

# A Proteogenomic Atlas of Clear Cell Renal Cell Carcinoma in a Chinese Population

**Chen Ding** (✉ [chend@fudan.edu.cn](mailto:chend@fudan.edu.cn))

Fudan University

**Yuan-Yuan Qu**

Fudan University Shanghai Cancer Center

**Jinwen Feng**

Fudan University <https://orcid.org/0000-0003-1683-4280>

**Xiaohui Wu**

Fudan University

**Lin Bai**

Fudan University

**Wenhao Xu**

Fudan University

**Lingli Zhu**

Fudan University

**Yang Liu**

Fudan University

**Fujiang Xu**

School of Life Sciences, Fudan University <https://orcid.org/0000-0002-7900-5319>

**Xuan Zhang**

Fudan University

**Guojian Yang**

Fudan University

**Jiacheng Lv**

Fudan University

**Xiuping Chen**

Fudan University

**Guohai Shi**

Fudan University

**Hongkai Wang**

Fudan University

**Dalong Cao**

Fudan University

**Hang Xiang**

Fudan University

**Lingling Li**

Fudan University

**Subei Tan**

Fudan University

**Hualei Gan**

Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China

**Menghong Sun**

Fudan University

**Hailiang Zhang**

Fudan University

**Jian-Yuan Zhao**

Fudan University

**Dingwei Ye**

Fudan University

---

**Article**

**Keywords:** clear cell renal cell carcinoma, proteogenomics, nicotinamide N-methyltransferase, lysine homocysteinylation

**Posted Date:** February 26th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-221123/v1>

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

[Read Full License](#)

---

# Abstract

Renal cell carcinoma (RCC) is among the top 10 malignant carcinomas<sup>1</sup>. Clear cell (cc)RCC, accounting for ~ 75% of RCC cases, is an aggressive histological RCC subtype. In the last decade, large-scale multiomics studies have profoundly enhanced our understanding of this disease<sup>2,3</sup>. However, despite the differences of genomic alterations between Western and Eastern ccRCC<sup>4,5</sup>, these studies mostly focused on patients in Western populations. Here we conducted a comprehensive proteogenomic analysis of 232 tumor and adjacent non-tumor tissue pairs from Chinese ccRCC patients. Genomic analysis revealed unique genetic features of Chinese ccRCC and distinct mutation patterns associated with copy number alterations. Based on proteomic profiles, ccRCC showed extensive metabolic dysregulation, especially in one-carbon metabolism. We classified ccRCC into three subtypes (GP1–3), among which the most aggressive GP1 exhibited dominant immune response, metastasis, and metabolic imbalance, linking the proteomic features, genomic alterations, and clinical outcomes of ccRCC. Nicotinamide N-methyltransferase (NNMT) and NNMT mediated protein homocysteinylation were identified as a poor prognosis indicator and a drug target for GP1, respectively. We demonstrated that NNMT induces DNA-dependent protein kinase catalytic subunit (DNA-PKcs) homocysteinylation, increases DNA repair, and promotes tumor growth in ccRCC. Treatment of N-acetyl-cysteine (NAC), an inhibitor of homocysteinylation, markedly reduced the NNMT overexpression induced radioresistance of tumor cells. This study provided valuable insights into the biological underpinnings and prognosis assessment of ccRCC, revealing a targetable metabolic vulnerability.

## Introduction

Multiomics strategies encompassing genome and expression profiling of multiple tumor types have elucidated novel molecular subtypes and abnormally activated signaling pathways, as well as potential therapeutic targets<sup>3,6-12</sup>. The Cancer Genome Atlas (TCGA) and Clinical Proteomic Tumor Analysis Consortium (CPTAC) have published landmark multiomics studies<sup>2,3</sup>, improving our cognition of ccRCC. CPTAC conducted an integrated proteogenomics analysis in 103 ccRCC cases, which revealed the immune signature of ccRCC, highlighting the necessity and significance of proteogenomics research.

In this study, we conducted genomic and proteomic profiling of 232 paired tumor and adjacent non-tumor samples of Chinese ccRCC patients with a median follow-up of 85 months (range, 3–138 months). Our study revealed distributions of genetic lesions unique to the Chinese cohort and shared between the Chinese and Western populations. Integrated data analysis disclosed the associations among genetic aberrations, proteomic features and clinical outcomes of ccRCC. Further, we identified NNMT as biomarker for poor prognosis, and verified that NNMT overexpression mediated homocysteine metabolism dysregulation was a potential therapeutic opportunity for renal cell carcinoma.

## Results

### Proteogenomic Landscape of Chinese ccRCC

We collected 232 paired tumor and adjacent non-tumor tissues from Chinese ccRCC patients based on strict criteria (**Figure S1A**). Clinicopathological indicators, including age at surgery, sex, clinical manifestation, laterality, tumor size, chronic diseases status, tumor node metastasis (TNM) stage, and International Society of Urological Pathology (ISUP) grading classification are summarized in **Table S1**. Tumor and adjacent tissue regions were determined by pathological examinations. Freshly frozen tissues were used for proteomics analysis and whole-exome sequencing (WES). WES was conducted in 224 paired samples; samples from 8 patients were excluded due to low DNA quality (Fig. 1A).

WES data of tumor adjacent tissues were used as a reference to detect genetic variants in ccRCC. The mean sequencing coverage in the hg38 reference genome was 120.5× for tumor tissues and 68.72× for adjacent tissues (**Table S2, Figure S1B**). Among the 224 sample pairs, 10,475 non-silent mutations in 6,875 genes and 1,203 silent mutations were detected. *VHL* was the most frequently mutated gene in this cohort (64.3%), followed by *PBRM1* (24.5%), *BAP1* (10.7%) and *SETD2* (8.9%) (Fig. 1B), consistent with previous studies<sup>2,4,5,13</sup> (Fig. 1C). Interestingly, mutation frequencies of these genes exhibited ethnic and geographic variations in the TCGA<sup>2</sup>, and European<sup>5</sup> cohorts. Specifically, *VHL* had the lowest mutation frequency in the TCGA cohort (64.3% in Chinese vs. 46.2% in TCGA vs. 73.4% in European). In contrast, *PBRM1* (24.5% in Chinese vs. 42.1% in TCGA vs. 39.4% in European), *SETD2* (8.0% in Chinese vs. 13.9% in TCGA vs. 19.1% in European), and *MTOR* (4.5% in Chinese vs. 9.2% in TCGA vs. 8.5% in European) had lower mutation frequencies in the Chinese cohort (Fig. 1C). Mutational spectra revealed that C > T transversion (27.0%) was the dominant mutation in the Chinese and TCGA cohorts (**Figure S1C**). The frequency of A > T transversions was higher in the Chinese cohort than in the TCGA cohort (21.0% vs. 10.6%, **Figure S1C**). When we decomposed the mutation spectra using the Catalogue of Somatic Mutations in Cancer (COSMIC) database<sup>14</sup>, five single-base substitution (SBS) signatures (SBS1, SBS5, SBS22, SBS40, SBS52) were detected (**Figure S1D**). Signatures SBS1, SBS5, and SBS40 were considered to be correlated with patient age. SBS22 was associated with exposure to aristolochic acid (AA), a Chinese herbal ingredient associated with renal injury and ccRCC carcinogenesis<sup>4,5,15</sup>, corroborating that AA exposure is a carcinogenic factor for Chinese ccRCC. Moreover, patients with the AA signature showed higher mutational burden ( $p < 0.0001$ ), but lower CNA values ( $p < 0.05$ ) (**Figures S1E-F**).

For proteomic data analysis, Spearman's correlation coefficient was calculated for all quality control (QC) runs using HEK293T cell samples (**Figure S1G**). The average correlation coefficient of the QC samples was 0.95 (range, 0.82–0.99), demonstrating consistent stability of the mass spectrometry (MS) platform. The correlation coefficients of the 232 tumor and adjacent samples were 0.35–0.89 (mean, 0.78) and 0.71–0.91 (mean, 0.84), respectively (**Figures S1H-I**). The lower correlation values among tumor samples indicated a high level of tumor heterogeneity. We identified 8,193–10,273 proteins in each sample (Fig. 1D). Significantly more proteins were identified in tumors (median, 9,697) than in paired adjacent tissues (median, 8,915) (paired t test,  $p < 0.0001$ , **Figure S1J**), indicating the complexity of tumor microenvironment. In total, 16,915 proteins were detected in the 232 paired samples (**Figure S1K**), among which 14,159 proteins were in common between tumor and adjacent tissues, whereas 1,606 and 1,150 proteins were detected specifically in tumor and adjacent tissues, respectively (Fig. 1E, **Table S3**). When

we compared our data with those of the CPTAC ccRCC study<sup>3</sup>, 10,581 proteins were found in both cohorts, whereas 6,332 and 774 proteins were detected specifically in the Chinese cohort and the CPTAC cohort, respectively (Fig. 1E). Proteome quantification was conducted using the iBAQ algorithm, followed by normalization to fraction of total (FOT) as reported previously<sup>16,17</sup> (**Figure S1L**). The dynamic range of proteins detected spanned eight orders of magnitude (**Figure S1L**). In summary, this study provided a comprehensive landscape of Chinese ccRCC at both the genomic and the proteomic levels.

### Genomic Alterations and Their Proteomic Consequences

Somatic (S)CNAs in Chinese ccRCC were identified using GISTIC<sup>18</sup> (**Table S2**). The most frequently arm-level deleted chromosomal regions in our cohort were 3p (83%), 14q (44%), and 9q (25%), whereas the most frequently arm-level amplified regions were 5q (49%), 7q (26%), 5p (24%) and 7p (24%) (Fig. 2A, **Table S2**), which is consistent with previous reports<sup>2,3</sup>. SCNA profiles were shown in **Figures S2A**; 3p 25.1 loss was found in 91% of tumors. We used Cox regression to identify associations between arm-level CNAs and clinical outcomes. Loss of 3p was associated with better survival (hazard ratio [HR] = 0.47; 95% confidence interval [CI], 0.24–0.907;  $p = 0.0245$ ), whereas 9p loss (HR = 2.2; 95%CI, 1.18–4.02;  $p = 0.0133$ ), 9q loss (HR = 1.9; 95%CI, 1.01–3.45;  $p = 0.0472$ ), 14q loss (HR = 1.9; 95%CI, 1.04–3.47;  $p = 0.0359$ ), and 12q gain (HR = 2.2; 95%CI, 11.1–4.52;  $p = 0.0256$ ) were associated with poorer survival (Fig. 2B).

Genomic alterations that affect gene expression levels at the same locus are said to act in *cis*, whereas an impact of another locus is defined as a *trans* effect<sup>10,19</sup>. Diagonal patterns in Fig. 2C represent *cis* effects of CNAs, and vertical patterns indicate *trans* effects. We found that 89,992 CNA–protein pairs showed significant correlation (multiple test-adjusted  $p < 0.01$ , Spearman's correlation). The CNAs with *cis* effect were centered around 3q, 5q, 12p, and 12p, whereas those with *trans* effect were centered around 3p, 3q, 9p, 9q, 14p, and 14q (Fig. 2C). Among CNAs with a strong *trans* effect, 3p CNA values were associated with prognosis (HR = 4.7; 95%CI, 1.36–16.4;  $p = 0.015$ ) (Fig. 2B). Specifically, high 3p loss burden (H-3PLB) was associated with favorable overall survival (OS) compared to low 3p loss burden (L-3PLB) (log-rank test,  $p = 0.038$ , Fig. 2D). This result was confirmed in the TCGA cohort (log-rank test,  $p = 0.009$ , Fig. 2D). Next, we investigated clinical information and mutation frequency differences between the H-3PLB and L-3PLB groups (Fig. 2E). 3p loss burden was negatively associated with TNM stage (Fisher's exact test,  $p = 0.042$ ). *BAP1*, a ccRCC driver gene, was mutated more frequently in L-3PLB than in H-3PLB (Fisher's exact test,  $p = 0.024$ ), whereas *PBRM1*, another driver gene, had a significantly lower mutation frequency in L-3PLB than in H-3PLB (Fisher's exact test,  $p = 0.019$ ). This phenomenon was also found in the TCGA dataset ( $p < 0.05$ , **Table S2**). As 3p loss is an initial genetic event in ccRCC<sup>20,21</sup>, this result suggested that clinical phenotypes might be determined at the time when 3p loss occurs. Integrated analysis of the genomic and proteomic data identified that 67 proteins, positively correlated with 3p CNA values (Spearman's correlation,  $q < 0.05$ ), converged on complement and coagulation cascades (Figs. 2E, S2B). Meanwhile, 1,343 proteins, negatively correlated with 3p CNA values, enriched in pathways including glycolysis, metabolism of lipids and signaling by receptor tyrosine kinases (**Figures S2B-C**).

We assessed the impact of somatic mutations on encoded protein expression levels. Thirty-six genes, including two tumor-suppressor genes (TSGs; *PBRM1* and *PTEN*), exhibited differential protein levels in mutated vs. WT tumors (mutation frequency > 1%, fold change [FC] > 2,  $p < 0.05$ , **Figures S3A-B**). We investigated the impact of the mutations on the OS in the Chinese ccRCC cohort. Mutations in eight genes with a mutation frequency > 2% (*BAP1*, *FBN1*, *ABCA13*, *CEP128*, *DST*, *COL6A6*, *PTEN*, and *SON*) were associated with poor prognosis (log-rank test,  $p < 0.05$ , **Figures S3C-D**). The impact of *BAP1* mutations on OS in ccRCC was confirmed in the TCGA cohort (log-rank test,  $p = 0.0106$ , **Figure S3E**). *FBN1*, *ABCA13*, *CEP128*, *DST*, *COL6A6*, *PTEN*, and *SON* mutations were correlated to poor prognosis only in the Chinese cohort. These results demonstrated the unique connection between mutation panels and clinical outcomes in the Chinese cohort.

*BAP1* mutations were mutually exclusive with *PBRM1* mutations in the Chinese cohort (Fig. 2F). Patients with *PBRM1* mutations had improved clinical outcomes compared to patients with *BAP1* mutations (log-rank test,  $p = 0.0053$ , Fig. 2G). Consistent with previous reports<sup>20,22,23</sup>, *BAP1* mutations were associated with higher TNM stage (Fisher's exact test,  $p = 0.0021$ ) and ISUP grade (Fisher's exact test,  $p = 0.0032$ ) than *PBRM1* mutations. Considering that *BAP1* and *PBRM1* mutations represented distinct evolutionary trajectories of ccRCC, we evaluated proteome alterations caused by these two mutations using Gene Set Enrichment Analysis (GSEA) (Fig. 2H). Complement and coagulation cascades were enriched in *BAP1*-mutated tumors (normalized enrichment score [NES] = 3.04, false discovery rate [FDR] = 0), while Glycolysis/gluconeogenesis were enriched in *PBRM1*-mutated tumor (NES = -1.95, FDR = 5.12E-3) (Figs. 2H-I). Furthermore, higher levels of most complement and coagulation cascade components (e.g., C1, C5, C6, C7, C8, C9, FGA, FGB, FGG, and PLG) were indicators of poor prognosis (log-rank test,  $p < 0.05$ , **Figure S4**). Activation of C1 and C5 promotes angiogenic processes and drives immunosuppression in ccRCC<sup>24,25</sup>. Fibrinogen genes, including *FGA*, *FGB*, and *FGG*, are associated with cell proliferation regulation, and increased fibrinogen gene expression has been shown in various malignancies<sup>26</sup>. In conclusion, we demonstrated the chromosome 3p loss is associated with ccRCC clinical outcomes. Chromosome 3p loss drives the distinct mutations of *BAP1* and *PBRM1* and regulates the expression of proteins in the complement and coagulation cascades in a *trans*-regulatory mode, which further influences clinical outcomes.

### Proteomic Alterations in ccRCC Compared to Adjacent Tissues

To obtain a general insight into the proteomic alterations in ccRCC tumor tissues compared to adjacent tissues, 6,111 proteins detected in > 25% of the patients were further analyzed (**Table S4**). Principle component analysis (PCA) and hierarchical clustering analysis revealed a clear distinction between the proteomes of tumor and adjacent tissues (Figs. 3A, S5A). PCA distances among tumor tissues were significantly lower than those among tumor adjacent tissues, corroborating tumor heterogeneity. In total, 1,995 differentially expressed proteins (DEPs) were identified in tumor tissues compared with adjacent tissues (Benjamini-Hochberg-adjusted  $p < 0.05$ , t test, FC > 2), including 1,296 downregulated and 699 upregulated proteins (Fig. 3B, **Table S4**). Oncogene/TSG information from the OncoKB database<sup>27</sup> was used to determine whether predicted effects would manifest in the tumor based on the observed protein

expression alternations of these genes. We identified 13 significantly upregulated oncogene products (AXL, BCL2, CCND1, CDK4, CDK6, DNMT1, EGFR, FLT1, LCK, LYN, NOTCH3, PIK3CD, and RBM15) and 10 significantly downregulated TSG products (ATP6V1B2, CDH1, EPCAM, FH, PTPRD, PTPRS, SDHA, SDHB, SDHC, and SFRP1) in the tumors (**Figure S5B**), among which 8 out of the 13 upregulated oncogene products and all 10 downregulated TSG products were also identified in the CPTAC cohort. AXL, BCL2, CCND1, LCK, and NOTCH3 were significantly upregulated in the Chinese cohort, but not in the CPTAC cohort, revealing the similarities and distinctions in the carcinogenesis of ccRCC between the Chinese and CPTAC cohorts. Notably, the higher protein expression levels of three oncogene products (DNMT1, LYN, and NOTCH3) and lower protein expression levels of two TSG products (EPCAM, SDHC), were related to poor prognosis (log-rank test,  $p < 0.05$ , **Figure S5C**).

To paint a full picture of cellular process alterations in ccRCC, we performed pathway enrichment analysis. Tumor-upregulated proteins significantly converged on pathways including glycolysis/gluconeogenesis (e.g., HK2, PFKP, ALDOA, PGK1), DNA replication (e.g., MCM2, POLD1, POLD4), extracellular matrix (e.g., CD36, COL2A1, LAMA4, THBS1), NF- $\kappa$ B signaling (e.g., BCL2, LCK, LYN, NFKB2), HIF-1 signaling (e.g., EGFR, FLT1, HK2, PDK1, PIK3CD), PI3K-AKT signaling (e.g., GYS2, ITGA5, PDK1, PIK3CD, TCR2) and chemokine signaling (e.g., CCR5, HCK, LYN, STAT1, STAT2). Proteins overrepresented in normal tissues were mainly involved in pathways related to kidney functions (renin-angiotensin system, proximal tubule bicarbonate reclamation), PPAR signaling (e.g., ACOX2, FABP3, CPT2, EHHADH, RXRA), the citrate cycle (e.g., CS, FH, IDH2, ACO2, SDHA), oxidative phosphorylation (e.g., ATPs, NDUFABs, NDs), pyruvate metabolism (e.g., LDHB, PCK2, DLAT), biosynthesis of amino acids (e.g., DDC, AFMID, PSAT1, GLS), and fatty acid metabolism (e.g., HADH, EHHADH, ACAT1) (Fig. 3C). Representative DEPs in these pathways are shown in Fig. 3D. Among these, GYS2, ITGA5, TLR2, STAT2, LYN, HCK, ACOX2, FABP3, CPT2, EHHADH, and ACE2 were associated with clinical outcomes (log-rank test,  $p < 0.05$ , Fig. 3E).

ccRCC is suggested to arise from epithelial cells of the proximal tubules<sup>21,28</sup>. We observed that 17 kidney-specific proteins annotated in the Human Protein Atlas<sup>29</sup> were significantly downregulated (Benjamini-Hochberg-adjusted  $p < 0.05$ ,  $FC > 2$ ) in tumor tissues, including three glomerulus-specific proteins (FGF1, NPHS1, NPHS2), two collecting duct-specific proteins (AQP2, ATP6V0D2), four distal tubule-specific proteins (CALB1, UMOD, SLC12A1, and SLC12A3), and eight proximal tubule-specific proteins (GGT1, BHMT, LRP2, DPYS, AGMAT, SLC22A8, SLC22A13, and HRSP12) (Figs. 3F, S5D). In addition, patients with lower expression of proximal tubule-specific proteins, GGT1 and BHMT in tumors appeared to have poorer clinical outcomes (log-rank test,  $p < 0.05$ ) (Fig. 3G). These results suggested that loss of kidney tissue identity and gain of oncogenic characteristics, such as cell cycle and ECM, are important features of ccRCC, indicating the malignancy of this disease.

### **Dysregulation of Metabolic Bioprocesses in ccRCC**

Kidney is a metabolic organ. As ccRCC is characterized by aberrant metabolic pathways that control energetics and biosynthesis, it is important to learn how metabolic bioprocesses are altered at the proteome level in ccRCC. To this end, we surveyed metabolism-related pathways annotated in KEGG<sup>30</sup>

and Reactome<sup>31</sup> (**Table S5**). Glycogen metabolism, and glycolysis were upregulated in tumor tissues. In contrast, most metabolic pathways, including tricarboxylic acid (TCA) cycle, oxidative phosphorylation (OXPHOS), amino acid metabolism, lipids metabolism, one-carbon metabolism, and metabolism of vitamins and cofactors were downregulated (Fig. 4A). Correspondingly, only 149 metabolism-related proteins were upregulated in tumor tissues, comparing with 502 downregulated metabolism-related proteins (Benjamini–Hochberg-adjusted  $p < 0.05$ ,  $FC > 2$ ). In contrast, ccRCC is characterized by downregulation of most metabolic bioprocesses<sup>32,33</sup>.

In normal cells, glucose is a major source of pyruvate, which feeds the TCA for energy production under normoxia. In contrast, cancer cells predominantly produce energy by lactic acid fermentation under hypoxia<sup>34</sup>. We observed that the expression of the glucose transporter GLUT1 (SLC2A1) was increased in ccRCC tumors vs. adjacent tissues ( $T/TA = 9.49$ ), suggesting increased glucose uptake. Moreover, glycolytic enzymes, including HK2 ( $T/TA = 36.69$ ), PFKP ( $T/TA = 15.97$ ), PKM ( $T/TA = 4.08$ ), and LDHA ( $T/TA = 8.58$ ) were upregulated in ccRCC, indicating increased glucose utilization for lactate fermentation. Further, pyruvate carboxylase (PC), which converts pyruvate to oxaloacetate in the TCA cycle, was significantly decreased in ccRCC ( $T/TA = 0.09$ ) (**Figure S6A**), corroborating the increased glucose flux into lactate fermentation. In ccRCC tissues, the TCA cycle (CS, ACO2, IDH2, OGDH, SUCLG2, SDHA, FH, and MDH2) and OXPHOS (COX6C, ATPs, NDUFBs, NDs) were downregulated. Notably, the upregulation of HK2 and downregulation of COX6C in tumors were associated with poor prognosis (log-rank test,  $p < 0.05$ ) (Fig. 4B). In conclusion, we observed a significant Warburg effect in ccRCC, which enables tumor cells to rapidly proliferate and survive under nutrient depletion and hypoxia<sup>34</sup>.

Fatty acid metabolism mainly involves  $\beta$ -oxidation and fatty acid synthesis<sup>3535393939393939</sup>. Our proteome data showed decreased levels of enzymes involved in  $\beta$ -oxidation (ECHS1, HADH, HADHA, EHHADH, ACAT1) in ccRCC. Conversely, levels of fatty acid synthesis-related enzymes (ACLY, ACACA, FASN) were increased in ccRCC (**Figure S6A**). This imbalance indicated lipid accumulation, which may determine the characteristic phenotypic feature of ccRCC, the clear cell morphology. Furthermore, the upregulation of FASN ( $T/TA = 2.03$ ) and downregulation of ECHS1 ( $T/TA = 0.07$ ) in tumors were associated with poor prognosis (log-rank test,  $p < 0.05$ ) (Fig. 4B). Consistent with our previous study, decreased ECHS1 was associated with fatty acid accumulation and poor prognosis in ccRCC<sup>36</sup>.

Dysregulation of amino acid metabolism was another feature of ccRCC. The enzymes AFMID ( $T/TA = 0.18$ ), KMO ( $T/TA = 0.22$ ), KYNU ( $T/TA = 0.46$ ), HAAO ( $T/TA = 0.45$ ), and QPRT ( $T/TA = 0.21$ ), which participate in tryptophan (Trp) catabolism, were downregulated. The tryptamine pathway via dopa decarboxylase (DDC) ( $T/TA = 0.08$ ), an alternative Trp catabolic pathway, was also downregulated in tumor tissues (**Figure S6A**). In addition, PSAT1 ( $T/TA = 0.09$ ), PSPH ( $T/TA = 0.48$ ), and GLS ( $T/TA = 0.35$ ), GLUD1 ( $FC = 0.24$ ), which participate in serine and glutamine metabolism, were downregulated in ccRCC (**Figure S6A**). Downregulation of AFMID, AOX1, and DDC in tumors were associated with poor prognosis (log-rank test,  $p < 0.05$ ) (**Figure S6B**).



One-carbon metabolism, which supports various bioprocesses, including nucleotide biosynthesis, amino acid homeostasis, epigenetic maintenance, and redox defense, plays central roles in carcinogenesis and tumor progression<sup>37</sup>. Loss of SHMT1 (T/TA = 0.19) and ALDH1L1 (T/TA = 0.22) attenuated formate clearance. Overexpression of MTHFD1L (T/TA = 4.37) and MTHFD2 (T/TA = 6.89, identified in 52 samples) in turn resulted in increased one-carbon flux to generate excess formate (Fig. 4C). NNMT (T/TA = 17.37) and DNMT1 (T/TA = 3.97), upstream enzymes of homocysteine (Hcy) metabolism, were overexpressed in tumors, leading to enhanced Hcy generation. Hcy can be removed through catabolic processes via different enzymes, such as methionine synthase (MTR), betaine-Hcy S-methyltransferase (BHMT, T/TA = 0.20; BHMT2, T/TA = 0.17), and cystathionine-beta-synthase<sup>38</sup>. In tumor tissues, we observed impairment of the cytosolic one-carbon cycle (ALDH1L, MTHFD1, and MTHFR) (Fig. 4C), limiting generation of CH<sub>3</sub>-THF, coenzyme of MTR. Thus, enhanced production and diminished removal of Hcy resulted in Hcy accumulation in ccRCC. To verify this hypothesis, we examined the levels of Hcy metabolites in tumor and adjacent tissues. Hcy was 2.7-fold more concentrated in ccRCC tumors than in adjacent tissues (paired t test,  $p < 0.0001$ ) (Fig. 4D). By survey the CPTAC data<sup>3</sup>, we found that uncoupling of mRNA and protein level was not only observed in OXPHOS, but also in one-carbon metabolism (**Figure S6C**). More excitingly, by investigating the enzymes involved in formate metabolism, we found that patients with higher expression of ALDH1L1 and SHMT1 appeared to have better prognostic outcomes, whereas patients with higher expression of MTHFD1L appeared to have poorer prognostic outcomes (log-rank test,  $p < 0.05$ , Fig. 4E). As for Hcy metabolism, higher expression of NNMT and DNMT1 was negatively associated with good prognosis, whereas higher expression of BHMT and BHMT2 was positively associated with good prognosis (log-rank test,  $p < 0.05$ , Fig. 4F). Our study interrogated concrete protein expression alterations in one-carbon metabolism in ccRCC, highlighting the significance of one-carbon metabolism dysregulation during ccRCC pathogenesis.

### Proteomic Subtypes of Chinese ccRCC

TNM staging of ccRCC reflects tumor size, position, lymph node involvement, and metastasis. ISUP grading of ccRCC is based on tumor differentiation and morphology<sup>39</sup>. Both TNM staging and ISUP grading were associated with clinical outcomes in this cohort (**Figures S7A-B**). We conducted separate differential analyses to reveal proteomic differences in different stages and grades. Compared to early-stage (TNM stages I&II), advanced-stage (TNM stages III&IV) ccRCC displayed elevated expression of complement and coagulation cascades (FGA, PLG), neutrophil degranulation (CEACAM8, CD177), membrane trafficking (KIF2A, SRC), and translation (EIF4E, EEF1A1) (**Figure S7C, Table S6**). Consistently, complement and coagulation cascades, neutrophil degranulation, membrane trafficking, and translation were upregulated in high-grade vs. low-grade tumors (**Figure S7D, Table S6**).

Given the inter-tumoral heterogeneity, it is important to perform molecular subtyping. We employed consensus clustering<sup>40</sup> to identify ccRCC proteomic subtypes. ccRCC was classified into three subtypes, GP1, GP2 and GP3, comprising 55, 99, and 78 cases, respectively (Figs. 5A, S7E-F). Remarkably, the proteomic subtypes significantly differed in OS (log-rank test,  $p < 0.001$ , Figs. 5B-C) and progression-free survival (log-rank test,  $p < 0.001$ , **Figures S7G-H**). Among the three subtypes, GP1 showed a particularly

high mortality risk (HR = 7.8; 95% CI, 4.33–14.1;  $p = 9.23e-12$ , **Table S1**) and 80% GP1 cases eventually developed progressive disease (**Figure S7G**). The results of multivariable analysis after adjusting for TNM stage and ISUP grade authenticated subtype GP1 as an independent prognostic factor (HR, 3.15; 95% CI, 1.59–6.21;  $p = 9.6e-4$ ; **Table S1**). After patient stratification according to TNM stage, proteomics subtypes were still significantly associated with patient survival, regardless of tumor stage (log-rank test,  $p < 0.0001$ ), supporting the superior prognostic power of our proteomic subtyping (Figs. 5C, S7H). To confirm the association between proteomic subtyping and clinical outcomes, we extracted the 20 most representative proteins of each proteome subtype in the Chinese ccRCC cohort to classify CPTAC data<sup>3</sup> into three subtypes (**Figures S8A-B; STAR Methods**). Consistently, CPTAC-GP1 was significantly associated with poorer survival than CPTAC-GP2 and CPTAC-GP3 (log-rank test,  $p = 0.001$ , **Figure S8C**), in line with the survival differences among the three subtypes in the Chinese cohort (Fig. 5B), indicating the robustness of proteomic subtyping based on the Chinese ccRCC cohort.

When we separately decomposed the mutation spectra of the three proteome subtypes, we found that 7 patients in GP1 contained SBS23 (Fig. 5D), which was not found in the overall mutation signatures. Significant arm-level CNA events varied among the three subtypes, reflecting profound genomic effects on the ccRCC proteome. Multiple arm-level CNA events, including 8p loss, 9p loss, 9q loss, 18q loss, 12q gain, and 20q gain, were aggregated in GP1 (Fisher's exact test,  $p < 0.05$ , Figs. 5A, E), accounting for 37.3%, 47.1%, 49.0%, 25.5%, 25.4%, and 31.4% of GP1 respectively. Among these, 9p loss, 9q loss, and 12q gain were risk factors for survival (Fig. 2B). *BAP1* and *ABAC13* mutations, which were associated with poor survival (**Figures S3C-D**) occurred more frequently in GP1, accounting for 23.5% and 9.8% of GP1, respectively (Fig. 5F). In contrast, *PBRM1* mutations, mutually exclusive *BAP1* mutations, were enriched in GP2 (Fisher's exact test,  $p = 0.038$ , Fig. 5F).

We used ESTIMATE<sup>41</sup> to deconvolute tumor microenvironment (TME) compositions (**Figure S9A**) and conducted overrepresentation analysis of elevated proteins in each subtype (Fig. 5G). In total, 641, 1,838, and 97 proteins were upregulated in GP1, GP2, and GP3, respectively (**Table S7**). GP1 was characterized by a high degree of immune infiltration (Kruskal–Wallis test,  $p < 0.0001$ , **Figure S9A**), as indicated by the enrichment of multiple immune-associated pathways, including innate immune system, complement and coagulation cascades, antigen processing-cross presentation, interferon signaling, and T cell receptor (TCR) signaling ( $q < 0.05$ , Fig. 5G). Consistently, GP1 had the highest immunosuppression, CD 8 cluster, and MHC I antigen-presenting machinery (APM) scores (Kruskal–Wallis test,  $p < 0.05$ , **Figure S9B**). GP2 displayed high tumor purities (Kruskal–Wallis test,  $p < 0.0001$ , **Figure S9A**) and increased metabolism-related pathways, including the TCA cycle and respiratory chain, amino acid metabolism, mitochondrial translation, lipid metabolism, and glycolysis/gluconeogenesis ( $q < 0.05$ , Fig. 5G). GP3 featured the highest stromal scores (Kruskal–Wallis test,  $p < 0.0001$ , **Figure S9A**), corresponding to upregulation of ECM-related pathways, including ECM organization, collagen formation, elastic fiber formation, and focal adhesion ( $q < 0.05$ , Fig. 5G). Classification of our proteome data according to established CPTAC subtyping signatures<sup>3</sup> provided further support for the diverse characteristics of the proteome subtypes in the Chinese ccRCC cohort (**STAR Methods**). Specifically, GP1 were mainly CD8<sup>+</sup> inflamed tumors, GP2

were mainly of the metabolic immune-desert subtype, and GP3 were mainly CD8<sup>-</sup> inflamed tumors (**Figure S9C**). The allocation of CPTAC subtypes in our data was also indicated by the immune and stromal scores (**Figures S9A, C**).

Given that GP1 represented the most aggressive ccRCC subtype, we investigated detailed molecular expression patterns of GP1 by comparing proteins upregulated in GP1 with those in GP2 and GP3 (FC > 2, p < 0.05, Fig. 5H). The molecular characteristics of GP1 tumors were summarized into three categories: immunity, metastasis, and metabolic dysregulation. Specifically, proteins involved in TCR signaling (e.g., CD8A, ZAP70, LCK, SYK, PTPRC) and macrophage signatures (e.g., ITGAM, MSR1, FCGR1A, TLR2) were significantly upregulated in GP1. CD163, a marker of tumor-associated macrophages<sup>42,43</sup>, was also elevated in GP1 (FC = 2.55, p = 0.001). Additionally, inflammasome components (PYCARD, CASP1, CASP4, NAIP, NLRC4, IL18) and inflammation-related molecules (CRP, ALOX5) were highly expressed in GP1. Correspondingly, GP1 had the highest immunosuppression scores among the three subtypes (Kruskal–Wallis test, p < 0.0001, **Figure S9B**). The metastatic potential of GP1 was reflected by the upregulation of ECM remodeling (e.g., MMP9, SPARC, PLOD3, P4HB) and cytoskeleton rearrangement (e.g., VTN, VIM, CFL1, ARPC2). Moreover, GP1 displayed increased expression of angiogenic features (e.g., ANGPTL4, EGFR, AAMP) and complement and coagulation cascade components (e.g., C1, C7, C8, C9, PLG, SERPINE1), which have been associated with aggravated tumor invasion<sup>44,45</sup>. Metabolic dysregulation, including prominent overactivation of the oxidative pentose phosphate (OxPPP) pathway (G6PD, PGD) and dysregulation of Hcy metabolism (NNMT, DNMT1), was another characteristic of GP1 (Fig. 5H).

Since GP1 patients had the poorest prognosis and were supposed to be assigned into a clinically high-risk category, they deserved to receive further therapy. To determine the potential therapeutic drug targets, we mapped GP1-representative proteins to Drugbank<sup>46</sup> druggable proteins (Fig. 5I, **Table S7**). Ten candidates, involved in metabolism (IL4I1, NNMT), immunity (C1QC, HM13, CASP1), and metastasis (P4HA1, P4HB, PLOD3, S100A4, PML), were identified. In brief, we identified three novel proteomic subtypes of Chinese ccRCC with distinct molecular features that connect the proteomic, genomic, and clinical features of ccRCC.

### **NNMT Promotes Cancer Cell Proliferation Through Hcy Accumulation**

We conducted supervised analysis to identify robust and representative prognostic proteins, and we anticipated to screen out drug targets (Fig. 5I). NNMT, an important enzyme in Hcy metabolism, was overexpressed in ccRCC tumors (Fig. 6A) and significantly associated with poor prognosis (Fig. 6B). Furthermore, western blotting (Fig. 6C) and immunohistochemistry (IHC) (Figs. 6D, S10) confirmed that NNMT was overexpressed in ccRCC. Historically, ccRCC has been considered resistant to conventional chemo- and radiotherapy, indicating tolerance to genotoxic stress. Moreover, ccRCC cells are able to proliferate rapidly in a nutrient-depleted microenvironment<sup>34</sup>. Thus, we tested whether high NNMT expression increased the viability of ccRCC under various stresses. NNMT overexpression promoted the proliferation of ACHN and 786-O cells and profoundly enhanced cell proliferation during nutritional stress or genotoxic stress (Figs. 6E-F). Nutritional and genotoxic stresses induce DNA damage in cells. NNMT

overexpression reduced DNA damage in stressed ACHN, 786-O, and 769-P cells as evidenced by the levels of  $\gamma$ -H<sub>2</sub>AX detected using immunofluorescence staining (Fig. 6G) and western blotting (Fig. 6H) and by DNA damage detection using the comet assay (Fig. 6I), in cultured renal cancer cell lines. These results indicated that NNMT overexpression may contribute to proliferation promotion under stress. NNMT catalyzes methyl transfer from S-adenosyl methionine (SAM) to nicotinamide (NAM) and generates S-adenosyl homocysteine (SAH) and 1-methylenicotinamide (1MNA). Increased NNMT in cells resulted in a decrease in SAM and increases in SAH and 1MNA (Fig. 6J). The level of Hcy, the hydrolysis product of SAH, was also increased significantly (Fig. 6J). Supplementation of SAH or Hcy, but not supplementation of 1MNA or a reduction in SAM through knockdown of MAT, reduced DNA damage (Fig. 6K) and promoted cell proliferation under stress (Fig. 6L). Furthermore, blockade of SAH hydrolysis by knockdown of S-adenosylhomocysteine hydrolase (SAHH) in NNMT-overexpressing cells abrogated the DNA damage-reducing effect of NNMT (Figs. 6M-N), suggesting that Hcy, but not SAH, plays a role in the DNA repair- and proliferation-promoting effects of NNMT.

### Lysine Homocysteinylation of DNA-PKcs Enhances DNA Repair

When present at high levels, intracellular Hcy modifies protein lysine residues, which results in protein lysine-homocysteinylation in cells<sup>47</sup>. We observed increased protein lysine-homocysteinylation (K-Hcy) levels in NNMT-overexpressing (Fig. 7A) and SAH-supplemented cultured renal cancer-derived ACHN, 786-O, 769-P, and A-498 cells (Fig. 7B). Meanwhile, Hcy (Fig. 4D) and K-Hcy (Fig. 6C) levels were increased in ccRCC tumors compared to adjacent tissues. In our previous cell-wide proteomics screen for K-Hcy substrates in HEK293T cells, we observed that DNA-PKcs, a protein required for the non-homologous end-joining pathway of DNA repair, was heavily modified by K-Hcy<sup>47,48</sup>. In ccRCC tumors, we validated that three different lysine residues (K122, K712, K868, and K902) were modified by K-Hcy (Fig. 7C), suggesting that K-Hcy regulates DNA-PKcs-mediated DNA repair. Among the three lysine residues in DNA-PKcs (K122, K712, K868, and K902) that were modified by homocysteinylation, K122 is located within the interface between DNA-PKcs and KU70/KU80, while K712, K868, and K902 are located within the intramolecular interaction region of DNA-PKcs<sup>49</sup> (Fig. 7C). Proteins that physically interact with methionyl-tRNA synthetase (MARS) are more prone to being modified and regulated by K-Hcy<sup>47</sup>. Accordingly, an interaction between DNA-PKcs and MARS was confirmed by co-immunoprecipitation assays using either exogenous DNA-PKcs and MARS in ACHN and 769-P cells (**Figure S11A**), or endogenous DNA-PKcs and MARS in 786-O cells (Fig. 7D). Elevated NNMT, SAH, Hcy, or MARS levels led to dose-dependent increases in DNA-PKcs homocysteinylation in ACHN and 769-P cells (**Figures S11B-E**). These results confirmed that DNA-PKcs is subject to MARS-mediated K-Hcy modification. Increased NNMT expression resulted in the activation of DNA-PKcs as indicated by increased phosphorylation of DNA-PKcs and its downstream target protein p53 (at Ser15) in ACHN and 769-P cells (Fig. 7E). Moreover, supplementation of either SAH or Hcy or overexpression of MARS induced elevated K-Hcy level of DNA-PKcs and activated the DNA-PKcs pathway in ACHN and 769-P cells (**Figures S11F-H**). In contrast, reducing K-Hcy modification through knockdown of NNMT, SAHH, or MARS inhibited DNA-PKcs activity in ACHN and 769-P cells (**Figures S11I-K**). Furthermore, phosphorylation levels of DNA-PKcs and p53 were markedly increased in ccRCC tumors vs. adjacent tissues (Fig. 6C). These results were consistent with the comet assay results, which revealed

that compared with adjacent tissues, tumors exhibited decreased DNA damage (Fig. 7F) and reduced  $\gamma$ -H<sub>2</sub>AX levels (Fig. 6C). The results also indicated that DNA-PKcs is inactivated in ccRCC tumors and that NNMT-induced hyper lysine-homocysteinylation might promote ccRCC by activating DNA-PKcs.

### **Lysine-homocysteinylation facilitates the formation of DNA-PKcs complex**

We next investigated the mechanism by which lysine-homocysteinylation activates DNA-PKcs. DNA-PKcs–KU70/KU80 interaction was significantly enhanced by increased cellular K-Hcy levels induced by NNMT overexpression in ACHN and 769-P cells (Figs. 7G, H). As a result, the activity of DNA-PKcs was enhanced in NNMT-overexpressing cells as determined by monitoring its kinase activity in phosphorylating its substrate p53 *in vitro* (Fig. 7I) and measuring ADP formation in an ADP-Glo-DNA-PK assay (Fig. 7J). We validated that at increased levels, K-Hcy activates DNA-PKcs, as determined by adding homocysteine thiolactone (HTL) to the *in vitro* DNA-PKcs assay (Fig. 7K) and measuring intracellular ADP formation in an ADP-Glo-DNA-PK assay (Fig. 7L). Moreover, to mimic the bulky side chain effects of K-Hcy<sup>47</sup>, we created two mutant DNA-PKcs constructs in which either Lys122 within the KU70/KU80-binding interface or all three modifiable lysine residues were mutated to tryptophan (“KW” and “4KW” constructs, respectively). Relative to wild-type DNA-PKcs, the mutant DNA-PKcs showed increased binding affinity to KU70/KU80 (Figs. 7M, N) and enhanced DNA-PKcs activity, as determined by an *in vitro* DNA-PKcs assay (Fig. 7O). In addition, increased NNMT expression in 786-O and ACHN promoted xenograft tumor growth in nude mice, especially in IR-treated cell xenografts, whereas inhibition of K-Hcy by intraperitoneal injection of N-acetyl-cysteine (NAC)<sup>47,48</sup> delayed xenograft growth (Figs. 7P, Q). Together, these results confirmed that NNMT upregulation induces hyper K-Hcy, which activates DNA-PKcs and promotes tumor growth (Fig. 7R).

## **Discussion**

Our comprehensive proteogenomic study in 232 ccRCC tumor and tumor adjacent tissue pairs of Chinese patients provided new insights into ccRCC protein profiles and biology. The genomic profile of ccRCC revealed the somatic mutations and CNAs in Chinese ccRCC patients. Comparison of genome alterations in Chinese and Western ccRCC cohorts emphasized the genetic diversity across geographic regions and revealed features of Chinese ccRCC. Loss of chromosome 3p has long been regarded the initial event of ccRCC<sup>20,21</sup>. Our study demonstrated that copy-neutral and low-degree 3p loss might be a novel predictor of poor prognosis in ccRCC. The degree of 3p loss was associated with the frequencies of driver mutations in *BAP1* and *PBRM1*. Further studies to dissect out the association of chromosome 3p copy number alterations and driven mutations will promisingly promotes our cognations for the carcinogenesis and development of this disease.

To provide metabolic insights into ccRCC, we survey the expression of metabolism-related proteins in the 232 ccRCC tumor/adjacent tissue pairs. The results revealed three major metabolic imbalances in ccRCC, including energy metabolism, lipid metabolism and one-carbon metabolism. Concretely, the Warburg effect, manifested by upregulated glucose uptake, glycolysis, a downregulated TCA cycle, OXPHOS, led to energy metabolic imbalance. Lipid metabolism imbalance manifested as increased fatty acid synthesis

and decreased  $\beta$ -oxidation, suggesting lipid accumulation in ccRCC. One-carbon metabolism imbalance in ccRCC resulted in the accumulation of two oncometabolites, formate and Hcy. In particular, we detected Hcy accumulation in tumor and adjacent tissues, confirming our finding based on proteomic data. Overall, metabolic alterations were not only ccRCC features, but were also associated with advanced disease and poor clinical outcomes. Metabolic reprogramming facilitates the identification of novel and repurposed drugs that could potentially be used to treat ccRCC.

Based on the proteome profiles of tumor tissues, we conducted molecular subtyping of ccRCC, which revealed three subtypes. The subtypes exhibited dramatic diversity in proteomic signatures, genetic alterations, and patient survival. Among the three subtypes, GP1 was associated with the poorest survival, and 80% cases of GP1 finally progressed. GP1 exhibited a dominant immune signature, with the highest CD8<sup>+</sup> T cell infiltration and immunosuppression scores (**Figure S9B**), indicating adaptive immune resistance. Accordingly, GP1 showed higher APM scores (**Figure S9B**), which reportedly are associated with the immunogenicity of ccRCC tumors<sup>50</sup>. Therefore, we hypothesized that GP1 patients might benefit from immune checkpoint inhibitor therapy.

For the targeted therapeutic strategy, we paid more attention to GP1 patients because they had the poorest prognosis. Ten druggable candidates were screened, including IL4I1, NNMT, C1QC, HM13, CASP1, P4HA1, P4HB, PLOD3, S100A4 and PML. NNMT, a metabolic enzyme, was identified as an important carcinogenic factor and drug target. Some clinical studies using the candidate protein survey strategy<sup>51–55</sup> have reported that NNMT is overexpressed in various tumors, including lung, liver, bladder, colon, and kidney cancers. One proteomics study revealed NNMT as a master metabolic regulator of cancer-associated fibroblasts<sup>56</sup>. In cultured cells, NNMT promotes cancer cell survival, proliferation, migration, and invasion<sup>51–55</sup>. However, the exact oncogenic role of NNMT in ccRCC as well as its metabolic functions in cancer cells have not been determined.

Our recent study demonstrated that Hcy can modify protein lysine residues and turn the metabolic status to cell signaling in colorectal cancer<sup>47</sup>. As NNMT is an upstream metabolic enzyme in Hcy metabolism, we linked NNMT and Hcy in ccRCC tumorigenesis and development. Mechanistically, NNMT overexpression increases Hcy and K-Hcy modification in tumor cells and promoted tumor proliferation. K-Hcy modification of DNA-PKcs, enhancing DNA-PKcs–KU70/KU80 interaction, and finally activates DNA-PK complex. Xenograft experiments revealed that NNMT overexpression empowered the resistance to radiation therapy in renal cell carcinoma. Inhibiting K-Hcy modification by NAC rescued the injuring effect of radiation on tumor. The current study suggested that the NNMT-K-Hcy-DNA-PKcs axis can partially explain the radiotherapy resistance of ccRCC and be considered a potential therapeutic target.

In summary, our study provided a comprehensive proteogenomic landscape of Chinese ccRCC. The dominant pathways that were altered in the ccRCC proteome subtypes revealed the potential molecular mechanism underlying clinical phenotypes and outcomes. We identified a potential druggable protein, NNMT, and demonstrated the value of this multiomics approach. We believe that this study provides valuable information regarding ccRCC biology and paves the way to novel therapeutic strategies.

# Methods

## *Clinical Sample Collection*

We screened 1,556 consecutive patients who underwent radical or partial nephrectomy for the treatment of renal tumors at the Department of Urology of Fudan University Shanghai Cancer Center (FUSCC, Shanghai, China) from January 2007 to March 2014. Electronic medical records were screened retrospectively. In total, 232 eligible ccRCC patients who had undergone radical nephrectomy at the FUSCC were consecutively enrolled. Median follow-up was 85 months (range, 3–138 months). At the last follow-up, 79 patients (34.1%) had progressive disease and 49 patients (21.1%) had died of ccRCC. Clinicopathological indicators, including age at surgery, sex, clinical manifestation, laterality, tumor size, chronic diseases status, TNM stage, and ISUP grading classification are summarized in **Table S1**. Tumor and adjacent non-tumor tissue samples were collected during surgery and are available from the FUSCC tissue bank. Samples were collected according to following criteria: 1) tumor adjacent tissues were collected >2cm from the tumor margin; 2) each tumor/adjacent sample was checked by an expert pathologist to confirm the sample quality. Hematoxylin and eosin (H&E)-stained slides of tumor and tumor adjacent tissues were uploaded to the Mendeley data (<https://data.mendeley.com/datasets/pb5tbs2by5/draft?a=1b8aa955-40a7-4df0-970a-e936666ffd99>).

Among the 1,324 excluded patients, 161 patients were diagnosed with benign renal tumor, 118 with urinary tract carcinoma, 326 with non-clear RCC, and 89 with other simultaneous or heterochronous malignancies. Further, 577 patients (mainly those who underwent partial nephrectomy) were excluded because of unavailable adjacent normal tissues, and 53 samples failed to pass pathological quality check, such as tumor cell rate < 90% (**Figure S1A**). All cases were staged according to the 2010 American Joint Committee on Cancer TNM staging system. H&E-stained sections were reviewed by an experienced genitourinary pathologist to determine the ISUP grade, and frozen sections were reviewed to determine the tumor cell rate of the ccRCC tissues. The study was compliant with the ethical standards of Helsinki Declaration II and was approved by the institutional review board of FUSCC (050432-4-1212B). Written informed consent was obtained from each patient before any study-specific investigation was conducted.

## *DNA Extraction and WES*

WES was conducted at Life Healthcare Clinical Laboratory (China). DNA isolated from fresh or frozen tumor tissue samples was used for WES, and matched germline DNA was obtained from adjacent non-tumor tissue samples. DNA was isolated from fresh tissues using DNeasy Blood & Tissue Kit (Qiagen, 69504) according to the manufacturer's instructions. Purified DNA was quantified using a Qubit 3.0 Fluorometer (Life Technologies). For matched germline and tumor tissues, 100 ng of DNA was sheared to 200–300-bp fragments using a Covaris M220 system. Tumor and matched germline DNA libraries were constructed using Accel-NGS 2S HYB DNA LIBRARY KIT (Swift Biosciences, 23096) and Accel-NGS 2S MID S1-S4 (Swift Biosciences, 279384). xGen Exome Research Panel v1.0 (IDT, 1056115) and xGen Lockdown reagents (IDT, 1072281) were used for exome enrichment. Dynabeads M-270 Streptavidin (Thermo, 65306) was used for library purification, P5/P7 primers (Nanodigmbio, ND10010) and HotStart

ReadyMix (KAPA, KK2612) were used for library amplification. The amplified libraries were purified using SPRISELECT (Beckman, B23319). DNA quality was assessed using a Bioanalyzer High Sensitivity DNA Analysis kit (Agilent Technologies, 5067-4626). Samples underwent paired-end sequencing on a Nextseq CN500 platform (Illumina), with a 150-bp read length. The WES target region was 33 M. A mean coverage of 100×, a capture rate of 95%, and a dup rate of 40% were achieved for tumor sequencing.

### ***Somatic Variant Detection***

Read-depth statistics were calculated using the DepthOfCoverage function in the Genome Analysis Toolkit (GATK v3.8.1.0)<sup>57</sup>. Paired-end reads in Fastq format were aligned to a reference human genome<sup>58</sup> (UCSC Genome Browser, hg38) using Burrows-Wheeler Aligner. Variant calling was conducted following GATK best practices. Somatic single-nucleotide variations and small insertions and deletions were detected using MuTect2 (GATK v4.1.2.0) and were annotated using ANNOVAR<sup>59</sup> based on UCSC known genes. Two longest genes, *TTN* and *MUC16*, were excluded as they tended to acquire numerous mutations by chance in large-scale genome/exome sequencing experiments. The Maftools R package<sup>60</sup> was used to display mutant genes with non-synonymous mutations. MutSigCV<sup>61</sup> was used to identify significantly mutated genes with default parameters. Genes with Benjamini–Hochberg-adjusted  $p < 0.01$  were identified as significantly mutated genes.

### ***Mutation Frequency Variances Across Regions***

TCGA ccRCC genome data were downloaded from xenabrowser.net<sup>62</sup> and data for a European ccRCC cohort were obtained from<sup>5</sup>. The top 10 most frequently mutated genes in our Chinese cohort and these two cohorts were compared using Fisher's exact test.

### ***Mutual Exclusivity and Mutation Co-occurrence Analysis***

Mutually exclusive or co-occurring sets of genes were detected using the somaticInteractions function in the Maftools R package, using pair-wise Fisher's exact test to detect significant gene pairs.  $p < 0.05$  was used as a threshold for statistical significance.

### ***Mutational Signature***

SBSs are defined as a replacement of a certain nucleotide base. There are six possible substitutions: C>A, C>G, C>T, T>A, T>C, and T>G. Considering the nucleotide context, these SBS classes can be further expanded to 96 possible mutation types. The frequencies of the 96 mutation types were estimated for each sample. The non-negative matrix factorization algorithm of SigProfiler<sup>63</sup> was used to estimate the minimal components that could explain maximum variance among samples. *De novo* mutation signatures were decomposed using COSMIC v3<sup>14</sup>. After decomposing a matrix of the 96 substitution classes of the samples into five signatures, the contribution of each signature in each sample was estimated.



## ***CNA Calling***

CNAs were called following somatic CNA best practice, using the Calculate Target Coverage function in GATK (v4.1.2.0). We applied Genomic Identification of Significant Targets in Cancer (GISTIC2.0)<sup>18</sup> to identify significantly amplified or deleted focal-level and arm-level events, with  $q < 0.05$  considered significant. The following parameters were used: amplification threshold = 0.1; deletion threshold = 0.1; cap value = 1.5; broad length cutoff = 0.90; remove X-chromosome = 0; confidence level = 0.95; join segment size = 4; arm-level peel off = 1; maximum sample segments = 2,000; gene GISTIC = 1.

Each gene in each sample is assigned a threshold copy number that reflects the magnitude of its deletion or amplification. These are integer values ranging from  $-2$  to  $2$ , where  $0$  means no amplification or deletion of a magnitude greater than the threshold parameters described above. Amplifications are represented by positive numbers:  $1$  indicates amplification above the amplification threshold;  $2$  indicates amplification larger than the arm-level amplifications observed in the sample. Deletions are represented by negative numbers:  $-1$  indicates deletion beyond the threshold;  $-2$  indicates deletions greater than the minimum arm-level copy number observed in the sample.

## ***Protein Extraction and Trypsin Digestion***

Samples were minced and lysed in lysis buffer (8 M urea, 100 mM Tris hydrochloride, pH 8.0) containing protease and phosphatase inhibitors (Thermo Scientific) and then sonicated for 1 min (3 s on and 3 s off, amplitude 25%). The lysates were centrifuged at  $14,000 \times g$  for 10 min and supernatants were collected as whole-tissue extracts. Protein concentrations were determined by the Bradford protein assay (TaKaRa, T9310A). Extracts (100  $\mu$ g protein) were reduced with 10 mM dithiothreitol at  $56^\circ\text{C}$  for 30 min and alkylated with 10 mM iodoacetamide at room temperature in the dark for 30 min. The samples were digested with trypsin using a filter-aided sample preparation method<sup>64</sup>. Tryptic peptides were separated in a home-made reverse-phase C18 column. Peptides were eluted and separated into nine fractions using an acetonitrile gradient (6%, 9%, 12%, 15%, 18%, 21%, 25%, 30%, and 35%) at pH 10. The nine fractions were pooled into three fractions (6%+15%+25%; 9%+18%+30%; 12%+21%+35%), vacuum-dried (Concentrator Plus, Eppendorf), and analyzed by liquid chromatography tandem MS (LC-MS/MS).

## ***LC-MS/MS***

Samples were analyzed on a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific) coupled with a high-performance liquid chromatograph (EASY-nLC 1200 System, Thermo Fisher Scientific). Dried peptide samples were dissolved in solvent A (0.1% formic acid in water) and loaded onto a trap column (100  $\mu$ m  $\times$  2 cm, home-made; particle size, 3  $\mu$ m; pore size, 120  $\text{\AA}$ ; SunChrom) with a maximum pressure of 280 bar using solvent A, then separated on a home-made 150  $\mu$ m  $\times$  12 cm silica microcolumn (particle size, 1.9  $\mu$ m; pore size, 120  $\text{\AA}$ ; SunChrom) with a gradient of 5%–35% mobile phase B (acetonitrile and 0.1% formic acid) at a flow rate of 600 nL/min for 75 min. MS analysis was conducted with one full scan (300–1,400 m/z,  $R = 120,000$  at 200 m/z) at an automatic gain control target of  $3e6$  ions, followed by up to 20 data-dependent MS/MS scans with higher-energy collision dissociation (target  $5e4$  ions, max

injection time 20 ms, isolation window 1.6 m/z, normalized collision energy of 27%). Detection was done using Orbitrap (R = 7,500 at 200 m/z). Data were acquired using the Xcalibur software (Thermo Fischer Scientific).

### ***MS Platform QC and ccRCC Proteome Quality Assessment***

For QC of MS performance, tryptic digests of HEK293T cell lysates were measured as a QC standard every 2 days. The QC standard was made and run using the same method, conditions, software, and parameters as those used for ccRCC samples. Pairwise Spearman's correlation coefficients were calculated using the R package *corrplot*<sup>65</sup> for all QC runs, and the results are shown in **Figure S1B**. The average correlation coefficient among standards was 0.95, with a maximum of 0.82 and minimum of 0.99. Log10-transformed fractions of total (FOTs) for each ccRCC sample (**Figure S1C-D**) were plotted to show consistency of data quality. The Sva R package v3.34.0<sup>66</sup> was used to evaluate batch effects. We found no significant batch effect in the proteome data. Moreover, PCA plots showed that the batch effects were negligible for batch number, but significant for sample types (**Figure 3A**).

### ***Proteome Identification and Quantification***

Raw files were processed in Firmiana<sup>17</sup> and searched against the human National Center for Biotechnology Information (NCBI) RefSeq protein database (updated on 04-07-2013, 32,015 entries) using the Mascot 2.4 search engine (Matrix Science Inc). Mass tolerances were 20 ppm for precursor and 50 mmu for product ions. Up to two missed cleavages were allowed. Cysteine carbamidomethylation was set as a fixed modification and methionine N-acetylation and oxidation as variable modifications. Precursor ion score charges were limited to +2, +3, and +4. The data were also searched against a decoy database so that protein identifications were accepted at FDR of 1%. Label-free protein quantifications were calculated using a label-free, intensity-based absolute quantification (iBAQ) approach<sup>16</sup>. Match between runs<sup>67</sup> was used to improve parallelism between tumor/adjacent samples. We built a dynamic regression function based on common peptides in tumor/adjacent samples. Based on the correlation value  $R^2$ , Firmiana chooses a linear or quadratic function for regression to calculate the retention time (RT) of corresponding hidden peptides and checks the existence of the extracting ion current (XIC) based on the m/z and calculated RT. The program determines the peak area values of existing XICs. We calculated peak area values as parts of corresponding proteins. Proteins with at least 1 unique peptide with a 1% FDR at the peptide level were selected for further analysis. The FOT was used to represent the normalized abundance of a particular protein across samples. FOT was defined as a protein's iBAQ divided by the total iBAQ of all proteins identified in each sample. FOT values were multiplied by  $10^5$  for ease of presentation and missing values were assigned  $10^{-5}$  (**Table S3**).

### ***Protein and Pathway Alterations in Tumor vs. Adjacent Tissues***

PCA was conducted to visualize the separation of tumor and tumor-adjacent proteomes using the R package *factoextra* v1.0.6<sup>68</sup>. In total, 6,111 proteins identified in both >25% of tumor and tumor-adjacent

samples were used for subsequent analysis. Volcano plots were used to display DEPs in tumor and adjacent tissues by applying thresholds of fold change >2 and Benjamini–Hochberg-adjusted  $p < 0.05$ . Among the DEPs, 1,296 proteins were significantly upregulated and 699 proteins were significantly downregulated in ccRCC tumor tissues. The DEPs were then subjected to KEGG pathway enrichment analyses in DAVID<sup>69</sup>, with a  $p$  value cutoff of 0.05 (**Table S4**). Signature proteins of the nephrons (including glomerulus, proximal tubule, distal tubule and collecting duct) were obtained from the Human Protein Atlas database (<https://www.proteinatlas.org/humanproteome/tissue/kidney>).

### ***Tumor Purity, Immune, Stromal, APM, Immunosuppression, CD8 cluster, and Metabolic Pathway Scores***

Tumor purity, immune, and stromal scores were inferred using the R package ESTIMATE v1.0.11<sup>41</sup>. Although the ESTIMATE algorithm was designed to analyze transcriptome data, some studies have used it for proteome analysis<sup>3,7</sup>. The results indicate the feasibility to evaluate the engagement of each subtype of immune cells. APM, immunosuppression, and CD8 cluster signatures were obtained from previous reports<sup>50,70</sup> and computed using single-sample GSEA<sup>71</sup>. Metabolic pathway scores for 232 paired ccRCC samples were computed using the R package GSVA v1.34.0<sup>72</sup> (**Table S4**). KEGG and Reactome gene sets downloaded from the Molecular Signatures Database (MSigDB v7.1, <http://software.broadinstitute.org/gsea/msigdb/index.jsp>) were set as background.

### ***GSEA***

GSEA was conducted using the GSEA 4.0.3 software (<http://software.broadinstitute.org/gsea/index.jsp>)<sup>73</sup>. KEGG, Reactome, and HALLMARK gene sets downloaded from the MSigDB v7.1 were set as background. FDR < 0.05 was used as a cutoff. The normalized enrichment score was used to reflect the degree of pathway overrepresentation.

### ***Associations Between Clinical Characteristics and the ccRCC Proteome***

Specific clinical information is presented in **Table S1**. TNM stage- and ISUP grade-specific proteins were screened out based on a fold change > 1.5 and  $p < 0.05$ . Specific proteins of each TNM stage and ISUP grade were subjected to over-representation analysis using ConsensusPathDB (<http://cpdb.molgen.mpg.de/>)<sup>74</sup>. Clinical characteristics-associated pathways are listed in **Table S6**.

### ***Proteomic Subtyping of ccRCC, and Subtype Features***

Consensus clustering was conducted using the R package Consensus Cluster Plus<sup>75</sup> using Pearson correlation as the distance measure. The 1,000 proteins with the highest median absolute deviation in tumor samples were used for k-means clustering with up to five groups. Consensus matrices for  $k = 2, 3, 4, 5$  clusters are shown in **Figure S8E-F**. The consensus matrix for  $k = 3$  showed clear separation among clusters. The cumulative distribution function of the consensus matrix for each  $k$ -value was also measured (**Figure S8F**). The relative change in area under the cumulative distribution function curve increased by 33% from 2 clusters to 3 clusters, whereas others exhibited no appreciable increase. Thus,

proteome clusters were defined using k-means consensus clustering with  $k = 3$ . Subtype-specific upregulated proteins are: (1) detected in  $\geq 25\%$  tumor samples; (2) expressed higher than other subtypes ( $FC > 2$ ,  $t$  test  $p < 0.05$ ). Subtype-specific upregulated proteins were further analyzed in ConsensusPathDB<sup>74</sup>. DEPs of each subtype and relevant enriched pathways are listed in **Table S7**.

### ***Validation of Proteomic Subtyping Performance***

GSEA was conducted to identify signature proteins of each proteomic subtype using GSEA v4.0.3<sup>73</sup>, and the 20 proteins with the highest scores in each subtype were selected. Hierarchical clustering of CPTAC ccRCC cohort<sup>3</sup> (available follow-up is three years at present) proteome data with signature proteins also classified the CPTAC cohort into three subgroups with a similar survival curve in our population, with GP1 showing distinctly worse survival than the other two subtypes (log-rank test,  $p = 0.001$ ) (**Figure S8**).

### ***Correlations Between Subtypes and Clinical Features***

To evaluate correlations between proteomic subtypes and clinical features, Fisher's exact test was conducted on categorical variables, including driver gene mutations, significant arm-level CNA events, age, sex, hypertension status, obesity status, cardiovascular and cerebrovascular disease status, family history of cancer, TNM stage, ISUP grade, and CPTAC subtype. Only variables that varied significantly among the three proteome subtypes are shown in **Figure 5A**. Scaled CPTAC ccRCC proteome data were used to identify signature proteins of each subtype by GSEA. The 20 proteins with the highest GSEA scores were selected as support vectors to build a support vector machine classifier. Chinese ccRCC cohort was divided into four CPTAC subtype using this classifier.

### ***Effects of CNAs***

Spearman's correlations between CNA values (gene level) and protein abundances were calculated using 14,538 genes quantified at both CNA and proteome levels. CNAs with significant correlation with proteins were selected based on  $FDR < 0.01$ . In total, 89,992 CNA and protein pairs showed significant correlation. Correlations were visualized using the R package multiOmicsViz. Genomic alterations that affect gene expression at the same locus are said to act in *cis*, whereas an impact of another locus is defined as a *trans* effect (vertical patterns in **Figure 2C**), whereas the impact of other locus was defined as a *trans* effect (diagonal patterns in **Figure 2C**).

### ***Survival Analysis***

The Kaplan–Meier method was used for survival analyses, and groups were compared using the log-rank test. The R survival package 3.2-3<sup>76</sup> and survminer 0.4.8 were used for statistical tests and visualization. The HR was calculated by Cox proportional hazards regression analysis. Variates with  $p < 0.05$  were considered to significantly impact prognosis. OS was used as a primary endpoint. Clinical and molecular variates with  $p < 0.05$  in single variant analysis were selected for Cox regression multivariate analysis (**Table S1**).

## ***Drug Target Analysis***

Target proteins were selected based on three criteria: significantly upregulated in tumor vs. adjacent (Benjamini–Hochberg-adjusted  $p < 0.05$ ,  $FC > 2$ ), upregulated in GP1 compared to GP2&3 (Benjamini–Hochberg-adjusted  $p < 0.05$ ,  $FC > 2$ ), and associated with poor prognosis ( $HR > 1.8$ ,  $p < 0.05$ ). The proteins were mapped using the Drugbank druggable protein database (<https://www.drugbank.ca/>). Druggable proteins are listed in **Table S7**.

## ***Cell Culture***

Human HEK293T (ATCC, CRL-11268) A-498 (ATCC, HTB-44; RRID: CVCL\_1056) and ACHN (ATCC, CRL-1611) cells were cultured in high-glucose Dulbecco's modified Eagle's medium (DMEM; HyClone) supplemented with 10% fetal bovine serum (FBS; Invitrogen), 100 units/mL penicillin (Invitrogen), and 100  $\mu$ g/mL streptomycin (Invitrogen). 769-P (ATCC, CRL-1933) and 786-O cells (ATCC, CRL-1932) were maintained in RPMI 1640 medium (Invitrogen) containing 10% FBS. Cells were incubated in 5%  $CO_2$  at 37°C. Cells were transfected using polyethylenimine (linear, 25 KDa) or Lipofectamine 2000 (Invitrogen). To generate a cell model of nutrition stress, ACHN and 786-O cells were cultured in medium without serum and glucose for 12 h before the assays. To generate a cell model of genotoxic stress, cultured cells were irradiated with 4 Gy X-ray radiation using a linear accelerator (Oncor, Siemens) before the experiments.

## ***Plasmid Construction and Transfection***

Whole-length NNMT, MARS, KU70, KU80, and p53 cDNA clones were purchased from Origene. A whole-length DNA-PKcs cDNA clone was obtained from Prof. Yanhui Xu<sup>49</sup>. After confirming the sequences by Sanger sequencing, DNA-PKcs, NNMT, and p53 were amplified and subcloned into the NheI and EcoRI restriction sites of the pcDNA3.1-Flag vector, using ClonExpress MultiS One Step Cloning Kit (#C113-02, Vazyme). KU70 and KU80 were amplified and subcloned into the NheI and EcoRI restriction sites of the pcDNA3.1-Myc vector using the same kit. DNA-PKcs mutants were generated by site-directed mutagenesis using the MutanBEST kit (TaKaRa). The primers used were as follows: NNMT: forward, 5'- ggg aga ccc aag ctg gct agc ATG GAA TCA GGC TTC ACC TCC -3', and reverse, 5'- tag tcc agt gtg gtg gaa ttc CAG GGG TCT GCT CAG CTT CC -3'; p53: forward, 5'- ggg aga ccc aag ctg gct agc ATG GAG GAG CCG CAG TCA G -3', and reverse, 5'- tag tcc agt gtg gtg gaa ttc GTC TGA GTC AGG CCC TTC TGT C -3'; KU70: forward, 5'- ggg aga ccc aag ctg gct agc ATG TCA GGG TGG GAG TCA TAT TAC A -3', and reverse, 5'- tag tcc agt gtg gtg gaa ttc GTC CTG GAA GTG CTT GGT GAG G -3'; KU80: forward, 5'- ggg aga ccc aag ctg gct agc ATG GTG CGG TCG GGG AAT -3', and reverse, 5'- tag tcc agt gtg gtg gaa ttc TAT CAT GTC CAA TAA ATC GTC CAC A -3'; DNA-PKcs: forward, 5'- ggg aga ccc aag ctg gct agc ATG GCG GGC TCC GGA GCC G -3', and reverse, 5'- tag tcc agt gtg gtg gaa ttc CAT CCA GGG CTC CCA TCC T -3'; DNA-PKcs K122W: forward, 5'- GA GCT GCT tgg TGT AAA ATT CCA GCC CTG GAC C -3', and reverse, 5'- TTT ACA cca AGC AGC TCT ATC TTT TGT ATA AAC ACT G -3'; DNA-PKcs K712W: forward, 5'- AA TTT GGC tgg GAG GTG GCA GTT AAA ATG AAG CA -3', and reverse, 5'- CAC CTC cca GCC AAA TTT CAC AAA TAA AGC AAA -3';

DNA-PKcs K868W: forward, 5'- C tgg AAT CTT CTG ACA GTC ACG TCC TCA GAT G -3', and reverse, 5'- C TGT CAG AAG ATT cca GTT TAT TTG TCC TCC TAG AGA TCC AAG -3'; DNA-PKcs K902W: forward, 5'- A GAG ATG tgg CCT GTC ATT TTC CTG GAT GTG TT -3', and reverse, 5'- T GAC AGG cca CAT CTC TCT AAA GGG CAC TGC AA -3'. For transient transfection, 1 µg of each plasmid was transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions.

### ***RNA Interference***

Synthetic oligos were used for siRNA-mediated silencing of *NNMT* (5'-CCTCTCTGCTTGTGAATCCTT-3') and *SAHH* (5'-GTCAGGAGGGCA ACATCTTTG-3'), and scramble siRNA was used as a control. Cells were transfected with siRNAs using Lipofectamine 2000 according to the manufacturer's protocol. Knockdown efficiency was verified by qRT-PCR or western blotting.

### ***Gene Silencing and Overexpression***

For *NNMT* stable shRNA knockdown or overexpression, cells were co-transfected with pCMV-VSV-G, pCMV-Gag-Pol, and plasmids using the calcium phosphate method<sup>77</sup>. Transfected cells were cultured in DMEM containing 10% FBS for 6 h. Twenty-four hours after transfection, culture supernatant was collected and used for retrovirus preparation to infect cells at 10% confluency in 90-mm-diameter dishes. Cells were re-infected 48 h after the initial infection and selected using 5 µg/mL puromycin (Amresco). *NNMT* shRNA was cloned into the AgeI and EcoRI restriction sites of the pMKO vector. *NNMT* was subcloned into the BamHI and EcoRI restriction sites of the pBABE vector using ClonExpress MultiS One Step Cloning Kit. The sequences of primers used were as follows: *shNNMT*-Forward: 5'- CCG GCC TCT CTG CTT GTG AAT CCT TCT CGA GAA GGA TTC ACA AGC AGA GAG GTT TTT G -3', *shNNMT*-Reverse: 5'- AAT TCA AAA ACC TCT CTG CTT GTG AAT CCT TCT CGA GAA GGA TTC ACA AGC AGA GAG G -3'; *NNMT* (pBABE): forward, 5'- ctc tag gcg ccg gcc gga tcc ATG GAA TCA GGC TTC ACC TCC -3', and reverse, 5'- ggt ctt ctc gtc cat gaa ttc CAG GGG TCT GCT CAG CTT CC -3'.

### ***Western Blot Analysis***

Cultured cells or cells from human ccRCC and matched normal tissues were lysed with 0.5% NP-40 buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5% Nonidet P-40, and a mixture of protease inhibitors (Sigma-Aldrich). After centrifugation at 13,800 × *g* and 4°C for 15 min, supernatants were collected for western blotting according to standard procedures. Antibodies against DNA-PKcs (#38168), phospho-Ser2056 DNA-PKcs (#68716), H<sub>2</sub>AX (#7631), γ-H<sub>2</sub>AX (#9718), p53 (#9282), phospho-Ser15 p53 (#9284), KU70 (#4103), and KU80 (#2753) were purchased from Cell Signaling Technology. Antibody against *NNMT* was purchased from Abcam (#ab58743). Antibody against Actin was purchased from Genscript (#A00702). Anti-K-Hcy antibody was generated as described previously<sup>47</sup>. Chemiluminescence was measured on a Typhoon FLA 9500 instrument (GE Healthcare).

### ***IHC***

Sections of ccRCC and adjacent tissues were obtained from formalin-fixed, paraffin-embedded tissue blocks (not enrolled in the proteogenomic cohort). Immunostaining was carried out as reported previously<sup>78,79</sup>. Sections were stained using relevant antibodies and the Envision detection kit (Dako). Immunostaining quantified based on the number of immunoreactive cells (quantity score) and the staining intensity (intensity score), as reported<sup>78,79</sup>.

### ***Metabolite Quantification***

Human tissues were homogenized in ice-cold phosphate-buffered saline (PBS) and centrifuged, and supernatants were collected for Hcy quantification. Hcy concentrations were determined using an Axis Homocysteine Enzyme Immunoassay Kit (Axis-Shield). To assay HTL, cells were harvested by PBS washing and denatured in pre-chilled 60% methanol (in ddH<sub>2</sub>O, pre-cooled at -80°C for 1–2 h). Cell lysates were centrifuged (10,000 × *g*) at 4°C for 5 min. Supernatants were vacuum-dried, re-dissolved in ddH<sub>2</sub>O, and subjected to ultrafiltration on a polyvinylidene fluoride low protein binding membrane (Millex-GV4 and Millex-HV4, Millipore). Metabolites were extracted and HTL was analyzed using LC-MS. SAM and SAH levels were detected using a SAM & SAH ELISA Combo Kit (Cell Biolabs). 1-Methylnicotinamide was measured using a UHPLC-QTOF-MS System (Agilent Technologies, 1290 LC, 6550 MS) as described previously<sup>80</sup>. NAD<sup>+</sup> levels were determined according using an NAD/NADH assay kit (Abcam) per the manufacturer's instructions. Each assay was repeated in triplicate, and means were used for analysis.

### ***Lysine-homocysteinylation Site Identification in ccRCC Tissues***

To identify lysine-homocysteinylation sites in tissue samples, ccRCC tumor and non-tumor tissues were ground in 0.5% NP-40 buffer, and supernatants were immunoprecipitated with anti-DNA-PKcs antibody and digested with trypsin. LC-MS/MS experiments were conducted on an EASY-nLC100 chromatograph coupled with an Orbitrap Elite (both from Thermo Fischer Scientific) equipped with an online nano-electrospray ion source. Peptides were desalted and suspended in 10 µL solvent A (A: water with 0.1% formic acid; B: acetonitrile with 0.1% formic acid). Each sample was loaded onto a self-packed C18 column (100 µm × 2 cm, 5-µm particle size), with a flow rate of 5 µL/min for 5 min and subsequently separated on the analytical column (C18, 75 µm × 20 cm) with a linear gradient from 5% B to 90% over 120 min. The column was re-equilibrated at initial conditions for 15 min. The column flow rate was maintained at 200 nL/min. The mass spectrometer was set as follows: ion-transfer capillary, 275°C; spray voltage, 2 kV; and full MS range, 400–2,000 *m/z*. Full mass spectra were acquired at 60,000 resolution with a target ion setting of 10<sup>6</sup>. One full MS scan was followed by 15 MS/MS scans, and multistage activation was enabled. The dynamic exclusion function was set as follows: repeat count, 2; repeat duration, 30 seconds; and exclusion duration, 60 s.

### ***DNA-PKcs In Vitro Kinase Assay***

*In vitro* DNA-PKcs kinase assays were conducted as described previously<sup>49</sup>. In brief, 200 ng DNA-PKcs and 3 µg p53 were incubated in a buffer containing 50 mM HEPES (pH 7.4), 100 mM KCl, 10 mM MgCl<sub>2</sub>,

2 mM EGTA, 0.1 mM EDTA, and 1 mM ATP at 30°C for 30 min. Y-shape DNA and KU70/KU80 were added as indicated. Reactions were terminated by addition of sodium dodecyl sulfate (SDS) sample loading buffer and boiling for 5 min. Samples were subjected to SDS-polyacrylamide gel electrophoresis and immunoblotting using site-specific antibody against p53.

### ***DNA-PK Kinase Assay***

DNA-PKcs activity was measured using the ADP-Glo™ + DNA-PK kinase system (Promega, Cat#4107). Briefly, we isolated DNA-PKcs protein from cells subjected to various treatments. To measure DNA-PKcs activity, 1 µL 5% DMSO, 2 µL of enzyme, and 2 µL of substrate/ATP mix were added to the wells of a 384-well plate. The plate was incubated at room temperature for 60 min. Then, 5 µL of ADP-Glo™ reagent was added and the plate was incubated at room temperature for 40 min. Consequently, 10 µL of kinase detection reagent was added and the plate was incubated at room temperature for 30 min. Luminescence was recorded with an integration time of 0.5–1 s.

### ***Cell Proliferation Assay***

Cell proliferation was assessed using the Cell Counting Kit-8 (Dojindo Laboratories). In brief, cells were seeded in a 96-well plate at  $4 \times 10^3$  cells/well and allowed to adhere. Cell Counting Kit-8 solution (10 µL) was added to each well, and the cells were incubated in 5% CO<sub>2</sub> at 37°C for 2 h. Cell proliferation was determined by measuring the absorbance at 450 nm.

### ***Comet Assay***

A Comet Assay Kit (Trevigen) was used to detect single- and double-stranded DNA breaks in cultured cells and tissues. Slides were examined under a Leica DMI 4000B epifluorescence microscope (425–500-nm excitation). Comet slides were used for each condition. In normal cells, fluorescence is mostly confined to the nucleus because intact DNA cannot migrate. In DNA-damaged cells, DNA is denatured with an alkaline or neutral solution to detect single- or double-stranded breaks, respectively; negatively charged DNA fragments are released from the nucleus and migrate toward the anode.

### ***In Vivo Xenograft studies***

Four-to-six-week-old Balb/C nude mice were obtained from Shanghai SLAC Laboratory Animal Co., Ltd. Control and NNMT-overexpressing ACHN and 786-O cell lines were subcutaneously transplanted into the left and right flanks of each mouse. For the IR group, irradiated control and NNMT-overexpressing cells were transplanted into the left and right flanks of each mouse. For the IR+NAC group, irradiated control and NNMT-overexpressing cells were transplanted into the left and right flanks of each mouse, and the mice were intraperitoneally injected with NAC (500 mg/kg) every other day. At the end of the experiment, following euthanasia, tumors were excised, weighed, and imaged. All procedures were approved by the Animal Care Committee at Fudan University.

## **QUANTIFICATION AND STATISTICAL ANALYSIS**



Quantification methods and statistical analysis methods for proteomic and integrated analyses were mainly described and referenced in the respective Method Details subsections.

Additionally, standard statistical tests were used to analyze the clinical data, including but not limited to Student's t test, Fisher's exact test, Kruskal-Wallis test, log-rank test. All statistical tests were two-sided, and statistical significance was considered when p value < 0.05. To account for multiple-testing, the p values were adjusted using the Benjamini–Hochberg FDR correction. Kaplan–Meier plots (log-rank test) were used to describe overall survival. Variables associated with overall survival were identified using univariate Cox proportional hazards regression models. Significant factors in univariate analysis were further subjected to a multivariate Cox regression analysis. All the analyses of clinical data were performed in R and GraphPad Prism. For functional experiments, each was repeated at least three times independently, and results were expressed as mean ± standard error of the mean (SEM). Statistical analysis was performed using GraphPad Prism.

## **DATA AVAILABILITY**

Proteome raw datasets are publicly available at the iProX data portal:

<https://www.iprox.org/page/PSV023.html?url=1605014585802S8oG>, with a password rEiV. WES data files can be accessed at <https://www.biosino.org/node/review/detail/OEV000128?code=XZPGYZGS>.

## **Declarations**

### **Acknowledgements**

This work is supported by Chinese Ministry of Science and Technology (2016YFA0502500); National International Cooperation Grant (2014DFB30010); National Program on Key Basic Research Project (973 Program, 2012CB910300, 2014CBA02000, 2019YFA0801900, 2019YFC1316000), National High-tech R&D Program of China (863 program, 2014AA020201), National Natural Science Foundation of China (31270822, 81722021, 81771627, 81722021, 81771627, 81802525, 31521003, 81672544, 81872099), Beijing Natural Science Foundation (Z131100005213003), Science and Technology Municipal Commission of Shanghai, China (18511108000, 2018ZHYL0201), Natural Science Foundation of Shanghai (No. 20ZR1413100), and Shanghai "Rising Stars of Medical Talent" Youth Medical Talents–Specialist Program. We thank the Lifehealthcare Clinical Laboratories (Hangzhou, China) for whole exome sequencing service and related technical assistance.

### **Competing interests**

The authors declare no competing interests.

### **Correspondence and requests for materials**

Further information and requests should be directed to and will be fulfilled by the lead Contact, Chen Ding ([chend@fudan.com.cn](mailto:chend@fudan.com.cn)).

## References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J Clin* **70**, 7–30, doi:10.3322/caac.21590 (2020).
2. Cancer Genome Atlas Research, N. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49, doi:10.1038/nature12222 (2013).
3. Clark, D. J. *et al.* Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* **179**, 964–983 e931, doi:10.1016/j.cell.2019.10.007 (2019).
4. Wang, X. M. *et al.* Integrative genomic study of Chinese clear cell renal cell carcinoma reveals features associated with thrombus. *Nat Commun* **11**, 739, doi:10.1038/s41467-020-14601-9 (2020).
5. Scelo, G. *et al.* Variation in genomic landscape of clear cell renal cell carcinoma across Europe. *Nat Commun* **5**, 5135, doi:10.1038/ncomms6135 (2014).
6. Gillette, M. A. *et al.* Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. *Cell* **182**, 200–225 e235, doi:10.1016/j.cell.2020.06.013 (2020).
7. Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62, doi:10.1038/nature18003 (2016).
8. Dou, Y. *et al.* Proteogenomic Characterization of Endometrial Carcinoma. *Cell* **180**, 729–748 e726, doi:10.1016/j.cell.2020.01.026 (2020).
9. Vasaikar, S. *et al.* Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* **177**, 1035–1049 e1019, doi:10.1016/j.cell.2019.03.030 (2019).
10. Gao, Q. *et al.* Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell* **179**, 561–577.e522, doi:10.1016/j.cell.2019.08.052 (2019).
11. Jiang, Y. *et al.* Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* **567**, 257–261, doi:10.1038/s41586-019-0987-8 (2019).
12. Chen, Y. J. *et al.* Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression. *Cell* **182**, 226–244 e217, doi:10.1016/j.cell.2020.06.012 (2020).
13. Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* **45**, 860–867, doi:10.1038/ng.2699 (2013).
14. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941–D947, doi:10.1093/nar/gky1015 %J Nucleic Acids Research (2018).
15. DeBelle, F. D., Vanherweghem, J. L. & Nortier, J. L. Aristolochic acid nephropathy: a worldwide problem. *Kidney international* **74**, 158–169, doi:10.1038/ki.2008.129 (2008).
16. Zhang, W. *et al.* LFQuant: a label-free fast quantitative analysis tool for high-resolution LC-MS/MS proteomics data. *Proteomics* **12**, 3475–3484, doi:10.1002/pmic.201200017 (2012).
17. Feng, J. *et al.* Firmiana: towards a one-stop proteomic cloud platform for data processing and analysis. *Nat Biotechnol* **35**, 409–412, doi:10.1038/nbt.3825 (2017).

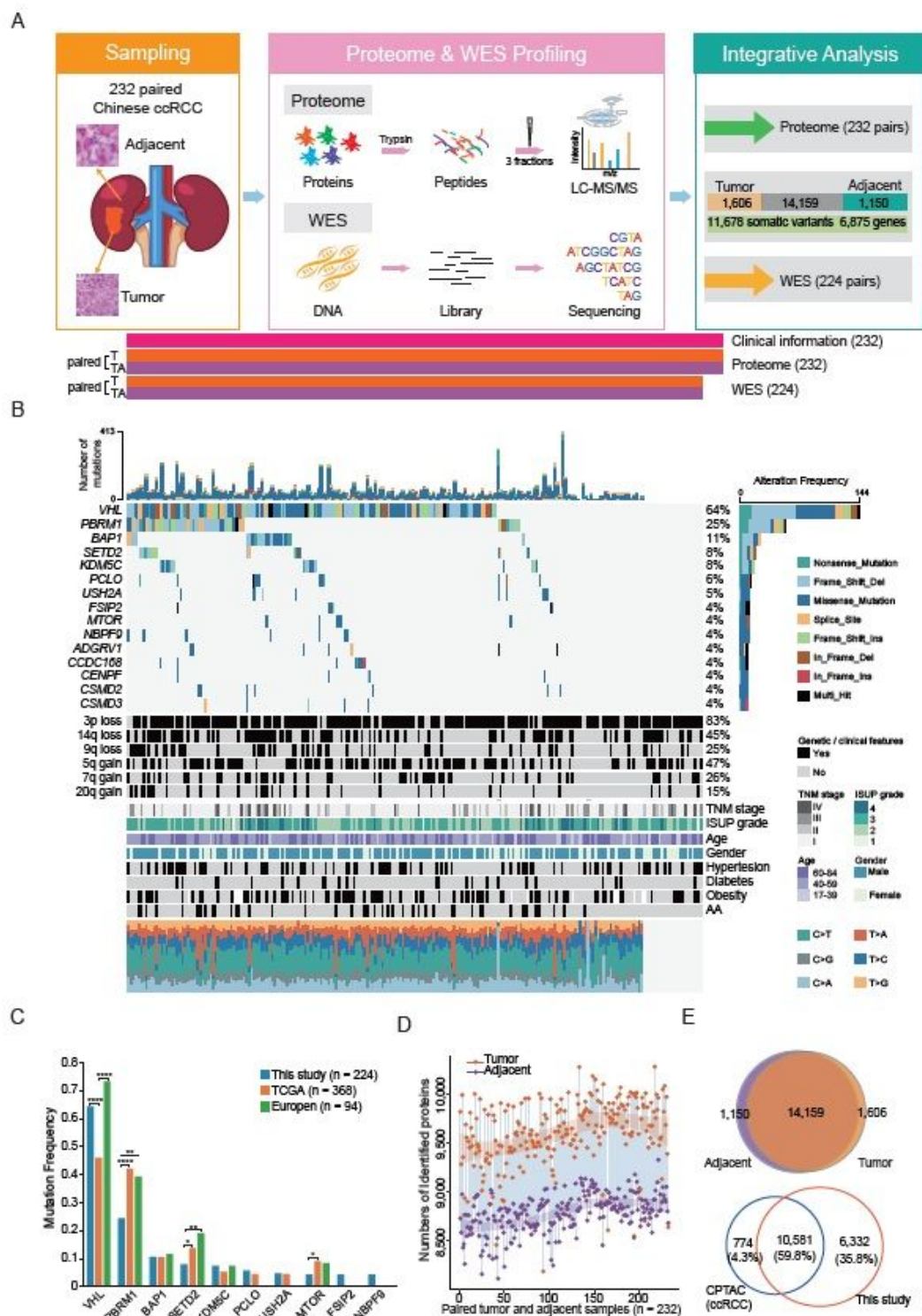
18. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41, doi:10.1186/gb-2011-12-4-r41 (2011).
19. Zhang, B. *et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382–387, doi:10.1038/nature13438 (2014).
20. Brugarolas, J. Molecular genetics of clear-cell renal cell carcinoma. *J Clin Oncol* **32**, 1968–1976, doi:10.1200/JCO.2012.45.2003 (2014).
21. Hsieh, J. J. *et al.* Chromosome 3p Loss-Orchestrated VHL, HIF, and Epigenetic Deregulation in Clear Cell Renal Cell Carcinoma. *J Clin Oncol* **36**, Jco2018792549, doi:10.1200/jco.2018.79.2549 (2018).
22. Kapur, P. *et al.* Effects on survival of BAP1 and PBRM1 mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation. *The Lancet Oncology* **14**, 159–167, doi:10.1016/s1470-2045(12)70584-3 (2013).
23. Hakimi, A. A. *et al.* Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: a report by MSKCC and the KIRC TCGA research network. *Clin Cancer Res* **19**, 3259–3267, doi:10.1158/1078-0432.CCR-12-3886 (2013).
24. Mangogna, A. *et al.* Is the Complement Protein C1q a Pro- or Anti-tumorigenic Factor? Bioinformatics Analysis Involving Human Carcinomas. *Front Immunol* **10**, 865, doi:10.3389/fimmu.2019.00865 (2019).
25. Roumenina, L. T. *et al.* Tumor Cells Hijack Macrophage-Produced Complement C1q to Promote Tumor Growth. *Cancer Immunol Res* **7**, 1091–1105, doi:10.1158/2326-6066.Cir-18-0891 (2019).
26. Perroud, B., Ishimaru, T., Borowsky, A. D. & Weiss, R. H. Grade-dependent proteomics characterization of kidney cancer. *Mol Cell Proteomics* **8**, 971–985, doi:10.1074/mcp.M800252-MCP200 (2009).
27. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. 1–16, doi:10.1200/po.17.00011 (2017).
28. Dizman, N., Philip, E. J. & Pal, S. K. Genomic profiling in renal cell carcinoma. *Nat Rev Nephrol* **16**, 435–451, doi:10.1038/s41581-020-0301-x (2020).
29. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419, doi:10.1126/science.1260419 (2015).
30. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30, doi:10.1093/nar/28.1.27 (2000).
31. Fabregat, A. *et al.* Reactome graph database: Efficient access to complex pathway data. *PLoS computational biology* **14**, e1005968, doi:10.1371/journal.pcbi.1005968 (2018).
32. Hakimi, A. A. *et al.* An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. *Cancer Cell* **29**, 104–116, doi:10.1016/j.ccell.2015.12.004 (2016).
33. Gatto, F., Nookaew, I. & Nielsen, J. Chromosome 3p loss of heterozygosity is associated with a unique metabolic network in clear cell renal carcinoma. *Proc Natl Acad Sci U S A* **111**, E866–875, doi:10.1073/pnas.1319196111 (2014).

34. Wettersten, H. I., Aboud, O. A., Lara, P. N., Jr. & Weiss, R. H. Metabolic reprogramming in clear cell renal cell carcinoma. *Nat Rev Nephrol* **13**, 410–419, doi:10.1038/nrneph.2017.59 (2017).
35. La, F. L. *et al.* Expression of scavenger receptor MARCO defines a targetable tumor-associated macrophage subset in non-small cell lung cancer. (2018).
36. Qu, Y. Y. *et al.* Inactivation of the AMPK-GATA3-ECHS1 Pathway Induces Fatty Acid Synthesis That Promotes Clear Cell Renal Cell Carcinoma Growth. *Cancer Res* **80**, 319–333, doi:10.1158/0008-5472.CAN-19-1023 (2020).
37. Ducker, G. S. & Rabinowitz, J. D. One-Carbon Metabolism in Health and Disease. *Cell Metab* **25**, 27–42, doi:10.1016/j.cmet.2016.08.009 (2017).
38. Stipanuk, M. H. Sulfur amino acid metabolism: pathways for production and removal of homocysteine and cysteine. *Annual review of nutrition* **24**, 539–577, doi:10.1146/annurev.nutr.24.012003.132418 (2004).
39. Delahunt, B. *et al.* The International Society of Urological Pathology (ISUP) grading system for renal cell carcinoma and other prognostic parameters. *The American journal of surgical pathology* **37**, 1490–1504, doi:10.1097/PAS.0b013e318299f0fb (2013).
40. Monti, S., Tamayo, P., Mesirov, J. & Golub, T. J. M. L. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. **52**, 91–118 (2003).
41. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* **4**, 2612, doi:10.1038/ncomms3612 (2013).
42. Etzerodt, A. *et al.* Specific targeting of CD163 + TAMs mobilizes inflammatory monocytes and promotes T cell-mediated tumor regression. **216**, 2394–2411, doi:10.1084/jem.20182124 %J The Journal of Experimental Medicine (2019).
43. Kovaleva, O. V., Samoilo, D. V., Shitova, M. S. & Gratchev, A. Tumor Associated Macrophages in Kidney Cancer. *Anal Cell Pathol (Amst)* 2016, 9307549, doi:10.1155/2016/9307549 (2016).
44. Reis, E. S., Mastellos, D. C., Ricklin, D., Mantovani, A. & Lambris, J. D. Complement in cancer: untangling an intricate relationship. *Nat Rev Immunol* **18**, 5–18, doi:10.1038/nri.2017.97 (2018).
45. Roumenina, L. T., Daugan, M. V., Petitprez, F., Sautes-Fridman, C. & Fridman, W. H. Context-dependent roles of complement in cancer. *Nat Rev Cancer* **19**, 698–715, doi:10.1038/s41568-019-0210-0 (2019).
46. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* **46**, D1074-D1082, doi:10.1093/nar/gkx1037 (2018).
47. Wang, D. *et al.* Colonic Lysine Homocysteinylation Induced by High-Fat Diet Suppresses DNA Damage Repair. *Cell Rep* **25**, 398–412 e396, doi:10.1016/j.celrep.2018.09.022 (2018).
48. Mei, X. *et al.* Inhibiting MARSs reduces hyperhomocysteinemia-associated neural tube and congenital heart defects. *EMBO molecular medicine* **12**, e9469, doi:10.15252/emmm.201809469 (2020).

49. Yin, X., Liu, M., Tian, Y., Wang, J. & Xu, Y. Cryo-EM structure of human DNA-PK holoenzyme. *Cell research* **27**, 1341–1350, doi:10.1038/cr.2017.110 (2017).
50. Senbabaoglu, Y. *et al.* Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol* **17**, 231, doi:10.1186/s13059-016-1092-z (2016).
51. Roessler, M. *et al.* Identification of nicotinamide N-methyltransferase as a novel serum tumor marker for colorectal cancer. *Clin Cancer Res* **11**, 6550–6557, doi:10.1158/1078-0432.CCR-05-0983 (2005).
52. Wu, Y., Siadaty, M. S., Berens, M. E., Hampton, G. M. & Theodorescu, D. Overlapping gene expression profiles of cell migration and tumor invasion in human bladder cancer identify metallothionein 1E and nicotinamide N-methyltransferase as novel regulators of cell migration. *Oncogene* **27**, 6679–6689, doi:10.1038/onc.2008.264 (2008).
53. Kim, J. *et al.* Expression of nicotinamide N-methyltransferase in hepatocellular carcinoma is associated with poor prognosis. *J Exp Clin Cancer Res* **28**, 20, doi:10.1186/1756-9966-28-20 (2009).
54. Tomida, M., Mikami, I., Takeuchi, S., Nishimura, H. & Akiyama, H. Serum levels of nicotinamide N-methyltransferase in patients with lung cancer. *J Cancer Res Clin Oncol* **135**, 1223–1229, doi:10.1007/s00432-009-0563-y (2009).
55. Tang, S. W. *et al.* Nicotinamide N-methyltransferase induces cellular invasion through activating matrix metalloproteinase-2 expression in clear cell renal cell carcinoma cells. *Carcinogenesis* **32**, 138–145, doi:10.1093/carcin/bgq225 (2011).
56. Eckert, M. A. *et al.* Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts. *Nature* **569**, 723–728, doi:10.1038/s41586-019-1173-8 (2019).
57. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297–1303, doi:10.1101/gr.107524.110 (2010).
58. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921, doi:10.1038/35057062 (2001).
59. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164, doi:10.1093/nar/gkq603 (2010).
60. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C. & Koeffler, H. P. J. G. r. Maftools: efficient and comprehensive analysis of somatic variants in cancer. **28**, 1747–1756 (2018).
61. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218, doi:10.1038/nature12213 (2013).
62. Goldman, M. J. *et al.* Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* **38**, 675–678, doi:10.1038/s41587-020-0546-8 (2020).
63. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* **3**, 246–259, doi:10.1016/j.celrep.2012.12.008 (2013).

64. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359–362, doi:10.1038/nmeth.1322 (2009).
65. Wei, T. *et al.* Package 'corrplot'. *Statistician* **56**, 316–324 (2017).
66. Leek, J. (2019).
67. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301–2319, doi:10.1038/nprot.2016.136 (2016).
68. Kassambara, A. & Mundt, F. Factoextra: extract and visualize the results of multivariate data analyses. *R package version 1*, 2016 (2016).
69. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57, doi:10.1038/nprot.2008.211 (2009).
70. Ricketts, C. J. *et al.* The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep* **23**, 313–326 e315, doi:10.1016/j.celrep.2018.03.075 (2018).
71. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112, doi:10.1038/nature08460 (2009).
72. Hänzelmann, S., Castelo, R. & Guinney, J. J. B. B. GSVA: Gene set variation analysis for microarray and RNA-Seq data. **14**, 7 (2013).
73. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. **102**, 15545–15550, doi:10.1073/pnas.0506580102 %J Proceedings of the National Academy of Sciences (2005).
74. Kamburov, A. *et al.* ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res* **39**, D712-717, doi:10.1093/nar/gkq1156 (2011).
75. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573, doi:10.1093/bioinformatics/btq170 (2010).
76. Therneau, T. M. & Lumley, T. Package 'survival'. *R Top Doc* **128** (2015).
77. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308, doi:10.1016/0092-8674(81)90413-x (1981).
78. Qu, Y. *et al.* Constitutively active AR-V7 plays an essential role in the development and progression of castration-resistant prostate cancer. *Sci Rep* **5**, 7654, doi:10.1038/srep07654 (2015).
79. Qu, Y. *et al.* Phosphorylated 4EBP1 is associated with tumor progression and poor prognosis in Xp11.2 translocation renal cell carcinoma. *Sci Rep* **6**, 23594, doi:10.1038/srep23594 (2016).
80. Strom, K. *et al.* N(1)-methylnicotinamide is a signalling molecule produced in skeletal muscle coordinating energy metabolism. *Sci Rep* **8**, 3016, doi:10.1038/s41598-018-21099-1 (2018).

## Figures



**Figure 1**

Proteogenomic Landscape of Chinese ccRCC. A, Schematic representation of the multiomics analyses of ccRCC, including sample preparation, protein identification, WES, and function verification. B, Genomic profile and associated clinical features of 224 ccRCC patients. C, Comparison of frequently mutated genes among Chinese, European, and TCGA cohorts (Fisher's exact test). D, Overview of proteomic profiles of pairwise ccRCC samples. The dashed curves fitted by lasso regression show the distribution of

protein identifications. The shading that underlies the lasso curves denotes the 95% confidence intervals. E, The upper Venn diagram shows the overlap of proteins identified in tumors and adjacent normal tissues. The lower Venn diagram shows that proteins identified in this study cover most of the proteins identified in the CPTAC ccRCC cohort.

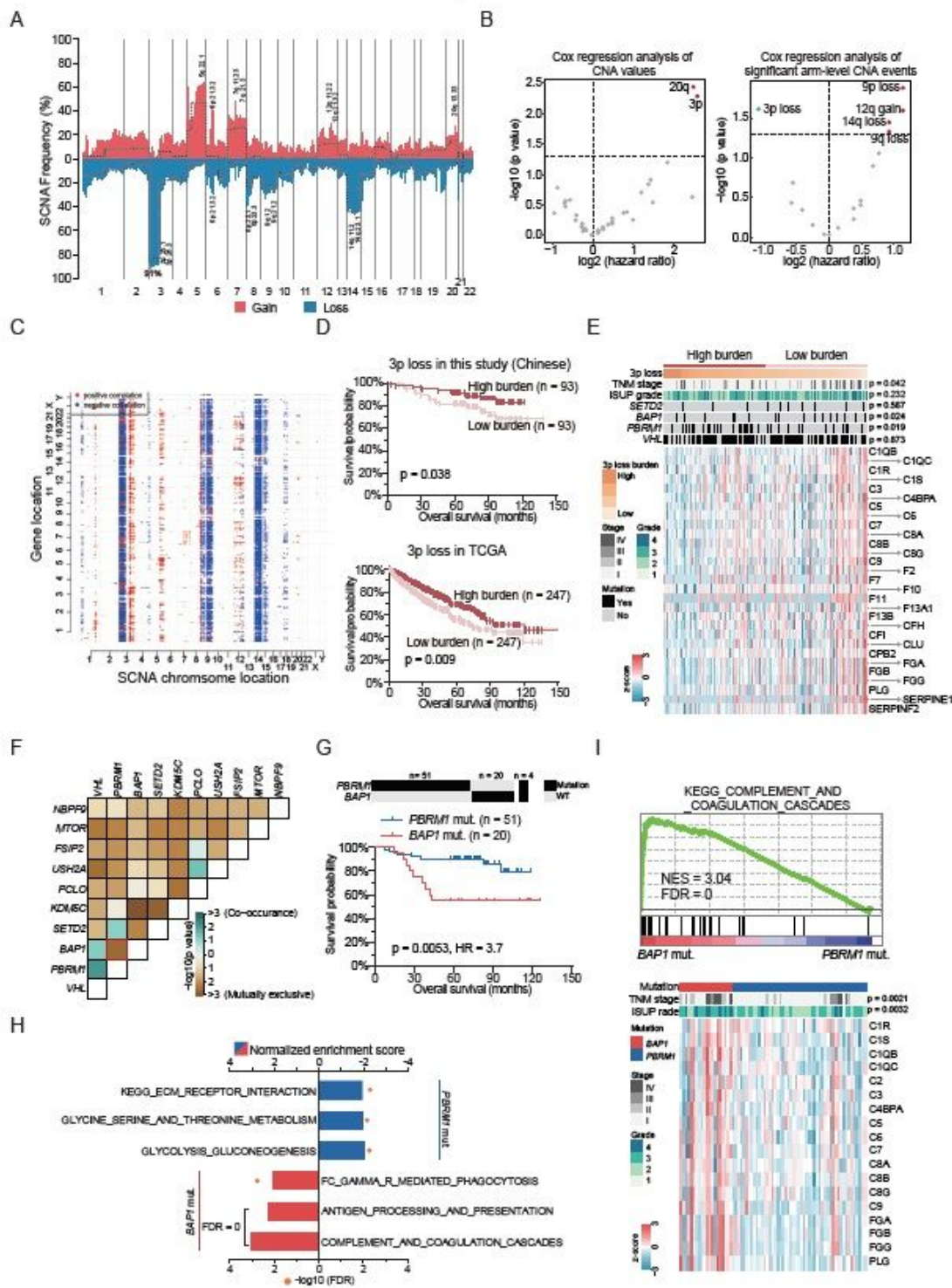
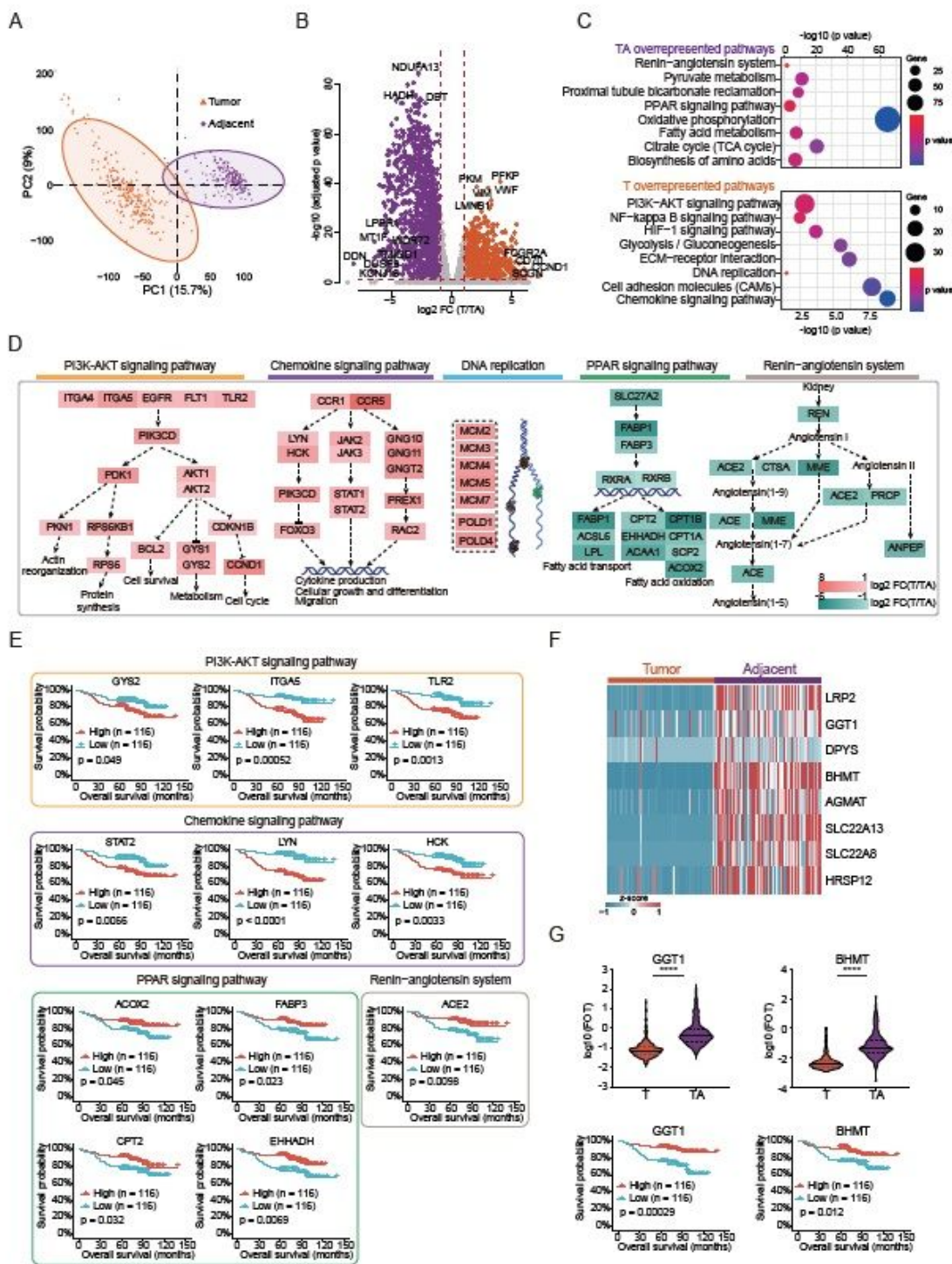


Figure 2



Profiles of CNAs and Effects of CNA on Somatic Mutations, Proteome, and Clinical outcomes. A, Frequency of SCNAs. Copy number gains and losses are indicated in red and blue, respectively. The dotted line indicates the frequency of arm-level CNA events. B, Cox regression analysis of CNA values and significant arm-level CNA events. C, Correlations of CNA (x axes) with protein abundance (y axes). Significant (FDR < 0.01) positive (red) and negative (blue) correlations are shown. D, Kaplan–Meier curves of OS for patients with different 3p loss burden in the Chinese and TCGA cohorts. E, Heatmap of proteins positively correlated with Chromosome 3p CNAs. The upper panel shows the association of 3p loss burden with clinical and mutation signatures. F, Mutual exclusivity and co-occurrence analysis of the 10 most frequent mutations (pair-wise Fisher’s exact test). G, Kaplan–Meier curves of OS for patients with only PBRM1 mutations or BAP1 mutations (log-rank test). H, GSEA of patients BAP1 mutations (n = 20) compared to patients with PBRM1 mutations (n = 51). NES, normalized enrichment score. I, Heatmap of protein expression abundances of complement and coagulation cascades components between BAP1 and PBRM1 mutant ccRCC. The upper panel shows enrichment plot of complement and coagulation cascades.



**Figure 3**

Proteomic Alterations in ccRCC Compared to Adjacent Tissues. A, PCA of 6,111 proteins in 232 paired tumor and adjacent tissue samples. Orange, tumor tissue; purple, tumor adjacent tissue. B, Volcano plot showing DEPs (Benjamini–Hochberg-adjusted p value < 0.05, FC > 2) in tumor and adjacent tissues. Proteins that were significantly overexpressed in tumor/adjacent tissues are presented with red/blue filled asterisks. C, DEPs in tumors and adjacent tissues, and their associated biological pathways. D-E,

Representative overrepresented pathways in ccRCC tumor and adjacent tissues and associations of proteins involved in these pathways with prognosis (log-rank test). F, Expression of proximal tubule signatures in ccRCC tumor and adjacent tissues. G, Three significantly downregulated renal proximal tubule signatures (GGT1, BHMT, LRP2) in ccRCC tumors were associated with poor prognosis (log-rank test).

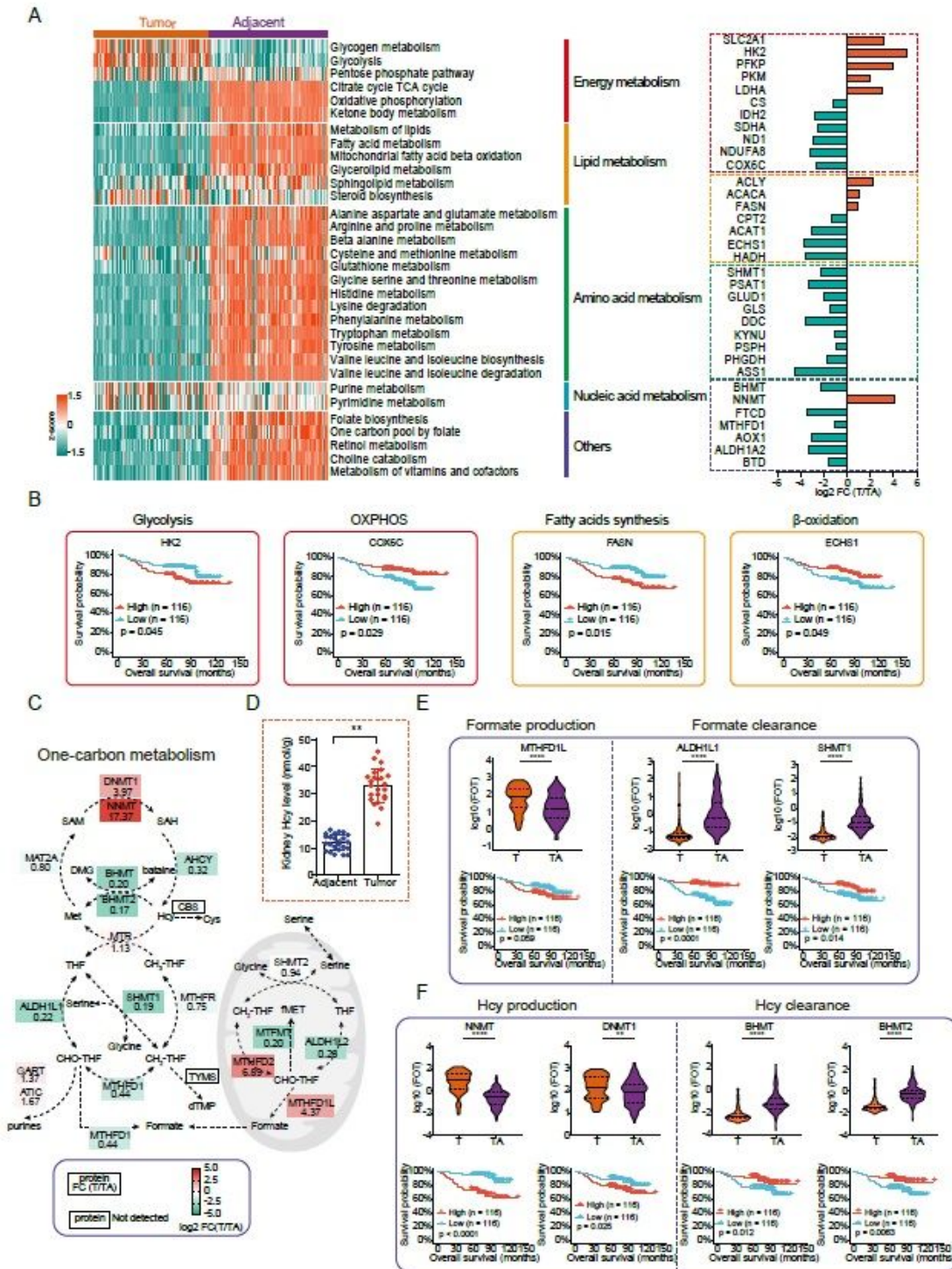
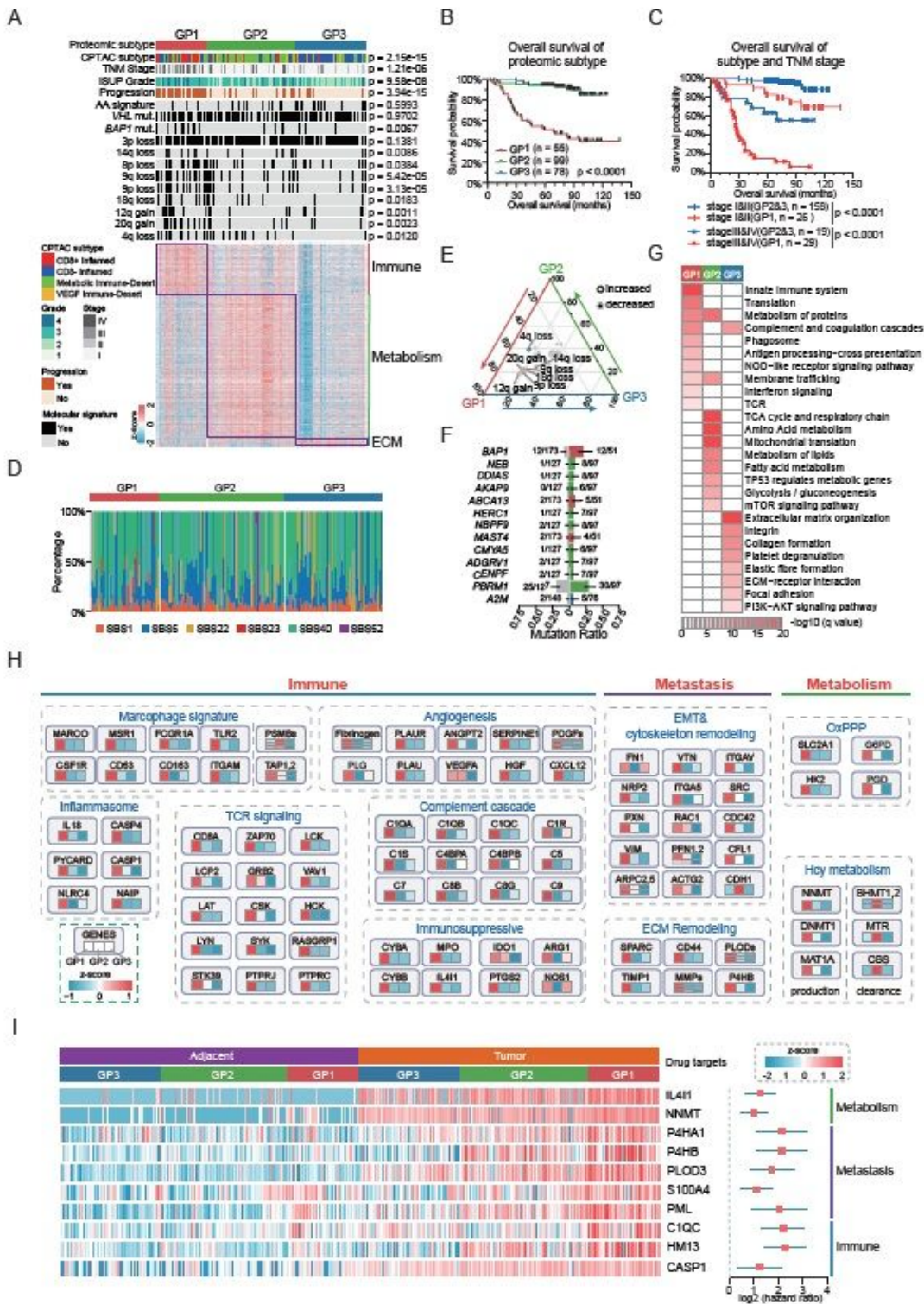


Figure 4

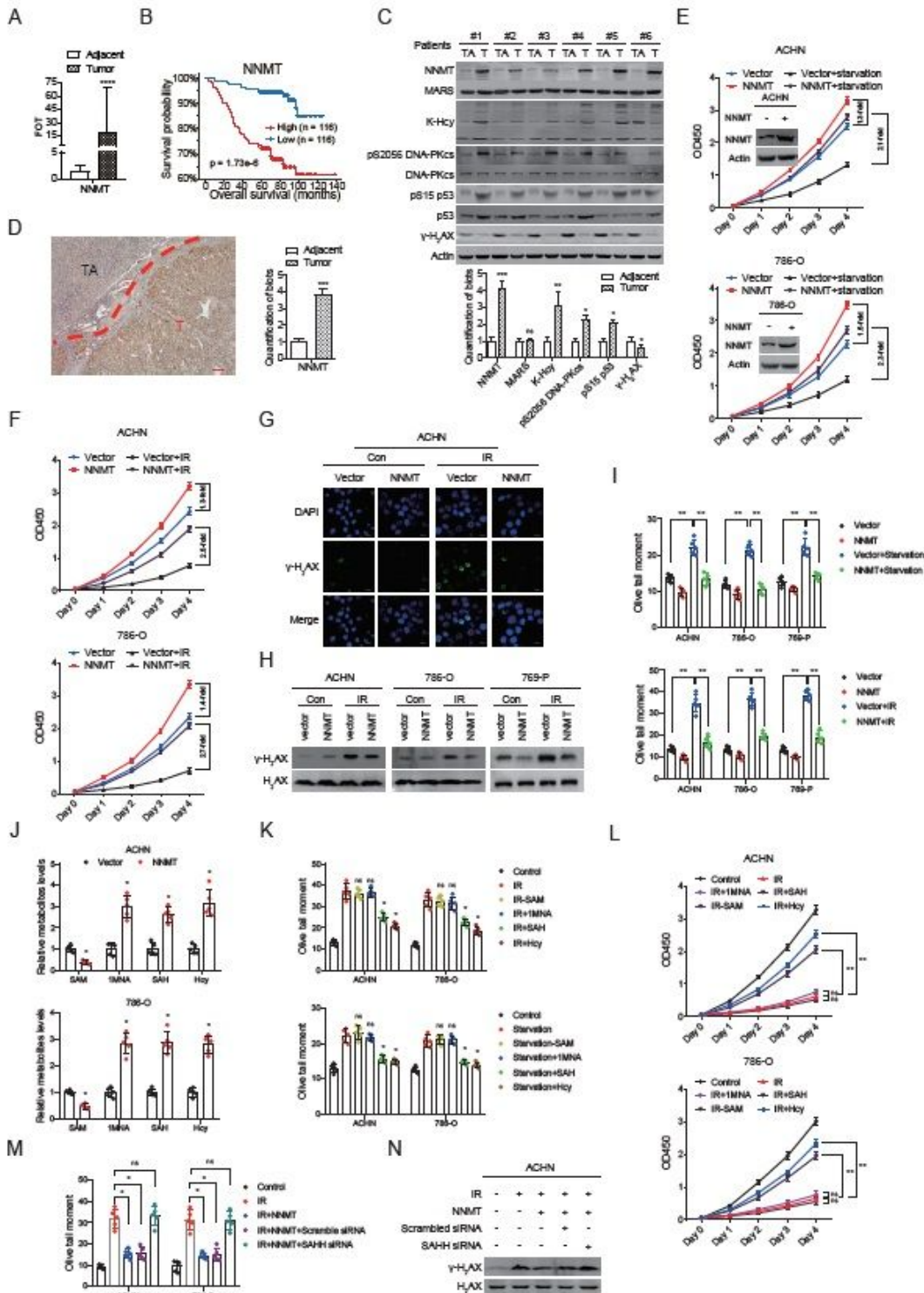
Dysregulation of Metabolic Bioprocesses in ccRCC. A, Left panel, GSVA was used to estimate metabolism-related pathway scores. Right panel, alteration scores of each protein depicted as  $\log_2$  FC (T/TA). B, Representative proteins from the metabolic pathways and their association with prognosis (log-rank test). C, One-carbon metabolism dysregulation in ccRCC. Alteration scores of each protein depicted as FC (T/TA). The color gradient indicates alteration scores of each protein depicted as  $\log_2$  FC (T/TA). Red, upregulated proteins in tumor; green, downregulated proteins in tumor. Proteins undetected are indicated with black boxes. D, Hcy concentrations in tumor and adjacent tissues (paired t test). E, Differentially expressed one-carbon metabolic enzymes between tumor and adjacent tissues and their associations with clinical outcomes (log-rank test).



**Figure 5**

Proteomic Subtypes of ccRCC and Associations with Genetic Features and Clinical Outcomes. A, Relative abundances of upregulated proteins in the three proteomic subtypes and associations of proteomic subtypes with multiple variables, including CPTAC subtype, TNM stage, ISUP grade, status of progression and genetic features (Fisher's exact test). B, Kaplan–Meier curves of OS for the three subtypes (log-rank test). C, Kaplan–Meier curves of OS for subtypes GP1 and GP2&3 at different TNM stages (stage I&II vs.

III&IV) (log-rank test). D, Relative percentage of each mutation signatures in the three subtypes. E, Genes with differential mutation rates in each subtype (Fisher's exact test). F, Ternary plot showing the distribution of significant arm-level events in the three subtypes. G, Upregulated pathways enriched in the three subtypes. H, Summary of signature proteins and pathways involved in tumor aggressiveness present in GP1. I, Drug target candidates for ccRCC. Druggability information was derived from DrugBank. Left, z-scored protein abundance. Middle, HR of each protein. Right, proteins involved in biological functions.



## Figure 6

NNMT Promotes Cancer Cell Proliferation through Hcy Accumulation. A, NNMT expression based on our proteomic data. B, Association between NNMT protein expression and OS. C, Left panel, representative western blots of NNMT, MARS, K-Hcy, DNA-PKcs, DNA-PKcs (pS2056), and  $\gamma$ -H2AX in tumor and tumor tissues. Right panel, quantified western blot results. TA = tumor adjacent, T = tumor. D, Left panel, IHC results of NNMT expression in tumors and adjacent tissues (scale bars: 50  $\mu$ m). Right panel, quantified IHC results of 12 sample pairs. TA = tumor adjacent, T = tumor. Results for other samples are shown in Figure S10. E-F, Cell proliferation associated with various treatments (n = 5 repeats per group). G, Immunofluorescence staining of  $\gamma$ -H2AX in cells under various treatments (scale bars: 20  $\mu$ m). H, Western blot analysis of  $\gamma$ -H2AX and H2AX in cells under various treatments. I, Comet assay of DNA damage levels in cells subjected to various treatments. For each group, DNA damage levels in a total of 30 cells from five independent repeats were measured. J, Relative metabolite levels in cells subjected to various treatments. K, Comet assay of DNA damage levels in cells subjected to various treatments. L, Cell proliferation associated with various treatments (n = 5 repeats per group). M, Comet assay of DNA damage levels in cells subjected to various treatments. N, Western blot analysis of  $\gamma$ -H2AX and H2AX in cells under various treatments. Data are the mean  $\pm$  SE. \*\*p < 0.01, \* p < 0.05.





treatments. F, Comet assay of DNA damage levels in ccRCC tumor vs. adjacent normal tissues. G-H, Hcy treatment weakens the interaction between endogenous DNA-PKcs and endogenous KU70 (a) or KU80 (b) in ACHN and 769-P cells. I, In vitro DNA-PKcs activity assayed by monitoring its kinase activity in phosphorylating its substrate p53. J, DNA-PKcs activity indicated by measuring ADP formation in an ADP-Glo-DNA-PK assay. K, HTL increases the in vitro DNA-PKcs activity. L, DNA-PK activity under different treatments. M-N, DNA-PKcs KW and 4KW mutants exhibit enhanced binding affinity for KU70 (g) and KU80 (h) compared to wild-type DNA-PKcs. O, DNA-PKcs 4KW mutant exhibits enhanced kinase activity. P-Q, Tumor size of cell xenografts under different treatments in normal and NAC-administered nude mice. R, Model depicting NNMT-mediated DNA repair and cell proliferation in ccRCC. Data are the mean  $\pm$  SE. \*\*p < 0.01, \*p < 0.05.

## Supplementary Files

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

- [FigureS1.pdf](#)
- [FigureS2.pdf](#)
- [FigureS3.pdf](#)
- [FigureS4.pdf](#)
- [FigureS5.pdf](#)
- [FigureS6.pdf](#)
- [FigureS7.pdf](#)
- [FigureS8.pdf](#)
- [FigureS9.pdf](#)
- [FigureS10.pdf](#)
- [FigureS11.pdf](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)