation Analysis

The TIMER3 Immune Mutation dataset was used to evaluate infiltration differences between mutant and wild-type tumors. log2 fold changes (log2FC) were visualized with violin and forest plots, showing distribution and effect sizes for each gene–immune cell pair. Analyses were performed in R, focusing on biologically relevant immune populations.

Results

Differential Gene Expression Across Datasets

Analysis across five independent datasets consistently revealed significant downregulation of *BAP1*, *SETD2*, and *VHL* in renal tumor samples compared to adjacent normal tissues, underscoring their tumor suppressive roles in RCC (table 1, figure 2). While *BAP1* and *VHL* were not significantly altered in GSE11151, their recurrent downregulation in the remaining datasets reinforces their relevance in tumorigenesis. *SETD2* showed highly consistent and robust downregulation across all datasets, further supporting its critical role in RCC biology.

By contrast, expression patterns of *FLCN*, *PBRM1*, and *PIK3CA* were more variable. *FLCN* showed upregulation in several datasets, suggesting a potential context-dependent role, whereas *PBRM1* was significantly downregulated only in TCGA. *PIK3CA* exhibited significant upregulation in multiple GEO datasets but not in TCGA, highlighting dataset-specific differences. Finally, *TP53* was consistently upregulated in most datasets, linking it to RCC progression.

Clinical Associations of Gene Expression in RCC

Clinical variables assessed included sex, vital status, AJCC pathologic stage, race, tumor grade, sample type (normal vs. primary tumor), and histological diagnosis.

Table 1: Differential expression status of key genes across renal cell carcinoma datasets

Gene	TCGA (n=613)	GSE6344 (n=40)	GSE11151 (n=67)	GSE46699 (n=130)	GSE53757 (n=144)	Overall trend
<i>BAP1</i>	↓ (FDR<0.001)	↓ (FDR=0.06)	↓ (P=0.118)	↓ (FDR<0.001)	↓ (FDR<0.001)	Downregulated
<i>FLCN</i>	↑ (FDR<0.001)	↑ (P=0.53)	↑ (P=0.766)	↑ (FDR=0.02)	↑ (FDR<0.001)	Upregulated/ mixed
<i>PBRM1</i>	↓ (FDR=0.02)	↓ (P=0.12)	↓ (P=0.247)	↓ (P=0.66)	↓ (P=0.13)	No significant change
<i>PIK3CA</i>	↓ (P=0.13)	↑ (FDR=0.06)	↓ (P=0.990)	↑ (FDR< 0.001)	↑ (FDR=0.04)	Upregulated/ mixed
<i>SETD2</i>	↓ (FDR<0.001)	↓ (FDR=0.002)	↓ (FDR=0.343)	↓ (FDR=1.73e-7)	↓ (FDR< 0.001)	Strongly downregulated
<i>TP53</i>	↑ (FDR<0.001)	↑ (FDR=0.06)	↑ (P=0.747)	↑ (FDR<0.001)	↑ (FDR<0.001)	Upregulated
<i>VHL</i>	↓ (FDR<0.001)	↓ (FDR=0.05)	↑ (P=0.952)	↓ (FDR<0.01)	↓ (FDR<0.001)	Strongly downregulated

Upregulation (↑) and downregulation (↓) in tumor tissue compared with normal kidney tissue. P values represent either raw values or False Discovery Rate (FDR) adjusted values as indicated in the primary analysis. Statistical significance was set at $P < 0.05$. RCC: Renal cell carcinoma; TCGA: The cancer genome atlas; KIRC: Kidney renal clear cell carcinoma; GEO: Gene expression omnibus; FDR: False discovery rate; P: probability value

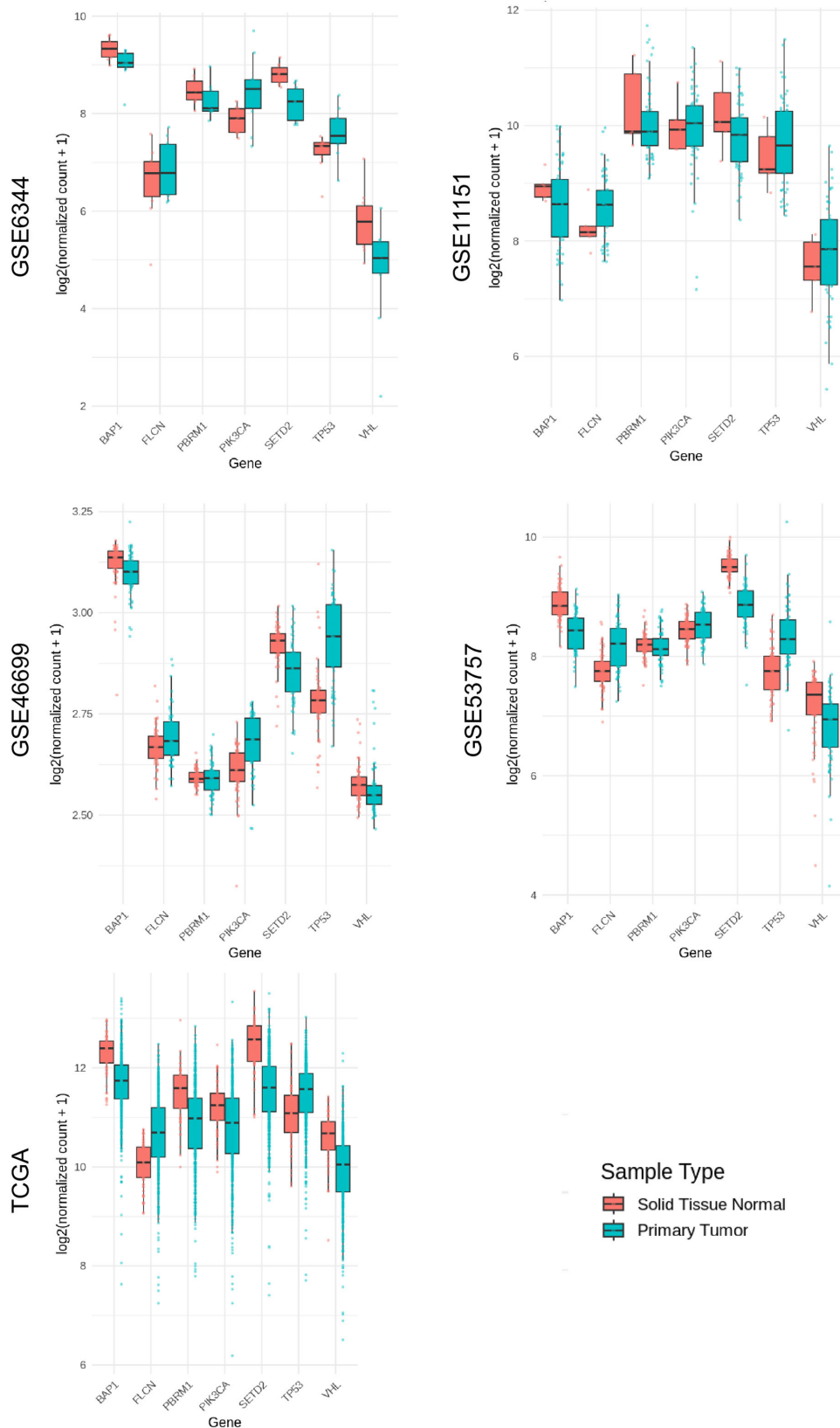


Figure 2: This figure presents a comprehensive overview of the differential gene expression patterns for seven key genes across multiple Renal Cell Carcinoma (RCC) datasets. The box plots demonstrate the expression profiles of *BAP1*, *FLCN*, *PBRM1*, *PIK3CA*, *SETD2*, *TP53*, and *VHL* in TCGA and four GEO datasets (GSE6344, GSE11151, GSE46699, and GSE53757), comparing solid tissue normal samples with primary tumor samples. RCC: Renal cell carcinoma; TCGA: The cancer genome atlas; GSE: Gene expression omnibus series; $\log_2$: Base-2 logarithm; *BAP1*: BRCA1 associated protein 1; *FLCN*: Folliculin; *PBRM1*: Polybromo 1; *PIK3CA*: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; *SETD2*: SET domain containing 2; *TP53*: Tumor protein P53; *VHL*: Von Hippel-Lindau.

Associations were evaluated using the Wilcoxon rank-sum test (for two groups) and the Kruskal-Wallis test (for multiple groups), with a significance threshold of $P < 0.05$. The cohort included 613 samples in total (sex: 208 female, 406 male; vital status: 412 alive, 202 dead; AJCC stage: 510 in various stages; race: 8 Asian, 59 Black/African American, 10 Not Reported, 537 White; tumor grade: 499 in various grades; sample type: 1 additional primary, 541 primary tumor, 72 solid tissue normal).

PIK3CA expression was significantly associated with vital status ($P < 0.001$), race ($P < 0.001$), tumor grade ($P = 0.02$), and sample type ($P < 0.001$), but not with sex ($P = 0.36$) or AJCC stage ($P = 0.34$). *SETD2* expression was linked to vital status ($P = 0.01$) and sample type ($P = 2.17 \times 10^{-22}$), but not with sex ($P = 0.82$), AJCC stage ($P = 0.17$), race ($P = 0.08$), or tumor grade ($P = 0.31$). *TP53* expression showed significant associations with AJCC stage ($P = 0.02$), race ($P = 0.01$), tumor grade ($P = 0.01$), and sample type ($P = 1.31 \times 10^{-6}$), but not with sex ($P = 0.71$) or vital status ($P = 0.05$). *BAP1* expression was significantly associated with race ($P = 0.03$) and sample type ($P = 3.58 \times 10^{-17}$), but not with sex ($P = 0.61$), vital status ($P = 0.14$), AJCC stage ($P = 0.09$), or tumor grade ($P = 0.47$). *FLCN* expression was significantly associated with sex ($P < 0.001$), race ($P < 0.001$), and sample type ($P = 1.94 \times 10^{-12}$), but not with vital status ($P = 0.78$), AJCC stage ($P = 0.38$), or tumor grade ($P = 0.34$). *PBRM1* expression was associated with vital status ($P < 0.001$), race ($P = 0.02$), and sample type ($P = 1.02 \times 10^{-10}$), but not with sex ($P = 0.99$), AJCC stage ($P = 0.29$), or tumor grade ($P = 0.18$). *VHL* expression showed significant associations with race ($P = 0.02$) and sample type ($P = 5.69 \times 10^{-14}$), but not with sex ($P = 0.47$), vital status ($P = 0.11$), AJCC stage ($P = 0.30$), or tumor grade ($P = 0.68$) (figure 3A–G).

Survival Analysis

Kaplan–Meier analysis showed no significant association between expression of the seven genes and overall survival (*BAP1* [$P = 0.42$], *FLCN* [$P = 0.96$], *PIK3CA* [$P = 0.69$], *SETD2*

[$P = 0.15$], *VHL* [$P = 0.39$], *PBRM1* [$P = 0.52$], *TP53* [$P = 0.33$]), indicating that their expression alone is not prognostic in this cohort. These findings suggest that the impact of these genes on survival may depend on additional molecular or clinical factors (figure 4).

Univariate and Multivariable Cox Regression Analysis

Univariate Cox regression suggested that *BAP1*, *FLCN*, *TP53*, and *PBRM1* were associated with overall survival. In multivariable analysis, *FLCN* and *TP53* remained independent prognostic factors, highlighting their potential roles as risk determinants in RCC (table 2 and figure 5A).

Prognostic Modeling and LASSO Signature

A seven-gene Cox model initially stratified patients into high- and low-risk groups, showing significantly different survival outcomes. To refine this model, LASSO-Cox regression was applied, resulting in a six-gene signature (*BAP1*, *SETD2*, *TP53*, *PBRM1*, *FLCN*, *VHL*), which demonstrated improved prognostic accuracy with a C-index of 0.673. This model consistently separated patients based on survival risk ($P < 0.001$; figure 5B–D).

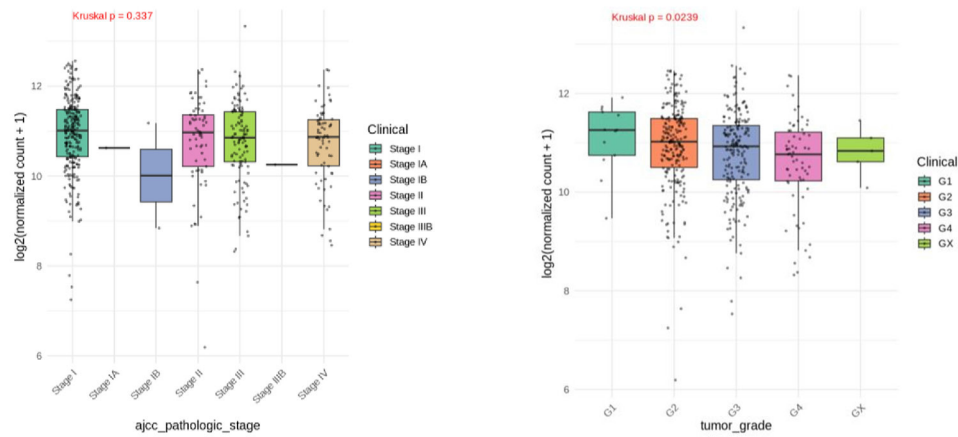
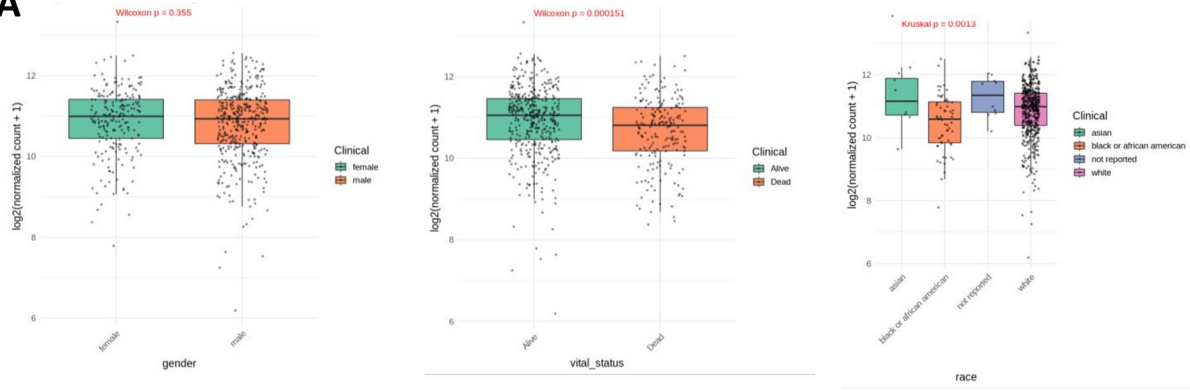
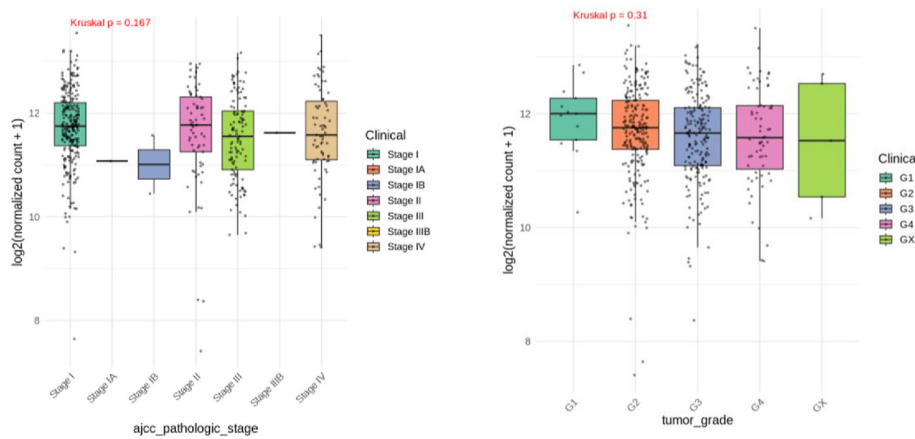
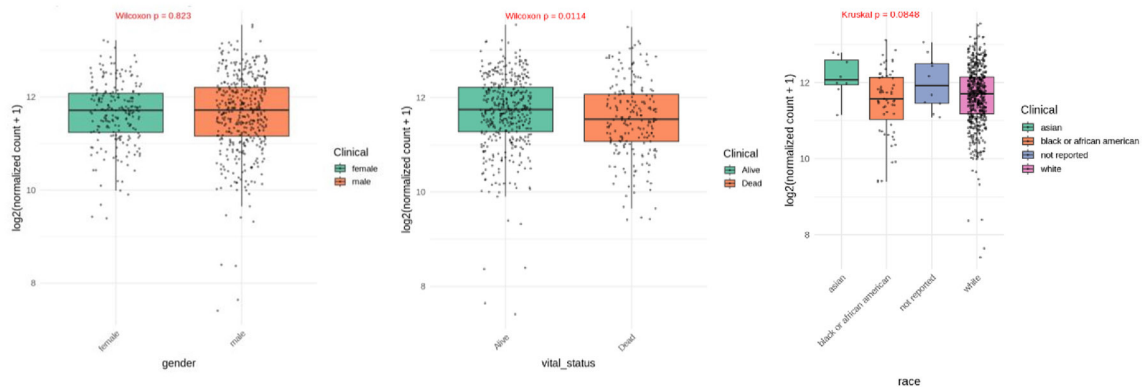
The LASSO-derived model showed enhanced prognostic accuracy compared to the original seven-gene signature (C-index=0.673), significantly improving risk stratification ($P < 0.001$, figure 5E). Time-dependent ROC curve analysis demonstrated stable predictive performance across different time points, with AUC values consistently above 0.70 at 1-, 3-, and 5-year follow-ups (AUC: 0.74, 0.76, 0.77; 95% CI: [0.70–0.78], [0.72–0.80], [0.73–0.79]; figure 5F). These results indicate that the LASSO-based signature is a robust and reliable tool for risk stratification of RCC patients.

To further evaluate the independence of the LASSO-derived risk score, a multivariable Cox regression model incorporating clinical variables such as age, sex, tumor grade, and stage was performed.

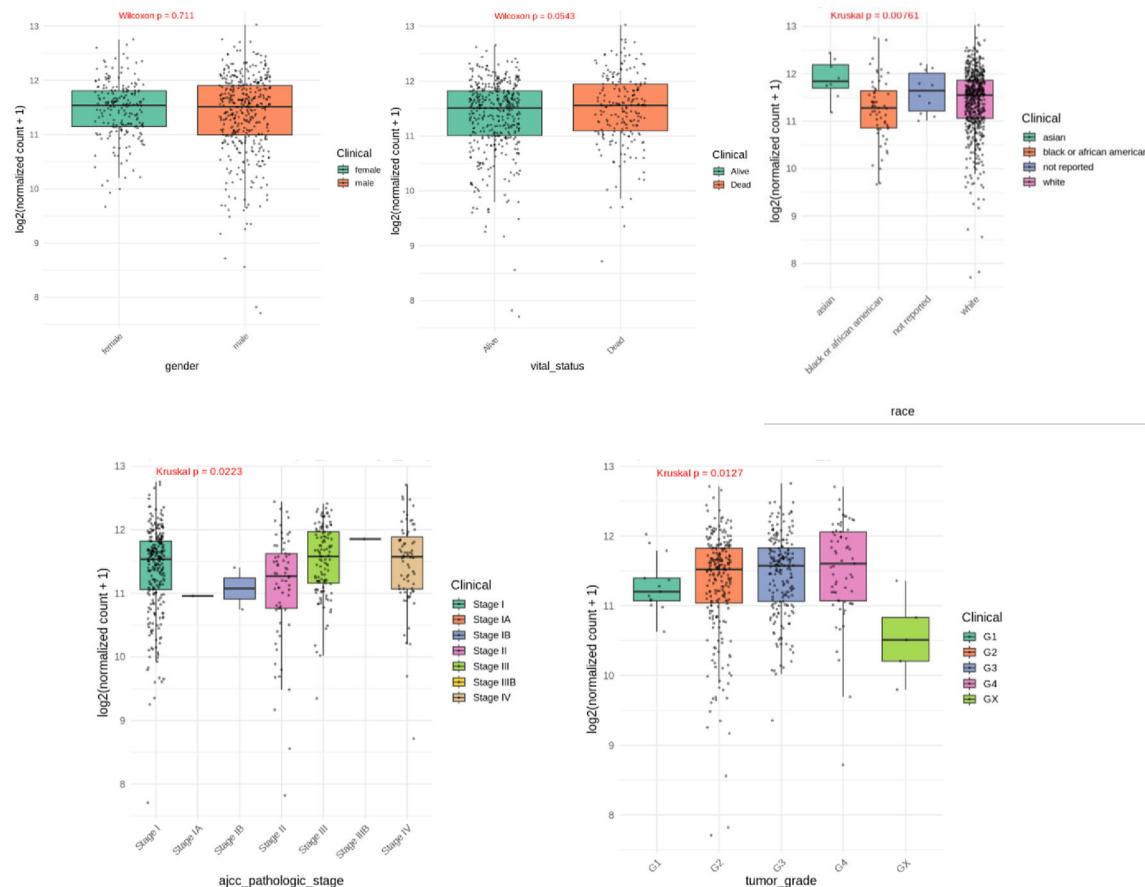
Table 2: Multivariable Cox regression analysis of RCC-associated genes

Gene	Coef	HR (exp[coef])	SE (coef)	z value	P value
<i>BAP1</i>	-2.40×10^{-4}	0.9998	1.35×10^{-4}	-1.774	0.08
<i>FLCN</i>	1.03×10^{-3}	1.0010	2.21×10^{-4}	4.684	< 0.001
<i>TP53</i>	1.99×10^{-4}	1.0002	6.55×10^{-5}	3.039	0.0024
<i>PIK3CA</i>	-5.65×10^{-5}	0.9999	9.54×10^{-5}	-0.593	0.55
<i>SETD2</i>	-1.06×10^{-4}	0.9999	8.27×10^{-5}	-1.284	0.20
<i>PBRM1</i>	-1.03×10^{-4}	0.9999	1.58×10^{-4}	-0.653	0.51
<i>VHL</i>	-1.13×10^{-4}	0.9999	1.03×10^{-4}	-1.098	0.27

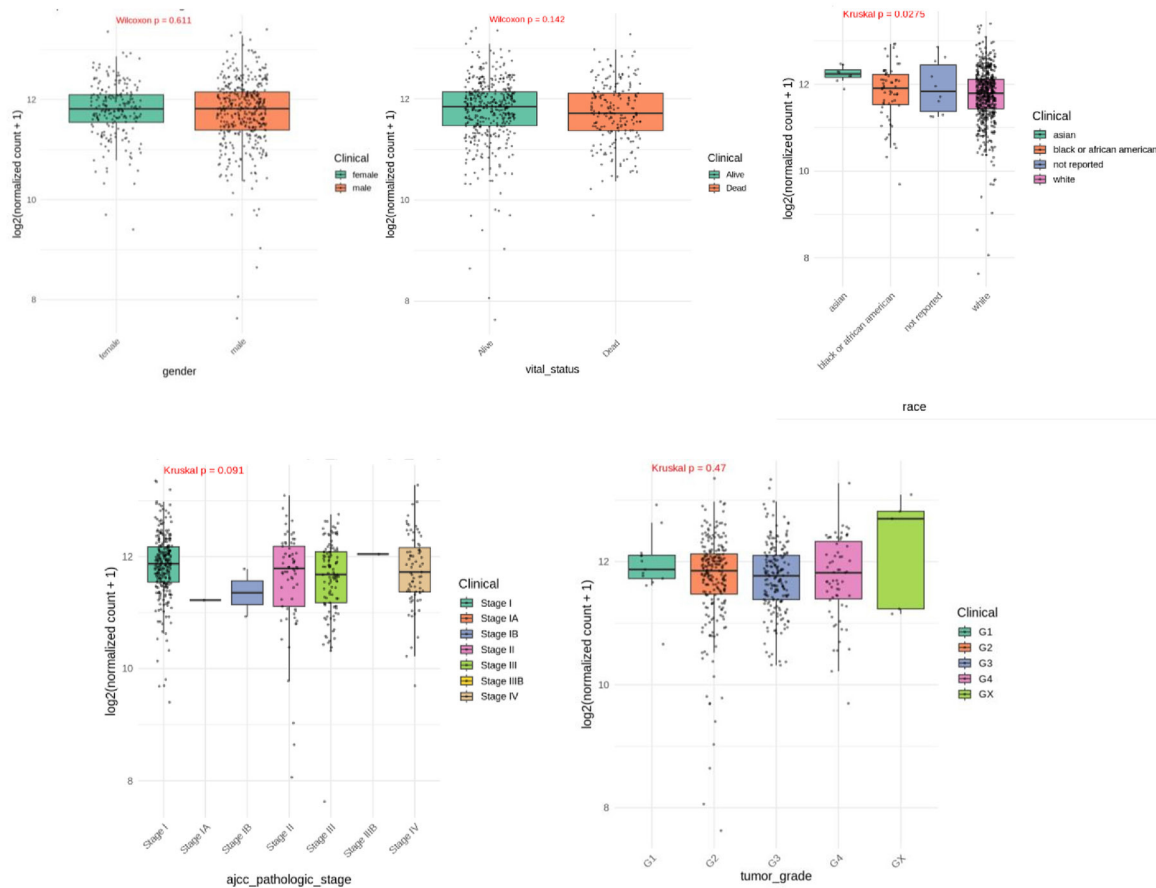
RCC: Renal cell carcinoma; Coef: Regression coefficient; HR: Hazard ratio; exp: Exponential function; SE: Standard error; P: Probability value

A**B**

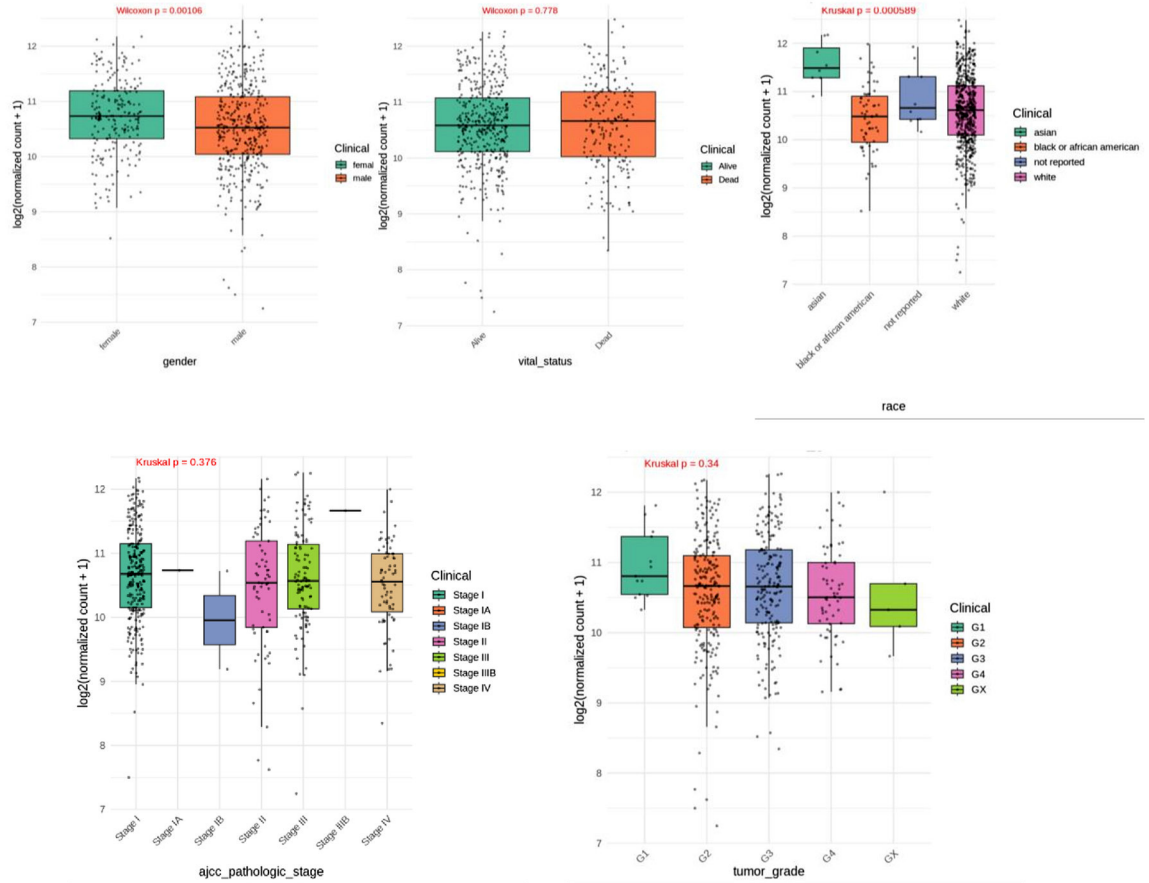
C



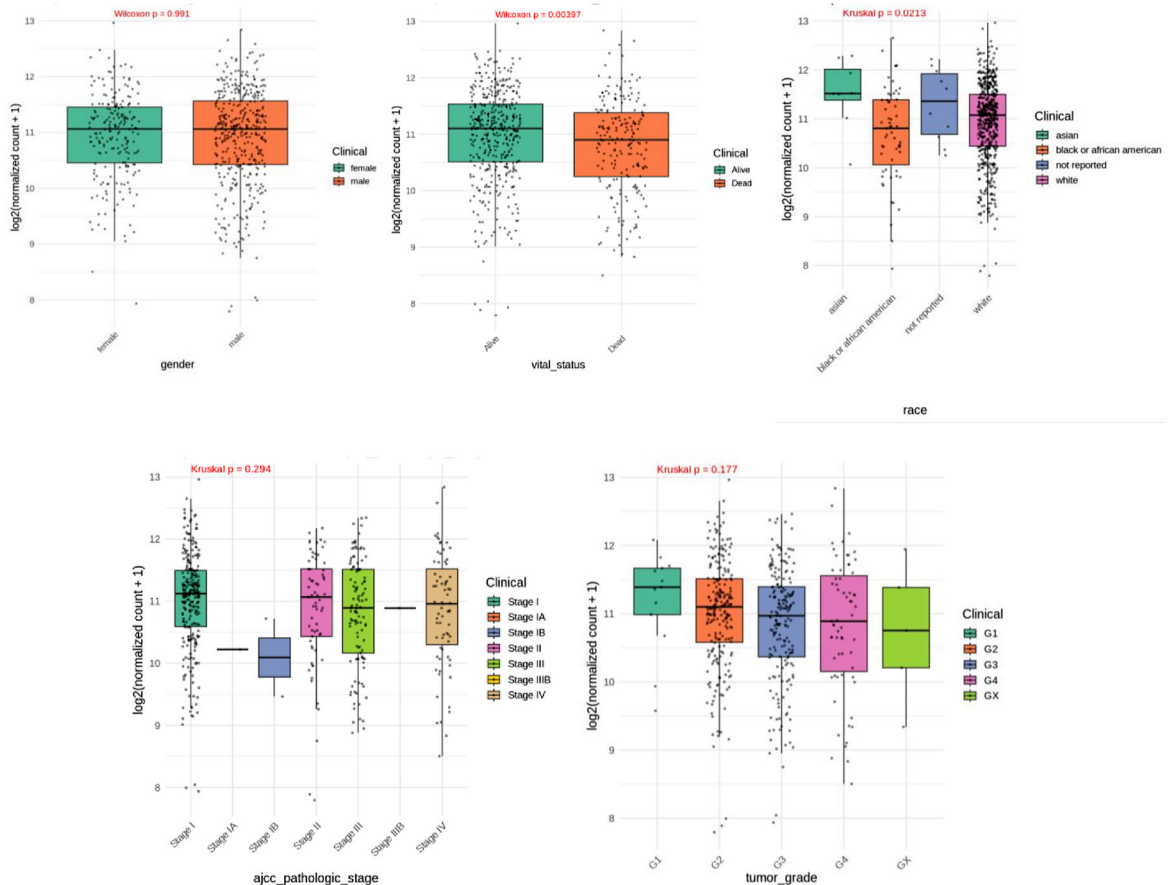
D



E



F



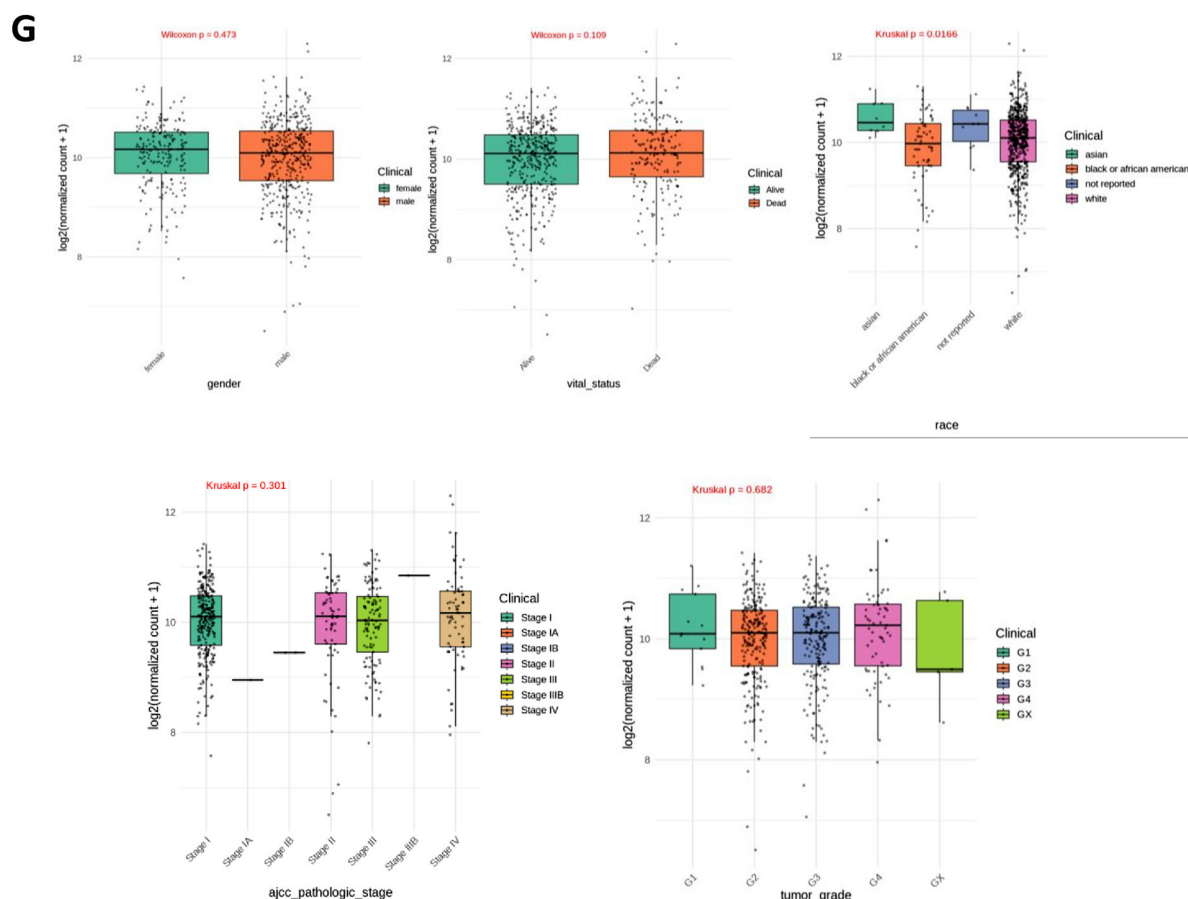


Figure 3: This figure displays the distribution of expression levels for key genes across various clinicopathologic variables within the TCGA-KIRC cohort. The box plots illustrate how the expression of (A) *PIK3CA*, (B) *SETD2*, (C) *TP53*, (D) *BAP1*, (E) *FLCN*, (F) *PBRM1*, and (G) *VHL* correlates with clinical factors, including sex, vital status, race, AJCC pathologic stage, and tumor grade. TCGA: The cancer genome atlas; KIRC: Kidney renal clear cell carcinoma; AJCC: American joint committee on cancer; log2: Base-2 logarithm; G1–G4: Tumor grade 1–4; GX: Grade cannot be assessed; *PIK3CA*: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; *SETD2*: SET domain containing 2; *TP53*: Tumor protein P53; *BAP1*: BRCA1 associated protein 1; *FLCN*: Folliculin; *PBRM1*: Polybromo 1; *VHL*: Von Hippel-Lindau

Table 3: Multivariable Cox regression of risk score and clinical variables

Variable	Coef	HR (exp[coef])	SE (coef)	z value	P value
RiskScoreLASSO	0.8	2.14	0.158	4.81	<0.001
Age	9.7×10^{-5}	1.00	1.96×10^{-5}	4.95	<0.001
Sex	0.05	1.05	0.157	0.33	0.74

RCC: Renal cell carcinoma; LASSO: Least absolute shrinkage and selection operator; RiskScoreLASSO: LASSO-derived composite risk score; Coef: Regression coefficient; HR: Hazard ratio; exp: Exponential function; SE: Standard error; P: Probability value

The analysis confirmed that both the LASSO risk score (HR=2.14, 95% CI=1.57–2.92, $P=1.51 \times 10^{-6}$) and age (HR=1.00, 95% CI=1.0001–1.0002, $P=7.30 \times 10^{-7}$) were significant independent predictors of overall survival, while sex (HR=1.05, $P=0.744$) did not show a significant effect (table 3). The model achieved a C-index of 0.673, reinforcing the clinical relevance and predictive value of the LASSO-based risk score (figure 5G).

Validation and Nomogram Construction

The prognostic risk model was further

validated in training and test subsets, confirming its reproducibility. A nomogram integrating the risk score and clinical variables provided individualized 1-, 3-, and 5-year overall survival predictions, demonstrating good calibration (figure 5H).

Cross-Validation of Expression Signatures and Dataset Integration

Meta-analysis of GEO datasets confirmed robust downregulation of core tumor suppressors, particularly *VHL*, *BAP1*, *PBRM1*, and *SETD2*.

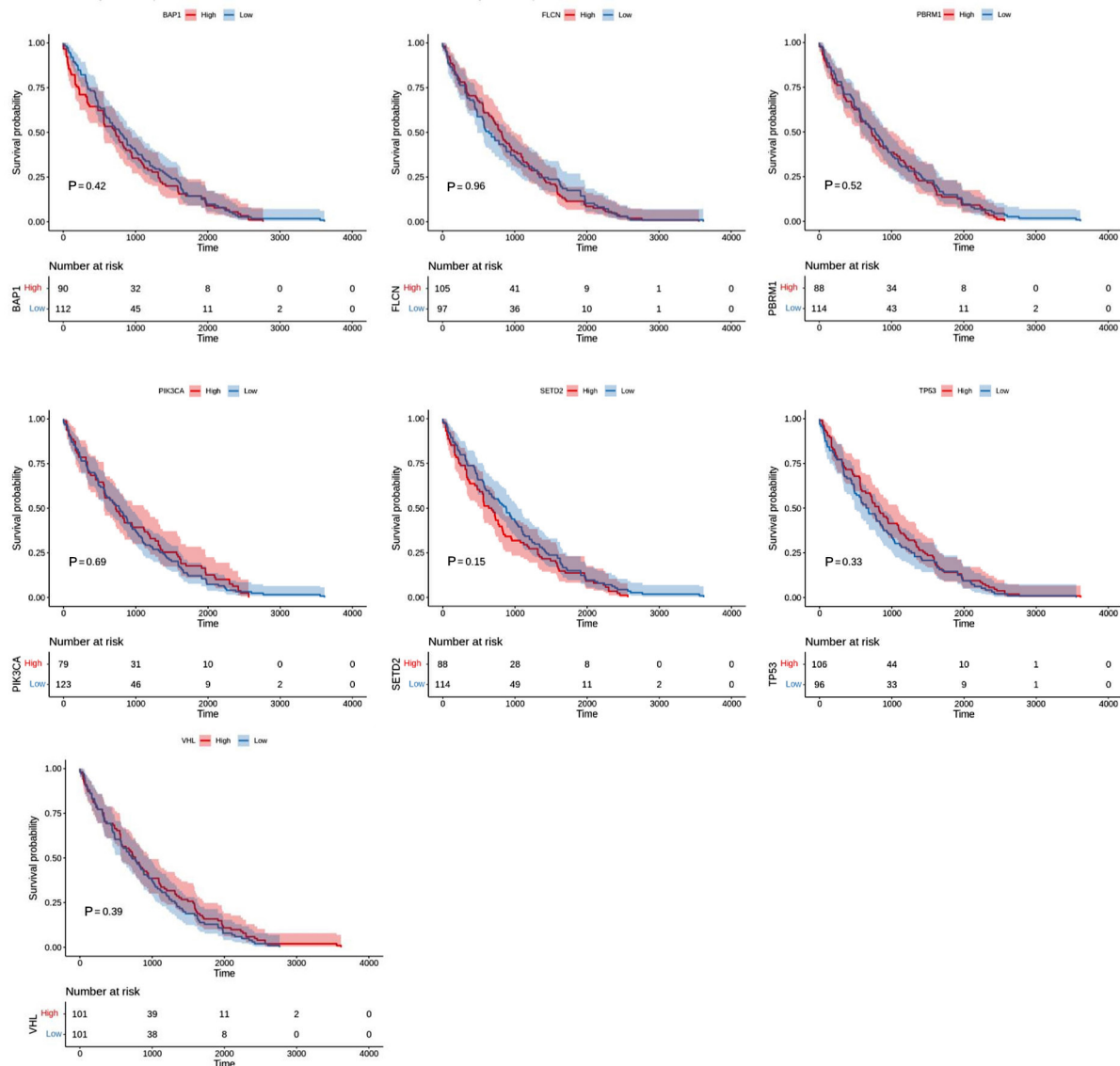


Figure 4: This figure illustrates the overall survival analysis for seven core genes using Kaplan-Meier curves and log-rank tests. Each panel demonstrates the survival probability over time (days) for patients stratified into high and low expression groups for *VHL*, *SETD2*, *PBRM1*, *BAP1*, *FLCN*, *TP53*, and *PIK3CA*, accompanied by "number at risk" tables. *BAP1*: BRCA1 associated protein 1; *FLCN*: Folliculin; *PBRM1*: Polybromo 1; *PIK3CA*: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; *SETD2*: SET domain containing 2; *TP53*: Tumor protein P53; *VHL*: Von Hippel-Lindau

Principal component analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) visualizations demonstrated effective batch correction, ensuring reliable cross-cohort integration for downstream analyses (figure 6A–B).

Nonsynonymous RNA Editing Events in Key Tumor Suppressor and Oncogene Genes

Analysis of the REDportal identified 25 nonsynonymous RNA editing sites across the seven targets, indicating potential impacts on protein function (table 4). Notably, multiple nonsynonymous editing events were found in the *TP53* gene, including variants such as p.H283R, p.K281R, and p.I279M, which occur

predominantly in exon 7 and exon 9 of different transcript isoforms. Additional nonsynonymous editing sites were detected in *SETD2* (e.g., p.Q9R, p.T8A) and *BAP1* (e.g., p.L259P, p.V255A), highlighting the potential functional relevance of RNA editing in these cancer-associated genes. No nonsynonymous editing sites were observed in the other target genes within this dataset.

Domain Localization of RNA Editing-Derived Mutations

Out of 25 identified nonsynonymous RNA editing events, 15 were located within known functional domains. In *TP53*, all 12 mutations were located within two functional domains.

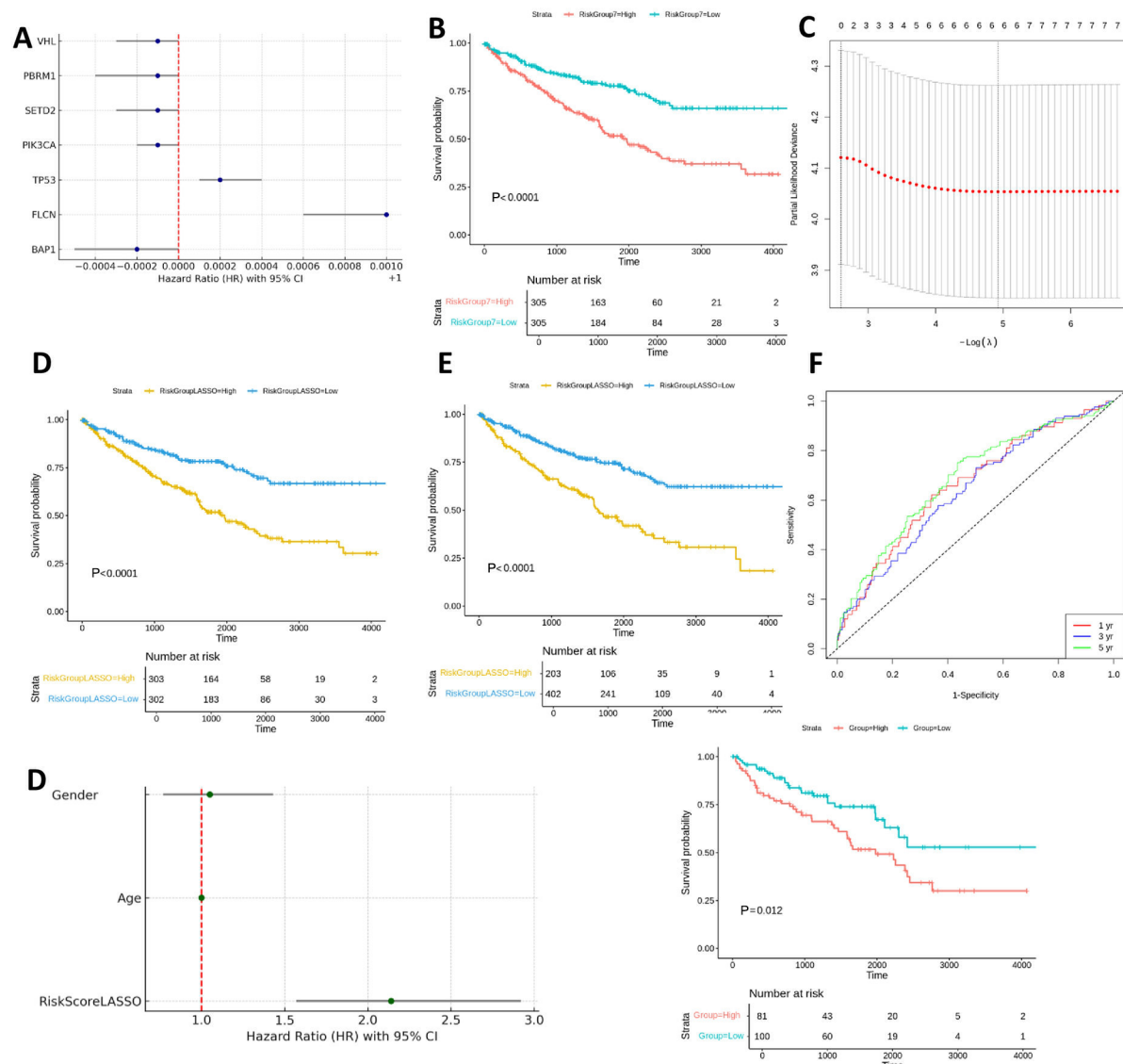


Figure 5: This figure presents the construction and validation of prognostic models based on key genes in patients with Renal Cell Carcinoma (RCC). The panels include: (A) Forest plot of multivariable Cox regression; (B) Kaplan-Meier curves for risk groups based on the seven-gene model; (C) LASSO coefficient profiles and cross-validation; (D-E) Survival analysis and predictive performance based on the LASSO model; (F) Time-dependent ROC curves at 1, 3, and 5 years; (G) Multivariable Cox regression including clinical variables; and (H) Nomogram for individualized survival prediction. RCC: Renal cell Carcinoma; LASSO: Least absolute shrinkage and selection operator; ROC: Receiver operating characteristic; HR: Hazard ratio; CI: Confidence interval; log: Logarithm; *BAP1*: BRCA1 associated protein 1; *FLCN*: Folliculin; *PBRM1*: Polybromo 1; *PIK3CA*: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; *SETD2*: SET domain containing 2; *TP53*: Tumor protein P53; *VHL*: Von Hippel-Lindau

Specifically, nine mutations (e.g., p.H283R, p.K281R, p.Q264R) mapped to the DNA-binding domain (residues 102–292), while three mutations (p.H342R, p.H338R, p.R337G) were within the Tetramerization domain (325–356). In *SETD2*, two out of eight mutations (p.R1643G and p.P1461L) were located in the SET domain (residues 1407–1706), which plays a role in histone methylation. In *BAP1*, only one mutation (p.R227C) was found within the UCH domain (residues 1–240), a region implicated in the protein's deubiquitylation activity (table 5).

Clinical Annotation of DomainLocalized RNA Editing Events

Two RNA editing-derived mutations had documented clinical significance. *TP53* p.R337C/G, within the tetramerization domain, is classified as pathogenic and associated with Li-Fraumeni syndrome and multiple cancers. *BAP1* p.R227C has uncertain significance in ClinVar. No RCC-relevant RNA editing events were recorded in COSMIC (table 6). These observations indicate that select editing events may have potential clinical relevance.

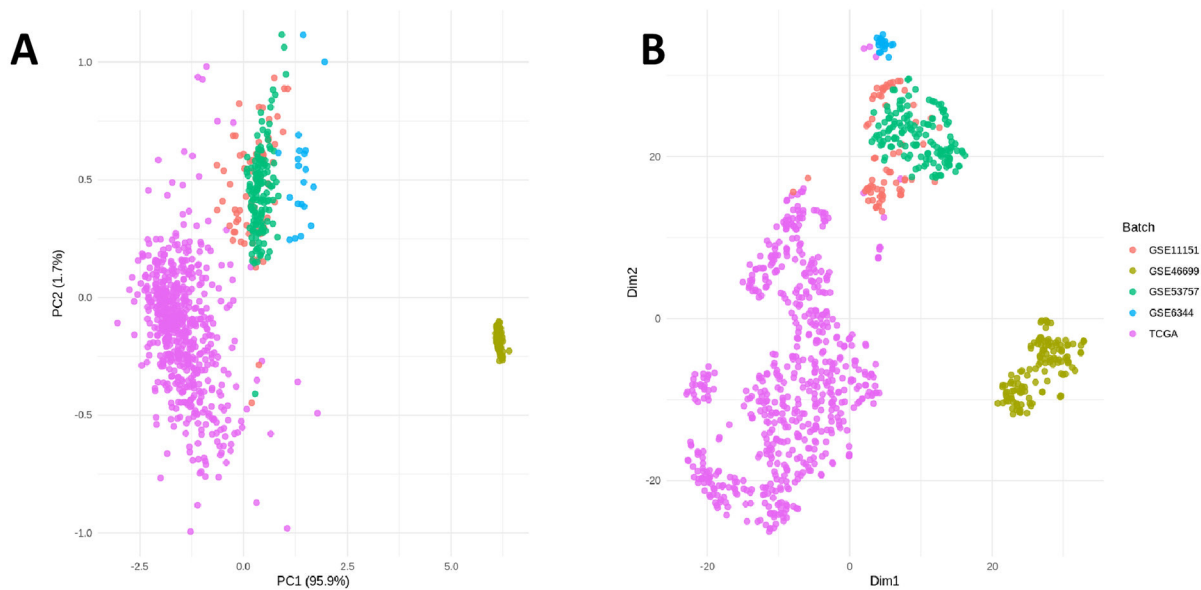


Figure 6: This figure illustrates the evaluation of batch effects and data integration across different datasets using dimensionality reduction techniques. (A) Principal Component Analysis (PCA) used to assess distribution and potential batch effects across TCGA and GEO datasets. (B) t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization showing the integration of samples from multiple sources (GSE11151, GSE46699, GSE53757, GSE6344, and TCGA). PCA: Principal component analysis; t-SNE: t-distributed stochastic neighbor embedding; PC1/PC2: Principal component 1 and 2; Dim1/Dim2: Dimension 1 and 2; TCGA: The cancer genome atlas; GSE: Gene expression omnibus series

Table 4: Nonsynonymous RNA editing sites in RCC-associated genes

Region	Position	Ref	Ed	Gene	Transcript isoform	Exon	Amino acid change
chr17	7565267	T	C	<i>TP53</i>	uc002gig.1	7	p.H283R
chr17	7565273	T	C	<i>TP53</i>	uc002gig.1	7	p.K281R
chr17	7565278	T	C	<i>TP53</i>	uc002gig.1	7	p.I279M
chr17	7565282	T	C	<i>TP53</i>	uc002gig.1	7	p.N278S
chr17	7565283	T	C	<i>TP53</i>	uc002gig.1	7	p.N278D
chr17	7565292	T	C	<i>TP53</i>	uc002gig.1	7	p.S275G
chr17	7565297	T	C	<i>TP53</i>	uc002gig.1	7	p.Q273R
chr17	7565324	T	C	<i>TP53</i>	uc002gig.1	7	p.Q264R
chr17	7565328	T	C	<i>TP53</i>	uc002gig.1	7	p.R263G
chr17	7569531	T	C	<i>TP53</i>	uc002gih.3	9	p.H342R
chr17	7569543	T	C	<i>TP53</i>	uc002gih.3	9	p.H338R
chr17	7569547	T	C	<i>TP53</i>	uc002gih.3	9	p.R337G
chr3	47082859	T	C	<i>SETD2</i>	uc003cqr.3	3	p.Q9R
chr3	47082863	T	C	<i>SETD2</i>	uc003cqr.3	3	p.T8A
chr3	47082865	T	C	<i>SETD2</i>	uc003cqr.3	3	p.H7R
chr3	47082868	T	C	<i>SETD2</i>	uc003cqr.3	3	p.H6R
chr3	47082872	T	C	<i>SETD2</i>	uc003cqr.3	3	p.R5G
chr3	47082881	T	C	<i>SETD2</i>	uc003cqr.3	3	p.T2A
chr3	47082884	T	C	<i>SETD2</i>	uc003cqr.3	3	p.M1V
chr3	47147269	T	C	<i>SETD2</i>	uc003cqv.3	7	p.R1643G
chr3	52440276	A	G	<i>BAP1</i>	uc003ddx.4	9	p.L259P
chr3	52440285	A	G	<i>BAP1</i>	uc003ddx.4	9	p.L256P
chr3	52440288	A	G	<i>BAP1</i>	uc003ddx.4	9	p.V255A
chr3	52440303	A	G	<i>BAP1</i>	uc003ddx.4	9	p.V250A
chr3	52440331	A	G	<i>BAP1</i>	uc003ddx.4	9	p.Y241H

Data sourced from the REDportal database. Nonsynonymous sites resulting in amino acid substitutions are shown for renal cell carcinoma associated genes. RCC: Renal cell carcinoma; Ref: Reference nucleotide; Ed: Edited nucleotide; HGVS: Human genome variation society

Table 5: RNA editing-induced amino acid substitutions in protein functional domains

Gene	Mutation	Position	Domain name
TP53	p.H283R	283	p53 DNA-binding domain
TP53	p.K281R	281	p53 DNA-binding domain
TP53	p.I279M	279	p53 DNA-binding domain
TP53	p.N278S	278	p53 DNA-binding domain
TP53	p.N278D	278	p53 DNA-binding domain
TP53	p.S275G	275	p53 DNA-binding domain
TP53	p.Q273R	273	p53 DNA-binding domain
TP53	p.Q264R	264	p53 DNA-binding domain
TP53	p.R263G	263	p53 DNA-binding domain
TP53	p.H342R	342	Tetramerization domain
TP53	p.H338R	338	Tetramerization domain
TP53	p.R337G	337	Tetramerization domain
SETD2	p.R1643G	1643	SET domain
SETD2	p.P1461L	1461	SET domain
BAP1	p.R227C	227	UCH domain

Substitutions are derived from RNA editing events located within functional domains of TP53, SETD2, and BAP1. HGVS: Human genome variation society; TP53: Tumor protein p53; SETD2: SET domain containing 2; BAP1: BRCA1 associated protein 1; UCH: Ubiquitin carboxy-terminal hydrolase

Table 6: Clinical significance of RNA editing-derived mutations

Gene	Mutation	ClinVar ID	Clinical significance	Disease association	Domain
TP53	p.R337G/C	142536/ 12379	Pathogenic/ Likely Pathogenic	LiFraumeni syndrome, various cancers (breast, adrenocortical)	Tetramerization
BAP1	p.R227C	237938	Uncertain significance	(Check ClinVar detail)	UCH

Clinical annotations and disease associations were retrieved from the ClinVar database. HGVS: Human genome variation society; ClinVar: Clinical variation database; ID: Identifier; TP53: Tumor protein p53; BAP1: BRCA1 associated protein 1; UCH: Ubiquitin carboxy-terminal hydrolase

Table 7: Candidate RNA editing-derived neoantigens across five human populations

Gene	Population	AA_Change	Peptide Length	Pep_Start	Pep_End	HLA	IC50_ED	IC50_WT	Delta_IC50
BAP1	Iranian	p.L256P	9	254	262	C*04:01	63.50	1387.71	1324.21
	Asian								
	American								
	European								
	Iranian		9	254	262	C*07:01	104.92	716.22	611.29
	European								
	Iranian		10	253	262	C*15:02	96.21	504.62	408.41
	African		10	253	262	A*34:01	78.097	1016.02	937.93
			9	254	262	C*18:02	97.58	2006.97	1909.39
			9	258	266	B*42:01	98.68	25177.9	25079.2
			10	253	262	C*17:01	107.014	873.11	766.094
			8	255	262	B*42:01	119.66	18418.5	18298.8
			10	253	262	C*02:10	126.85	1041.49	914.65
	Iranian	p.V255A	9	254	262	B*35:01	94.37	1400.8	1306.43
	European								
TP53	African	p.Y241H	9	254	262	B*53:01	84.6	1249.53	1164.91
	Iranian								
	African	p.Y241H	9	240	248	B*38:01	62.77	7686.55	7623.78
			9	240	248	B*39:01	79.49	7125.92	7046.43
	Iranian	p.R337G	9	336	344	B*51:01	113.47	10469.9	10356.4
	Asian								
	American		9	336	344	C*12:03	145.95	861.29	715.34

HLA: Human leukocyte antigen; AA: Amino acid; Pep: Peptide; IC50: Half-maximal inhibitory concentration; ED: Edited; WT: Wild type; Delta: Difference; TP53: Tumor protein p53; BAP1: BRCA1 associated protein 1

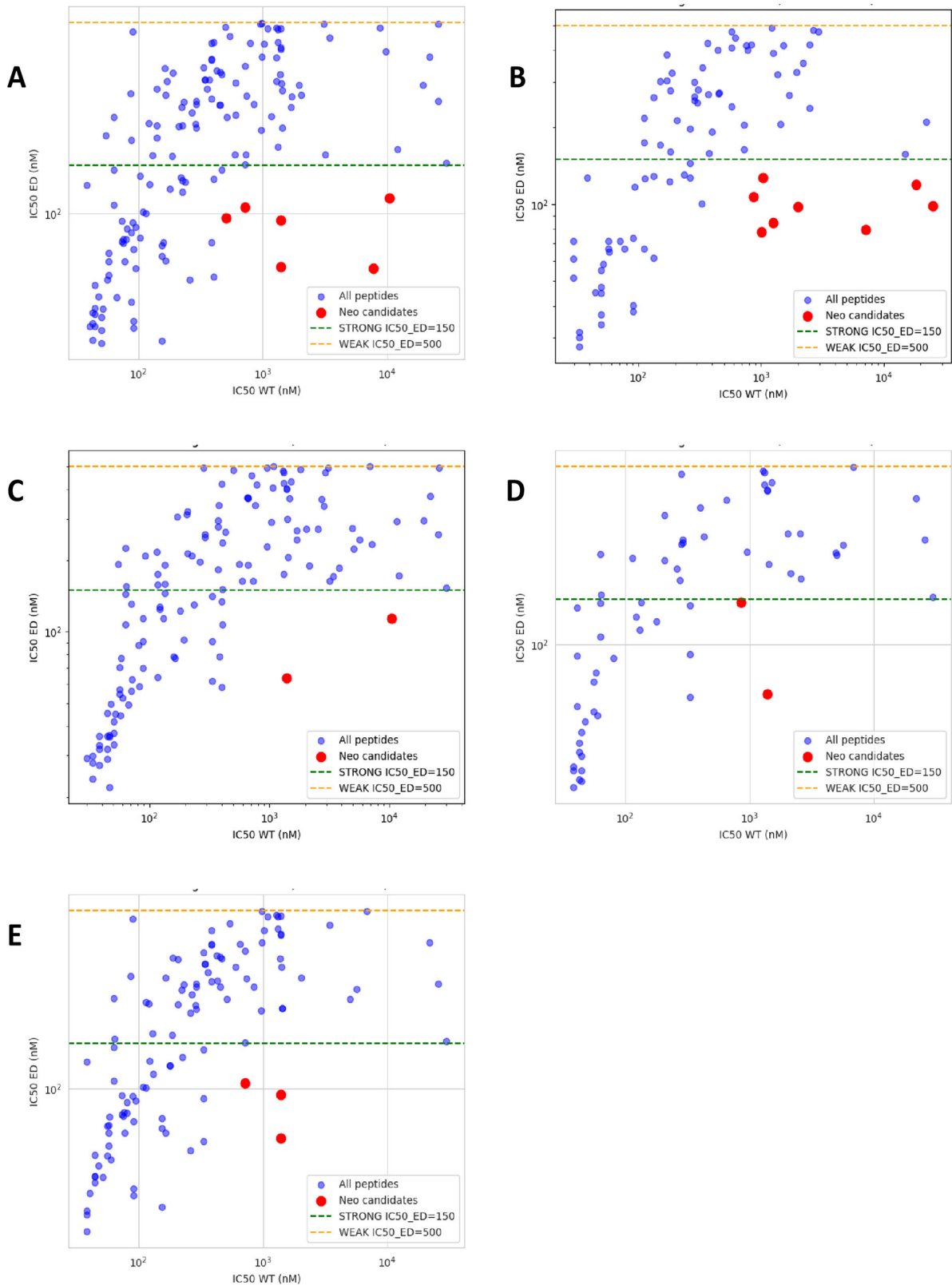


Figure 7: This figure presents comparative scatterplots of predicted binding affinities (IC₅₀) for edited (ED) versus wild-type (WT) peptides across five distinct global populations. Panels (A) to (E) represent Iranian (A), African (B), Asian (C), American (D), and European (E) populations. Each point corresponds to a peptide-HLA pair with predicted binding affinities, plotted as IC₅₀ for WT (x-axis) and ED (y-axis), both on a log scale. The blue dots represent all evaluated peptides, while the red dots indicate candidate neoantigens that meet the filtering criteria (IC₅₀_ED ≤ 150 nM and IC₅₀_WT ≥ 500 nM). Dashed lines mark the thresholds for strong (≤ 150 nM) and weak (≥ 500 nM) binding affinities. IC₅₀: Half maximal inhibitory concentration; ED: Edited; WT: Wild-type; nM: Nanomolar; HLA: Human leukocyte antigen; log: Logarithm

Population-Specific and Shared RNA Editing-Derived Neoantigens

RNA editing-derived peptides were evaluated for predicted HLA class I binding across five distinct populations (Iranian, African, Asian, American, and European). Application of stringent affinity thresholds ($IC_{50_ED} \leq 150$ nM; $IC_{50_WT} \geq 500$ nM) yielded population-specific and shared candidate neoantigens (table 7). Across populations, *BAP1* emerged as the predominant source of candidate neoantigens, with recurrent events at residues Y241H, V255A, L256P, and L259P. In contrast, *TP53* contributed a recurrent candidate at R337G, observed in Iranian, Asian, and American cohorts, while *SETD2* did not yield any peptides meeting selection criteria. A key finding was the identification of a public neoantigen: *BAP1* p.L256P (TVPEALQQL), which demonstrated strong binding (IC_{50_ED} range 63–107 nM) across all five populations. This peptide consistently showed large ΔIC_{50} values compared to the corresponding wild-type sequence, indicating enhanced immunogenic potential across multiple HLA alleles and ethnic

groups. Population-specific “private” candidates were also identified. For example, *BAP1* p.L259P was exclusive to African alleles, while *BAP1* p.Y241H and p.V255A were restricted to subsets of Iranian, African, and European HLA types. The *TP53* p.R337G neoantigen, although recurrent, was absent in the African and European datasets, highlighting differential population-level immunogenic landscapes. Scatter plot analyses of IC_{50} values (ED vs. WT) revealed a clear separation between strong-binding edited peptides and their weak-binding wild-type counterparts (figure 7A–E).

Mutational Landscape of Selected Genes

An oncoprint analysis of the seven selected genes was performed on 411 TCGA-KIRC samples. The analysis revealed that 288 samples (70.07%) harbored at least one mutation in these genes. The most frequently mutated genes were *VHL* (44% of samples) and *PBRM1* (41%), followed by *SETD2* (12%) and *BAP1* (10%). *TP53* and *PIK3CA* showed lower mutation frequencies (3% and 1%, respectively),

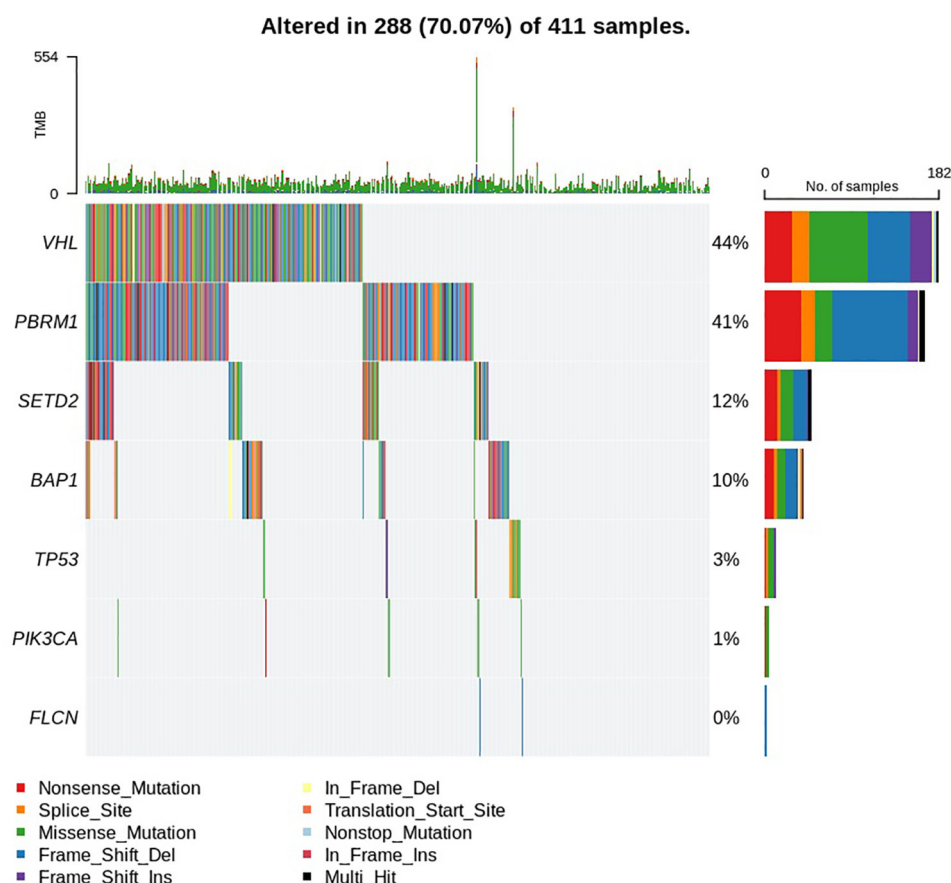


Figure 8: This Oncoprint visualization summarizes the mutation landscape of seven selected genes across 411 TCGA-KIRC samples. The plot details the distribution and frequency of various mutation types, including nonsense, missense, and frame-shift mutations, along with the Tumor Mutation Burden (TMB) for individual samples and the overall percentage of cases altered for each gene. TCGA: The cancer genome atlas; KIRC: Kidney renal clear cell carcinoma; TMB: Tumor mutation burden; *VHL*: Von hippel-lindau; *PBRM1*: Polybromo 1; *SETD2*: SET domain containing 2; *BAP1*: BRCA1 associated protein 1; *TP53*: Tumor protein P53; *PIK3CA*: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; *FLCN*: Folliculin

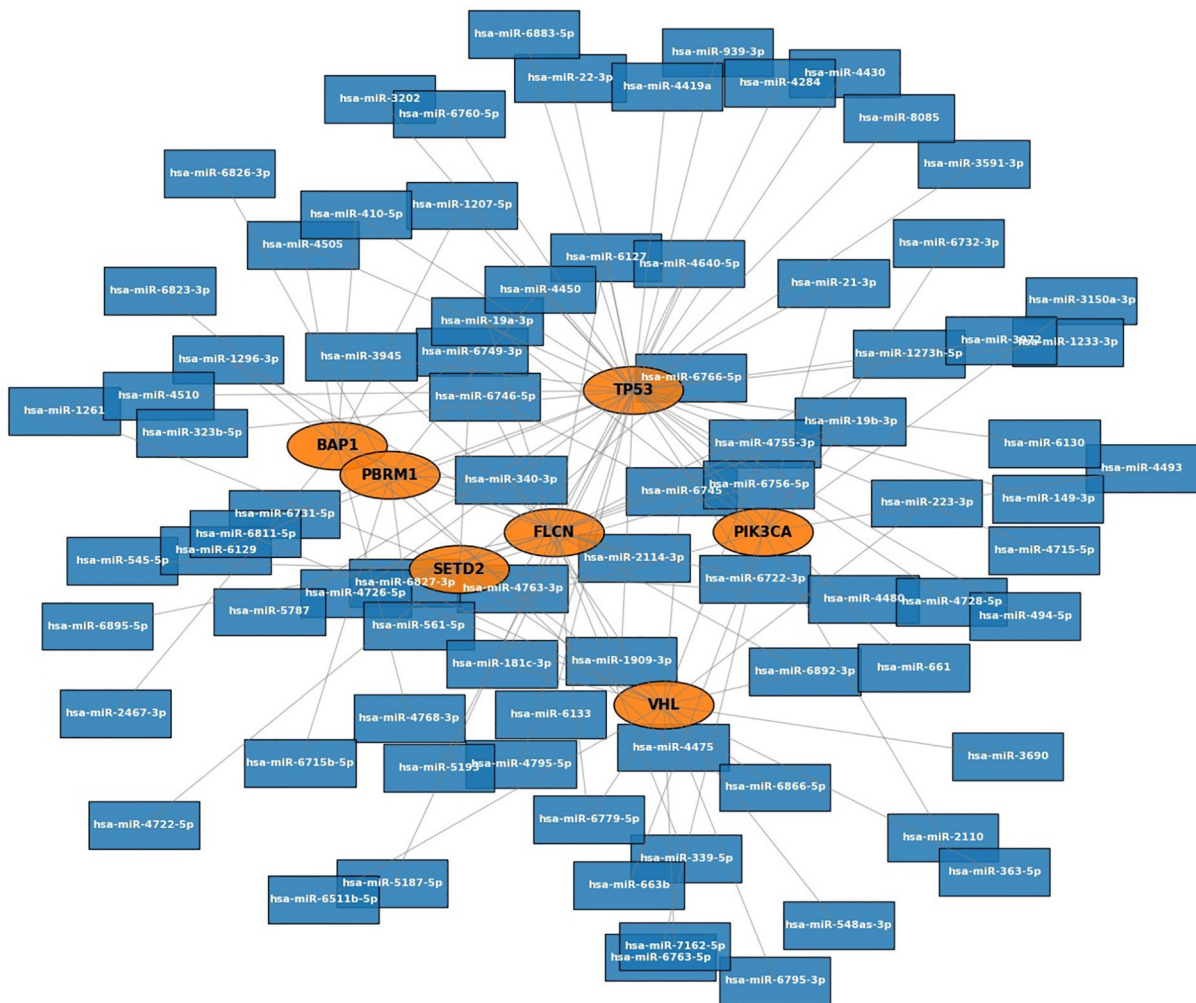


Figure 9: This figure illustrates the bipartite miRNA–mRNA regulatory network, identifying hub miRNAs that target multiple core genes in KIRC. The network visualizes the complex interactions between miRNAs (blue rectangles) and their target mRNAs (orange ovals), including *BAP1*, *FLCN*, *PBRM1*, *PIK3CA*, *SETD2*, *TP53*, and *VHL*. miRNA: MicroRNA; mRNA: Messenger RNA; KIRC: Kidney renal clear cell carcinoma; *BAP1*: BRCA1 associated protein 1; *FLCN*: Folliculin; *PBRM1*: Polybromo 1; *PIK3CA*: Phosphatidylinositol-4,5-bisphosphate 3-kinase Catalytic Subunit Alpha; *SETD2*: SET Domain Containing 2; *TP53*: Tumor Protein P53; *VHL*: Von Hippel-Lindau

while *FLCN* was not found to be mutated in this cohort. The oncoprint plot shows the tumor mutation burden (TMB), (figure 8) provides a detailed visualization of the mutation types and their distribution across the samples. Each column represents a single patient, and each row represents a gene. Notably, co-occurrence patterns were observed, suggesting a potential cooperative role between key mutated genes such as *VHL* and *PBRM1* in KIRC.

Identification of Hub miRNAs Regulating Core Genes

Network analysis identified 83 hub miRNAs targeting ≥ 3 core genes, retaining 119 high-confidence interactions. *VHL* and *PIK3CA* were the most frequently regulated, with several miRNAs coordinating tumor suppressors *TP53* and *BAP1*, highlighting central post-transcriptional regulators in RCC (figure 9).

GO Biological Process Enrichment Analysis

The seven core genes were enriched in cellular regulation and metabolism, including autophagy, negative regulation of catabolism, and transcription elongation, as well as TOR signaling, epigenetic regulation, and embryonic development, indicating broad functional impact in RCC (figure 10A).

Correlation Between RCC Driver Gene Expression and Immune Cell Infiltration

Most RCC driver genes showed weak correlations with immune cells, but *PBRM1*, *PIK3CA*, and *SETD2* had moderate positive associations ($p=0.3-0.4$) with endothelial cells and weaker links to fibroblasts and dendritic cells. *BAP1* showed weak negative correlations with CD8+ and exhausted T cells. Other genes showed minimal correlations, suggesting gene-specific modulation of the tumor immune

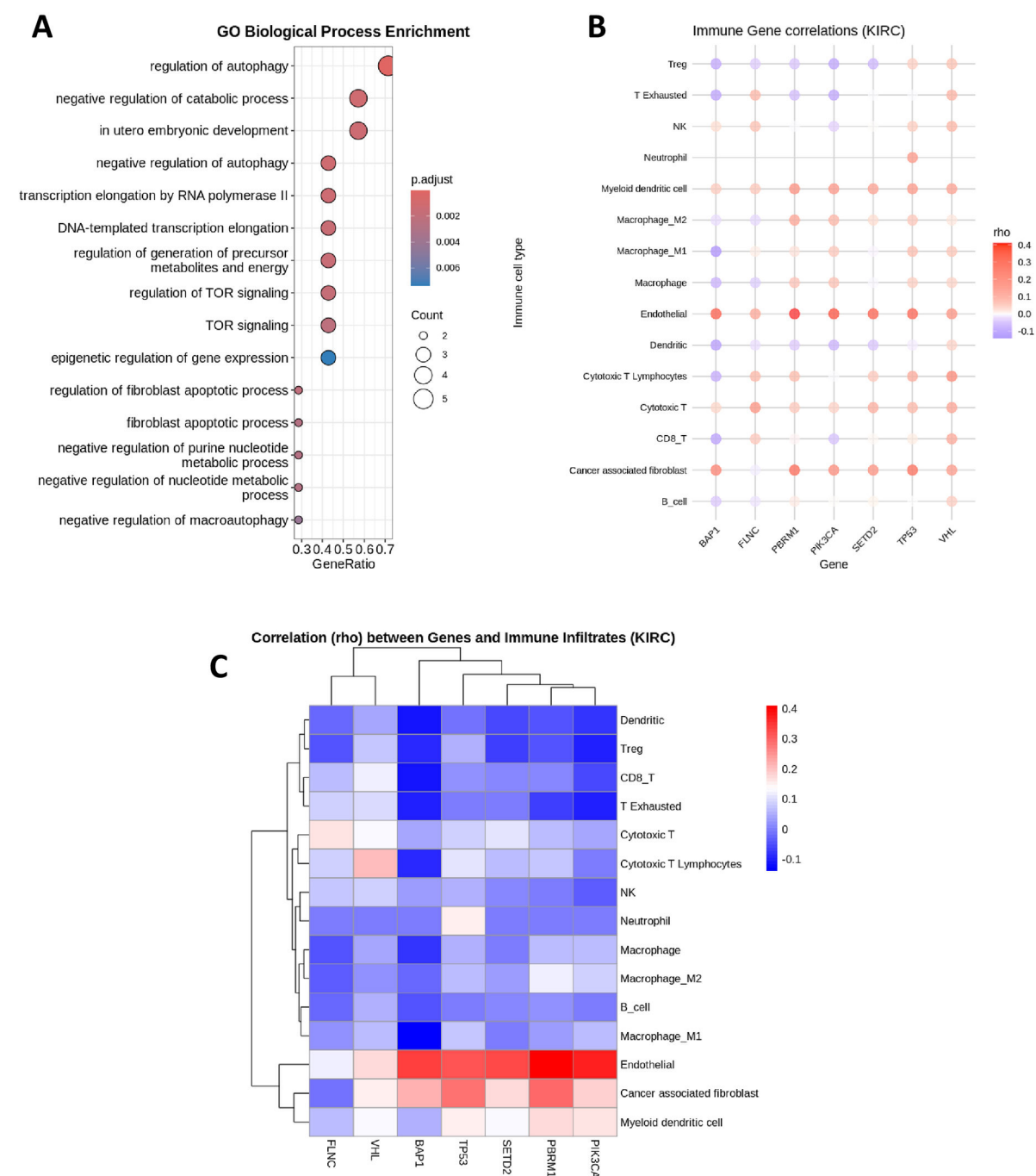


Figure 10: This figure presents the functional enrichment and immune infiltration patterns associated with RCC driver genes. (A) Dot plot of top enriched Gene Ontology (GO) Biological Processes. (B) Gene-immune cell correlation patterns in KIRC. (C) Heatmap showing the correlation between driver gene expression and various immune cell infiltrates in the TCGA-KIRC cohort. GO: Gene ontology; KIRC: Kidney renal clear cell carcinoma; TCGA: The cancer genome atlas; rho: Spearman's rank correlation coefficient; Treg: Regulatory T cell; NK: Natural killer cell; B_cell: B cell; *BAP1*: BRCA1 associated protein 1; *FLCN*: Folliculin; *PBRM1*: Polybromo 1; *PIK3CA*: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; *SETD2*: SET domain containing 2; *TP53*: Tumor protein P53; *VHL*: Von Hippel-Lindau

microenvironment (figure 10B–C).

Impact of Somatic Mutations on Immune Cell Infiltration

Immune infiltration differed between wild-type and mutant tumors. *BAP1* and *VHL* mutations strongly reduced NK ($\log_2\text{FC} < -4$) and regulatory

T cell infiltration. *TP53* mutations decreased Tregs in most cases, with occasional increases. *PBRM1*, *SETD2*, and *PIK3CA* mutations had modest, variable effects, while *FLCN* mutations had minimal impact. These results highlight *VHL* and *TP53* mutations as key contributors to immune evasion in RCC (figure 11A–B).

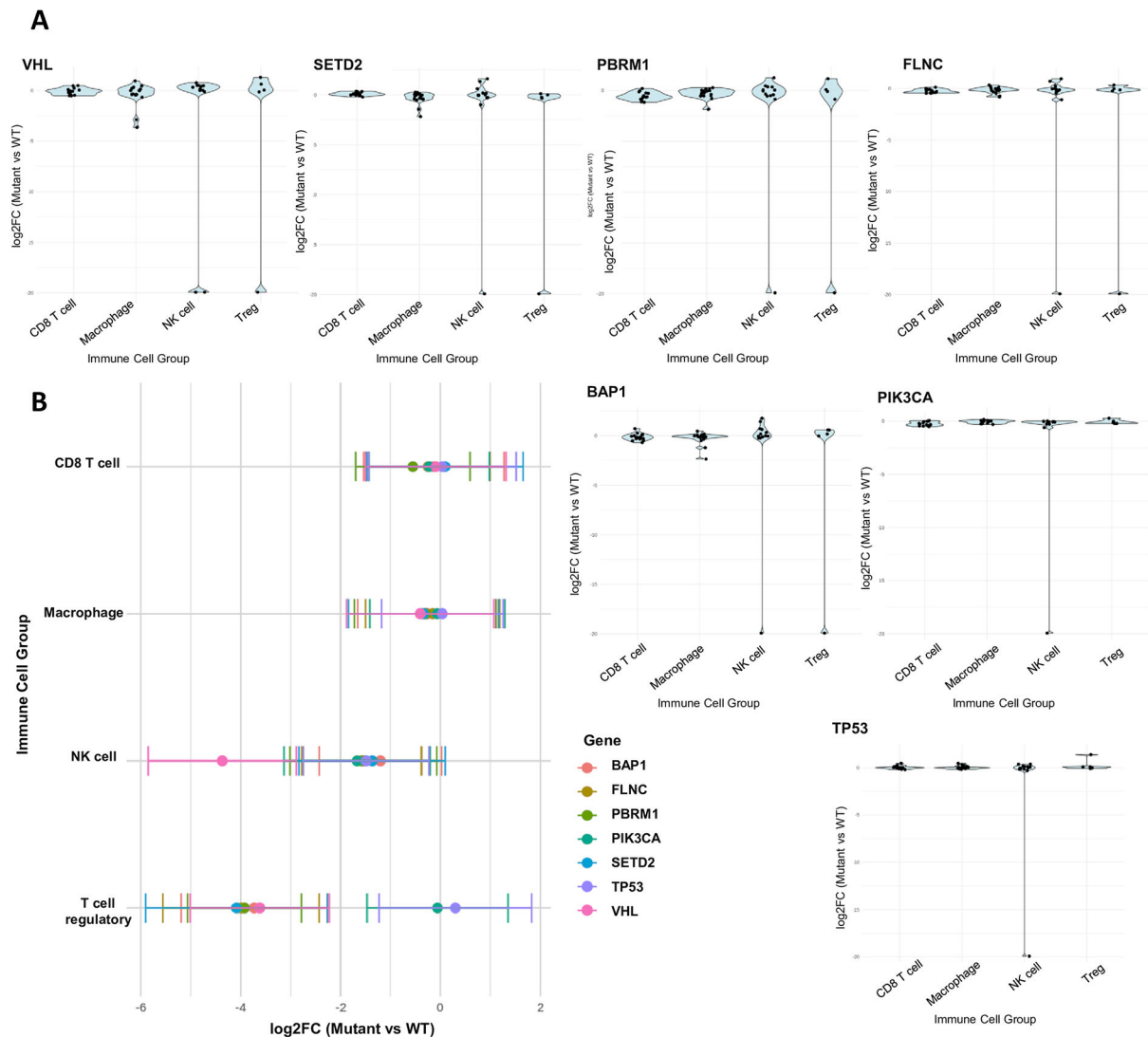


Figure 11: This figure illustrates the impact of somatic mutations on immune cell infiltration in KIRC through comparative fold change analysis. (A) Violin plots show the distribution of log₂ fold change (log₂FC) values for seven individual gene mutations across four key immune cell populations compared to wild-type samples. (B) A forest plot summarizes the effect size and statistical significance for these mutations, with horizontal bars representing 95% confidence intervals. log₂FC: Base-2 logarithm of fold change; WT: Wild-type; NK cell: Natural killer cell; Treg: Regulatory T cell; CD8 T cell: CD8+ Cytotoxic T lymphocyte; CI: Confidence interval; *BAP1*: BRCA1 associated protein 1; *FLCN*: Folliculin; *PBRM1*: Polybromo 1; *PIK3CA*: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; *SETD2*: SET domain containing 2; *TP53*: Tumor protein P53; *VHL*: Von Hippel-Lindau

Discussion

In this study, we integrated transcriptomic, mutational, and RNA-editing analyses of seven key RCC-associated genes (*VHL*, *BAP1*, *SETD2*, *PBRM1*, *FLCN*, *TP53*, and *PIK3CA*) using TCGA-KIRC and four GEO datasets. By combining multi-cohort gene expression, clinical associations, mutational profiling, RNA editing-derived neoantigen prediction, and immune infiltration analyses, we provide a comprehensive overview of how canonical genetic drivers and post-transcriptional modifications shape RCC biology and immunogenicity.

Consistent with prior studies, our results revealed marked downregulation of *VHL*, *BAP1*,

and *SETD2* in tumor tissues, in line with their well-established tumor suppressor roles. *VHL* inactivation remains a canonical initiating event in clear cell RCC, driving HIF stabilization and angiogenic signaling.¹⁷ Similarly, *SETD2* and *BAP1* downregulation reflects their 3p loss, epigenetic deregulation, and association with aggressive histology and poor clinical outcome.^{7,18} Our transcript-level findings reinforce previous protein-level and genomic studies, suggesting these tumor suppressors retain prognostic significance at multiple molecular layers. In contrast, *TP53* showed consistent upregulation despite its relatively low mutation frequency in RCC. This pattern may reflect p53 pathway activation under cellular stress rather than

canonical mutation-driven stabilization.^{9, 19} Such findings are consistent with other cancers where *TP53* overexpression is observed in aggressive phenotypes even in the absence of hotspot mutations. *FLCN* displayed heterogeneous expression across cohorts, suggesting context-dependent contributions, while *PBRM1* also showed variable transcript-level deregulation despite being a frequent mutational target. *PIK3CA* was variably upregulated, consistent with its context-specific role in activating the PI3K/AKT/mTOR pathway in RCC.¹⁹

While the expression of individual genes such as *PIK3CA*, *SETD2*, and *TP53* correlates with clinical outcomes, their predictive power in isolation is somewhat limited. This limitation can be attributed to the multifactorial nature of RCC, where gene expression is influenced by a variety of factors, including mutations, epigenetic modifications, and post-transcriptional regulation. Among clinical associations, *PIK3CA*, *SETD2*, and *TP53* demonstrated the strongest correlations with prognosis, stage, and grade, underscoring their clinical utility.²⁰⁻²² In contrast, *BAP1*, *FLCN*, *PBRM1*, and *VHL* exhibited more limited associations, suggesting a context-dependent prognostic value.²³ This observed heterogeneity in prognostic strength suggests that single-gene expression levels may be insufficient to fully capture the complexity of RCC progression. Therefore, the combined LASSO signature was essential to aggregate the weighted effects of the six genes. This approach integrates the cumulative risk stemming from the collective deregulation of multiple, interconnected tumor suppressor and oncogenic pathways (such as chromatin remodeling, hypoxia signaling, and cell cycle control) rather than relying on the noisy signal of a single marker. The resulting signature provides a more robust and stable measure for comprehensive risk stratification. Multivariable Cox regression identified *FLCN* and *TP53* as independent prognostic factors, supporting their potential as biomarkers of risk stratification. Furthermore, our refined six-gene LASSO signature (*BAP1*, *SETD2*, *TP53*, *PBRM1*, *FLCN*, *VHL*) achieved moderate predictive accuracy (C-index~0.63), comparable to other transcriptomic prognostic models.^{24, 25} Although the C-index is moderate, we argue that the combination of multiple genes significantly improves the model's predictive ability, as individual gene expression does not fully account for the diverse molecular mechanisms driving RCC. By integrating these seven genes, we are able to account for both genetic alterations and post-transcriptional

modifications that contribute to tumor progression and immune escape. This combined approach, by incorporating genetic and RNA-level data, offers a more comprehensive understanding of the disease than examining individual genes in isolation.

RNA editing analysis highlighted an additional layer of transcriptomic complexity. We identified 25 nonsynonymous RNA editing events, particularly within *TP53*, *SETD2*, and *BAP1*. Many of these events mapped to functional protein domains, and in some cases, overlapped with pathogenic variants in ClinVar, suggesting potential functional relevance. Importantly, several editing-derived peptides showed strong HLA class I binding. For example, *BAP1* p.L256P was predicted as a “public” neoantigen, consistently immunogenic across multiple populations, representing a promising candidate for vaccine development. In contrast, epitopes such as *TP53* p.R337G displayed population specificity, underscoring the importance of tailoring immunotherapies to population genetics.²⁶⁻²⁸ In our other study on Iranian RCC patients, we demonstrated that population-specific genetic profiling and neoantigen prediction reveal distinct immunogenic patterns, which support the current findings and emphasize the importance of integrating RNA editing-derived signatures with HLA diversity.²⁹ Our mutational analysis confirmed the predominance of *VHL* and *PBRM1* alterations, often co-occurring, consistent with their synergistic roles in chromatin remodeling and hypoxia signaling pathways.³⁰ Mutations in *SETD2*, *BAP1*, and *TP53* were less frequent but clinically relevant, while *FLCN* mutations were absent, suggesting that its role in sporadic RCC may be limited compared to hereditary syndromes. These patterns reinforce the notion that RCC is driven by both recurrent genetic events and context-specific modulators.

Beyond mutations, our miRNA–mRNA regulatory network revealed 83 hub miRNAs targeting multiple driver genes, including *VHL*, *TP53*, and *PIK3CA*. Such hubs highlight the critical contribution of post-transcriptional regulation in shaping RCC biology and suggest that miRNAs may function as master regulators coordinating multiple oncogenic and tumor suppressor pathways.^{31, 32} Functional enrichment analyses implicated autophagy, TOR signaling, and chromatin remodeling, aligning with prior reports linking these pathways to RCC progression and therapy resistance.^{33, 34}

Immune profiling revealed distinct associations between gene alterations and immune cell infiltration. *PBRM1* and *SETD2* correlated

with endothelial cell infiltration, whereas *BAP1* mutations were linked to reduced NK and Treg infiltration, consistent with prior immunogenomic studies.³⁵⁻³⁷ These findings suggest that specific genetic drivers influence immune evasion in gene-specific manners, offering potential explanations for heterogeneous immunotherapy responses in RCC.³⁸

Despite multi-omics analysis, residual batch effects and discordance between mRNA and protein levels (e.g., *VHL*, *PBRM1*) may bias results. Prognostic models showed moderate predictive power (C-index~0.63–0.67), underscoring the need for additional molecular or clinical covariates. Future studies should integrate proteomic/phosphoproteomic data, validate RNA-editing–derived neoantigens, and develop comprehensive multi-omics models to enhance risk stratification and guide personalized immunotherapy in RCC.

Conclusion

This study establishes an integrative multi-omics framework linking genetic alterations, RNA editing, and miRNA-mediated regulation with immune modulation in renal cell carcinoma. We identified a six-gene prognostic signature (*BAP1*, *SETD2*, *TP53*, *PBRM1*, *FLCN*, *VHL*) that robustly stratifies patients into clinically relevant risk groups. Recurrent RNA editing events, particularly within *TP53*, *BAP1*, and *SETD2*, generated immunogenic neoantigens with both population-wide and ethnicity-specific relevance, underscoring their potential for precision immunotherapy. In parallel, miRNA regulatory networks and immune profiling highlighted additional layers of tumor progression and immune evasion. Collectively, these findings refine the molecular understanding of RCC heterogeneity and provide actionable insights for prognostic modeling and therapeutic development. This work lays a foundation for precision oncology strategies, including biomarker-guided risk assessment, population-tailored immunotherapies, and next-generation vaccine design in RCC.

Acknowledgment

We would like to thank all individuals who contributed to and supported this study.

Authors' Contribution

All authors contributed in the study concept and design; Z.M, G.A.K participated in formal analysis, investigation, data curation,

visualization, and writing original draft preparation. G.A.K participated in project administration, supervision and funding acquisition. All authors participate in reviewing final version of the manuscript. All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

AI Declaration

No artificial intelligence tools or technologies were used in the design, analysis, writing, or preparation of this manuscript.

Conflict of Interest: None declared.

References

- 1 Padala SA, Barsouk A, Thandra KC, Saginala K, Mohammed A, Vakiti A, et al. Epidemiology of Renal Cell Carcinoma. *World J Oncol*. 2020;11:79-87. doi: 10.14740/wjon1279. PubMed PMID: 32494314; PubMed Central PMCID: PMC7239575.
- 2 Li J, Luo P, Liu S, Fu M, Lin A, Liu Y, et al. Effective strategies to enhance the diagnosis and treatment of RCC: The application of biocompatible materials. *Mater Today Bio*. 2024;27:101149. doi: 10.1016/j.mtbio.2024.101149. PubMed PMID: 39100279; PubMed Central PMCID: PMC11296058.
- 3 Xie D, Li G, Zheng Z, Zhang X, Wang S, Jiang B, et al. The molecular code of kidney cancer: A path of discovery for gene mutation and precision therapy. *Mol Aspects Med*. 2025;101:101335. doi: 10.1016/j.mam.2024.101335. PubMed PMID: 39746268.
- 4 Kim E, Zschiedrich S. Renal Cell Carcinoma in von Hippel-Lindau Disease-From Tumor Genetics to Novel Therapeutic Strategies. *Front Pediatr*. 2018;6:16. doi: 10.3389/fped.2018.00016. PubMed PMID: 29479523; PubMed Central PMCID: PMC5811471.
- 5 Sumiyoshi T, Yamasaki T, Takeda M, Mizuno K, Utsunomiya N, Sakamoto H, et al. Detection of von Hippel-Lindau gene mutation in circulating cell-free DNA for clear cell renal cell carcinoma. *Cancer Sci*. 2021;112:3363-74. doi: 10.1111/cas.14972. PubMed PMID: 34009695; PubMed Central PMCID: PMC8353948.
- 6 Mazumder S, Higgins PJ, Samarakoon R. Downstream Targets of VHL/HIF- α Signaling

- in Renal Clear Cell Carcinoma Progression: Mechanisms and Therapeutic Relevance. *Cancers* (Basel). 2023;15. doi: 10.3390/cancers15041316. PubMed PMID: 36831657; PubMed Central PMCID: PMC9953937.
- 7 Bihr S, Ohashi R, Moore AL, Rüschoff JH, Beisel C, Hermanns T, et al. Expression and Mutation Patterns of PBRM1, BAP1 and SETD2 Mirror Specific Evolutionary Subtypes in Clear Cell Renal Cell Carcinoma. *Neoplasia*. 2019;21:247-56. doi: 10.1016/j.neo.2018.12.006. PubMed PMID: 30660076; PubMed Central PMCID: PMC6355619.
- 8 Walton J, Lawson K, Prinios P, Finelli A, Arrowsmith C, Ailles L. PBRM1, SETD2 and BAP1 - the trinity of 3p in clear cell renal cell carcinoma. *Nat Rev Urol*. 2023;20:96-115. doi: 10.1038/s41585-022-00659-1. PubMed PMID: 36253570.
- 9 Croessmann S, Wong HY, Zabransky DJ, Chu D, Rosen DM, Cidado J, et al. PIK3CA mutations and TP53 alterations cooperate to increase cancerous phenotypes and tumor heterogeneity. *Breast Cancer Res Treat*. 2017;162:451-64. doi: 10.1007/s10549-017-4147-2. PubMed PMID: 28190247; PubMed Central PMCID: PMC6100768.
- 10 Lu YR, Yuan Q, Liu J, Han X, Liu M, Liu QQ, et al. A rare occurrence of a hereditary Birt-Hogg-Dubé syndrome: A case report. *World J Clin Cases*. 2021;9:7123-32. doi: 10.12998/wjcc.v9.i24.7123. PubMed PMID: 34540968; PubMed Central PMCID: PMC8409184.
- 11 Mai TL, Chuang TJ. A-to-I RNA editing contributes to the persistence of predicted damaging mutations in populations. *Genome Res*. 2019;29:1766-76. doi: 10.1101/gr.246033.118. PubMed PMID: 31515285; PubMed Central PMCID: PMC6836733.
- 12 Zhang S, Xiao X, Yi Y, Wang X, Zhu L, Shen Y, et al. Tumor initiation and early tumorigenesis: molecular mechanisms and interventional targets. *Signal Transduct Target Ther*. 2024;9:149. doi: 10.1038/s41392-024-01848-7. PubMed PMID: 38890350; PubMed Central PMCID: PMC11189549.
- 13 Kurkowiak M, Arcimowicz Ł, Chruściel E, Urban-Wójciuk Z, Papak I, Keegan L, et al. The effects of RNA editing in cancer tissue at different stages in carcinogenesis. *RNA Biol*. 2021;18:1524-39. doi: 10.1080/15476286.2021.1877024. PubMed PMID: 33593231; PubMed Central PMCID: PMC8582992.
- 14 Wang Z, Gu Y, Sun X, Huang H. Computation strategies and clinical applications in neoantigen discovery towards precision cancer immunotherapy. *Biomark Res*. 2025;13:96. doi: 10.1186/s40364-025-00808-9. PubMed PMID: 40629481; PubMed Central PMCID: PMC12239460.
- 15 Richard G, Princiotta MF, Bridon D, Martin WD, Steinberg GD, De Groot AS. Neoantigen-based personalized cancer vaccines: the emergence of precision cancer immunotherapy. *Expert Rev Vaccines*. 2022;21:173-84. doi: 10.1080/14760584.2022.2012456. PubMed PMID: 34882038.
- 16 Roerden M, Castro AB, Cui Y, Harake N, Kim B, Dye J, et al. Neoantigen architectures define immunogenicity and drive immune evasion of tumors with heterogenous neoantigen expression. *J Immunother Cancer*. 2024;12. doi: 10.1136/jitc-2024-010249. PubMed PMID: 39521615; PubMed Central PMCID: PMC11552027.
- 17 Shirole NH, Kaelin WG, Jr. von-Hippel Lindau and Hypoxia-Inducible Factor at the Center of Renal Cell Carcinoma Biology. *Hematol Oncol Clin North Am*. 2023;37:809-25. doi: 10.1016/j.hoc.2023.04.011. PubMed PMID: 37270382; PubMed Central PMCID: PMC11315268.
- 18 Ma C, Liu M, Feng W, Rao H, Zhang W, Liu C, et al. Loss of SETD2 aggravates colorectal cancer progression caused by SMAD4 deletion through the RAS/ERK signalling pathway. *Clin Transl Med*. 2023;13:e1475. doi: 10.1002/ctm2.1475. PubMed PMID: 37962020; PubMed Central PMCID: PMC10644329.
- 19 Wang H, Guo M, Wei H, Chen Y. Targeting p53 pathways: mechanisms, structures, and advances in therapy. *Signal Transduct Target Ther*. 2023;8:92. doi: 10.1038/s41392-023-01347-1. PubMed PMID: 36859359; PubMed Central PMCID: PMC9977964.
- 20 Shigaki H, Baba Y, Watanabe M, Murata A, Ishimoto T, Iwatsuki M, et al. PIK3CA mutation is associated with a favorable prognosis among patients with curatively resected esophageal squamous cell carcinoma. *Clin Cancer Res*. 2013;19:2451-9. doi: 10.1158/1078-0432.ccr-12-3559. PubMed PMID: 23532889.
- 21 Yang X, Chen R, Chen Y, Zhou Y, Wu C, Li Q, et al. Methyltransferase SETD2 inhibits tumor growth and metastasis via STAT1-IL-8 signaling-mediated epithelial-mesenchymal transition in lung adenocarcinoma. *Cancer Sci*. 2022;113:1195-207. doi: 10.1111/cas.15299. PubMed PMID: 35152527; PubMed Central PMCID: PMC8990294.
- 22 Qayoom H, Haq BU, Sofi S, Jan N, Jan A, Mir MA. Targeting mutant p53: a key player in breast cancer pathogenesis and beyond.

- Cell Commun Signal. 2024;22:484. doi: 10.1186/s12964-024-01863-9. PubMed PMID: 39390510; PubMed Central PMCID: PMC11466041.
- 23 Hudler P, Urbancic M. The Role of VHL in the Development of von Hippel-Lindau Disease and Erythrocytosis. *Genes (Basel)*. 2022;13. doi: 10.3390/genes13020362. PubMed PMID: 35205407; PubMed Central PMCID: PMC8871608.
 - 24 Zhang H, Xu J, Long Y, Maimaitijiang A, Su Z, Li W, et al. Unraveling the Guardian: p53's Multifaceted Role in the DNA Damage Response and Tumor Treatment Strategies. *Int J Mol Sci*. 2024;25. doi: 10.3390/ijms252312928. PubMed PMID: 39684639; PubMed Central PMCID: PMC11641486.
 - 25 Feng T, Zhao J, Wei D, Guo P, Yang X, Li Q, et al. Immunogenomic Analyses of the Prognostic Predictive Model for Patients With Renal Cancer. *Front Immunol*. 2021;12:762120. doi: 10.3389/fimmu.2021.762120. PubMed PMID: 34712244; PubMed Central PMCID: PMC8546215.
 - 26 Zhang Y, Li L, Mendoza JJ, Wang D, Yan Q, Shi L, et al. Advances in A-to-I RNA editing in cancer. *Mol Cancer*. 2024;23:280. doi: 10.1186/s12943-024-02194-6. PubMed PMID: 39731127; PubMed Central PMCID: PMC11673720.
 - 27 Li X, You J, Hong L, Liu W, Guo P, Hao X. Neoantigen cancer vaccines: a new star on the horizon. *Cancer Biol Med*. 2023;21:274-311. doi: 10.20892/j.issn.2095-3941.2023.0395. PubMed PMID: 38164734; PubMed Central PMCID: PMC11033713.
 - 28 Ding YP, Liu CC, Yu KD. RNA modifications in the tumor microenvironment: insights into the cancer-immunity cycle and beyond. *Exp Hematol Oncol*. 2025;14:48. doi: 10.1186/s40164-025-00648-1. PubMed PMID: 40176140; PubMed Central PMCID: PMC11963313.
 - 29 Mehmandoostli Z, Ashkezari MD, Seifati SM, Farzin A, Kardar GA. Exploring neoantigens and genetic profiles in renal cell carcinoma: a study of Iranian patients. *J Curr Oncol Med Sci*. 2025;5:1103-14.
 - 30 Peña-Llopis S, Vega-Rubín-de-Celis S, Liao A, Leng N, Pavía-Jiménez A, Wang S, et al. BAP1 loss defines a new class of renal cell carcinoma. *Nat Genet*. 2012;44:751-9. doi: 10.1038/ng.2323. PubMed PMID: 22683710; PubMed Central PMCID: PMC3788680.
 - 31 Page PM, Dastous SA, Richard PO, Pavic M, Nishimura T, Riazalhosseini Y, et al. MicroRNA profiling identifies VHL/HIF-2 α dependent miR-2355-5p as a key modulator of clear cell Renal cell carcinoma tumor growth. *Cancer Cell Int*. 2025;25:71. doi: 10.1186/s12935-025-03711-3. PubMed PMID: 40016765; PubMed Central PMCID: PMC11869434.
 - 32 Yang W, Ma J, Zhou W, Cao B, Zhou X, Zhang H, et al. Reciprocal regulations between miRNAs and HIF-1 α in human cancers. *Cell Mol Life Sci*. 2019;76:453-71. doi: 10.1007/s00018-018-2941-6. PubMed PMID: 30317527; PubMed Central PMCID: PMC11105242.
 - 33 Peng S, Xie Z, Jiang H, Zhang G, Chen N. Revealing the characteristics of SETD2-mutated clear cell renal cell carcinoma through tumor heterogeneity analysis. *Front Genet*. 2024;15:1447139. doi: 10.3389/fgene.2024.1447139. PubMed PMID: 39119581; PubMed Central PMCID: PMC11306021.
 - 34 Zamora-Fuentes JM, Hernández-Lemus E, Espinal-Enríquez J. Methylation-related genes involved in renal carcinoma progression. *Front Genet*. 2023;14:1225158. doi: 10.3389/fgene.2023.1225158. PubMed PMID: 37693315; PubMed Central PMCID: PMC10486271.
 - 35 Liu T, Xia Q, Zhang H, Wang Z, Yang W, Gu X, et al. CCL5-dependent mast cell infiltration into the tumor microenvironment in clear cell renal cell carcinoma patients. *Aging (Albany NY)*. 2020;12:21809-36. doi: 10.18632/aging.103999. PubMed PMID: 33177244; PubMed Central PMCID: PMC7695370.
 - 36 Xu D, Gao Y, Yang H, Spils M, Marti TM, Losmanová T, et al. BAP1 Deficiency Inflames the Tumor Immune Microenvironment and Is a Candidate Biomarker for Immunotherapy Response in Malignant Pleural Mesothelioma. *JTO Clin Res Rep*. 2024;5:100672. doi: 10.1016/j.jtocr.2024.100672. PubMed PMID: 38715965; PubMed Central PMCID: PMC11070913.
 - 37 Chang H, Li M, Zhang L, Li M, Ong SH, Zhang Z, et al. Loss of histone deubiquitinase Bap1 triggers anti-tumor immunity. *Cell Oncol (Dordr)*. 2025;48:183-203. doi: 10.1007/s13402-024-00978-y. PubMed PMID: 39141316; PubMed Central PMCID: PMC11850471.
 - 38 Zhang A, Miao K, Sun H, Deng CX. Tumor heterogeneity reshapes the tumor microenvironment to influence drug resistance. *Int J Biol Sci*. 2022;18:3019-33. doi: 10.7150/ijbs.72534. PubMed PMID: 35541919; PubMed Central PMCID: PMC9066118.