Spring 5-15-2022

Multi-Omics Investigation of Tumor Heterogeneity, Oncogenic Signaling, and Treatment Response in Human Cancers

Yige Wu
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Division of Biology and Biomedical Sciences
Human and Statistical Genetics

Dissertation Examination Committee:
Li Ding, Chair
John Edwards
Ramaswamy Govindan
Michael Province
John Welch

Multi-Omics Investigation of Tumor Heterogeneity, Oncogenic Signaling, and Treatment Response in Human Cancers

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2022
St. Louis, Missouri

Table of Contents

List of Figures

Acknowledgments

I would like to thank my family, especially my parents. Without their support, their patience, and the freedom they gave me, I wouldn't be the person I am today.

I also would like to thank my thesis mentor Li. She showed me how to keep the big picture in mind when I get into the nitty-gritty of the research. She taught me the importance of planning and timeline. She taught me how to take overwhelmingly big datasets, sort them out systematically and then weave the results into a story.

I would also like to thank my committee members. Dr. Michael Province, Dr. John Edwards, Dr. John Welch, and Dr. Ramaswamy Govindan. I'm grateful for the time they spent talking with and listening to me about my thesis proposal and thesis update. I'm also deeply grateful for their constructive feedback and warm support.

I love my Ding lab family. They are one major reason I joined the lab. Wenwei is one of the most important people who recruited me to the lab. She was also one of the first friends I made in the lab. Reyka and Sohini showed me how to have fun and help other people. Kuan taught me how to be self-reliant and resourceful. Bobo taught me how to take better notes. Dan and Fernanda showed me how to communicate and present clearly. Alla showed how to be straightforward, and how to be thoughtful in interpreting the results with a clear understanding of biology in mind. Steven showed how to think the statistical assumptions behind each analysis. All of you showed me how science can be fun, and how collaborating and helping each other can be fun.

Finally, I would like to thank my boyfriend Rick and my cat zaza. They always cheer me up when I am at home.

Yige Wu

Washington University in St. Louis

August 2022

Dedicated to my parents.

ABSTRACT

Multi-Omics Investigation of Tumor Heterogeneity, Oncogenic Signaling, and Treatment Response in human cancers

by

Yige Wu

Doctor of Philosophy in Biology and Biomedical Sciences

Human and Statistical Genetics

Washington University in St. Louis, 2022

Professor Li Ding, Chair

Cancer is a highly complex disease with aberrations at the genetic, epigenetic, transcriptomic, and protein levels that drove its phenotypic diversity. Clear cell renal cell carcinoma (ccRCC) is the most common form of kidney cancer, comprising roughly 80% of cases. To define the epigenetic and transcriptomic regulation of ccRCC at the single nucleus (sn) level, we performed snRNA-seq and snATAC-seq in 34 and 28 samples respectively, including primary tumors and normal adjacent tissues, and matched them with bulk proteogenomics data. We identified tumor-specific markers and tumor subpopulations using snRNA-seq, which demonstrated diverse pathway activity within and across patients. *PBRM1* and *BAP1* are two of the most frequently mutated genes in ccRCC, and both encode epigenetic regulators. However, the consequences of *BAP1* and *PBRM1* mutations on chromatin accessibility and downstream transcriptional networks remain largely unknown. Utilizing the combined analysis of snATAC-seq and snRNA-seq, we dissected chromatin accessibility and transcriptome changes associated with *BAP1* and *PBRM1* mutations, illuminating molecular alterations underlying differential phenotypes between *BAP1*- and *PBRM1*-mutant patients.

For the treatment of RCC, patients with metastatic or inoperable tumors typically receive systemic treatment with targeted therapy and/or immunotherapy. Although these drugs have been proven effective to some extent, resistance eventually develops, and combinational therapy will be necessary to overcome such resistance. Patient-derived xenograft (PDX) models have proven valuable in studying treatment mechanisms and novel therapeutics for cancer, including renal cell carcinoma. Hence, we performed a series of drug tests on a set of RCC PDX models, in which cabozantinib and sapanisertib are the two most effective drugs, and found the combination of two drugs is effective for all six models. We collected PDX tumors at baseline and under treatments and performed bulk whole-exome sequencing, bulk RNA-seq, bulk proteomics and phosphoproteomics, and snRNA-seq. We revealed the pathways affected by the combination therapy and identified treatment-affected proteins that are associated with patient survival. We also identified baseline protein markers that may serve to predict treatment response, such as MET, with support from snRNA-seq data. This study proposed a potential new combination for RCC patients and revealed potential molecular alterations underlying tumor reduction induced by the combination treatment.

Chapter 1: Introduction

1.1 Cancer hallmarks

Cancer is a complex disease with diverse phenotypes. To rationalize the complex cancer phenotypes under a conceptual scaffold, Weinburg and Hanahan et al[1] have proposed a common set of hallmarks shared by human cancers. These cancer hallmarks define a set of functional capabilities acquired by normal human cells crucial for their ability to form malignant tumors. After the initial publication in 2000 and the later 2011 sequel[1,2], the 8 cancer hallmarks, comprise "sustaining proliferative signaling", "evading growth suppressors", "resisting cell death", "enabling replicative immortality", "genome instability and mutation", "activating invasion and metastasis", "deregulating cellular metabolism" and "avoiding immune destruction". They also introduced the concept of "enabling characteristic", a concept focusing on the molecular and cellular mechanisms by which hallmarks are acquired rather than the hallmarks themselves. The first enabling characteristic, inducing or accessing vasculature, is defined as the capability to induce or access vasculature that supports tumor growth, primarily by invasion and metastasis. The second enabling characteristic, tumor-promoting inflammation, complements the hallmark "genome instability and mutation and is fundamentally involved in activating the eight hallmark capabilities. Most recently, Hanahan et al[3] proposed new hallmarks and enabling characteristics, including "unlocking phenotypic plasticity", "nonmutational epigenetic reprogramming", "polymorphic microbiomes", and "senescent cells", with the growing support from published evidence. In this conceptual framework, they also highlighted the importance of the tumor microenvironment (TME), which is composed of a heterogeneous population of cancer cells,

cancer stem cells, and a large array of different stromal and immune cell types. As of now, eleven years later, the concept of TME has been widely appreciated and known to play an essential role in tumorigenesis and progression[4,5].

1.2 Cancer genomics and new dimensions

Since the landmark publication of the Human Genome Project[6] and the publication of the first cancer genome in 2001[7], cancer genomics has made great strides because of the continuous advances in sequencing technologies and computational tools, establishing it as a mainstay in cancer research. The larger tumor sample collections and more robust statistical methods made it possible for researchers to identify the pathogenic variants and somatic cancer-driving mutations that are less frequent. These efforts have led to the discovery of more driver mutations[8,9], recurrent mutations with patterns of co-occurrences or mutual exclusivity[10,11], molecular subtypes[12], and patterns of clonal evolution[13]. Notably, the Cancer Genome Atlas (TCGA) project genomically characterized over 10,000 patients from more than two dozen cancer types and helped deepen our understanding of the cancer genome[14].

Moving forward, single-cell and single-nucleus sequencing[15,16] are becoming two of the most important technologies in cancer genomics. As the conventional bulk RNA-seq results in averaged gene expression quantification, the single-cell methods dissect the transcriptome at the resolution of individual cells or nuclei. Fan, Slowikowski & Zhang et al[17] outlined several Analyses single-cell and single-nucleus sequencing data can provide for a better understanding of cancer genome[17], including distinguishing neoplastic from non-neoplastic cells using marker genes and snRNA-based somatic mutation calling, copy number inference, and fusion gene detection and identifying common cell type and cell states across patients and disease states.

With these computational analyses developed, single-cell and single-nucleus sequencing are now applied for the studying of cancer subtyping[18], chemotherapy resistance[19], immunotherapy[20], and metastasis[21].

While genomics helped decipher the key genetic changes driving tumorigenesis and cancer progression, mass spectrometry-based protein analysis opened up new dimensions in protein expression profiling, and protein post-translational modifications[22,23]. Moreover, proteomics enabled the identification of protein markers and protein signatures that are useful for tumor early detection, prognosis, classification, and identifying tumors sensitive to particular drugs[24–27]. Proteomics technologies help researchers understand the specific alterations in the signaling pathway of cancer cells, therefore, help improve the understanding of the disease etiology and potentially the therapeutic strategy to target different pathways in cancer[28–31].

The Clinical Proteomic Tumor Analysis Consortium (CPTAC) has published proteomics and phosphoproteomics studies of many cancer types, including colon and rectal, breast, ovarian, lung adenocarcinoma, and renal cancers[32–35,29]. More importantly, these studies were conducted using a multi-omics approach by integrating leveraged genomic, proteomic, and phosphoproteomic data. This multi-omics CPTAC data[35] revealed four immune-based subtypes for clear cell renal cell carcinoma (ccRCC) with unique genomic and microenvironment signatures and signaling pathways that could be used to predict therapeutic response. The CPTAC lung adenocarcinoma study revealed new molecular subtypes and therapeutic opportunities associated with KRAS, EGFR, and ALK mutations[29]. As the CPTAC continues to generate data for an expanding number of samples, its publically available data serves as a great resource for the research community.

1.3 Clear cell renal cell carcinoma

Renal cell carcinoma (RCC) is a type of cancer that originates from the renal epithelium and accounts for over 90% of cancers in the kidney[36]. It includes several subtypes and clear cell RCC (ccRCC) is the most common subtype (75%). The name "clear cell" denotes its distinct morphological feature that the cytoplasm is filled with lipids and glycogen, which are dissolved during histological processing, leading to a clear cytoplasm surrounded by a distinct cell membrane[37]. Kidney cancer, as a whole, accounts for 2% of all cancer diagnoses and cancer deaths worldwide[38]. RCC incidence is higher for men than for women (a 2:1 ratio of new diagnoses) and increases markedly with age[36]. Ethnicity and genetic factors also contribute to RCC risk. In the United States, different ethnic groups have varied incidence, with rates highest in Native Americans, Indigenous Alaskans, and African Americans and lowest in Asian Americans and people of Pacific Island descent[39]. Hereditary kidney cancer syndrome accounts for 3 - 5% of all RCCs[40]. And other major established risk factors for RCC include excess body weight, hypertension, and cigarette smoking[41].

Genes and pathways altered in ccRCC. Genes and pathways that are involved in ccRCC include the VHL-HIF pathway, epigenetic regulators and chromatin remodeling genes, and the PI3K-AKT-mTOR signaling pathway. The VHL tumor suppressor gene is the most frequently mutated gene in ccRCC. It is affected by the loss of heterozygosity (LOH) at chromosome 3p, which occurs in > 90% of ccRCC tumors, and somatic mutation or epigenetic silencing, which occurs in >80% of these tumors[42–44]. Complete loss of VHL constitutes the earliest truncal driver event in ccRCC[45,46].

The VHL-HIF pathway. pVHL controls the protein levels of the HIF transcription factors (including two isoforms HIF-1a and HIF-2a) in an oxygen-dependent manner[47]. Under normal oxygen tension, HIF-a is hydroxylated at two conserved proline residues[48], a process mediated by pVHL, and results in its rapid degradation[47]. Under inadequate oxygen tension or hypoxia, HIF-a hydroxylation does not occur, which stabilizes HIF-a and promotes binding to HIF-b instead of pVHL. Inactivating VHL mutations also result in the stabilization of HIF-a and formation of a HIF-a- HIF-b complex, leading to the increased transcription of downstream targets involved in angiogenesis, glycolysis, cell proliferation, invasion, and metastasis[49].

Intratumor heterogeneity. As the tumor progresses, the genomes of tumor cells often become increasingly unstable. When the new mutant alleles are generated at a much higher rate than the elimination rate for phenotypically less-fit clones, the tumor mass may develop an increasing number of distinct sectors, each dominated by a genetically distinct subclone. Such intratumor heterogeneity (ITH), as Nowell first described 46 years ago[50], is thought to enable tumors to adapt to metabolic demands within themselves and pressures from the microenvironment. The extent of ITH in ccRCC has been demonstrated by several bulk-based studies. In a multi-region genetic analysis of four patients with sporadic primary ccRCC, the tumors were characterized by a small number of truncal or clonal alterations[51]. Only VHL mutation and chromosome 3p LOH were found to be ubiquitous events across all regions sampled. Most somatic alterations were found to be subclonal and present in varying degrees throughout the tumor. Frequent driver events, such as PBRM1, SETD2, KDM5C, MTOR, PIK3CA and PTEN mutations, were found in some regions but not others within the primary tumor and metastatic sites. In a follow-up

study of eight ccRCC patients, ITH was found in all cases and 73%-75% of driver alterations were found to be subclonal[52].

ccRCC Treatment. The standard of care for RCC patients with surgically resectable tumors has been surgical excision by either partial or radical nephrectomy. Approximately 20-30% of patients present regional or distant metastases and an additional 20% will relapse in distant sites after radical surgical treatment[53,54]. Those with metastatic or inoperable RCC typically were treated with targeted therapy and/or immune checkpoint inhibitors, as RCC are generally not very responsive to chemotherapy (response rate 6%) and radiotherapy[37].

Targeted therapy for ccRCC. The major classes of targeted therapy for ccRCC include VEGF ligand antibody (bevacizumab), tyrosine kinase inhibitors, and mTOR inhibitors, while the immunotherapy agent is a programmed death-1 inhibitor (nivolumab). The first-line treatment includes tyrosine kinase inhibitors sunitinib, pazopanib, VEGF ligand antibody bevacizumab, and mTOR inhibitor temsirolimus, and the rest are second-line or later treatment[55]. Although all of these drugs have been proven effective to some extent, a big proportion of patients display intrinsic resistance to targeted therapies[56], and those who do not initially will acquire resistance to these treatments later on[56,57].

Preventing drug resistance. The fact that a large proportion of ccRCC patients will develop resistance to single-agent treatment leads to the realization that combinational therapy will be necessary to overcome such resistance. Clinical experiences and studies of model systems have shown combinational drug regimen for RCC is superior to the single-agent regimens[58]. However,

avoiding higher toxicities (including on-target and off-target effects) is needed while pursuing higher efficacy. For instance, the sunitinib plus everolimus combination produced severe toxicity when treating metastatic RCC[59]. In contrast, bevacizumab plus everolimus is well tolerated and has been shown to be efficacious in treating non-ccRCC with papillary features[60]. The success of combinational drug treatment relies on efficient and correct targeting of both primary and secondary pathways. In ccRCC, VEGF is the primary pathway due to the universal VHL loss; secondary targets can include C-MET and the mammalian target of rapamycin complex 1 (mTORC1), given the available clinical[61–64] and preclinical studies[65–68].

1.4 References

1. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
2. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
3. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **12**, 31–46 (2022).
4. De Palma, M., Biziato, D. & Petrova, T. V. Microenvironmental regulation of tumour angiogenesis. *Nat. Rev. Cancer* **17**, 457–474 (2017).
5. Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**, 541–550 (2018).
6. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
7. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
8. Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
9. Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* **360**, 765–773 (2009).
10. Barbieri, C. E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
11. Gainor, J. F. *et al.* ALK rearrangements are mutually exclusive with mutations in EGFR or KRAS: an analysis of 1,683 patients with non-small cell lung cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **19**, 4273–4281 (2013).
12. Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).
13. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
14. Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
15. Eberwine, J., Sul, J.-Y., Bartfai, T. & Kim, J. The promise of single-cell sequencing. *Nat. Methods* **11**, 25–27 (2014).
16. Wu, H., Kirita, Y., Donnelly, E. L. & Humphreys, B. D. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J. Am. Soc. Nephrol. JASN* **30**, 23–32 (2019).
17. Fan, J., Slowikowski, K. & Zhang, F. Single-cell transcriptomics in cancer: computational challenges and opportunities. *Exp. Mol. Med.* **52**, 1452–1465 (2020).
18. Gan, Y., Li, N., Zou, G., Xin, Y. & Guan, J. Identification of cancer subtypes from single-cell RNA-seq data using a consensus clustering method. *BMC Med. Genomics* **11**, 117 (2018).
19. Sharma, A. *et al.* Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. *Nat. Commun.* **9**, 4931 (2018).
20. Zhang, L. & Zhang, Z. Recharacterizing Tumor-Infiltrating Lymphocytes by Single-Cell RNA Sequencing. *Cancer Immunol. Res.* **7**, 1040–1046 (2019).
21. Lawson, D. A., Kessenbrock, K., Davis, R. T., Pervolarakis, N. & Werb, Z. Tumour heterogeneity and metastasis at single-cell resolution. *Nat. Cell Biol.* **20**, 1349–1360 (2018).
22. Pandey, A. & Mann, M. Proteomics to study genes and genomes. *Nature* **405**, 837–846 (2000).
23. Hyung, S.-J. & Ruotolo, B. T. Integrating mass spectrometry of intact protein complexes

into structural proteomics. *Proteomics* **12**, 1547–1564 (2012).
24. Hanash, S. & Taguchi, A. Application of proteomics to cancer early detection. *Cancer J. Sudbury Mass* **17**, 423–428 (2011).
25. Yadav, M. *et al.* Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515**, 572–576 (2014).
26. Chen, F., Chandrashekar, D. S., Varambally, S. & Creighton, C. J. Pan-cancer molecular subtypes revealed by mass-spectrometry-based proteomic characterization of more than 500 human cancers. *Nat. Commun.* **10**, 5679 (2019).
27. Enroth, S. *et al.* High throughput proteomics identifies a high-accuracy 11 plasma protein biomarker signature for ovarian cancer. *Commun. Biol.* **2**, 221 (2019).
28. Chang, L. *et al.* Identification of protein biomarkers and signaling pathways associated with prostate cancer radioresistance using label-free LC-MS/MS proteomic approach. *Sci. Rep.* **7**, 41834 (2017).
29. Gillette, M. A. *et al.* Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. *Cell* **182**, 200-225.e35 (2020).
30. Nanjundan, M. *et al.* Proteomic profiling identifies pathways dysregulated in non-small cell lung cancer and an inverse association of AMPK and adhesion pathways with recurrence. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* **5**, 1894–1904 (2010).
31. Shruthi, B. S., Vinodhkumar, P. & Selvamani, null. Proteomics: A new perspective for cancer. *Adv. Biomed. Res.* **5**, 67 (2016).
32. Zhang, H. *et al.* Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **166**, 755–765 (2016).
33. Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).
34. Vasaikar, S. *et al.* Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* **177**, 1035-1049.e19 (2019).
35. Clark, D. J. *et al.* Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* **179**, 964-983.e31 (2019).
36. Hsieh, J. J. *et al.* Renal cell carcinoma. *Nat. Rev. Dis. Primer* **3**, 17009 (2017).
37. Frew, I. J. & Moch, H. A clearer view of the molecular complexity of clear cell renal cell carcinoma. *Annu. Rev. Pathol.* **10**, 263–289 (2015).
38. Cancer, I. A. for R. on. GLOBOCAN 2012 v1. 0, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11. *Lyon Available Httpglobocan Iarc Fr Accessed* **6**, (2017).
39. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA. Cancer J. Clin.* **66**, 7–30 (2016).
40. Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E. & Ulbright, T. M. The 2016 WHO classification of tumours of the urinary system and male genital organs—part A: renal, penile, and testicular tumours. *Eur. Urol.* **70**, 93–105 (2016).
41. Schottenfeld, D. & Fraumeni Jr, J. F. *Cancer epidemiology and prevention*. (Oxford University Press, 2006).
42. Gnarra, J. R. *et al.* Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nat. Genet.* **7**, 85–90 (1994).
43. Banks, R. E. *et al.* Genetic and epigenetic analysis of von Hippel-Lindau (VHL) gene alterations and relationship with clinical variables in sporadic renal cancer. *Cancer Res.* **66**, 2000–2011 (2006).

44. Le, M. *et al.* Von Hippel-Lindau (VHL) inactivation in sporadic clear cell renal cancer: associations with germline VHL polymorphisms and etiologic risk factors. *PLoS Genet.* **7**, (2011).
45. Hakimi, A. A., Pham, C. G. & Hsieh, J. J. A clear picture of renal cell carcinoma. *Nat. Genet.* **45**, 849–850 (2013).
46. Linehan, W. M., Srinivasan, R. & Schmidt, L. S. The genetic basis of kidney cancer: a metabolic disease. *Nat. Rev. Urol.* **7**, 277–285 (2010).
47. Gossage, L., Eisen, T. & Maher, E. R. VHL, the story of a tumour suppressor gene. *Nat. Rev. Cancer* **15**, 55–64 (2015).
48. Chan, D. A., Sutphin, P. D., Yen, S.-E. & Giaccia, A. J. Coordinate regulation of the oxygen-dependent degradation domains of hypoxia-inducible factor 1 alpha. *Mol. Cell. Biol.* **25**, 6415–6426 (2005).
49. Harris, A. L. Hypoxia--a key regulatory factor in tumour growth. *Nat. Rev. Cancer* **2**, 38–47 (2002).
50. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
51. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
52. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).
53. Athar, U. & Gentile, T. C. Treatment options for metastatic renal cell carcinoma: a review. *Can. J. Urol.* **15**, 3954–3966 (2008).
54. Gupta, K., Miller, J. D., Li, J. Z., Russell, M. W. & Charbonneau, C. Epidemiologic and socioeconomic burden of metastatic renal cell carcinoma (mRCC): a literature review. *Cancer Treat. Rev.* **34**, 193–205 (2008).
55. Choueiri, T. K. & Motzer, R. J. Systemic Therapy for Metastatic Renal-Cell Carcinoma. *N. Engl. J. Med.* **376**, 354–366 (2017).
56. Rini, B. I., Campbell, S. C. & Escudier, B. Renal cell carcinoma. *The Lancet* **373**, 1119–1132 (2009).
57. Singer, E. A., Gupta, G. N. & Srinivasan, R. Update on targeted therapies for clear cell renal cell carcinoma. *Curr. Opin. Oncol.* **23**, 283–289 (2011).
58. Hsieh, J. J. *et al.* Overcome tumor heterogeneity-imposed therapeutic barriers through convergent genomic biomarker discovery: A braided cancer river model of kidney cancer. *Semin. Cell Dev. Biol.* **64**, 98–106 (2017).
59. Molina, A. M. *et al.* Phase 1 trial of everolimus plus sunitinib in patients with metastatic renal cell carcinoma. *Cancer* **118**, 1868–1876 (2012).
60. Voss, M. H. *et al.* Phase II Trial and Correlative Genomic Analysis of Everolimus Plus Bevacizumab in Advanced Non-Clear Cell Renal Cell Carcinoma. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **34**, 3846–3853 (2016).
61. Motzer, R. J. *et al.* Lenvatinib, everolimus, and the combination in patients with metastatic renal cell carcinoma: a randomised, phase 2, open-label, multicentre trial. *Lancet Oncol.* **16**, 1473–1482 (2015).
62. Choueiri, T. K. *et al.* Cabozantinib versus Everolimus in Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* **373**, 1814–1823 (2015).
63. Powles, T. *et al.* Randomized Open-Label Phase II Trial of Apitolisib (GDC-0980), a Novel Inhibitor of the PI3K/Mammalian Target of Rapamycin Pathway, Versus Everolimus in Patients With Metastatic Renal Cell Carcinoma. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.*

34, 1660–1668 (2016).
64. Carlo, M. I. *et al.* A Phase Ib Study of BEZ235, a Dual Inhibitor of Phosphatidylinositol 3-Kinase (PI3K) and Mammalian Target of Rapamycin (mTOR), in Patients With Advanced Renal Cell Carcinoma. *The Oncologist* **21**, 787–788 (2016).
65. Huang, D. *et al.* Interleukin-8 mediates resistance to antiangiogenic agent sunitinib in renal cell carcinoma. *Cancer Res.* **70**, 1063–1071 (2010).
66. Bergers, G. & Hanahan, D. Modes of resistance to anti-angiogenic therapy. *Nat. Rev. Cancer* **8**, 592–603 (2008).
67. Nakaigawa, N. *et al.* Inactivation of von Hippel-Lindau gene induces constitutive phosphorylation of MET protein in clear cell renal carcinoma. *Cancer Res.* **66**, 3699–3705 (2006).
68. Mizukami, Y. *et al.* Induction of interleukin-8 preserves the angiogenic response in HIF-1alpha-deficient colon cancer cells. *Nat. Med.* **11**, 992–997 (2005).

Chapter 2: Epigenetic and transcriptomic characterization reveals progression markers and essential pathways in clear cell renal cell carcinoma

This chapter is adapted from a manuscript in revision. Contribution: I lead the project design, sample selection, sample request, coordinate data generation for snRNA-seq data, processing snRNA-seq data, downstream analysis for snRNA-seq and snATAC-seq data, manuscript writing, and figure generation as a first-author of the paper. In particular, analysis relating to snATAC-seq data was done in collaboration with Nadezhda Terekhanova. Details of other colleagues' contribution can be seen in the "Author Contributions" section.

2.1 Summary

To define the epigenetic and transcriptomic regulation of clear cell renal cell carcinoma (ccRCC), we performed snRNA-seq and snATAC-seq in 34 and 28 human RCC specimens respectively, and matched them with bulk proteogenomics data. Here, we show that tumor clusters identified by snRNA-seq exhibit a continuum of epithelial to mesenchymal features; clusters with low omics-derived epithelial scores and high EMT scores display upregulated expression and accessibility of *TGFBI* and enriched transcription factor motifs for TWIST1. We identified tumor-specific markers by a newly developed single-cell tiered approach coupled with multi-omics validation. Among the top 20 tumor-specific markers, Ceruloplasmin (CP) levels in tumor cells negatively correlate with patient survival. Inhibitors against CP and other markers we identified reduced the survival of tumor cells. Spatial transcriptomics analysis shows high *CP* expression is associated with mesenchymal histology, while high *PCSK6* expression is linked to epithelial features. Dysregulation of 13 genes in the glycolysis pathway at the epigenetic and/or

transcriptomic levels underscores the metabolic abnormalities in ccRCC. *BAP1* mutations show a moderate reduction of chromatin accessibility in a wide spectrum of genes while *PBRM*1 mutations generally increase chromatin accessibility and expression of a selective set of genes. These integrated multi-omics analyses reveal the cellular architecture of ccRCC, providing new insights into key markers and pathways in ccRCC tumorigenesis and suggesting new potential therapeutic targets.

2.2 Introduction

Clear cell renal cell carcinoma (ccRCC) is the most common form of kidney cancer, comprising roughly 80% of cases. Prevailing molecular alterations of ccRCC are 3p alterations, with concomitant loss of *VHL* functionality, and mutations in epigenetic regulators and chromatin remodelers, including *BAP1, PBRM1*, *SETD2,* and *KDM5C*[1]. Many aspects of epigenetic and transcriptomic regulations within ccRCC have been investigated, though few studies have examined the same cohorts[2–4]. And the consequences of *BAP1* and *PBRM1* mutations on chromatin accessibility and downstream transcriptional networks remain largely unknown. ccRCC is also known as a metabolic disease, accompanied by reprogramming of glucose and fatty acid metabolism[5–9]. Studies using genomic[1], proteomic[5,6,10], and metabolomic[11,12] profiling also uncovered a metabolic shift in aggressive ccRCCs that involves the tricarboxylic acid cycle (TCA), pentose phosphate, and phosphoinositide 3-kinase pathways. However, the current understanding of the transcriptional regulation of important metabolic enzymes is incomplete. Although there is a discordance between metabolome and transcriptome for certain pathways, a careful examination might elucidate additional disease mechanisms and potential therapeutic targets. Aside from these regulation and remodeling dynamics, ccRCC is known for its

substantial genetic heterogeneity, parallel evolution of subclones[13], and abundant genetic alterations, as revealed by bulk sequencing-based studies[14,15]. We still do not completely understand how these properties arise and whether there is also a high level of heterogeneity in the transcriptome and chromatin accessibility of ccRCC cells.

Single-cell RNA-seq (scRNA-seq) studies of relatively small numbers of ccRCC samples have been reported, shedding light on the molecular attributes of cells of origin[16,17] and the tumor microenvironment. A few studies have also identified tumor markers[16,18] or therapeutic targets[19]. Important aspects remain largely unexplored, including tumor markers of sufficient strength within and across samples, the degree of the transcriptional heterogeneity within the tumor cells, and the common features separating the tumor subpopulations. In addition, scRNA-seq has been limited by the availability of fresh tissue, especially since clinical samples are normally cryo-preserved[20]. Single nucleus RNA-seq (snRNA-seq) can analyze frozen specimens and avoids the cell dissociation process that promotes stress-related alterations[21]. To our knowledge, no snRNA-seq nor snATAC-seq studies of ccRCC samples have yet been reported. Bulk epigenomic profiling data (e.g. ATAC-seq and DNase-seq) for ccRCC have shown some ccRCC-specific regulatory elements and TFs[2,4,22] and the chromatin accessibility changes associated with *SETD2* mutation[23]. However, the transcription factors having significantly higher accessibility in ccRCC genomes than in non-tumor cells and which genes are regulated by these TFs remain largely unknown. Furthermore, there has not yet been adequate integration of snRNA-seq, snATAC-seq, bulk omics (including proteomics), and spatial transcriptomics (ST) to comprehensively investigate the epigenetic and transcriptomic landscape of ccRCC.

In this study, we bring these aspects together in the broadest multi-omics analysis on ccRCC specimens yet undertaken. We identify key genes and pathways specifically altered epigenetically and transcriptionally in tumor cells, finding correlations between these alterations and various ccRCC features. We performed spatial transcriptomics, discovering intriguing patterns of the spatial distribution of gene expression and histopathological features of ccRCC. Finally, we dissected chromatin accessibility changes associated with *BAP1* and *PBRM1* mutations, further illustrating the multi-level interplay between mutational, global, and specific epigenetic alterations and transcriptomic changes in ccRCC.

2.3 Results

2.3.1 Overview of clinical features and datasets

We performed snRNA-seq on 34 samples (25 patients) and matched snATAC-seq on 28 of these samples (24 patients) from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) ccRCC collection[10] (Supplementary Table 1). These were procured from the same pool of pulverized powder that has previously produced WES, bulk RNA-seq, and proteomics data for these samples (Fig. 1a, Extended Data Fig. 1a). Using the same pool of pulverized sample powder ensures the highest level of comparability among datasets from different platforms and enables tight integration of diverse omics data sets. We also performed spatial transcriptomics (ST) on 2 patient tumor samples and 1 patient-derived xenograft (PDX) sample collected in-house using FFPE Visium ST technology (10x Genomics).

For the snRNA-seq data, we obtained 141,950 nuclei from 34 samples, comprising 30 primary tumor samples and 4 normal adjacent tissue (NAT) samples (Fig. 1b). These samples had distinct

tumor, immune, and stromal populations based on canonical markers curated from the literature (Methods, Supplementary Table 2, Extended Data Fig. 1c). The snRNA-seq data indicated tumor cell content averaged 71% per sample, which correlated strongly with the bulk mRNA data estimate (Pearson's $R = 0.72$, $p < 0.001$, Extended Data Fig. 1b). We also generated snATAC-seq data for 211,497 nuclei from 24 of these same tumors and 4 NATs (Fig. 1a). We detected peaks of accessible chromatin in snATAC data across all samples, ranging from 86K to 220K instances per sample. As expected, the majority of peaks appeared in intronic and intergenic regions, while an average of 24K peaks were located in gene promoter regions (Extended Data Fig. 1e). snRNA and snATAC paired samples yielded comparable cell type content estimates (Extended Data Fig. 8b; Pearson's $R=0.77$, $P < 0.0001$)

2.3.2 Single-cell-based ccRCC tumor marker discovery and epigenetic regulation of tumor markers

Although bulk sequencing studies have reported markers altered between ccRCC and adjacent normal tissue that presumably reflect the changes in tumor cells[1], these studies were limited by the confounding effects from non-tumor cells or had limited discovery power for subpopulations of tumor cells expressing unique markers. As only a few canonical tumor-cell markers such as CA9 (carbonic anhydrase 9) have been widely used in ccRCC, we aim to identify additional markers that show higher specificity to ccRCC tumor cells than established markers that may also have prognostic/diagnostic values or potential to become therapeutic targets. Here we leveraged a 4-stage process to identify bona fide markers for ccRCC (Extended Data Fig. 2a): (1) compare expression levels across cell types strictly within samples to discern markers characteristic of tumor cells, identifying those that hold more generally across our 30 ccRCC samples, (2) narrow to those exclusive of proximal tubule (PT) cells and epithelial cell types (as

they were scarce in tumor samples), (3) confirm their chromatin accessibility changes using snATAC-seq, and (4) validate in a larger cohort from bulk RNA and protein data, and further characterize using spatial transcriptome. Using this approach, we identified 131 candidate surface markers overexpressed in tumor cells compared to all the other cell types in a majority of individual samples (step 1; Supplementary Table 3), prioritizing 20 that were also overexpressed in ccRCC cells compared to normal proximal tubule cells and other epithelial cell types (step 2; Fig. 1c, adding the canonical ccRCC marker *CA9*), thereby bolstering specificity to ccRCC. 19 of these markers showed higher chromatin accessibility (gene activity, fold change > 1) in tumor cells using snATAC-seq data, suggesting higher chromatin accessibility may contribute to their higher expression in tumor cells (step 3). 17 were further supported by the bulk RNA-seq and proteomics data, by comparing the tumors to the normal adjacent tissues (step 4; Fig. 1c). Of the 20 markers, 7 have not been previously associated with ccRCC, including *SNAP25*, *PHKA2*, *EPHA6*, *ABLIM3*, *SHISA9, PCSK6*, *UBE2D2*. 4 have been associated with ccRCC but have not been shown to be expressed in ccRCC tumor cells, including *ABCC3*[24], *KCTD3*[25], *SEMA6A*[26], *PLEKHA1*[27]. 9 have been shown to be expressed in ccRCC tumor cells, including *TGFA*[28], *PLIN2*[29], *FTO*[30], *SLC6A3*[31], *NDRG1*[32], *CP*[33], *EGFR*[34], *ENPP3*[35], *COL23A1*[36]. However, none of the studies have carefully compared the expression of these markers in tumor cells to non-tumor cell types.

We investigated correlations between expression levels of these markers and clinical features of the tumors using snRNA-seq data. *CP*, *PCSK6*, *UBE2D2*, *MGST1*, *NDRG1*, and *KCTD3* tumor-cell expression levels were associated with higher tumor grades, results which were further supported by the bulk RNA-seq and protein data from the larger ccRCC cohort (n = 103)

(Extended Data Fig. 2). Ceruloplasmin (CP) is particularly interesting because CP is specifically expressed in tumor cells (Fig. 1d) and high CP expression in ccRCC cells is associated with shorter survival in this study cohort and the larger ccRCC CPTAC cohort with bulk RNA-seq data (Fig. 1e-f). CP helps stabilize and transport HIF1A to the nucleus[37], thus promoting VEGF expression and angiogenesis[38,39]. Consequently, elevated CP expression may benefit ccRCC cells by enhancing HIF signaling. Due to the lack of detection for HIF1A bulk protein levels (n = 8 out of 103), we looked at the association between VEGFA and CP protein levels, finding a significant positive correlation, which further supports CP's regulation of the HIF1A/VEGF axis (Extended Data Fig. 2e, $R = 0.4$, $P < 0.0001$).

We asked whether differences in chromatin accessibility might be responsible for higher expression of ccRCC markers in tumor vs proximal tubule cells. We found that *CP*, *UBE2D2*, and *CA9* promoter regions were more accessible in tumor cells (Fig. 1g, Supplementary Table 3). We also found motifs for HIF1A, ARNT, NFKB1, RELA, MXI1, KLF9, and SREBF2 transcription factors, which were significantly enriched in tumor cells using snATAC-seq (Supplementary Table 3), in differentially accessible promoter regions of *CP*, *CA9*, and *UBE2D2*. As *MXI1* expression itself is up-regulated and displays higher accessibility binding in ccRCC cells, it is likely an active transcription factor in ccRCC cells. We also found that MXI1 may regulate many other genes up-regulated in ccRCC, such as *GATM*, *PKM*, *OBSCN*, *TRIB3*, and *SERPINE1* (fold change > 2) based on the presence of MXI1 motifs in their promoter regions (Supplementary Table 3), the latter two being respectively implicated as master regulators of ccRCC pathogenesis[4] and associated with poor prognosis in ccRCC. MXI1 appears to have an important role in ccRCC progression.

2.3.3 Spatial characterization of tumor markers identified by single-cell omics

We generated immunofluorescence (IF) and spatial transcriptomics (ST) data to validate and further characterize the spatial distribution of novel tumor markers discovered by snRNA-seq. Specifically, we investigated protein level and spatial expression in 2 ccRCC human tumors (282 and 293) and 2 ccRCC patient-derived xenografts (PDX) (RESL5 and RESL10) samples. Tumor cell-specific *CP* expression (tumor cells vs. non-tumor cells fold change = 3.6) was validated by IF staining of ceruloplasmin and canonical ccRCC tumor-cell marker CA9 and the co-localization of the two proteins in ccRCC PDX model sections (Fig. 2b, ID: RESL5). We also validated the tumor cell expression of PCSK6 (Fig. 2a-b).

To understand the heterogeneous and spatial distribution of tumor markers, as well as their potential impact on cellular morphological features, we utilized ST expression to validate selected tumor markers, namely *CP* (Fig. 2c), *CA9*, *PCSK6*, *ABCC3*, *MGST1*, and *NDRG1* (Fig. 2d) across 2 human tumor samples (ID: 282, 293) and 1 PDX tumor (ID: RESL10). We found *CP* expression exhibited a spatially-dependent enrichment pattern in both human tumor samples (Fig. 2c). In human tumor 293, *CP* expression was enriched in an area showing a relative sparsity of tumor cells embedded in an abundant background of hyalinized stroma (Fig. 2c, location A) compared to the rest of the tumor (Fig. 2c, location B). In tumor 282, while lacking areas of complete regression, there was a focus of the tumor showing a higher hyalinization-to-cell ratio than in the rest of the tumor, indicating partial regressive changes. *CP* showed greater concentration in these areas (Fig. 2c, location C), compared to the rest of the tumor (Fig. 2c, location D). These observations suggest a role of CP, as a secreted protein, in mediating tumor-

stroma interactions and the hyalinization of the microenvironment. This pattern was not as prominent in the PDX (RESL10) sample, perhaps due to the unique tumor microenvironment in the PDX mouse compared to the original human tissue.

In addition to *CP*, we also validated the pattern of canonical ccRCC tumor marker *CA9*, as well as 4 other novel tumor markers. We observed various degrees of intra-tumor heterogeneity in spatial expression for these 4 specific markers. *PCSK6* was the highest in human tumor 293 among these tumor samples. Expression of *ABCC3* was highest in PDX sample RESL10 and lowest in human tumor 282. Expressions of novel markers *MGST1* and *NDRG1* were relatively homogenous across all 3 tumor samples, with *MGST1* exhibiting a slightly higher intra-tumoral, spatial-dependent expression pattern enriched in the same region *PCSK6* was highest in sample 293. Indeed, such inter-tumor heterogeneous expression patterns can also be observed via the scaled average expression of all tumor markers using ST and snRNA-seq among all 3 tumor samples (Fig. 2e). Of note, mesenchymal and proximal tubular (PT) epithelial cell markers indicate overall higher epithelial gene expression in both human tumor samples and a yet stronger mesenchymal expression profile in the PDX sample. The lower expressions of some tumor markers, such as *COL23A1* or *SEMA6A* in both ST and snRNA-seq, could indicate their different roles in these 2 tumor states..

We next set out to test if some of the tumor-cell markers are good treatment targets for ccRCC. We prioritized 3 druggable targets, including ABCC3, CP, and SLC6A3, and treated 4 RCC cell lines and 2 control cell lines with 3 inhibitors against these markers. The control cell lines include an HK-2 adult proximal tubule cell line and HEK-293. The drugs include MK571

(targets ABCC3), Benztropine (targets SLC6A3), and Tetrathiomolybdate, which is a copper-lowering agent that therefore interrupts CP function. The bulk gene expressions of these target genes were higher in the cancer cell lines versus non-cancer controls (Extended Data Fig. 3a). Each drug produced stronger inhibition of growth of 2-4 RCC lines compared to the control line HK-2. Fig. 2f depicts 4 cancer cell lines that showed higher sensitivity compared to the control cell line HK-2 (full results in Extended Data Fig. 3). For example, we estimated the respective IC50 values for MKI571 in RCC-4, RCC-4-VHL+, 786-O, and 786-O-VHL+ lines to be 68, 90, 134, 196 **μ**M, which are much lower than the IC50 of 246 **μ**M for the control HK-2 (Fig. 2f). Both RCC-4 and 786-O were more sensitive to MK571 than their VHL-reconstituted counterparts, suggesting VHL deficiency might increase the cell line sensitivity to MKI571. Nonetheless, the IC50 for the control HEK-293 for MK571 is 64 **μ**M, which is lower than that of the cancer cell lines. We also found RCC-4, RCC4-VHL+, 786-O, and 786-O-VHL+ cells to be more sensitive to tetrathiomolybdate than the control HK-2. Our results suggest these three markers could play important roles in ccRCC development and may serve as potential therapeutic targets for future preclinical studies.

2.3.4 New transcription factors mediating high expression of glycolytic genes in tumor cells compared to normal proximal tubule cells

To ascertain the cell type in the normal adjacent tissue most similar to the ccRCC cells, we examined transcription factor (TF) motif enrichment in different epithelial cell types based on TF motif binding accessibility (Methods). Among the 6 nephron epithelial cell types examined, we found tumor cells had the strongest correlations with proximal tubule cells in TF binding accessibility (Extended Data Fig. 4a). This was supported by the correlation analysis among different epithelial cell types using snRNA-seq data (Extended Data Fig. 4a). Our data thus have

gone beyond gene expression and mutational analyses to use similarity in epigenetic regulation between tumor cells and PT cells to support the hypothesis that ccRCC derives from PT cells[16,40].

To identify differentially enriched TF motifs in tumor cells, we compared the TF binding accessibilities between tumor cells from each tumor and combined PT cells from the 4 NATs (Methods, Supplementary Table 3). 23 TF motifs showed consistently higher accessibility (tumor cells vs PT cells) in over half of the tumor samples (Fig. 3a). 16 TF motifs, including HIF1A/ARNT, NF-κB TFs (NFKB1, NFKB2, REL, RELA), RBPJ, MXI1, KLF9, ZNF75D, HSF2, NEUROD1, SREBF2, NEUROG2, RREB1, and TBXT showed higher accessibility (tumor cells vs PT cells) in all tumor samples (Fig. 3a). High HIF1A motif accessibility is consistent with the activation of HIF1A downstream transcriptional programs associated with *VHL* loss[41]. 14 of the top ccRCC-specific TF motifs (except for SREBF2 and TBXT) were reported to be ccRCC-specific in previously published bulk pan-cancer ATAC-seq data[42] (Extended Data Fig. 4b, Methods). The absence of SREBF2 and TBXT in the published data could be the result of lower detection sensitivity when pooling different cell types together. In addition, we found that the expression of *MXI1*, *RBPJ*, and *NFKB1* was significantly up-regulated in ccRCC cells using snRNA-seq data, and the up-regulation of RBPJ and NFKB1 was further confirmed by bulk protein data (Supplementary Table 3), suggesting the activity of these TFs in ccRCC is not only enhanced by increased binding accessibility but also increased TF abundance. Finally, HNF and RXR family TFs, which were more enriched in PT cells compared to ccRCC cells (Fig. 3a), were previously associated with ccRCC by bulk ATAC analysis[2] (Extended Data Fig. 4c). These results highlight the important utility of snATAC-seq in

discerning motifs specific to tumor cells as bulk ATAC analysis may confuse TFs specific to the normal PT cells with TFs specific to tumor cells.

To further investigate genes regulated by the ccRCC-specific TFs, we identified 1,161 overexpressed and 171 down-regulated genes in ccRCC tumor cells in comparison to PT (Fig. 3b; Methods, Supplementary Table 3). The tumor-cell-overexpressing genes include genes in glycolysis (*PFKP, PKM*), hypoxia (*VEGFA, ANGPTL4*), and solute carrier (SLC) transporter disorder pathways (*SLC6A3, SLC9A9*; Fig. 3c), and known ccRCC markers, such as *CA9*. On the other hand, the genes down-regulated in tumor cells were concentrated in SLC-mediated transmembrane transport (*SLC13A3, SLC47A3*), late response to estrogen (*SLC27A2, SLC22A5*), and extracellular matrix proteins (*ITGA6, COL4A2*).

ccRCC is characteristic of the Warburg effect and is known to have glycogen and lipid accumulation[43]. Therefore, to understand the transcriptional regulation mediating the phenotypic transition from normal PT cells to ccRCC cells, we further focused on glycolysis, TCA cycle, and glycogen and fatty acid synthesis pathways. A majority of the glycolysis enzymes were overexpressed in ccRCC cells, as were *PDK1* and *PDK4*, which inhibit the conversion of pyruvate to acetyl-CoA in the mitochondria. It is worth noting that several gluconeogenesis enzymes, including fructose-1,6-bisphosphatase (*FBP1*), phosphoenolpyruvate carboxykinase (*PCK1*), and pyruvate carboxylase (*PC*), were down-regulated in ccRCC cells (Supplementary Table 3). Together, they shunt pyruvate away from the mitochondria and thus reduce the flux through the TCA cycle. These results suggest glycolysis activation may mediate the transition from normal PT cells to ccRCC cells. We also observed up-regulated gene expression for *ACLY*,

PYGL, and *GBE1*, which are required for the synthesis of fatty acids or glycogen. Among the glycolysis enzymes, *HK1*, *HK2*, *PFKP*, *TPI1*, *ENO2*, and *PKM* showed significantly increased accessibility in their respective promoter regions in cancer cells (Fig. 3e). We observed motifs of HIF1A in the *HK1, PFKP,* and *ENO2* open promoter regions that are more accessible in tumor cells, consistent with previous reports[44–47] (Fig. 3d, Supplementary Table 3). We additionally identified many other TFs, such as RBPJ, NFKB1/2, KLF9, and MXI1, that may also regulate genes involved in glycolysis (Fig. 3f), as their binding sites were significantly more accessible in ccRCC cells compared to the PT cells. The activation of RBPJ, NFKB1 and MXI1 is supported by significantly increased expression levels in tumor cells compared to PT cells. Our analysis of the snRNA-seq and snATAC-seq data illustrates the up-regulation of many glycolytic genes in ccRCC cells compared to normal PT cells. Our study also identified several new TFs, beyond HIF1A, that may regulate the transcription of these glycolytic genes in ccRCC.

2.3.5 Intratumor signaling heterogeneity revealed by single-cell tumor subclustering

Among the 30 tumor samples, we identified 95 total clusters of tumor cells (at least 50 cells per cluster) with substantial inter-cluster transcriptional differences (Fig. 4a, Methods). Each sample averaged 3 tumor-cell clusters. For those genes and pathways most differentially expressed among intrapatient tumor subclusters, we performed an unbiased search across the MSigDB Hallmark gene sets (Methods) and found cell cycle control, DNA damage repair, epithelial-mesenchymal transition (EMT), and mTORC1 signaling genes over-represented (Fig. 4a, Extended Data Fig. 5a). Some pathways, like EMT, maybe are the first steps for tumor metastasis[48,49], while others such as mTORC1 signaling may contribute to cell growth[50]. Cell cycle control and DNA damage repair pathways are important drivers for cancer progression[51,52]

and their overrepresentation here hints at key roles in producing and amplifying tumor heterogeneity and driving clonal evolution in ccRCC progression.

We further scored the tumor subclusters based on the expression of these pathway genes and identified four major, differentially-enriched gene expression modules, including cell cycle, immune, EMT, and mTOR signaling (Fig. 4a, Supplementary Table 4). 16 tumor subclusters (in 15 tumors), including C3 from sample C3L-00010-T1 (Fig. 4b), were enriched with cell cycle gene set modules, which include genes in the mitotic spindle and G2M checkpoint pathways, E2F targets, as well as DNA repair pathway genes and MYC targets. Based on cell cycle scoring, these tumor clusters appear to have significantly higher fractions of cells in the S phase compared to other tumor clusters (Extended Data Fig. 5b). Recurrent up-regulated genes in these clusters include *ABR*, *MYO9B*, *NIN*, *CNTRL*, and *CAPZB* (Extended Data Fig. 5c), which are involved in centrosome regulation and cytoskeleton organization. Immune expression modules were enriched in 14 tumor subclusters (in 9 tumors), such as C3 from tumor C3N-00733-T2 (Fig. 4b). They showed high overall expression of genes within the inflammatory response, interferon-alpha response, complement and allograft rejection pathways, and KRAS signaling up-regulated genes. Recurrent up-regulated immune-regulated genes in these tumor clusters include anti-phagocytosis signal β2-microglobulin (*B2M*) and other components of the MHC-I complex (*HLA-A* and *HLA-E*) (Extended Data Fig. 5), while their corresponding receptor *LILRB1* is mostly expressed in macrophages (Extended Data Fig. 5e). These observations suggest different tumor subpopulations appear to interact with the tumor microenvironment differently.

Finally, we identified 9 tumor subclusters with high EMT gene expression and 10 subclusters with high mTORC1 signaling gene expression. For example, C4 from tumor C3L-00079-T1 was EMT-module enriched, while C2 from the *MTOR*-mutated tumor C3L-00583-T1 was mTOR enriched (Fig. 4b). We also found that mTORC1 signaling activation was significantly associated with *BAP1* mutation (Fig. 4c, $P < 0.01$ compared to the tumors without *BAP1* or *PBRM1* mutation), a relationship that has been reported in ccRCC[53,54]. EMT is a cellular program essential for embryogenesis, wound healing, and cancer progression[55,56]. In the context of cancer, EMT could increase the tumor-initiating and metastatic potential for tumor cells[119–121]. We found significantly higher EMT activity (maximum score per tumor) in the tumor subclusters of stage III and IV patients compared to stage I/II patients (Fig. 4d, $P < 0.05$), suggesting higher EMT activity may be indicative of RCCs progression.

2.3.6 Tumor subgroups with distinct epithelial and mesenchymal features

To better understand and characterize EMT in ccRCC progression, we built a panel of markers to calculate mesenchymal and epithelial feature scores for 95 tumor-cell clusters and 9 proximal-tubule clusters (Fig. 5a, Methods, Supplementary Table 4). We identified 4 major tumor subgroups, including three subgroups with strong, medium, and low epithelial features, all of which have relatively low mesenchymal features (denoted as Epi-H, Epi-M, and Epi-L tumor clusters, respectively); and one subgroup with outlier mesenchymal feature scores (denoted as EMT tumor clusters). The continuum of epithelial and mesenchymal features across tumor clusters was further supported by the chromatin-accessibility-based gene activities derived from snATAC-seq data. Specifically, five Epi-H tumor clusters with high epithelial gene expression scores also showed high epithelial gene activity scores and two EMT tumor clusters behaved

similarly (Fig. 5a). Regardless of the epithelial/mesenchymal feature scores, the tumor subgroups overall showed higher expression of tumor-cell markers, such as *CA9*, *CP*, and *PCSK6*, while PT clusters overall showed higher proximal tubule markers, such as *CUBN*, *GLYAT*, and *LRP2*. We further characterized the PT segmental identities of these clusters according to the commonly used PT S1/2 and S3 classification. Most of the PT clusters did not fit into the clear-cut S1/2 and S3 classifications. Only a small subset of the tumor clusters (mostly Epi-H, and Epi-M tumor clusters) showed enriched PT S3 and/or S1/2 features, and most of the tumor clusters cannot be classified squarely into PT S1/2 or S3. Matters are similar for gene activity, as inferred from snATAC-seq. Tumor subgroups showed higher accessibility of tumor cell markers, while PT clusters were higher for their respective markers. Tumor clusters were likewise sporadic for S1/2 and S3 classifications. With the exception of sporadic clusters for *VIM* and a few for *TGFBI*, mesenchymal markers were not highly expressed across the epithelial subgroups. These results indicate that ccRCC tumor cells may originate from all three segments of the proximal tubule and additional PT cell states independent of the traditional PT classification.

To further identify genes that characterize the aforementioned tumor subgroups, we compared the gene expression profiles between the EMT tumor clusters and the Epi-H tumor clusters supported by both snRNA-seq and snATAC-seq data. We detected many known EMT regulators up-regulated in the EMT tumor population (Fig. 5b), such as *SERPINE1*[57], *TGFBI*[58], *WNT5B*[59], vimentin[60] (*VIM*), and fibronectin[61] (*FN1*). These up-regulated genes are evidence that the EMT population possesses the strong mesenchymal potential and may represent a pre-metastatic tumor population. We validated vimentin as well as WNT5B using immunofluorescence staining in a tumor with EMT tumor cells compared to another tumor without EMT tumor cells (Fig. 5c). We

believe the regulation of the key genes defining the epithelial-mesenchymal scores is likely epigenetically regulated due to the correlation between the upregulation of the genes whose promoters show higher accessibility. Indeed, the gene expression changes between the two tumor groups showed a significant positive correlation with the promoter accessibility changes. *SERPINE1* and *TGFBI* had the highest increased promoter accessibility and gene expression in EMT tumor clusters (fold change > 2). 8 genes, namely *LRP2*, *EPB41L4A*, *SLC6A3*, *FRMD3*, *PTGER3*, *ABI3BP*, *SLC28A1*, and *CIT* showed over 2-fold changes in increased promoter accessibility and expression in Epi-H tumor clusters (Fig. 5d).

To understand which TFs may be differentially enriched between the two tumor clusters and may regulate the transcription of the above genes, we compared the TF binding accessibilities between the EMT tumor clusters and the Epi-H tumor cluster and prioritized TFs that were differentially expressed between the two groups (Fig. 5b). The EMT tumor population showed increased binding accessibility for known positive regulators for EMT, such as TWIST1 and JUN (Fig. 5e, Supplementary Table 4). We also observed increased accessibility binding for the hepatocyte nuclear factors (HNF4A and HNF4G) in the Epi-H tumor clusters. These transcription factors are known to regulate kidney development[62,63]. To connect these differentially enriched TFs to the differentially expressed genes, We subsequently searched for the binding motifs of these in the promoter regions of these genes. One example is *TGFBI*, which showed increased promoter accessibility and gene expression in the two EMT tumor clusters. The accessibility peak within its promoter harbors motifs for TWIST1 and JUN, consistent with the reported roles of these TFs regulating *TGFBI* transcription[64,65]. Conversely, *EPB41L4A* showed increased promoter accessibility and gene expression in the Epi-H tumor clusters. Taken

together, these data indicate that many genes distinguishing tumor groups with distinct epithelial and mesenchymal features, such as the *EPB41L4A* and *TGFBI*, are controlled epigenetically by an array of transcriptional factors delineated above and chromatin accessibility changes.

2.3.7 Chromatin accessibility landscape of *BAP1* and *PBRM1* mutant tumors

It has been reported that 80% of ccRCC tumors carry non-synonymous mutations in epigenetic regulators and chromatin remodeling genes[66], highlighting the central role of epigenome dysregulation in ccRCC. We next sought to understand the expression signatures of tumors harboring *BAP1* (BRCA1-associated protein 1) and *PBRM1* (Polybromo 1) mutations and the impact of these mutations on chromatin accessibility that may underpin such expression signatures. These genes are two of the most recurrently mutated genes in ccRCC, after *VHL*. Mutations in them are largely mutually exclusive[53,67–69], although double mutants have been observed. Previous studies have suggested *BAP1* and *PBRM1* function in different processes, as *BAP1* and *PBRM1* mutations are associated with non-overlapping gene expression signatures, differential mTORC1 activation, and different patient outcomes[53,67]. For snATAC-seq analysis, we selected 4 *BAP1*-mutant tumors (15,786 total nuclei; Fig. 6a, Extended Data Fig. 6) and 9 *PBRM1*-mutant tumors (48,774 total nuclei), 2 tumors with both *BAP1* and *PBRM1* mutations (13,580 nuclei), and 8 tumors without mutations in either *PBRM1* or *BAP1* (32,255 nuclei). All of these samples have matching snRNA-seq data and all mutant samples carrying *VHL* mutation (except for one with *VHL* promoter hypermethylation) and 3p loss opposite to the mutated alleles (Extended Data Fig. 1a).

To understand the impact of *BAP1* deficiency on chromatin accessibility, we used snATAC-seq data to analyze differentially accessible chromatin regions (DACRs) by comparing the tumor cells of *BAP1*-mutant tumors (6 tumors including 2 with both *BAP1* and *PBRM1* mutations) versus tumors without *PBRM1* or *BAP1* mutations (8 tumors). We identified 4,554 DACRs. Interestingly, most of these regions showed reduced accessibility in *BAP1* mutants (3,829 peaks, 84%; Fig. 6a, Supplementary Table 5), which is consistent with the known BAP1 function of promoting transcription[70,71]. *BAP1*-associated DACRs with decreased accessibility were distributed across all chromosomes (Fig. 6b), with hotspots in chromosomes 11 and 19 (Extended data Fig. 6b). DACRs with increased accessibility in *BAP1* mutants were more sparsely distributed, with hotspots in chromosome 5 (Extended data Fig. 6b). Moreover, *BAP1* mutation seems to have a dominant effect compared to *PBRM1* mutation, as the two tumors with both *BAP1* and *PBRM1* mutation showed more similar patterns in chromatin accessibility to *BAP1*-only mutated tumors than the *PBRM1*-only mutated tumors (Fig. 6a). We also analyzed DACRs for *PBRM1* mutants (9 tumors, not including 2 with both *BAP1* and *PBRM1* mutations) versus tumors without *PBRM1* or *BAP1* mutations (again using only tumor cells) and identified 646 DACRs, with the majority of those (87%, 561 DACRs) having increased accessibility in *PBRM1*-mutants (Fig. 6a, Supplementary Table 5). Through these analyses of snATAC-seq data, we observed that *BAP1*-deficient tumors undergo more global changes in chromatin accessibility compared to the *PBRM1*-deficient tumors in ccRCC.

2.3.8 Impact of *BAP1* and *PBRM1* mutations on transcriptional network in ccRCC

In order to better understand the functional impact of *BAP1* and *PBRM1* mutations in ccRCC, we performed pathway analysis on the genes with promoters/potential enhancers overlapping

identified DACRs. The potential enhancers were identified by the co-accessibility with the promoter DACRs. We observed that genes with promoters/enhancers displaying increased accessibility in *BAP1* mutants (compared to the non-*BAP1*/*PBRM1* mutants) were over-represented in focal adhesion (MAPK9), TNFA signaling (DUSP1), and EPHA forward signaling pathways (EPHA6; Fig. 7a-b), while genes with more closed promoters/enhancers were enriched in mTORC1 (DDIT4), EGF-EGFR signaling (PEBP1), and Rho GTPase cycle pathways (DLC1). We found enriched pathways of focal adhesion, Rac1 GTPase cycle, and hypoxia for the genes with promoter DACRs having increased accessibility in *PBRM1* mutants, and no enriched pathways for more closed genes in *PBRM1* mutants due to the small number. This suggests *BAP1* and *PBRM1* control different sets of cellular signaling pathways in ccRCC.

We further focused our analysis on the gene expression changes associated with *BAP1* mutations that can be linked to the changes in DNA accessibility. We compared the tumor-cell expression profiles of *BAP1*-mutant tumors (10 tumors including 6 tumors mutated in both *BAP1* and *PBRM1*, 31,002 cells) versus non-*BAP1*, non-*PBRM1*-mutant tumors (9 tumors, 60,658 cells) using snRNA-seq data and identified 563 differentially expressed genes (DEGs; Methods, Supplementary Table 5). By cross-checking with the *BAP1*-associated DACRs, we observed that the changes in gene expression and associated promoter/enhancer peak accessibility are significantly positively correlated (Fig. 7c). Genes with both decreased accessibility and expression in *BAP1* mutants were associated with the Rho GTPase cycle (*DLC1*, *ARHGAP24/28/32/42*), genes down-regulated by KRAS activation (*PTPRJ*, *CDH16*, *CPEB3*, *NR6A1*, and *ZBTB16*), and nuclear receptor meta-pathway (*CES3*, *PDK4*, *SERPINA1*, *SLC5A1*, and *TGFBR3*). Genes with increased accessibility and expression in *BAP1* mutants were

associated with Ras signaling (*RAPGEF5* and *MAPK9*) and EPHA forward signaling (*EPHA6* and *EFNA5*).

We performed a similar analysis to understand the gene expression and chromatin accessibility changes associated with *PBRM1* mutation. We identified 469 *PBRM1*-associated DEGs by comparing the tumor-cell expression profiles of *PBRM1*-mutant tumors (10 tumors, 29,656 cells) versus tumors without *BAP1* or *PBRM1* mutations (9 tumors, 31,002 cells) using snRNA-seq data (Supplementary Table 5). Compared to *BAP1*-associated DACRs, accessibility changes of *PBRM1*-associated DACRs were also positively correlated with the gene expression changes (Pearson's R = 0.81, P = 7.5e-11; Extended Data Fig. 6). Genes with decreased accessibility and expression in *PBRM1* mutants include *PBX1*, *RP1*, and *NCALD* (expression fold change > 2). Genes with increased accessibility and expression in *PBRM1* mutants were associated with hypoxia signaling (*ANGPTL4*, *CAV1*, and *PPFIA4*), focal adhesion (*CAV1/2*), and TNFA signaling (*TNFAIP8* and *EDN1*).

We further used the larger bulk gene expression and protein datasets from ccRCCto validate identified *BAP1*-specific DEGs (Fig. 6b). We found that 224 DEGs were consistently down/up-regulated in bulk gene expression and that 21 of them also showed consistent patterns in bulk protein data (Supplementary Table 5). Two of the most striking examples by this analysis were *CES3* and *PTPRJ*, which showed both reduced gene expression and reduced accessibilities of associated promoter/enhancer peaks in *BAP1*-mutant tumor cells (Fig. 7c). We further looked into sample-level accessibilities for the *PTPRJ* promoter peak and found that it displayed consistently lower accessibility in tumor cells of all *BAP1* mutants compared with other tumors

and PT cells from NATs (Fig. 7d, Extended Data Fig. 6c). *PTPRJ* is a member of the protein tyrosine phosphatases with tumor suppressor activity through inhibiting mitogenic signals[72]. Our results support the finding that *PTPRJ* is targeted by *BAP1*[73] and suggest *BAP1* mutation may contribute to ccRCC progression by suppressing PTPRJ activity. We also identified a potential *CES3* enhancer peak located ~5 Kb upstream of *CES3* transcriptional start site (TSS), displaying consistent lower accessibility in tumor cells of all *BAP1* mutants compared with other tumors and PT cells from NATs (Fig. 7d, Extended Data Fig. 6c). *CES3* encodes a carboxylesterase with crucial roles in xenobiotic metabolism. We found reduced *CES3* DNA accessibility and gene expression were associated with *BAP1* mutation, which was validated by bulk RNA and protein data. Although the exact function of CES3, especially in RCC, is still unknown, as a carboxylesterase, its downregulation may affect lipid metabolism and promote tumor progression in BAP1 mutants.

We also looked at the link between chromatin accessibility changes and methylation changes associated with *BAP1* mutation. We identified 2675 methylation probes differentially methylated between *BAP1*-mutant tumors and other tumors (using the bulk methylation data of a larger ccRCC cohort) and negatively correlated with the corresponding gene expression (FDR < 0.05). While these probes were associated with 1551 genes, a majority of them (1386) was associated with hypermethylation of these probes. 21.8% of the hypermethylated genes (302/1386) also showed decreased promoter accessibility in *BAP1* mutants versus other tumors (FDR < 0.05). In fact, the changes in methylation and promoter accessibility in *BAP1* mutants were negatively correlated (Extended Data Fig. 6d). These suggest the effect of *BAP1* mutation on chromatin accessibility may be related to its effect on DNA methylation.

These results provide new insights into the effects of *BAP1* and *PBRM1* mutations on chromatin accessibility in ccRCC (Fig. 6g). Compared to *PBRM1* mutations, *BAP1* mutations seem to exert moderate, but more widespread, effects on chromatin accessibility. The predominant *BAP1* mutation effect on chromatin appears to be decreasing chromatin accessibility, while *PBRM1* mutation is mostly associated with increased chromatin accessibility (Fig. 6a). Furthermore, we observed that *BAP1* mutations may control a different set of genes and signaling pathways compared to *PBRM1* mutations. These results suggest transcription regulations unique to *BAP1* and *PBRM1* mutations might produce differential phenotypes between *BAP1*- and *PBRM1*-mutant patients.

2.4 Discussions

We report the first combined snATAC-seq/snRNA-seq analysis on ccRCC. The single-nucleus approach has a number of advantages over single-cell analysis, including better representation of cells difficult to be dissociated in single-cell isolation and a dramatic reduction of artifacts from stress responses due to enzymatic treatment during tissue dissociation. These methods are also readily applied to archived material. However, we acknowledge that the single-nucleus approach generally does not cover transcripts from the small mitochondrial genome and bias toward active transcriptional activities.

Using a 4-stage process of within-sample snRNA-seq analysis, we identified 20 markers expressed specifically in ccRCC cells compared to all the other cell types. We then comprehensively characterized these markers, including their spatial distributions, using IF and

ST data. Several were associated with high tumor grade, including CP and PCSK6. CP is the only ccRCC marker that also predicts worse survival (Fig. 1e-f). Furthermore, we illustrated that the up-regulation of CP could be potentially attributed to the transcription factor MXI1 binding at its promoter and spatial distribution within individual tumors, furthering our understanding of CP in ccRCC beyond previous findings[33,74]. We also identified 3 druggable tumor markers including ABCC3, SLC6A3, and CP, and tested their therapeutic potentials in ccRCC cell lines using corresponding inhibitors. While ABCC3 and SLC6A3 may serve as candidate therapeutic targets, based on the RNA-seq data obtained for 5 of the 6 cell lines (Extended Data Fig. 3a), we observed ABCC3 expression is quite similar between the cancer cell lines. RCC-4 has the highest SLC6A3 expression and RCC-4-VHL+ has the highest CP expression. However, it's hard to draw any conclusions about if there is a link between the target gene expression and drug response, as we only have RNA-seq for 3 cancer cell lines, and the cell line with the highest target expression doesn't show the highest sensitivity to the corresponding drug.

When comparing tumor cells to various normal epithelial cell types in epigenetic and transcriptional patterns, we found the highest similarity with normal proximal tubule (PT) cells. In the context of previous reports suggesting PT is the cell of origin for ccRCC[40,75–77], our results have gone beyond mutational and transcriptional similarities to support the PT origin of ccRCC with strong epigenetic evidence. When we compared chromatin accessibility patterns between tumor cells and normal PT cells, we found increased HIF1A binding accessibility in the tumor cells, as expected from the high prevalence of *VHL* deficiency in these tumors. We also uncovered many other ccRCC-specific TFs. MXI1, KLF9, RBPJ, and NFKB1/2 have been previously implicated in renal cancer tumorigenesis[78–83], while HSF2 and SREBF2 have been

linked to renal tubular cell injury[84,85]. We observed a dozen genes up-regulated in the glycolysis pathway in ccRCC cells compared to normal PT cells, which is expected as the Warburg effect has been well-known in ccRCC. Several transcription factors have been known to play a direct role in the regulation of the Warburg effect, such as HIF-1, c-MYC, p53[86], and SIX1[87]. Nonetheless, transcriptional regulation of the Warburg effect remains largely unknown.

In this study, we discovered that MXI1, RBPJ, and NFKB1 may contribute to the regulation of the Warburg effect based on the presence of their binding motifs in the open promoter regions of glycolytic genes and their increased binding accessibility and expression in ccRCC cells. MXI1 is a member of the Mad family proteins that are known to antagonize c-Myc dependent transcription. It has been discovered that MXI1 is a direct HIF target gene. Furthermore, MXI1 knockdown impaired kidney cancer xenograft formation in nude mice[78]. While some of the previous studies had focused on the potential MXI1 connection with c-Myc[88,89] and promoting mitochondrial energy metabolism[90], our results suggest MXI1 may promote ccRCC cell growth through regulating glycolytic genes

On the other hand, the activation of NF-kappaB has been demonstrated by several studies in RCC, and NFKB1 (p50) was shown to be the subunit involved in tumorigenesis[82,91,92]. Increased NF-kappaB is associated with apoptosis and angiogenesis[82,93,94], and knockdown of NFKB1 inhibits the growth of RCC in vitro and in vivo[95]. Tanaka and colleagues showed that NF-kappaB activation in p50-null mouse embryonic fibroblasts increased glucose uptake by up-regulating the glucose transporter 3 (GLUT3) expression and maintaining a high glycolytic flux[96]. In gastric cancer and sarcoma, NF-kappaB has also been shown to promote glycolysis[27,97].

However, none of the previous studies showed NFKB1 may regulate glycolysis in ccRCC through transactivating TPI1 and ENO2, which is shown in this study and remain to be validated experimentally. Finally, RBPJ is the main transcription factor for all canonical Notch signaling, and it also functions beyond Notch signaling[98–100]. RBPJ has shown to have an oncogenic role in glioblastoma and lung cancer[101,102] and a tumor-suppressor role in breast cancer[103]. In ccRCC, RBPJ has shown elevated expression in cancer stem cells. but it remains largely unknown what the downstream targets of activated RBPJ are in ccRCC. Our results suggest that RBPJ may contribute to ccRCC pathogenesis through activating PKM. Future molecular investigation of these TFs in relation to glycolysis is needed to confirm whether and how these TFs regulate glycolytic genes and activity in ccRCC.

The discovery of averaged 3 to 4 tumor-cell clusters suggests tremendous heterogeneity within the tumor cells, as expected. We identified 4 distinct tumor subgroups based on a marker panel and the 95 tumor-cell and 9 PT-cell clusters. Epithelial tumor clusters are sporadic with respect to the traditional PT S1/2 and S3 group classification markers, suggesting these tumor cells might not come from one group of S1/2 or S3 proximal tubule cells. We believe our work showcased the dissection of the tumor subpopulation with EMT potential using single-cell omics within the primary tumor, which allows future studies to apply our methodology to connect the EMT tumor subpopulation with metastasis and drug response. Another value-add of our finding is that WNT5B is up-regulated in the ccRCC EMT tumor subpopulation. WNT5B is a member of the WNT5 protein subfamily and signals through the non-canonical beta-catenin-independent signaling pathway. It is required for cell migration, proliferation, or differentiation in many cell types[104] and has an emerging role in mediating EMT in breast cancer and pancreatic cancer[105,106].

Overexpression of WNT5B promotes colorectal cancer migration and invasion[104]. However, not much has been reported about whether WNT5B has a role in RCC. Here our results suggest WNT5B is upregulated in ccRCC tumor subpopulations with EMT features. We validated the expression of WNT5B in a patient tumor we predicted with EMT subpopulation using western blot (Fig. 5c), suggesting WNT5B may mediate/activate the EMT process in ccRCC. And we believe our results may serve as a foundation for future investigations into whether WNT5B directly or indirectly regulates the EMT process in ccRCC.

The fact that *BAP1* and *PBRM1* mutations are associated with distinct overall survival led to the first molecular classification of sporadic ccRCC[53,67]. However, not much is known about the epigenetic alterations brought upon by *BAP1* mutation and *PBRM1* mutation leading to the distinct phenotypes in ccRCC. Our snATAC-seq analysis revealed an overwhelming portion of genomic loci displaying decreased accessibility in ccRCC with *BAP1* mutations, consistent with the role of BAP1 in deubiquitination of H2AK119Ub and transcriptional activation[107]. Nonetheless, the presence of both increases and decreases in chromatin accessibility suggests that BAP1-mediated regulation of chromatin accessibility depends on the epigenetic landscape. For example, in a previous study of BAP1 function in human liver organoids, *BAP1* knockout was associated with more ATAC peaks with increased accessibility rather than decreased accessibility, suggesting BAP1 function might vary based on cell type as well as the corresponding epigenetic state. On the other hand, *PBRM1* encodes BAF180, a subunit of nucleosome remodeling complex PBAF. In contrast to the *BAP1* mutation, *PBRM1* mutation was primarily associated with ATAC-peaks exhibiting increased accessibility, suggesting PBRM1

may have a role in gene silencing. These results provide hypotheses for future functional studies of *BAP1* and *PBRM1* in ccRCC.

This report represents the first combined application of snRNA-seq and snATAC-seq in ccRCC to study respective transcriptional profiles and chromatin accessibility patterns at the single-nucleus level. Confirmation of known mutational and transcriptional alterations in ccRCC validates these single-nuclei methods as powerful tools to study cancer cell behavior at high resolution. More importantly, we have revealed previously underrecognized tumor-specific molecular changes and spatial distribution patterns of ccRCC subpopulations, underlying transcriptional regulatory mechanisms, and potential targets for treatment.

2.5 Acknowledgement

We are grateful to the patients, families, and professionals who have contributed to this study and the CPTAC ccRCC study. We also thank the Washington University Flow Cytometry and Fluorescence Activated Cell Sorting Facility for their expertise and sorting services, especially Erica Lantelme; members of McDonnell Genome Institute for providing sequencing service, especially Bob Fulton, Catrina Fronick, Jennifer Ponce, Jessica Klette, Kevin Haub, and Christopher Sawyer; Lisa Thammavong for sample processing; Di Liu for cell line handling, and Liang-Bo Wang for manuscript feedback. Kazuhito Sato has received overseas research fellowships from the Japan Society for the Promotion of Science (JSPS) and the Uehara Memorial Foundation. This work was done in collaboration with the U.S. National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) and supported by the NCI R01HG009711, U24CA211006, and U24CA210976 Funds to L.D. and F.C.

2.6 Author Contributions

L.D. and F.C. conceived the project and led project design. F.C. and W.C. developed the snRNA-seq protocol with assistance from B.D.H., X.Y., and Y.Z.. Y.W. and L.D. selected the samples. C.N. coordinated sample distribution. W.C. performed the sample preparation for snRNA-seq and snATAC-seq. R.G.J. and Y.W. coordinated the FACS and sample preparation schedule. P.L., W.C., S.Chen, Y.Z., A.S., and M.S. performed the immunostaining experiments. S.Chen and C.W. reviewed histological images. J.J.H. provided the RCC samples for spatial transcriptomics. S.Chen performed the spatial transcriptomics experiments. N.A.D performed the cell line treatment. M.S. handled the cell line culture. K.S. handled the patient-derived xenograft samples. Y.W. performed snRNA-seq data processing, quality control, clustering, and cell-type annotation with assistance from R.L., D.C.Z., and L.Y. Y.L. assisted with immune cell characterization and survival analysis. C.M performed the spatial transcriptome analysis. A.K. annotated immune cell type classification. N.V.T. developed snATAC-seq data processing and analysis pipelines. L.Y. developed the tumor-cell specific marker discovery pipeline. H.Zhu performed mutation inspections. Y.W., N.V.T., S.Cao, A.K., L.D., and F.C. interpreted the results. Y.W., N.V.T., C.M., S.Cao, A.K. generated the figures with assistance from M.W. L.D., F.C., S.J., H.R., M.M., C.K., J.J.H., D.C., T.W., Y.R., D.W.C., and H.Zhang supervised and supported the work. Y.W., N.V.T., S.Cao, A.K., and F.C. wrote the manuscript. M.C.W., M.G.C., A.C., M.M., L.D., Y.W., D.M., F.C., D.C.Z., A.K., S.Cao., Y.R, and S.Chen. provided critical reviews of the manuscript.

2.7 Methods

Selection of CPTAC3 clear cell renal cell carcinoma (ccRCC) samples

The recent large-scale CPTAC ccRCC proteomics study[10] included 103 treatment-naive ccRCC samples, which had already been comprehensively characterized. Data consisted of proteomics and phosphoproteomics, whole-exome sequencing (WES), whole-genome sequencing (WGS), RNA-seq, and histology. We selected 35 samples from this corpus, with a balanced representation of mutation status, immune subtypes, and druggable events. More specifically, we selected comparable numbers of samples with *PBRM1* and *BAP1* mutations alone and samples without mutations in either of these genes, samples with immune inflamed and immune dessert subtypes, and samples with c-MET overexpression, the last being a promising druggable target in the lab. Finally, we requested the remaining cryo-pulverized tissue (the very same pool of tissue powder that was used for the original bulk sequencing) for single nuclei RNA-seq (snRNA-seq) and single nuclei ATAC-seq (snATAC-seq). Additional tumor segments were selected by the CPTAC Biospecimen Core Resource based on availability. Additional tumor segments processed for snRNA-seq were selected based on the successes of the original tumor segment snRNA-seq, the weight of additional tumor samples, and mutation status (the four cases selected had different *PBRM1* and *BAP1* mutation status).

Clinical information

Clinical data used in this paper were retrieved via the CPTAC Data Portal at: https://cptac-data-portal.georgetown.edu/study-summary/S044.

Internal patient-derived xenograft sample collection

All human tissues acquired for experiments were processed in compliance with NIH regulations and institutional guidelines, as approved by the Institutional Review Board at Washington University in St. Louis (WUSTL). All tumor materials from patients were obtained either via core needle biopsy or surgical resection after informed consent. All animal procedures were reviewed by and received ethical approval from the Institutional Animal Care and Use Committee at WUSTL. Animal environments were maintained at 68-72°F and 45-55% relative humidity. The sample tissues for these PDX models were obtained from archived, cryopreserved PDX harvests. Final tumor passages in mice were kept cold and harvested into RPMI-1640 with antibiotic and antimycotic additives.

Cell lysis

15-25 mg of pulverized tissue were placed in a 5mL Eppendorf tube on ice. Using a wide-bore pipette tip (Rainin), a lysis buffer prepared from the Nuclei Isolation protocol (10x Genomics) and SuperRNase inhibitor (Invitrogen) was added to the tube. The tissue solution was gently pipetted until the lysis liquid turned a slightly cloudy color. (The number of pipetting iterations depended on the specific tissue.) The tissue homogenate was then filtered through a 40-micron strainer (pluriSelect) and washed with a BSA wash buffer (2% BSA + 1x PBS + RNase inhibitor). The filtrate was collected, centrifuged at 500g for 6 minutes at 4°C, and resuspended with a BSA wash buffer.

Fluorescence-activated cell sorting (FACS)

100 uL of cell lysis solution was set aside for unstained reference, while the rest was stained with DRAQ5 or 7AAD for RNA or ATAC sequencing, respectively. Namely, snRNA-seq nuclei were

stained with 1 uL of DRAQ5 per 300 uL of the sample, and snATAC-seq nuclei were stained with 1uL of 7AAD per 500 uL of the sample. Sorting gates were based on size, granularity, and dye staining signal.

10X library preparation and sequencing of snRNA-seq and snATAC-seq

Nuclei and barcoded beads were isolated in oil droplets via the 10x Genomics Chromium instrument. Single nuclei suspensions were counted and adjusted to a range of 500 to 1800 nuclei/μL using a hemocytometer. Reverse transcription was subsequently performed to incorporate cell and transcript specific barcodes. All snRNA-seq samples were run using the Chromium Next GEM Single Cell 3' Library and Gel Bead Kit v3.1 (10x Genomics). For snATAC-seq, Chromium Next GEM Single Cell ATAC Library and Gel Bead Kit v1.1 prep (10x Genomics) were used for all samples. Barcoded libraries were then pooled and sequenced on the Illumina NovaSeq 6000 system with specific flow cell types (snRNA-seq: S4; snATAC-seq: S1).

Immunofluorescence (IF) staining

5-micron thickness cut ccRCC Formalin-Fixed Paraffin-Embedded (FFPE) sections were deparaffinized and rehydrated using xylene, high to low percentages of ethanol, and finally placed in 1x PBS. Heat antigen retrieval method was applied using 1mM EDTA for at least 25 minutes. 5% Donkey serum and 1% BSA was used as blocking buffer and as primary and secondary antibodies diluent. CA9 (#NB100-417; Novus Bio) at 1:350, CA9 (#PA5-47268, Invitrogen) at 1:50, VIM (#NB300-223, Novus) at 1:150, WNT5a/b (#55184-1-AP, Proteintech) at 1:100, CP (#A80-124A; Bethyl Laboratory) at 1:100, and PCSK6 (#PA5-32966; Invitrogen) at 1:100 were applied on sections and later detected with specific fluorescent secondary antibodies

conjugated with alexa-fluors (594 & 488) emitting at red, green, and white light wavelengths. All IF images were taken using a Lecia DMi8 fluorescence microscope.

FFPE Spatial transcriptomics prep and sequencing

The RNA quality of FFPE tissue blocks were evaluated by calculating DV200 of RNA extracted from FFPE tissue sections following the Qiagen RNeasy FFPE Kit protocol. After the Tissue Adhesion Test, 5 μm sections were placed on the Visium Spatial Gene Expression Slide following Visium Spatial Protocols-Tissue Preparation Guide (10X Genomics, CG000408 Rev A). After overnight drying, slides were incubated at 60°C for 2 h. Deparaffinization was then performed following Visium Spatial for FFPE – Deparaffinization, H&E Staining, Imaging & Decrosslinking Protocol (10X Genomics, CG000409 Rev A). Sections were stained with hematoxylin and eosin and imaged at 20x magnification using the brightfield imaging setting on a Leica DMi8 microscope. After that, decrosslinking was performed immediately for H&E stained sections. Next, human whole transcriptome probe panels were then added to the tissue. After these probe pairs hybridized to their target genes and ligated to one another, the ligation products were released following RNase treatment and permeabilization. The ligated probes were then hybridized to the spatially barcoded oligonucleotides on the Capture Area. Spatial Transcriptomics libraries were generated from the probes and sequenced on the S4 flow cell of the Illumina NovaSeq 6000 system.

FFPE Spatial transcriptomics quantification and analysis

After cDNA library construction and sequencing, we use the short read probe alignment algorithm for FFPE ‘count’ method in Space Ranger (v1.3.0) from the 10X Genomics to align

probe reads to the human reference genome (GRCh38). The resulting count matrix and associated H&E physiological images were then used by the R package Seurat (v.4.0.4)[108] for subsequent analysis. The filtered gene-count matrices of all 3 samples were normalized using SCTransform before being merged into one object for joint processing and analysis using the FindNeighbors, and FindClusters function in Seurat using standard processing parameters (30 PCs, original Louvain algorithm).

In vitro cell line drug treatment and growth inhibition assessment using alamarBlue Cell Viability assay

Two ccRCC cell lines 786-O-VHL+ and 786-O were maintained in RPMI culture medium (Sigma Aldrich - R8758-1L) supplemented with 10% fetal bovine serum (FBS; Sigma, F-9665) and 1% Penicillin-Streptomycin antibiotic (Pen Strep; Gibco, 10,000 U/mL - 15140122). The ccRCC cell lines RCC-4 and RCC-4-VHL+ and a control cell line HK-2 were maintained in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F-12) culture medium (Gibco – 11320033) supplemented with 10% FBS (Sigma, F-9665) and 1% Pen Strep (Gibco, 10,000 U/mL - 15140122). All cell lines were seeded at 5000 cells/well in triplicates in 96-well plates at day 0 and were treated with either MK-571 (MRP inhibitor targeting ABCC3, Sigma Aldrich - M7571-5MG) or tetrathiomolybdate (chelator lowering copper level targeting CP, Sigma Aldrich -323446-1G) upon reaching 50-60% confluency at day 1 in culture. For MK-571 treatment, the main stock was prepared at a 1mM concentration dissolved in water and the working concentrations were used at 500µM, 100µM, 20µM, and 4µM. For tetrathiomolybdate treatment, the main stock was prepared at 15mM dissolved in Dimethyl Sulfoxide (DMSO, Sigma Aldrich - D5879-1L) and the working concentrations were used at 75µM, 15µM, 3µM,

and 0.6µM. Treatment was maintained in culture and growth inhibition assessment was performed using alamarBlue™ Cell Viability Reagent (Invitrogen - DAL1025) on day 4 post-treatment at a ratio of 1:10for 4 hours according to the manufacturer's protocol. Plots and IC50 concentrations were produced in Prism GraphPad (version 9.2.0) by plotting the percent growth inhibition on the y-axis and the Log(concentration) on the x-axis. The corresponding IC50 was extracted from the nonlinear regression curve fitting analysis using Prism GraphPad. Cells treated with only a growth medium without any drugs or with a growth medium with 0.5% DMSO were used as negative controls for cells treated with MK-571 and tetrathiomolybdate, respectively. Cisplatin (Sigma Aldrich - 479306-1G) was used as a positive control at 400µM.

Bulk omics data

Raw and processed proteomic data used in this paper can be accessed via the CPTAC Data Portal at: https://cptac-data-portal.georgetown.edu/study-summary/S044. Genomic and transcriptomic data files can be accessed via Genomic Data Commons (GDC) at: https://portal.gdc.cancer.gov/projects/CPTAC-3. Processed bulk mutation, methylation, tumor purity, and immune subtype data were downloaded from the Clark et al. ccRCC study[10].

Bulk copy number calling

Copy number variation was detected using BIC-seq2[109], a read depth-based CNV calling algorithm for WGS tumor data. Briefly, BIC-seq2 divides genomic regions into disjoint bins and counts uniquely aligned reads for each bin. It then combines neighboring bins into genomic segments with similar copy numbers iteratively based on Bayesian information criteria (BIC). We used paired-sample CNV calling that takes a pair of samples as inputs and detects genomic regions with

different copy numbers between the two samples. We used a bin size of 100bp and a lambda of 3 (smoothing parameter for CNV segmentation). Segments were called copy gain or loss when their log2 copy ratios were larger than 0.2 or smaller than −0.2, respectively. To further summarize the arm-level copy number change, we used a weighted sum approach[110], in which the segment-level log2 copy ratios for all the segments located in the given arm were added up with the length of each segment being weighted.

Sequencing read alignments and quality control (QC) of snRNA-seq data

After single-nuclei prep and sequencing, Cell Ranger (v3.1.0) from 10X Genomics (with Count functionality) was used for aligning reads to the human genome (GRCh38) with the addition of pre-mRNA references. The resulting gene-by-cell UMI count matrix was used by the R package Seurat (v.3.1.0)[108] for all subsequent processing.

Quality filters were applied to the data to remove barcodes that fell into any of the following categories: too few genes expressed (possible debris), too many associated UMIs (possibly more than one cell), and too high mitochondrial gene expression (possible dead cell). The cut-offs for these filters were based on recommendations by Seurat package documentation and manually adjusted to keep the number of cells after filtering between 3000 and 6000.

Normalization, feature selection, and dimensional reduction of snRNA-seq data

The filtered gene-count matrix was normalized for sequencing depth by dividing by the total gene counts in each cell. The value was then log-transformed using the Seurat NormalizeData function (default parameters). We calculated a subset of features (genes) that showed high cell-to-cell variation for downstream analysis. For the processing of individual samples, the Seurat function

FindVariableFeatures was used (with default parameters) to identify the top 2000 most variable features, which were then scaled using the Seurat function ScaleData (with the default parameters) to have respective mean expression and variance of 0 and 1 across cells. For the merging of datasets across all samples, the top 3000 most variable features were identified. Here, the "features" parameter for the ScaleData function was specified as all genes in the count matrix, whereby the downstream Principle Component Analysis (PCA) will take all features (with available scaled data) as inputs. For the merging of snRNA data from the same patient, we applied the Seurat function SCTransform with the parameter "vars.to.regress" specified as "nCount_RNA" and "percent.mito". The scaled data were then used directly as input for PCA using the Seurat function RunPCA (with the default parameters). The first 30 Principal Components (PCs) were used for downstream analysis. We also used the RunUMAP function (with default parameters) and the first 30 PCs to perform the Uniform Manifold Approximation and Projection (UMAP), a standard dimensional reduction step, to visualize the snRNA data. For the processing of tumor cells only in individual samples and immune cells (lymphoid and myeloid lineage immune cells separately) of all samples, the same functions (used ScaleData instead of SCTransform) and the same aforementioned parameters were used.

Clustering snRNA-seq data

Cells were clustered using a graph-based clustering (default of Seurat) approach. First, we utilized the Seurat function FindNeighbors to embed cells in a K-nearest neighbor (KNN) graph structure, based on the Euclidean distance in PCA space, with edges drawn between cells having similar expression patterns. We used the previously-defined first 30 PCs as inputs to the function, while other parameters were left as defaults. To cluster cells, we then applied modularity optimization

techniques (using the default Louvain algorithm from the Seurat function FindClusters) to iteratively group cells together in order to optimize the standard modularity function. We set the resolution parameter at 0.5, while other parameters were left as defaults. For defining tumor clusters with substantial transcriptional differences, tumor-cell clusters initially assigned by Seurat were visualized in UMAPs and manually inspected. Tumor-cell clusters without clear separation, suggesting a lack of transcriptional differences, were grouped into one cluster.

Merging of snRNA-seq data across samples

We used the Seurat function "merge" to combine the Seurat objects from multiple samples after quality control. Details for merging, normalization, feature selection, dimension reduction, and clustering of the all snRNA-seq datasets can be found at 30_aliquot_integration/docker_run_integration/integrate.R, and details for the merging and downstream analysis for multiple samples from the same patient can be found at merge_same_patient_segments/merge_same_patient_segments_C3L-00088.R, both being at our code archive https://github.com/ding-lab/ccRCC_snRNA_analysis/blob/master/integration/.

Cell-type annotation of snRNA-seq data

We curated from the literature a list of well-known markers, including CA9 for tumor cells (a downstream target of HIF and commonly upregulated in ccRCC cells, but not in normal kidney cells) and LRP2 for proximal tubule cells (Supplementary Table 2). Using the integrated snRNA data of all cells from the 31 samples, we filtered the marker genes down to those that were expressed in at least 10% of at least one cluster. We then labeled each cluster with cell type names by examining the expression values and the percentages expressed of all the filtered marker genes

across all clusters (using the "Dotplot" function of the Seurat package). Finally, we also corrected the cell type labels in individual samples based on marker gene expression, mutation, and CNV mapping evidence.

Tumor cell associated marker discovery

Tumor-specific marker discovery was done in Seurat by comparing gene expression between tumor cells and non-tumor cells in patient samples. A gene is labeled tumor cell specific if both the following criteria are satisfied: 1) the average expression of the gene is higher in tumor cells compared with any other cell type, respectively, for at least one sample, and that all the differences are of statistical significance (log(Fold Change) >0; adjusted p value<0.05); 2) the average expression of the gene is higher in tumor cells compared with non-tumor cells (as a combined population) for 90% of the samples and that such differences are found to be statistically significant in at least 75% of the samples. Here, all p values were adjusted stringently by Bonferroni correction.

Tumor-cell associated marker subcellular location annotation

To find potential antigens, we further annotated tumor cell-specific genes by their subcellular location and tissue specificity. We used 3 databases to curate the subcellular location information: 1) Gene Ontology Term 0005886; 2) Mass Spectrometric-Derived Cell Surface Protein Atlas[111] (CSPA); 3) The Human Protein Atlas (HPA) subcellular location data based on HPA version 19.3 and Ensembl version 92.38.

Average expression of given genes by cell type and sample

For this analysis, we utilized the merged Seurat object with all the nuclei from all patients, and grouped nuclei by the combination of sample ID and cell type (set it as the identity of the nuclei using the “Idents” function). Then we used the “AverageExpression” function to calculate the average expression using the “SCT” assay and “data” slot of the Seurat object (“data” slot stored the normalized expression values).

Survival Analysis

The R package "survival" was used to perform survival analysis. Standard multivariate Cox-proportional hazards modeling was applied to estimate the hazard ratio among subtypes (function coxph). Kaplan-Meier curves of overall survival were used to compare prognoses among subtypes (function survfit). The expression-high and expression low groups were defined as those with expression level of the studied gene in the top and bottom 30% quantile respectively.

Differential expression analysis

Differential expression analysis was performed in order to compare the tumor cells of each tumor sample vs. proximal tubule (PT) cells of four NATs with snRNA-seq data using the default test (Wilcoxon Rank Sum test) of function FindMarkers (from the Seurat package) with the specified parameters: min.pct=0.1, min.diff.pct=0.1, logfc.threshold = 0, and only.pos = F. Additionally, in order to correct for CNV, we performed a comparison of all tumor cells vs. normal PT cells using each pre-filtered gene and the corresponding CNV-value calculated from bulk WGS data as latent variables. We removed from the final list of genes those that were insignificant after performing CNV-correction. For comparing EMT vs. Epi-H tumor clusters and comparing the tumor cells of each of the *PBRM1*-mutant and *BAP1*-mutant tumors to the combined non-*BAP1*/*PBRM1*-mutated

tumors, the following specified parameters were used: min.pct=0.1, min.diff.pct=0, logfc.threshold = 0, and only.pos = F.

For the filtering of differentially expressed genes (DEGs) consistently up-regulated in tumor cells of individual tumors vs. combined PT cells, we require the DEGs to be significantly up-regulated (p_val_adj < 0.05, avg_logFC >0) in ≥50% of the comparisons. The filtering of DEGs consistently down-regulated in tumor cells, and DEGs specific to *BAP1*- and *PBRM1*-mutant tumors individually and together was similar to the filtering strategy described above.

Pairwise correlation of the gene expression of tumor cells and normal nephron epithelial cell types

First, we modified the tumor-cell-associated marker discovery pipeline to identify a set of markers specific to each of the 6 nephron epithelial cell types. Then we collected the top 100 markers for each of the nephron epithelial cell types and tumor cells. The average expression of the genes in the united gene list was used for the pairwise correlation of the cell groups.

Calculating the pathway activity score

To identify the top pathways that can best explain the variations among tumor subclusters within individual samples, we first identified differentially expressed genes (DEGs, positive only) for each tumor subcluster (over 50 cells) for each tumor using the Seurat FindMarkers function (default parameters). Secondly, we ran over-representation tests for DEGs for each tumor subcluster using the hallmarker gene set from MSigDB (to avoid gene redundancy) using the clusterProfiler package in R. Thirdly, we counted the frequency of a pathway over-represented in

the subcluster-associated DEGs across tumors and focused on the top 15 pathways (Extended Data Fig. 5a). Then we calculate the pathway scores for each tumor subcluster for each of the top pathways. For this step, we ran the AverageExpression function (Seurat package) to get the average expression of DEGs by tumor subclusters ("SCT" assay, "data" slot). For each DEG, its expression was scaled across all tumor clusters. And for each pathway, the pathway score is the average of the scaled expression of the pathway-associated DEGs for each tumor subcluster. For the pathway modules consisting of multiple pathways, tumor clusters with pathway scores in the upper 25% quantile for each member of the pathway module were considered enriched in the corresponding pathway module. For the mTOR pathway module, we require the tumor clusters to be in the upper 10% quantile for the pathway score of the "HALLMARK_MTORC1_SIGNALING" gene set to be considered enriched in the mTOR pathway module. For the EMT pathway module, we require the tumor clusters to be in the upper 10% quantile for the pathway score of the "HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION" gene set and those with epithelial scores lower than 20% quantile to be considered enriched in the EMT pathway module. For comparisons of the pathway scores across patient groups, we took the highest pathway score across tumor subclusters in the same patient to be tested and visualized in these figures.

Calculating epithelial score and assigning epithelial group

For the epithelial score, we used the markers for the proximal tubule cells and epithelial cells listed in Supplementary Table 2 that were also down-regulated in the EMT-enriched tumor clusters vs. other tumor clusters (FDR < 0.05). We obtained their average expression by tumor subclusters and PT clusters using the AverageExpression function (Seurat package, "SCT" assay, "data" slot). The expression for each marker was scaled across all tumor-cell and PT clusters. And for each pathway,

the epithelial score is the average of the scaled expression of the markers for each tumor subcluster. Tumor clusters with epithelial scores higher than 70% quantile were assigned as Epi-H tumor clusters. Tumor clusters with epithelial scores lower than 70% quantile and higher than 40% were assigned as Epi-M tumor clusters. Tumor clusters that were in neither of the above two groups nor EMT-enriched tumor clusters were assigned as Epi-L tumor clusters.

Sequencing read alignments and quality control (QC) of snATAC-seq

To process sequenced snATAC-seq data, the Cell Ranger ATAC tool (v.1.2.0, 10X Genomics) was used. We utilized the cellranger-atac count pipeline to filter and map snATAC-reads and to identify transposase cut sites. The GRCh38 human reference was used for the reads mapping. Next, we performed peak calling using MACS2[112]. All peaks were resized to 501 bp centered at the peak summit defined by MACS2. After this, we combined all peaks and removed the ones overlapping with the peaks with greater signal, in order to get the set of non-overlapping peaks, as described in Schep et al[113]. The resulting set of sample peaks was used to calculate the peak-count matrix using FeatureMatrix from the Signac package v.1.2.0 (https://github.com/timoast/signac), which was also used for downstream analysis. QC-filtering of the snATAC-datasets was performed using functions from the Signac package. Filters that were applied for the cell calling include: 1,000 < number of fragments in peaks < 20,000, percentage of reads in peaks > 15, ENCODE blacklist regions percentage < 0.05 (https://www.encodeproject.org/annotations/ENCSR636HFF/), nucleosome banding pattern score < 10, and enrichment-score for Tn5-integration events at transcriptional start sites > 2.

Normalization, feature selection, and dimension reduction of snATAC-seq data

The filtered peak-count matrix was normalized using term frequency-inverse document frequency (TF-IDF) normalization implemented in the Signac package. This procedure normalizes across cells, accounting for differences in coverage across them and across peaks, giving higher values to the more rare peaks. All the peaks were used as features for the dimensional reduction. We used the RunSVD Signac function to perform singular value decomposition on the normalized TF-IDF matrix, which is known as Latent Semantic Indexing (LSI) dimension reduction. The resulting 2:30 LSI components were used for non-linear dimension reduction using the RunUMAP function from the Seurat package.

Clustering of snATAC-seq data

The nuclei were clustered using a graph-based clustering approach implemented in Seurat. First, we utilized the Seurat function FindNeighbors to construct a Shared Nearest Neighbor graph using the 2:30 LSI components. Next, we used the FindClusters function to iteratively group nuclei together while optimizing modularity using the Louvain algorithm.

Merging of snATAC-seq data across samples

Merging of snATAC-seq datasets was performed using functions from the Signac and Seurat packages. In order to get the set of peaks for merging, we first combined peaks from all samples, and then for overlapping peaks, we performed an iterative removal procedure, the same as was used for creating individual sample sets of peaks. The resulting list of peaks was quantified in each dataset and was used to create a peak-cell matrix so that the set of features was the same across all snATAC datasets. After that, the merge function from the Seurat package was used to merge snATAC datasets. Next, we performed TF-IDF normalization. The LSI dimensional reduction was

performed using the RunSVD function. Non-linear dimension reduction was performed using the RunUMAP function with the first 2:50 LSI components.

Cell type label transfer from snRNA-seq to snATAC-seq data

Cell type label transfer was performed using functions from Signac and Seurat. First, we quantified chromatin accessibility associated with each gene by summing the reads overlapping the gene body and its upstream region of 2Kb, thus creating the gene by cell matrix. Coordinates for the genes were used from the Ensembl database v.86 (EnsDb.Hsapiens.v86 package). Next, we performed log-normalization of the resulting matrices using the NormaliseData function. The integration of paired snATAC-seq and snRNA-seq datasets was performed using the FindTransferAnchors function with the Canonical Correlation Analysis (CCA) option for the dimensional reduction. We then utilized the TransferData function to transfer cell type labels from the snRNA-seq dataset to the snATAC-seq dataset using the obtained set of anchors from the previous step.

Peak annotation

Peaks were annotated using R package ChiPseeker using transcript database TxDb.Hsapiens.UCSC.hg38.knownGene. The promoter region was specified (-1000,100) relative to the transcription start site.

Annotating differentially accessible chromatin regions with cis-regulatory elements

In order to annotate DACRs with cis-regulatory elements, we used the CICERO package[114]. Peaks co-accessible with the promoter peaks (co-accessibility cutoff 0.25) were annotated as potential enhancer elements.

Calculation of TF motif scores using snATAC-seq data

To evaluate TF binding accessibility profiles in the snATAC-seq data, we used the chromVAR tool[113], which calculates biased-corrected deviations (motif scores) corresponding to gain or loss of accessibility for each TF motif relative to the average cell profile.

Identifying differential TF binding accessibilities between cell groups for snATAC-seq data

To compare the differences in the binding accessibility profiles between cell groups we used a two-sided Wilcoxon rank-sum test, applying FDR correction for the resulting p-values. For the cell-type-specific TF motifs, we compared cells from each group vs. all other cells. For the comparison of tumor cells vs. the proximal tubule (PT) cells, we compared tumor cells for each individual sample vs. the PT cells pooled from two NAT samples.

Identifying ccRCC-specific TF motifs using bulk ATAC-seq data

In order to identify ccRCC-specific TFs in bulk ATAC-seq data we first performed an analysis to find KIRC-cohort-specific peaks. We performed two comparisons: between samples from KIRC-cohort and samples from all other cancer types, and between samples from KIRC-cohort vs samples from KIRP-cohort. For downstream analysis, we used only significant peaks with positive fold change found in both comparisons. To calculate motif enrichment, we used TFmotifView[115] with the default parameters.

Identifying differentially accessible chromatin regions using snATAC-seq data

To identify differentially accessible chromatin regions (DACRs) between tumor cells and normal PT cells, we performed a comparison for tumor cells from each tumor sample vs. PT cells pooled from four NAT samples using the FindMarkers function from the Seurat package, with logistic regression test and the fraction of fragments peaks as a latent variable to reduce the effect of different sequencing depths across cells. Bonferroni correction was applied for P-value adjustment using all peaks from the dataset. We required the peak to be significant (FDR<0.05) in at least 50% of comparisons, and the same fold-change direction in all comparisons. Additionally, in order to correct for CNV, we performed a comparison of all tumor cells vs normal PT cells using a fraction of fragments in peaks and CNV-value calculated from bulk WGS data as latent variables. We removed from the final list of peaks the ones that were insignificant after performing CNV-correction.

Next, to identify DACRs specific to *BAP1*-mutant and *PBRM1*-mutant tumors, we used the sets of samples for each category described above. We performed comparisons for each *BAP1*/*PBRM1*-mutant tumor sample vs pooled tumor cells from non-mutant samples. DACRs specific to *BAP1*-mutant and *PBRM1*-mutant groups of samples were selected if they were significantly more accessible in ≥50% of the samples from the respective groups compared to the non-mutant samples. We also required that a DACR should have the same fold-change direction in all comparisons. Finally, we removed peaks that were insignificant after CNV-correction.

Mapping TF motif to DACRs of DEGs

Then we filtered out genes lacking DACRs overlapping their short promoter regions (-1000 to 100 relative to the TSS). Next, we searched for motifs of top cell-type-specific TFs in the DACRs of selected DEGs for cell types of interest. We then divided the TF-DEG interactions into two categories based on the coordinates, relative to TSS, of the motifs found in the DACRs overlapping a promoter: (1) promoter motif, if the motif was found in the short promoter region (-1,000 to 100 from TSS) and (2) distant motif, if the motif was found outside the promoter region. We used the described procedure to study the mechanisms of transcriptional regulation in both normal PT and tumor cells. Mapping of the motifs to the DACRs was performed using the motifmatchr R package.

Visualizing the coverage of snATAC-seq for individual cell types

For snATAC coverage plots, we used the CoveragePlot function from the Signac package. For tumor samples, we plotted coverage for tumor cells only, and for NAT samples we plotted coverage for normal PT cells only.

Over-representation test for differentially expressed genes and differentially accessible chromatin regions

For over-representation tests other than tumor subcluster-associated DEGs, we used the hallmark gene set and the canonical gene set from MSigDB[116,117], and the enricher function from the clusterProfiler R package[118]. For the over-representation test for DEGs between tumor cells and PT cells, genes that are expressed in at least 10% of either cell group were used as background. For the over-representation test for the differentially accessible peaks associated with BAP1 and PBRM1, the nearest genes associated with all detected peaks were used as background.

Data availability

Clinical data and raw/processed proteomic data reported in this paper can be accessed via the CPTAC Data Portal at: https://cptac-data-portal.georgetown.edu/study-summary/S044. Bulk genomic and transcriptomic data files can be accessed via Genomic Data Commons (GDC) at: https://portal.gdc.cancer.gov/projects/CPTAC-3.

Code availability

Detailed codes for the analysis and figures are provided at https://github.com/ding-lab/ccRCC_snRNA_analysis.

2.8 Main Figures

Figure 1

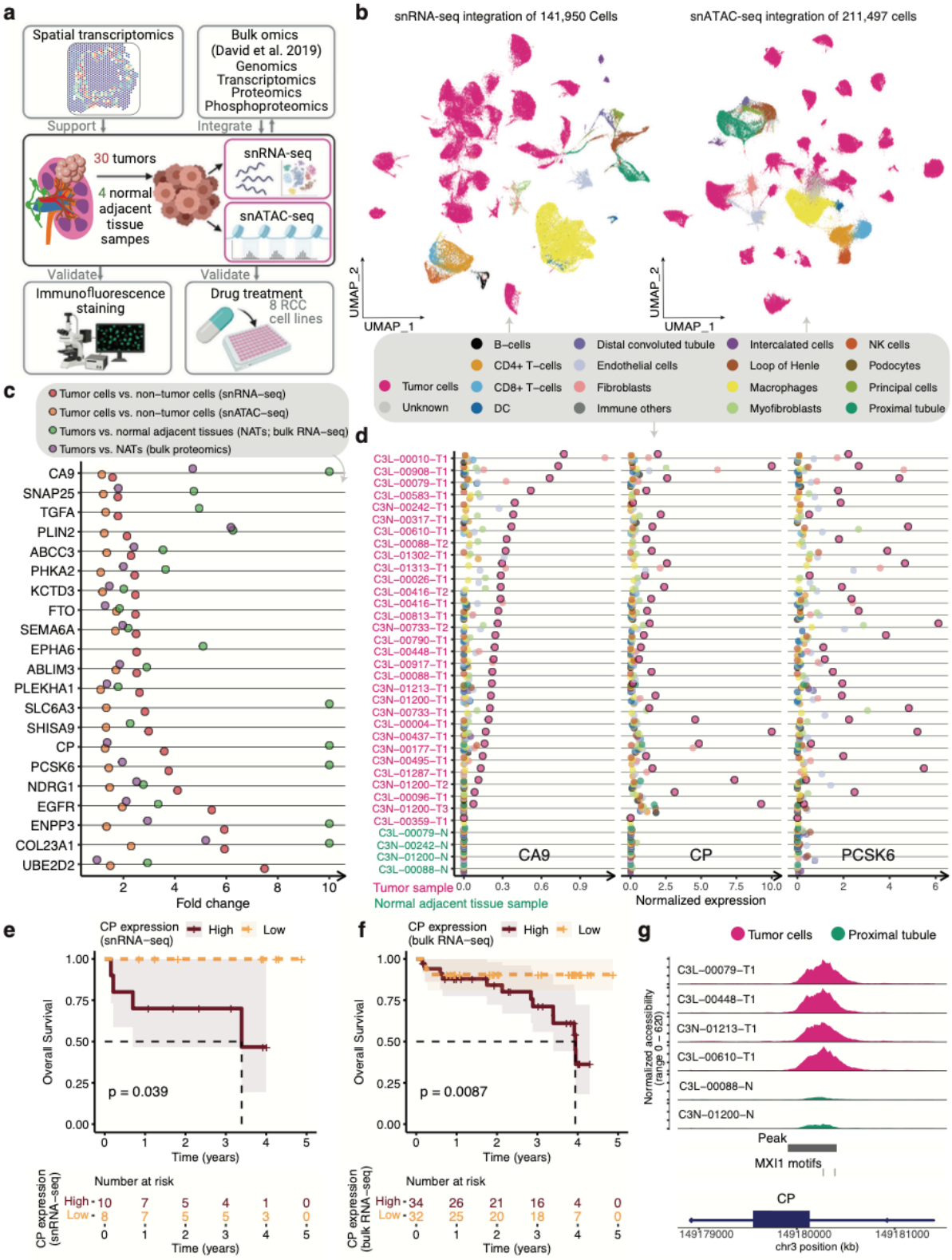


Figure 1. snRNA-seq analysis identifies tumor-cell-specific markers.
a, Schematic for integrating snRNA-seq, snATAC-seq, bulk omics, and spatial transcriptomics data, and validating the omics findings with immunofluorescence staining and targeted drug treatment. **b**, UMAP visualization of 146,425 nuclei and 211,497 nuclei profiled by snRNA-seq and snATAC-seq, respectively, colored by major cell groups. The "immune others" cell group include basophils, mast cells, lymphoid lineage immune cells with proliferating signature, and immune cells with ambiguous myeloid/lymphoid identity. **c**, Dot plot showing the fold changes of expression of 21 tumor-cell markers in ccRCC (capped at 10X). Red dots denote the gene expression fold changes between tumor cells vs. non-tumor cells using snRNA-seq data. Orange dots denote the gene accessibility fold changes between tumor cells vs. non-tumor cells using snATAC-seq data. Green dots denote the gene expression fold changes between bulk tumor and normal adjacent tissue (NAT) samples using bulk RNA-seq data. Purple dots denote the bulk protein level changes (spectrum intensity) between tumors vs. NAT samples. **d**, Dot plot showing the expression levels of *CA9*, *CP*, and *PCSK6* in each cell type and each sample (non-log space). Expression levels for tumor cells are highlighted by black outlined circles. **e-f**, Kaplan-Meier survival analysis showing recurrence-free survival after initial pathologic diagnosis. Patients with high tumor-cell expression of *CP* (n = 10, top 35% percentile) displayed significantly lower chance of event-free survival compared to patients with low tumor-cell expression of *CP* (n = 8, bottom 35% percentile) using snRNA-seq data (left plot). This result is validated by a similar analysis using bulk gene expression of *CP* in the larger CPTAC ccRCC discovery cohort (n = 66). P-value was calculated by log-Rank test. **g**, Genomic region near *CP* gene promoter (in ccRCC cells). The plots show the normalized accessibility by snATAC-seq around these regions in proximal tubule cells (green) from NAT samples and ccRCC cells (pink) from representative tumor samples. Similar results have been obtained by comparing tumor cells from all tumor samples to the proximal tubule cells.

Figure 2

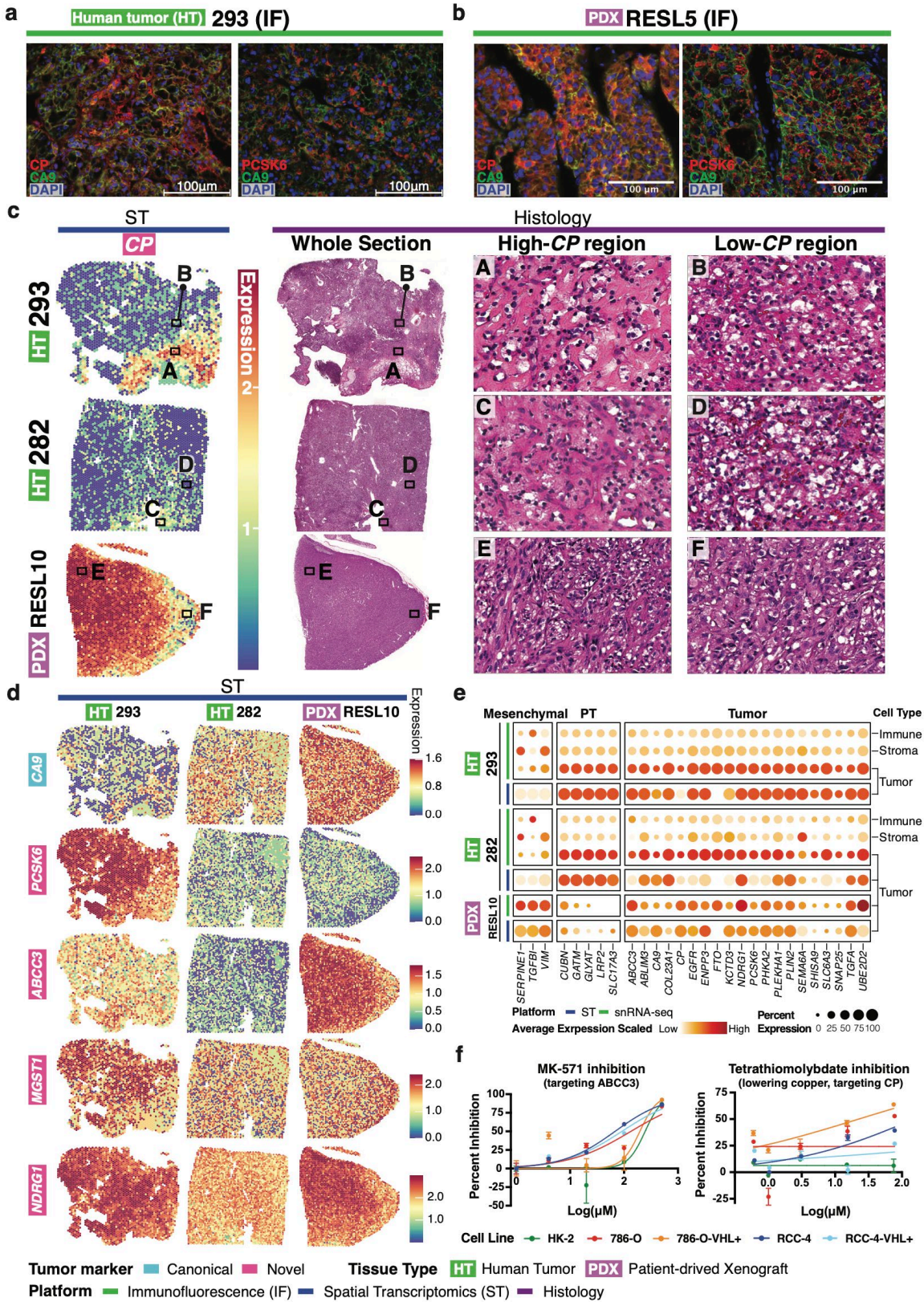

a
Human tumor (HT) 293 (IF)
CP
CA9
DAPI
100μm
PCSK6
b
PDX RESL5 (IF)
100 μm
c
ST
CP
Histology
Whole Section
High-CP region
Low-CP region
HT 293
HT 282
PDX RESL10
Expression
d
ST
HT 293
HT 282
PDX RESL10
CA9
PCSK6
ABCC3
MGST1
NDRG1
Expression
e
Mesenchymal
PT
Tumor
Cell Type
Immune
Stroma
Tumor
SERPINE1
TGFBI
VIM
CUBN
GATM
GLYAT
LRP2
SLC17A3
ABCC3
ABLIM3
CA9
COL23A1
CP
EGFR
ENPP3
FTO
KCTD3
NDRG1
PCSK6
PHKA2
PLEKHA1
PLIN2
SEMA6A
SHISA9
SLC6A3
SNAP25
TGFA
UBE2D2
Platform ST snRNA-seq
Average Exrpession Scaled Low High
Percent Expression 0 25 50 75 100
f
MK-571 inhibition
(targeting ABCC3)
Tetrathiomolybdate inhibition
(lowering copper, targeting CP)
Percent Inhibition
Log(μM)
Cell Line HK-2 786-O 786-O-VHL+ RCC-4 RCC-4-VHL+
Tumor marker Canonical Novel
Tissue Type HT Human Tumor PDX Patient-drived Xenograft
Platform Immunofluorescence (IF) Spatial Transcriptomics (ST) Histology

Figure 2. ccRCC tumor-cell markers CP and PCSK6 are validated by immunofluorescence staining and spatial transcriptomics.
Immunofluorescence (IF) staining of a ccRCC **a.** PDX tumor (ID: RESL5) and **b.** human tumor sample (ID: 293). Left panel, markers CP (novel ccRCC marker, red), CA9 (canonical ccRCC marker, green), and DAPI (nucleus, blue). Right panel, markers PCSK6 (novel ccRCC marker, red), CA9 (green), and DAPI (nucleus, blue). **c.** Spatial transcriptomics and H&E histology of 2 ccRCC human tumors (ID: 293, and 282) and a PDX tumor sample (ID: RESL10). Two regions were selected for each sample (293, 282, and RESL10) indicating morphology of high-*CP* (regions A, C, and E) and low-*CP* (regions B, D, and F). **d.** Spatial transcriptomics gene expression of canonical ccRCC marker *CA9* and novel tumor markers: *PCSK6, ABCC3, MGST1, and NDRG1.* **e.** Bubble plot of mesenchymal, proximal tubule (PT) and tumor markers gene expression from snRNA-seq (green) and ST (blue) in sample 293, 282, and RESL10. **f.** *In vitro* cell line growth inhibition using MK-571 (left panel, MRP inhibitor targeting ABCC3) and tetrathiomolybdate (right panel, chelator lowering copper level targeting CP). The error bars represent the standard error of the mean (SEM).

Figure 3

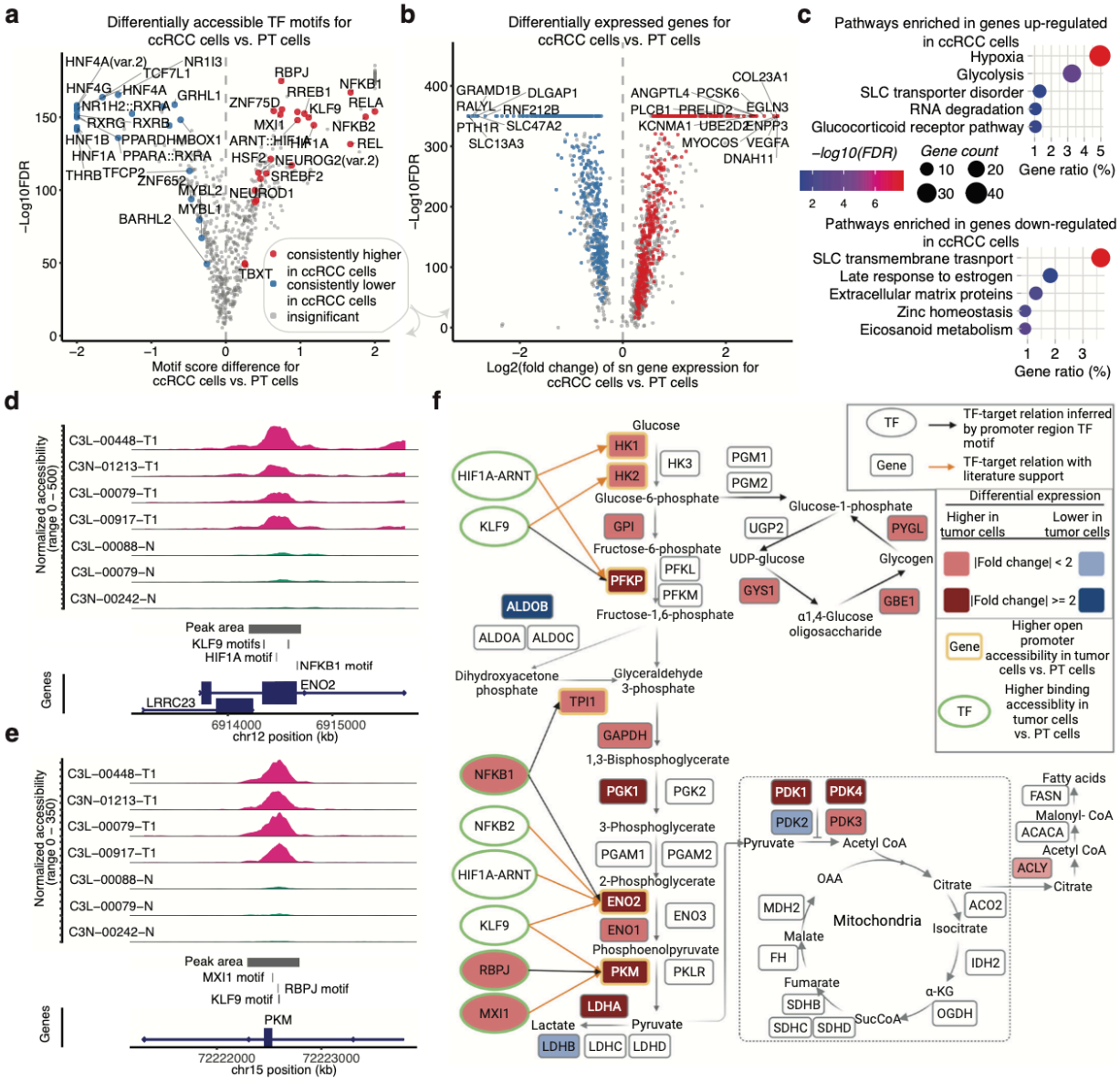


Figure 3. Glycolysis pathway displays significant changes in ccRCC tumor cells compared to the proximal tubule cells.

a, Volcano plot showing differentially enriched TF motifs between ccRCC (tumor) cells (n = 118,409) and the combined proximal tubule (PT) cells from the NATs (n = 9,676). X-axis shows the motif score difference, while Y axis shows the -Log10(adjusted p-value). Color denotes whether a motif is consistently higher or lower in tumor cells when tumor cells from individual tumor samples were compared to the PT cells or it has insignificant or inconsistent fold changes (Methods). The motifs that have consistent higher/lower TF binding accessibilities in all comparisons of individual tumor vs. PT cells are highlighted. **b,** Volcano plot showing differentially expressed genes between ccRCC cells (n = 88,536) and the combined PT cells from the NATs (n = 4,269). X-axis shows the log2 fold change of the sn gene expression of the ccRCC cells compared to PT cells; Y axis shows the -Log10(adjusted p-value). Color denotes whether a gene is consistently expressed higher or lower in tumor cells, or has insignificant or inconsistent

fold changes (Methods). Genes with log2 fold change larger than 2 or smaller than -2 are highlighted. **c**, Bubble plot showing the pathways over-represented in genes up-regulated (top) and down-regulated (bottom) in ccRCC cells compared to the PT cells. **d-e,** Genomic regions near two up-regulated genes in ccRCC cells compared to PT cells. The plots show the normalized accessibility by snATAC-seq around these regions in proximal tubule cells (green) from NAT samples and ccRCC cells (pink) from representative tumor samples. Similar results have been obtained by comparing tumor cells from all tumor samples to the proximal tubule cells. **f**, When viewed in the context of important metabolic pathways in ccRCC, ccRCC cells displayed an overall up-regulation of genes encoding glycolysis enzymes as well as other metabolic proteins (rounded rectangle) at sn gene expression level (red and blue filled colors represent significantly increased or decreased sn expression in ccRCC cells vs. proximal tubule cells) validated by bulk RNA-seq and/or proteomics data. Among them, genes showing increased promoter accessibility are highlighted by the yellow border. Eclipses with green borders represent transcription factors with significantly enriched accessibility binding in ccRCC cells vs. PT cells. Lines connecting TFs and genes represent TF-target relations inferred by the presence of TF motif in the more accessible promoter region of the genes using snATC-seq. Orange lines denote those TF-target relations with literature support.

Figure 4

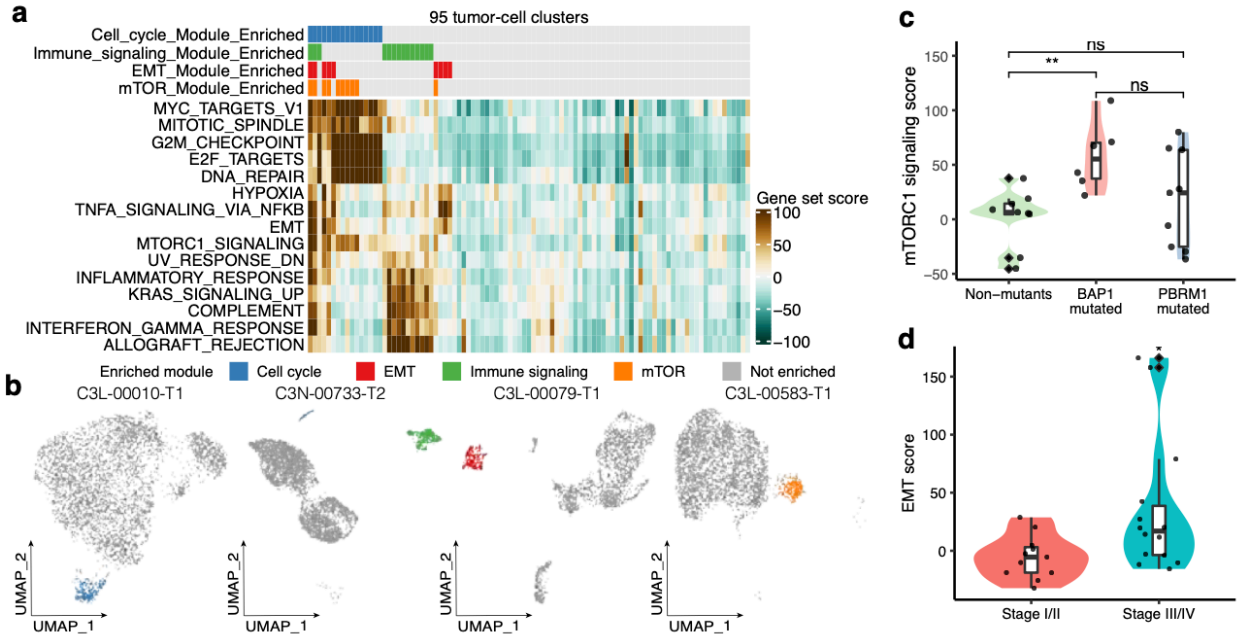


Figure 4. Intratumor signaling heterogeneity revealed by single-cell tumor subclustering. **a**, Heatmap showing the gene set scores for 95 tumor subclusters (columns). Co-clustered gene sets were grouped into one module, e.g. the cell cycle module includes G2M checkpoint, E2F targets, DNA repair, mitotic spindle, and MYC target gene sets. The black bar represents whether a tumor cluster showed enriched gene expression in each module. Each tumor cluster is annotated by the fraction of the tumor cells with 3p (VHL/SETD2/BAP1/PBRM1) loss by snRNA-based CNV calling. **b**, UMAP visualizations of the tumor-cell-only clustering of 4 tumors, colored by different activated pathway modules. **c**, Boxlot showing maximum mTORC1 signaling score per tumor sample, grouped by *BAP1/PBRM1* mutation status. **d**, Boxplot showing maximum EMT scores per tumor sample, grouped by tumor stage.

Figure 5

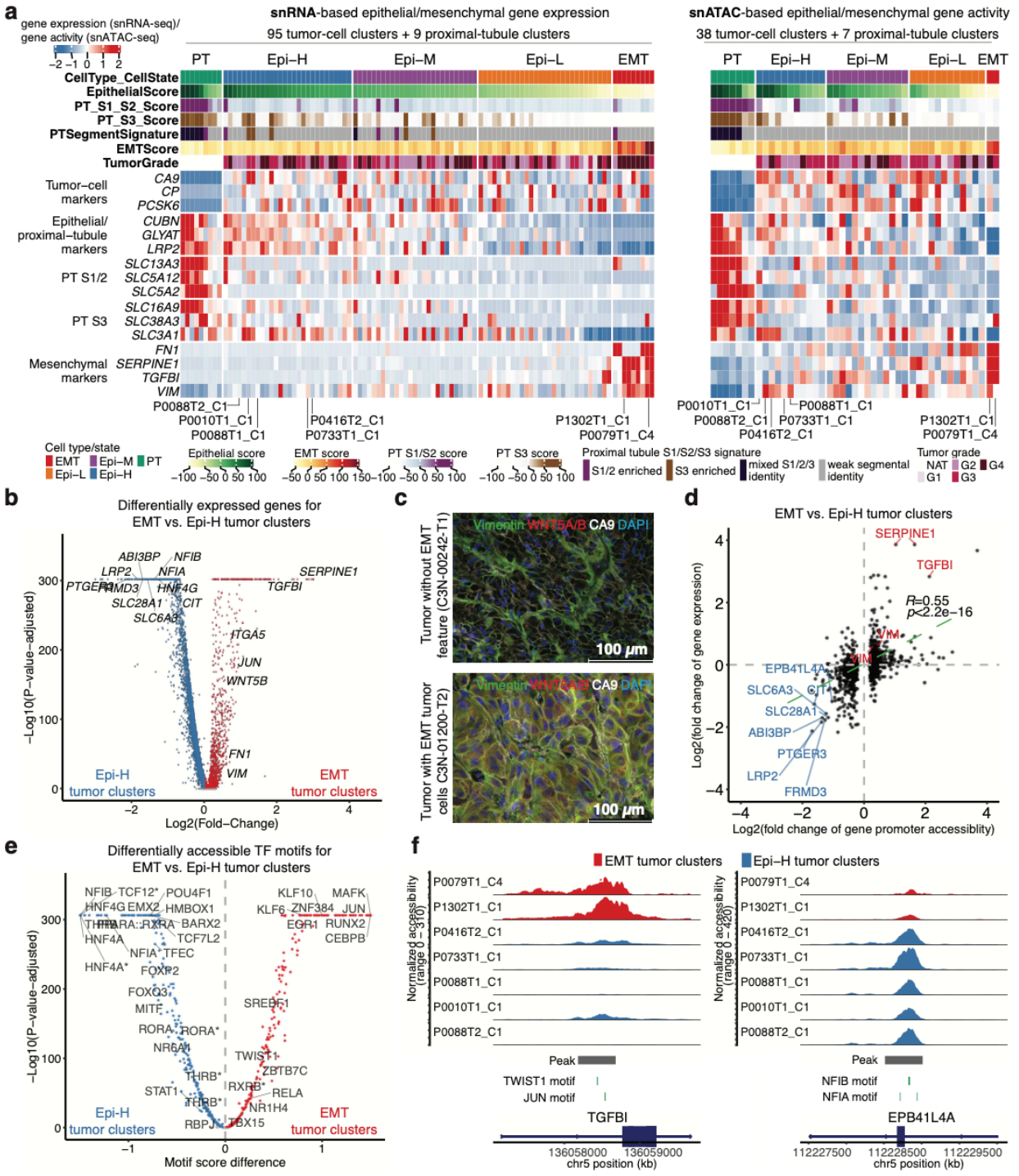
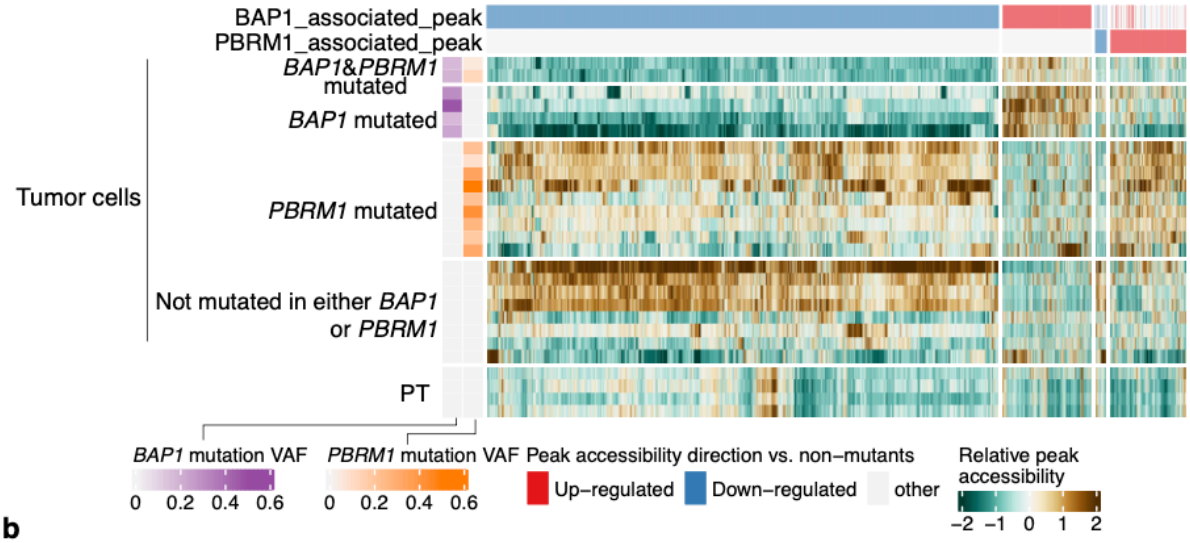


Figure 5. Four tumor subgroups with distinct epithelial and mesenchymal features.
a, Left: Heatmap showing gene expression of the epithelial and mesenchymal marker genes for 95 tumor clusters and 9 proximal tubule (PT) clusters (> 50 cells) using snRNA-seq data. Right: Heatmap showing gene activity of the epithelial and mesenchymal marker genes for 38 tumor clusters and 9 PT clusters (> 50 cells) using snATAC-seq data. **b**, Volcano plot showing

differentially expressed genes between the EMT tumor clusters and Epi-H tumor clusters highlighted in a. Labels on the right denote known mesenchymal markers, while those on the left denote known markers for PT cells. **c**, Immunofluorescence staining of vimentin (VIM), CA9, WNT5A/B, and DAPI, showing VIM and WNT5A/B in CA9+ cells in the cross-sections of tumor with EMT tumor cells (C3N-01200-T2), but not in the control tumor (C3N-00242-T1). Scale bar, 100 μm. **d**, Scatter plot displaying the Log2 transformed fold change for gene promoter accessibility versus Log2 gene expression change in EMT tumor clusters vs. Epi-H tumor clusters highlighted in a. **e**, Volcano plot showing differentially accessible TF motifs between the EMT tumor clusters and Epi-H tumor clusters. Asterisks denote the "var.2" version of the TF motif based on the JASPAR database. **f**, Genomic regions near *TGFBI* (up-regulated in EMT tumor clusters) and *EPB41L4A* (up-regulated in Epi-H tumor clusters). The plots show the normalized accessibility by snATAC-seq around these regions in EMT tumor clusters (red) and Epi-H tumor clusters (blue).

Figure 6

a


BAP1_associated_peak
PBRM1_associated_peak
BAP1&PBRM1 mutated
BAP1 mutated
PBRM1 mutated
Not mutated in either BAP1 or PBRM1
Tumor cells
PT
BAP1 mutation VAF
0 0.2 0.4 0.6
PBRM1 mutation VAF
0 0.2 0.4 0.6
Peak accessibility direction vs. non−mutants
Up−regulated
Down−regulated
other
Relative peak accessibility
−2 −1 0 1 2

b

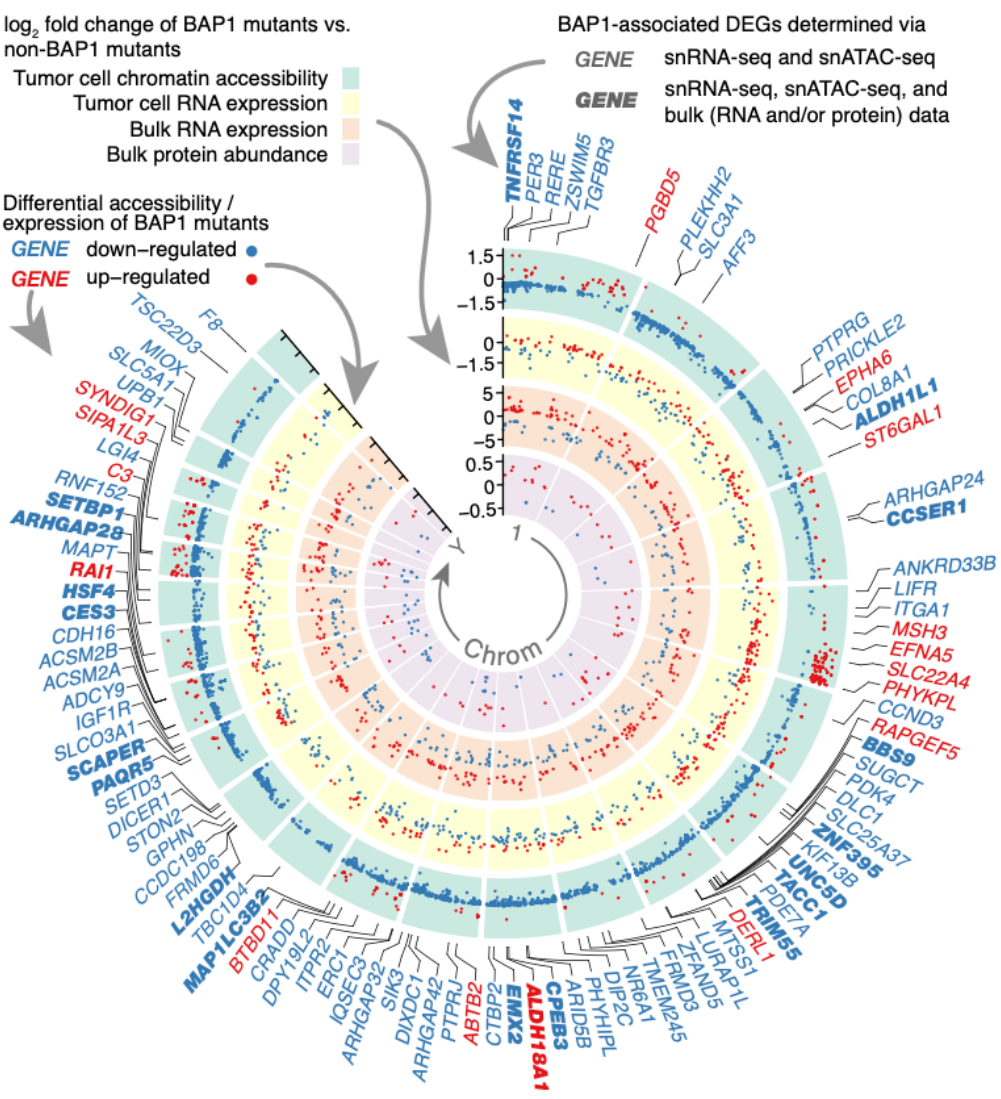

log2 fold change of BAP1 mutants vs. non-BAP1 mutants
Tumor cell chromatin accessibility
Tumor cell RNA expression
Bulk RNA expression
Bulk protein abundance
Differential accessibility / expression of BAP1 mutants
GENE down−regulated
GENE up−regulated
BAP1-associated DEGs determined via
GENE snRNA-seq and snATAC-seq
GENE snRNA-seq, snATAC-seq, and bulk (RNA and/or protein) data
Chrom
TNFRSF14
PER3
RERE
ZSWIM5
TGFBR3
PGBD5
PLEKHH2
SLC3A1
AFF3
PTPRG
PRICKLE2
EPHA6
COL8A1
ALDH1L1
ST6GAL1
ARHGAP24
CCSER1
ANKRD33B
LIFR
ITGA1
MSH3
EFNA5
SLC22A4
PHYKPL
CCND3
RAPGEF5
BBS9
SUGCT
PDK4
DLC1
SLC25A37
KIF13B
ZNF395
UNC5D
TACC1
TRIM55
PDE7A
DERL1
MTSS1
LURAP1L
ZFAND5
FRMD3
TMEM245
NR6A1
DIP2C
PHYHIPL
ARID5B
CPEB3
ALDH18A1
EMX2
CTBP2
ABTB2
PTPRJ
ARHGAP42
DIXDC1
SIK3
ARHGAP32
IQSEC3
ERC1
ITPR2
DPY19L2
CRADD
BTBD11
MAP1LC3B2
L2HGDH
TBC1D4
FRMD6
CCDC198
GPHN
STON2
DICER1
SETD3
PAQR5
SCAPER
SLCO3A1
IGF1R
ADCY9
ACSM2A
ACSM2B
CDH16
CES3
HSF4
RAI1
MAPT
ARHGAP28
SETBP1
RNF152
C3
LGI4
SIPA1L3
SYNDIG1
UPB1
SLC5A1
MIOX
TSC22D3
F8

Figure 6. Chromatin accessibility landscape of BAP1 and PBRM1 mutant tumors.
a, Heatmap showing the relative changes in ATAC-peak accessibility for peaks differentially accessible between the tumor cells of *BAP1*-mutated tumors (6 tumors, including 2 *BAP1*- and *PBRM1*-mutated tumors, 29,366 cells) vs. non-*BAP1/PBRM1*-mutated tumors (non-mutants) and peaks differentially accessible between tumor cells of *PBRM1*-mutated tumors (9 tumors, 32,255 cells). Each column is an ATAC-peak and only significantly and consistently changed peaks are plotted (FDR < 0.05, Methods). Only samples with > 8% mutation VAF in either *BAP1* or *PBRM1* are shown (1 sample excluded). **b**, Circos plot showing the genome wide chromatin accessibility changes associated with *BAP1* mutation. The green circle represents the fold change of significantly different ATAC-peaks in *BAP1*-mutated tumors vs. non-*BAP1*-mutated tumors (in red peaks with higher accessibility and in blue peaks with lower accessibility). The yellow circle plots the fold changes of the differentially expressed genes (DEGs) associated with *BAP1* mutation discovered by snRNA-seq data. The orange circle displays the fold changes of the DEGs by bulk RNA-seq. The innermost purple circle plots the fold changes of differentially expressed proteins.

Figure 7

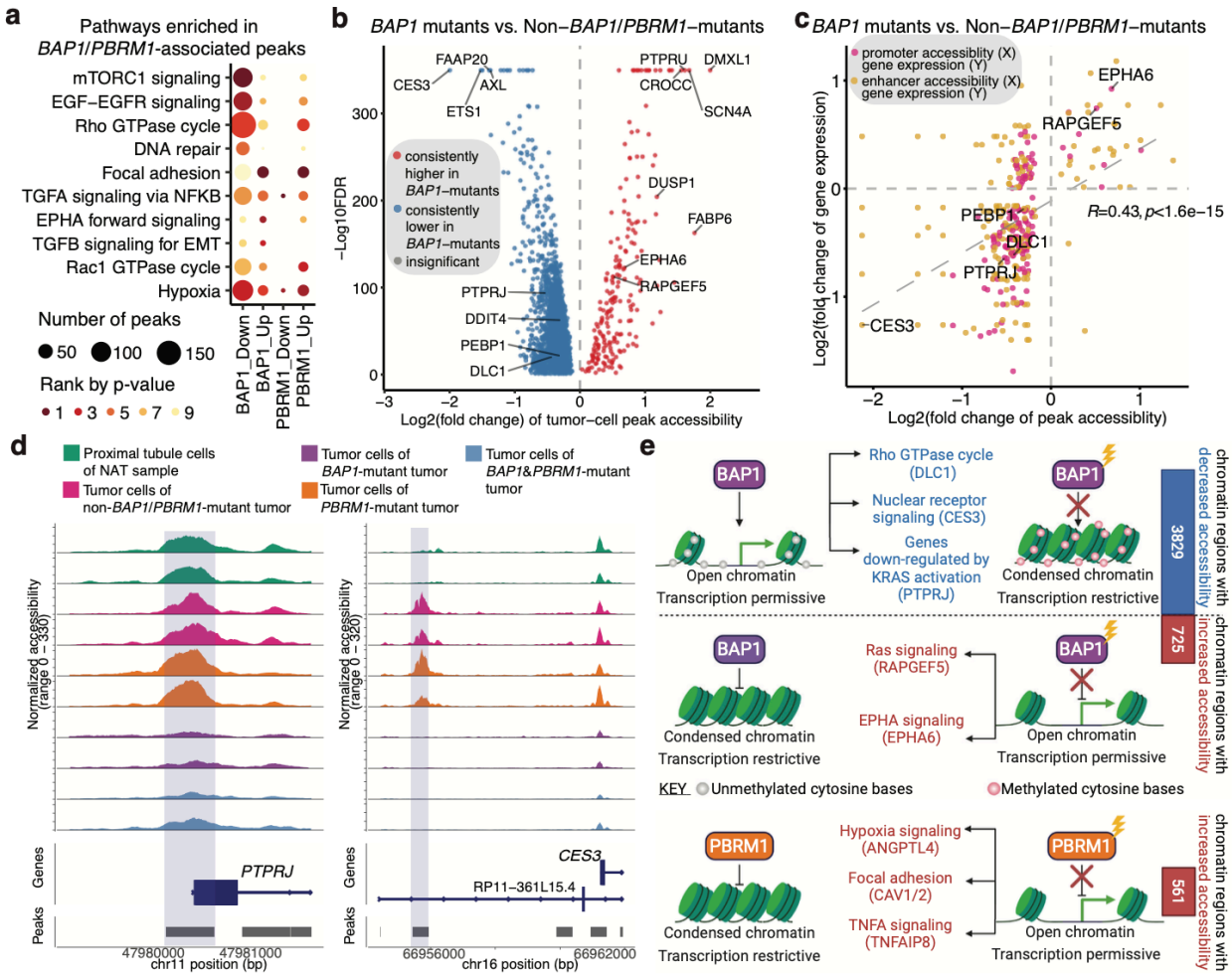


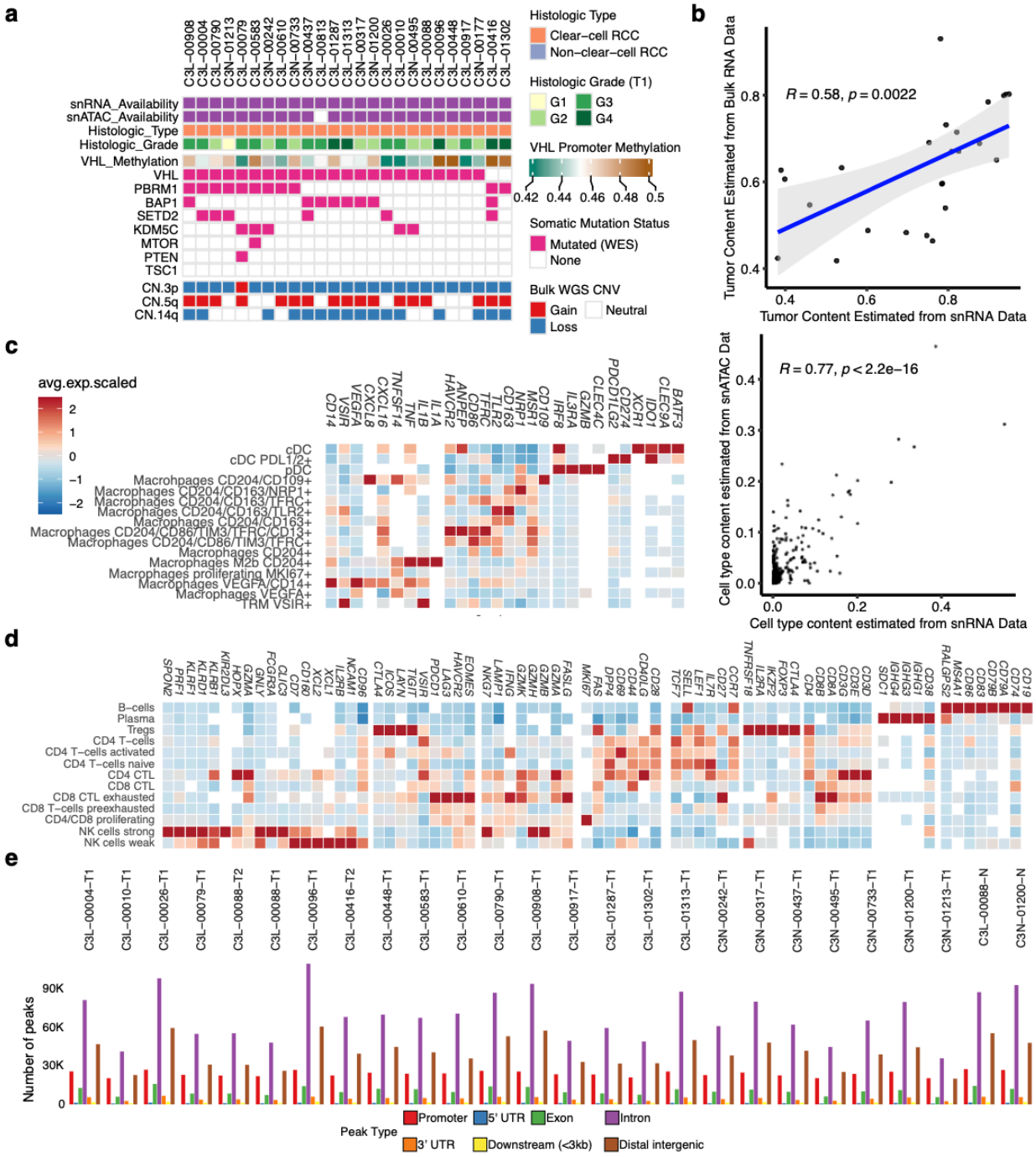
Figure 7. Impact of BAP1 and PBRM1 mutations on chromatin accessibility and transcriptional networks.

a, Bubble plot showing the pathways over-represented in the differentially accessible peaks associated with *BAP1* mutation. **b**, Volcano plot displaying the differentially accessible peaks between the tumor cells of *BAP1*-mutated tumors (29,366 cells) vs. tumor cells of non-*BAP1/PBRM1*-mutated tumors (32,255 cells) by snATAC-seq data. Dots are colored by whether the genes showed significant and consistent fold changes in individual comparisons of each *BAP1*-mutated tumor vs. non-*BAP1/PBRM1*-mutated tumors. **c**, Scatter plot showing the positive correlation of chromatin accessibility and transcriptional changes. The fold change (log2) of the snRNA-seq expression for each gene (mRNA) is plotted against the fold change (log2) in the relative snATAC-seq peaks (for all the genes/peaks with significant fold change in over 50% of the comparisons for individual *BAP1*-mutated tumor vs. non-*BAP1/PBRM1*-mutated tumors). Each dot represents a gene-peak pair. Dots are colored by whether the peak overlaps the gene promoter or is a potential enhancer (co-accessible with the promoter peak). **d**, Genomic regions near two down-regulated genes (*PTPRJ* and *CES3*) in BAP1-mutated tumors vs. non-*BAP1*-mutated tumors. The plots show the normalized accessibility signal by snATAC-seq around these regions in tumor cells of *BAP1*-mutant tumor (purple), tumor cells of *PBRM1*-mutant tumor (orange), tumor cells of non-*BAP1*/*PBRM1*-mutant tumors (pink), and proximal tubule

cells (green) from NAT samples from representative tumor samples. Similar results have been obtained by comparing tumor cells from all *BAP1*-mutant tumor samples to other tumors (Extend Data Fig. 6). **e**, Schematic diagram showing the differential effects of *BAP1* and *PBRM1* mutations on chromatin accessibility.

2.9 Supplemental Figures

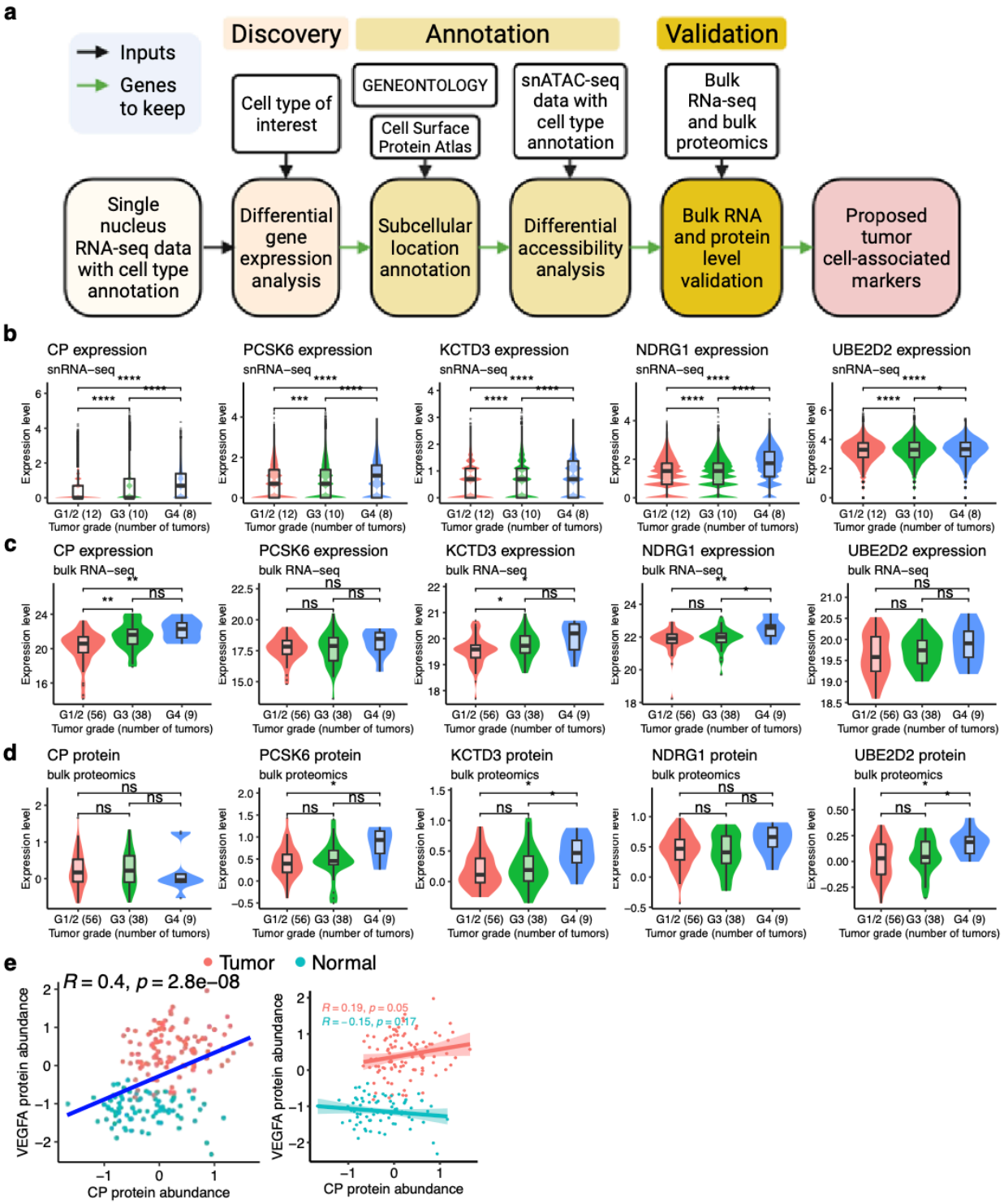
Extended Data Figure 1



Extended Data Figure 1. Overview of snRNA-seq, snATAC-seq, and bulk omics data. **a**, Genomic and clinical features for 25 ccRCC patients used in this study. **b**, Top: Scatter plot showing that tumor cell percentages estimated by the snRNA-seq (x-axis) were significantly correlated with the tumor purity estimates by ESTIMATE using bulk RNA data (y-axis) among 25 ccRCC tumors. Bottom: Scatter plot showing that cell type content estimated by snRNA-seq

(x-axis) is significantly correlated with cell type content estimated by snATAC-seq (y-axis). **c**, Heatmap showing the snRNA-seq expression of known marker genes (x-axis) in various lymphoid cell types (y-axis). Expression was averaged across cells within each cell type and then scaled across all lymphoid cell types. Expression is shown only for cell types expressing a given gene in at least 1% of cells. **d**, Expression of known marker genes (x-axis) in various myeloid cells (y-axis). Expression was averaged across cells within each cell type and then scaled across all myeloid cell types. Expression is shown only for cell types expressing a given gene in at least 1% of cells. **e**, Bar plot showing the numbers of peaks, divided and colored by peak type, across the snATAC samples.

Extended Data Figure 2

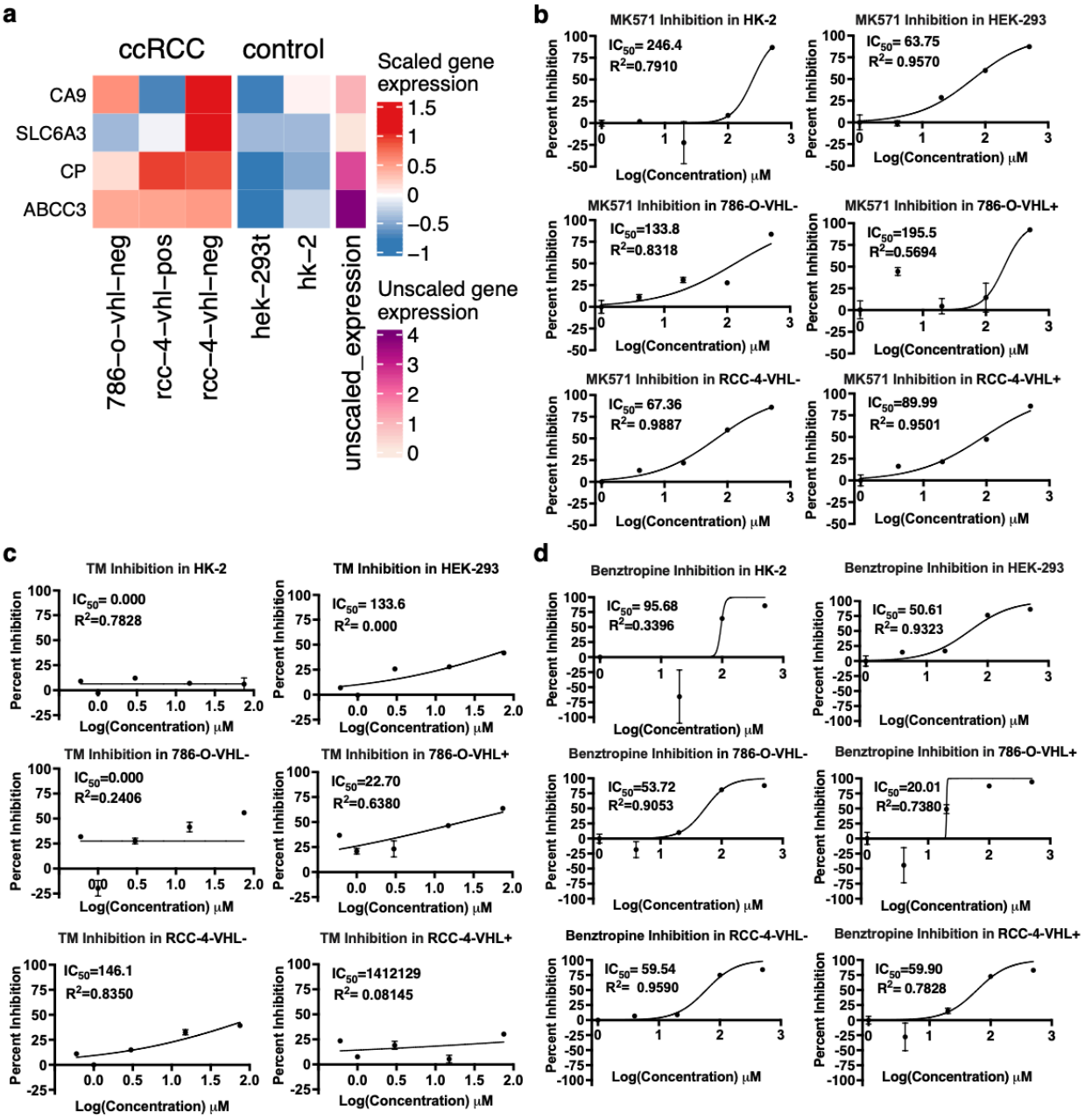


Extended Data Figure 2. *CP*, *PCSK6*, *KCTD3*, *NDRG1*, and *UBE2D2* expressions are associated with higher tumor grades.

a, Schematic showing the tumor-cell discovery pipeline. **b**, Violin plots showing the tumor-cell expression of *CP*, *PCSK6*, *KCTD3*, *NDRG1*, and *UBE2D2* for grade 1/2, grade 3, and grade 4

tumors (by cell). Horizontal connecting segments indicate levels of statistically significant differences between grades (* = 0.05, ** = 0.01, *** = 0.001, **** = 0.0001, “ns” = not significant) **c**, Violin plots showing the bulk RNA expression of the aforementioned genes for tumors with different grades. **d**, Violin plots showing the bulk protein abundance of the aforementioned genes for tumors with different grades. **e**, Scatter plot showing the significant correlation between CP and VEGFA protein levels when considering tumor and normal data jointly (left), though correlation weakens when they are considered individually.

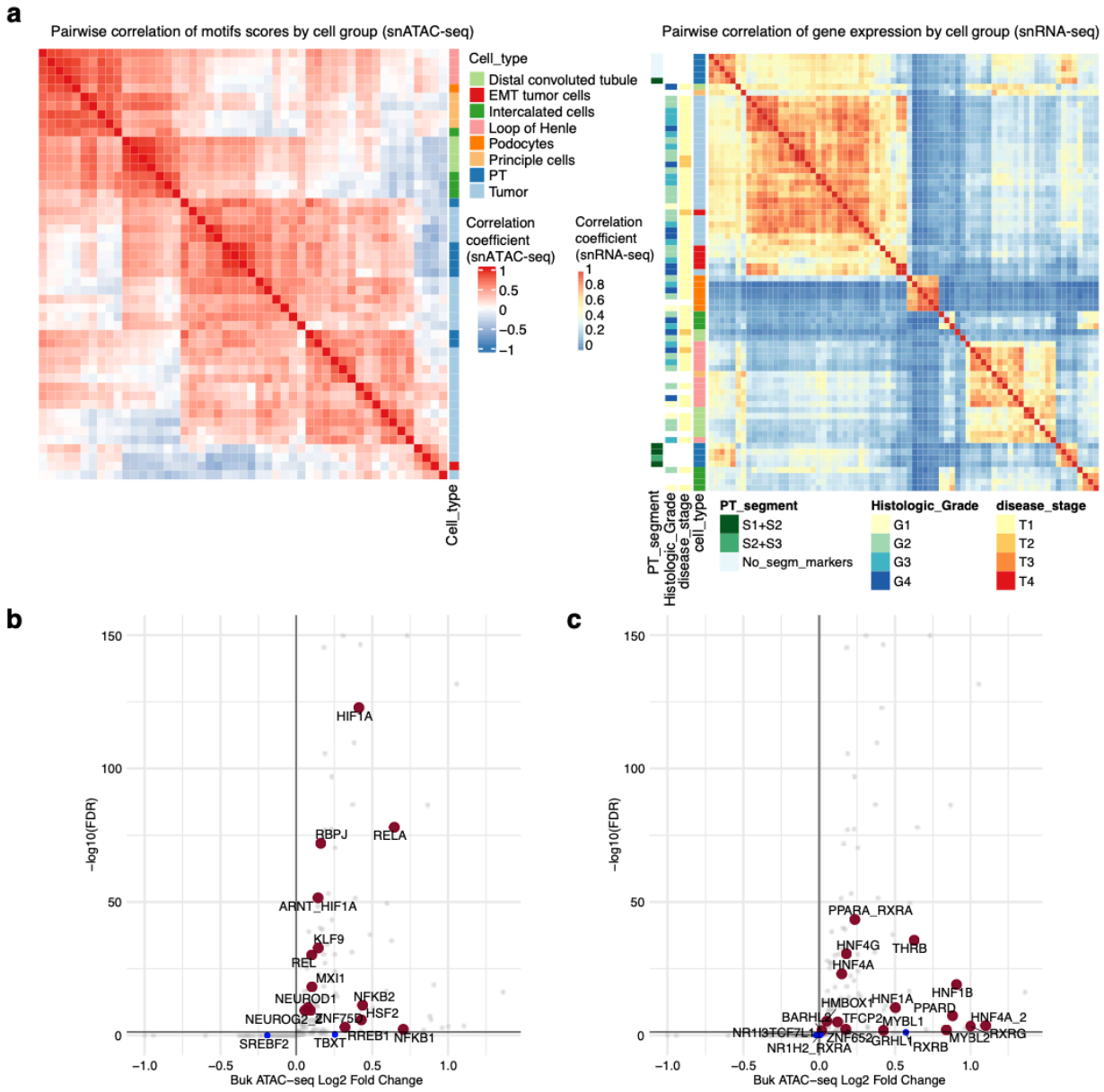
Extended Data Figure 3



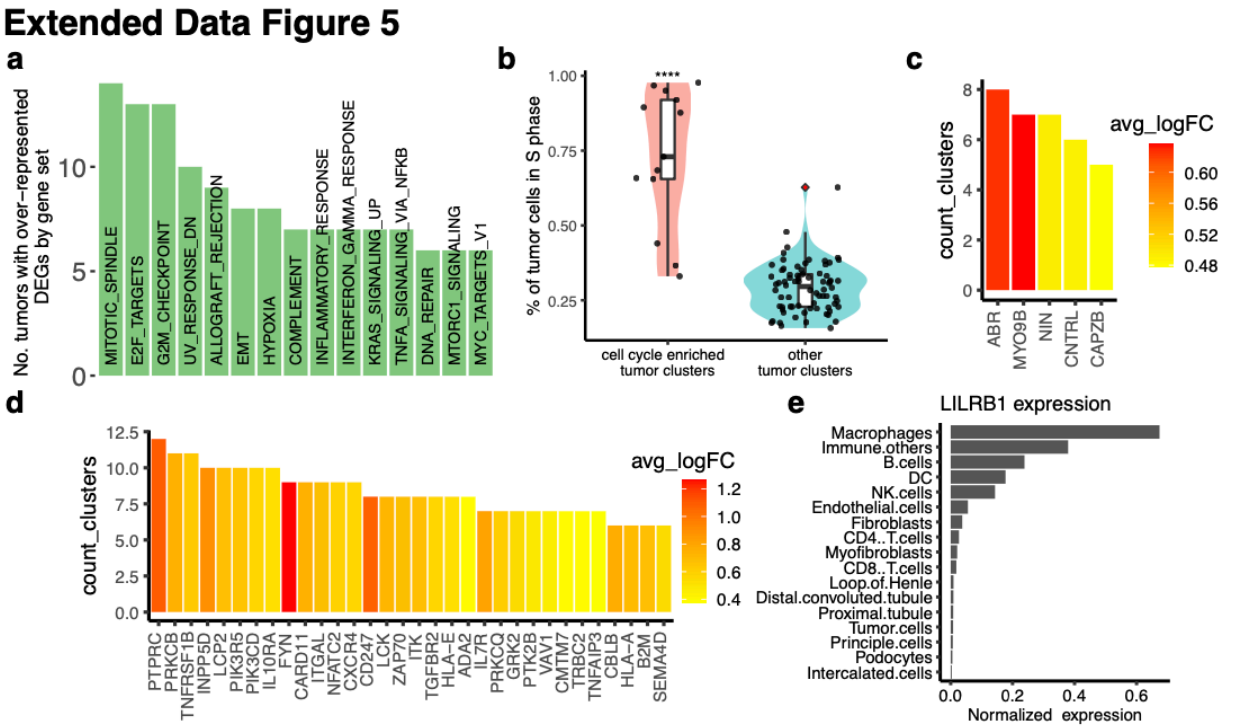
Extended Data Figure 3. Gene expression of CP, ABCC3, SLC6A3, and growth inhibition for the corresponding inhibitors in RCC cell lines and controls.

a, Heatmap showing the gene expression of *CP*, *ABCC3*, *SLC6A3*, and *CA9* in ccRCC cell lines (786-O-VHL+ RNA-seq data not available) and control cell lines. **b-d**, Growth inhibition curves in ccRCCC and control cell lines of **b**, MK571 (targeting ABCC3), **c**, tetrathiomolybdate (TM; targeting copper, related to CP signaling) and **d**, Benztropine (targeting SLC6A3) treatments. The error bars represent the standard error of the mean (SEM).

Extended Data Figure 4

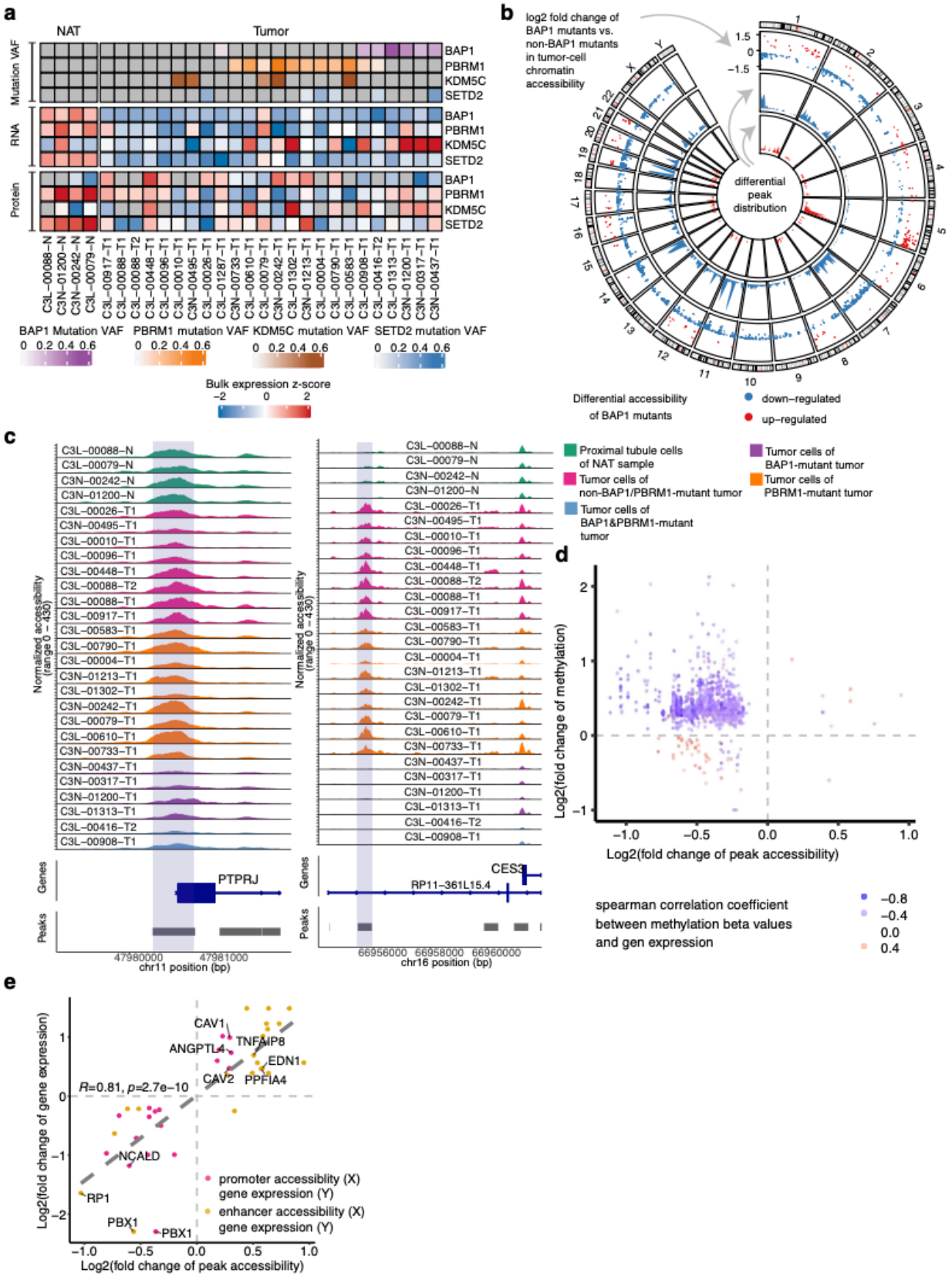


Extended Data Figure 4. Pairwise correlation between tumor cells and other normal nephron epithelial cell types, and bulk ATAC-seq validation of ccRCC-specific TF motifs. **a**, Left: Heatmap showing the pairwise correlation between tumor cells and other normal nephron epithelial cell types using snATAC-based TF motif scores. Right: Heatmap showing the pairwise correlation between tumor cells and other normal nephron epithelial cell types using snRNA-seq based gene expression. **b**, Volcano plot showing the snATAC-based ccRCC-specific TF motifs validated by the TCGA bulk ATAC-seq data (enriched in ccRCC compared to other cancer types, highlighted in red dots). Those that are not validated are in blue dots. **c**, Volcano plot showing the snATAC-based PT-specific TF motifs in the TCGA bulk ATAC-seq data. Red dots denote those that are significantly enriched in ccRCC compared to other cancer types. Those that are not enriched in ccRCC are in blue dots.



Extended Data Figure 5. Differential pathway activities among tumor subclusters.
a, Barplot showing the number of tumors in which certain pathways are over-represented in the differentially expressed genes among tumor subclusters. **b**, Violin plot showing tumor clusters annotated as cell-cycle-enriched have significantly higher fraction of tumor cells in S phase. **c**, Bar plot showing the genes frequently upregulated in the cell-cycle enriched tumor clusters compared to other tumor clusters in the same tumor, and the number of cell-cycle enriched tumor clusters with up-regulated expression of these genes. **d**, Bar plot showing the genes frequently upregulated in the immune-signaling enriched tumor clusters compared to other tumor clusters in the same tumor, and the number of immune-signaling enriched tumor clusters with up-regulated expression of these genes. **e**, Bar plot showing the expression of *LILRB1* is highest in macrophages compared to other cell types.

Extended Data Figure 6



Extended Data Figure 6. Overview of the snATAC-seq samples and *BAP1/PBRM1*-associated chromatin accessibility changes.
a, Heatmap showing the mutation status, gene expression and protein levels of the BAP1, PBRM1, KDM5C, and SETD2 among the snATAC-seq samples. **b**, Circos plot showing the distribution of ATAC-peaks associated with *BAP1* mutation. The outermost circles show the genomic location of individual ATAC-peaks associated with *BAP1* mutation, with red dots denoting more accessible peaks and blue dots denoting more closed peaks. The middle circle shows the density of more closed *BAP1*-associated peaks across the genome. The innermost circle shows the density of more *BAP1*-associated peaks across the genome. **c**, Genomic regions near two down-regulated genes (*PTPRJ* and *CES3*) in *BAP1*-mutated tumors vs. non-*BAP1*-mutated tumors. The plots show the normalized accessibility signal by snATAC-seq around these regions in tumor cells of *BAP1*-mutant tumor (purple), tumor cells of *PBRM1*-mutant tumor (orange), tumor cells of non-*BAP1*/*PBRM1*-mutant tumors (pink), and proximal tubule cells (green) from NAT samples from representative tumor samples. **d**, Scatter plot showing genes with both differential methylation and differential promoter accessibility in *BAP1*-mutated tumors vs. other tumors. The majority of these genes show hypermethylation and decreased promoter accessibility. The color of the dot denotes the correlation coefficient between the probe methylation (bulk) and the expression of the corresponding genes (bulk RNA-seq). **e**, Scatter plot showing the positive correlation of chromatin accessibility and transcriptional changes associated with *PBRM1* mutation. The fold change (log2) of the snRNA-seq expression for each gene (mRNA) is plotted against the fold change (log2) in the relative snATAC-seq peaks (for all the genes/peaks with significant fold change in over 50% of the comparisons for individual *PBRM1*-mutated tumor vs. tumors not mutated in either *BAP1* or *PBRM1*). Each dot represents a gene-peak pair. Dots are colored by whether the peak overlaps the gene promoter or is a potential enhancer (co-accessible with the promoter peak).

2.10 References

1. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
2. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, (2018).
3. Vanharanta, S. *et al.* Epigenetic expansion of VHL-HIF signal output drives multiorgan metastasis in renal cancer. *Nat. Med.* **19**, 50–56 (2013).
4. Yao, X. *et al.* VHL Deficiency Drives Enhancer Activation of Oncogenes in Clear Cell Renal Cell Carcinoma. *Cancer Discov.* **7**, 1284–1305 (2017).
5. Perroud, B., Ishimaru, T., Borowsky, A. D. & Weiss, R. H. Grade-dependent Proteomics Characterization of Kidney Cancer. *Mol. Cell. Proteomics MCP* **8**, 971–985 (2009).
6. Perroud, B. *et al.* Pathway analysis of kidney cancer using proteomics and metabolic profiling. *Mol. Cancer* **5**, 64 (2006).
7. Hu, S. L. *et al.* The Nephrologist's Tumor: Basic Biology and Management of Renal Cell Carcinoma. *J. Am. Soc. Nephrol. JASN* **27**, 2227–2237 (2016).
8. Linehan, W. M. & Ricketts, C. J. The Metabolic Basis of Kidney Cancer. *Semin. Cancer Biol.* **23**, 46–55 (2013).
9. Linehan, W. M., Srinivasan, R. & Schmidt, L. S. The genetic basis of kidney cancer: a metabolic disease. *Nat. Rev. Urol.* **7**, 277–285 (2010).
10. Clark, D. J. *et al.* Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* **179**, 964-983.e31 (2019).
11. Hakimi, A. A. *et al.* An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. *Cancer Cell* **29**, 104–116 (2016).
12. Wettersten, H. I. *et al.* Grade-dependent metabolic reprogramming in kidney cancer revealed by combined proteomics and metabolomics analysis. *Cancer Res.* **75**, 2541–2552 (2015).
13. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613–628 (2017).
14. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
15. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).
16. Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).
17. Zhang, Y. *et al.* Single-cell analyses of renal cell cancers reveal insights into tumor microenvironment, cell of origin, and therapy response. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2103240118 (2021).
18. Su, C. *et al.* Single-Cell RNA Sequencing in Multiple Pathologic Types of Renal Cell Carcinoma Revealed Novel Potential Tumor-Specific Markers. *Front. Oncol.* **11**, 719564 (2021).
19. Kim, K.-T. *et al.* Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. *Genome Biol.* **17**, 80 (2016).
20. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
21. Grindberg, R. V. *et al.* RNA-sequencing from single nuclei. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 19802–19807 (2013).

22. Buck, M. J. *et al.* Alterations in chromatin accessibility and DNA methylation in clear cell renal cell carcinoma. *Oncogene* **33**, 4961–4965 (2014).
23. Simon, J. M. *et al.* Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. *Genome Res.* **24**, 241–250 (2014).
24. Meng, F., Xiao, Y., Xie, L., Liu, Q. & Qian, K. Diagnostic and prognostic value of ABC transporter family member ABCG1 gene in clear cell renal cell carcinoma. *Channels* **15**, 375–385.
25. Scanlan, M. J. *et al.* Antigens recognized by autologous antibody in patients with renal-cell carcinoma. *Int. J. Cancer* **83**, 456–464 (1999).
26. Dhanabal, M. *et al.* Recombinant semaphorin 6A-1 ectodomain inhibits in vivo growth factor and tumor cell line-induced angiogenesis. *Cancer Biol. Ther.* **4**, 659–668 (2005).
27. Zhou, L., Luo, Z., li, Z. & Huang, Q. Circadian clock is associated with tumor microenvironment in kidney renal clear cell carcinoma. *Aging* **12**, 14620–14632 (2020).
28. Pelletier, S. *et al.* TGF-alpha as a candidate tumor antigen for renal cell carcinomas. *Cancer Immunol. Immunother. CII* **58**, 1207–1218 (2009).
29. Cao, Q. *et al.* Overexpression of PLIN2 is a prognostic marker and attenuates tumor progression in clear cell renal cell carcinoma. *Int. J. Oncol.* **53**, 137–147 (2018).
30. Xiao, Y. *et al.* The m6A RNA demethylase FTO is a HIF-independent synthetic lethal partner with the VHL tumor suppressor. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 21441–21449 (2020).
31. Hansson, J. *et al.* Overexpression of Functional SLC6A3 in Clear Cell Renal Cell Carcinoma. *Clin. Cancer Res.* **23**, 2105–2115 (2017).
32. Zhang, Z.-Y. *et al.* The up-regulation of NDRG1 by HIF counteracts the cancer-promoting effect of HIF in VHL-deficient clear cell renal cell carcinoma. *Cell Prolif.* **53**, e12853 (2020).
33. Zimpfer, A. *et al.* Ceruloplasmin expression in renal cell carcinoma correlates with higher-grade and shortened survival. *Biomark. Med.* **15**, 841–850 (2021).
34. Minner, S. *et al.* Epidermal growth factor receptor protein expression and genomic alterations in renal cell carcinoma. *Cancer* **118**, 1268–1275 (2012).
35. Doñate, F. *et al.* AGS16F Is a Novel Antibody Drug Conjugate Directed against ENPP3 for the Treatment of Renal Cell Carcinoma. *Clin. Cancer Res.* **22**, 1989–1999 (2016).
36. Xu, F. *et al.* The Oncogenic Role of COL23A1 in Clear Cell Renal Cell Carcinoma. *Sci. Rep.* **7**, 9846 (2017).
37. Harned, J. *et al.* Ceruloplasmin alters intracellular iron regulated proteins and pathways: ferritin, transferrin receptor, glutamate and hypoxia-inducible factor-1α. *Exp. Eye Res.* **97**, 90–97 (2012).
38. Dai, L. *et al.* SARI inhibits angiogenesis and tumour growth of human colon cancer through directly targeting ceruloplasmin. *Nat. Commun.* **7**, 11996 (2016).
39. Pan, Q. *et al.* Copper deficiency induced by tetrathiomolybdate suppresses tumor growth and angiogenesis. *Cancer Res.* **62**, 4854–4859 (2002).
40. Harlander, S. *et al.* Combined mutation in Vhl, Trp53 and Rb1 causes clear cell renal cell carcinoma in mice. *Nat. Med.* **23**, 869–877 (2017).
41. Gudas, L. J., Fu, L., Minton, D. R., Mongan, N. P. & Nanus, D. M. The role of HIF1α in renal cell carcinoma tumorigenesis. *J. Mol. Med. Berl. Ger.* **92**, 825–836 (2014).

42. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
43. Wettersten, H. I., Aboud, O. A., Lara, P. N. & Weiss, R. H. Metabolic reprogramming in clear cell renal cell carcinoma. *Nat. Rev. Nephrol.* **13**, 410–419 (2017).
44. Iyer, N. V. *et al.* Cellular and developmental control of O2 homeostasis by hypoxia-inducible factor 1 alpha. *Genes Dev.* **12**, 149–162 (1998).
45. Ryan, H. E., Lo, J. & Johnson, R. S. HIF-1 alpha is required for solid tumor formation and embryonic vascularization. *EMBO J.* **17**, 3005–3015 (1998).
46. Lu, S., Gu, X., Hoestje, S. & Epner, D. E. Identification of an additional hypoxia responsive element in the glyceraldehyde-3-phosphate dehydrogenase gene promoter. *Biochim. Biophys. Acta* **1574**, 152–156 (2002).
47. Xia, X. & Kung, A. L. Preferential binding of HIF-1 to transcriptionally active loci determines cell-type specific response to hypoxia. *Genome Biol.* **10**, R113 (2009).
48. Heerboth, S. *et al.* EMT and tumor metastasis. *Clin. Transl. Med.* **4**, 6 (2015).
49. Vu, T. & Datta, P. K. Regulation of EMT in Colorectal Cancer: A Culprit in Metastasis. *Cancers* **9**, E171 (2017).
50. mTORC1 signaling and the metabolic control of cell growth - PubMed. https://pubmed.ncbi.nlm.nih.gov/28411448/.
51. Tu, Y. *et al.* The Ubiquitin Proteasome Pathway (UPP) in the regulation of cell cycle control and DNA damage repair and its implication in tumorigenesis. *Int. J. Clin. Exp. Pathol.* **5**, 726–738 (2012).
52. Casimiro, M. C., Crosariol, M., Loro, E., Li, Z. & Pestell, R. G. Cyclins and cell cycle control in cancer and disease. *Genes Cancer* **3**, 649–657 (2012).
53. Peña-Llopis, S. *et al.* BAP1 loss defines a new class of renal cell carcinoma. *Nat. Genet.* **44**, 751–759 (2012).
54. Gu, Y.-F. *et al.* Modeling Renal Cell Carcinoma in Mice: Bap1 and Pbrm1 Inactivation Drive Tumor Grade. *Cancer Discov.* **7**, 900–917 (2017).
55. Nieto, M. A. Epithelial-Mesenchymal Transitions in development and disease: old views and new perspectives. *Int. J. Dev. Biol.* **53**, 1541–1547 (2009).
56. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* **119**, 1420–1428 (2009).
57. Kang, J. *et al.* Plasminogen activator inhibitor-1 enhances radioresistance and aggressiveness of non-small cell lung cancer cells. *Oncotarget* **7**, 23961–23974 (2016).
58. Yoo, H.-J. *et al.* Genetic and expression alterations in association with the sarcomatous change of cholangiocarcinoma cells. *Exp. Mol. Med.* **41**, 102–115 (2009).
59. Gujral, T. S. *et al.* A Noncanonical Frizzled2 Pathway Regulates Epithelial-Mesenchymal Transition and Metastasis. *Cell* **159**, 844–856 (2014).
60. Huang, R. Y.-J., Guilford, P. & Thiery, J. P. Early events in cell adhesion and polarity during epithelial-mesenchymal transition. *J. Cell Sci.* **125**, 4417–4422 (2012).
61. Maschler, S. *et al.* Tumor cell invasiveness correlates with changes in integrin expression and localization. *Oncogene* **24**, 2032–2041 (2005).
62. Singh, V., Singla, S. K., Jha, V., Puri, V. & Puri, S. Hepatocyte nuclear factor-1β: A regulator of kidney development and cystogenesis. *Indian J. Nephrol.* **25**, 70–76 (2015).
63. Martovetsky, G., Tee, J. B. & Nigam, S. K. Hepatocyte Nuclear Factors 4α and 1α Regulate Kidney Developmental Expression of Drug-Metabolizing Enzymes and Drug Transporters. *Mol. Pharmacol.* **84**, 808–823 (2013).

64. Bildsoe, H. *et al.* Transcriptional targets of TWIST1 in the cranial mesoderm regulate cell-matrix interactions and mesenchyme maintenance. *Dev. Biol.* **418**, 189–203 (2016).
65. He, H. *et al.* c-Jun/AP-1 overexpression reprograms ERα signaling related to tamoxifen response in ERα-positive breast cancer. *Oncogene* **37**, 2586–2600 (2018).
66. Scelo, G. *et al.* Variation in genomic landscape of clear cell renal cell carcinoma across Europe. *Nat. Commun.* **5**, 5135 (2014).
67. 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. *Lancet Oncol.* **14**, 159–167 (2013).
68. 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. Off. J. Am. Assoc. Cancer Res.* **19**, 3259–3267 (2013).
69. Peña-Llopis, S., Christie, A., Xie, X.-J. & Brugarolas, J. Cooperation and antagonism among cancer genes: the renal cancer paradigm. *Cancer Res.* **73**, 4173–4179 (2013).
70. Tamburri, S. *et al.* Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. *Mol. Cell* **77**, 840-856.e5 (2020).
71. Campagne, A. *et al.* BAP1 complex promotes transcription by opposing PRC1-mediated H2A ubiquitylation. *Nat. Commun.* **10**, 348 (2019).
72. D'Agostino, S. *et al.* The receptor protein tyrosine phosphatase PTPRJ negatively modulates the CD98hc oncoprotein in lung cancer cells. *Oncotarget* **9**, 23334–23348 (2018).
73. Yen, M. *et al.* Transposase mapping identifies the genomic targets of BAP1 in uveal melanoma. *BMC Med. Genomics* **11**, 97 (2018).
74. Bleu, M. *et al.* PAX8 activates metabolic genes via enhancer elements in Renal Cell Carcinoma. *Nat. Commun.* **10**, 3739 (2019).
75. Droz, D. *et al.* Expression of the human nephron differentiation molecules in renal cell carcinomas. *Am. J. Pathol.* **137**, 895–905 (1990).
76. Paraf, F. *et al.* Renal lesions in von Hippel-Lindau disease: immunohistochemical expression of nephron differentiation molecules, adhesion molecules and apoptosis proteins. *Histopathology* **36**, 457–465 (2000).
77. Davis, C. F. *et al.* The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* **26**, 319–330 (2014).
78. Tsao, C. C. *et al.* Inhibition of Mxi1 suppresses HIF-2alpha-dependent renal cancer tumorigenesis. *Cancer Biol. Ther.* **7**, 1619–1627 (2008).
79. Fendler, A. *et al.* Inhibiting WNT and NOTCH in renal cancer stem cells and the implications for human patients. *Nat. Commun.* **11**, 929 (2020).
80. Zhou, Q. *et al.* SNX5 suppresses clear cell renal cell carcinoma progression by inducing CD44 internalization and epithelial-to-mesenchymal transition. *Mol. Ther. - Oncolytics* **24**, 87–100 (2022).
81. Huang, C. *et al.* The miR-140-5p/KLF9/KCNQ1 axis promotes the progression of renal cell carcinoma. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **34**, 10623–10639 (2020).
82. Meteoglu, I., Erdogdu, I. H., Meydan, N., Erkus, M. & Barutca, S. NF-KappaB expression correlates with apoptosis and angiogenesis in clear cell renal cell carcinoma tissues. *J. Exp. Clin. Cancer Res.* **27**, 53 (2008).
83. Morais, C., Gobe, G., Johnson, D. W. & Healy, H. The emerging role of nuclear factor kappa B in renal cell carcinoma. *Int. J. Biochem. Cell Biol.* **43**, 1537–1549 (2011).

84. Eickelberg, O. *et al.* Functional activation of heat shock factor and hypoxia-inducible factor in the kidney. *J. Am. Soc. Nephrol. JASN* **13**, 2094–2101 (2002).
85. Lhoták, S. *et al.* ER stress contributes to renal proximal tubule injury by increasing SREBP-2-mediated lipid accumulation and apoptotic cell death. *Am. J. Physiol. Renal Physiol.* **303**, F266-278 (2012).
86. Yeung, S. J., Pan, J. & Lee, M.-H. Roles of p53, MYC and HIF-1 in regulating glycolysis - the seventh hallmark of cancer. *Cell. Mol. Life Sci. CMLS* **65**, 3981–3999 (2008).
87. Li, L. *et al.* Transcriptional Regulation of the Warburg Effect in Cancer by SIX1. *Cancer Cell* **33**, 368-385.e7 (2018).
88. Corn, P. G. *et al.* Mxi1 is induced by hypoxia in a HIF-1-dependent manner and protects cells from c-Myc-induced apoptosis. *Cancer Biol. Ther.* **4**, 1285–1294 (2005).
89. Brunelle, J. K. *et al.* c-Myc sensitization to oxygen deprivation-induced cell death is dependent on Bax/Bak, but is independent of p53 and hypoxia-inducible factor-1. *J. Biol. Chem.* **279**, 4305–4312 (2004).
90. Zhang, H. *et al.* HIF-1 inhibits mitochondrial biogenesis and cellular respiration in VHL-deficient renal cell carcinoma by repression of C-MYC activity. *Cancer Cell* **11**, 407–420 (2007).
91. Oya, M. *et al.* Increased nuclear factor-kappa B activation is related to the tumor development of renal cell carcinoma. *Carcinogenesis* **24**, 377–384 (2003).
92. Oya, M. *et al.* Constitutive activation of nuclear factor-kappaB prevents TRAIL-induced apoptosis in renal cancer cells. *Oncogene* **20**, 3888–3896 (2001).
93. Djordjević, G. *et al.* Relationship between vascular endothelial growth factor and nuclear factor-kappaB in renal cell tumors. *Croat. Med. J.* **49**, 608–617 (2008).
94. An, J., Sun, Y., Fisher, M. & Rettig, M. B. Maximal apoptosis of renal cell carcinoma by the proteasome inhibitor bortezomib is nuclear factor-kappaB dependent. *Mol. Cancer Ther.* **3**, 727–736 (2004).
95. Ikegami, A. *et al.* Knockdown of NF-κB1 by shRNA Inhibits the Growth of Renal Cell Carcinoma In Vitro and In Vivo. *Oncol. Res.* **26**, 743–751 (2018).
96. Kawauchi, K., Araki, K., Tobiume, K. & Tanaka, N. p53 regulates glucose metabolism through an IKK-NF-kappaB pathway and inhibits cell transformation. *Nat. Cell Biol.* **10**, 611–618 (2008).
97. Liu, X. *et al.* Warburg effect revisited: an epigenetic link between glycolysis and gastric carcinogenesis. *Oncogene* **29**, 442–450 (2010).
98. Díaz-Trelles, R. *et al.* Notch-independent RBPJ controls angiogenesis in the adult heart. *Nat. Commun.* **7**, 12088 (2016).
99. Hori, K. *et al.* A nonclassical bHLH Rbpj transcription factor complex is required for specification of GABAergic neurons independent of Notch signaling. *Genes Dev.* **22**, 166–178 (2008).
100. Masui, T., Long, Q., Beres, T. M., Magnuson, M. A. & MacDonald, R. J. Early pancreatic development requires the vertebrate Suppressor of Hairless (RBPJ) in the PTF1 bHLH complex. *Genes Dev.* **21**, 2629–2643 (2007).
101. Zhang, G. *et al.* RBPJ contributes to the malignancy of glioblastoma and induction of proneural-mesenchymal transition via IL-6-STAT3 pathway. *Cancer Sci.* **111**, 4166–4176 (2020).
102. Lv, Q., Shen, R. & Wang, J. RBPJ inhibition impairs the growth of lung cancer. *Tumor Biol.* **36**, 3751–3756 (2015).

103. Kulic, I. *et al.* Loss of the Notch effector RBPJ promotes tumorigenesis. *J. Exp. Med.* **212**, 37–52 (2014).
104. Suthon, S., Perkins, R. S., Bryja, V., Miranda-Carboni, G. A. & Krum, S. A. WNT5B in Physiology and Disease. *Front. Cell Dev. Biol.* **9**, 667581 (2021).
105. Samanta, S. *et al.* IMP3 Stabilization of WNT5B mRNA Facilitates TAZ Activation in Breast Cancer. *Cell Rep.* **23**, 2559–2567 (2018).
106. Harada, T. *et al.* Wnt5b-associated exosomes promote cancer cell migration and proliferation. *Cancer Sci.* **108**, 42–52 (2017).
107. Szczepanski, A. P. & Wang, L. Emerging multifaceted roles of BAP1 complexes in biological processes. *Cell Death Discov.* **7**, 20 (2021).
108. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
109. Xi, R., Lee, S., Xia, Y., Kim, T.-M. & Park, P. J. Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants. *Nucleic Acids Res.* **44**, 6274–6286 (2016).
110. Vasaikar, S. *et al.* Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* **177**, 1035-1049.e19 (2019).
111. Bausch-Fluck, D. *et al.* A mass spectrometric-derived cell surface protein atlas. *PloS One* **10**, e0121314 (2015).
112. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
113. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
114. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* **71**, 858-871.e8 (2018).
115. Leporcq, C. *et al.* TFmotifView: a webserver for the visualization of transcription factor motifs in genomic regions. *Nucleic Acids Res.* **48**, W208–W217 (2020).
116. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
117. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
118. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **16**, 284–287 (2012).
119. Ye, X. *et al.* Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256–260 (2015).
120. Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349–361 (2012).
121. Krebs, A. M. *et al.* The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. *Nat. Cell Biol.* **19**, 518–529 (2017).

Chapter 3: Sapanisertib and cabozantinib combination results in potent antitumor activity in renal cell carcinoma patient-derived xenografts

This chapter is adapted from a manuscript in preparation. Contribution: I lead the project management, data analysis and figure generation for the drug-treated tumor volumes, downstream analysis for all the omics-related results as a first-author of the paper. Details of other colleagues' contribution can be seen in the "Author Contributions" section.

3.1 Abstract

Patient-derived xenograft (PDX) models have proven valuable in studying treatment effects, mechanisms, and novel therapeutics for cancer, including renal cell carcinoma (RCC). We performed a series of omics-guided drug tests on a set of RCC PDX models, in which cabozantinib and sapanisertib are the two most effective drugs. The combination of cabozantinib and sapanisertib treatment induced tumor growth arrest or regression for all 6 models and it's better tolerated compared to treatment with single agents. Cabozantinib plus sapanisertib can inhibit tumor growth by reducing vascular density. Proteomics analysis reveals increased epithelial-mesenchymal transition pathway may be the potential drug resistance mechanism in the residual tumor after the combination treatment. Finally, we identified MET baseline protein levels that may serve to predict treatment response to cabozantinib, with support from single-nucleus RNA-seq data. This study proposed a potential new treatment option for RCC patients and revealed molecular alterations underlying tumor reduction induced by the combination of cabozantinib and sapanisertib treatment.

3.2 Introduction

Renal cell carcinoma (RCC) originates from cells in the renal epithelium and accounts for over 90% of cancers in the kidney[1]. It includes several subtypes, among which clear cell RCC (ccRCC) is the most common subtype (75%). For the past 20 years, several targeted agents and one immunotherapy agent have been approved for the treating metastatic RCC. The major classes of targeted therapy include Vascular Endothelial Growth Factor (VEGF) ligand antibody (bevacizumab), tyrosine kinase inhibitors, and mammalian target of rapamycin (mTOR) inhibitors, while the immunotherapy agent is a programmed death-1 inhibitor (nivolumab). Although all of these drugs are effective to some extent, a big proportion of ccRCC patients display intrinsic resistance to targeted therapies[2]. For those who responded well at first, the vast majority will eventually develop acquired resistance to these treatments[2,3]. This fact points to combinational therapy as a way to overcome such resistance.

Patient-derived xenograft (PDX) models have proven valuable in the quest toward personalized medicine for studying treatment effects, mechanisms, and novel therapeutics for cancer. Here, we performed a series of drug treatment effect tests on a set of RCC PDX models, ultimately identifying cabozantinib and sapanisertib as the two most effective agents. Cabozantinib (Cabometyx, by Exelixis) is a relatively new receptor tyrosine kinase (RTK) inhibitor approved for advanced renal cell carcinoma targeting VEGFR, mesenchymal-epithelial transition factor (c-MET), and AXL[4]. It was recently approved as a first-line treatment for advanced RCC. Although cabozantinib rescues sunitinib resistance in some patients, resistance eventually develops.

Sapanisertinib (TAK-228, MLN0128, by Takeda/Millenium) is an experimental small-molecule inhibitor of mTOR, targeting both mTORC1 and mTORC2[5,6]. There are several phase-I and

phase-II studies of sapanisertib for multiple non-RCC solid cancer types (NCT03047213, NCT02417701, NCT02244463, NCT03430882, NCT03017833, NCT02159989). Sapanisertib is tolerable and has some activity in NFE2L2 and KEAP1 mutant squamous cell lung cancer patients[7]. On the other hand, sapanisertib did not result in an objective response in bladder cancer[8]. Although mTOR activation is well known in RCC, and mTOR inhibitors such as rapamycin analogs (rapalog) have been used to treat advanced RCC in the clinic, they mostly target mTOR complex 1 (mTORC1) rather than mTOR complex 2 (mTORC2), and the corresponding patient response varies[9]. Therefore, we hypothesized that a newer class of drugs that target both mTORC1 and mTORC2, such as Sapanisertib, would likely show good efficacy by inhibiting mTOR more completely. To provide potential novel treatment options to the clinic, we also studied the treatment effects of combined drugs and explored potential molecular mechanisms for such effects. Our results showed the combination of cabozantinib plus sapanisertib has the best efficacy among all combinations tested. One model even showed close to a 70% reduction of tumor volume after 8-week treatment. The combination inhibits tumor growth by reducing vascular density and the cell cycle-related targets of E2F transcription factors. Based on the omics results, the resistance to the combination treatment may arise from the activation of the epithelial-mesenchymal transition pathway. In general, the omics results are valuable in explaining the molecular alterations underlying treatment effects and in predicting potential drug resistance mechanisms.

3.3 Results

3.3.1 Study overview and PDX treatment response

We applied 10 drugs to a set of 6 RCC PDX models, based on evidence of activated genes and factors in a number of pathways in ccRCC from previous studies[10–19]. The clinical and genetic features of the 6 PDX models are described in Tables 1 and 2. The drugs we have used in this study were RTK inhibitors cabozantinib (targets VEGFRs, c-MET, RET, KIT, AXL) and sunitinib (targets VEGFRs, PDGFRb), mTOR inhibitor sapanisertib (targets mTORC1 and mTORC2), HDAC inhibitors panobinostat (targets broad spectrum HDACs) and beta-hydroxybutyrate (BHB; suspected to have HDAC inhibitor and other effects), HIF inhibitor acriflavine hydrochloride (targets HIF1a and HIF2a), CDK inhibitor abemaciclib, MEK inhibitor selumetinib, IAP inhibitor birinapant, and angiotensin receptor blocker losartan potassium (Fig. 1A). Among these 10 drugs, only cabozantinib and sapanisertib showed effective responses across all PDX lines (Fig. 1A), meaning the drugs can inhibit tumor growth or keep it at a lower growth rate than the control. This promising result prompted us to test the combined effects of these two drugs in comparison to the single treatment and perform sequencing to further explore the mechanisms underlying the tumor reduction (Fig. 1B).

Model ID	Patient age	Patient gender	Tumor stage	Clinical event point at collection	Site of collection	Subtype
RESL3	55-59	Male	pT3aNxMx	Metastasis	Lung	ccRCC
RESL4	60-64	Male	pT3aN0	Metastasis	Brain	ccRCC
RESL5	50-54	Female	pT2N0	Metastasis	Lymphnode	ccRCC
RESL10	45-49	Male	pT3aNxMx	Metastasis	Kidney	ccRCC
RESL11	90-94	Male	pT4NxMx	Primary		pRCC
RESL12	65-69	Male	cT3aNxM1	Metastasis	Brain	ccRCC

Table 1. Clinical features of the patients

Model ID	Somatic mutation (amino acid change)					Somatic copy number altaration (log2 copy number ratio) [1]				
	VHL	PBRM1	PIK3CA	KDM5C	ARID1A	Chr. 3p	Chr. 5q	Chr. 14q	Chr. 7	Chr. 17
RESL3	L89H			R68Efs*5		-1.00	0.98	0.00	-2.01	0.00
RESL4			D350G		Q594*	-0.64	0.70	0.65	1.08	1.32
RESL5						-0.70	0.16	-0.95	2.18	-0.01
RESL10	H125Qfs*7		H1047R			-0.76	0.38	-0.72	1.64	0.63
RESL11						0.00	0.00	0.00	0.00	0.99
RESL12	V166D	Q779*				0.02	0.03	-0.98	1.62	-0.23

Table 2. Genetic features of the PDX models

For cabozantinib treatment, all models showed some extent of response when compared to vehicle-treated controls. Cabozantinib-treated tumor volumes were lower than the vehicle-treated tumor volumes in all 6 models after 1-month treatment (day 27-30) and the differences were significant for RESL4, RESL5, RESL10, and RESL12 (FDR < 0.1), albeit at variable levels (Fig. 1C). For example, when comparing PDX models RESL5 and RESL10 specifically, there was an obvious difference: cabozantinib caused tumor regression in RESL5, while only slowing tumor growth rate in RESL10. When taking a closer look at days 47/48, cabozantinib reduced tumor growth with tumor growth inhibition (TGI) of 84.6% and 43.8% in RESL5 and RESL10, respectively, relative to the control (Methods), resulting in tumor regression in RESL5 with a

relative tumor volume (RTV, relative to day 0) of 75.8% and sustained tumor growth in RESL10 with an RTV of 274.2% (Table 3). On the other hand, sapanisertib showed reversed treatment effects in RESL5 and RESL10, when compared to cabozantinib. To be specific, RESL10 showed tumor growth arrest to sapanisertib with an RTV of 121.2% (TGI = 75.1%), while RESL5 showed sustained tumor growth with an RTV of 212.0% (TGI = 57.1%) on days 47/48. Overall, Sapanisertib-treatment tumor volumes were lower than the vehicle-treated tumor volumes in all 6 models after 1-month treatment (day 27-30) and the differences were significant for 4 out of 6 models (RESL4, RESL5, RESL10, and RESL12; FDR < 0.1). When we combined cabozantinib and sapanisertib, 5 out of 6 models showed significantly lower RTVs compared to controls on days 27-31 (FDR < 0.1; except for RESL11 due to only 1 data point). And the combinational therapy was significantly more effective than at least one of the monotherapies in 5 models (RESL5, RESL10, RESL3, RESL4, and RESL12; FDR < 0.1). On days 47/48, RESL12 and RESL5 showed 72.2% and 57.3% reduction in RTV compared to day 0, respectively, while RESL10 and RESL4 showed tumor stabilization (RTV = 88.9% and 110.0% respectively). While toxicity is usually a major concern for combinational therapy, the combination of cabozantinib and sapanisertib appears to be well-tolerated as the co-treatment did not induce dramatically reduced body weight compared to the control (Fig. 1D).

Model	Treatment length (X days)	Relative tumor volume (day X vs. day 0)				Tumor growth inhibition		
		vehicle	Cab	Sap	Cab+Sap	Cab	Sap	Cab+Sap
RESL3	28	351.2%	161.2%	132.4%	80.2%	54.1%	62.3%	77.2%
RESL4	28	510.7%	187.2%	137.1%	78.7%	63.3%	73.2%	84.6%
RESL5	27	229.5%	84.4%	160.5%	51.5%	63.2%	30.1%	77.6%
RESL10	28	283.8%	184.1%	136.6%	96.5%	35.1%	51.9%	66.0%
RESL11	28	214.8%	118.1%	114.7%	71.3%	45.0%	46.6%	66.8%
RESL12	28	223.7%	118.9%	140.4%	41.0%	46.8%	37.2%	81.7%
RESL4	46		248.3%	310.1%	110.0%			
RESL5	48	493.6%	75.8%	212.0%	42.7%	84.6%	57.1%	91.3%
RESL10	47	487.5%	274.2%	121.2%	88.9%	43.8%	75.1%	81.8%
RESL12	48	456.0%	169.0%	166.0%	27.8%	62.9%	63.6%	93.9%

Table 3. Relative tumor volume and tumor growth inhibition statistics

We also identified some potential connections between the patient's clinical and genetic features and the corresponding PDX model responses to the two drugs. RESL10 was derived from a patient who relapsed after a 10-month treatment of axitinib (targets VEGFRs) plus avelumab (targets PD-L1) and a 2.5-month cabozantinib treatment (Table 4). RESL4 was derived from a patient receiving lenvatinib and everolimus concurrent to the tumor collection. The other 4 patients had no recorded targeted therapy before tumor collection. RESL10 and RESL4 showed the highest RTV during the two-month cabozantinib treatment (Table 3), suggesting a similar suboptimal response to cabozantinib in both PDX models RESL10 and RESL4 and the corresponding patients. We also observed that RESL10 carries a 'hotspot' PIK3CA mutation (H1047R) and RESL4 carries a PIK3CA D350G mutation (Table 2). The activation of PI3K signaling in RESL10 and RESL4 was supported by the increased AKT (Ser 473) phosphorylation compared to the PIK3CA wild type RESL5 (Fig. 1E). The PIK3CA H1047R mutation has been associated with resistance to cabozantinib in NIH3T3 fibroblast cells and head and neck cancer cells[20], which is consistent with observations in our RCC PDX model RESL10. On the other hand, RESL10 showed the lowest RTV to sapanisertib. This is consistent with a

previous study that showed PIK3CA H1047R mutation is sensitive to sapanisertib[21]. For RESL4, although it also carries a PIK3CA activating mutation, its lack of response to sapanisertib may be potentially due to the resistances developed from its previous everolimus (mTORC1 inhibitor) treatment (Table 4).

Case ID	Model ID	Event Index	Regimen	Status	Duration (Months)	Reason Stopped	Clinical Response	Relation to PDX Collection	Comments
WUR-016	RESL3	1	SUNITINIB MALATE	Ongoing	-	-	Non-evaluable	After	
WUR-016	RESL3	2	Axitinib	Ended	2.63	Adverse events/side effects	Non-evaluable	After	
WUR-016	RESL3	3	Pazopanib hydrochloride	Ongoing	-	-	Non-evaluable	After	
WUR-065	RESL4	1	Lenvatinib, Everolimus	Ended	12.68	Adverse events/side effects	Non-evaluable	Concurrent	
WUR-065	RESL4	2	Ipilimumab, Nivolumab	Ended	0.26	Adverse events/side effects	Off study	After	Off this treatment due to patient preference.
WUR-065	RESL4	3	Lenvatinib, Everolimus	Ongoing	-	-	Non-evaluable	After	
WUR-014	RESL5	1	Ipilimumab, Nivolumab	Ended	8.97	Disease progression	Non-evaluable	After	
WUR-014	RESL5	2	-	Ongoing	-	-	Not applicable	After	
WUR-012	RESL10	1	Axitinib, Avelumab	Ended	10.35	Disease progression	Non-evaluable	Before	
WUR-012	RESL10	2	Cabozantinib	Ended	2.5	Disease progression	Non-evaluable	Before	
WUR-012	RESL10	3	Radiation	Ended	0.23	Completed therapy	Non-evaluable	Before	
WUR-012	RESL10	4	Gemcitabine	Ended	0.72	Death on study	Non-evaluable	After	
WUR-106	RESL11	1	Pazopanib hydrochloride	Ended	1.05	Death on study	Non-evaluable	After	
WUR-154	RESL12	1	Ipilimumab, Nivolumab	Ended	2.01	Completed therapy	Non-evaluable	After	

Table 4. Treatment histories of the patients

3.3.2 Cabozantinib and sapanisertib combination reduces vascular density

To further understand how the 6 PDX lines are organized and structured at the tissue level, we performed H&E staining on collected tumor samples (Fig. 2A). As shown in the clinical information table (Table 1), RESL3, RESL4, RESL5, and RESL12 are T2-T3 ccRCCs; RESL10 is a ccRCC line with sarcomatoid and rhabdoid features; RESL11 is papillary renal cell carcinoma. In concordance with the clinical information, RESL3 and RESL5 exhibited typical clear cell characteristics while RESL4 and RESL12 are less typical; the sarcomatoid and

papillary structures in RESL10 and RESL11 look representative and convincing. Based on the treatment response curve in Figure 1, we observed that RESL5 responded well to cabozantinib while RESL10 showed resistance. Among the 6 PDX models, RESL5 and RESL10 showed the most differential architectural patterns. RESL5 exhibited a low-grade, classical nested clear cell pattern, with the optically transparent cytoplasm and a rich capillary vascular network enclosing the cells (Figure 2B). Vascularity for RESL10 was not as easy to identify as for RESL5 by H&E staining. But the spindle-shaped cells and nuclei underscored the classical sarcomatoid feature for this line, as indicated by the triangles in Figure 2B. Furthermore, by applying immunofluorescence staining, the positive staining for CA9 in the spindle cell areas also supported that RESL10 is a PDX line retained clear cell RCC with sarcomatoid features (Figure 3A). In addition to this, the positivity for Ki67 highlighted the proliferating spindle-shaped nuclei, which co-stained with CA9 within the same cells (Figure 3A). In comparison to RESL10, the majority of the nuclei in RESL5 were more round-shaped and there was significantly higher expression of CA9 on the cell membrane.

As cabozantinib has been reported to have an anti-angiogenic effect[22], we want to investigate vascular density changes in response to cabozantinib and sapanisertib. Specifically, we performed immunofluorescence staining with CD31 on control and treated RESL5/10 samples. We first noted that the positive CD31 staining in RESL10 exposed the rich vascular network, which was somewhat obscured in H&E staining. This further demonstrated the diagnostically essential characteristic for sarcomatoid tumors with clear cell RCC origin (Figure 3B). Next, we compared CD31 expression densities between control and after treatment. As shown in Figure 3B, for both RESL5 and RESL10, cabozantinib significantly reduced the vascular density, while

sapanisertib treatment did not show obvious changes in vessels. This is consistent with the fact that cabozantinib inhibits the tyrosine kinase activity of MET, VEGFRs, and several other receptors that are actively involved in tumor angiogenesis. As an mTOR inhibitor, sapanisertib mainly works on regulating cell metabolism and proliferation. So it is expected that sapanisertib would not cause the change in vascular density. When cabozantinib and sapanisertib were used together, the decrease in vessel density was still obvious. This indicates that adding sapanisertib does not generally affect the excellent performance of cabozantinib inhibiting tumor angiogenesis. Lastly, by quantifying the expression density of CD31, we found that RESL5 was more responsive to cabozantinib than RESL10 (Figure 3C; $P < 0.0001$). After cabozantinib treatment, the expression density of CD31 in RESL5 decreased by 94% compared to control; as for RESL10, the density dropped only by 75%.

3.3.3 Cabozantinib plus sapanisertib treatment inhibits cell cycle related proteins while inducing proteins related to epithelial-mesenchymal transition

To obtain a mechanistic understanding of the observed tumor regressions after cabozantinib-sapanisertib co-treatment, we collected tumors at baseline and under four treatment regimens, namely vehicle, cabozantinib, sapanisertib, and the combination therapy, for 1-month and two-month periods, subsequently performing bulk whole-exome sequencing (WES), bulk RNA sequencing (RNA-seq), bulk proteomics and phosphoproteomics, and single-nucleus RNA-seq (snRNA-seq) on selected samples (Fig. 1B). The combination resulted in differential expression of 310 human proteins compared to the control (P-value < 0.05, difference in log2 intensity > 0.1), inducing more down-regulation than up-regulation (97 up + 213 down) (Fig. 4A). To identify the pathways altered by the combination treatment, we performed pathway over-representation tests on the differentially expressed proteins and examined the top 5 pathways

enriched in differentially expressed proteins (Fig. 4B). Proteins that decreased after the combination treatment are enriched in translation, oxidative phosphorylation, protein localization, E2F targets, and bile acid metabolism pathways. Decreased proteins in the translation include a large number of mitochondrial ribosomal proteins (MRPL1/3/9/11/14/16/19/41/43 and MRPS9/10/31). Oxidative phosphorylation is another mitochondrial process having a myriad of decreased proteins (TIMM8B, GOT2, HCCS, UQCRQ, SLC25A20, NDUFB1, COX7C, TOMM22, SLC25A5, UQCRB, UQCR10, COX4I1, COX17, MRPL11, DLD). Mitochondrial oxidative phosphorylation is already reduced in primary ccRCC[23], our results suggest the translational and respiratory capacities of mitochondria are further reduced after combination therapy. We also observed decreased protein levels of cell cycle-related targets of E2F transcription factors (TK1, CDK1, KPNA2, MCM2, PRDX4, PCNA, TBRG4, MCM3, RFC3, MCM7), which may explain the tumor growth arrest/regression induced by the combination treatment.

On the other hand, proteins increased after the combination treatment are enriched in epithelial-mesenchymal transition (EMT), amino and nucleotide sugar metabolism, G2M checkpoint, nuclear envelope breakdown, and cellular trafficking proteins pathways (Fig. 4B). In particular, the combination treatment increased the protein levels of EMT pathway members TGFBI, LRP1, COL7A1, TAGLN, SCG2, THBS1, and TFPI2 compared to the control. Moreover, the combination treatment also increased the protein levels of G2M checkpoint proteins CDK4, TNPO2, RPS6KA5, PRMT5, and NUP98. EMT has been known to play an important role in cancer progression, metastasis, and drug resistance[24,25]. The up-regulation of EMT and G2M

checkpoint proteins may be a survival adaptation to the blockade of RTK and mTOR signaling by the combination treatment.

We were interested to see the specific effects of combination treatment on the tumor proteome, so we compared the combination-treated PDX tumor proteome to post-monotherapy proteomes. 8 proteins showed significantly lower levels in combination-treated tumors compared to both cabozantinib-treated and sapanisertib-treated tumors, namely IGF2BP3, DNAJC7, PYCR1, CHMP6, MRM3, ERO1B, CLIC6, and COBLL1 (Fig. 4C). Using the proteomics data from an independent ccRCC cohort (81 cases), we found that IGF2BP3, PYCR1, and ERO1B showed significantly higher protein levels in tumor samples than in the matching normal adjacent tissue ($P < 1e\text{-}10$, Fig. 4D). Kaplan–Meier analysis showed a significant association between IGF2BP3 protein levels and survival (Fig. 4E; $P < 0.05$). Increased IGFBP3 protein may have an oncogenic role in ccRCC, which would suggest that their reduced levels caused by combination treatment contribute to the more effective tumor inhibition as compared to the two monotherapies.

3.3.4 Baseline MET protein levels predict response to cabozantinib

We would like to know whether the baseline levels of proteins and phosphorylation sites in the models may be associated with their response to the drug treatment, especially proteins within key pathways affected by cabozantinib and sapanisertib respectively. These proteins/pathways include RTKs and their corresponding ligands (such as MET, HGF, VEGFRs, and VEGFs), PI3K/AKT/mTOR pathway, and Ras/Raf/MAPK pathway[26] (Fig. 5A-B). Among them, we found MET protein levels are highest in RESL10, which is the least sensitive to cabozantinib (TGI = 35.1% at day 28), and lowest in RESL5, which is most sensitive to cabozantinib (TGI = 65.3% at

day 28). Overall, MET protein level is significantly negatively associated with cabozantinib TGI (Fig. 5C; P = 0.0013, R = -0.97). We observed a similar trend for MET mRNA levels (Fig. 5C). The higher MET level in RESL10 compared to RESL5 is supported by the bulk RNA-seq and single-nuclei RNA-seq (Fig. 5D), and it's validated by western blot (Fig. 5E).

3.4 Discussion

Systemic treatment with targeted agents and/or immune checkpoint inhibitors have been the pillars for treating RCC patients with inoperable or metastatic tumors, as RCCs are generally not very responsive to chemotherapy and radiotherapy. Immunotherapy has shown good efficacy in RCC and is also widely used to treat many other types of cancers. And the other major category of drugs for RCC are targeted therapies. RCCs are highly vascular, thus tyrosine kinase inhibitors targeting the VEGF signaling have been major approved first-line and second-line targeted agents. Although these single agents have been proven effective to some extent, resistance eventually develops, leading to the realization that multiple drugs with specific different targets are needed to overcome such resistance. Here we tested 10 drugs as well as selected combinations across 6 RCC models, finding cabozantinib and sapanisertib to be the most effective single agents. The combination of cabozantinib and sapanisertib is also the most effective combinational therapy we tested, and its effect superior to either of the single agents across all 6 lines. Compared to a similar combination that has been FDA-approved – lenvatinib plus everolimus, our preliminary results showed that the combination of cabozantinib plus sapanisertib performs similarly to lenvatinib plus everolimus. However, comparisons in more models need to be done to ascertain if there is a difference in their tumor inhibition effect as well as toxicity.

We hypothesized that RCCs with different mutational profiles may respond very differently to cabozantinib and sapanisertib. We indeed found various degrees of response to the two drugs across the 6 lines, which have differential mutation profiles (Table 2). Interestingly, tumor RESL10, which harbors a PIK3CA H1047R mutation, showed the least tumor inhibition by cabozantinib (at least at day 28 and day 47). It has been known that mutations in members of the PI3K pathway are highly frequent in various human cancer types, including clear cell renal cell carcinoma[27–29]. These mutations are most commonly found in the PIK3CA gene, which encodes the p110alpha catalytic subunit of PI3K. 5% of ccRCC harbor PIK3CA mutations or amplifications[30]. PIK3CA mutations cluster at hotspot H1047 in the kinase domain, as well as E542 and E545 within the helical domain. Previous in vitro studies have shown the PIK3CA H1047R mutation increases kinase activity and confers variable oncogenic features[28,31]. Nisa et al reported that PIK3CA H1047R confers resistance to MET inhibition (using tepotinib) in head and neck cancer cells, and combined MET/PI3K inhibition led to enhanced anti-tumor activity in tumors with PIK3CA H1047R[20]. Our results for RESL10 seem to agree with the Nisa et al study, considering cabozantinib also targets MET and sapanisertib targets mTORC1/2, downstream of PI3K signaling. Further investigation with an expanded RCC sample set and/or with isogenic cell lines is needed to explore the relation between PIK3CA mutation and cabozantinib response in RCC.

The inhibition of cell cycle-related proteins by the combination therapy is expected and maybe the molecular alterations underlying the tumor regression. The decreased mitochondrial proteins

may be a sign of mitochondrial damage, which is usually considered a drug's "off-target" effect, contributing to adverse reactions[32]. The mTORC1 is known as a regulator of mitochondrial functions[33–35]. As a mTORC1/2 inhibitor, sapanisertib alone decreased several proteins in the oxidative phosphorylation pathway (ATP6V1G1, NDUFA2, COX17, COX7C, NDUFB3) compared to the control ($P < 0.05$, difference in log2 intensity > 0.1). Cabozantinib has been reported to increase mitochondrial membrane potential (an indicator of mitochondrial activity) in medullary thyroid carcinoma[36]. However, here in our data, we observed cabozantinib decreased cytochrome C (CYCS) and OPA1 mitochondrial dynamin-like GTPase proteins, which function in the oxidative phosphorylation pathway and did not lead to a dramatic increase in proteins related to mitochondrial processes. It's interesting to see that the combination of sapanisertib and cabozantinib seems to decrease mitochondrial activity based on our proteomics analysis, suggesting the combination therapy may contribute to adverse reactions through damaging mitochondria.

We found an over-representation of EMT-related proteins increased after the cabozantinib plus sapanisertib treatment, including LRP1 and TGFBI. Low-density lipoprotein (LDL)-related protein-1 (LRP1) has been shown to bind tissue-type plasminogen activator and activate extracellular ERK1/2 to stimulate MMP9 production in kidney fibroblasts[37], which degrades the basement membrane of epithelial cells and initiates EMT[38,39]. LRP1 was also reported to induce MMP2 and MMP9 to promote glioblastoma cell migration and invasion[40]. Knockdown of LRP1 in ccRCC cells has been shown to compromise the recombinant lactotransferrin-mediated suppression of cell migration and EMT marker expression[41]. Although lactotransferrin (LTF) was not differentially expressed after the combination treatment and tissue-type plasminogen

activator (PLAT) was not detected in our dataset, we did observe a slight increase in human MMP2 protein (P = 0.05, protein change in log2intensity = 0.86). We also observed an increase in the mouse Mmp19 protein after combination treatment (P = 0.03, protein change in log2intensity = 1.02), a matrix metalloproteinase that was reported to induce EMT in vitro[42], promote metastasis in lung cancer[43], and it is associated with ccRCC patient survival[44], although there is no report on direct association between MMP19 and tissue-type plasminogen activator nor LRP1. Transforming growth factor-beta-induced (TGFBI) protein is a downstream member of the TGF-beta signaling pathway[45], and has been reported to promote EMT in gastric cancer and cholangiocarcinoma. Although TGFBI has not been directly linked to EMT in RCC, TGF-beta 1 treatment was reported to promote RCC EMT[46]. EMT has been associated with drug resistance in cancers. Although how EMT induced resistance to targeted therapies is largely unknown, studies have shown EMT-regulating transcription factors, such as ZEB1 and TWIST1, are correlated with resistance to EGFR inhibitors in *EGFR* mutant lung cancer through inhibiting pro-apoptotic protein BIM and therefore suppressing apoptosis[47,48]. Yochum and colleagues[47] showed that BCL-2/BCL-XL inhibitor (ABT-737) and a TWIST1 inhibitor, harmine, desensitize EMT *EGFR* mutant lung cancers via apoptosis sensitization. We did not find differential expression of BIM protein after the combination treatment in our dataset. But we did observe decreased pro-apoptotic proteins CASP8, IRAK1, and anti-apoptotic proteins BIRC6 after combination treatment. Further studies would be needed to evaluate the apoptosis activity for the residual tumor after combination treatment, as it may provide a new therapeutic option to combat potential resistance to the combination treatment.

IGF2BP3 is a member of the IGF2BP protein family, which is involved in RNA localization, transportation, and stability, and plays functional roles in embryonic development and cell metabolism, proliferation, migration, and invasion[49]. The IGF2BP proteins were known to be expressed in developing human tissues but absent in normal adult tissues. IGF2BP3, however, was reported to be re-expressed in a variety of tumors, including kidney cancer[50,51]. Previous studies showed tissue and circulating IGF2BP3 levels may be used as a prognosis marker as well as a predictor for metastasis in ccRCC[52–55]. IGF2BP3 has been shown to promote RCC progression by activating NK-kappaB signaling pathway[52] and by stabilizing lncRNA CDKN2B-AS1[56]. In this study, we found decreased IGF2BP3 proteins in RCC PDX tumors after cabozantinib plus sapanisertib treatment, a significant change compared to the single agent-treated tumors ($P < 0.05$). And we also found high IGF2BP3 protein level is associated with worse survival in the CPTAC ccRCC cohort. We did not find significant differential expression of NFKB1/2 protein nor CDKN2B-AS1 RNA. However, by examining the reported IGF2BP3 targets in cancer summarized by Mancarella, Scotlandi, and colleagues[57], we found combination treatment also decreased mRNA levels of CCND3 compared to the control ($P < 0.05$), cabozantinib treatment ($P < 0.05$) as well as sapanisertib treatment ($P < 0.1$), suggesting the coordinated down-regulation of IGF2BP3 and downstream targets CCND3 may contribute to the superior tumor inhibition effect of cabozantinib plus sapanisertib treatment. Pyrroline-5-carboxylate reductase 1 (PYCR1) is the last key enzyme that catalyzes the synthesis of proline, and it also functions to increase mitochondrial reactive oxygen species production, promoting EMT through inducing mitochondrial ion protease, thus also activating M2 macrophage polarization and angiogenesis[58]. Previous studies have shown that PYCR1 up-regulation is correlated with shorter overall survival (OS) in RCC[59,60]. In ccRCC cell lines, PYCR1 is

associated with resistance to multiple drugs, including RTK inhibitor lapatinib[61]. Cabozantinib treatment leads to increased PYCR1 protein compared to the control ($P < 0.05$). Interestingly, the combination of cabozantinib plus sapanisertib seems to decrease PYCR1 protein levels compared to the cabozantinib-alone and sapanisertib-alone treatments, bringing it to a similar level compared to the control. Our results indicate the cabozantinib plus sapanisertib combination may be able to block the increase in PYCR1 protein, thus preventing PYCR1-mediate drug resistance.

We found high MET protein expression (as well as gene expression) predicts worse tumor growth inhibition by cabozantinib, which targets many RTKs including MET. High MET expression in advanced RCC patients has been associated with poor prognosis and prior exposure to VEGFR tyrosine-kinase inhibitors[62,63]. Of note, RESL10 in our study shows the highest MET protein expression and was derived from a patient who has previously received VEGFR inhibitors axitinib and cabozantinib. RESL4 was also derived from a patient who is undergoing VEGFR inhibitor lenvatinib plus everolimus treatment, however, its MET protein expression is lower than other RESL3, RESL11, and RESL12, which did not receive prior VEGFR inhibitors. In preclinical models, high MET expression is associated with resistance to VEGFR tyrosine-kinase inhibitor treatment[64]. MET expression level does not affect cabozantinib treatment outcome in the phase 3 METEOR RCC clinical trial[65], nor in breast cancer or cholangiocarcinoma clinical trials for cabozantinib treatment[66,67]. However, in a recent study of the plasma biomarkers for the RCC METEOR trial, decreased MET levels, when analyzed with log2 of baseline protein levels as a continuous variable, are prognostic for improved progression-free survival (PFS) and OS[68]. In phase 3 CELESTIAL trial of cabozantinib treatment in

advanced hepatocellular carcinoma, high levels of MET were associated with shorter OS[69]. These results suggest additional studies are needed to confirm the relation between MET protein and cabozantinib response in RCC as our sample size is limited.

3.5 Acknowledgement

This project has been funded by the National Cancer Institute under award U54-CA224083 to Li Ding and Shunqiang Li. We thank the members of McDonnell Genome Institute for providing sequencing service, especially Bob Fulton, Catrina Fronick, Jennifer Ponce, Jessica Klette, Kevin Haub, and Christopher Sawyer. Kazuhito Sato. has received overseas research fellowships from the Japan Society for the Promotion of Science (JSPS) and the Uehara Memorial Foundation. Finally, we are grateful to the patients, families, and professionals who have contributed to this study.

3.6 Author Contributions

L.D. and F.C. conceived the project and led project design. Yige.W. performed data processing and data analysis and generated figures and tables. Y.W. and S.C. wrote the manuscript. X.Y., P.L., and K.S. contributed to the PDX model development, maintenance, drug treatment, sample measurement and collection, sample preparation for bulk WXS and RNA-seq sequencing, and manuscript writing. X.Y. prepared the samples for single-nucleus sequencing. S.C. prepared the samples for proteomics. Yuefan.W. and H. Z. contributed to the proteomics data generation. S.C. and A.S. performed immunofluorescence staining and imaging. H.S. contributed to the sequencing data QC and processing. S.L. and S.R.D. contributed to sample collection and sequencing data generation. S.C., P.L., and K.S. reviewed and edited the manuscript.

3.7 Methods

Tumor materials

All human tissues acquired for experiments were processed in compliance with NIH regulations and institutional guidelines, as approved by the Institutional Review Board at Washington University in St. Louis (WUSTL). All tumor materials from patients were obtained either via core needle biopsy, skin punch biopsy, or surgical resection after informed consent. Freshly collected tumors were maintained in Dulbecco's Modified Eagle Medium (DMEM)/F-12 (Gibco, 11330032) on ice until implantation. In addition to the implanted piece, another piece of each tumor tissue was fixed in 4% paraformaldehyde at 4°C and processed the next day; several small pieces (3 mm cubes) were frozen in 10% dimethylsulfoxide (Sigma-Aldrich, D2660) and 90% fetal bovine serum (Gibco, 10437028) at –80°C and stored in liquid nitrogen for future implantation; several small pieces were snap-frozen in liquid nitrogen and stored at –80°C for subsequent analyses; all leftover tissues were dissociated and used for organoid culture.

PDX model establishment

All animal procedures were reviewed by and received ethical approval from the Institutional Animal Care and Use Committee at WUSTL. This study followed the protocol approved by American Association for Laboratory Animal Science (IACUC). All immunodeficient NSG mice (NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}*/SzJ*, Stock No: 005557) were purchased from The Jackson Laboratory. Mice were kept in a temperature-controlled facility on a 12/12-hour light/dark schedule with normal food and water supplies. 6~8 week old female NSG mice were placed under isoflurane anesthesia and received tumor implantation subcutaneously on both flanks. Mice were checked daily after surgery for one week. Upon signs of solid tumor establishment, growth was monitored

weekly. Tumor growth was determined by a two-dimensional measurement with calipers. Tumor volume [mm^3] = length [mm] x $width^2$ [mm^2] x 0.5. When tumors reached 2 cm in diameter or the mouse was weak, mice were euthanized and tumors were collected immediately. Collected tumors were cut into 3 mm cubes in PBS for another implantation immediately. The remaining tumor pieces were assigned and stored in the same way as described in the "Tumor materials" section. For each tumor, volumes are expressed relative to the initial volume (day 0), as a relative tumor volume (RTV). The % tumor growth inhibition (TGI) is defined as (1 – (mean volume of treated tumors)/(mean volume of control tumors)) x 100%. The statistical significance of tumor inhibition by drug treatment was assessed by comparing RTVs between the treated vs. control groups or combinational tharapy vs. monotherapy groups in unpaired Student's t tests.

Treatment cohort and tumor sample collection

Implantation was performed similarly to PDX establishment. A batch of 7 week-old female NSG mice received bilateral tumor implants. Tumor growth and body weight were monitored once there were signs of solid tumor growth. When most tumors from one cohort reached a volume of 150~250 mm^3, the mice were assigned into 4 groups randomly with comparable median and mean tumor volumes. The day of grouping was designated as day 0 of an experiment and was also the first day of treatment.

Mice in the vehicle group were given corn oil (Sigma, C8267) 5 days/week (Monday to Friday) by oral gavage using feeding tubes (Instech, FTP-20-38). Mice in the treatment groups were treated with 30 mg/kg cabozantinib (Selleck, S4001), 1 mg/kg sapanisertib (Sigma, L21907), or 30 mg/kg cabozantinib + 1 mg/kg sapanisertib, 5 days a week by oral gavage. Tumor size and body weight

were measured twice a week. The treatment lasted about 30 days, or stopped when the tumor reached 2 cm in diameter. Tumors were harvested immediately after the host was euthanized. One small piece of tissue from each tumor was fixed in 4% paraformaldehyde at 4°C and processed the next day; all other tissues were snap-frozen in liquid nitrogen and stored at –80°C for subsequent analyses.

Genomic DNA and total RNA extraction

A small piece (~10 mg) of tumor tissue was cut from each sample and the genomic DNA was isolated using the QIAamp DNA mini kit (Qiagen, 51304) following the Quick-Start Protocol.

Total RNA was isolated with TRI Reagent (Sigma, T9424) and cleaned up with RNeasy MinElute Cleanup Kit (Qiagen, 74204). Briefly, about 15 mg of tumor tissue was chopped on a petri dish into finer pieces with a scalpel blade on ice and transferred to a 1.5 ml Eppendorf tube with 800 µl TRI Reagent. After vortexing, the mixture stayed at room temperature for 10 minutes. Then 200 µl chloroform (Sigma, 319988) was added to the tube, vortexed for 1 minute, and incubated at room temperature for 5 minutes. The mixture was centrifuged at 12,500 rpm for 10 minutes. The upper aqueous layer (500 µl) was transferred into a new tube and mixed with 500 µl of ice-cold 2-propanol (Sigma, 190764). After vortexing, the mixture was incubated at –20°C for 2 hours and then was transferred to an RNeasy MinElute spin column. After that, the sample was processed following the Quick-Start Protocol of the RNeasy MinElute Cleanup Kit. Total RNA was run on TapeStation (Agilent Technologies) and a sample that had a minimum RNA integrity number equivalent to 7 was subjected to RNA sequencing.

Whole-exome sequencing

Genomic DNA quantity was assessed by fluorometry using the Qubit dsDNA HS Assay (Q32854) according to the manufacturer's instructions (Thermo Fisher Scientific, Waltham, MA). 100–250 ng of genomic DNA was fragmented on the Covaris LE220 instrument targeting 250 bp inserts. Automated dual-indexed libraries were constructed with the KAPA Hyper library prep kit (Roche) on the SciClone NGS platform (Perkin Elmer). Up to ten libraries were pooled at an equimolar ratio by mass prior to the hybrid capture targeting a 5-µg library pool. The library pools were hybridized with the xGen Exome Research Panel v1.0 reagent (IDT Technologies) that spans a 39Mb target region (19,396 genes) of the human genome. The libraries were hybridized for 16–18 h at 65°C followed by a stringent wash to remove spuriously hybridized library fragments. Enriched library fragments were eluted and PCR cycle optimization was performed to prevent over-amplification. The enriched libraries were amplified with the KAPA HiFi master mix (Roche) prior to sequencing. The concentration of each captured library pool was accurately determined through qPCR utilizing the KAPA Library Quantification Kit according to the manufacturer's protocol (Roche) to produce cluster counts appropriate for the Illumina NovaSeq-6000 instrument. 2x150 paired-end reads were generated targeting 12Gb of sequence to achieve ~100x coverage per library.

RNA sequencing

Library preparation was performed with 500 ng to 1 µg of total RNA. Ribosomal RNA was blocked using FastSelect reagents (Qiagen) during cDNA synthesis. RNA was fragmented in reverse transcriptase buffer with FastSelect reagent and heated to 94°C for 5 min, 75°C for 2 min, 70°C for 2 min, 65°C for 2 min, 60°C for 2 min, 55°C for 2 min, 37°C for 5 min, 25°C for 5 min. mRNA was reverse transcribed to yield cDNA using SuperScript III RT enzyme (Life

Technologies, per manufacturer's instructions) and random hexamers. A second strand reaction was performed to yield ds-cDNA. cDNA was blunt ended, had an A base added to the 3′ ends, and then had Illumina sequencing adapters ligated to the ends. Ligated fragments were then amplified for 15 cycles using primers incorporating unique dual index tags. Fragments were sequenced on an Illumina NovaSeq-6000 S4 instrument generating approximately 30M paired end 2x150 reads per library.

Sample cryopulverization

Sample cryo-pulverization was performed at Washington University in St. Louis (Ma Lab). Flash-frozen tissues were transferred into pre-cooled Covaris Tissue-Tube 1 Extra (TT01xt) bags (Covaris #520007) on the dry ice and quickly processed in a Covaris CP02 Cryoprep device using an impact setting of 3. Tissue powder was then transferred to an aluminum weighing dish (VWR #1131-436) and thoroughly mixed with a pre-cooled metal spatula. Next, the tissue powder was partitioned into three aliquots, aliquot 1 = ~25mg, aliquot 2 = ~ 25mg and excess tissue in aliquot 3, into separate precooled cryovials (Corning #430487). All procedures were carried out on dry ice and liquid nitrogen to maintain tissues and all tools in the frozen state. Aliquot 1 was used for proteomics analysis in this study.

Sample processing for protein extraction and tryptic digestion

All samples for the current study were prospectively collected as described above and processed for mass spectrometric (MS) analysis at Johns Hopkins University. Tissue lysis and downstream sample preparation for global proteomic and phosphoproteomic analysis were carried out as previously described[70]. Approximately 25-120 mg of each cryo-pulverized renal tumor tissues or NATs were homogenized separately in an appropriate volume of lysis buffer (8 M urea, 75 mM

NaCl, 50 mM Tris, pH 8.0, 1 mM EDTA, 2 g/mL aprotinin, 10 g/mL leupeptin, 1 mM PMSF, 10 mM NaF, Phosphatase Inhibitor Cocktail 2 and Phosphatase Inhibitor Cocktail 3 [1:100 dilution], and 20 mM PUGNAc) by repeated vortexing. Lysates were clarified by centrifugation at 20,000 x g for 10 min at 4°C, and protein concentrations were determined by BCA assay (Pierce). Lysates were diluted to a final concentration of 8 mg/ml with lysis buffer, and 800 g of protein was reduced with 5 mM dithiothreitol (DTT) for 1 h at 37°C and subsequently alkylated with 10 mM iodoacetamide for 45 min at RT (room temperature) in the dark. Samples were diluted 1:3 with 50 mM Tris-HCl (pH 8.0) and subjected to proteolytic digestion with LysC (Wako Chemicals) at 1 mAU:50 g enzyme-to-substrate ratio for 2h at RT, followed by the addition of sequencing-grade modified trypsin (Promega) at a 1:50 enzyme-to-substrate ratio and overnight incubation at RT. The digested samples were then acidified with 50% trifluoroacetic acid (TFA, Sigma) to a pH value of approximately 2.0. Tryptic peptides were desalted on reversed-phase C18 SPE columns (Waters), followed by aliquoting 20 g of digested peptides for global proteomic analysis, dried in a Speed-Vac, and resuspended in 3% ACN/0.1% formic acid prior to ESI-LC-MS/MS analysis.

ESI-LC-MS/MS for global proteome and phosphoproteome using DIA-MS analysis

Individual global proteome and phosphoproteome samples were analyzed using the same instrumentation and methodology; albeit with varied gradient settings. Unlabeled, digested peptide material from individual tissue samples was spiked with index Retention Time (iRT) peptides (Biognosys) and subjected to data-independent acquisition (DIA) analysis. Peptides (~0.8 g; ~1 ug for glycopeptides) were separated on an Easy nLC 1200 UHPLC system (Thermo Scientific) on an in-house packed 20 cm x 75 m diameter C18 column (1.9 m Reprosil-Pur C18-AQ beads (Dr. Maisch GmbH); Picofrit 10 m opening (New Objective)). The column was heated

to 50°C using a column heater (Phoenix-ST). The flow rate was 0.200 µl/min with 0.1% formic acid and 3% acetonitrile in water (A) and 0.1% formic acid, 90% acetonitrile (B).

For global proteomic characterization, the peptides were separated using the following LC gradient: 0-3 min (2% B, isocratic), 3-103 min (7%-20% B, linear), 103-121 min (20-30% B, linear), 121-125 min (30-60% B, linear), 125-126 min (60-90% B, linear), 126-130 min (90% B, isocratic), 130-131 min (90-50% B, linear), 131-140 min (50% B, isocratic). Samples were analyzed using the Thermo Fusion Lumos mass spectrometer (Thermo Scientific).

For global and phosphoproteome, the DIA segment consisted of one MS1 scan (350-1650 m/z range, 120K resolution) followed by 30 MS2 scans (variable m/z range, 30K resolution) as described previously[71]. Additional parameters were as follows: MS1: RF Lens – 30%, AGC Target 4.0e5, Max IT – 60 ms, charge state include - 2-6; MS2: isolation width (m/z) – 0.7, AGC Target – 3.0e6, Max IT – 80 ms.

Single nuclei isolation

Each tumor sample (25~30 mg) stored at -80°C was cut into 2 mm pieces on a petri dish with a scalpel blade on ice and loaded into a 2 ml glass homogenizer (Fisher, K8853030002) with 1 ml of ice-cold lysis buffer, which was made freshly with 10 mM Tris-HCl (pH 7.4, Thermo, 15567027), 10mM NaCl (Thermo; AM9759), 3 mM $MgCl_2$ (Thermo, AM9530G), 0.01% NP-40 substitute (Sigma, 74385), nuclease-free water (Sigma, W4502), and 0.2 U/ul RNase inhibitor. Each sample was homogenized with the pestle (Fisher, K8853020002) by 4 pushes and 4 pulls. The homogenate was then incubated on ice for 1 minute with an additional 1 ml of cold lysis buffer, then pipetted gently for 4 times, and incubated on ice for another 1.5 minutes. After incubation, the homogenate went through a 40 um cell strainer, and the cell strainer was washed with 1 ml wash buffer, which was freshly made with 2% BSA in PBS and 0.2 U/ul RNase inhibitor. Both

filtrates were collected in the same tube and centrifuged at 500 *g* for 6 minutes at 4°C. The pellet was resuspended in a 300 ml ice-cold wash buffer gently and 1 ml DRAQ5 was added for staining. After 10 minutes of incubation on ice, the sample was sorted by FACS and ~200,000 nuclei were collected in the ice-cold wash buffer. Then the sorted nuclei in the wash buffer were centrifuged at 500 *g* for 6 minutes at 4°C and the pellet was resuspended in ~35 ml ice-cold wash buffer gently. We then extracted a 1 ul nuclei sample for counting on a hemocytometer with trypan blue staining. For single nuclei sample submission, the nuclei were diluted into 1,500~2,000 nuclei/ml in the ice-cold wash buffer.

H&E staining

Fresh tumor tissues were fixed in 10% neutral buffered formalin (Epredia, 5725) at room temperature overnight but for less than 24 hours. Tissues were then dehydrated, infiltrated with wax, and embedded into paraffin blocks. Formalin-fixed paraffin-embedded (FFPE) samples were cut into 4-µm tissue sections and heated to 55°C prior to staining. Next, deparaffinization was performed by incubating in xylene, followed by rehydration incubation in 100%, 95%, 70%, 50%, and 25% ethanol. Then, slides were stained with hematoxylin (Vector Labs, H-3401-500) for 6 minutes followed by triple-washing with tap water and bluing in Scott's Tap Water (Sigma, S5134) gently for 30 seconds. After that, tissues were counterstained with eosin (Biosciences, 786-1063) for 30 seconds followed by washing with tap water. Finally, the stained tissues were dried overnight at room temperature. The next day, sections were dehydrated with 100% ethanol and xylene. Finally, slides were sealed with nail polish, and images were taken using the Leica DMi8 microscope.

Immunofluorescence Staining of FFPE Slides

Fresh samples were processed into FFPE blocks as described in the "H&E staining" section. FFPE slides were first heated to 55°C for 10 minutes prior to staining. After this, section dehydration was performed via incubation subsequently in xylene, 100%, 95%, 70%, 50%, and 25% ethanol. After another 2-minute wash in ddH2O, antigen retrieval was performed in a hot-water bath using a Tris EDTA buffer pH 9 (Genemed, 10-0046) at 80-90°C for 22 minutes. The slides were then cooled to room temperature. After that, glycine blocking was then performed using two 9-minute washes in 100mM glycine, followed by two 2-minute washes in PBST. Sections were circled with a PAP pen, then blocked for one hour in a blocking solution consisting of 10% normal donkey serum in 1% BSA/PBS. The sections were dried with a Kimwipe, but the blocking solution was not washed. Next, primary antibodies, diluted to the appropriate concentration in the same blocking buffer as above, were applied to the sections and allowed to incubate at 4°C overnight. The next day, the slides were washed twice for 5 minutes in 1x PBST. Secondary antibodies, also diluted to the appropriate concentration in the same blocking buffer as before, were applied to the sections and allowed to incubate on the bench for one hour. Following two more 5-minute washes in PBST, sections were treated with Hoechst stain at a concentration of 1:2000 for 8 minutes. Sections were then washed in 1x PBS twice for 3-minutes, and then mounted in an aqueous mounting medium and covered with a coverslip. Pressure was applied to the sections to ensure total coverage by the mounting medium. Finally, slides were sealed with nail polish and stored at 4°C until ready for use.

Immunoblotting

Flash-frozen tissue was washed with 1x PBS and was submerged in RIPA buffer (#9806, CST) and lysed using PRO200 Biogen homogenizer, and centrifuged at 13300 rpm for 15 min.

Supernatants were quantified using Bio-rad DC protein assay and equal amounts of proteins were loaded and separated using 10% polyacrylamide gel. Proteins were transferred onto the PVDF membrane (Immobilon-FL Merck Millipore) and later blocked using Odyssey blocking buffer. Primary antibodies were incubated at appropriate concentrations O/N and the next day were incubated with Licor IR 680 (Donkey anti-Mouse) and IR700 (Donkey anti-Rabbit) fluorescent antibodies from Licor. The blot was developed using the Bio-rad Chemidoc MP imaging system.

Bulk WES and RNA-seq data processing

The bulk WES and RNA-seq data were processed as previously described by Sun et al. PDX study[72].

Proteomics data analysis

Proteome and phosphoproteome DIA data quantification

All the DIA files of the global proteome were analyzed via a library-free directDIA approach embedded in Spectronaut (version 14.10, Biognosys) with precursor and protein Qvalue cutoff at 1%. Forty raw files acquired from 1 μg injections were analyzed together in one directDIA search. The phospho analyses were conducted by searching global DIA data with phospho modification via directDIA approach.

Proteomics data normalization

For all the DIA data, we use Spectronaut™ to analyze and use the software function of "Cross Run Normalization" to normalize by the median at the peptide level as previously described by Callister et al[73]. The resulting quantified protein and phosphorylation spectral intensity values

were log2 transformed. The protein abundance level was further quantile normalized using the preprocessCore R package[74,75].

Protein changes associated with drug treatment

The comparison of the protein abundance level between different treatment groups was performed using paired Student's t-test. The false discovery rates were derived based on the Benjamini & Hochberg correction.

Gene set over-representation test

For over-representation tests, we used the hallmark gene set and the canonical gene set from MSigDB[76,77], and the enricher function from the clusterProfiler R package[78].

Independent ccRCC cohort proteomics data analysis

Global proteomics data for the independent ccRCC samples were downloaded from the Clark et al. ccRCC study[70].

snRNA-seq quantification and analysis

First, the combined human genome (GRCh38) and mouse genome (mm10) references were made using Cell Ranger (v5.0.1) mkref functionality. The raw snRNA-seq FASTA files were first aligned to the human-mouse combined reference using the Cell Ranger count function (specifying –include-introns) to generate the barcode classification (from gem_classification.csv), which classifies each barcode to either mouse cell or human cell or multiplet. Then the FASTA files were aligned to the human-only reference using Cell Ranger as described above. The resulting gene-by-cell UMI count matrix was used to construct a Seurat object for each sample using the R package Seurat[79] (v.3.1.0). Mouse cells and multiplets were

filtered out based on the gem_classification.csv. Doublets were filtered out using Scrublet (https://github.com/AllonKleinLab/scrublet). Scrublet was run on each sample separately with the following parameter settings: expected_doublet_rate=0.06, min_counts=2, min_cells=3, min_gene_variability_pctl=85, n_prin_comps=30. The doublet score threshold was adjusted manually, which can separate the two peaks of a bimodal simulated doublet score histogram. Quality filters were applied to the data to remove barcodes that fell into any of the following categories: possible debris with too few genes expressed (<200) or too few UMIs (<1,000), possibly more than one cell with too many genes expressed (>10,000) or too associated UMIs (>80,000), and possible dead cell with too high mitochondrial gene expression over the total transcript counts (> 10%). Each cell was scored based on the cell cycle S phase and G2M phase gene markers from https://raw.githubusercontent.com/hbc/tinyatlas/master/cell_cycle/Homo_sapiens.csv using the CellCycleScoring function from the Seurat package. Each sample was scaled and normalized using Seurat's 'SCTransform' function to correct for batch effects (with parameters: vars.to.regress = c(""mitoRatio", 'nFeature_RNA', "nCount_RNA", 'S.Score', 'G2M.Score'), variable.features n = 2000). Any merged analysis or subsequent subsetting of cells/samples underwent the same scaling and normalization method. Cells were clustered using the original Louvain algorithm[80] and top 30 PCA dimensions via 'FindNeighbors' and 'FindClusters' (with parameters: resolution = 0.5) functions. Any integrated analysis underwent the same scaling and normalization method. The resulting objects were then merged and integrated (using SelectIntegrationFeatures, PrepSCTIntegration, FindIntegrationAnchors, and IntegrateData functions) for the subsequent analysis.

3.8 Main figures

Figure 1

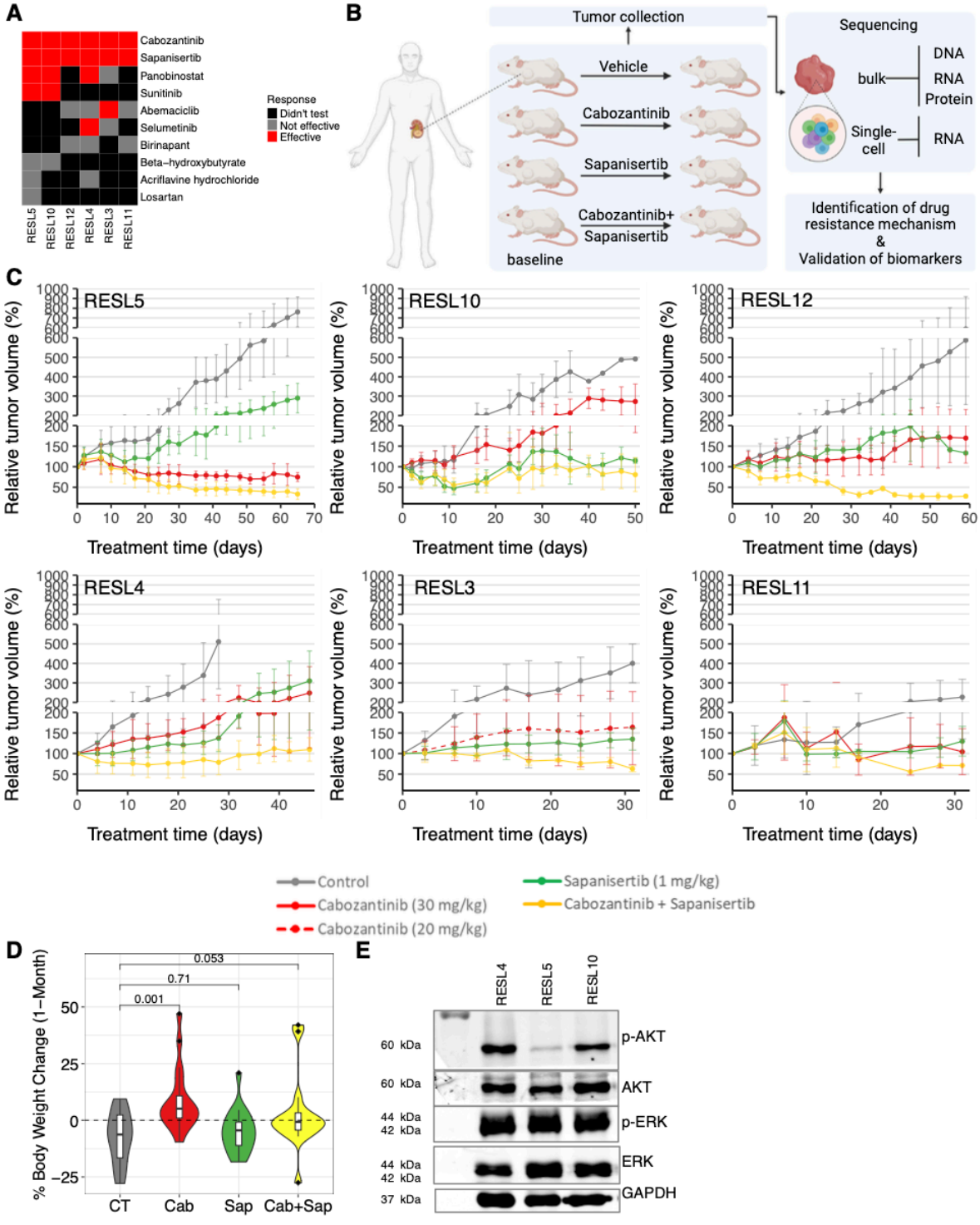


Figure 1. Study design and PDX relative tumor volume curves.
A, Heatmap summary of the treatment response for 10 single-agent drugs tested across 6 PDX lines. **B**, Schematic of the study design. **C**, Relative tumor volume curves for mice treated with

vehicle, cabozantinib, sapanisertib, and the combination (cabozantinib plus sapanisertib) across 6 PDX lines. **D**, Bodyweight changes (compared to day 0) for mice treated with vehicle, cabozantinib, sapanisertib, and the combination. P values were derived from the Wilcoxon test. **E**, Western blot of phospho-AKT (Ser473), total AKT, phospho-ERK (Thr202/Tyr204), total ERK, and total GAPDH using untreated PDX tumor samples of RESL4, RESL5, and RESL10.

Figure 2

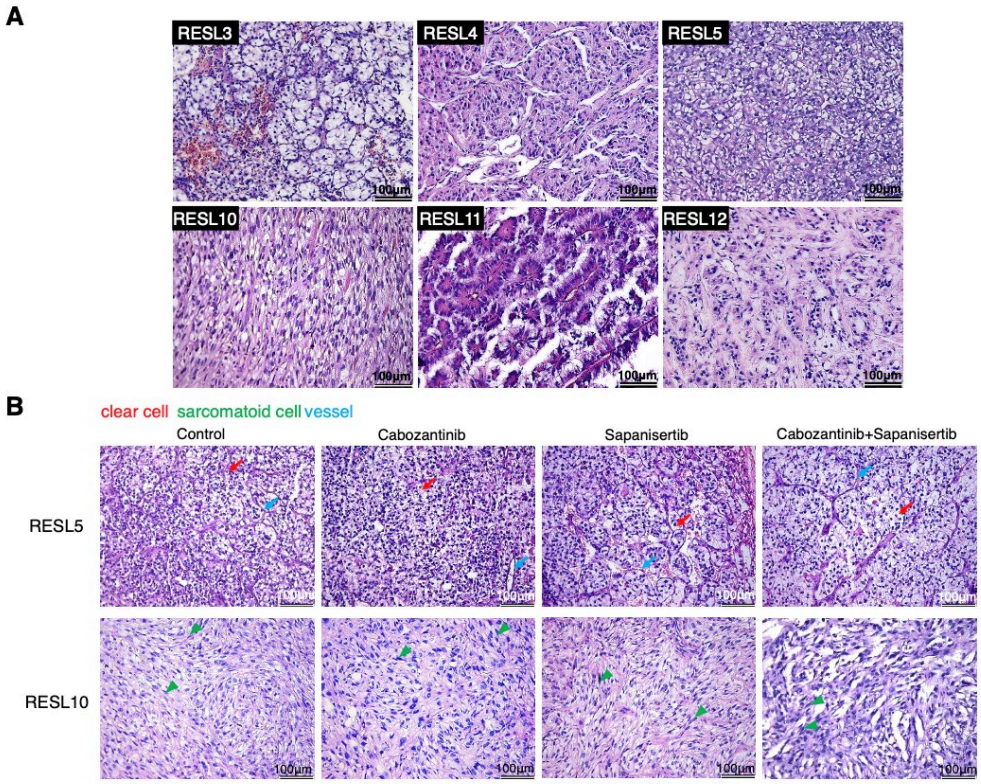


Figure 2. H&E images for the PDX models.
A, H&E images for the 6 PDX models. **B**, H&E images for tumors in mice treated with vehicle, cabozantinib, sapanisertib, and the combination (cabozantinib plus sapanisertib) in PDX line RESL5 and RESL10. Red arrows denote the clear cells. Green arrows denote sarcomatoid cells. Blue arrows denote vessels.

Figure 3

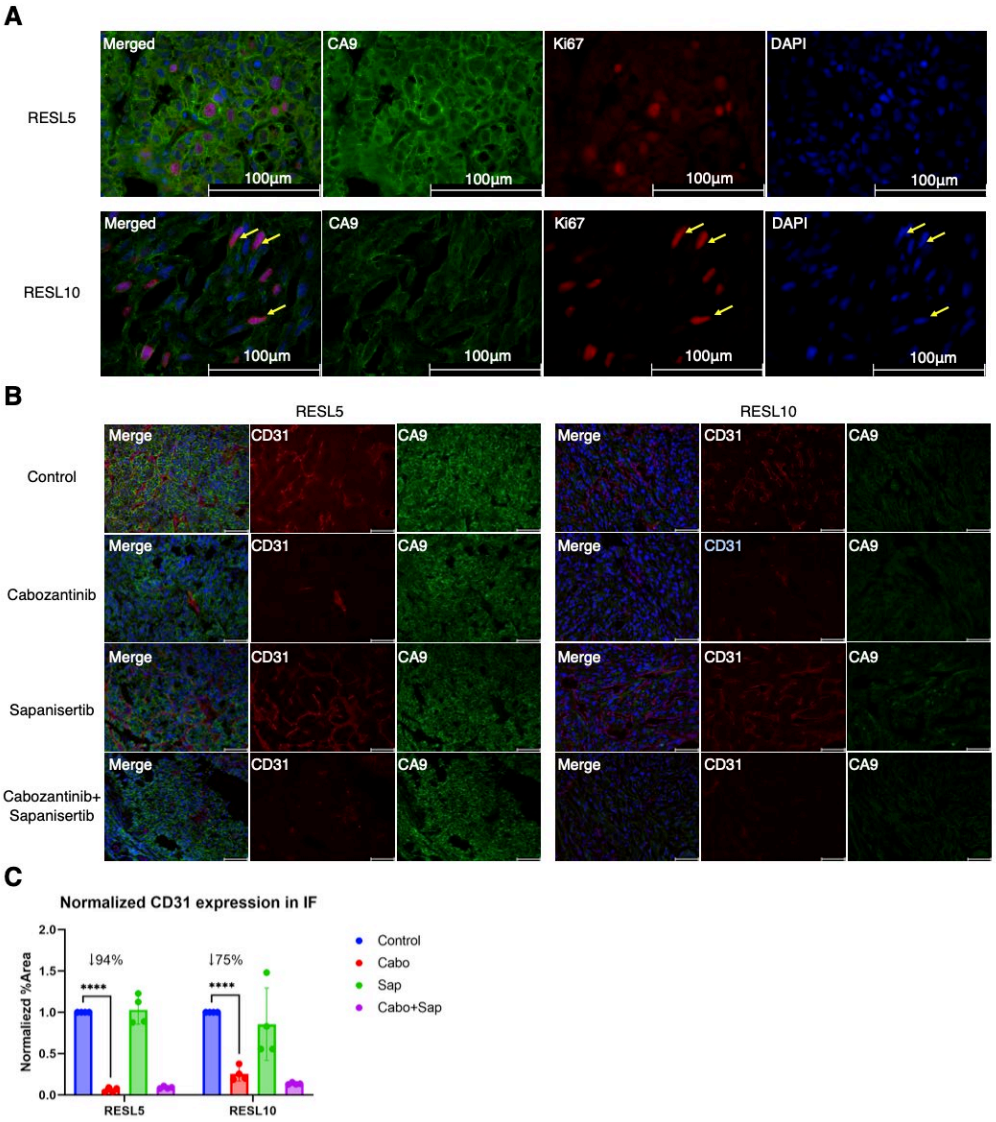


Figure 3. Immunofluorescence staining for tumor cells and vasculature in PDX lines RESL5 and RESL10.

A. IF images of untreated tumor samples stained with anti-CA9 antibody (green), anti-Ki67 antibody (red), and DAPI (blue) for PDX lines RESL5 and RESL10. Scale bars represent 100 µm. **B**, IF images of tumor samples stained with anti-CD31 antibody (red) and anti-CA9 antibody (green) in mice treated with vehicle, cabozantinib, sapanisertib, and the combination (cabozantinib plus sapanisertib) in PDX line RESL5 and RESL10. **C**, Bar plot showing the normalized CD31 expression from IF images for mice treated with vehicle, cabozantinib, sapanisertib, and the combination in PDX line RESL5 and RESL10. **** indicates $P < 0.0001$ using Student's T-test.

Figure 4

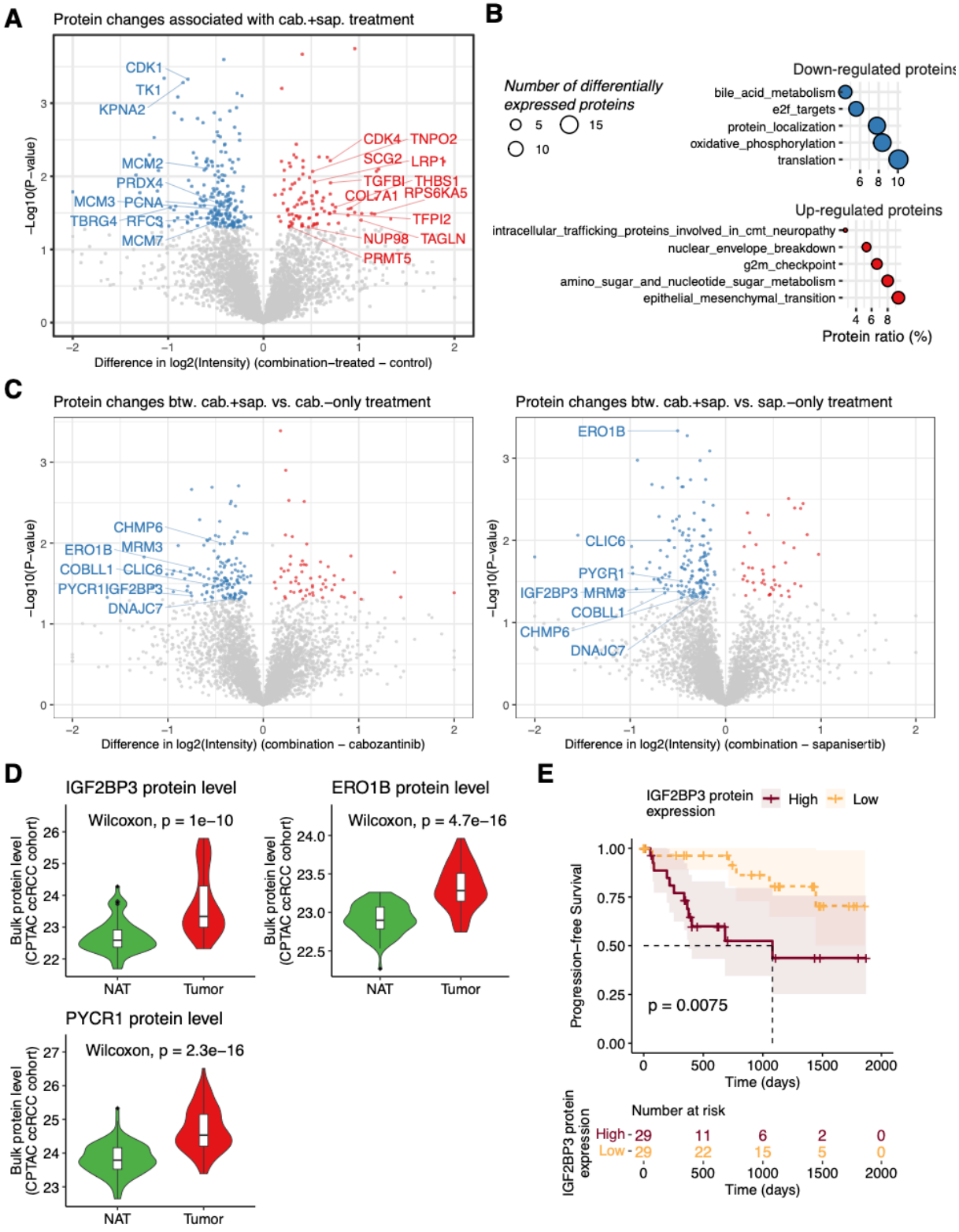


Figure 4. Proteomic analysis of the treatment effect.

A, Volcano plot showing differentially expressed proteins between tumors after combination treatment of cabozantinib plus sapanisertib compared to the vehicle-treated controls, mapping 97 up-regulated proteins (red dots) and 213 down-regulated proteins (blue dots). **B**, Dot plot showing the overrepresented gene sets in proteins down-regulated (upper panel) and up-regulated (bottom panel) after combination treatment. **C**, Volcano plots showing differentially expressed proteins between tumors after combination treatment vs. single-agent treatments. The left panel represents the comparison between combination-treated and cabozantinib-treated tumors, mapping 59 up-regulated and 146 down-regulated proteins. The right panel represents the comparison between combination-treated and sapanisertib-treated tumors, mapping 42 up-regulated and 154 down-regulated proteins. **D**, Violin plots showing protein levels of IGF2BP3, ERO1B, and PYCR1 in the tumor and normal adjacent tissue (NAT) samples in the CPTAC ccRCC discovery cohort. **E**, Kaplan-Meier curves displaying progression-free survival probability for 2 groups of patients in the CPTAC ccRCC discovery cohort. The two groups of patients are selected based on the IGF2BP3 protein levels in their primary tumors. High IGF2BP3 protein expression represents those with protein expression in the upper 35% quantile. Low IGF2BP3 protein expression represents those with protein expression in the bottom 35% quantile.

Figure 5

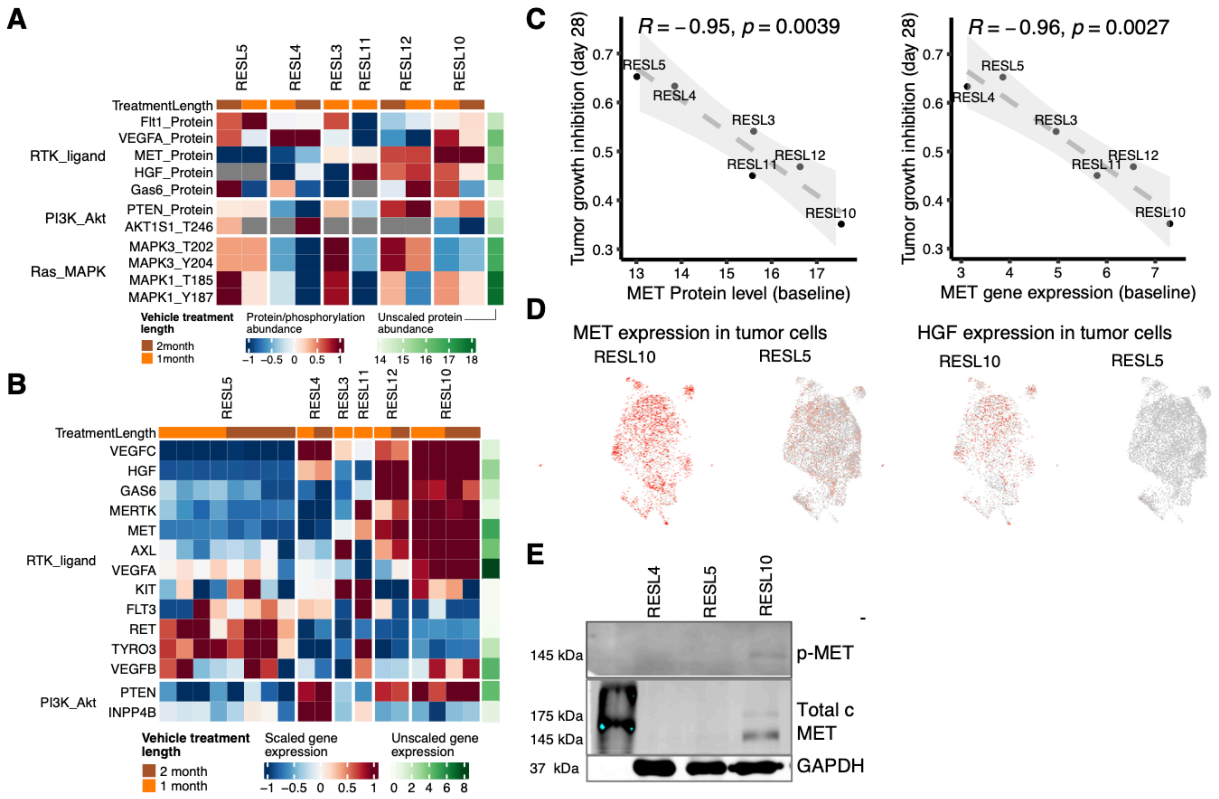


Figure 5. Protein markers associated with treatment effect.
A, Heatmap showing the scaled protein/phosphorylation abundance of the key members in the PI3K-mTOR pathway and RTKs targeted by cabozantinib (that were detected in the proteomics and phosphoproteomics datasets) in the control tumor samples across 6 PDX lines. **B**, Heatmap showing the scaled gene expression of the key members in the PI3K-mTOR pathway and RTKs targeted by cabozantinib in the control tumor samples across 6 PDX lines. The unscaled gene expression represents log2(TPM+1). **C**, Scatter plot showing the association between baseline MET protein level (left panel) or MET gene expression (right panel) with the tumor growth inhibition at day 28. **D**, MET (left panel), and HGF (right panel) gene expression in the human tumor cells in the snRNA-seq data of the RESL10 and RESL5 control tumor samples. **E**, Western blot of the phospho-MET (Tyr1234/1235), total c-MET, and total GAPDH in untreated RESL4, RESL5, and RESL10 tumor samples.

3.9 Reference

1. Hsieh, J. J. *et al.* Renal cell carcinoma. *Nat. Rev. Dis. Primer* **3**, 17009 (2017).
2. Rini, B. I., Campbell, S. C. & Escudier, B. Renal cell carcinoma. *The Lancet* **373**, 1119–1132 (2009).
3. Singer, E. A., Gupta, G. N. & Srinivasan, R. Update on targeted therapies for clear cell renal cell carcinoma. *Curr. Opin. Oncol.* **23**, 283–289 (2011).
4. Choueiri, T. K. *et al.* Cabozantinib versus Everolimus in Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* **373**, 1814–1823 (2015).
5. Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55–61 (2012).
6. García-García, C. *et al.* Dual mTORC1/2 and HER2 blockade results in antitumor activity in preclinical models of breast cancer resistant to anti-HER2 therapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **18**, 2603–2612 (2012).
7. Paik, P. K., Ahn, L. S. H., Plodkowski, A. J., Fan, P.-D. & Rudin, C. M. Phase II study of TAK228 in patients with advanced non-small cell lung cancer (NSCLC) harboring NFE2L2 and KEAP1 mutations. *J. Clin. Oncol.* **38**, 9607–9607 (2020).
8. Kim, J. W. *et al.* Sapanisertib, a dual mTORC1/2 inhibitor, for TSC1- or TSC2-mutated metastatic urothelial carcinoma (mUC). *J. Clin. Oncol.* **39**, 431–431 (2021).
9. Hua, H. *et al.* Targeting mTOR for cancer therapy. *J. Hematol. Oncol.J Hematol Oncol* **12**, 71 (2019).
10. Choueiri, T. K. & Motzer, R. J. Systemic Therapy for Metastatic Renal-Cell Carcinoma. *N. Engl. J. Med.* **376**, 354–366 (2017).
11. Posadas, E. M., Limvorasak, S. & Figlin, R. A. Targeted therapies for renal cell carcinoma. *Nat. Rev. Nephrol.* **13**, 496–511 (2017).
12. Juengel, E. *et al.* HDAC-inhibition counteracts everolimus resistance in renal cell carcinoma in vitro by diminishing cdk2 and cyclin A. *Mol. Cancer* **13**, 152 (2014).
13. Hongo, F. *et al.* CDK1 and CDK2 activity is a strong predictor of renal cell carcinoma recurrence. *Urol. Oncol.* **32**, 1240–1246 (2014).
14. Hedberg, Y., Davoodi, E., Roos, G., Ljungberg, B. & Landberg, G. Cyclin-D1 expression in human renal-cell carcinoma. *Int. J. Cancer* **84**, 268–272 (1999).
15. Chen, D., Sun, X., Zhang, X. & Cao, J. Inhibition of the CDK4/6-Cyclin D-Rb Pathway by Ribociclib Augments Chemotherapy and Immunotherapy in Renal Cell Carcinoma. *BioMed Res. Int.* **2020**, 9525207 (2020).
16. Diaz-Montero, C. M. *et al.* MEK inhibition abrogates sunitinib resistance in a renal cell carcinoma patient-derived xenograft model. *Br. J. Cancer* **115**, 920–928 (2016).
17. Yang, W. Z., Zhou, H. & Yan, Y. XIAP underlies apoptosis resistance of renal cell carcinoma cells. *Mol. Med. Rep.* **17**, 125–130 (2018).
18. Dolley-Hitze, T. *et al.* Angiotensin-2 receptors (AT1-R and AT2-R), new prognostic factors for renal clear-cell carcinoma? *Br. J. Cancer* **103**, 1698–1705 (2010).
19. Derosa, L., Izzedine, H., Albiges, L. & Escudier, B. Hypertension and Angiotensin System Inhibitors in Patients with Metastatic Renal Cell Carcinoma. *Oncol. Rev.* **10**, 298 (2016).
20. Nisa, L. *et al.* PIK3CA hotspot mutations differentially impact responses to MET targeting in MET-driven and non-driven preclinical cancer models. *Mol. Cancer* **16**, 93 (2017).
21. Janku, F. *et al.* Assessing PIK3CA and PTEN in early-phase trials with PI3K/AKT/mTOR inhibitors. *Cell Rep.* **6**, 377–387 (2014).

22. Yakes, F. M. *et al.* Cabozantinib (XL184), a novel MET and VEGFR2 inhibitor, simultaneously suppresses metastasis, angiogenesis, and tumor growth. *Mol. Cancer Ther.* **10**, 2298–2308 (2011).
23. Nilsson, H. *et al.* Primary clear cell renal carcinoma cells display minimal mitochondrial respiratory capacity resulting in pronounced sensitivity to glycolytic inhibition by 3-Bromopyruvate. *Cell Death Dis.* **6**, e1585–e1585 (2015).
24. Du, B. & Shim, J. S. Targeting Epithelial-Mesenchymal Transition (EMT) to Overcome Drug Resistance in Cancer. *Mol. Basel Switz.* **21**, E965 (2016).
25. Shibue, T. & Weinberg, R. A. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat. Rev. Clin. Oncol.* **14**, 611–629 (2017).
26. Abdelaziz, A. & Vaishampayan, U. Cabozantinib for the treatment of kidney cancer. *Expert Rev. Anticancer Ther.* **17**, 577–584 (2017).
27. Fruman, D. A. & Rommel, C. PI3K and cancer: lessons, challenges and opportunities. *Nat. Rev. Drug Discov.* **13**, 140–156 (2014).
28. Samuels, Y. *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554 (2004).
29. Thorpe, L. M., Yuzugullu, H. & Zhao, J. J. PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting. *Nat. Rev. Cancer* **15**, 7–24 (2015).
30. Guo, H. *et al.* The PI3K/AKT Pathway and Renal Cell Carcinoma. *J. Genet. Genomics Yi Chuan Xue Bao* **42**, 343–353 (2015).
31. Samuels, Y. *et al.* Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell* **7**, 561–573 (2005).
32. Young, M. J. Off-Target Effects of Drugs that Disrupt Human Mitochondrial DNA Maintenance. *Front. Mol. Biosci.* **4**, 74 (2017).
33. Toschi, A., Lee, E., Gadir, N., Ohh, M. & Foster, D. A. Differential Dependence of Hypoxia-inducible Factors 1α and 2α on mTORC1 and mTORC2 *. *J. Biol. Chem.* **283**, 34495–34499 (2008).
34. Hardwick, J. S., Kuruvilla, F. G., Tong, J. K., Shamji, A. F. & Schreiber, S. L. Rapamycin-modulated transcription defines the subset of nutrient-sensitive signaling pathways directly controlled by the Tor proteins. *Proc. Natl. Acad. Sci.* **96**, 14866–14870 (1999).
35. Arsham, A. M., Howell, J. J. & Simon, M. C. A novel hypoxia-inducible factor-independent hypoxic response regulating mammalian target of rapamycin and its targets. *J. Biol. Chem.* **278**, 29655–29660 (2003).
36. Starenki, D., Hong, S.-K., Wu, P.-K. & Park, J.-I. Vandetanib and cabozantinib potentiate mitochondria-targeted agents to suppress medullary thyroid carcinoma cells. *Cancer Biol. Ther.* **18**, 473–483 (2017).
37. Hu, K. *et al.* Tissue-type plasminogen activator acts as a cytokine that triggers intracellular signal transduction and induces matrix metalloproteinase-9 gene expression. *J. Biol. Chem.* **281**, 2120–2127 (2006).
38. Wang, X. *et al.* Mice lacking the matrix metalloproteinase-9 gene reduce renal interstitial fibrosis in obstructive nephropathy. *Am. J. Physiol. Renal Physiol.* **299**, F973-982 (2010).
39. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* **119**, 1420–1428 (2009).
40. Song, H., Li, Y., Lee, J., Schwartz, A. L. & Bu, G. Low-Density Lipoprotein Receptor-related Protein 1 Promotes Cancer Cell Migration and Invasion by Inducing the Expression of Matrix Metalloproteinase 2 and 9. *Cancer Res.* **69**, 879–886 (2009).

41. Chiu, I.-J. *et al.* Lactotransferrin Downregulation Drives the Metastatic Progression in Clear Cell Renal Cell Carcinoma. *Cancers* **12**, 847 (2020).
42. Yu, G. *et al.* Role Of MMP19 In Epithelial-Mesenchymal Transition And Fibrosis. in *A28. REGULATION OF EPITHELIAL-MESENCHYMAL TRANSITION* A1219–A1219 (American Thoracic Society, 2010).
43. Yu, G. *et al.* Matrix Metalloproteinase-19 Promotes Metastatic Behavior In Vitro and Is Associated with Increased Mortality in Non–Small Cell Lung Cancer. *Am. J. Respir. Crit. Care Med.* **190**, 780–790 (2014).
44. Gobin, E. *et al.* A pan-cancer perspective of matrix metalloproteases (MMP) gene expression profile and their diagnostic/prognostic potential. *BMC Cancer* **19**, 581 (2019).
45. Skonier, J. *et al.* cDNA cloning and sequence analysis of beta ig-h3, a novel gene induced in a human adenocarcinoma cell line after treatment with transforming growth factor-beta. *DNA Cell Biol.* **11**, 511–522 (1992).
46. Tretbar, S. *et al.* TGF-β inducible epithelial-to-mesenchymal transition in renal cell carcinoma. *Oncotarget* **10**, 1507–1524 (2019).
47. Yochum, Z. A. *et al.* Targeting the EMT transcription factor TWIST1 overcomes resistance to EGFR inhibitors in EGFR-mutant non-small-cell lung cancer. *Oncogene* **38**, 656–670 (2019).
48. Song, K.-A. *et al.* Epithelial-to-Mesenchymal Transition Antagonizes Response to Targeted Therapies in Lung Cancer by Suppressing BIM. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **24**, 197–208 (2018).
49. Bell, J. L. *et al.* Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? *Cell. Mol. Life Sci. CMLS* **70**, 2657–2675 (2013).
50. Lederer, M., Bley, N., Schleifer, C. & Hüttelmaier, S. The role of the oncofetal IGF2 mRNA-binding protein 3 (IGF2BP3) in cancer. *Semin. Cancer Biol.* **29**, 3–12 (2014).
51. Xu, W. *et al.* Increased IGF2BP3 expression promotes the aggressive phenotypes of colorectal cancer cells in vitro and vivo. *J. Cell. Physiol.* **234**, 18466–18479 (2019).
52. Pei, X. *et al.* Enhanced IMP3 Expression Activates NF-κB Pathway and Promotes Renal Cell Carcinoma Progression. *PloS One* **10**, e0124338 (2015).
53. Tschirdewahn, S. *et al.* Circulating and tissue IMP3 levels are correlated with poor survival in renal cell carcinoma. *Int. J. Cancer* **145**, 531–539 (2019).
54. Hoffmann, N. E. *et al.* External validation of IMP3 expression as an independent prognostic marker for metastatic progression and death for patients with clear cell renal cell carcinoma. *Cancer* **112**, 1471–1479 (2008).
55. Jiang, Z. *et al.* Analysis of RNA-binding protein IMP3 to predict metastasis and prognosis of renal-cell carcinoma: a retrospective study. *Lancet Oncol.* **7**, 556–564 (2006).
56. Xie, X. *et al.* LncRNA CDKN2B-AS1 stabilized by IGF2BP3 drives the malignancy of renal clear cell carcinoma through epigenetically activating NUF2 transcription. *Cell Death Dis.* **12**, 1–16 (2021).
57. Mancarella, C. & Scotlandi, K. IGF2BP3 From Physiology to Cancer: Novel Discoveries, Unsolved Issues, and Future Perspectives. *Front. Cell Dev. Biol.* **7**, (2020).
58. Kuo, C.-L. *et al.* Mitochondrial oxidative stress by Lon-PYCR1 maintains an immunosuppressive tumor microenvironment that promotes cancer progression and metastasis. *Cancer Lett.* **474**, 138–150 (2020).
59. Li, H. & Mo, Z. Prognostic Value of Metabolism-Related Genes and Immune Infiltration in Clear Cell Renal Cell Carcinoma . *Int. J. Gen. Med.* **14**, 6885–6898 (2021).

60. Hu, F., Zeng, W. & Liu, X. A Gene Signature of Survival Prediction for Kidney Renal Cell Carcinoma by Multi-Omic Data Analysis. *Int. J. Mol. Sci.* **20**, E5720 (2019).
61. Wu, Y. *et al.* A Mitochondrial Dysfunction and Oxidative Stress Pathway-Based Prognostic Signature for Clear Cell Renal Cell Carcinoma. *Oxid. Med. Cell. Longev.* **2021**, e9939331 (2021).
62. Gibney, G. T. *et al.* c-Met is a prognostic marker and potential therapeutic target in clear cell renal cell carcinoma. *Ann. Oncol.* **24**, 343–349 (2013).
63. Zhou, L. *et al.* Targeting MET and AXL overcomes resistance to sunitinib therapy in renal cell carcinoma. *Oncogene* **35**, 2687–2697 (2016).
64. Ciamporcero, E. *et al.* Combination strategy targeting VEGF and HGF/c-met in human renal cell carcinoma models. *Mol. Cancer Ther.* **14**, 101–110 (2015).
65. Choueiri, T. K. *et al.* Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomised, open-label, phase 3 trial. *Lancet Oncol.* **17**, 917–927 (2016).
66. Tolaney, S. M. *et al.* Phase II and Biomarker Study of Cabozantinib in Metastatic Triple-Negative Breast Cancer Patients. *The Oncologist* **22**, 25–32 (2017).
67. Goyal, L. *et al.* A Phase II and Biomarker Study of Cabozantinib in Patients with Advanced Cholangiocarcinoma. *Cancer* **123**, 1979–1988 (2017).
68. Powles, T. *et al.* Outcomes based on plasma biomarkers in METEOR, a randomized phase 3 trial of cabozantinib vs everolimus in advanced renal cell carcinoma. *BMC Cancer* **21**, 904 (2021).
69. Rimassa, L. *et al.* Outcomes Based on Plasma Biomarkers for the Phase 3 CELESTIAL Trial of Cabozantinib versus Placebo in Advanced Hepatocellular Carcinoma. *Liver Cancer* **11**, 38–47 (2022).
70. Clark, D. J. *et al.* Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* **179**, 964-983.e31 (2019).
71. Cho, K.-C. *et al.* Deep Proteomics Using Two Dimensional Data Independent Acquisition Mass Spectrometry. *Anal. Chem.* **92**, 4217–4225 (2020).
72. Sun, H. *et al.* Comprehensive characterization of 536 patient-derived xenograft models prioritizes candidates for targeted treatment. *Nat. Commun.* **12**, 5086 (2021).
73. Callister, S. J. *et al.* Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* **5**, 277–286 (2006).
74. bmbolstad.com>, B. B. <bmb at. *preprocessCore: A collection of pre-processing functions*. (Bioconductor version: Release (3.15), 2022). doi:10.18129/B9.bioc.preprocessCore.
75. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinforma. Oxf. Engl.* **19**, 185–193 (2003).
76. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
77. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
78. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **16**, 284–287 (2012).
79. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

80. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).

Chapter 4: Conclusions and Future Directions

My dissertation made several contributions to cancer research: (1) providing unique value resources of paired snRNA-seq and snATAC-seq of 30 ccRCC tumors and 4 normal adjacent tissues, the largest single-nucleus datasets for ccRCC to our knowledge; (2) a better understanding of intratumor heterogeneity, transcription regulation and transcriptomic and epigenetic changes associated with *BAP1* and *PBRM1* mutations in primary human ccRCCs; (3) discovering a novel combination treatment for RCC and elucidated protein changes associated with treatment effect. Moving forward, I will discuss some of the future directions below.

4.1 Experimental validation for the single-nucleus RNA/ATAC analysis of primary human ccRCCs

In Chapter 2, we generated an impressive ccRCC dataset consisting of snRNA-seq data of 34 samples (30 tumors + 4 normal adjacent tissue (NATs)) and snATAC-seq data of 28 samples (24 tumors + 4 NATs). By mapping the transcriptomic and epigenetic landscape specifically altered in ccRCC tumor cells, we identified ccRCC tumor-specific markers with therapeutic potential, ccRCC-specific alterations in the glycolytic pathway, heterogeneous tumor subpopulations, and tumor clusters exhibiting a continuum of epithelial to mesenchymal features. In addition, we also investigated the transcriptomic and epigenetic changes associated with BAP1 and PBRM1 mutations, which frequently occur in ccRCC. However, this study would benefit greatly from additional experimental validation of the computational findings, especially for the ccRCC tumor-specific markers and the relation between transcription factors and their putative target genes.

We identified 20 ccRCC tumor-specific markers, among which we highlighted CP and PCSK6 as the two most promising tumor markers. Ceruloplasmin (CP) as a ccRCC marker has been reported and studied by multiple groups previously[1–3]. My work in chapter 2 provided additional insights into the upstream regulation of CP by studying its promoter's chromatin accessibility and potential TF regulators, the spatial expression of CP, and its potential role in mediating tumor-stroma interactions based on the spatial transcriptomics data. However, both of these analyses were performed with a lack of validation. For example, our results suggest that transcription factor MXI1 may drive the transcription of CP and other genes. One of our future directions is to knock down MXI1 in ccRCC cell lines using vector-based short hairpin RNA (shRNA). Similarly, we can also perform shRNA-mediated knockdown of CP gene in ccRCC cell lines and perform bulk RNA-seq and western blot for the knockdown and control cell lines. Specifically, a western blot can be performed to validate the knockdown status. For cells with MXI1 knocked down, western blot can further confirm whether the putative MXI1 target CP is down-regulated in the MXI1 knockdown cell lines. By comparing the bulk RNA-seq data of the knockdown and control cell lines, we can evaluate genes and pathways affected by the knocked-down gene, such as CP and MXI1, and examine whether CP knockdown affects VEGF expression and pathways related to tumor-stroma interactions. Similarly, as we identified transcription factors MXI1 and KLF9 might drive the transcription of multiple glycolytic genes, we test whether MXI1 or KLF9 knockdown affects the glycolysis activation in ccRCC cells by measuring the glucose intake and lactate production of the knockdown and control cell lines. Finally, we also identified several druggable tumor markers, some of which we have tested using specific inhibitors in ccRCC cell lines. Although the results suggest some tumor inhibition effect,

the difference between tumor cell lines and one control cell line HEK293 is not very clear, potentially due to the high passage number of HEK293, making it more like tumor cells. Additional testing with a new low-passage HEK293 and ccRCC cell lines could be done, as well as testing more inhibitors against druggable markers such as PCSK6 and testing using ccRCC patient-derived organoids we have in-house.

4.2 Deeper analysis of the ccRCC tumor microenvironment using the single-nucleus sequencing data

Clear cell renal cell carcinoma is one of the cancer types that respond to immune checkpoint inhibitors[4,5]. And the ccRCC tumor microenvironment (TME) is characterized by the extensive infiltration of regulatory T cells and various myeloid lineage immune cell types, such as tumor-associated macrophages and myeloid-derived suppressor cells (MDSC), which have been recognized to suppress the immune system both dependent and independent of the interaction between PD1 and its ligands[6]. Previous studies have identified several immunosuppression mechanisms in the ccRCC TME, such as amino acid depletion, NO and kynurenine production[7–13]. In Chapter 2, we identified 13 lymphoid lineage immune cell types and 16 myeloid lineage immune cell types, including Treg and various tumor-associated macrophage populations. Although we have done some analysis focusing on the immune cells, due to the space limitation and the plethora of topics we want to explore for tumor cells, we did not include those results in Chapter 2. Going forward, our future direction will be to reevaluate the existence of MDSC in our dataset as well as new immunosuppressive mechanisms in light of existing literature with therapeutic potentials.

In addition to the immune cells, the ccRCC is also one of the most highly vascularized of all cancer types. It is thought to be a result of the overproduction of VEGF and other pro-angiogenic cytokines by tumor cells driven by the loss of VHL and the subsequent accumulation of transcription factors HIF-1 and HIF-2[14]. The ccRCC vasculature has been reported to consist of multiple types of microvessels with varying levels of differentiation. Specifically, Qian and colleagues[15,16] identified two types of RCC endothelium, including differentiated vessels that express both CD31 and CD34, and a less differentiated type of vessel that expresses only CD31. These two types of vessels have different localization within the tumor mass and pericyte coverage. More importantly, the undifferentiated vessels are more abundant in high-grade RCC and they are associated with shorter survival in patients. Similar studies have also found six distinct vessel types[17–20], among which all but two types lost the dependency on VEGF and may persist despite the VEGFR inhibition, limiting the efficacy of VEGFR inhibitors in the treatment of RCC patients. In Chapter 2, we identified the endothelial cell population as a whole across our 34 samples. However, deeper characterization of the endothelial cells in light of the reported endothelial subpopulations will help us better understand the landscape of vasculature in primary ccRCC.

Furthermore, the accessory cells in the TME, including endothelial cells, cancer-associated fibroblasts, and innate immune cells, are thought to enhance their tumor-promoting ability not through the mutational programming and genetic instability but rather are epigenetically reprogrammed upon their recruitment by factors produced by the tumor cells and TME[21,22]. A recent study[23] showed such reprogramming can involve epigenome modification in the infiltrating myeloid cells in lung and oesophageal cancer cells, which can be blocked by drugs

targeting the epigenome. Thus, another future direction is to use our paired snRNA-seq and snATAC-seq dataset to investigate how normal accessory cells in the ccRCC TME may be corrupted by epigenetic regulation to support tumor development and progression.

4.3 Experimental validation for the sapanisertib plus cabozantinib treatment

In Chapter 3, we identified the mTORC1/2 inhibitor sapanisertib potentiates the antitumor effect of the FDA-approved RTK inhibitor cabozantinib, inducing tumor growth arrest or regression after 1-month treatment. Moving from the preclinical studies to future clinical trials, we need to compare our new drug combination with the FDA-approved combination of RTK inhibitor lenvatinib and mTORC1 inhibitor everolimus. We need to evaluate whether cabozantinib plus sapanisertib performs better than lenvatinib plus everolimus in terms of efficacy, toxicity, and other parameters. We have some preliminary results on this comparison using two PDX models (not shown), and it seems to suggest the two combinations have similar tumor inhibition effects as well as tolerance. One of our future directions would be to test more PDX models for these two combinations and systemically evaluate whether there is a differential effect on cellular proliferation, migration, and invasion between the two combinations.

We also observed in Chapter 3 one interesting case with PIK3CA hotspot mutation H1047R showing the highest resistance to cabozantinib. PIK3CA H1047R mutation was known to increase the PI3K kinase activity and confer multiple oncogenic features[24,25]. It was associated with resistance to MET inhibition in head and neck cancer[26]. So it can be anticipated that PIK3CA mutation may also confer resistance to cabozantinib in RCC. To validate this

hypothesis, we will collect more PIK3CA-mutated RCC models or use isogenic cell lines, perform cabozantinib testing and compare their response to PIK3CA wild-type counterparts. Similarly, additional experimental validation is needed to validate the relation between MET overexpression and cabozantinib resistance. Based on our proteomics analysis, increased EMT-related proteins may mediate resistance to cabozantinib plus sapanisertib co-treatment. We can confirm their treatment-associated change using western blot and potentially perform knockdown or inhibitor study in cell lines to see if blocking EMT may result in further tumor inhibition.

4.4 Identifying new therapeutic options in cancer

The major treatment categories for patients with inoperable or metastatic RCC are systemic treatments using targeted agents and/or immune checkpoint inhibitors. In Chapter 3, we took advantage of the valuable resources being a part of the WashU PDX Development and Trial Center (PDTC) and tested a variety of drugs in the RCC PDX models. However, PDX models only contain a microenvironment of murine cells, lacking immune components in the immunocompromised mice to test potential immunotherapy. Our lab has been developing autochthonous mouse models for ccRCC with conditional knockout of a single allele (Pax8-rtTA, tetO-Cre, Vhl f/f) and multiple alleles among Vhl, Pbrm1, and Setd2 (Pax8-rtTA, tetO-Cre, Vhl f/f, Pbrm1 f/f; Pax8-rtTA, tetO-Cre, Vhl f/f, Pbrm1 f/f, Setd2 f/f). The generation of ccRCC genetic engineered mouse models that resemble human ccRCC, if successful, will allow testing of immune checkpoint inhibitors as well as combinations with the targeted agents.

5.5 References

1. Maruschke, M. *et al.* Putative biomarker genes for grading clear cell renal cell carcinoma. *Urol. Int.* **87**, 205–217 (2011).
2. Takahashi, M. *et al.* Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9754–9759 (2001).

3. Zhang, Y. *et al.* Ceruloplasmin overexpression is associated with oncogenic pathways and poorer survival rates in clear-cell renal cell carcinoma. *FEBS Open Bio* **11**, 2988–3004 (2021).
4. Motzer, R. J. *et al.* Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* **373**, 1803–1813 (2015).
5. Motzer, R. J. *et al.* Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* **378**, 1277–1290 (2018).
6. Mier, J. W. The tumor microenvironment in renal cell cancer. *Curr. Opin. Oncol.* **31**, 194–199 (2019).
7. Rodriguez, P. C. *et al.* Arginase I-producing myeloid-derived suppressor cells in renal cell carcinoma are a subpopulation of activated granulocytes. *Cancer Res.* **69**, 1553–1560 (2009).
8. Markowitz, J. *et al.* Nitric oxide mediated inhibition of antigen presentation from DCs to CD4+ T cells in cancer and measurement of STAT1 nitration. *Sci. Rep.* **7**, 15424 (2017).
9. Delage, B. *et al.* Arginine deprivation and argininosuccinate synthetase expression in the treatment of cancer. *Int. J. Cancer* **126**, 2762–2772 (2010).
10. Klysz, D. *et al.* Glutamine-dependent α-ketoglutarate production regulates the balance between T helper 1 cell and regulatory T cell generation. *Sci. Signal.* **8**, ra97 (2015).
11. Gross, M. I. *et al.* Antitumor activity of the glutaminase inhibitor CB-839 in triple-negative breast cancer. *Mol. Cancer Ther.* **13**, 890–901 (2014).
12. Hornyák, L. *et al.* The Role of Indoleamine-2,3-Dioxygenase in Cancer Development, Diagnostics, and Therapy. *Front. Immunol.* **9**, 151 (2018).
13. Mitchell, T. C. *et al.* Epacadostat Plus Pembrolizumab in Patients With Advanced Solid Tumors: Phase I Results From a Multicenter, Open-Label Phase I/II Trial (ECHO-202/KEYNOTE-037). *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **36**, 3223–3230 (2018).
14. Hsieh, J. J. *et al.* Renal cell carcinoma. *Nat. Rev. Dis. Primer* **3**, 17009 (2017).
15. Qian, C.-N., Huang, D., Wondergem, B. & Teh, B. T. Complexity of tumor vasculature in clear cell renal cell carcinoma. *Cancer* **115**, 2282–2289 (2009).
16. Cao, Y. *et al.* Pericyte coverage of differentiated vessels inside tumor vasculature is an independent unfavorable prognostic factor for patients with clear cell renal cell carcinoma. *Cancer* **119**, 313–324 (2013).
17. Sitohy, B., Nagy, J. A., Jaminet, S.-C. S. & Dvorak, H. F. Tumor-surrogate blood vessel subtypes exhibit differential susceptibility to anti-VEGF therapy. *Cancer Res.* **71**, 7021–7028 (2011).
18. Sitohy, B., Nagy, J. A. & Dvorak, H. F. Anti-VEGF/VEGFR therapy for cancer: reassessing the target. *Cancer Res.* **72**, 1909–1914 (2012).
19. Nagy, J. A. & Dvorak, H. F. Heterogeneity of the tumor vasculature: the need for new tumor blood vessel type-specific targets. *Clin. Exp. Metastasis* **29**, 657–662 (2012).
20. Nagy, J. A., Chang, S.-H., Dvorak, A. M. & Dvorak, H. F. Why are tumour blood vessels abnormal and why is it important to know? *Br. J. Cancer* **100**, 865–869 (2009).
21. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
22. Hanahan, D. & Coussens, L. M. Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. *Cancer Cell* **21**, 309–322 (2012).
23. Lu, Z. *et al.* Epigenetic therapy inhibits metastases by disrupting premetastatic niches. *Nature* **579**, 284–290 (2020).

24. Samuels, Y. *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554 (2004).
25. Samuels, Y. *et al.* Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell* **7**, 561–573 (2005).
26. Nisa, L. *et al.* PIK3CA hotspot mutations differentially impact responses to MET targeting in MET-driven and non-driven preclinical cancer models. *Mol. Cancer* **16**, 93 (2017).
27. Jaykumar, A. B. *et al.* WNK1 Enhances Migration and Invasion in Breast Cancer Models. *Mol. Cancer Ther.* **20**, 1800–1808 (2021).