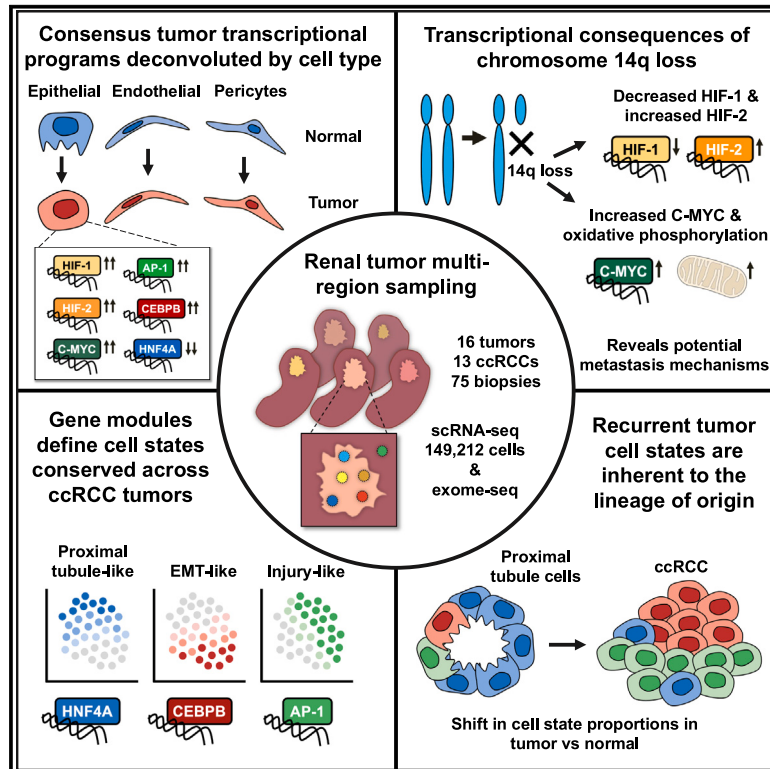


Conserved patterns of transcriptional dysregulation, heterogeneity, and cell states in clear cell kidney cancer

Graphical abstract



Authors

Olivia Lombardi, Ran Li, Faiz Jabbar, ..., Hani Choudhry, Peter J. Ratcliffe, David R. Mole

Correspondence

david.mole@ndm.ox.ac.uk

In brief

Lombardi et al. identify conserved dysregulated transcriptional programs within ccRCC and associated vasculature. They describe the transcriptional consequences of metastasis-associated Chr14q loss. They identify additional prognostically significant gene modules reflecting conserved intratumor transcriptional heterogeneity and marking co-existing cell states and lineage plasticity found in both cancer and normal cells.

Highlights

- ccRCC exhibits hallmark transcriptional programs in cancer and stromal cells
- Conserved cancer-stroma interactions characterize ccRCC tumors
- Chr14q loss is associated with altered HIF-1/2 and C-MYC transcriptional responses
- Cancer cell states recapitulate normal cell subpopulations, reflecting plasticity



Article

Conserved patterns of transcriptional dysregulation, heterogeneity, and cell states in clear cell kidney cancer

Olivia Lombardi,^{1,7} Ran Li,^{1,7} Faiz Jabbar,¹ Hannah Evans,¹ Silvia Halim,¹ Joanna D.C.C. Lima,^{1,4} Lisa Browning,² Helen M. Byrne,^{3,4} Hani Choudhry,⁵ Peter J. Ratcliffe,^{4,6} and David R. Mole^{1,8,*}

¹NDM Research Building, University of Oxford, Old Road Campus, Headington, Oxford OX3 7FZ, UK

²Department of Cellular Pathology, Oxford University Hospitals NHS Foundation Trust, Headington, Oxford OX3 9DU, UK

³Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

⁴Ludwig Institute for Cancer Research, University of Oxford, Old Road Campus, Headington, Oxford OX3 7DQ, UK

⁵Department of Biochemistry, Faculty of Science, Center of Innovation in Personalized Medicine, King Fahd Center for Medical Research, King Abdulaziz University, Jeddah 3270, Saudi Arabia

⁶The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

⁷These authors contributed equally

⁸Lead contact

*Correspondence: david.mole@ndm.ox.ac.uk

<https://doi.org/10.1016/j.celrep.2024.115169>

SUMMARY

Clear cell kidney cancers are characterized both by conserved oncogenic driver events and by marked intra-tumor genetic and phenotypic heterogeneity, which help drive tumor progression, metastasis, and resistance to therapy. How these are reflected in transcriptional programs within the cancer and stromal cell components remains an important question with the potential to drive novel therapeutic approaches to treating cancer. To better understand these programs, we perform single-cell transcriptomics on 75 multi-regional biopsies from kidney tumors and normal kidney. We identify conserved patterns of transcriptional dysregulation and their upstream regulators within the tumor and associated vasculature. We describe recurrent subclonal transcriptional consequences of Chr14q loss linked to metastatic potential. We identify prognostically significant conserved patterns of intratumor transcriptional heterogeneity. These reflect co-existing cell states found in both cancer cells and normal kidney cells, indicating that rather than arising from genetic heterogeneity they are a consequence of lineage plasticity.

INTRODUCTION

Kidney cancer is among the 10 commonest cancer types with approximately 295,000 new cases and 134,000 deaths worldwide each year.¹ Clear cell renal cell carcinoma (ccRCC) is the most frequent subtype comprising over 80% of cases.² Tumors are associated with truncal mutation of the von Hippel-Lindau (*VHL*) tumor suppressor gene and constitutive activation of hypoxia-inducible factor (HIF)-driven transcriptional pathways.³ Both HIF-1 and HIF-2 isoforms are stabilized and have overlapping, but distinct, roles in ccRCC tumorigenesis.⁴ Additional mutations, affecting the epigenetic regulators polybromo 1 (PBRM1), SET domain containing 2 (SETD2), and BRCA1-associated deubiquitinase 1 (BAP1), often co-occur with *VHL* inactivation.^{5–9} These may be associated with changes in chromatin accessibility, the HIF pathway, and immune/angiogenic pathways. Additionally, copy-number alterations are common, including Chr9p and Chr14q loss, which are both associated with metastasis.¹⁰ However, these large-scale chromosomal alterations encompass hundreds of genes and are technically challenging to model *in vitro*, hindering elucidation of their downstream molecular effects.

Two distinguishing hallmarks of ccRCC are marked intratumor heterogeneity and cancer-stromal interactions. Importantly, current therapeutic options for treating advanced kidney cancer largely target these interactions.¹¹ Specifically, belzutifan (an HIF-2 inhibitor) and tyrosine kinase inhibitors block the secretion and effects of vascular growth factors on their endothelial cell receptors, and immune checkpoint inhibitors alter interactions with immune effectors.

Intratumor heterogeneity helps drive tumor progression, metastasis, and resistance to therapy.^{12–15} Indeed, ccRCCs exhibit a high degree of subclonal genetic and chromosomal diversity that correlate with clinical outcomes.^{10,16} Extensive intratumor heterogeneity is also observed histologically.¹⁷ However, the transcriptional consequences of genetic heterogeneity, the transcriptional drivers of morphological diversity, and the role of epigenetic modifications and tumor microenvironment remain unclear.¹⁸ Over time, these heterogeneous transcriptional programs may allow cellular subpopulations to outcompete other cells, generating accelerated tumor growth or resistance to therapy. Therefore, it is important to understand both the crosstalk between different cell types within the tumor and the transcriptional



programs underlying heterogeneous cancer cell subpopulations. Unfortunately, bulk RNA sequencing approaches, by averaging across all cells, are unable to distinguish the transcriptional programs of individual cell types or subclones.

We have undertaken multi-region single-cell RNA sequencing (scRNA-seq) of kidney cancers and surrounding normal renal tissue. This identified conserved programs of transcriptional dysregulation in malignant cancer cells driven by activation of specific transcription factors. These signal to stromal cells within the tumor leading to conserved patterns of transcriptional dysregulation and cellular crosstalk. Combining scRNA-seq data with whole exome sequencing and bulk RNA-seq, we characterized the transcriptional effects of Chr14q loss, a driver of ccRCC metastasis. Analysis of transcriptional heterogeneity within the malignant cancer cells, agnostic to genetics, identified three gene modules representing conserved cell states found in all ccRCC tumors. These comprise two mutually exclusive states (differentiated proximal-tubule-like, and epithelial-mesenchymal transition [EMT]-like states) that exist in a continuum and an uncoupled cell-injury-like state. Each module reflects cell states also observed in the cancer-founder cell type (normal proximal tubular cells) indicating that they reflect non-genetic heterogeneity and cellular plasticity inherent to the lineage of origin.

RESULTS

Multi-region tumor sampling and single-cell transcriptomics

Multi-region scRNA-seq of primary kidney tumors removed at nephrectomy was performed, focusing on ccRCC (Figure 1A). 55 samples were obtained from 16 tumors in 15 patients (median 3 samples per tumor, range 1–8) (Figures 1B and Table S1). Thirteen were clear cell (ccRCC), one (T22) was papillary (pRCC), one (T23) was renal oncocyoma (oRCC), and one (T35) was unclassified (uRCC). All patients were treatment naive with the exception of patient 35, who was on first-line systemic immunotherapy with nivolumab. Patient 26 had VHL syndrome with an R167W germline *VHL* mutation. 20 background kidney samples were also obtained from macroscopically “normal” kidney tissue within each nephrectomy specimen (range 1–4). Samples from 12 patients (Pt18/22/23/24/25/26/29/30/31/32/33/35) were analyzed immediately following tissue dissociation (biopsy samples). Samples from 3 patients (Pt3/17/20), and normal kidney from Pt18 were used to generate primary cultures prior to separate scRNA-seq analysis.

After filtering, data were obtained from 149,212 cells (132,424 from biopsy samples and 16,788 from cultured samples, Tables S2 and S3). The median number of detected features per cell was 1,778. Hierarchical clustering was performed, and marker genes for each cluster were used to assign broad cell types (Figures 1C, 1D, S1, and S2).^{19–22} Between-patient cell-mixing scores, stratified by cell type, were low for tumor epithelial cells but, in contrast, were high for stromal tumor cells and for normal samples processed in parallel (Figures S1I and S1J). This signifies inter-tumor heterogeneity in tumor epithelial cells in the absence of systemic batch effects affecting other cell types.

Tumor samples had fewer epithelial cells and more pericytes and immune cells than normal samples (Figure S1K), while

epithelial cells were the most abundant (median 38.9%, Figure 1E). Tumor samples from Pt35 (on immunotherapy), yielding only 4 epithelial cells, and normal samples from Pt26 (VHL syndrome) containing visible cysts were excluded from further analysis. Following culture, epithelial cells accounted for a greater proportion (median 96.3%) of cells (Figure 1F) likely due to the growth medium selectively promoting growth of these cells. However, these cultures provided material for confirmatory whole exome sequencing and to study the effects of hypoxia/HIF-2 inhibition (Figures S2I–S2P).

Genetic characterization of tumor samples was undertaken using InferCNV (<https://github.com/broadinstitute/inferCNV>) and Numbat²³ (Figure 1B) focusing on previously described recurrent renal cell carcinoma (RCC) “driver” events.¹⁶ All ccRCC epithelial cell clusters exhibited loss of heterozygosity of Chr3p encompassing the *VHL* gene locus. Tumors T18A and T18B, from patient 18, showed distinct patterns of copy-number variation (CNV), indicating they were separate synchronous tumors. pRCC and oRCC epithelial cells exhibited loss of heterozygosity characteristic of these cancer subtypes²⁴ (chr7/17 gain and chr1/14 loss, respectively).

Conserved transcriptional program dysregulation in cancer epithelial cells

We first sought to explore dysregulated gene expression programs in ccRCC cancer epithelial cells relative to normal renal tubular epithelial cells. Epithelial cells from the tumor and normal biopsy samples were reclustered and remapped (Figure 2A). Marker gene analysis of epithelial cell clusters identified a range of expected tubular and glomerular epithelial cell types in normal samples^{19,20,25,26} (Figures S3A–S3I). Normal epithelial cells from different patients clustered closely together with high between-patient cell-mixing scores (Figure S1I). Epithelial cell clusters from tumor samples were distinguished by Chr3p loss (Figure S3J) and high levels of the HIF target and ccRCC marker gene, carbonic anhydrase 9 (CA9)²⁷ (Figure S3A).

Notably, normal and cancer epithelial cells clustered distinctly from one another, demonstrating a profound transcriptional change associated with malignant transformation (Figure 2A). However, while normal epithelial cells from different patients overlapped considerably, the malignant epithelial cell clusters demonstrated much less overlap and low between-patient cell-mixing scores (Figures S1I and S1J). This, together with distinct clustering of T18A and T18B (separate tumors from the same patient), indicates inter-tumor heterogeneity, due to true biological variability.

We next examined for conserved patterns of gene dysregulation in tumor epithelial cells compared to normal tubular cells. First, epithelial dysregulated genes in each of the 10 freshly biopsied tumors were separately identified using the Wilcoxon rank-sum test. Proximal tubular cells, the most likely cell of origin for ccRCC,^{20,28} were used as the normal reference. 3,264 genes were upregulated and 2,337 downregulated in at least one of the 10 ccRCCs. Gene set intersection analysis revealed that dysregulated genes were statistically much more likely ($p < 0.001$) to be either unique to an individual tumor or conserved across all 10 ccRCCs (Figures 2B and 2C) and identified conserved gene sets comprising 153 upregulated and 203 downregulated genes

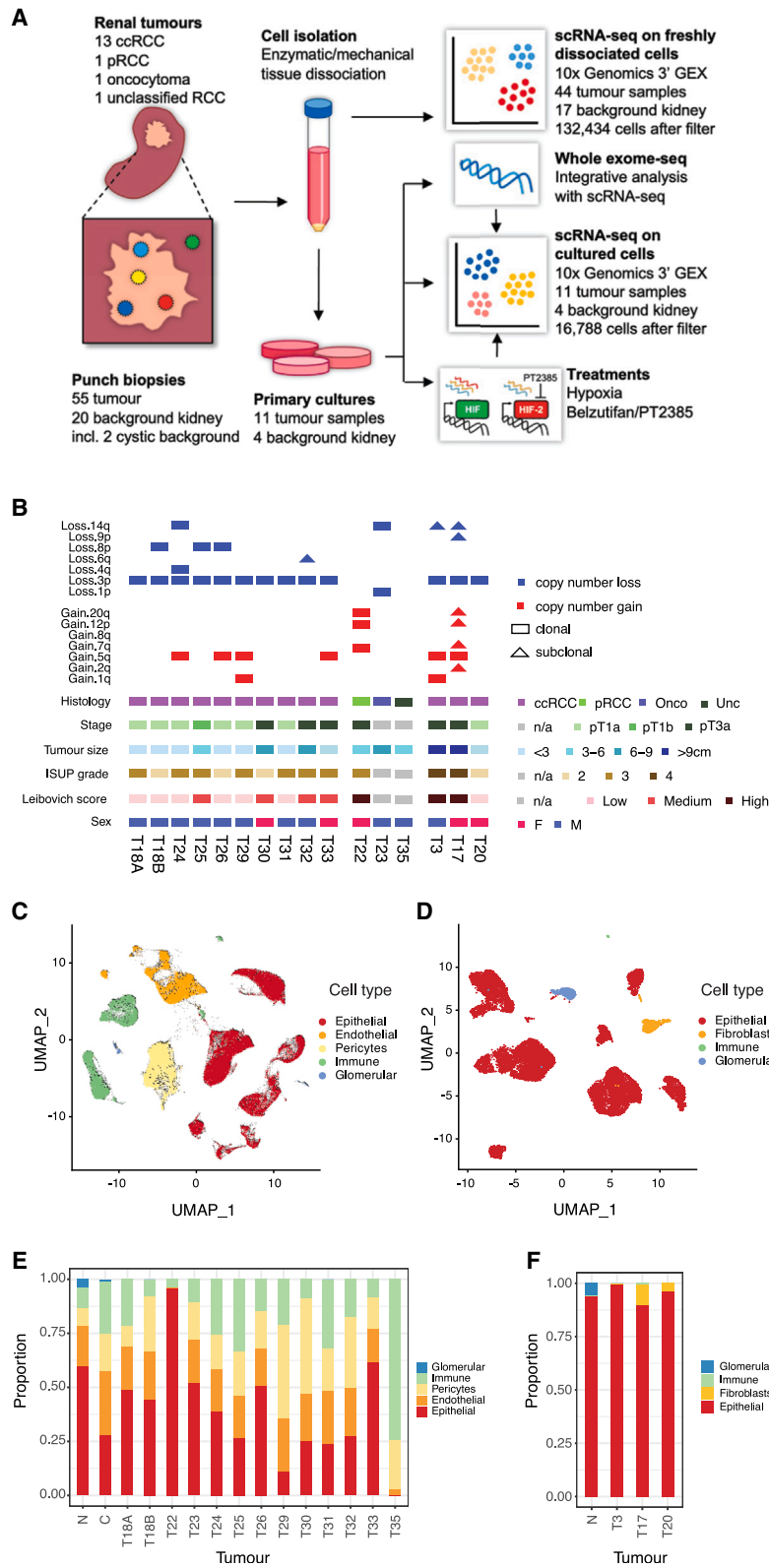


Figure 1. Study overview

(A) Schematic of sample analysis.

(B) Summary of tumors analyzed, showing CNV (inferred from scRNA-seq data), histological subtype, tumor stage, tumor size, International Society of Urological Pathology (ISUP) grade, Leibovich score, and patient sex.

(C and D) Uniform manifold approximation and projection (UMAP) plots of cells from (C) freshly biopsied tumors and (D) primary tumor cultures showing major cell types.

(E and F) Stacked bar charts showing major cell type proportions for cells from (E) freshly biopsied tumors and (F) primary tumor cultures. See also [Figures S1](#) and [S2](#).

(Table S4). Comparatively few genes were upregulated/downregulated in subsets of ccRCC indicating an absence of common ccRCC subtypes defined by large programs of dysregulated genes in cancer epithelial cells. However, subsets of these genes were deregulated in the pRCC and/or oRCC (Figure 2D). Analysis of conserved genes in The Cancer Genome Atlas (TCGA-KIRC) bulk RNA-seq database confirmed upregulation/downregulation in ccRCC tissue samples compared to normal kidney samples and high expression of the 153 upregulated genes in ccRCC tumors relative to other tumor types (Figures S4A–S4C).

Genes upregulated in cancer cells from all 10 ccRCC tumors included the classical HIF target genes *CA9* and *VEGFA*, as well as the atypical mitochondrial subunit *NDUFA4L2*²⁰ (Figure 2E). Gene set enrichment analysis (GSEA) using hallmark gene sets (The Molecular Signatures Database [MSigDB]) revealed positive enrichment of hypoxia-associated genes in ccRCC cancer cells (Figure 2F). Additional enriched pathways included tumor necrosis factor alpha (TNF- α)/nuclear factor κ B (NF- κ B), interferon, angiogenesis, and Notch signaling. Negatively enriched (downregulated) pathways included oxidative phosphorylation, peroxisome, and fatty acid metabolism. Since proximal tubule (PT) cells rely on fatty acid oxidation metabolism via peroxisomes and mitochondria,²⁹ this is consistent with loss of physiological cellular function in tumors.

We next identified the transcription factors underlying these dysregulated gene programs using epigenetic landscape *in silico* deletion analysis (LISA).³⁰ HIF-1 α , endothelial PAS domain protein 1 (HIF-2 α /EPAS1), and aryl hydrocarbon receptor nuclear translocator (HIF-1 β /ARNT) binding were enriched at upregulated genes consistent with their known stabilization in ccRCC (Figure 2G). MAF BZIP transcription factor B (MAFB), P53, V-Rel avian reticuloendotheliosis viral oncogene homolog A (RELA), and signal transducer and activator of transcription 3 (STAT3) binding were also highly enriched. MAFB is a bZIP transcription factor important for podocyte differentiation during development and protects renal tubules from apoptosis.³¹ Consistent with previous bulk studies,^{32,33} activities of hepatocyte nuclear factor 4 (HNF4) isoforms, which have roles in PT development/differentiation,^{34–36} are depleted, suggesting cellular dedifferentiation.

While HIF regulation is predominantly post-translational, other transcription factors may be regulated transcriptionally. Cross-referencing our LISA analysis with expression analysis identified

transcription factors for which both expression and activity were dysregulated in all 10 ccRCCs (Figures 2H and Table S5). Activator protein 1 (AP-1) transcription factors (FOS, FOSL2, JUN, JUNB, and JUND), other immediate-early transcription factors (ATF3, EGR1, and MYC), CCAAT enhancer binding proteins (CEBPB and CEBPD), another avian musculoaponeurotic fibrosarcoma (MAF) family transcription factor (MAFF), immune regulatory transcription factors (IRF1, NFIL3, and STAT3), an EMT master regulator (ZEB2), chromatin remodelers (BRD2, CHD1, and CHD2), epigenetic modifiers (EPC1, JMJD1C, JMJD6, KDM6B, KMT2A, and SAP30), Kruppel-like factor (KLF) transcription factors (KLF4 and KLF10), and avian erythroblastosis virus E26 oncogene homolog (ETS) transcription factors (ETS1 and ETS2) showed upregulation of their mRNA and downstream target genes. The only transcription factor whose expression and activity were downregulated in ccRCC was HNF4A, which directly controls genes critical for PT function^{37,38} again suggesting dedifferentiation in ccRCC cancer cells.

Conserved transcriptional programs in non-epithelial cancer stromal cells

Beside epithelial cells, tumor and normal biopsy samples contain blood vessels (comprising endothelial cells and pericytes) and immune cells (predominantly macrophages and T cells) (Figure 1G). We examined for conserved programs of gene expression in these non-epithelial stromal tumor cells. Endothelial cells from normal and ccRCC samples clustered separately on uniform manifold approximation and projection (UMAP) plots, suggesting different transcriptional profiles (Figure 3A). Examining genes dysregulated in individual tumors compared to normal samples identified 8,011 genes upregulated and 6,242 genes downregulated in endothelial cells in at least one ccRCC. Gene set intersection analysis (Figures 3B and 3C) revealed that, again, most deregulated genes were unique to individual tumors. However, 218 genes were upregulated and 16 downregulated in endothelial cells in all 10 tumors (Table S6)—significantly more than expected ($p < 0.001$). Both gene signatures were deregulated compared to normal kidney in analysis of bulk RNA-seq data from TCGA-KIRC (Figure 3D). GSEA of endothelial subtype markers³⁹ showed enrichment of tip cell markers among ccRCC upregulated genes (Figure 3E). LISA analysis showed enrichment for binding sites of TAL1, GATA2, and FLI, as well as

Figure 2. Analysis of conserved patterns of gene dysregulation in epithelial cells from ccRCC samples compared to proximal tubular cells from surrounding normal kidney samples

- (A) UMAP plot of epithelial cells (freshly biopsied samples) showing tumor/normal samples for each patient (multiple normal or tumor regions combined).
 (B) Intersection analysis showing overlap between genes upregulated in epithelial cells from each of the 10 freshly biopsied ccRCC tumors compared to normal proximal tubular cells.
 (C) The same analysis for downregulated genes. In both cases the number of dysregulated genes common to all 10 tumors was significantly higher than expected ($p < 0.001$, by bootstrapping).
 (D) Heatmap showing expression of the same 153 individual genes in different cell types.
 (E) Volcano plot showing average \log_2 (fold change) versus average $-\log_{10}(p$ value) for each gene in epithelial cells from the 10 ccRCCs compared to normal proximal tubular cells. Average expression was calculated for each tumor and then averaged across the 10 tumors, so each tumor was weighted equally. Red denotes genes with \log_2 (fold change) >0.25 or <-0.25 and Bonferroni-adjusted p value <0.05 .
 (F) GSEA enrichment of hallmark pathways among genes up- or downregulated in epithelial cells from the 10 ccRCCs compared to normal PTs.
 (G) Scatterplot showing $-\log_{10}(p$ value) from LISA analysis of transcription factor binding enrichment at up- vs. downregulated genes in epithelial cells from the 10 ccRCCs compared to normal PTs.
 (H) Scatterplot showing $-\log_{10}(p$ value) from LISA analysis of transcription factor binding enrichment at genes up- (positive values) or downregulated (negative transformed values) plotted against average \log_2 (fold change) in mRNA level for the given transcription factor. See also Figure S3 and S4.

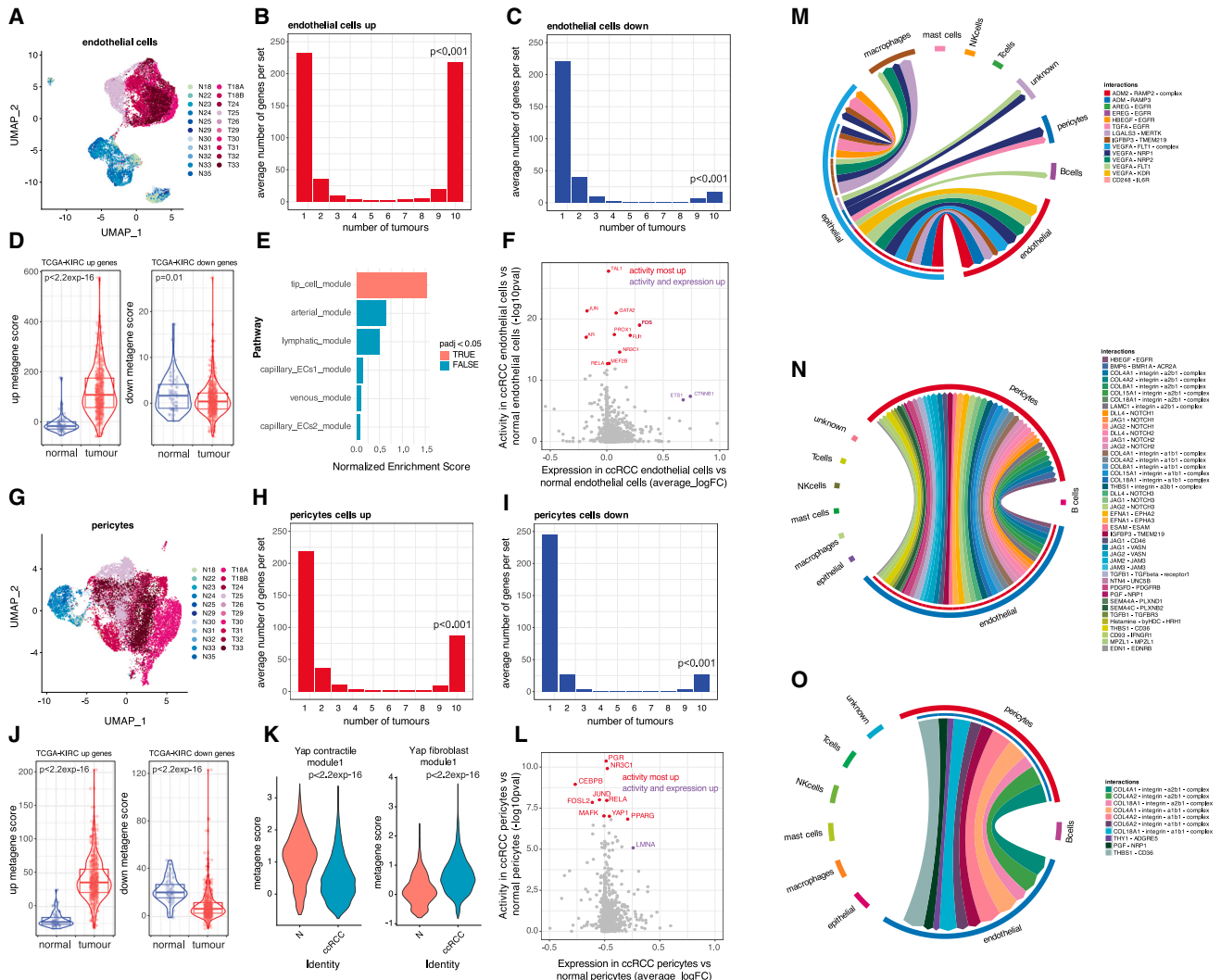
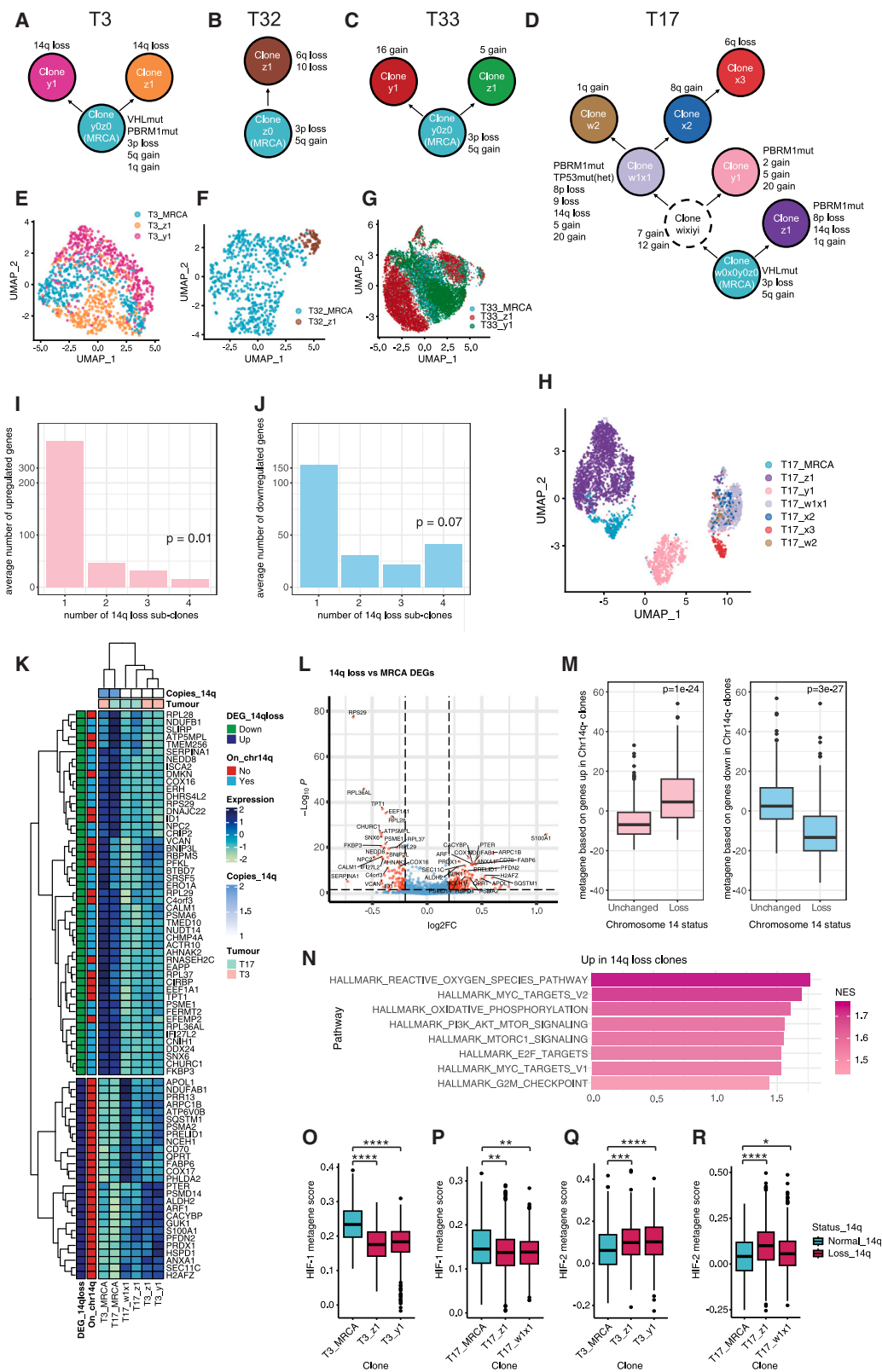


Figure 3. Analysis of conserved patterns of gene dysregulation in endothelial and pericyte cell populations

(A) UMAP plot of endothelial cells (freshly biopsied samples) showing sample type and patient.
 (B) Intersection analysis showing overlap between genes upregulated in endothelial cells from each of the 10 ccRCC tumors compared to normal kidney samples.
 (C) The same analysis for downregulated genes. In both cases the number of genes dysregulated in all 10 ccRCCs was more than expected ($p < 0.001$ by bootstrapping).
 (D) Violin and box-and-whisker plots showing composite metagene expression, in bulk RNA analysis of normal and tumor samples from the TCGA-KIRC cohort, for genes up- or downregulated in endothelial cells from all 10 ccRCC compared to normal samples.
 (E) GSEA showing enrichment of endothelial tip cell marker genes among genes upregulated in endothelial cells.
 (F) Scatterplot showing $-\log_{10}(p \text{ value})$ from LISA analysis of transcription factor binding at genes upregulated in endothelial cells from all 10 ccRCCs compared to normal samples plotted against the fold change in mRNA for each transcription factor.
 (G) UMAP plot of pericytes (freshly biopsied samples) showing sample type and patient.
 (H and I) Intersection analysis to identify genes (H) up- and (I) downregulated in pericytes from all 10 ccRCCs compared to normal samples ($p < 0.001$ by bootstrapping).
 (J) Violin and box-and-whisker plots showing expression of composite metagenes based on these up- and downregulated genes in bulk RNA-seq analysis of normal and tumor samples from the TCGA-KIRC cohort.
 (K) Violin plots showing expression of composite metagenes reflecting contractile or fibroblast-like behavior.
 (L) Scatterplot showing $-\log_{10}(p \text{ value})$ from LISA analysis versus average $-\log_{2}(\text{fold change})$ in mRNA level for transcription factors enriched at genes upregulated in tumor compared to normal pericytes.
 (M) Chord diagram showing ligand-receptor interactions (identified in CellphoneDB) from ccRCC epithelial cells to other cell types involving genes upregulated in epithelial cells in all 10 ccRCCs.
 (N) The same analysis for interactions from endothelial cells involving genes upregulated in endothelial cells from all 10 ccRCCs.
 (O) The same analysis for pericytes. Box-and-whisker plots show median, inter-quartile range, and range. See also [Figure S5](#) and [S6](#).



(legend on next page)

ETS1 and CTNNB1, whose mRNA was also upregulated (Figure 3F). All these transcription factors have been implicated in orchestrating the angiogenic response in endothelial cells.^{40–42}

Pericytes from normal and ccRCC samples also cluster separately on UMAP plots (Figure 3G). Gene set intersection analysis again showed statistically significant sets of conserved upregulated/downregulated genes (88/27) in tumor versus normal pericytes (Figures 3H, 3I and Table S7). Metagene signatures derived from these gene sets confirmed upregulation/downregulation in bulk RNA-seq TCGA-KIRC data (Figure 3J). Notably, the atypical mitochondrial subunit NADH dehydrogenase 1 alpha subcomplex subunit 4-like 2 (NDUFA4L2) was upregulated, while NDUFA4 was downregulated (Figures S5A and S5B). While in some settings this switch results from direct transcriptional regulation by HIF,⁴³ we did not observe HIF activation in ccRCC pericytes (Figure S5C). This suggests that the switch from NDUFA4 to NDUFA4L2 in ccRCC pericytes is due to paracrine signaling from HIF-expressing, ccRCC epithelial cells. We also observed an increase in collagen subunits and genes associated with a previously defined,²⁵ fibroblast-like vascular smooth muscle cell phenotype together with reduced expression of genes associated with a contractile phenotype indicating a functional change in pericytes associated with tumorigenesis (Figure 3K). Enrichment of binding sites for several transcription factors was also observed (Figure 3L).

Conversely, macrophages and T cells from normal and tumor samples clustered together in UMAP plots (Figures S5D–S5I) and no significant overlap in dysregulated macrophage or T cells genes was observed between each of the 10 ccRCC samples. Thus, our analysis did not distinguish conserved transcriptional profiles in ccRCC macrophages or T cells.

Analysis of cell-cell interactions in ccRCC tumors

We then examined intratumor cell-cell communication using the CellPhoneDB database⁴⁴ of receptor-ligand interactions. We first analyzed conserved features of cell-cell signaling from ccRCC epithelial cells to other cell types, using the 153 conserved ccRCC upregulated genes (Figure 3M). Direct signaling interactions were observed with several other cell types predominantly between ccRCC epithelial cells and tumor endothelial cells. These involved signaling by the HIF target genes adrenomedullin (ADM) and its paralog adrenomedullin 2 (ADM2) to endothelial receptor activity-modifying protein

(RAMP) receptors: galectin 3 (LGALS3) to myeloid epithelial reproductive proto-oncogene tyrosine kinase (MERTK), insulin-like growth factor binding protein 3 (IGFBP3) to transmembrane protein 219 (TMEM219), and vascular endothelial growth factor alpha (VEGFA) to various endothelial receptors (see Figure S6 for additional immunostaining). Direct signaling from ccRCC epithelial cells was also observed to pericytes, macrophages, and B cells as well as cell-type-autonomous signaling back to ccRCC epithelial cells.

Based on the 88 conserved genes upregulated in all 10 ccRCCs, tumor endothelial cells communicated exclusively with tumor pericytes (Figure 3N). Signaling pathways involved included delta-like and jagged canonical notch ligands (DLL4, JAG1, and JAG2) signaling to NOTCH1/2; platelet-derived growth factor D (PDGFD) signaling to its receptor (PDGFRB); binding of collagen (COL4A1, COL4A2, COL8A1, COL15A1, and COL18A1), laminin (LAMC1), and thrombospondin (THBS1) molecules to cell surface integrins; epidermal growth factor (EGF) family members, heparin-binding EGF-like growth factor (HBEGF), and transforming growth factor (TGF)- α ; semaphorin 4A and 4C (SEMA4A and SEMA4C); placental growth factor (PGF); netrin 4 (NTN4); and endothelin (EDN1). Communication between ccRCC endothelial cells and pericytes was bidirectional with pericytes expressing high levels of collagens, PGF, THBS1, and Thy-1 cell surface antigen (THY-1) able to interact with cell surface molecules on endothelial cells (Figure 3O).

Effect of intratumor genetic heterogeneity on transcriptional profile

ccRCC is characterized by high intratumor genetic heterogeneity with recurrent subclonal mutations and chromosomal aberrations. However, it is unclear how these alter the transcriptional output of epithelial cancer cells. We examined the effect of genetic heterogeneity on transcriptional programs in ccRCC. Genetic subclones, marked by different patterns of CNV within the epithelial cells, were identified in tumors T3/17/32/33 (Figures S7, S8, and S9), and phylogenetic trees, ordering subclones by evolutionary stage, were constructed (Figures 4A–4H) aided by exome sequencing (exome-seq) of T3/17 (Table S8). Genetic abnormalities often exhibited parallel evolution, occurring independently in multiple subclones, including previously described driver events: *PBRM1* mutation, 14q loss, 1q gain, and 8p loss.^{10,16,45,46}

Figure 4. Analysis of intratumor genetic heterogeneity

(A–D) Analysis of CNV and loss of heterogeneity (LOH) identified subclonal architecture and tumor phylogenies in epithelial cells from 4 ccRCC tumors. Subclone T17w_xy_i (D) represents a hypothetical “intermediate” antecedent of T17w_x and T17y_i.
(E–H) UMAP plots of epithelial cells from each of the 4 tumors showing individual subclones.
(I and J) Intersection analysis of genes (I) up- or (J) downregulated in the 4 subclones with Chr14q loss compared to their antecedent chr14q wild-type subclones.
(K) Heatmap of genes up- or downregulated in all 4 subclones with chr14q loss.
(L) Volcano plot showing average log₂(fold change) and average $-\log_{10}(p$ value) for differential gene expression between subclones with chr14q loss and their antecedent chr14q wild-type subclones. Average expression was calculated for each subclone and then averaged across the 4 subclones, so each subclone was weighted equally. Red denotes genes with log₂(fold change) >0.25 or <−0.25 and Bonferroni-adjusted p value <0.05.
(M) Box-and-whisker plot showing expression of composite metagenes, based on genes up- or downregulated in subclones with Chr14q loss, in bulk RNA-seq analysis of tumor samples from the TCGA-KIRC cohort segregated according to Chr14q status.
(N) GSEA enrichment of hallmark pathways among genes upregulated in Chr14q loss subclones compared to antecedent 14q wild-type subclones.
(O–R) Box-and-whisker plot showing expression of composite metagenes, based on genes up- or downregulated in subclones with Chr14q loss in scRNA-seq analysis of these subclones. * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001, (Holm-adjusted) Wilcoxon rank-sum test. Box-and-whisker plots show median, inter-quartile range, and range. See also Figures S7–S10.

Chr3p loss, a known truncal event in ccRCC,⁴⁷ was consistently observed in all subclones except a T17 subclone, which had copy neutral Chr3p loss of heterozygosity. Chr14q loss, a putative driver of metastasis,¹⁰ typically encompasses *HIF1A*, which acts as a tumor suppressor in ccRCC.^{48–52} Notably, its paralog, *EPAS1* (*HIF2A*), which promotes ccRCC tumorigenesis,^{48,49,52} is on Chr2p. However, Chr14q loss encompasses hundreds of genes, and therefore the mechanism associated with tumor metastasis remains unclear. We identified four subclonal incidences of Chr14q loss: two discrete events in T3 and two in T17. Two Chr14q loss subclones had additional CNVs. However, the only common CNV was Chr14q loss. Differential gene expression analysis for each Chr14q loss subclone compared to the most recent common ancestor subclone from the corresponding tumor identified genes dysregulated in all four subclones to delineate the core transcriptional response to Chr14q loss (Figures 4I–4L and Table S9).

49 genes were downregulated in all four subclones. 31 were located on Chr14q and 18 on other chromosomes, distinguishing direct effects of Chr14q loss from secondary effects. 27 genes were upregulated in all four subclones. These included *APOL1*, one of four genes previously associated with high tumor grade and poor prognosis in ccRCC⁵³; *SQSTM1* (encoding p62), associated with high-grade ccRCCs and ccRCC oncogenesis⁵⁴; and *CD70*, implicated in immune evasion.⁵⁵ Composite metagenes based on these gene signatures reflected Chr14q status in bulk RNA-seq analysis of tumors from the TCGA-KIRC cohort (Figure 4M), demonstrating the generalizability of our findings.

Both oxidative phosphorylation and myelocytomatosis oncogene (*MYC*) targets were significantly enriched among genes upregulated by Chr14q loss, while no hallmark gene sets were enriched among downregulated genes (Figure 4N). Mitochondrial respiration is upregulated in metastatic compared to primary ccRCC, and ccRCC cells with elevated oxidative phosphorylation have increased metastatic potential in mice.⁵⁶ Therefore, the association between Chr14q loss and mitochondrial respiration may, at least partially, explain its association with metastasis. *MYC*, a proposed oncogene in ccRCC,⁵⁷ is a transcriptional target of HIF-2 α in ccRCC⁵⁸ and associated with HIF-2-dominant tumors lacking HIF-1.⁵⁹

Despite Chr14q encompassing *HIF1A*, the “hypoxia” hallmark gene set was not significantly enriched among Chr14q loss downregulated genes. However, some genes in the hypoxia gene set were downregulated, while others were upregulated. To test whether HIF-1 and HIF-2 target genes behave differently upon 14q loss, we performed targeted small interfering RNA (siRNA) knockdown of *HIF1A/2A* in RCC4 cells, followed by RNA-seq analysis to identify HIF-1/2 target genes (Figure S10). Isoform-specific target genes were identified (excluding genes regulated by both isoforms) and leveraged as metagenes to deconvolute HIF-1/2 activity (Tables 10 and S11). Their validity in scRNA-seq analysis was confirmed in primary cultures through their regulation by hypoxia and the HIF-2-specific inhibitor (belzutifan) (Figures S10D and S10E). HIF-1 metagene activity was depleted upon Chr14q loss, consistent with copy-number loss of the *HIF1A* gene (Figures 4O and 4P). Conversely, HIF-2 metagene activity was elevated in response to Chr14q loss

(Figures 4Q and 4R). HIF-2 synergizes with *MYC*^{4,59} and may help drive its increased activity. Together, this indicates that Chr14q loss causes decreased HIF-1 activity, with a reciprocal increase in HIF-2 activity, *MYC*, and oxidative phosphorylation.

Conserved patterns of intratumor transcriptional heterogeneity in ccRCC epithelial cancer cells

We next investigated conserved patterns of intratumor (transcriptional) heterogeneity. Epithelial cells from 8 ccRCCs, with over 500 captured cancer cells, were clustered individually, and “subcluster” markers within each were identified (Figure 5A). Marker genes were more likely to be unique to one tumor or common to all 8 tumors ($p < 0.001$). 57 genes marked cancer cell subpopulations in all 8 ccRCCs (Figure 5B) representing hallmarks of ccRCC intratumor heterogeneity.

Hierarchical clustering of expression of these 57 genes across subclusters in the 8 ccRCCs identified three major groups (modules) of genes (Figures 5C and Table S12). Module-1 genes included PT marker genes (e.g., brush border proteins) determined by cross-referencing markers of normal epithelial cells in our dataset (Table S13). Module-2 genes included several markers of a previously described EMT “meta program”⁶⁰ and ceruloplasmin (CP), previously associated with high ccRCC grade and poor patient prognosis and which may mediate tumor-stroma crosstalk.⁹ Module-3 genes were characteristic of damaged/injured PTs, including immediate-early transcription factors and inflammation markers.^{19,61,62} Immunostaining for representative genes from each module confirmed heterogeneous protein expression (Figure S11).

There was striking negative correlation between expression of Module-1 and -2 genes (Figures 5D and 5G–5O), suggesting that they represent mutually exclusive cellular properties, similar to the findings of a previous study, which noted a negative correlation between PT- and EMT-meta programs.⁶⁰ Expression of Module-3 genes varied on a different axis in UMAP plots (Figures 5I and 5L), suggesting that it represents a distinct cellular property, uncoupled from Module-1 or -2. Module-1 correlated with previously derived *PAX8* and *HNF1B* gene programs⁶³ consistent with this module representing a PT-like state (Figure S12). Module-2 negatively correlated with these modules, while Module-3 showed no significant correlation. However, our modules only partially aligned with *PAX8* and *HNF1B* modules in principal component analysis (PCA) analysis indicating that they are at least partially distinct from these programs.

Expression of each module varied continuously across tumors (Figures 5D and 5G–5O) and was leveraged to identify additional genes that covaried (Spearman rank correlation) with each core module in each tumor (Table S14). GSEA, using gene lists ranked according to these correlations, identified oxidative phosphorylation and fatty acid metabolism (metabolic processes related to normal PT function) as most enriched among Module-1-associated genes (Figure S13A). EMT and angiogenesis were most enriched among Module-2-associated genes. Positively enriched Module-1 pathways were negatively enriched among Module-2-associated genes, and vice versa. Pathways most enriched among Module-3-associated genes were TNF- α signaling via NF- κ B, interleukin-6 (IL-6)-JAK-STAT3 signaling, and TGF- β signaling.

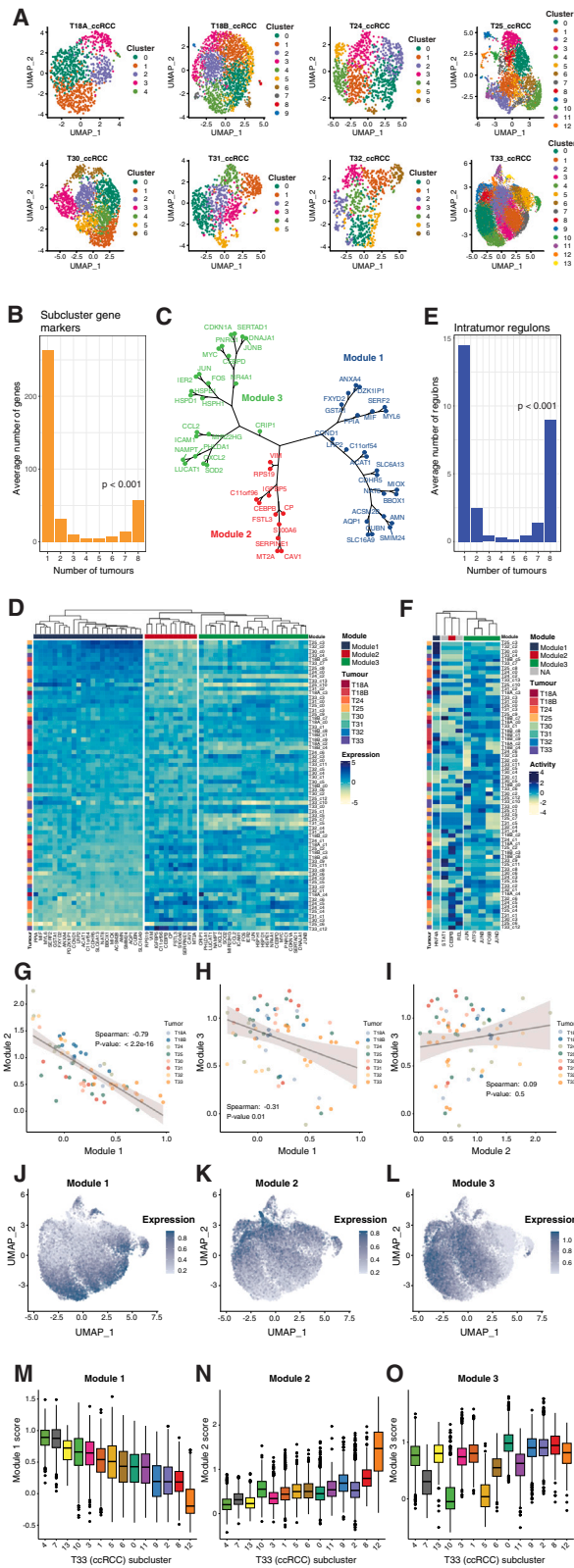


Figure 5. Conserved patterns of transcriptional heterogeneity in epithelial cells from ccRCC samples

(A) UMAP plots showing epithelial cells from each ccRCC, remapped and clustered separately. Only tumors with >500 epithelial cells from freshly biopsied samples were analyzed.

(B) Intersection analysis of genes marking subclusters in epithelial cells from each of the 8 ccRCCs. Genes marking subclusters in all 8 tumors were enriched ($p < 0.001$ by bootstrapping).

(C) k-means hierarchical clustering of cluster markers common to all 8 ccRCC, based on normalized average expression in each cluster, identifying 3 main gene modules.

(D) Heatmap showing expression of each common cluster marker gene in individual clusters ordered by the sum of Module-1 gene values.

(E) Intersection analysis of transcriptional regulons (defined by SCENIC) exhibiting intratumor heterogeneity in epithelial cells from each of the 8 ccRCCs. Regulons variable within all 8 tumors were enriched ($p < 0.001$ by bootstrapping).

(F) Heatmap showing activity of each common cluster marker gene in individual clusters ordered by the sum of Module-1 gene values.

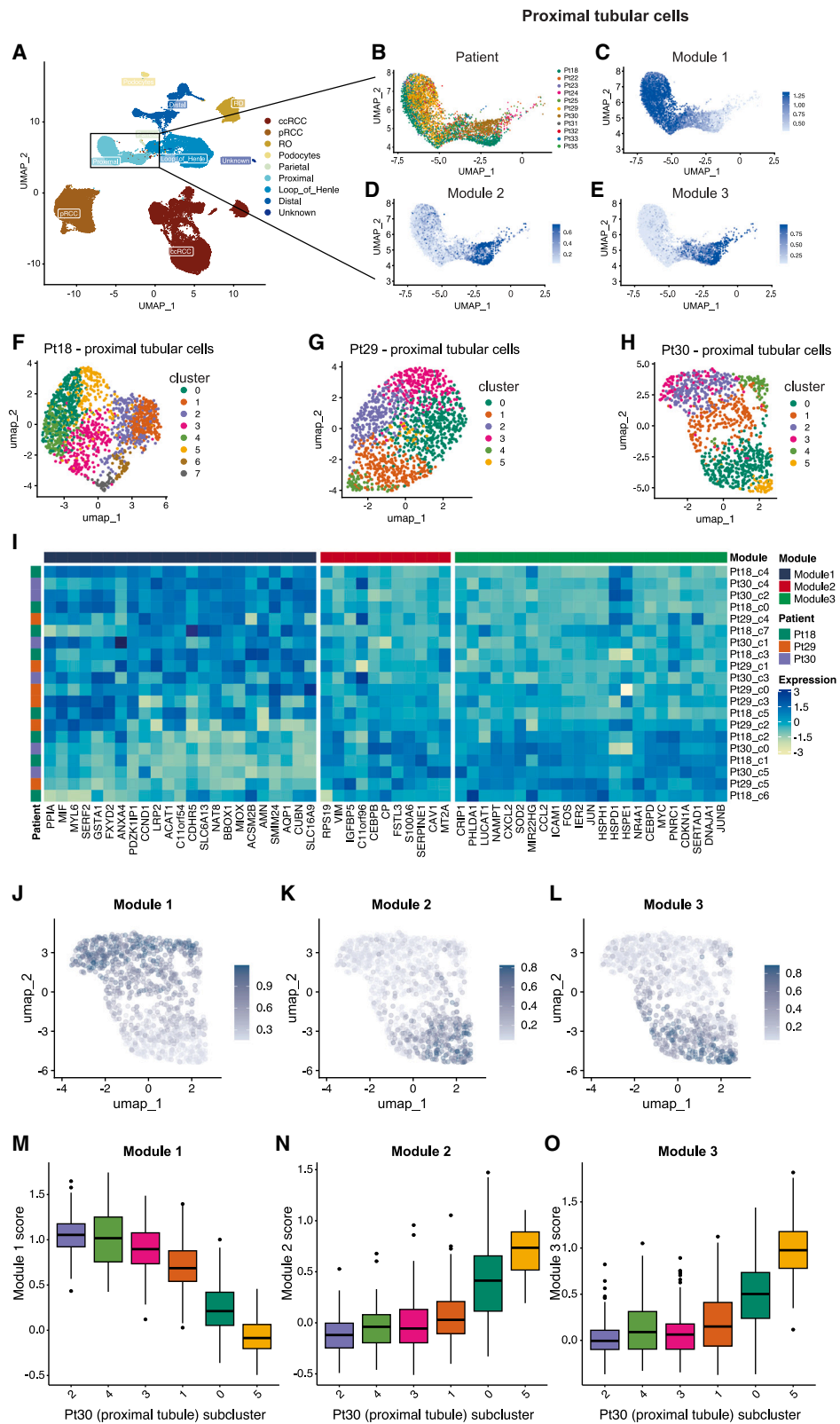
(G–I) Scatterplots showing mean expression of composite metagenes based on Module-1, -2, and -3 genes in individual tumor clusters. R = Pearson correlation coefficient.

(J–L) UMAP plots showing expression of composite metagenes based on Module-1, -2, and -3 genes in epithelial cells from tumor T33.

(M–O) Box-and-whisker plots showing T33 cluster-level composite module scores for Module-1, -2, and -3 in clusters ordered by median Module-1 score. Box-and-whisker plots show median, inter-quartile range, and range. See also Figures S11–S14.

Barkley et al. reported recurrent cancer cell subpopulations agnostic to tumor type in pan-cancer scRNA-seq analysis.¹⁸ Barkley oxidative phosphorylation genes were most enriched among Module-1-associated genes (likely reflecting mitochondrial activity related to PT function) and negatively enriched among Module-2-associated genes (Figure S13B). Barkley metal module genes were most enriched among Module-2-associated genes and negatively enriched among Module-1-associated genes. The Barkley stress module was most enriched among Module-3-associated genes.

We then performed single-cell regulatory network interference clustering (SCENIC) analysis⁶⁴ to identify transcription factors with heterogeneous intratumor activity in each of the 8 tumors (Figure S14). 9 transcription factors common to all 8 ccRCCs were identified (Figure 5E, $p < 0.001$). These included HNF4A (Figure 5F), which is required for PT differentiation^{37,38} and is in the extended list of Module-1-correlated genes, and CEBPB in the core Module-2. Consistently, single-nuclei assay for transposase-accessible chromatin using sequencing (ATAC-seq) showed greater chromatin accessibility at CEBPB motifs in EMT-like ccRCC populations and at HNF4A motifs in epithelial-like populations.⁹ Activities of AP-1 family immediate-early transcription factors (JUN/JUNB also in the core Module-3 and ATF3/FOSB/JUND in the extended list of Module-3-correlated genes) also exhibited intratumor heterogeneity across all 8 tumors. Lastly, activity of the inflammatory transcriptional factors, REL and STAT1, exhibited common intratumor heterogeneity, although regulation is likely post-translational since their mRNAs did not vary. HNF4A activity was highest in subclusters with the highest Module-1 expression, whereas CEBPB activity was highest in subclusters with the highest Module-2 expression and exhibited an inverse correlation between each other



(legend on next page)

(Figure 5F). STAT1 and REL activities also clustered with CEBPB, suggesting that they may be additional transcriptional regulators associated with Module-2.

Recurrent cell states were identified in every ccRCC tumor irrespective of stage, grade, or size, with modules exhibiting graded expression, suggesting transitional cells with intermediate, dynamic phenotypes. Since Module-1 marks PT differentiation and Module-3 denotes PT injury, we postulated that these cancer cell states might represent lineage plasticity inherent to the cell of origin. We therefore examined expression of each of the modules in normal proximal tubular cells. Expression of all 3 was heterogeneous, varying at both single-cell and cluster level (Figure 6). The existence of these cell states in both normal progenitor cells and in ccRCC cells provides evidence that they are not driven solely by genetic factors in the tumor and indicates cellular plasticity since each tumor will have derived from a single founder cell. Inverse correlation was again observed between modules 1 and 2 (Figures 6C, D, 6I–K, 6M, and 6N). However, modules 2 and 3 were highly correlated (Figures 6D, 6E, 6I, 6K, 6L, 6N, and 6O), indicating tight coupling in normal cells, unlike cancer cells. Although the modules were heterogeneous in both ccRCC and normal PT cells, overall, modules 2/3 had higher and Module-1 had lower expression in cancer cells versus normal cells (Figures 7A–7C). Thus, tumor cells were more likely to adopt EMT/injured-like states underpinned by modules 2/3 and less likely to adopt the PT-like state underpinned by Module-1. Consistent with this, analysis of deregulated transcriptional regulators (Figure 2H) showed downregulation of the Module-1-associated transcriptional regulator (HNF4A) at both the RNA and activity level in ccRCC cells and upregulation of Module-2 (CEBPB) and Module-3 (ATF3, FOS/JUN isoforms)-associated transcriptional regulators.

Module-1 also varied in pRCC, also derived from PT, but not in oRCC, derived from distal tubule (Figure S15). Some variability in modules 2 and 3 was observed in both pRCC and oRCC, suggesting that the associated EMT-like and injury-like cell states are not specific to cancers deriving from the PT lineage. However, while expression of Module-3 is elevated in all 8 ccRCCs, Module-3 expression in the pRCC was comparable to that in the PT (Figure 7C), suggesting that the shift in cell states proportions may be specific to ccRCC.

While the three modules varied within ccRCCs, they also varied between ccRCCs, suggesting different proportions of each cell state (Figures 7A–7C). We leveraged this variation to examine associations with tumor grade and prognosis in the TCGA-KIRC cohort. High Module-1 expression was associated

with low tumor grade, while Module-2 increased in higher-grade tumors (Figures 7D–7F). Modules 1 and 2 were prognostic in ccRCC with Module-1 predictive of good and Module-2 predictive of poor prognosis (Figures 7G and 7H).

DISCUSSION

scRNA-seq allows deconvolution of complex multicellular environments into individual cell types and facilitates examination of transcriptional and genetic heterogeneity within populations of cells. Differential gene expression and gene set intersection analysis identified conserved patterns of transcriptional dysregulation within epithelial ccRCC cells common to all ccRCCs studied. These gene programs are driven by multiple transcription factors in addition to HIF and reflect those modulated in PTs during kidney fibrosis or injury.^{19,26,65–67} This suggests that most ccRCC malignant cells adopt transcriptional programs analogous to injured cell states in the PT epithelium that converge on a dedifferentiated, inflammatory phenotype. Binding sites for many of these transcriptional regulators have previously been reported to overlap with HIF binding sites⁶⁸ suggesting that they may modulate the HIF response.

Similarly, conserved patterns of transcriptional dysregulation were also identified in endothelial cells and pericytes from ccRCC samples. Conversely, we did not identify core transcriptional patterns that distinguished ccRCC macrophages or T cells. This may result from reduced statistical power due to low cell numbers, high stringency of our analysis, or the comparison with immune cells invading the peritumor “normal” kidney, which may resemble those found within the tumors. ccRCC endothelial cells showed upregulation of tip cell marker genes consistent with active sprouting angiogenesis. Pericytes exhibited downregulation of contractile genes with activation of a more fibroblast-like gene profile and a switch from typical to atypical mitochondrial subunit expression. Activation of these genes was greater than in most other cancer types and likely results from paracrine signaling from the tumor epithelial cells involving HIF target genes. Additionally, bidirectional signaling between endothelial and pericyte cell populations was also identified.

ccRCC cancers are characterized by high intratumor genetic heterogeneity, including many putative “driver” events,^{16,45} although their transcriptional effects are unclear. We identified tumors with subclonal Chr14q loss (a putative driver of ccRCC metastasis) allowing its transcriptional effects to be studied. Chr14q contains the *HIF1A* gene locus encoding for one of two

Figure 6. Conserved patterns of epithelial cell heterogeneity are reflected in proximal tubular cells from normal kidney samples

- (A) UMAP plot of epithelial cells from normal and tumor samples (freshly biopsied) showing cell type.
 (B) UMAP plot of proximal tubular cells from normal kidney samples showing patient.
 (C–E) UMAP plots showing expression of composite metagenes based on (C) Module-1, (D) Module-2, and (E) Module-3 genes in proximal tubular cells from normal kidney samples.
 (F–H) UMAP plots showing proximal tubular cells from normal kidney samples from each patient, remapped and clustered separately. Only patients with >500 PT cells were analyzed.
 (I) Heatmap showing average cluster-level expression of Module-1, -2, and -3 genes in proximal tubular cells from normal kidney samples with clusters ordered by the sum of Module-1 gene values.
 (J–L) UMAP plots showing expression of composite metagenes based on Module-1, -2, and -3 genes in normal proximal tubular cells from patient 30 (N30).
 (M–O) Box-and-whisker plots showing N30 cluster-level composite module scores for modules 1, 2, and 3 in clusters arranged by median Module-1 score. Box-and-whisker plots show median, inter-quartile range, and range.

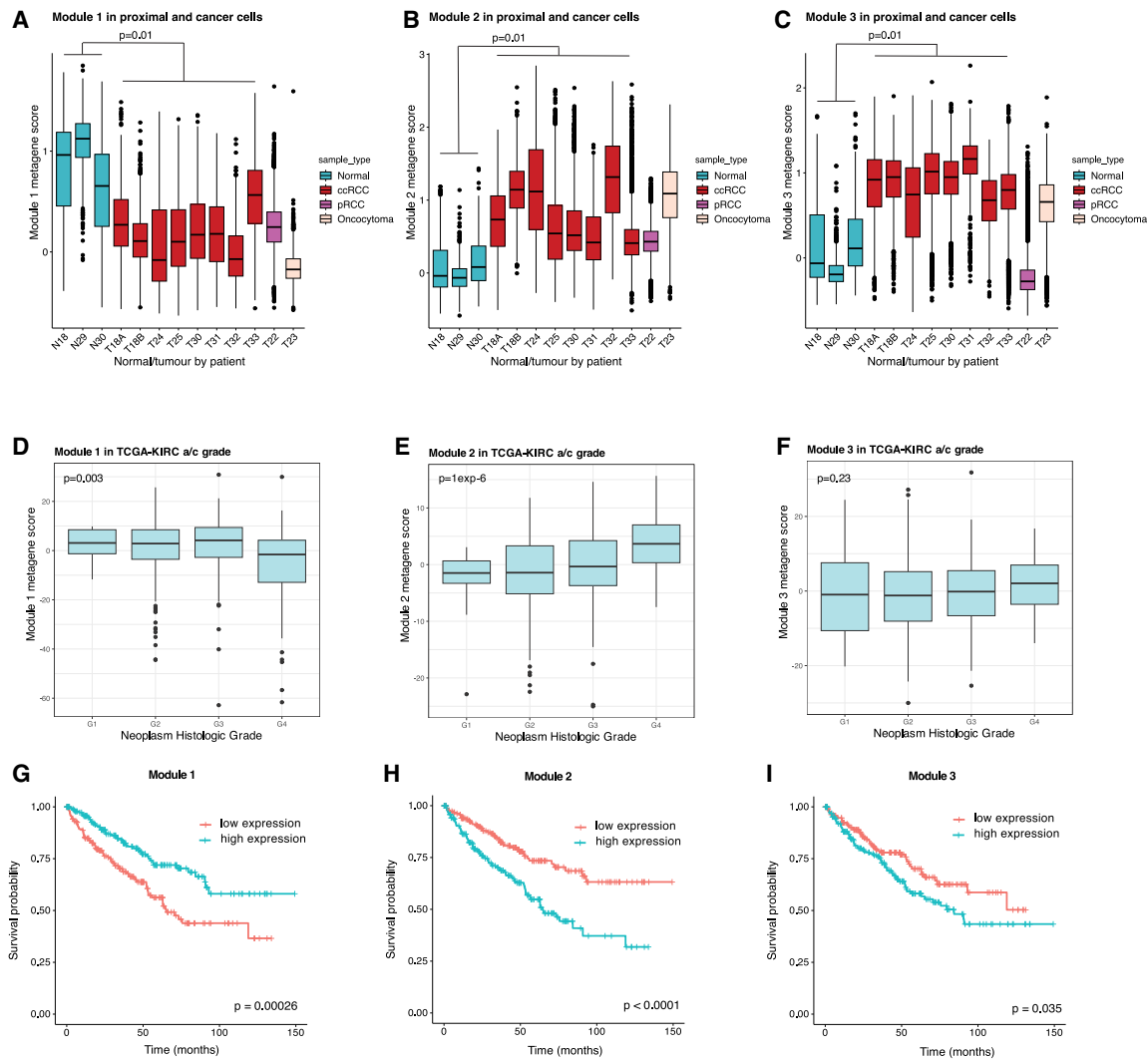


Figure 7. Analysis of Module-1, -2, and -3 gene expression in single-cell and bulk RNA-seq analysis from the TCGA-KIRC cohort

Box-and-whisker plots showing expression of composite metagenes based on (A) Module-1, (B) Module-2, and (C) Module-3 in freshly biopsied normal PT cells (blue) and epithelial cells from individual tumors, showing only tumors/patients with >500 cells per group. Box-and-whisker plots showing expression of composite metagenes reflecting (D) Module-1, (E) Module-2, and (F) Module-3 activity in ccRCC tumors (TCGA-KIRC) stratified according to tumor grade (p value from Kruskal-Wallis test). Kaplan-Meier plots showing overall survival in TCGA-KIRC patients stratified according to expression of (G) Module-1, (H) Module-2 and (I) Module-3 composite metagenes. Box-and-whisker plots show median, inter-quartile range, and range. See also [Figure S15](#).

HIF transcription factors that are constitutively activated in ccRCC. Previous studies indicate a restrictive role for HIF-1 in ccRCC progression, while HIF-2 promotes tumor development.^{48,49,51,52,69} In our study, loss of Chr14q was associated with reduction in HIF-1 activity with concomitant increase in HIF-2 activity. Oxidative phosphorylation, previously shown to be rate-limiting for metastasis in a mouse model,⁵⁶ was also upregulated by Chr14q loss. Furthermore, the ccRCC oncogene, SQSTM1, which correlates with both tumor grade and mitochondrial activity in ccRCCs and drives anchorage-independent cell growth,⁵⁴ was also upregulated by Chr14q loss.

Our analysis identifies an additional primary source of intratumor transcriptional heterogeneity in ccRCC epithelial cells character-

ized by three conserved gene programs representing differentiated (Module-1), EMT-like (Module-2), and damaged/injured-like (Module-3) cell states. These gene programs varied continuously across the ccRCC epithelial cell populations with the differentiated and EMT modules representing opposite ends of a continuous spectrum, while the damaged/injured gene module represented an independently varying cellular property. We observed similar heterogeneity in these modules in PT cells from normal kidney, which also resembled previously described modules identified in normal PTs.^{19,20,26,28,53,62,65,66,70} Generally, these separate normal PTs into two main cell states: a well-differentiated state and an injured-like state marked by VCAM1 expression, analogous to our recurrent cancer cell states marked by Module-1 and

Module-3, respectively. Although VCAM1 did not meet our stringent criteria to be included in the gene modules due to being variable in 7/8 (rather than 8/8) tumors, it correlated ($r = 0.40$, $p = 0.0006$) with expression of Module-3 (Figure S12).

The presence of these cell states in both normal and tumor cells suggests that they are not genetically defined. Importantly, in tumors from other tissues, cell states inherent to the corresponding lineage of origin have been described,¹⁸ implicating this type of heterogeneity as a pervasive feature across tumor biology. Since tumors are derived clonally from single founder cells, it indicates plasticity whereby cells can transdifferentiate between states. Although ccRCC cancer cells have been reported to be most like injured-like VCAM1+ PTs,²⁰ the presence of multiple cell states within the tumor cell population makes it difficult to determine the founder state.

The ability of cells to switch between multiple co-existing states has widespread implications for tumor biology, likely contributing to tumor initiation, progression, metastasis, and therapeutic resistance.^{71,72} Specifically, transitioning of tumor cells to and from an EMT-like state has an important role in metastasis. While EMT promotes tumor cell invasion and adherence-independent survival, reversion back to an epithelial-like state once it has reached its destination is also important.^{71–73} Indeed, high EMT gene signature expression has been associated with worse clinical outcomes across many cancers,⁷⁴ and our EMT-like signature correlated with both a higher tumor grade and poorer patient survival. Therefore, perturbing the equilibrium between PT-like (Module-1) and EMT-like (Module-2) cancer cell states represents an opportunity for differentiation therapy. Additionally, our EMT-like module correlated with increased angiogenic signaling, suggesting that EMT-like cells may have heightened sensitivity to antiangiogenic tyrosine kinase inhibitors (TKIs). Future functional studies will be required to understand more precisely how modules 1 and 2 influence cellular phenotype. However, an injury-like signature similar to our Module-3 was previously shown to be transiently induced in tumor thrombi (through which ccRCCs can metastasize) suggesting a role in invasion.⁷⁵

There is also increasing evidence that the presence of multiple cell states within the tumor (and the ability of cells to switch between them) contributes to therapeutic resistance in cancer.^{76,77} While our work has revealed potential drug sensitivities and targets associated with the EMT-like cell state, it is likely that, through plasticity, ccRCC tumors will be able to switch to other cell states, which may be less sensitive to the targeted treatment. Understanding cell state dynamics in ccRCC and the root cause of plasticity will pave the way for novel therapeutic intervention.

Limitations of the study

By nature, scRNA-seq is a non-spatial analysis of RNA. Further work will be required to study the spatial relationships of our findings at both the RNA and protein level. Genetic analysis of the single-cell data was limited to a few CNVs. New techniques and greater numbers of patients will be needed to study the effects of SNVs. Our methodology enabled a much better recovery of epithelial and blood vessel cells than previous publications, resulting in proportionately fewer immune cells. Our study focused on

primary tumors, and extension to tumor metastases will be important.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, David R. Mole (david.mole@ndm.ox.ac.uk).

Materials availability

The study did not generate new unique reagents.

Data and code availability

- **Data:** Seurat objects and bulk RNA-seq data are available at Gene Expression Omnibus (GEO): GSE269819 and GSE269826 (patient 3 data were previously uploaded individually to GSE200207). Original western blot images are available at Mendeley (<https://doi.org/10.17632/7p25n8gwj2.1>).
- **Code:** This paper does not report original code.
- **Other items:** Microscopy data reported in this paper will be shared by the [lead contact](#) upon request. Any additional information required to re-analyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

This study was supported by the National Institute for Health Research (D.R.M.; NIHR-RP-2016-06-004); Cancer Research UK (D.R.M.; A416016); the Deanship of Scientific Research, King Abdulaziz University, the Ministry of Higher Education for Saudi Arabia (H.C., D.R.M., and P.J.R.); the DoD CDMRP Concept Award (O.L. and D.R.M.; KC190033); the Ludwig Institute for Cancer Research (P.J.R.); the Wellcome Trust (P.J.R.; 106241/Z/14/Z); and The Francis Crick Institute (P.J.R.), which receives its core funding from Cancer Research UK (FC001501), the UK Medical Research Council (FC001501), and the Wellcome Trust (FC001501). We acknowledge the contribution to this study made by the Oxford Centre for Histopathology Research and the Oxford Radcliffe Biobank, which are supported by the University of Oxford, the Oxford CRUK Cancer Centre, the NIHR Oxford Biomedical Research Centre (Molecular Diagnostics Theme/Multimodal Pathology Sub-theme), and the NIHR CRN Thames Valley network. The computational aspects of this research were supported by the Wellcome Trust Core Award grant number 203141/Z/16/Z and the NIHR Oxford BRC. We thank Miriam O'Hanlon for assistance in collecting nephrectomy specimens; Hubert Slawinski and Joseph Chadwick for assistance with 10× Genomics and sequencing; Julie Adam and Veronique Lafleur for assistance with primary cultures, siRNA optimization, and scientific discussions; and Chris Pugh for scientific discussions. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. We are grateful to the patients and their families for participating in this research.

AUTHOR CONTRIBUTIONS

Conceptualization, O.L., R.L., H.C., H.M.B., P.J.R., and D.R.M.; methodology, O.L., R.L., F.J., H.E., S.H., J.D.C.C.L., L.B., and D.R.M.; software, R.L., F.J., S.H., and D.R.M.; formal analysis, O.L., R.L., F.J., H.E., S.H., and D.R.M.; investigation, O.L., R.L., F.J., H.E., S.H., J.D.C.C.L., L.B., and D.R.M.; data curation, R.L., S.H., and D.R.M.; writing – original draft, O.L., R.L., and D.R.M.; writing – review and editing, all authors; visualization, O.L., R.L., F.J., S.H., and D.R.M.; supervision, O.L., H.M.B., P.J.R., and D.R.M.; project administration, P.J.R. and D.R.M.; funding acquisition, O.L., H.C., P.J.R., and D.R.M.

DECLARATION OF INTERESTS

P.J.R. reports grants from Ludwig Institute for Cancer Research and The Francis Crick Institute during the conduct of the study and personal fees from Immunocore plc outside the submitted work.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
 - Tissue sampling
 - Tissue processing
 - Primary cultures
 - Cell hashing for multiplexing primary cultures
 - 10x Genomics GEM generation and library preparation
 - Whole exome sequencing
 - siRNA transfection and bulk RNA-sequencing
 - Sequencing
 - Western blotting
 - Immunohistochemistry
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Processing of scRNA-seq data
 - Visualization, clustering, and differential gene expression
 - Inferring copy number variations in scRNA-seq data using InferCNV
 - Inferring copy number variations in scRNA-seq data using numbat
 - Somatic copy number analysis of exome-seq data
 - Somatic variant analysis in exome-seq data
 - Allele frequency analysis in exome-seq data
 - Allele frequency analysis in scRNA-seq data
 - RNA-seq analysis of RCC4 cells
 - Gene set intersection analysis
 - Gene set enrichment analysis
 - Transcriptional regulator prediction using LISA
 - Transcriptional regulator prediction using SCENIC
 - Cell-cell interaction analysis
 - TCGA RNA-seq data and prognosis analysis
 - Deciphering intratumor expression programs and meta-programs
 - Generation of HIF-1 and HIF-2 specific gene metagenes

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2024.115169>.

Received: July 26, 2024
 Revised: November 13, 2024
 Accepted: December 17, 2024
 Published: January 9, 2025

REFERENCES

1. Hsieh, J.J., Purdue, M.P., Signoretti, S., Swanton, C., Albiges, L., Schmindinger, M., Heng, D.Y., Larkin, J., and Ficarra, V. (2017). Renal cell carcinoma. *Nat. Rev. Dis. Prim.* 3, 17009. <https://doi.org/10.1038/nrdp.2017.9>.
2. Warren, A.Y., and Harrison, D. (2018). WHO/ISUP classification, grading and pathological staging of renal cell carcinoma: standards and controversies. *World J. Urol.* 36, 1913–1926. <https://doi.org/10.1007/s00345-018-2447-8>.
3. Shen, C., and Kaelin, W.G., Jr. (2013). The VHL/HIF Axis in Clear Cell Renal Carcinoma. *Semin. Cancer Biol.* 23, 18–25. <https://doi.org/10.1016/j.semcancer.2012.06.001>.
4. Hoefflin, R., Harlander, S., Schafer, S., Metzger, P., Kuo, F., Schonberger, D., Adlesic, M., Peighambari, A., Seidel, P., Chen, C.Y., et al. (2020). HIF-1 α and HIF-2 α differently regulate tumour development and inflammation of clear cell renal cell carcinoma in mice. *Nat. Commun.* 11, 4111. <https://doi.org/10.1038/s41467-020-17873-3>.

5. Brugarolas, J., Rajaram, S., Christie, A., and Kapur, P. (2020). The Evolution of Angiogenic and Inflamed Tumors: The Renal Cancer Paradigm. *Cancer Cell* 38, 771–773. <https://doi.org/10.1016/j.ccell.2020.10.021>.
6. Gao, W., Li, W., Xiao, T., Liu, X.S., and Kaelin, W.G., Jr. (2017). Inactivation of the PBRM1 tumor suppressor gene amplifies the HIF-response in VHL-/- clear cell renal carcinoma. *Proc. Natl. Acad. Sci. USA* 114, 1027–1032. <https://doi.org/10.1073/pnas.1619726114>.
7. Nargund, A.M., Pham, C.G., Dong, Y., Wang, P.I., Osmangeyoglu, H.U., Xie, Y., Aras, O., Han, S., Oyama, T., Takeda, S., et al. (2017). The SWI/SNF Protein PBRM1 Restrains VHL-Loss-Driven Clear Cell Renal Cell Carcinoma. *Cell Rep.* 18, 2893–2906. <https://doi.org/10.1016/j.celrep.2017.02.074>.
8. Simon, J.M., Hacker, K.E., Singh, D., Brannon, A.R., Parker, J.S., Weiser, M., Ho, T.H., Kuan, P.F., Jonasch, E., Furey, T.S., et al. (2014). Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. *Genome Res.* 24, 241–250. <https://doi.org/10.1101/gr.158253.113>.
9. Wu, Y., Terekhanova, N.V., Caravan, W., Naser Al Deen, N., Lal, P., Chen, S., Mo, C.K., Cao, S., Li, Y., Karpova, A., et al. (2023). Epigenetic and transcriptomic characterization reveals progression markers and essential pathways in clear cell renal cell carcinoma. *Nat. Commun.* 14, 1681. <https://doi.org/10.1038/s41467-023-37211-7>.
10. Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Chambers, T., Lopez, J.I., Nicol, D., O'Brien, T., Larkin, J., Horswell, S., et al. (2018). Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell* 173, 581–594.e12. <https://doi.org/10.1016/j.cell.2018.03.057>.
11. Chen, Y.W., Wang, L., Panian, J., Dhanji, S., Derweesh, I., Rose, B., Bagrodia, A., and McKay, R.R. (2023). Treatment Landscape of Renal Cell Carcinoma. *Curr. Treat. Options Oncol.* 24, 1889–1916. <https://doi.org/10.1007/s11864-023-01161-5>.
12. Marusyk, A., and Polyak, K. (2010). Tumor heterogeneity: causes and consequences. *Biochim. Biophys. Acta* 1805, 105–117. <https://doi.org/10.1016/j.bbcan.2009.11.002>.
13. McGranahan, N., and Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* 168, 613–628. <https://doi.org/10.1016/j.cell.2017.01.018>.
14. Easwaran, H., Tsai, H.C., and Baylin, S.B. (2014). Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Mol. Cell* 54, 716–727. <https://doi.org/10.1016/j.molcel.2014.05.015>.
15. Alizadeh, A.A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., Caldas, C., Califano, A., Doherty, M., Elsner, M., et al. (2015). Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* 21, 846–853. <https://doi.org/10.1038/nm.3915>.
16. Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Horswell, S., Chambers, T., O'Brien, T., Lopez, J.I., Watkins, T.B.K., Nicol, D., et al. (2018). Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell* 173, 595–610.e11. <https://doi.org/10.1016/j.cell.2018.03.043>.
17. Cai, Q., Christie, A., Rajaram, S., Zhou, Q., Araj, E., Chintalapati, S., Cadeddu, J., Margulis, V., Pedrosa, I., Rakheja, D., et al. (2020). Ontological analyses reveal clinically-significant clear cell renal cell carcinoma subtypes with convergent evolutionary trajectories into an aggressive type. *EBioMedicine* 51, 102526. <https://doi.org/10.1016/j.ebiom.2019.10.052>.
18. Barkley, D., Moncada, R., Pour, M., Liberman, D.A., Dryg, I., Werba, G., Wang, W., Baron, M., Rao, A., Xia, B., et al. (2022). Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat. Genet.* 54, 1192–1201. <https://doi.org/10.1038/s41588-022-01141-9>.
19. Muto, Y., Wilson, P.C., Ledru, N., Wu, H., Dimke, H., Waikar, S.S., and Humphreys, B.D. (2021). Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. *Nat. Commun.* 12, 2190. <https://doi.org/10.1038/s41467-021-22368-w>.
20. Young, M.D., Mitchell, T.J., Vieira Braga, F.A., Tran, M.G.B., Stewart, B.J., Ferdinand, J.R., Collord, G., Botting, R.A., Popescu, D.M., Loudon, K.W.,

- et al. (2018). Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* 361, 594–599. <https://doi.org/10.1126/science.aat1699>.
21. Liao, J., Yu, Z., Chen, Y., Bao, M., Zou, C., Zhang, H., Liu, D., Li, T., Zhang, Q., Li, J., et al. (2020). Single-cell RNA sequencing of human kidney. *Sci. Data* 7, 4. <https://doi.org/10.1038/s41597-019-0351-8>.
 22. Franzén, O., Gan, L.-M., and Björkregren, J.L.M. (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database* 2019, baz046. <https://doi.org/10.1093/database/baz046>.
 23. Gao, T., Soldatov, R., Sarkar, H., Kurkiewicz, A., Biederstedt, E., Loh, P.R., and Kharchenko, P.V. (2023). Haplotype-aware analysis of somatic copy number variations from single-cell transcriptomes. *Nat. Biotechnol.* 41, 417–426. <https://doi.org/10.1038/s41587-022-01468-y>.
 24. Quddus, M.B., Pratt, N., and Nabi, G. (2019). Chromosomal aberrations in renal cell carcinoma: An overview with implications for clinical practice. *Urol. Ann.* 11, 6–14. https://doi.org/10.4103/UA.UA.32_18.
 25. Yap, C., Mieremet, A., de Vries, C.J.M., Micha, D., and de Waard, V. (2021). Six Shades of Vascular Smooth Muscle Cells Illuminated by KLF4 (Krüppel-Like Factor 4). *Arterioscler. Thromb. Vasc. Biol.* 41, 2693–2707. <https://doi.org/10.1161/ATVBAHA.121.316600>.
 26. Abedini, A., Levinsohn, J., Klötzer, K.A., Dumoulin, B., Ma, Z., Frederick, J., Dhillon, P., Balzer, M.S., Shrestha, R., Liu, H., et al. (2024). Spatially resolved human kidney multi-omics single cell atlas highlights the key role of the fibrotic microenvironment in kidney disease progression. Preprint at bioRxiv. <https://doi.org/10.1101/2022.10.24.513598>.
 27. Tostain, J., Li, G., Gentil-Perret, A., and Gigante, M. (2010). Carbonic anhydrase 9 in clear cell renal cell carcinoma: a marker for diagnosis, prognosis and treatment. *Eur. J. Cancer* 46, 3141–3148. <https://doi.org/10.1016/j.ejca.2010.07.020>.
 28. Lindgren, D., Eriksson, P., Krawczyk, K., Nilsson, H., Hansson, J., Veerla, S., Sjölund, J., Höglund, M., Johansson, M.E., and Axelson, H. (2017). Cell-Type-Specific Gene Programs of the Normal Human Nephron Define Kidney Cancer Subtypes. *Cell Rep.* 20, 1476–1489. <https://doi.org/10.1016/j.celrep.2017.07.043>.
 29. Gewin, L.S. (2021). Sugar or Fat? Renal Tubular Metabolism Reviewed in Health and Disease. *Nutrients* 13, 1580. <https://doi.org/10.3390/nu13051580>.
 30. Qin, Q., Fan, J., Zheng, R., Wan, C., Mei, S., Wu, Q., Sun, H., Brown, M., Zhang, J., Meyer, C.A., and Liu, X.S. (2020). Lisa: inferring transcriptional regulators through integrative modeling of public chromatin accessibility and ChIP-seq data. *Genome Biol.* 21, 32. <https://doi.org/10.1186/s13059-020-1934-6>.
 31. Moriguchi, T., Hamada, M., Morito, N., Terunuma, T., Hasegawa, K., Zhang, C., Yokomizo, T., Esaki, R., Kuroda, E., Yoh, K., et al. (2006). MafB is essential for renal development and F4/80 expression in macrophages. *Mol. Cell Biol.* 26, 5715–5727. <https://doi.org/10.1128/MCB.00001-06>.
 32. Sel, S., Ebert, T., Ryyfel, G.U., and Drewes, T. (1996). Human renal cell carcinogenesis is accompanied by a coordinate loss of the tissue specific transcription factors HNF4 alpha and HNF1 alpha. *Cancer Lett.* 101, 205–210. [https://doi.org/10.1016/0304-3835\(96\)04136-5](https://doi.org/10.1016/0304-3835(96)04136-5).
 33. Gao, Y., Yan, Y., Guo, J., Zhang, Q., Bi, D., Wang, F., Chang, Z., Lu, L., Yao, X., and Wei, Q. (2019). HNF-4alpha downregulation promotes tumor migration and invasion by regulating E-cadherin in renal cell carcinoma. *Oncol. Rep.* 42, 1066–1074. <https://doi.org/10.3892/or.2019.7214>.
 34. Yoshimura, Y., Muto, Y., Omachi, K., Miner, J.H., and Humphreys, B.D. (2023). Elucidating the Proximal Tubule HNF4A Gene Regulatory Network in Human Kidney Organoids. *J. Am. Soc. Nephrol.* 34, 1672–1686. <https://doi.org/10.1681/ASN.000000000000197>.
 35. Marable, S.S., Chung, E., and Park, J.S. (2020). Hnf4a Is Required for the Development of Cdh6-Expressing Progenitors into Proximal Tubules in the Mouse Kidney. *J. Am. Soc. Nephrol.* 31, 2543–2558. <https://doi.org/10.1681/ASN.2020020184>.
 36. Marable, S.S., Chung, E., Adam, M., Potter, S.S., and Park, J.S. (2018). Hnf4a deletion in the mouse kidney phenocopies Fanconi renotubular syndrome. *JCI Insight* 3, e97497. <https://doi.org/10.1172/jci.insight.97497>.
 37. Wang, J., Zuo, Z., Yu, Z., Chen, Z., Meng, X., Ma, Z., Niu, J., Guo, R., Tran, L.J., Zhang, J., et al. (2023). Single-cell transcriptome analysis revealing the intratumoral heterogeneity of ccRCC and validation of MT2A in pathogenesis. *Funct. Integr. Genomics* 23, 300. <https://doi.org/10.1007/s10142-023-01225-7>.
 38. Chen, L., Luo, S., Dupre, A., Vasoya, R.P., Parthasarathy, A., Aita, R., Malhotra, R., Hur, J., Toke, N.H., Chiles, E., et al. (2021). The nuclear receptor HNF4 drives a brush border gene program conserved across murine intestine, kidney, and embryonic yolk sac. *Nat. Commun.* 12, 2886. <https://doi.org/10.1038/s41467-021-22761-5>.
 39. Zhang, J., Lu, T., Lu, S., Ma, S., Han, D., Zhang, K., Xu, C., Liu, S., Gan, L., Wu, X., et al. (2023). Single-cell analysis of multiple cancer types reveals differences in endothelial cells between tumors and normal tissues. *Comput. Struct. Biotechnol. J.* 21, 665–676. <https://doi.org/10.1016/j.csbj.2022.12.049>.
 40. Yamada, Y., Zhong, Y., Miki, S., Taura, A., and Rabbitts, T.H. (2022). The transcription factor complex LMO2/TAL1 regulates branching and endothelial cell migration in sprouting angiogenesis. *Sci. Rep.* 12, 7226. <https://doi.org/10.1038/s41598-022-11297-3>.
 41. Payne, S., Neal, A., and De Val, S. (2024). Transcription factors regulating vasculogenesis and angiogenesis. *Dev. Dynam.* 253, 28–58. <https://doi.org/10.1002/dvdy.575>.
 42. Trimm, E., and Red-Horse, K. (2023). Vascular endothelial cell development and diversity. *Nat. Rev. Cardiol.* 20, 197–210. <https://doi.org/10.1038/s41569-022-00770-1>.
 43. Tello, D., Balsa, E., Acosta-Iborra, B., Fuertes-Yebra, E., Elorza, A., Ordóñez, Á., Corral-Escariz, M., Soro, I., López-Bernardo, E., Perales-Clemente, E., et al. (2011). Induction of the mitochondrial NDUFA4L2 protein by HIF-1alpha decreases oxygen consumption by inhibiting Complex I activity. *Cell Metabol.* 14, 768–779. <https://doi.org/10.1016/j.cmet.2011.10.008>.
 44. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* 15, 1484–1506. <https://doi.org/10.1038/s41596-020-0292-x>.
 45. Gerlinger, M., Horswell, S., Larkin, J., Rowan, A.J., Salm, M.P., Varela, I., Fisher, R., McGranahan, N., Matthews, N., Santos, C.R., et al. (2014). Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* 46, 225–233. <https://doi.org/10.1038/ng.2891>.
 46. Gerlinger, M., Rowan, A.J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* 366, 883–892. <https://doi.org/10.1056/NEJMoa1113205>.
 47. Mitchell, T.J., Turajlic, S., Rowan, A., Nicol, D., Farmery, J.H.R., O'Brien, T., Martincorena, I., Tarpey, P., Angelopoulos, N., Yates, L.R., et al. (2018). Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. *Cell* 173, 611–623.e17. <https://doi.org/10.1016/j.cell.2018.02.020>.
 48. Raval, R.R., Lau, K.W., Tran, M.G.B., Sowter, H.M., Mandriota, S.J., Li, J.L., Pugh, C.W., Maxwell, P.H., Harris, A.L., and Ratcliffe, P.J. (2005). Contrasting Properties of Hypoxia-Inducible Factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-Associated Renal Cell Carcinoma. *Mol. Cell Biol.* 25, 5675–5686. <https://doi.org/10.1128/MCB.25.13.5675-5686.2005>.
 49. Biswas, S., Troy, H., Leek, R., Chung, Y.L., Li, J.L., Raval, R.R., Turley, H., Gatter, K., Pezzella, F., Griffiths, J.R., et al. (2010). Effects of HIF-1alpha and HIF2alpha on Growth and Metabolism of Clear-Cell Renal Cell Carcinoma 786-O Xenografts. *JAMA Oncol.* 2010, 757908. <https://doi.org/10.1155/2010/757908>.

50. Monzon, F.A., Alvarez, K., Peterson, L., Truong, L., Amato, R.J., Hernandez-McClain, J., Tannir, N., Parwani, A.V., and Jonasch, E. (2011). Chromosome 14q loss defines a molecular subtype of clear-cell renal cell carcinoma associated with poor prognosis. *Mod. Pathol.* **24**, 1470–1479. <https://doi.org/10.1038/modpathol.2011.107>.
51. Shen, C., Beroukhi, R., Schumacher, S.E., Zhou, J., Chang, M., Signoretto, S., and Kaelin, W.G., Jr. (2011). Genetic and Functional Studies Implicate HIF1alpha as a 14q Kidney Cancer Suppressor Gene. *Cancer Discov.* **1**, 222–235. <https://doi.org/10.1158/2159-8290.CD-11-0098>.
52. Salama, R., Masson, N., Simpson, P., Sciesielski, L.K., Sun, M., Tian, Y.M., Ratcliffe, P.J., and Mole, D.R. (2015). Heterogeneous Effects of Direct Hypoxia Pathway Activation in Kidney Cancer. *PLoS One* **10**, e0134645. <https://doi.org/10.1371/journal.pone.0134645>.
53. Alchahin, A.M., Mei, S., Tsea, I., Hirz, T., Kfoury, Y., Dahl, D., Wu, C.L., Subtelny, A.O., Wu, S., Scadden, D.T., et al. (2022). A transcriptional metastatic signature predicts survival in clear cell renal cell carcinoma. *Nat. Commun.* **13**, 5747. <https://doi.org/10.1038/s41467-022-33375-w>.
54. Li, L., Shen, C., Nakamura, E., Ando, K., Signoretti, S., Beroukhi, R., Cowley, G.S., Lizotte, P., Liberzon, E., Bair, S., et al. (2013). SQSTM1 is a pathogenic target of 5q copy number gains in kidney cancer. *Cancer Cell* **24**, 738–750. <https://doi.org/10.1016/j.ccr.2013.10.025>.
55. Jacobs, J., Deschoolmeester, V., Zwaenepoel, K., Rolfo, C., Silence, K., Rottey, S., Lardon, F., Smits, E., and Pauwels, P. (2015). CD70: An emerging target in cancer immunotherapy. *Pharmacol. Ther.* **155**, 1–10. <https://doi.org/10.1016/j.pharmthera.2015.07.007>.
56. Bezwada, D., Lesner, N.P., Brooks, B., Vu, H.S., Wu, Z., Cai, L., Kasinton, S., Kelekar, S., Cai, F., Aurora, A.B., et al. (2023). Mitochondrial metabolism in primary and metastatic human kidney cancers. Preprint at bioRxiv. <https://doi.org/10.1101/2023.02.06.527285>.
57. Tang, S.W., Chang, W.H., Su, Y.C., Chen, Y.C., Lai, Y.H., Wu, P.T., Hsu, C.I., Lin, W.C., Lai, M.K., and Lin, J.Y. (2009). MYC pathway is activated in clear cell renal cell carcinoma and essential for proliferation of clear cell renal cell carcinoma cells. *Cancer Lett.* **273**, 35–43. <https://doi.org/10.1016/j.canlet.2008.07.038>.
58. Grampp, S., Platt, J.L., Lauer, V., Salama, R., Kranz, F., Neumann, V.K., Wach, S., Stöhr, C., Hartmann, A., Eckardt, K.U., et al. (2016). Genetic variation at the 8q24.21 renal cancer susceptibility locus affects HIF binding to a MYC enhancer. *Nat. Commun.* **7**, 13183. <https://doi.org/10.1038/ncomms13183>.
59. Gordan, J.D., Lal, P., Dondeti, V.R., Letrero, R., Parekh, K.N., Oquendo, C.E., Greenberg, R.A., Flaherty, K.T., Rathmell, W.K., Keith, B., et al. (2008). HIF-alpha effects on c-Myc distinguish two subtypes of sporadic VHL-deficient clear cell renal carcinoma. *Cancer Cell* **14**, 435–446. <https://doi.org/10.1016/j.ccr.2008.10.016>.
60. Li, R., Ferdinand, J.R., Loudon, K.W., Bowyer, G.S., Laidlaw, S., Muiyasa, F., Mamanova, L., Neves, J.B., Bolt, L., Fasouli, E.S., et al. (2022). Mapping single-cell transcriptomes in the intra-tumoral and associated territories of kidney cancer. *Cancer Cell* **40**, 1583–1599.e10. <https://doi.org/10.1016/j.ccell.2022.11.001>.
61. Abedini, A., Zhu, Y.O., Chatterjee, S., Halasz, G., Devalaraja-Narashimha, K., Shrestha, R., S Balzer, M., Park, J., Zhou, T., Ma, Z., et al. (2021). Urinary Single-Cell Profiling Captures the Cellular Diversity of the Kidney. *J. Am. Soc. Nephrol.* **32**, 614–627. <https://doi.org/10.1681/ASN.2020050757>.
62. Ide, S., Kobayashi, Y., Ide, K., Strausser, S.A., Abe, K., Herbek, S., O'Brien, L.L., Crowley, S.D., Barisoni, L., Tata, A., et al. (2021). Ferroptotic stress promotes the accumulation of pro-inflammatory proximal tubular cells in maladaptive renal repair. *Elife* **10**, e68603. <https://doi.org/10.7554/eLife.68603>.
63. Patel, S.A., Hirose, S., Rodrigues, P., Vojtasova, E., Richardson, E.K., Ge, J., Syafruddin, S.E., Speed, A., Papachristou, E.K., Baker, D., et al. (2022). The renal lineage factor PAX8 controls oncogenic signalling in kidney cancer. *Nature* **606**, 999–1006. <https://doi.org/10.1038/s41586-022-04809-8>.
64. Aibar, S., González-Bias, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086. <https://doi.org/10.1038/nmeth.4463>.
65. Wu, H., Lai, C.F., Chang-Panesso, M., and Humphreys, B.D. (2020). Proximal Tubule Translational Profiling during Kidney Fibrosis Reveals Proinflammatory and Long Noncoding RNA Expression Patterns with Sexual Dimorphism. *J. Am. Soc. Nephrol.* **31**, 23–38. <https://doi.org/10.1681/ASN.2019040337>.
66. Gerhardt, L.M.S., Liu, J., Koppitch, K., Cippà, P.E., and McMahon, A.P. (2021). Single-nuclear transcriptomics reveals diversity of proximal tubule cell states in a dynamic response to acute kidney injury. *Proc. Natl. Acad. Sci. USA* **118**, e2026684118. <https://doi.org/10.1073/pnas.2026684118>.
67. Yamaguchi, J., Tanaka, T., Eto, N., and Nangaku, M. (2015). Inflammation and hypoxia linked to renal injury by CCAAT/enhancer-binding protein delta. *Kidney Int.* **88**, 262–275. <https://doi.org/10.1038/ki.2015.21>.
68. Smythies, J.A., Sun, M., Masson, N., Salama, R., Simpson, P.D., Murray, E., Neumann, V., Cockman, M.E., Choudhry, H., Ratcliffe, P.J., and Mole, D.R. (2019). Inherent DNA-binding specificities of the HIF-1alpha and HIF-2alpha transcription factors in chromatin. *EMBO Rep.* **20**, e46401. <https://doi.org/10.15252/embr.201846401>.
69. Kondo, K., Kim, W.Y., Lechpammer, M., and Kaelin, W.G., Jr. (2003). Inhibition of HIF2alpha is sufficient to suppress pVHL-defective tumor growth. *PLoS Biol.* **1**, E83. <https://doi.org/10.1371/journal.pbio.0000083>.
70. Kurler, S., Lima, J.D.C.C., Li, R., Lombardi, O., Masson, N., Barros, A.B., Pontecorvi, V., Mole, D.R., Pugh, C.W., Adam, J., and Ratcliffe, P.J. (2024). Oncogenic cell tagging and single-cell transcriptomics reveal cell type-specific and time-resolved responses to Vhl inactivation in the kidney. *Cancer Res.* **84**, 1799–1816. <https://doi.org/10.1158/0008-5472.CAN-23-3248>.
71. Perez-Gonzalez, A., Bevant, K., and Blanpain, C. (2023). Cancer cell plasticity during tumor progression, metastasis and response to therapy. *Nat. Can. (Ott.)* **4**, 1063–1082. <https://doi.org/10.1038/s43018-023-00595-y>.
72. Stemmler, M.P., Eccles, R.L., Brabletz, S., and Brabletz, T. (2019). Non-redundant functions of EMT transcription factors. *Nat. Cell Biol.* **21**, 102–112. <https://doi.org/10.1038/s41556-018-0196-y>.
73. Yang, J., Mani, S.A., Donaher, J.L., Ramaswamy, S., Itzykson, R.A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R.A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927–939. <https://doi.org/10.1016/j.cell.2004.06.006>.
74. Gaude, E., and Frezza, C. (2016). Tissue-specific and convergent metabolic transformation of cancer correlates with metastatic potential and patient survival. *Nat. Commun.* **7**, 13041. <https://doi.org/10.1038/ncomms13041>.
75. Kim, K., Zhou, Q., Christie, A., Stevens, C., Ma, Y., Onabolu, O., Chintalapati, S., McKenzie, T., Tcheuyap, V.T., Woolford, L., et al. (2021). Determinants of renal cell carcinoma invasion and metastatic competence. *Nat. Commun.* **12**, 5760. <https://doi.org/10.1038/s41467-021-25918-4>.
76. Marine, J.C., Dawson, S.J., and Dawson, M.A. (2020). Non-genetic mechanisms of therapeutic resistance in cancer. *Nat. Rev. Cancer* **20**, 743–756. <https://doi.org/10.1038/s41568-020-00302-4>.
77. Shen, S., Vagner, S., and Robert, C. (2020). Persistent Cancer Cells: The Deadly Survivors. *Cell* **183**, 860–874. <https://doi.org/10.1016/j.cell.2020.10.027>.
78. Lombardi, O., Li, R., Halim, S., Choudhry, H., Ratcliffe, P.J., and Mole, D.R. (2022). Pan-cancer analysis of tissue and single-cell HIF-pathway activation using a conserved gene signature. *Cell Rep.* **41**, 111652. <https://doi.org/10.1016/j.celrep.2022.111652>.
79. Stoekius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., 3rd, Smibert, P., and Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224. <https://doi.org/10.1186/s13059-018-1603-1>.
80. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated

- analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
81. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput. Biol.* 12, e1004873. <https://doi.org/10.1371/journal.pcbi.1004873>.
 82. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
 83. Van der Auwera, G.A., and O'Connor, B.D. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O'Reilly Media, Inc.).
 84. Prashant, N.M., Alomran, N., Chen, Y., Liu, H., Bousounis, P., Movassagh, M., Edwards, N., and Horvath, A. (2021). SCReadCounts: estimation of cell-level SNVs expression from scRNA-seq data. *BMC Genom.* 22, 689. <https://doi.org/10.1186/s12864-021-07974-8>.
 85. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
 86. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
 87. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. <https://doi.org/10.1093/bioinformatics/btu638>.
 88. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
 89. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. <https://doi.org/10.1093/bioinformatics/btx364>.
 90. Korotkevich, G., Sukhov, V., and Sergushichev, A. (2019). Fast gene set enrichment analysis. Preprint at bioRxiv 888, 888. <https://doi.org/10.1101/060012>.
 91. Garcia-Alonso, L., Lorenzi, V., Mazzeo, C.I., Alves-Lopes, J.P., Roberts, K., Sancho-Serra, C., Engelbert, J., Marečková, M., Gruhn, W.H., Botting, R.A., et al. (2022). Single-cell roadmap of human gonadal development. *Nature* 607, 540–547. <https://doi.org/10.1038/s41586-022-04918-4>.
 92. Schmid, V., Lafleur, V.N., Lombardi, O., Li, R., Salama, R., Colli, L., Choudhry, H., Chanock, S., Ratcliffe, P.J., and Mole, D.R. (2019). Co-occurrence of RCC-susceptibility polymorphisms with HIF cis-acting sequences supports a pathway tuning model of cancer. *Sci. Rep.* 9, 18768. <https://doi.org/10.1038/s41598-019-55098-7>.
 93. The ENCODE Consortium (2016). ENCODE Guidelines and Best Practices for RNA-Seq: Revised December 2016. https://www.encodeproject.org/documents/cede0cbe-d324-4ce7-ace4-f0c3eddf5972/@download/attachment/ENCODE%20Best%20Practices%20for%20RNA_v2.pdf.
 94. Waltman, L., and van Eck, N.J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* 86, 471. <https://doi.org/10.1140/epjb/e2013-40829-0>.
 95. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* 20, 1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>.
 96. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550. <https://doi.org/10.1073/pnas.0506580102>.
 97. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., and Mesirov, J.P. (2007). GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics* 23, 3251–3253. <https://doi.org/10.1093/bioinformatics/btm369>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Purified mouse monoclonal to HIF-1 α	BD Transduction Laboratories	RRID:AB_398272; Cat# 610959
Purified rabbit monoclonal to HIF-2 α	Cell Signaling	RRID:AB_10898028; Cat# 7096S
Mouse monoclonal antisera to HIF-2 α	In house	190b
Purified rabbit polyclonal to HIF-1 β	Cell Signaling	RRID:AB_10694232; Cat# 5537S
Purified rabbit polyclonal to CA9	Abcam	RRID:AB_2066533; Cat# ab15086
Purified rabbit polyclonal to CA9 (D47G3)	Cell Signaling	RRID:AB_10706355; Cat#5649
Purified rabbit polyclonal to VHL	Cell Signaling	RRID:AB_2716279; Cat# 68547S
Purified mouse monoclonal to β -actin conjugated to HRP	Abcam	RRID:AB_867494; Cat# ab49900
Rabbit Recombinant Monoclonal to Cubilin C-terminal	Abcam	Cat# ab191073
Rabbit monoclonal to Caveolin-1	Abcam	RRID:AB_725987; Cat# ab32577
Anti-ICAM1	Abcam	Cat# ab282575
Rabbit polyclonal to CD31	Abcam	RRID:AB_726362; Cat# ab28364
Rabbit monoclonal to VEGF Receptor 1	Abcam	RRID:AB_778798; Cat# ab32152
TotalSeq-A anti-human Hashtag 1 (GTCAACTCTTTAGCG)	Biologend	RRID:AB_2750015; Cat# 394601
TotalSeq-A anti-human Hashtag 2 (TGATGGCCTATTGGG)	Biologend	RRID:AB_2750016; Cat# 394603
TotalSeq-A anti-human Hashtag 3 (TTCCGCCTCTTTG)	Biologend	RRID:AB_2750017; Cat# 394605
TotalSeq-A anti-human Hashtag 4 (AGTAAGTTCAGCGTA)	Biologend	RRID:AB_2750018; Cat# 394607
TotalSeq-A anti-human Hashtag 5 (AAGTATCGTTTCGCA)	Biologend	RRID:AB_2750019; Cat# 394609
Biological samples		
Renal tumor and background kidney samples	Oxford Radcliffe Biobank (ORB)	Available on request
Chemicals, peptides, and recombinant proteins		
DMEM/F12	Gibco	Cat# 11320033
Advanced DMEM/F12	Gibco	Cat# 12634010
Dulbecco's Modified Eagle's Medium (DMEM)	Sigma Aldrich	Cat# D6429
Glutamax	Gibco	Cat# 35050061
Anti-anti	Gibco	Cat# 15240062
Insulin-transferrin-sodium selenite	Gibco	Cat# 41400045
Triiodo-L-thyronine	Thermo Fisher	Cat# H34068.MD
Epidermal growth factor	Gibco	Cat# PHG0314
Hydrocortisone	Sigma-Aldrich	Cat# H0888
Hanks balanced salt solution (HBSS)	Gibco	Cat# 24020091
Fetal bovine serum	Sigma-Aldrich	Cat# F7524-500ML
Penicillin-streptomycin	Gibco	Cat# 15140122
Collagenase II	Thermo Fisher	Cat# 17101015
DNase I	Sigma-Aldrich	Cat# 11284932001

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Trypsin-EDTA	Sigma-Aldrich	Cat# T4049
Dulbecco's PBS (DPBS)	Gibco	Cat# 14190144
Bovine Serum Albumin (BSA) Fraction V	Sigma-Aldrich	Cat# 10735086001
DMSO	Sigma-Aldrich	Cat# D2650
PT2385 (Belzutifan)	MedChem Express	Cat# HY-12867
Trypan blue	Gibco	Cat# 15250061
Acridine Orange/Propidium Iodide	Logos Biosystems	Cat# F23001
HIF-1a/HIF1A siRNA	siTOOLS Biotech	Cat# si-K005-3091-HIF1A (Human)
HIF-2a/EPAS1 siRNA	siTOOLS Biotech	Cat# si-G050-2034-EPAS1 (Human)
Control siRNA	siTOOLS Biotech	Part of on-target siRNA pack
RNAiMAX transfection reagent	ThermoFisher Scientific	Cat# 13778075
cOmplete protease inhibitor cocktail	Roche	Cat# 11836145001

Critical commercial assays

MACS Tumor Dissociation Kit (human)	Miltenyi Biotec	Cat# 130-095-929
Nimblegen SeqCap EZ Exome v3	Roche	Discontinued
Twist human core exome	Twist Bioscience	Cat #102027
RNeasy Plus Mini Kit	Qiagen	Cat# 74134
RNase-free DNase Set	Qiagen	Cat# 79254
NEBNext Ultra II Directional RNA Library Prep Kit for Illumina	NEB	Cat# E7765 S/L
DNeasy Blood & Tissue Kit	Qiagen	Cat# 69504
RNase A	Qiagen	Cat# 19101
Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1	10x Genomics	Cat# PN-1000121
Chromium Next GEM Chip G Single Cell Kit	10x Genomics	Cat# PN-1000120
Dako Target Retrieval Solution, pH 6	Agilent	Cat# S2369
Dako Peroxidase Blocking solution	Agilent	Cat# S2023
Bovine serum albumin (BSA)	Sigma	Cat# 5482
Dako Envision system	Agilent	Cat# K4003
modified Harris Haematoxylin	Thermo Fisher Scientific	Cat# 72711
DPX mountant	Merck	Cat# 06522

Deposited data

Single cell RNA-seq analysis of gene expression in fresh and cultured normal and tumor samples	This paper	GEO accession: GSE269819
Bulk RNA-seq analysis of RCC4 cells treated with siRNAs targeting HIF-1alpha and/or HIF-2alpha	This paper	GEO accession: GSE269826
Single cell RNA-seq analysis of gene expression in normoxic/hypoxic primary normal kidney cultures and normoxic ccRCC tumor cultures	Lombardi et al. ⁷⁸	GEO accession: GSE200207
Original Western blot images	This paper	Mendeley: https://doi.org/10.17632/7p25n8gwj2.1

Experimental models: Cell lines

RCC4	Gift from C.H. Buys; validated by detection of the VHL gene mutation (chr3:10,183,841 G > del) in RNA-seq data	RRID: CVCL_0498
------	--	-----------------

(Continued on next page)

<i>Continued</i>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Software and algorithms</i>		
Cell Ranger analysis pipeline	10X Genomics	https://www.10xgenomics.com/support/software/cell-ranger/latest
Scrublet	Marine et al. ⁷⁶	https://github.com/swolock/scrublet
HTODemux	Stoeckius et al. ⁷⁹	https://satijalab.org/seurat/articles/hashing_vignette.html
Seurat (4.0.3)	Hao et al. ⁸⁰	https://satijalab.org/seurat/
R (4.0.5)	https://www.r-project.org/foundation/	https://www.r-project.org/foundation/
CellMixS (1.18.0)	https://bioconductor.org/packages/3.18/bioc/html/CellMixS.html	https://bioconductor.org/packages/3.18/bioc/html/CellMixS.html
InferCNV (1.6.0)	InferCNV of the Trinity CTAT Project	https://github.com/broadinstitute/infercnv
Numbat (1.3.2–1)	Gao et al. ²³	https://github.com/kharchenkolab/numbat
CNVkit (0.9.8)	Talevich et al. ⁸¹	https://cnvkit.readthedocs.io/en/stable/
BWA (0.7.15)	Li et al. ⁸²	https://github.com/lh3/bwa
Picard tools (2.18.20)	http://broadinstitute.github.io/picard/	http://broadinstitute.github.io/picard/
GATK (4.1.9)	Van der Auwera et al. ⁸³	https://gatk.broadinstitute.org/hc/en-us
gnomAD (4.0)	https://gnomad.broadinstitute.org/downloads	https://gnomad.broadinstitute.org/downloads
SCReadCounts	Prashant et al. ⁸⁴	https://horvathlab.github.io/NGS/SCReadCounts/
SAMtools (0.1.19)	Li et al. ⁸⁵	http://www.htslib.org/doc/samtools.html
TrimGalore (0.3.3)	https://github.com/FelixKrueger/TrimGalore	https://github.com/FelixKrueger/TrimGalore
HISAT2 (2.05)	Kim et al. ⁸⁶	http://daehwankimlab.github.io/hisat2/
HTSeq (0.5.4p3)	Anders et al. ⁸⁷	https://htseq.readthedocs.io/en/latest/
DESeq2	Love et al. ⁸⁸	https://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html
UpSetR (1.4.0)	Conway et al. ⁸⁹	https://github.com/hms-dbmi/UpSetR
fgsea	Korotkevich et al. ⁹⁰	https://github.com/ctlab/fgsea
MsigDB	https://www.gsea-msigdb.org/gsea/msigdb/	https://www.gsea-msigdb.org/gsea/msigdb/
LISA (2.3.0)	Qin et al. ³⁰	N/A
SCENIC (1.3.1)	Aibar et al. ⁶⁴	https://scenic.aertslab.org
CellphoneDB (4.1.0)	Garcia-Alonso et al. ⁹¹	https://github.com/Teichlab/cellphonedb
Ktplots (2.3.0)	https://github.com/zktuong/ktplots	https://github.com/zktuong/ktplots
<i>Other</i>		
<i>In Vivo</i> 2 400 Hypoxia Workstation	Ruskinn Technology	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The study was conducted in accordance with the Declaration of Helsinki, and protocols were approved by the Ethics Committee of Oxford University Hospitals NHS Foundation Trust under Oxford Center for Histopathology Research (OCHRe) application numbers 15/A233, 17/A145, 20/A106 and 20/A106b. Specimens were radical or partial nephrectomies collected from individuals undergoing surgery for known or suspected ccRCC (10 males and 5 females – age, ancestry, race and ethnicity not recorded). The RCC4 cell line was authenticated by STR genotyping and exon-sequencing of the VHL mutation and tested for mycoplasma contamination.

METHOD DETAILS

Tissue sampling

The study was conducted in accordance with the Declaration of Helsinki, and protocols were approved by the Ethics Committee of Oxford University Hospitals NHS Foundation Trust under Oxford Center for Histopathology Research (OCHRe) application numbers

15/A233, 17/A145, 20/A106 and 20/A106b. Specimens were radical or partial nephrectomies collected from individuals undergoing surgery for known or suspected ccRCC.

One or more punch biopsies (4–8 mm) were taken from fresh tumor and background macroscopically normal renal parenchyma as soon as feasible following surgical removal of the kidney and prior to formalin-fixation. Sampled punched out areas of tumor were marked using different colored inks and photographed as a record of sample locations. Where possible samples were taken from macroscopically distinct areas across the tumor, avoiding areas of necrosis. Sampling of tumors did not impact on the clinical diagnostic process.

Samples were annotated by tumor (T)/normal (N)/cystic (C), patient number and, if multiple regions were sampled, by an additional letter (except those from patient 18 who had two separate tumors from each of which one region was taken - annotated T18A and T18B). Biopsies were immediately transferred to storage medium (Advanced DMEM/F12 1:1, 1X Glutamax, 1X anti-anti, 1X insulin-transferrin-sodium selenite, 4 ng/mL triiodo-L-thyronine, 100 ng/mL epidermal growth factor, 36 ng/mL hydrocortisone and 20% fetal bovine serum) and put on ice. Biopsies were either processed immediately or kept in the fridge overnight and processed the next morning.

Tissue processing

Biopsies were washed twice in cold Hank's Balanced Salt Solution (HBSS) and minced into $\sim 1\text{mm}^3$ cubes. These were incubated, with shaking, for 1 h at 37°C in 7.5 mL HBSS containing enzymes from the MACS Tumor Dissociation Kit (37.5 μL enzyme A, 300 μL enzyme H and 70 μL enzyme R) together with collagenase II (193 U/ml) and DNase I (3.33 $\mu\text{g/ml}$). Cells were pelleted by centrifugation at 300g for 5 min at 4°C and resuspended in Trypsin-EDTA and incubated in a water bath for 3 min at 37°C. Trypsin was then inactivated by adding 10 mL of cold growth medium (Advanced DMEM/F12 1:1, 1X Glutamax, 1X anti-anti, 1X insulin-transferrin-sodium selenite, 4 ng/mL triiodo-L-thyronine, 100 ng/mL epidermal growth factor, 36 ng/mL hydrocortisone and 10% fetal bovine serum). Incompletely dissociated material was allowed to settle, and the suspension was passed sequentially through 100 μm , 70 μm and 40 μm cell strainers. Strained suspensions were repelleted at 300g for 5 min and then incubated in MACS red blood cell lysis buffer for 4 min at room temperature. Ice-cold growth medium was then added and cells repelleted at 300g for 5 min at 4°C. Cell pellets were washed once in 15 mL ice-cold Dulbecco's PBS (DPBS), containing 0.04% bovine serum albumin (BSA), repelleted and finally resuspended in 100 μL –500 μL of DPBS containing 0.04% BSA to achieve a final concentration of 1000–2000 cells/ μL .

Primary cultures

Tumor biopsies from Pt3, Pt17 and Pt20 were processed as previously described⁷⁸ to generate cell suspensions for primary cultures. Incompletely dissociated material from the normal kidney biopsy from Pt18 was also plated to derive a culture (N18). Non-adhered cells from all T17 tumor cultures (T17A–G) were pooled to create a 'T17 mix' culture, which was used for the DMSO and PT2385 (Belzutifan) treatments. Hypoxic incubations were performed using an *In Vivo* 2 400 Hypoxia Workstation (Ruskin Technology) in an atmosphere containing either 21% (normoxia), 5% ('physoxia') or 0.5% (hypoxia) oxygen for 16 h. Primary cultures were maintained in growth medium (Advanced DMEM/F12 1:1, 1X Glutamax, 1X anti-anti, 1X insulin-transferrin-sodium selenite, 4 ng/mL triiodo-L-thyronine, 100 ng/mL epidermal growth factor, 36 ng/mL hydrocortisone and 10% fetal bovine serum) and kept in physoxia (except Pt3 cultures from previously published experiments,^{78,92} which were maintained in normoxia). PT2385 (Belzutifan, HY-12867, MedChem Express) was used at a final concentration of 1 μM for 16 h, or 1:1000 DMSO as a vehicle control. Cultures were analyzed at early passage (passages 1–4).

Cell hashing for multiplexing primary cultures

Primary cell cultures were detached, made into single cell suspensions, and labeled with TotalSeq-A 'hashtag' antibodies as described previously.⁷⁸ Antibodies and associated hashtag sequences were as follows: anti-human 1 (GTCAACTCTTAGCG), anti-human 2 (TGATGGCCTATTGGG), anti-human 3 (TTCCGCCTCTCTTTG), anti-human 4 (AGTAAGTTCAGCGTA) and anti-human hashtag 5 (AAGTATCGTTTCGCA) (all Biolegend).

10x Genomics GEM generation and library preparation

Cell suspensions were counted (Bio-Rad TC20, Countess II, or LUNA-FL) using Trypan blue or Acridine Orange/Propidium Iodide to assess viability. Approximately 20,000 cells were loaded per channel on the 10x Genomics chip. Single cell 3' GEX RNA (version 3.1) libraries were prepared according to manufacturer's instructions (10x Genomics) with single indexing (Pt17, Pt18, Pt20 and Pt22) or dual indexing (Pt23, Pt24, Pt25, Pt26, Pt29, Pt30, Pt31, Pt32, Pt33 and Pt35).

Whole exome sequencing

Genomic DNA was prepared using the DNeasy Blood and Tissue kit (Qiagen) and treated with RNase (Qiagen) according to manufacturer's instructions. DNA integrity was assessed using Genomic DNA Screentape and Reagents on a TapeStation as part of quality control. Whole exome sequencing samples were prepared using the Nimblegen capture kit (Pt3) or the TWIST capture kit (Pt17, Pt18 and Pt20) according to manufacturer instructions.

siRNA transfection and bulk RNA-sequencing

RCC4 cells (gift from C.H. Buys) were cultured in high glucose Dulbecco's Modified Eagle's Medium (DMEM) (Sigma-Aldrich, D6429) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich F7524) and 1% 100X penicillin-streptomycin antibiotics (Gibco 15140122). Cells were reverse transfected with HIF-1 α (HIF1A), HIF-2 α (EPAS1), and control siPOOLS (siTOOLS Biotech) using Lipofectamine RNAiMAX Transfection Reagent (ThermoFisher Scientific, 13778075) according to manufacturer's instructions. For transfection, 2×10^5 cells were seeded on each 6cm dish in 3.5mL of DMEM (Sigma-Aldrich, D6429) containing 10% FBS (without antibiotic) and the final siRNA concentration for each condition was 2nM (control = 2nM controlSi; HIF-1 α knockdown = 1nM HIF1Asi + 1nM controlSi; HIF-2 α knockdown = 1nM HIF2Asi + 1nM controlSi; double knockdown = 1nM HIF1Asi + 1nM HIF2Asi; although the latter condition was not analyzed for the purpose of this study). After 24h, medium was replaced to fresh antibiotic-free medium. After a further 24h (48h-post transfection), cells were harvested for RNA-seq. RNA-seq was performed as previously described.⁷⁸ Briefly, Total RNA was prepared using the RNeasy Plus Mini kit (Qiagen) and treated with RNase-free DNase Set (Qiagen) according to manufacturer's instructions. RNA integrity was assessed using RNA Reagents and RNA Screentapes on a TapeStation as part of quality control. PolyA + RNA libraries were then prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina according to manufacturer's instructions. All RNA-seq experiments were performed in triplicate in accordance with ENCODE consortium guidelines.⁹³

Sequencing

Libraries were sequenced on the Illumina NovaSeq 6000 or NextSeq 2000 platforms, according to manufacturer instructions. Paired end sequencing was conducted using the following read configurations: Read 1 = 28bp and Read 2 = 98bp (Pt3, Pt17, Pt18, Pt22, Pt23, Pt24, Pt25, Pt26); Read 1 = 151bp and Read 2 = 151bp (Pt29, Pt30, Pt31, Pt32, Pt33); Read 1 = 28bp and Read 2 = 152bp (Pt35); or Read 1 = 111bp and Read 2 = 111bp (RCC4 bulk RNA-seq).

Western blotting

Cell lysates were prepared, and SDS-PAGE/Western blotting was performed as described previously.⁷⁸ Primary antibodies used were anti-HIF-1 α (BD cat no. 610959), anti-HIF-2 α (in-house 190b or Cell Signaling cat no. 7096S), anti-HIF-1 β (Novus Biologicals cat no. NB100-110 or Cell Signaling cat. no. 5537S), anti-CA9 (Abcam cat no. ab15086 or Cell Signaling cat no. 5649S), anti-VHL (Cell Signaling cat no. 68547S) and anti- β -actin (Abcam cat no. ab49900).

Immunohistochemistry

Immunohistochemistry was performed as previously described.⁷⁰ Briefly patient 34 FFPE tumor blocks were cut to 4 μ m, deparaffinized with xylene and ethanol, and rehydrated with double-distilled water. Sections were subjected to heat-induced epitope retrieval (HIER) using Target Retrieval Solution, pH 6, Dako in a pressure cooker for 20 min. Slides were blocked with Dako Peroxidase Blocking solution (Agilent, S2023) and 5% (w/v) bovine serum albumin (BSA; Sigma, 5482) for 10 and 40 min at room temperature respectively. Antibodies incubation was overnight at 4°C (Anti-Cubilin ab191073, Anti-Caveolin-1 ab32577, Anti-ICAM1 ab282575, Anti-CD31ab28364, Anti-VEGF Receptor 1ab32152) diluted 1:200 each. Signal detection was done using Dako Envision system (Agilent, K4003) with diaminobenzidine (DAB) for 10 min. Slides were counterstained with modified Harris Haematoxylin (Thermo Fisher Scientific, 72711) and a differentiation solution of 0.25% HCL in ethanol for 10 s. Ammonia water solution was used for bluing by immersion for 10 s. Slides were dehydrated and mounted with DPX mountant (Merck, 06522).

QUANTIFICATION AND STATISTICAL ANALYSIS

Processing of scRNA-seq data

Demultiplexing of sequencing results, barcode processing, read alignment, and UMI counting were performed using the Illumina 10x Cell Ranger analysis pipeline v6.1.1 with default parameters. Raw reads for each sample were aligned to GRCh38 reference genome, refdata-gex-GRCh38-2020-A.tar.gz, which was provided by 10X Genomics, using Cell Ranger. Cell hashing was used to pool cultured samples. Cells with detected genes <500, fraction of mitochondrially-encoded reads >0.5, or detected genes >3x median for each sample as well as Scrublet⁷⁶ predicted doublets (expected doublet rate 0.095) were removed. The threshold for mitochondrially-encoded reads was used to allow for the high mitochondrial content in renal tubular epithelial cells and was consistent with an observed bimodal distribution and with previous scRNA-seq analysis of renal tissue. For the pooled, hash tagged, cultured cells, a previously described hashing-based doublet detection strategy, implemented in HTODemux⁷⁹ was employed to identify doublets.

Visualization, clustering, and differential gene expression

Downstream analysis was conducted using the R package Seurat v4.⁸⁰ Each gene expression measurement was normalized to total expression in the corresponding cell, multiplied by a scaling factor of 10,000, and log₂-transformed. 2,000 variable features were identified based on stabilized variance and scaled for downstream analysis. The original log-normalized expression values were used for all differential expression and gene set level analyzes. Principal components analysis (PCA) was performed using the scaled expression matrix of the 2,000 variable features. The first 30 principal components (PCs) were used for Louvain clustering of cells with a resolution parameter of 1.0. Cell clusters were identified using the FindClusters function within Seurat, which uses a

shared nearest neighbor (SNN) modularity optimization based clustering algorithm.⁹⁴ Cell clusters were annotated into broad cell types and epithelial cell sub-types by reference to the expression of previously described marker genes (<https://panglaodb.se>),²² and summarized in Figures 1, S1 and S2. Uniform manifold approximation and projection (UMAP) was performed on the same PCs for visualization in two dimensions. CellMixS (v1.18.0) was used to calculate cell between-patient mixing metrics and to evaluate batch effects.

Differential expression analysis was performed using a two-sided Wilcoxon rank-sum test with Bonferroni FDR correction unless otherwise specified and filtered for an adjusted *p* value < 0.05 and log₂-fold change >0.25. Subsets of cells were re-clustered and remapped separately prior to downstream analysis.

Inferring copy number variations in scRNA-seq data using InferCNV

Chromosomal CNVs in individual cells were inferred from scRNA-seq data using InferCNV (<https://github.com/broadinstitute/inferCNV>) with default parameters (cutoff 0.1, cluster_by_groups = T, denoise = T, HMM = T, analysis_mode = "subclusters"). Epithelial cells from normal samples (randomly down sampled to 1,000 cells for the biopsy samples) were used as a reference. Predicted HMM results were then used to assign CNV status to each cell. From InferCNV results, only CNVs found to be recurrent ccRCC driver events in a previous study¹⁶ were reported.

Inferring copy number variations in scRNA-seq data using numbat

Epithelial cells from biopsy tumor samples were further analyzed for CNVs using Numbat (v1.3.2-1), a tool that integrates expression, allele and haplotype information derived from population-based phasing to characterize the CNV landscape in single-cell transcriptomes.²³ Normal epithelial cells from each patient were used to generate the expression reference and Numbat was applied with default parameters. This analysis confirmed CNVs identified by InferCNV. In addition, a focal copy loss at the telomeric end of 3p loss encompassing the VHL locus was identified in T33, which was not detected by InferCNV. One further subclonal copy loss was identified, using Numbat, in T32 (identifying the T32y₁ subclone) and two subclonal copy losses were identified in T33 (identifying T33y₁ and T33z₁ subclones), which were not detected by InferCNV. For visualization of subclones on UMAP plots, 'ambiguous' cells that could not be assigned to a subclone by Numbat (posterior probability <0.95) were excluded. Only CNVs found to be recurrent ccRCC driver events in a previous study¹⁶ were reported, except when non-driver CNVs were required to distinguish genetic subclones.

Somatic copy number analysis of exome-seq data

Exome sequencing was performed on Pt3, Pt17 and Pt20. CNVkit v0.9.8⁸¹ was used to identify CNVs from paired tumor-normal sequencing data using "--drop-low-coverage" to drop bins with low coverage and "--center mode" to recenter the log₂ values. Log₂ values were then converted to copy number to show copy losses or gains. For tumor samples with a single or dominant subclone, the CNVkit profile confirmed CNVs found by InferCNV. CNVkit analysis identified additional focal CNVs that were too small to be identified by InferCNV (e.g., losses on 8p or 6q) and demonstrated 5q tetraploidy in some samples.

Somatic variant analysis in exome-seq data

Fastq files were mapped to reference genome hg38 using bwa 0.7.15.⁸² Aligned bam files were sorted and duplicated reads marked by picard 2.18.20 (<http://broadinstitute.github.io/picard/>). bam files were then processed using the GATK 4.1.9 tools BaseRecalibrator and ApplyBQSR (<https://gatk.broadinstitute.org/hc/en-us>).⁸³ Bam files were ordered according to the reference genome using picard. Mutect2 from GATK 4.1.9 was used to call somatic variants in tumor samples, and filtering using gnomAD v4.0 (<https://gnomad.broadinstitute.org/downloads>). Potential somatic mutations were further filtered by FilterMutectCalls in GATK 4.1.9. The final mutation lists were annotated using Funcotator in GATK 4.1.9 using the data source funcotator_dataSources.v1.7.20200521s. Analysis focused on previously described driver mutations.¹⁶ Considering the clonal composition of samples (derived from scRNA-seq data) enabled mapping of variants to specific subclones.

Allele frequency analysis in exome-seq data

HaplotypeCaller in GATK 4.1.9 was used to call germline variants. Heterozygous germline variants identified in normal samples with a read depth >30 and balanced allele frequency of reference/alternative between 0.35 and 0.65 were leveraged to perform allele frequency analysis in patient-matched tumor samples. This identified mirrored subclonal allelic imbalance (differential loss of maternal and paternal chromosome copies in different subclones) indicating independent occurrences of the same CNVs. Specifically, this analysis identified loss of different 14q copies in samples from T3 (which was not possible using InferCNV) distinguishing T3y₁ and T3z₁ subclones from each other. Similarly, gains of different chr5 and chr20 copies were identified in different T17 samples, distinguishing them as events occurring independently in subclones T17y₁ and T17w₁x₁.

Allele frequency analysis in scRNA-seq data

SCReadCounts⁸⁴ was used to generate cell-SNV matrices containing the absolute variant- and reference-read counts from bar-coded scRNA-seq alignments. Specifically, NM and MD tags were added to scRNA-seq bam files using samtools⁸⁵ calmd -b, and the varLoci function from SCReadCounts used to get potential variants from the normal scRNA-seq samples (minimal

alternative reads = 10). Then, scReadCounts was used to generate the reference reads count and alternative reads count from both the normal and tumor scRNA-seq samples for the potential variants generated using varLocs with the parameter -C STARsolo.

To validate the mirrored subclonal allelic imbalance at 14q in T3 subclones (T3y₁ and T3z₁) suspected from exome-seq, heterozygous sites from the normal sample (N3) were selected using both exome-seq data and scRNA-seq data. From exome-seq data, sites with at least 10 alternative reads, 10 reference reads and balanced allele frequency (alternative reads frequency between 0.3 and 0.7) were selected. From the scRNA-seq data, cells from each T3 sample (T3A and T3B) were stratified by 14q status, according to inferCNV results. Then, using the reads counts from SCReadCounts, heterozygous sites were selected from N3 scRNA-seq data, requiring that the alleles must overlap with heterozygous sites identified in exome-seq and be balanced in N3 scRNA-seq data (alternative reads frequency between 0.3 and 0.7). Of those sites, those with over 50 reads collectively across T3 samples were selected, yielding 7 high confidence sites. Reference and alternative allele frequency was calculated at these sites in T3 cancer cells stratified by sample and 14q status, as well as N3 cells. This confirmed an allelic bias in cells with 14q loss from different T3 samples, indicating they lost different 14q copies.

This was also applied to investigate a cluster of T17 cells (mapping to clone T17y₁) in which 3p loss was not detected by InferCNV, despite expressing CA9 and other HIF target genes. To test for copy neutral loss of heterozygosity, heterozygous sites on 3p in the Pt17 normal sample (N17) were selected as above, specifically in the region undergoing copy loss in other ccRCCs (chr3:1-86900000). Sites from exome-seq were selected for those with a minimum 10 reference and 10 alternative reads, and allele frequency between 0.25 and 0.75. From scRNA-seq data, the sites were further selected for those having at least 10 reads from the N17 and T17 samples collectively, and at least 5 reads in the 3p copy neutral T17 cluster under investigation. Additionally, sites taken forward had at least 3 reference and 3 alternative reads in N17 scRNA-seq data. This yielded 9 high-confidence sites, which showed a balanced allele frequency in N17, yet the cluster under investigation only showed signal from one allele. This demonstrated that indeed the T17y₁ subclone underwent copy neutral loss of heterozygosity at 3p, whereby it originally lost 3p and then duplicated the remaining copy.

RNA-seq analysis of RCC4 cells

RNA-seq data was analyzed as previously described.⁷⁸ Briefly, Illumina adaptor sequences were trimmed using TrimGalore (0.3.3). Reads were aligned to Genome Reference Consortium GRCh37 (hg19) using HISAT2 (2.05) (<http://daehwankimlab.github.io/hisat2/>).⁸⁶ Non-uniquely mapped fragments were excluded using Picard tools (2.0.1) (<http://broadinstitute.github.io/picard/>). Total read counts for each UCSC-defined gene were extracted using HTSeq (0.5.4p3)⁸⁷ with 'intersection-strict' mode, and significantly regulated genes were identified using DESeq2, pairing conditions by biological replicate.⁸⁸

Gene set intersection analysis

Gene set intersection analysis was performed using UpSetR (<https://upset.app>).^{89,95} Significance was determined using a bootstrapping approach with 1,000 iterations.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed using 10,000 permutations, weighted enrichment score and pre-ranking of genes.^{90,96,97} The ranking metric (log₂ fold change) was first calculated for each tumor/subclone and then averaged across tumors/subclones to avoid biasing the analysis toward tumors/subclones with more cells. Genesets were accessed through MSigDB (<https://www.gsea-msigdb.org/gsea/msigdb/>) or published studies.^{18,39}

Transcriptional regulator prediction using LISA

Upstream transcriptional regulators of gene programs were identified using epigenetic Landscape In Silico deletion Analysis (LISA – v2.3.0 - <https://github.com/liulab-dfci/lisa2#>).³⁰ For epithelial cells, tumor (vs. normal) dysregulated transcription regulators were calculated by comparing the consistently upregulated ($n = 153$) and consistently downregulated ($n = 203$) genes in ccRCC cancer cells (vs. proximal tubules). The first sample p -value was selected for each transcriptional regulator. For transcriptional regulators enriched in the tumor downregulated (proximal tubule upregulated) genes, the p -value was negatively transformed so that it could be correlated with log₂FC in expression of the transcriptional regulator. The non-epithelial cell LISA analysis was conducted by inputting the consistently tumor upregulated genes only, since there were insufficient numbers of genes consistently downregulated (<50 genes).

Transcriptional regulator prediction using SCENIC

Single-cell regulatory network inference and clustering (SCENIC – v1.3.1 - <https://scenic.aertslab.org>)⁶⁴ was used to identify transcription regulators showing intratumor heterogeneity between cancer epithelial cells. For SCENIC analysis, variable regulons were calculated for each tumor, and the 9 regulons displaying intratumor heterogeneity in all 8 ccRCCs analyzed were described. Activities of the 9 regulons were mapped to individual cancer epithelial cells (each tumor separately), then scaled and centered across cells from each tumor. The average activity was calculated for each subcluster in each tumor, and then the 8 matrices merged by regulon. Activities for each regulon were scaled across subclusters before being visualized on a heatmap.

Cell-cell interaction analysis

Cell-cell interactions were analyzed using CellphoneDB v4.1.0⁹¹ and visualized using ktplots version 2.3.0 (<https://github.com/zktuong/ktplots>).

TCGA RNA-seq data and prognosis analysis

FPKM-UQ normalized RNA-seq data for 9,760 primary tumor samples and 730 normal samples were obtained from <https://portal.gdc.cancer.gov/repository> on 07.09.2021 using the gdc-client version 1.5.0 and the following advanced filters: cases.project.program.name in ["TCGA"], files.analysis.workflow_type in ["HTSeq - FPKM-UQ"], files.data_category in ["transcriptome profiling"], files.experimental_strategy in ["RNA-Seq"] and either cases.samples.sample_type in ["primary tumor"] or cases.samples.sample_type in ["solid tissue normal"].

Deciphering intratumor expression programs and meta-programs

Tumors with more than 500 captured cancer epithelial cells (8 of 10 ccRCCs) were analyzed for intratumor expression programs. Epithelial cells from each tumor were reclustered and remapped separately (to derive tumor 'subclusters'), as described above. Small subclusters suggestive of cell doublets (low expression of the ccRCC marker CA9 and high expression of non-epithelial cell markers) or poor-quality cells (e.g., low reads/features) were excluded. The remaining cells were then reclustered and remapped a second time. Subcluster marker genes were then distinguished for each tumor as described above. Analysis of intersecting gene sets identified 57 subcluster marker genes common to all 8 ccRCCs.

The expression of each of these 57 genes was then scaled and centered across cells from each tumor and used to calculate the average expression for each tumor subcluster. Data from the 8 tumors were then merged, and each rescaled across the subclusters. The Euclidean distance was calculated between genes, and hierarchical clustering (complete linkage) performed on the distance matrix to construct a dendrogram tree. The tree was cut into 3 clusters of genes, and the clusters of genes defined as gene modules. Gene module 'expression' was calculated for single cells using the Seurat "AddModuleScore" function. Cutting the dendrogram tree into 3, 4 and 5 clusters was tested, but the associated gene modules with 4/5 clusters exhibited high spearman correlation coefficients between certain pairs of modules in tumor subclusters, and therefore 3 clusters was selected.

Leveraging the three gene modules as 'core modules', we then performed spearman correlation coefficient rank tests to identify genes that correlated with each module in each tumor across individual cancer epithelial cells. Owing to the large number of cells, tumor T33 was randomly down sampled to 500 cells per sample (2,500 total) for this analysis. The average correlation coefficient across all 8 tumors was used as the gene ranking metric for subsequent GSEA. Positively correlated genes (adjusted $p < 0.05$) common across all 8 tumors defined the 'extended modules'.

Generation of HIF-1 and HIF-2 specific gene metagenes

Genes downregulated in RCC4 cells by HIF1A siRNA or HIF2A (EPAS1) siRNA (Benjamini-Hochberg corrected $p < 0.1$ and a \log_2 fold change < -0.263), relative to control siRNA, defined HIF-1 and HIF-2 target genes respectively. Non-overlapping HIF-1 and HIF-2 target genes defined the HIF-1 specific and HIF-2 specific metagenes (excluding HIF1A and HIF2A/EPAS1 themselves). HIF-1 and HIF-2-specific metagene 'expression' scores were assigned to single cells using the Seurat "AddModuleScore" function with "search = TRUE" to ensure the correct alias was used in the scRNA-seq data.

Cell Reports, Volume 44

Supplemental information

**Conserved patterns of transcriptional
dysregulation, heterogeneity, and cell
states in clear cell kidney cancer**

Olivia Lombardi, Ran Li, Faiz Jabbar, Hannah Evans, Silvia Halim, Joanna D.C.C. Lima, Lisa Browning, Helen M. Byrne, Hani Choudhry, Peter J. Ratcliffe, and David R. Mole

Figure S1

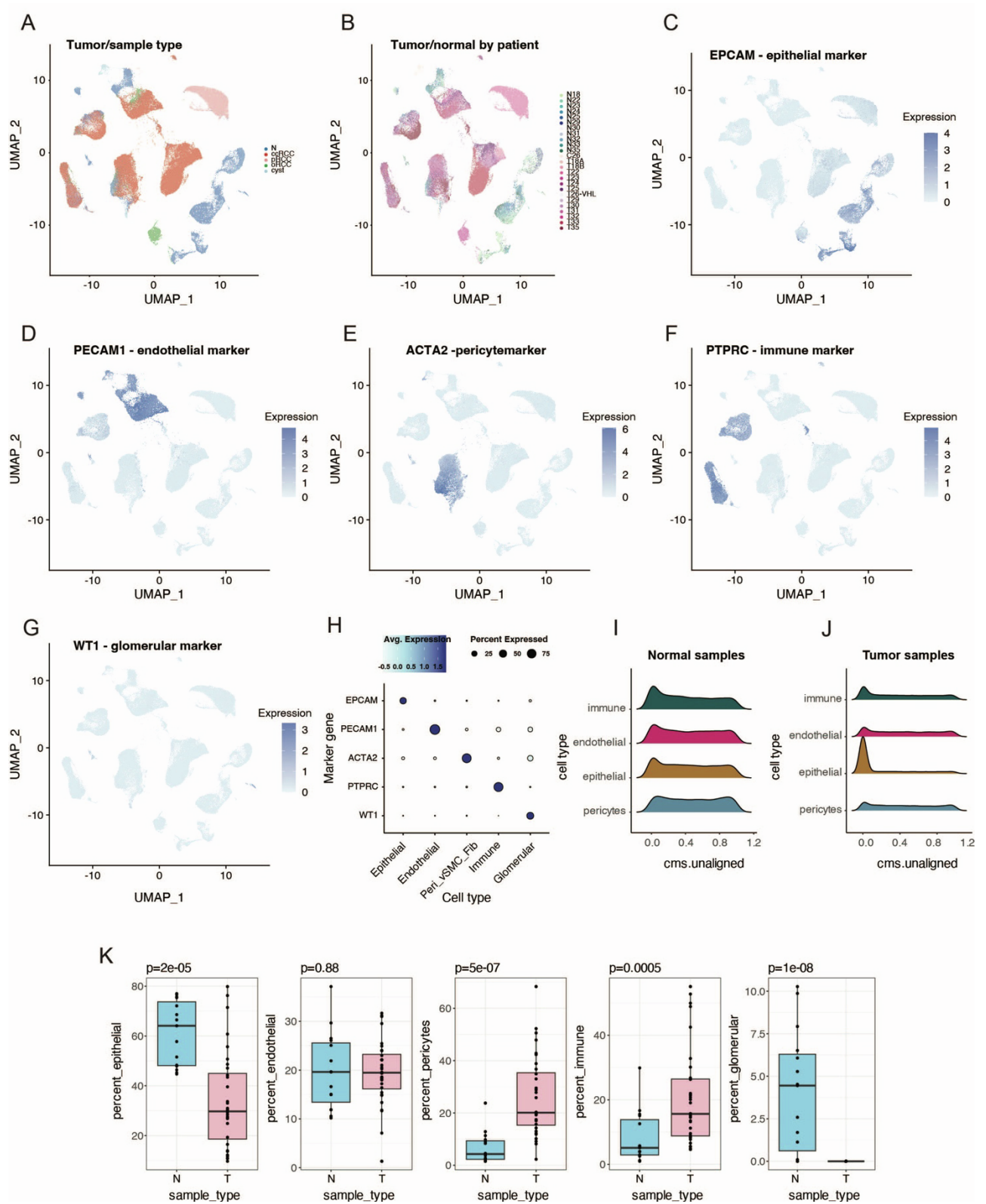


Figure S2

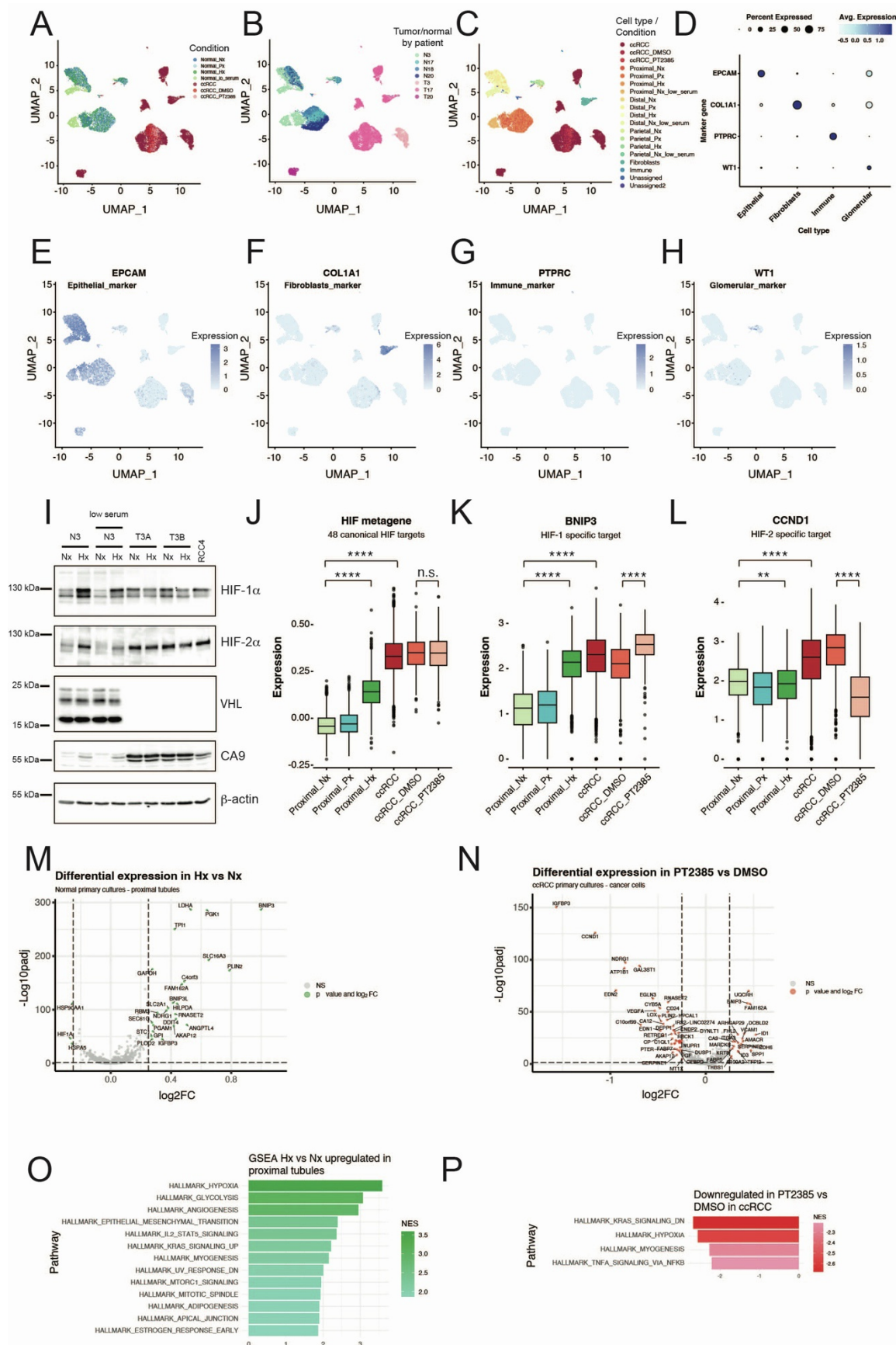


Figure S3

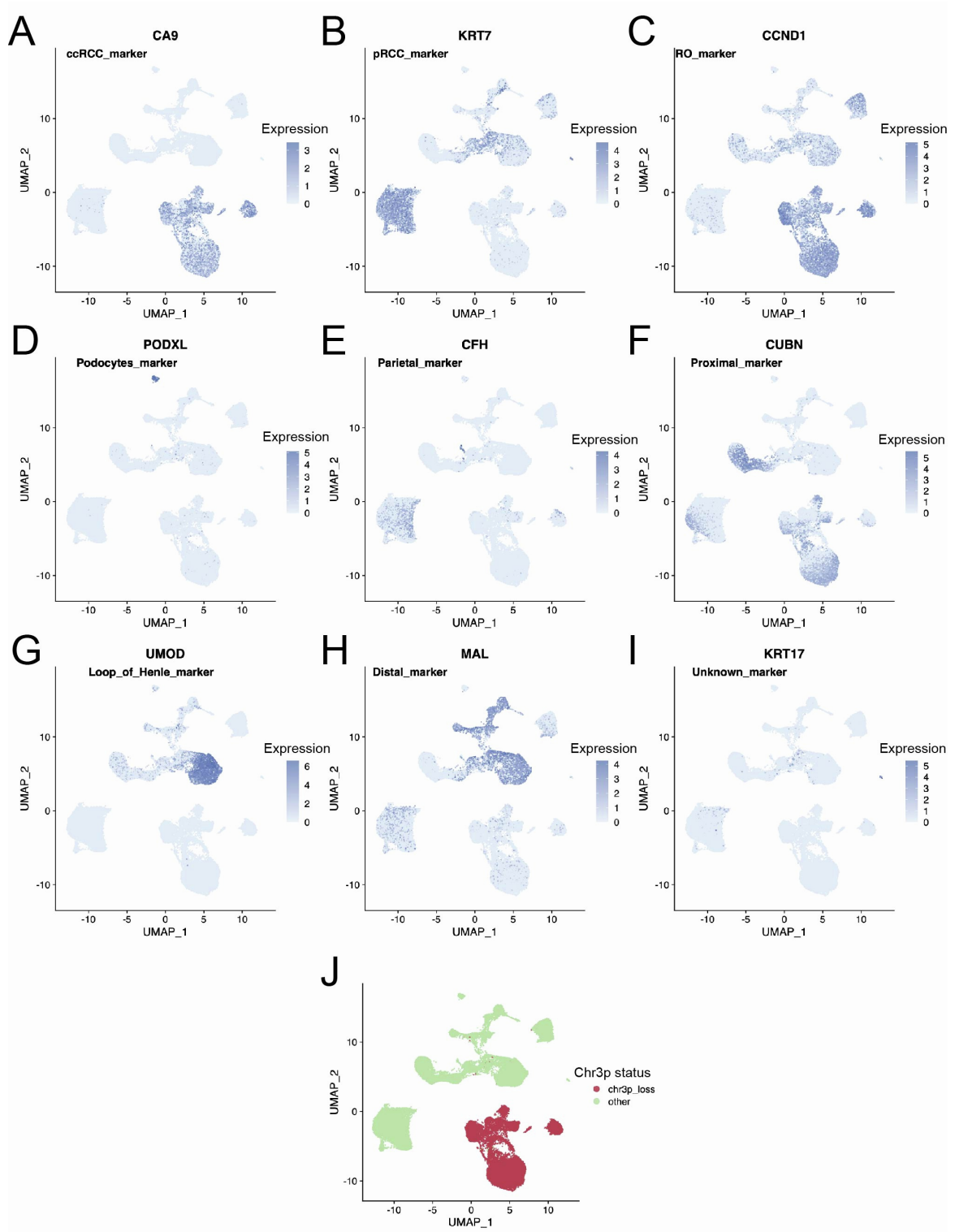


Figure S4

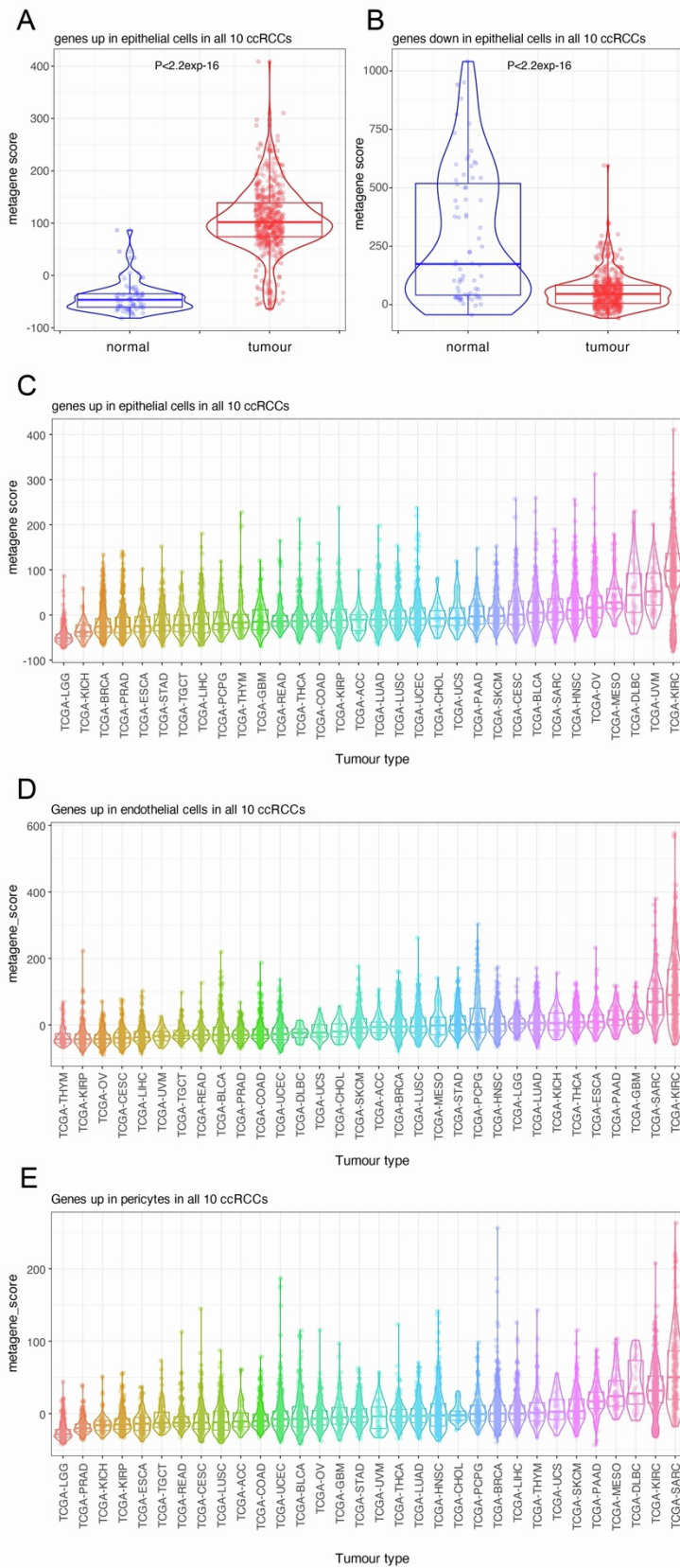


Figure S5

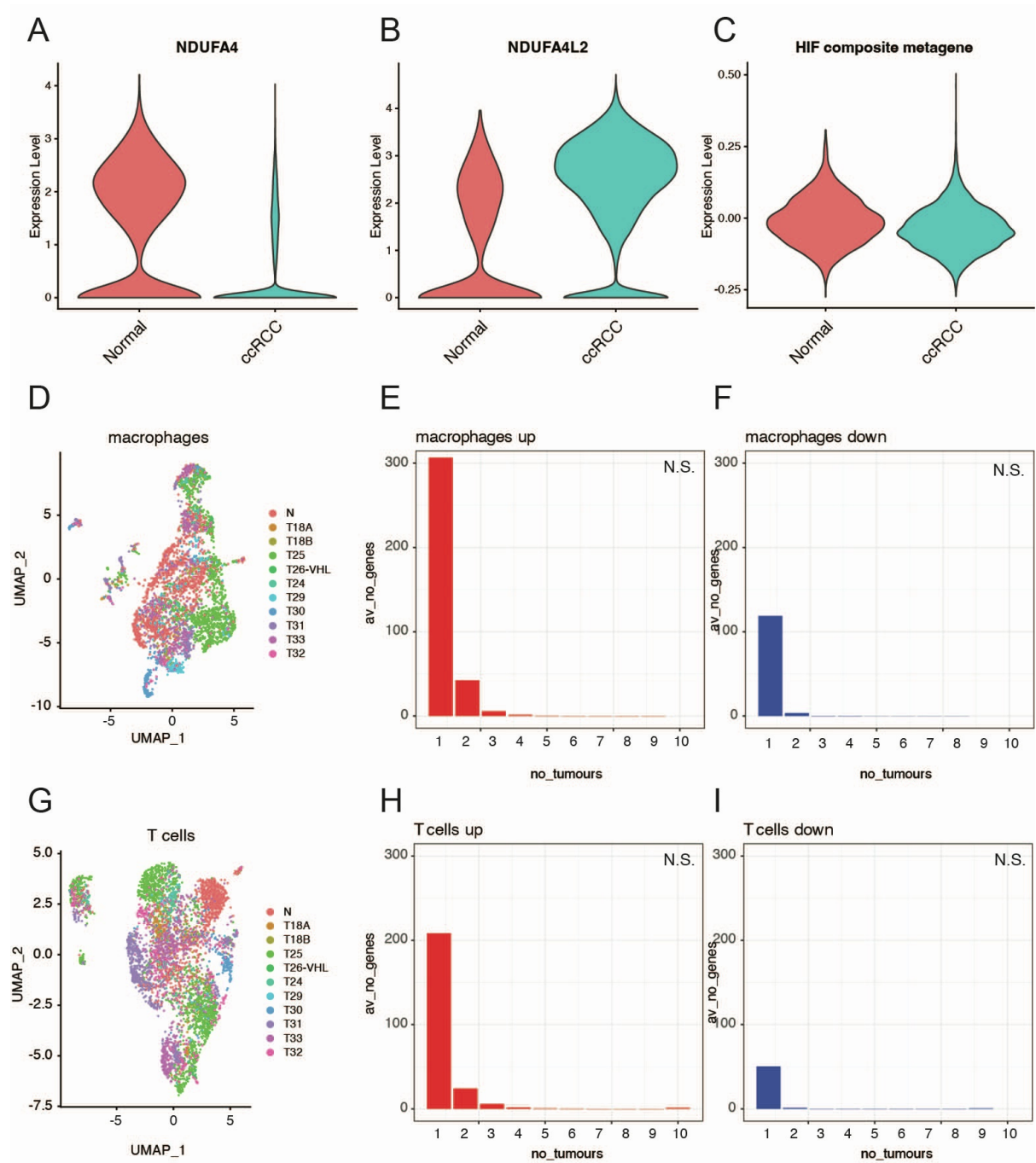


Figure S6

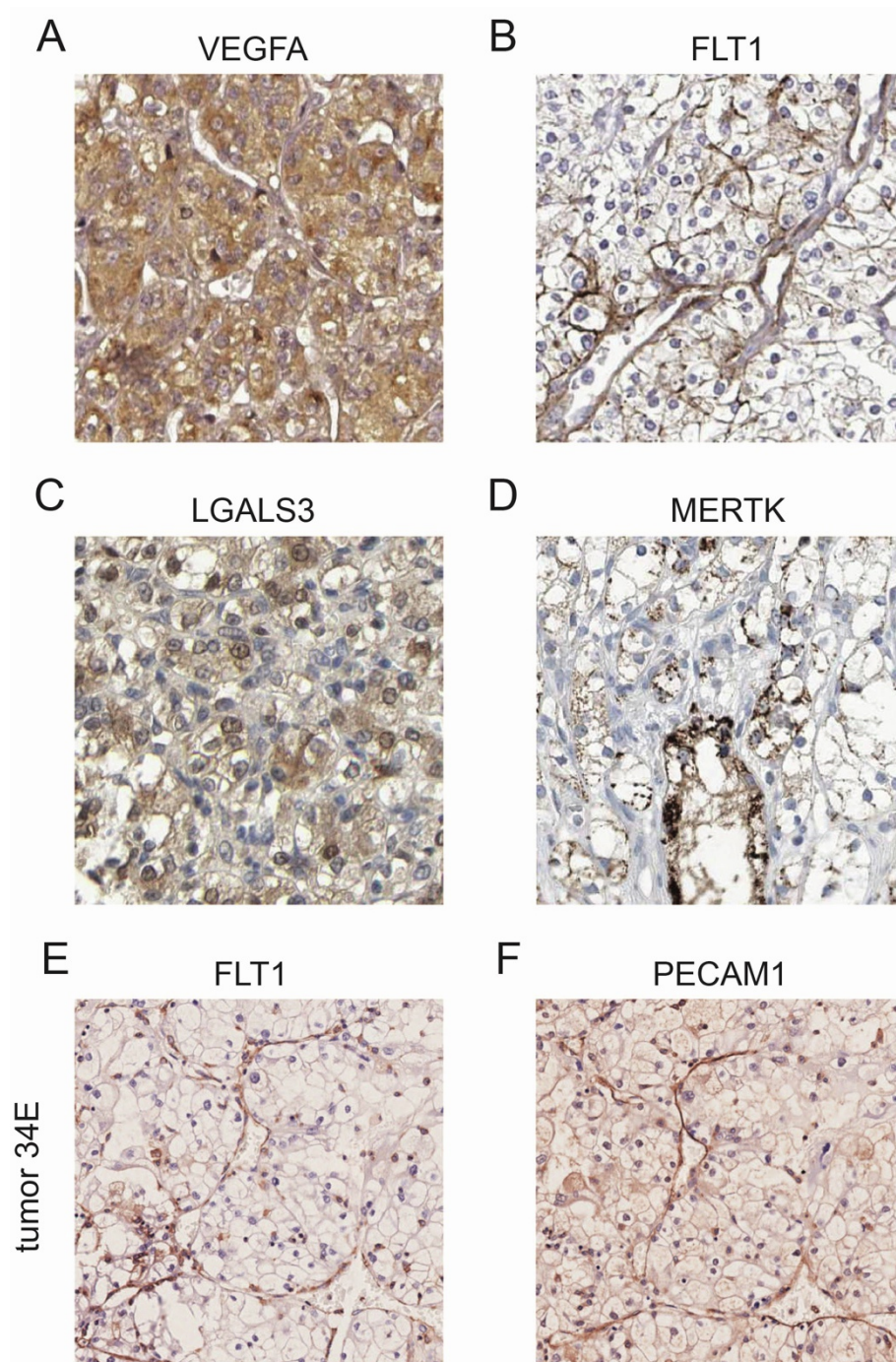


Figure S7

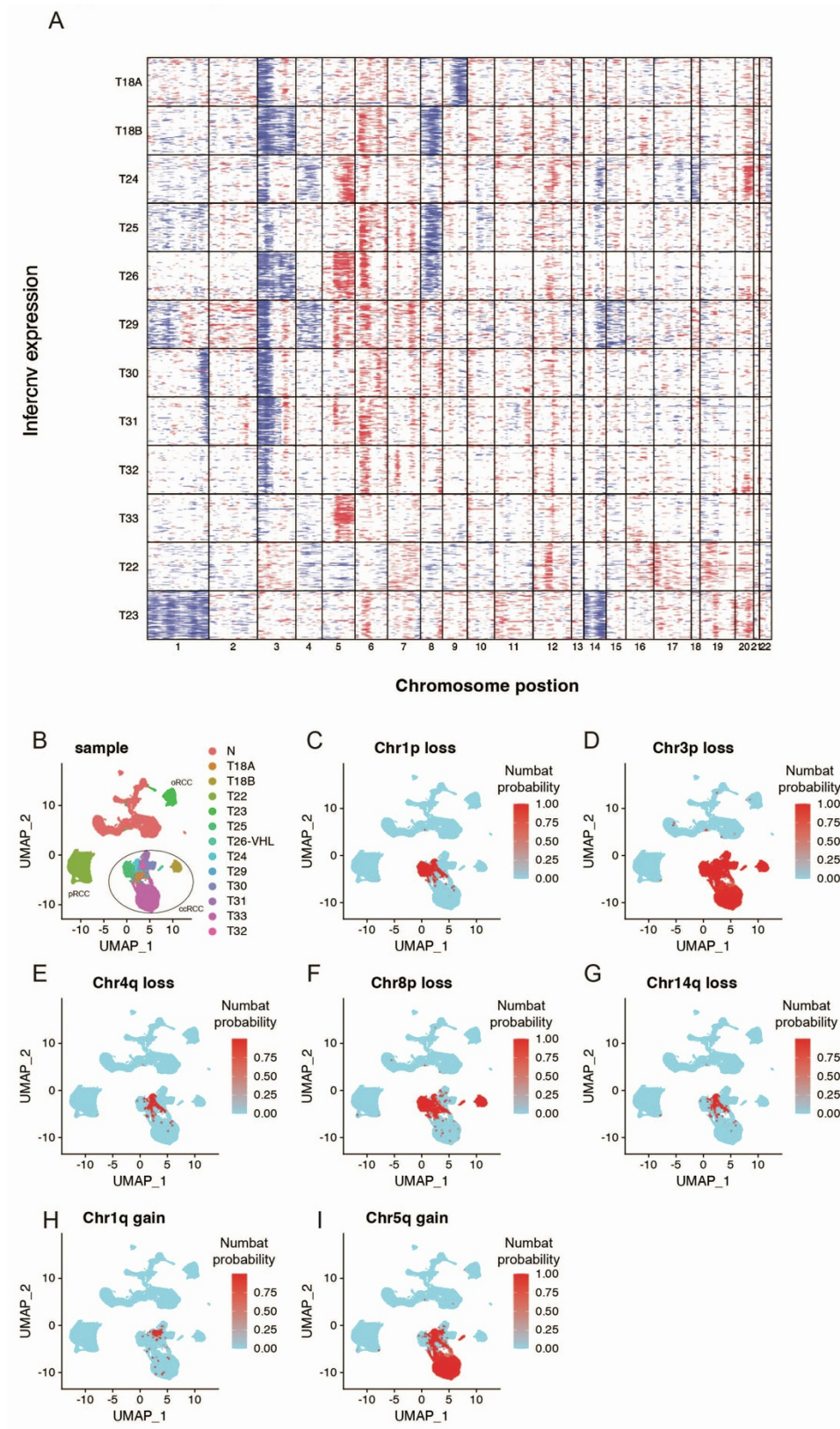


Figure S8

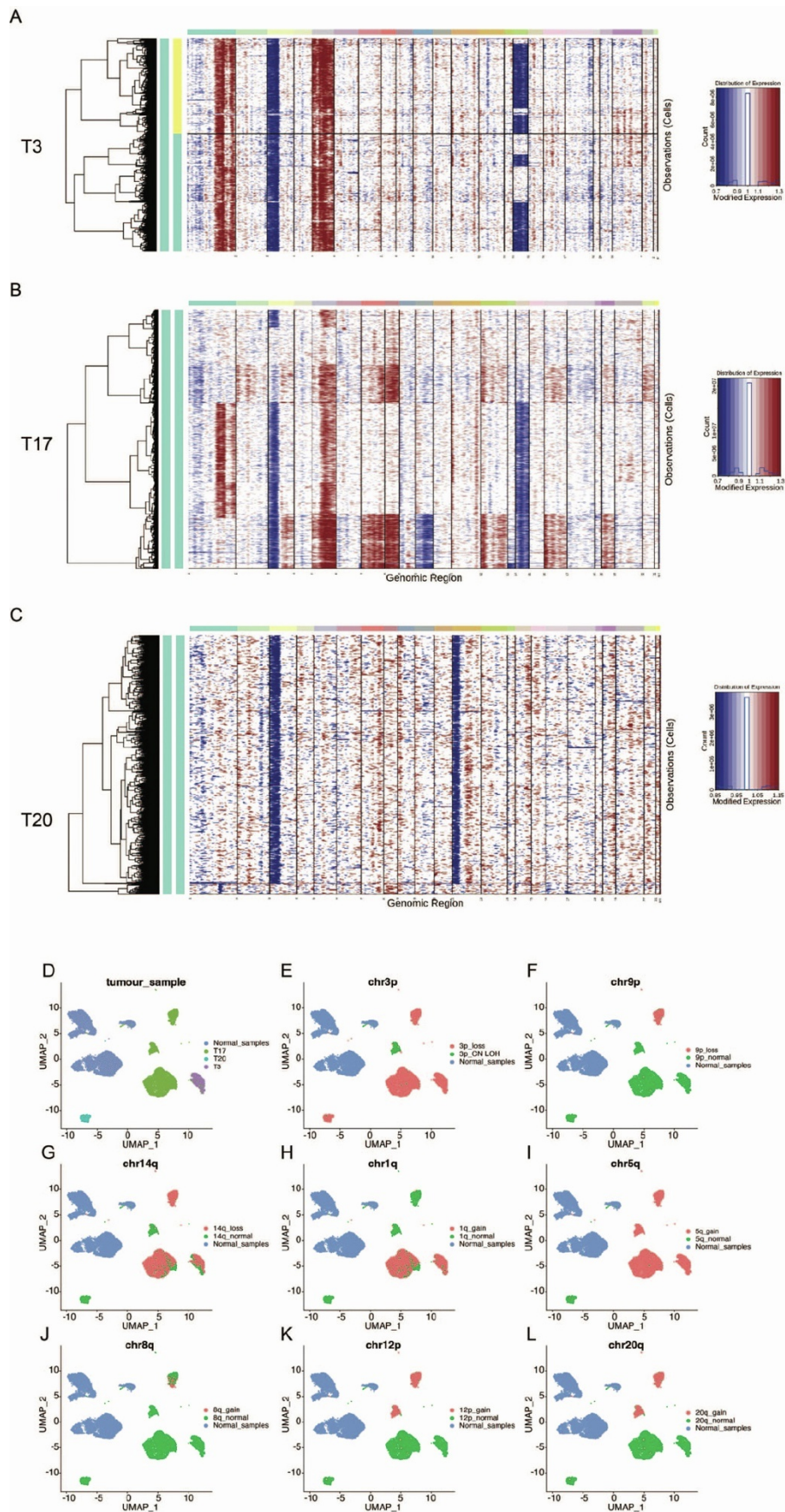


Figure S9

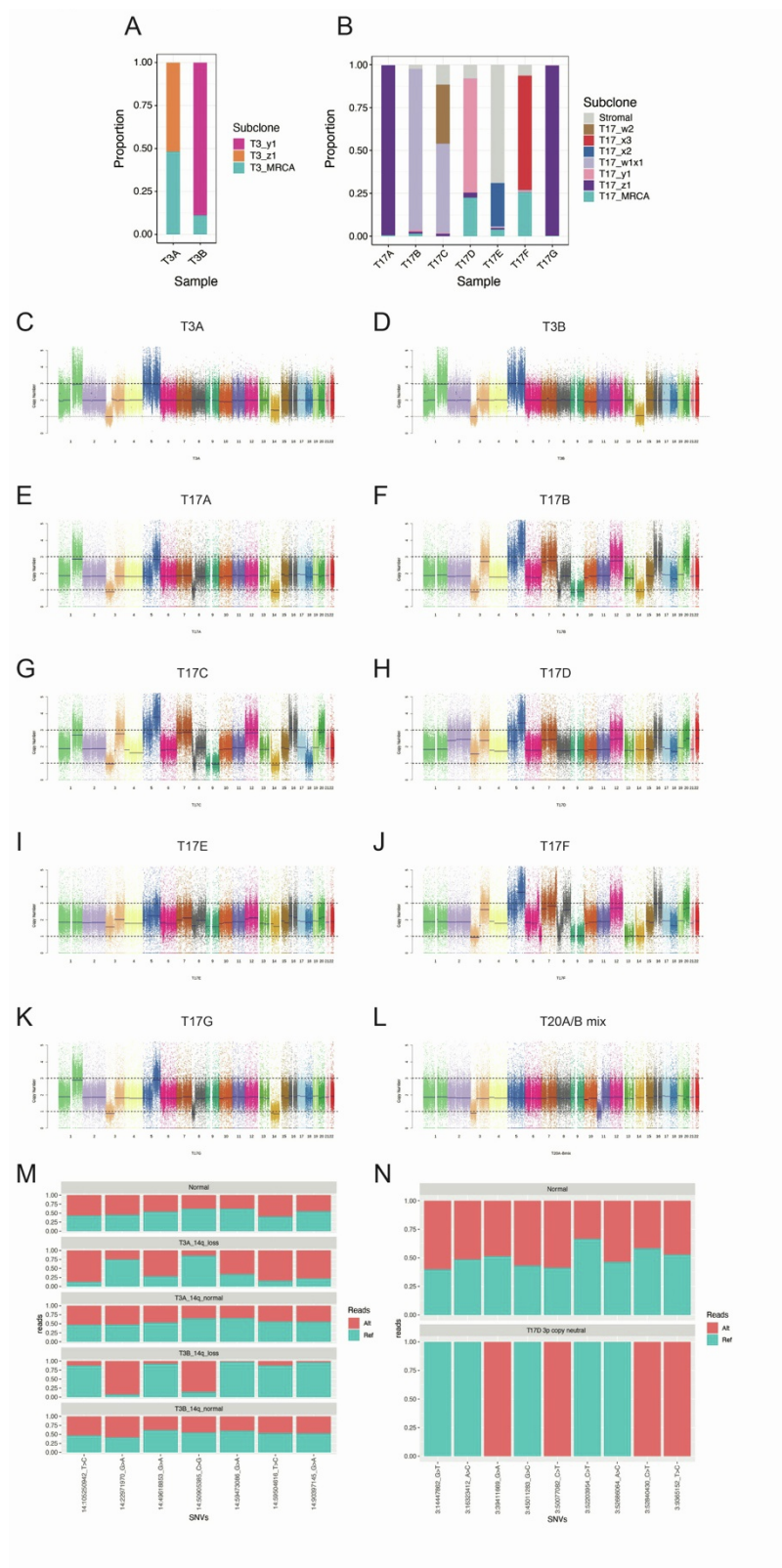


Figure S10

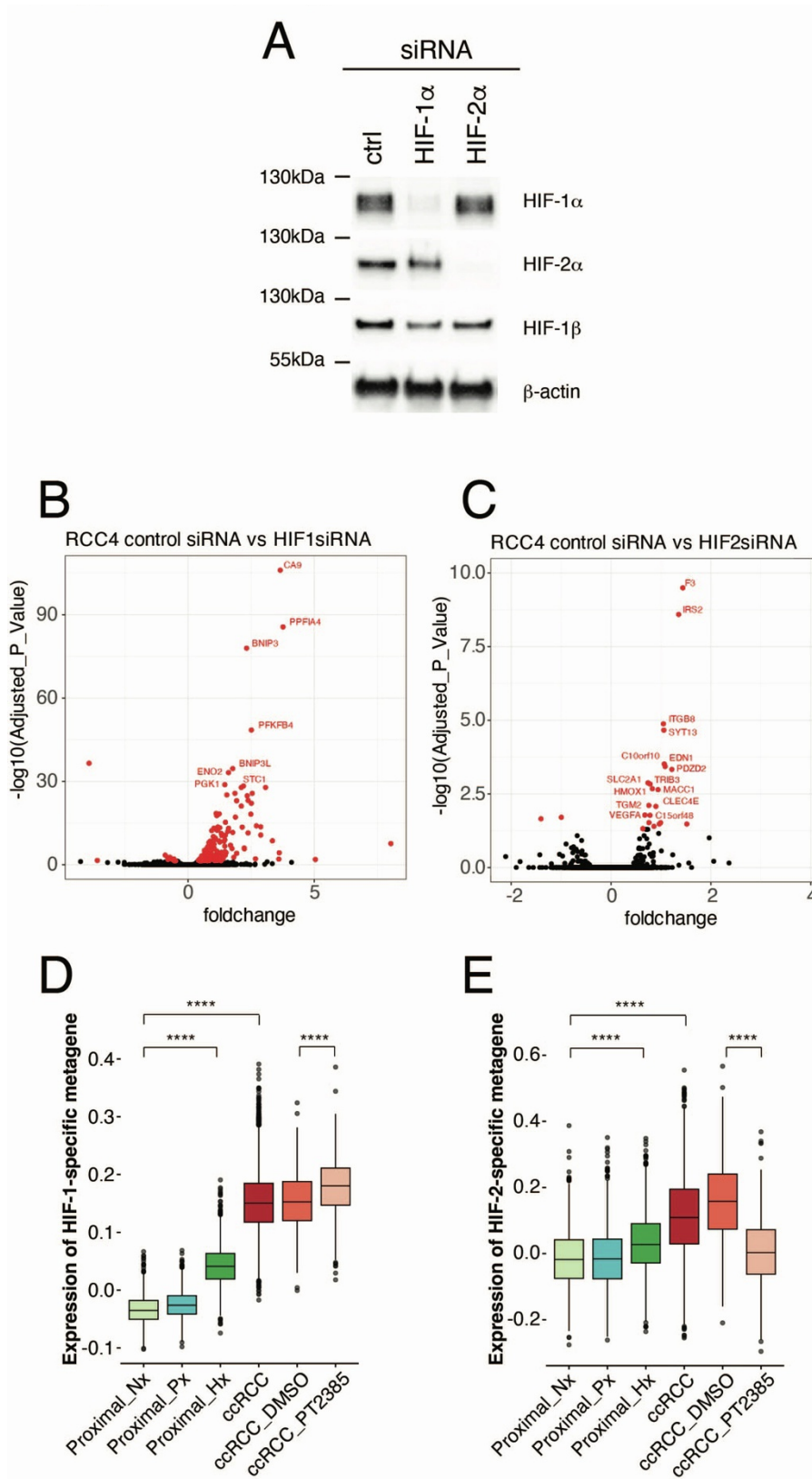


Figure S11

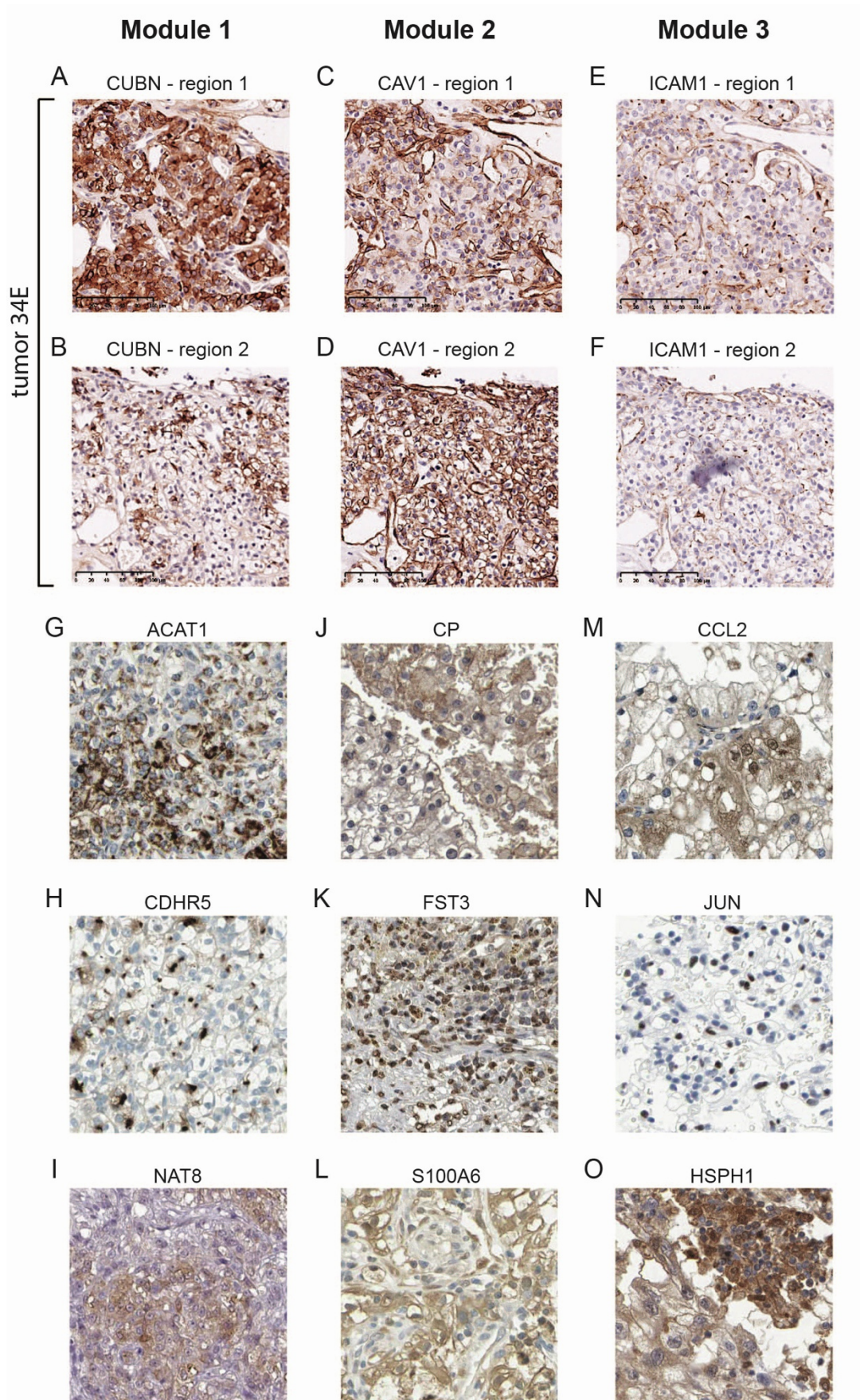


Figure S12

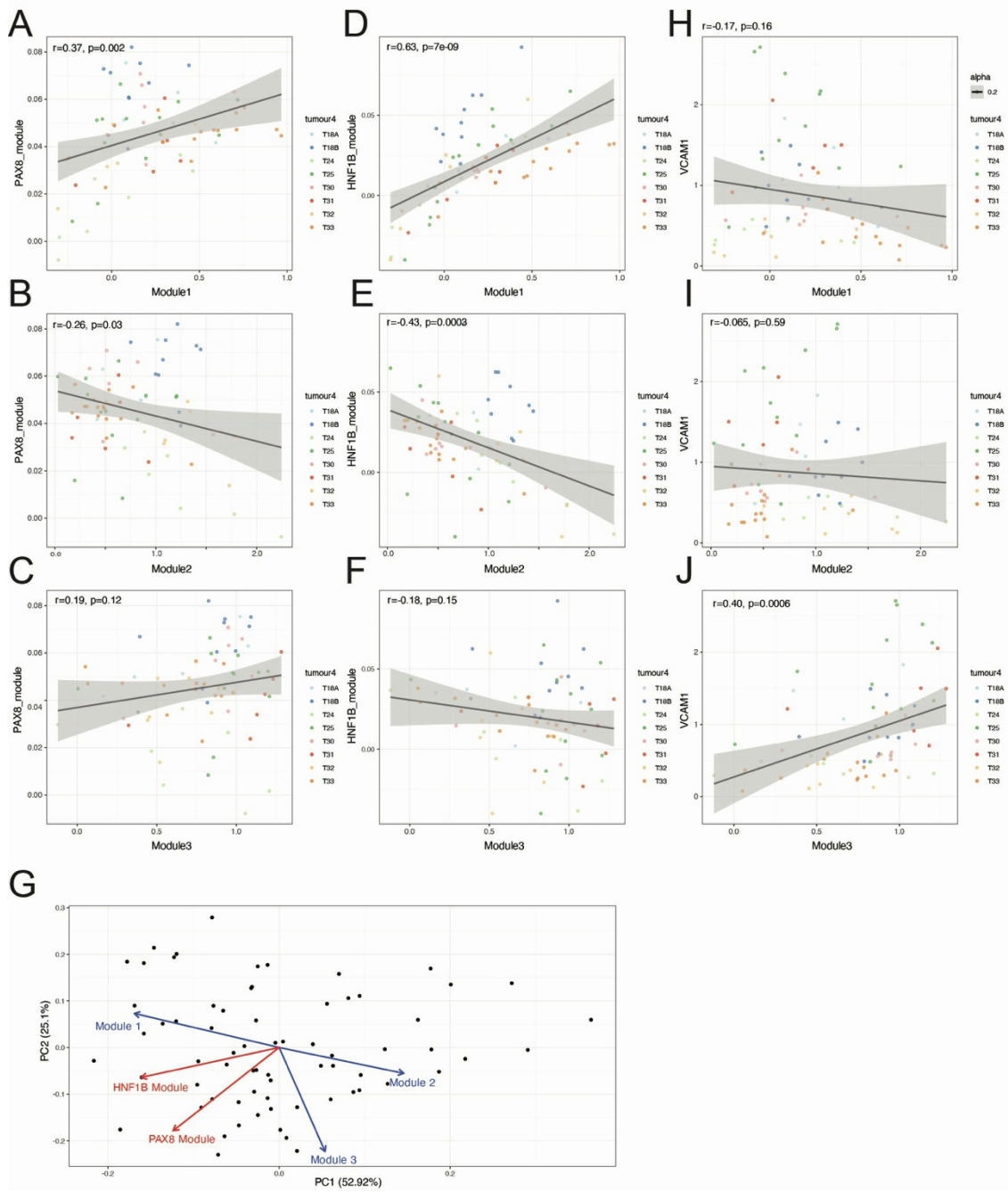


Figure S13

A



B

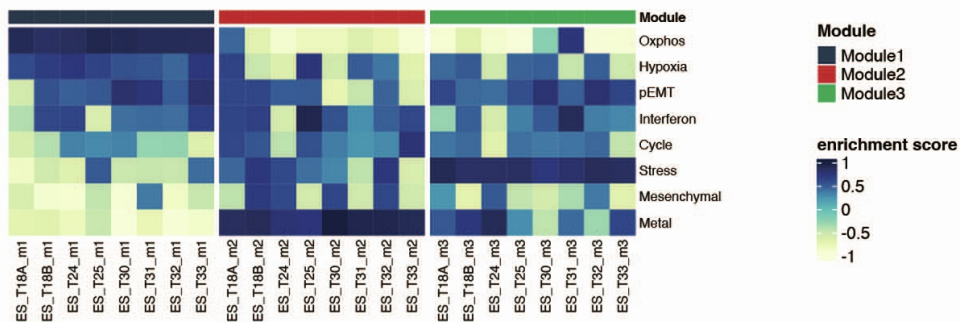


Figure S14

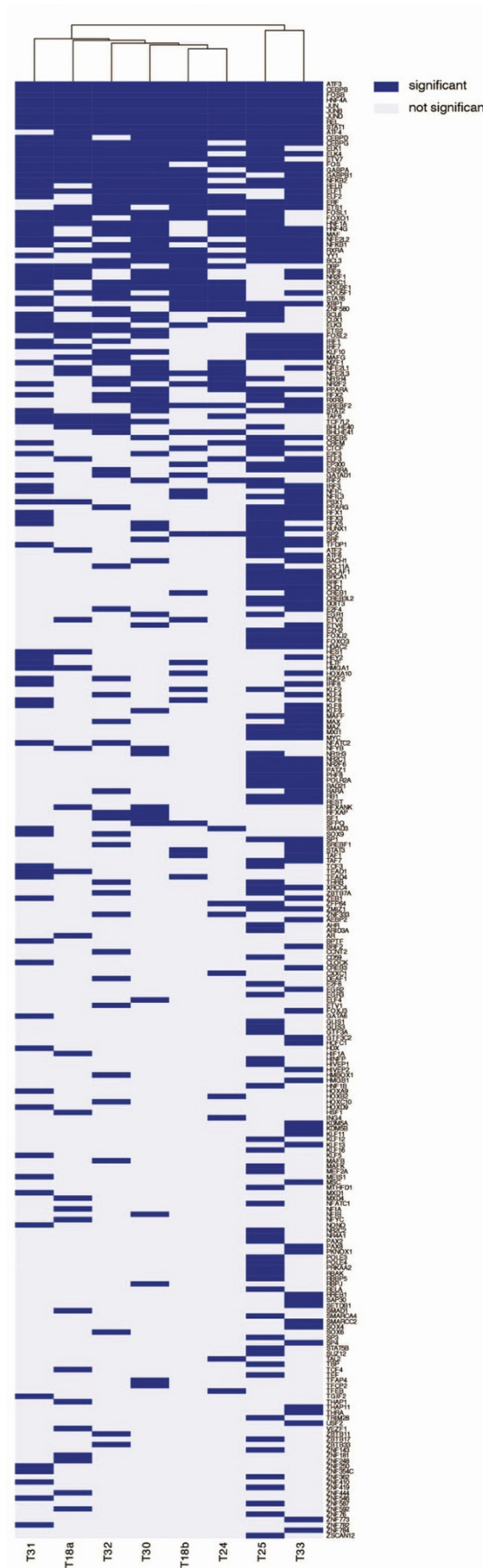
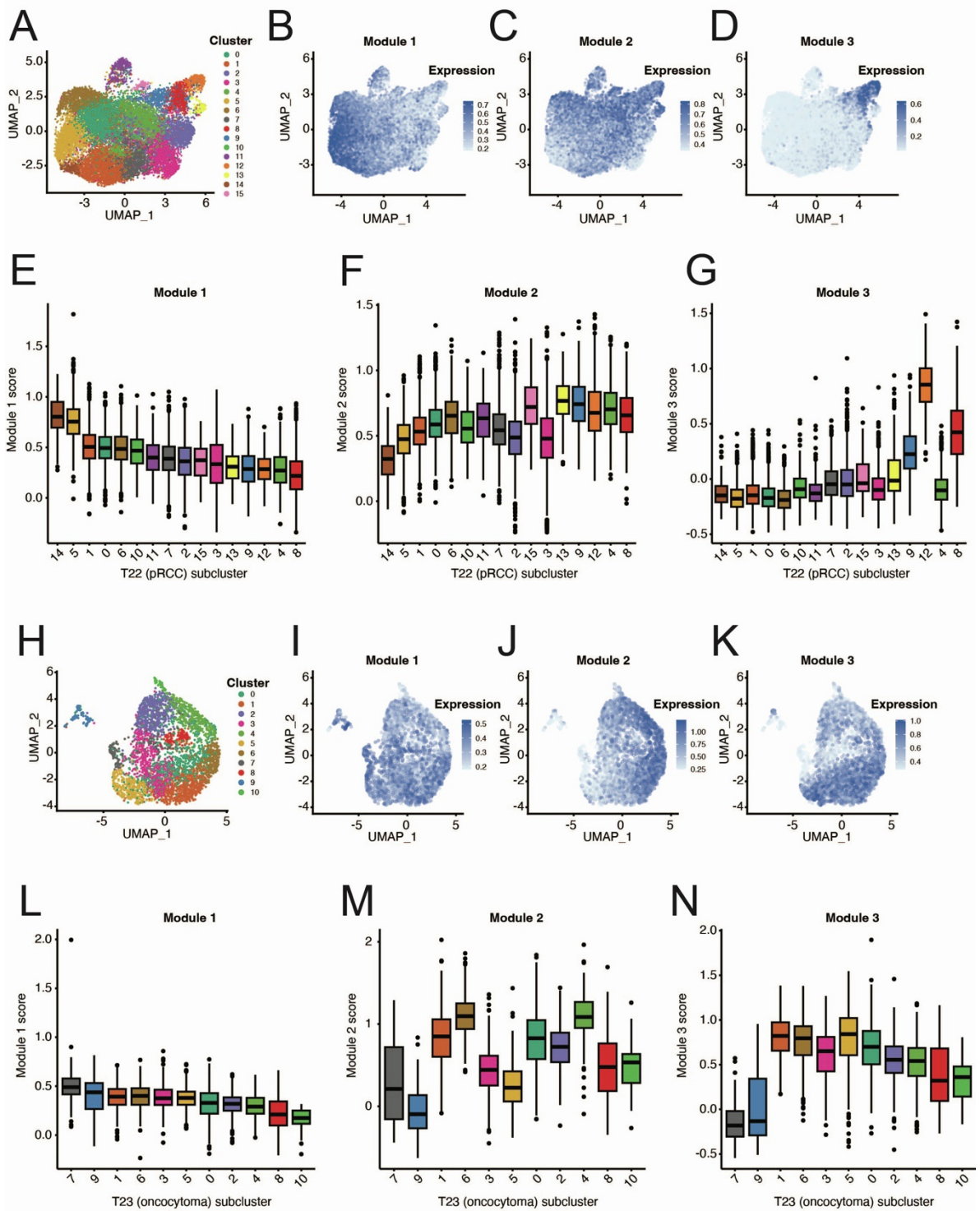


Figure S15



Supplemental Figure Legends

Figure S1. scRNA-seq analysis of freshly-dissociated biopsy samples, Related to Figure 1. UMAP plots of cells from freshly-biopsied tumor and background kidney biopsy samples showing (A) samples from different tumor types, normal background kidney and cystic background kidney; (B) sample type and patient; and expression of (C) EPCAM, (D) PECAM1, (E) ACTA2, (F) PTPRC, and (G) WT1. (H) Dotplots for cells from freshly-biopsied samples showing marker gene expression according to major cell type. Plots showing cellular frequency distribution against between-patient cell mixing score, stratified by cell type, for (I) normal and (J) tumor samples. (K) Box-and-whisker plots showing proportion of each cell type in normal and tumor samples. Box-and-whisker plots show median, inter-quartile range and range. See also Figure 1.

Figure S2. scRNA-seq analysis of cultured cells, Related to Figure 1. UMAP plots of cells from cultured tumor and normal kidney samples showing (A) sample type and intervention; (B) sample type and patient; and (C) cell type and intervention. (D) Dotplots for cells from primary cultures showing marker gene expression according to major cell type. UMAP plots showing expression of (E) EPCAM, (F) COL1A1, (G) PTPRC, and (H) WT1. (I) Immunoblots showing levels of HIF-1 α , HIF-2 α , VHL, CA9 and β -actin (loading control) protein in patient 3 cultured cells incubated in normoxia (Nx), hypoxia (Hx, 0.5% O₂), or low serum (0.5% FBS) conditions for 16 hours. Depicts hypoxic-induction of HIF-1 α , HIF-2 α and CA9 in primary normal kidney cultures, yet constitutive expression in primary VHL-deficient ccRCC cultures. RCC4 (established ccRCC cell line) were used as a reference. Violin plots showing RNA expression of (J) composite HIF metagene⁷⁵, (K) BNIP3 and (L) CCND1 in single cells cultured in normoxia or 0.5% hypoxia or with DMSO or 1 μ M PT2385 (Belzutifan, HIF-2 inhibitor) for 16 hours. n.s. not significant, **p<0.01, ****p<0.0001. (M) Volcano plot showing genes differentially expressed in normal proximal tubular cells cultured in 0.5% hypoxia vs normoxia for 16 hours. Green denotes genes with log₂(fold-change) >0.25 or <-0.25 and Bonferroni-adjusted-p-value <0.05; remaining genes colored grey. (N) Volcano plot showing genes differentially expressed in ccRCC epithelial cells cultured with 1 μ M PT2385 for 16 hours. Red denotes genes with log₂(fold-change) >0.25 or <-0.25 and Bonferroni-adjusted-p-value <0.05; remaining genes colored grey. (O) GSEA showing Hallmark pathways enriched amongst genes induced by hypoxia in normal proximal tubular cells. (P) GSEA showing Hallmark pathways enriched amongst PT2385-downregulated genes in ccRCC epithelial cells. Box-and-whisker plots show median, inter-quartile range and range. See also Figure 1.

Figure S3. Marker gene and genetic marker validation of epithelial cell types, Related to Figure 2. UMAP plots of epithelial cells from freshly-biopsied samples showing expression of (A) CA9, (B) KRT7, (C) CCND1, (D) PODXL, (E) CFH, (F) CUBN, (G) UMOD, (H) MAL, (I) KRT17, (J) chromosome 3p copy number loss. See also Figure 2.

Figure S4. Analysis of composite metagenes, based on scRNA-seq analysis, in bulk RNA-seq analysis from the TCGA cohort, Related to Figure 2. (A) Violin and box-and-whisker plots showing collective expression of genes upregulated in all 10 ccRCCs vs proximal tubule cells (as a composite metagene) shown in bulk RNA data from normal and tumor samples from the TCGA-KIRC cohort. (B) Same analysis for genes downregulated in all 10 ccRCCs vs proximal tubule cells. Violin and box-and-whisker plots showing expression of composite

metagenes based on genes up regulated in (C) epithelial cells (D) endothelial cells and (E) pericytes from all 10 ccRCCs in bulk RNA-seq analysis of tumor samples from 32 tumor types in the TCGA cohort. Tumor types are ranked according to their median expression of each metagene. Box-and-whisker plots show median, inter-quartile range and range. See also Figure 2.

Figure S5. Analysis of mitochondrial subunit switching in pericytes and conserved patterns of gene dysregulation in macrophages and T cell populations, Related to Figure 3. Violin plots showing expression of (A) NDUFA4, (B) NDUFA4L2, and (C) composite HIF-target metagene⁷⁵ in pericytes derived from freshly-biopsied normal kidney samples (red) and ccRCC samples (blue). (D) UMAP plot of macrophages from freshly-biopsied samples showing sample type and patient. (E) Intersection analysis showing the overlap between genes upregulated in macrophages from each of the 10 ccRCC tumors compared to macrophages from the normal kidney samples. (F) The same analysis for downregulated genes. (G) UMAP plot of T cells from freshly-biopsied samples showing sample type and patient. (H) Intersection analysis showing the overlap between genes upregulated in T cells from each of the 10 ccRCC tumors compared to T cells from the normal kidney samples. (I) The same analysis for downregulated genes. See also Figure 3.

Figure S6. Immunostaining of representative ligand-receptor pairings, Related to Figure 3. (A) VEGFA, (B) FLT1, (C) LGALS3 and (D) MERTK protein expression in ccRCC tumors from the Human Protein Atlas. Serial sections from tumor 34E showing co-localization of (E) FLT1 and (F) the endothelial marker PECAM1. See also Figure 3.

Figure S7. Analysis of chromosomal copy number variation (CNV) in freshly-biopsied samples, Related to Figure 4. (A) Heatmap showing InferCNV-normalized gene expression in freshly-biopsied tumor epithelial cells compared to normal epithelial cells across chromosomal regions (red = increased expression, blue = reduced expression). Cells are randomly downsampled to 100 cells per tumor. UMAP plots of epithelial cells from freshly-biopsied samples showing (B) sample type, and NUMBAT calculated probability of (C) chromosome 1p loss, (D) chromosome 3p loss, (E) chromosome 4q loss, (F) chromosome 8p loss, (G) chromosome 14q loss, (H) chromosome 1q gain and (I) chromosome 5q gain. See also Figure 4.

Figure S8. Analysis of chromosomal copy number variation (CNV) in cultured samples, Related to Figure 4. Heatmaps showing gene expression in cultured tumor sample epithelial cells compared to normal epithelial cells across chromosomal regions (red = increased expression - predicted copy number gain, blue = reduced expression - predicted copy number loss) for tumors (A) T3, (B) T17 and (C) T20. Each column represents a different chromosomal position and each row represents an individual cell. UMAP plots of cells from cultured samples showing (D) sample type, and InferCNV identification of (E) chromosome 3p loss, (F) chromosome 9p loss, (G) chromosome 14q loss, (H) chromosome 1q gain, (I) chromosome 5q gain, (J) chromosome 8q gain, (K) chromosome 12p gain and (L) chromosome 20q gain. One subclone of T17 exhibited copy neutral loss of heterozygosity of chromosome 3p (E). See also Figure 4.

Figure S9. Analysis of chromosomal copy number and loss of heterozygosity, Related to Figure 4. Stacked bar charts showing clonal and stromal composition of samples from (A) T3 and (B) T17 as determined by scRNA-seq, demonstrating how bulk exome-seq on the same samples could be used to map additional genetic information onto subclones. CNVkit analysis

of chromosomal copy number in bulk exome-seq data from (C) T3A, (D) T3B, (E) T17A, (F) T17B, (G) T17C, (H) T17D, (I) T17E, (J) T17F, (K) T17G, (L) T20 mix of samples A and B. (M) scRNA-seq allele frequency for germline heterozygous SNPs on chromosome 14q in cancer cells from T3A and T3B samples stratified by 14q (loss) status, showing independent loss of opposing alleles. (N) scRNA-seq allele frequency for germline heterozygous SNPs on chromosome 3p in a subcluster of cancer cells from sample T17D that was 3p diploid, demonstrating expression of single alleles and therefore copy neutral loss of heterozygosity. See also Figure 4.

Figure S10. Bulk RNA-seq analysis of HIF-1 and HIF-2 dependent genes in RCC4 cells, Related to Figure 4. (A) Western blots showing HIF-1 α , HIF-2 α , HIF-1 β and β -actin (loading control) protein levels in RCC4 cells treated with control, HIF-1 α or HIF-2 α siRNAs. Volcano plots showing log₂ fold-change and -log₁₀ adjusted p-values for (B) HIF-1a siRNA versus control siRNA and (C) HIF-2a siRNA versus control siRNA in bulk RNA-seq analysis of RCC4 cells. Red denotes genes with an adjusted p-value of <0.05; all other genes colored in black. Box-and-whisker plots showing expression of (D) HIF-1-specific metagene and (E) HIF-2-specific metagene in scRNA-seq analysis of cultured samples, demonstrating induction of both metagenes by hypoxia in proximal tubules, overexpression of both metagenes in ccRCC epithelial cells vs proximal tubules, and specific downregulation of the HIF-2-specific metagene by the HIF-2 inhibitor. **** p<0.0001. Box-and-whisker plots show median, interquartile range and range. See also Figure 4.

Figure S11. Immunostaining of representative module 1, 2 and 3 genes in ccRCC tumor, Related to Figure 5. Immunostaining of (A, B) CUBN, (C, D) CAV1 and (E, F) ICAM1 in two regions of tumor sample 34E. Expression of (G) ACAT1, (H) CDHR5, (I) NAT8, (J) CP, (K) FST3, (L) S100A6, (M) CCL2, (N) JUN, and (O) HSPH1 in ccRCC samples from the Human Protein Atlas. See also Figure 5.

Figure S12. Correlation of gene modules with PAX8 and HNF1b gene programs. (A-E), Related to Figure 5. Scatter plots showing mean expression of composite metagenes based on Module 1,2 and 3 genes and the 100 most PAX8- and HNF1b-dependent genes from Patel et al in tumor cell clusters from fresh biopsy samples. (G) Biplot showing PCA analysis of expression of these 5 gene modules across the individual tumor clusters. The arrows represent the original axes for the 5 modules in the first two principal components. See also Figure 5.

Figure S13. Gene set enrichment analysis of genes correlating with modules 1, 2 and 3, Related to Figure 5. (A) Heatmap showing enrichment scores for Hallmark pathway genes amongst genes correlating with Module 1, 2 and 3 genes in each of the 8 ccRCC tumors studied. Pathways are ordered by the average Module-1-correlating enrichment score across tumors. (B) Heatmap showing enrichment scores for Barkley genesets amongst genes correlating with Module 1, 2 and 3 genes. Genesets are ordered by the average Module-1-correlating enrichment scores across tumors. See also Figure 5.

Figure S14. Heatmap showing heterogenous transcriptional programs in individual tumors, Related to Figure 5. SCENIC analysis was used to identify transcription factors whose activity exhibited significant intratumor heterogeneity (dark blue) in each of the 8 tumors. See also Figure 5.

Figure S15. Expression of Module 1, 2 and 3 genes in pRCC and oncocytoma tumors, Related to Figure 7. UMAP plots showing (A) cell clusters, and (B) Module 1, (C) Module 2

and (D) Module 3 metagene expression in pRCC (T22) epithelial cells. (E-G) Box-and-whisker plots showing T22 cluster-level composite module scores for modules 1, 2 and 3 in clusters arranged in decreasing order of median Module 1 score. UMAP plots showing (H) cell clusters, and (I) Module 1, (J) Module 2 and (K) Module 3 metagene expression in oncocytoma (T23) epithelial cells. (L-M) Box-and-whisker plots showing T23 cluster-level composite module scores for modules 1, 2 and 3 in clusters arranged in decreasing order of median Module 1 score. Box-and-whisker plots show median, inter-quartile range and range. See also Figure 7.