Supporting Information

Supplementary text

Validation of cell-of-origin heuristic

Using our heuristic framework (Fig. 3*A*), we revisited known markers reported as being differentially expressed in RO. A canonical marker of the distal nephron, *SLC4A1*, would be expected to be upregulated in RO simply by cell-of-origin and consistent with this it was classified as T>N, P<D (Fig. S7*A*). Other genes in this same T>N, P<D category (apparent tumor upregulation confounded by cell-of-origin) include *KIT*, a widely accepted marker of RO (1), and genes involved in mitochondrial biogenesis such as *PPARGC1α* and the AMPK pathway genes (*PRKAG2*, and *ACACB*). In the opposite T<N, P>D category (apparent tumor downregulation confounded by cell-of-origin) we find *HNF4α*, which was spotlighted in a recent RNA-seq pathway analysis (Fig. S7*A*) (2). Using IHC we validate preferential expression of KIT in normal distal tubules and HNF4α in proximal tubules – hence confirming the RO expression pattern is attributable to cell-of-origin effects (Fig. S7*B*). Genes encoding OXPHOS subunits and their assembly factors revealed largely ambiguous tumor versus normal differences after accounting for cell-of-origin (Fig. S8).

Supplementary Materials & Methods

Renal cell carcinoma exome data analysis

Exome data from three previously published non-clear cell RCC subtypes were obtained with permission from the European Genome-phenome Archive. Exome fastq files were aligned to GRCh37 using BWA version 0.7.10. All tumor-normal pairs with mean mtDNA coverage >20X were included in analyses. Aligned exome data from ccRCC

were obtained with permission from the TCGA using the CGhub BAMslicer, and included in analysis were all tumor-normal pairs sequenced with paired-end reads at Baylor College of Medicine with mean mtDNA coverage>40x.

Analysis of mtDNA from exomes

Somatic mutations were identified using the following seven filters: (1) VAF≥0.05 in tumor; (2) VAF≤0.05 in adjacent normal; (3) if variant detected at low VAF in adjacent normal (i.e. 0<VAF≤0.05), then tumor VAF≥0.25; (4) Variant supported by >5 reads in tumor; (5) Depth >15x in tumor; (6) Depth >10x in adjacent normal; (7) No more than two alleles observed in tumor or normal (i.e. multiple alleles observed at a site excluded). Depth filters were derived from published power calculations. All potential loss-of-function variants were manually reviewed to remove artifacts, leading to removal of artifacts at position MT:14501, 14504, and 14440 in ccRCC (mapping errors with NUMTs). Indel callers left-aligned mutations in cases where homopolymeric tracts were involved. Deletions were searched for using the STAR aligner to identify split reads.

mtDNA copy number analysis

Briefly, 10 ng of DNA was loaded into 20 μL reactions with a custom synthesized assay to simultaneously quantitate nuclear and mitochondrial DNA copy number. Nuclear DNA copies were estimated using a Taqman assay against the AluYb8 repeat element consisting of forward primer 5'-CTTGCAGTGAGCCGAGATT-3', reverse primer 5'-GAGACGGAGTCTCGCTCTGTC-3', and the probe 5'-VIC-ACTGCAGTCCGCAGTCCGGCCT-MGBNFQ-3'. Mitochondrial copies were estimated with a Taqman assay against *MT-ND2* consisting of forward primer 5'-TGTTGGTTATACCCTTCCCGTACTA-3', reverse primer 5'-CCTGCAAAG-

ATGGTAGAGTAGATGA-3', and the probe 5'-6FAM-CCCTGGCCCAACCC-MGBNFQ-3'. An estimate of mtDNA copies per cell was calculated using a calibration of *MT-ND2* and AluYB8 PCR Ct values to dilution ladders of a chemically synthesized primer and total DNA extracted from human cells respectively.

Tumor/normal RNA-seq analysis

Sequencing reads were quality trimmed using Trimmomatic with parameters LEADING:20 TRAILING:20 MINLEN:23. They were then aligned using STAR to the human genome (hg19) with Gencode 19 transcript annotations. Gene-level read counts were obtained with HTSeq. Differential expression analysis between tumor and normal samples was performed in R using the DESeq2 package (3) with the design formula ~ pair + tumor/normal.

Metabolite profiling

Free fatty acids, bile acids, and metabolites of intermediate polarity were profiled using a Nexera X2 U-HPLC (Shimadzu, Marlborough, MA) coupled to a Q Exactive orbitrap mass spectrometer (Thermo Fisher Scientific; Waltham, MA). Homogenates (30 µL) were extracted using 90 µL of methanol containing PGE2-d4 as an internal standard (Cayman Chemical Co.; Ann Arbor, MI) and centrifuged (10 min, 9,000 x g, 4°C). The supernatants (2 µL) were injected onto a 150 x 2 mm ACQUITY BEH C18 column (Waters; Milford, MA). The column was eluted isocratically at a flow rate of 400 µL/min with 60% mobile phase A (0.1% formic acid in water) for 4 minutes followed by a linear gradient to 100% mobile phase B (acetonitrile with 0.1% acetic acid) over 8 minutes. MS analyses were carried out in the negative ion mode using electrospray ionization, full scan MS acquisition over 70-850 m/z, and a resolution setting of 70,000.

Other MS settings were: sheath gas 45, sweep gas 5, spray voltage -3.5 kV, capillary temperature 320°C, S-lens RF 60, and heater temperature 300°C. Raw data were processed using Progenesis QI software (NonLinear Dynamics) for feature alignment, nontargeted signal detection, and signal integration. Targeted processing of a subset of known metabolites was conducted using TraceFinder 3.3 software (Thermo Fisher Scientific; Waltham, MA). Compound identities were confirmed using reference standards and reference samples. CysGly and γGluCys dipeptide levels were evaluated by re-interrogating full scan metabolomics data, using a retention time standard for CysGly (Sigma catalog #C0166) and γGluCys (Sigma catalog #G0903). Metabolites were considered detected if found in at least 20% of samples (i.e. 4/17). Data for known metabolites were normalized by sample to column sum after adding a minimum value of 100 to empty cells. Fold-changes between the average of tumors and normals per metabolite were then calculated.

Immunohistochemistry

FFPE tissue sections were baked, deparaffinized, and rehydrated. Sections were incubated 10 min in 3% peroxide/methanol and then heat-induced epitope retrieval using either 0.5M EDTA (pH 8.0) or Borg Decloaker (BioCare Medical, BD1000, pH 9.5) was done. Sections were blocked and incubated overnight in primary antibody. Primary antibody dilutions and antigen retrieval conditions were as follows: P-AMPK (1:50, Borg Decloaker RTU 45 min), HNF4α (1:50, sodium citrate 30 min), MT-CO1 (1:500, EDTA 10 min), NDUFB8 (1:100, Borg 45 min), CD117 (RTU pretreatment ER2 [EDTA pH 9.0 20 min]), and cyclin D1 (RTU pretreatment ER2 [EDTA pH 9.0 20 min]). MT-CO1, and NDUFB8 sections were then put in biotinylated secondary antibody (30 min

RT; Vector Laboratories, BA-2000) and developed per Vector protocol (SK4100). All sections were counterstained with Hematoxylin 2, dehydrated, cleared, and coverslipped per standard protocol. Sections were then qualitatively evaluated in a blinded fashion by pathology review according to the following scoring system: complete absence of staining (score 0), mild positivity (score 1), moderate positivity (score 2), and strong positivity (score 3). In regard to cyclin D1 staining, the ratio of DAB/nuclear positivity was measured using the ImmunoRatio application (4). Tumors with >80% cyclin D1 nuclear positivity were considered to have strong expression; tumors with 30-80% nuclear positivity had moderate expression, and tumors with <30% nuclear positivity had absent expression. For cell-of-origin morphological interpretation, compared to the proximal tubule, the distal tubule had decreased cytoplasmic volume with a higher nuclear to cytoplasmic ratio, an absent brush border, and a clean tubular lumen. As we were unable to discriminate between the cortical thick ascending limb (cTAL), distal convoluted tubule (DCT), connecting tubule (CNT), and cortical collecting duct (CCD) by IHC, all of these nephron segments were considered to collectively represent distal tubules for our cell-of-origin analysis.

Statistical analysis

Differences in LoF mutation frequency by complex across different tumors was statistically tested using N-1 Chi-square test. Differences in gene expression between tumor and normal samples were tested for statistical significance using the Wald test and *P* values were corrected for multiple hypothesis testing using the method of Benjamini and Hochberg, as implemented in DESeq2. Nominal *P* values are reported for all metabolite comparisons.

Supplementary Figures

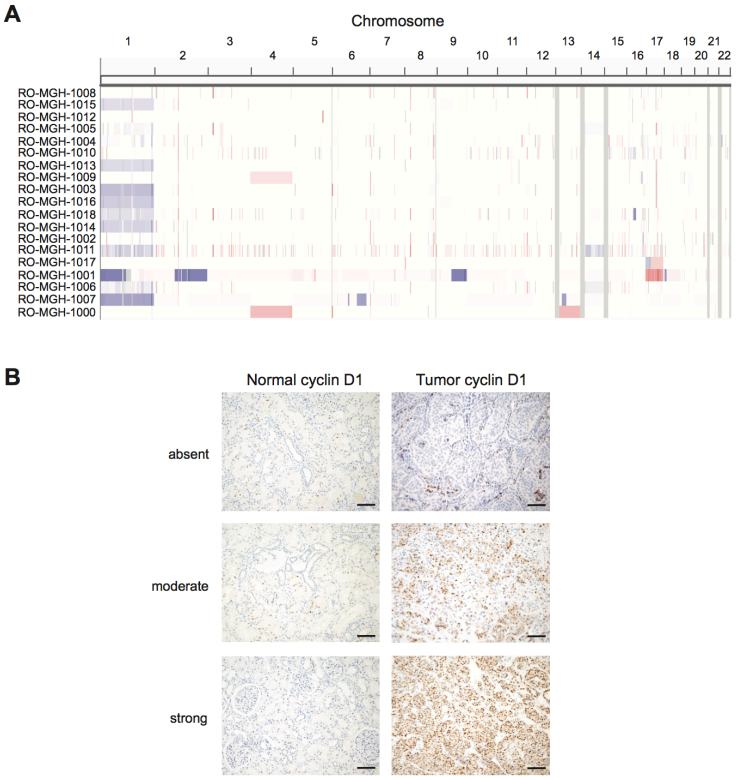

Fig. S1. Chromosome 1 loss and cyclin D1 expression in RO. (*A*) Copy number analysis by chromosome in RO^{MGH}. Chromosome numbers are indicated at the top of the figure. Blue indicates loss of chromosome material and red indicates gain of chromosome material. Each row corresponds to a patient from the RO^{MGH} cohort (N=19 normal kidney; N=19 tumors). (*B*) Representative IHC for cyclin D1 in matched normal kidney (N=19) and tumors (N=19) in RO^{MGH} with staining classification to the left of each picture. Samples were classified as having absent (<30% nuclei positive), moderate (30-80% nuclei positive), or strong (>80% nuclei positive) expression based on % of nuclei positive for cyclin D1 using the ImmunoRatio application. Images were captured at 200x magnification. Scale bar represents 50 μM.

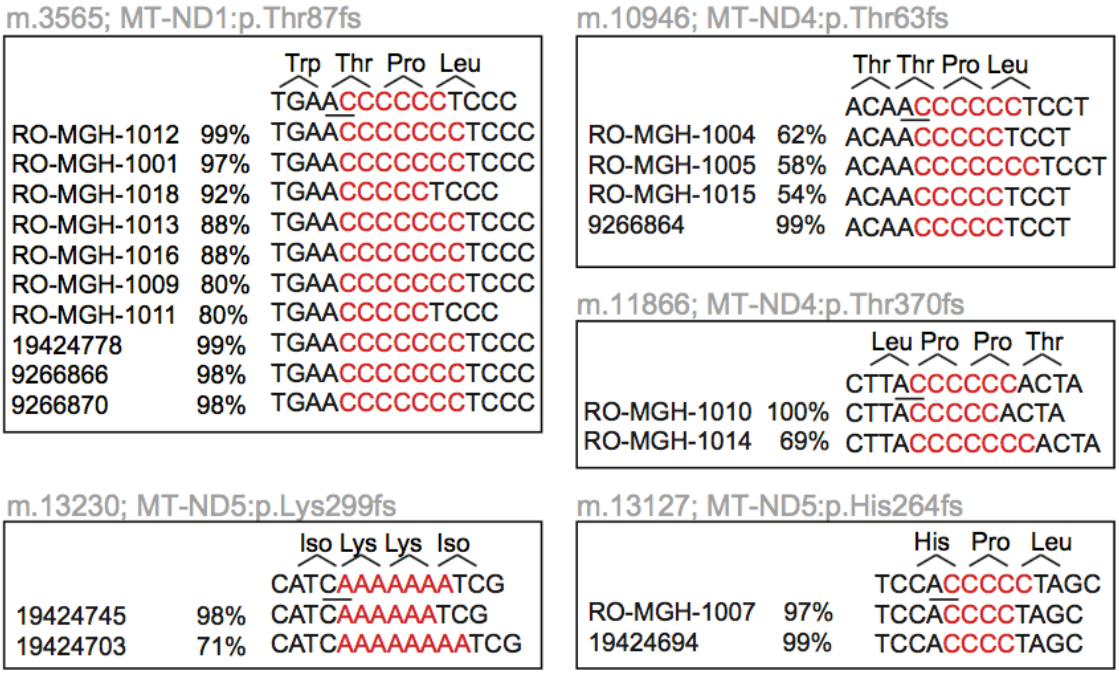

Fig. S2. Recurrent frameshift mutations in homopolymeric mtDNA tracts. Each box highlights a recurrent frameshift mutation with gray text indicating the mtDNA coordinate, gene, and amino acid position involved. Within each box, sample ID is shown for each cohort (RO^{MGH}, N=19; RO^{GEN}, N=16 samples), followed by VAF%, and the tumor mtDNA sequence including the indel event. Single nucleotide homopolymers are shown in red. Flanking DNA sequences are shown in black. The reference mtDNA sequence is shown at the top of each box with the mutation underlined and the involved amino acids highlighted by codon. Indel callers left-aligned mutations in homopolymeric tracts.

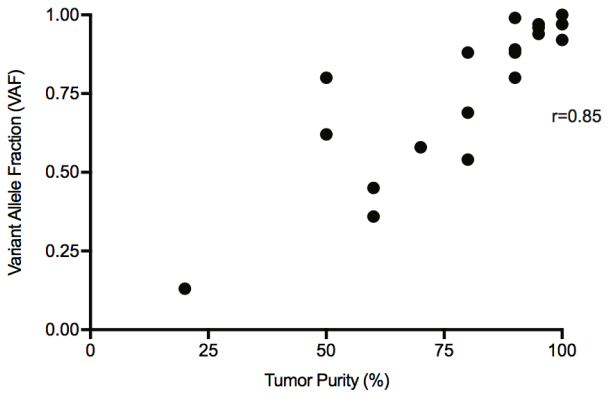

Fig. S3. Variant allele fraction (VAF) for complex I mtDNA mutations directly correlate with tumor purity. Tumor purity was estimated by pathologists for all N=19 RO^{MGH} samples as the amount of tumor versus non-tumor tissue in a sample. VAF of mtDNA mutations as determined from WES reads is plotted on the y-axis. Each black circle corresponds to a complex I mutation in a patient. Correlation coefficient (r) is indicated in the graph.

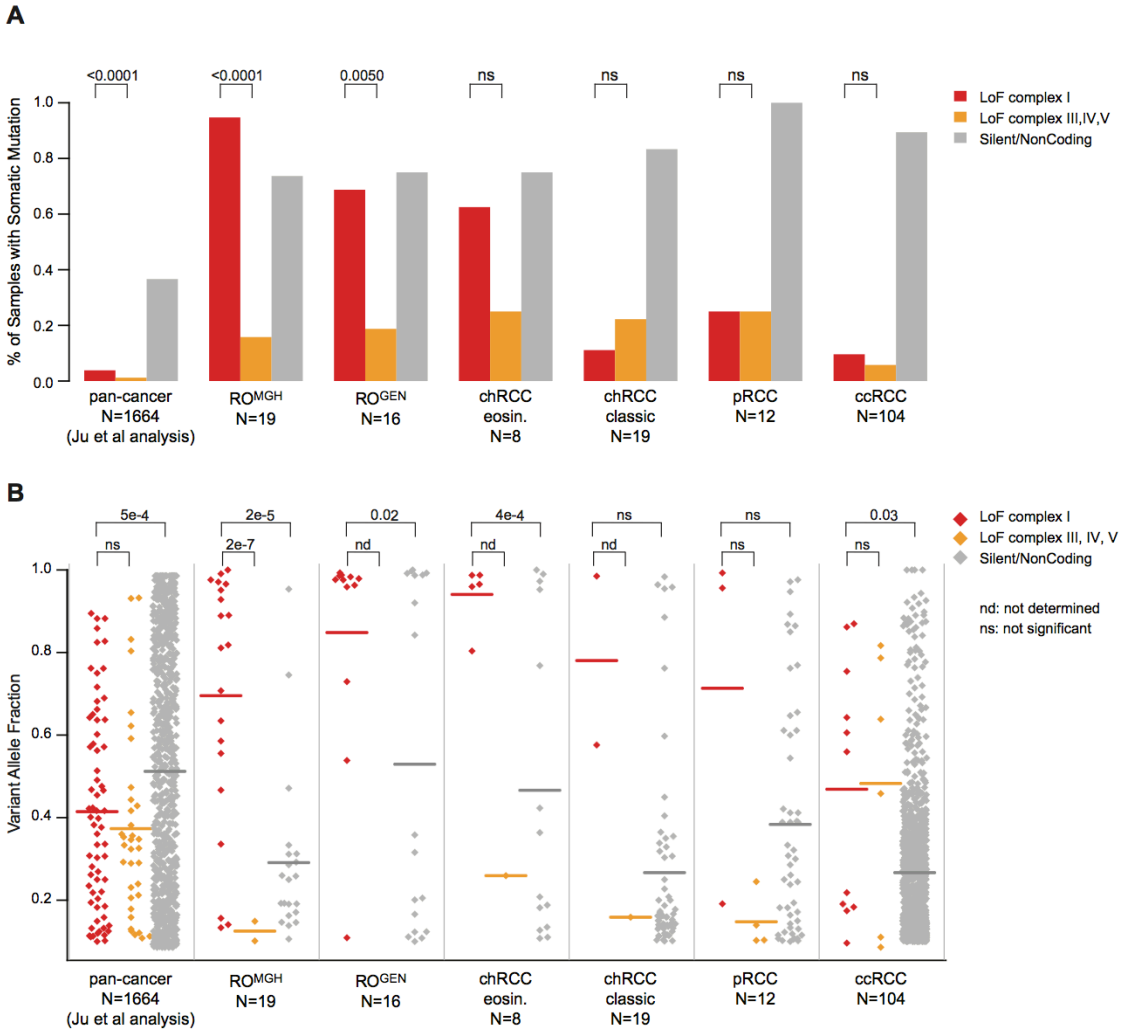

Fig. S4. LoF mtDNA mutations in kidney cancers compared to a pan-cancer analysis. (*A*) The percentage of samples with somatic mtDNA mutations VAF≥0.1 are shown for a pan-cancer analysis by Ju et al., and for all kidney tumors evaluated in this study. Mutations are categorized as LoF in complex I (red), LoF in complex III, IV, or V (orange), and silent/noncoding (gray). *P* values from N-1 Chi-square test; ns indicates $P > 0.05$. (*B*) VAF for mtDNA mutations are shown by study with the same color scheme as in panel *A*. *P* values from two-sided *t*-test; nd indicates not determined due to small sample size; ns indicates $P > 0.05$; horizontal bars indicate mean.

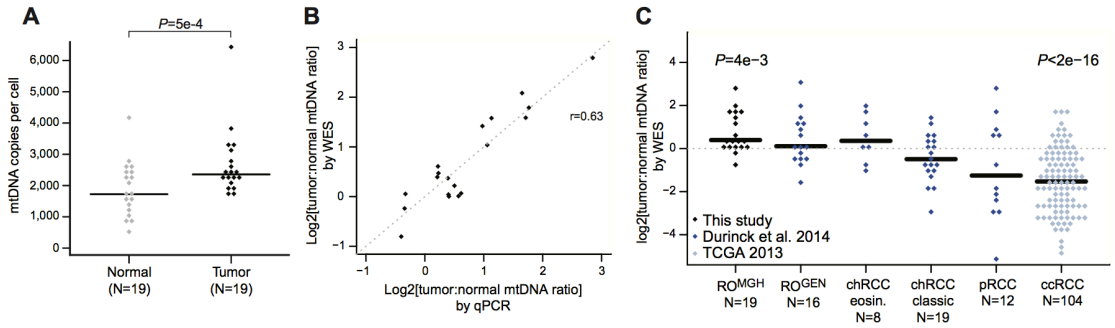

Fig. S5. mtDNA copy number in RO and RCC. (*A*) mtDNA copy number measured in RO^{MGH} samples (N=19 normals; N=19 tumors) via multi-plexed quantitative PCR (single technical replicate) using primers against *MT-ND2* and the AluYB8 nuclear DNA repeat element. *P* value from Wilcoxon paired rank sum test with horizontal bars indicating the median. (*B*) Correlation of tumor versus normal mtDNA copy number ratio estimated from quantitative PCR and WES data, using ratio of mean coverage in mtDNA versus nuclear DNA, for all 19 RO^{MGH} samples. Correlation coefficient (r) indicated in the graph. (*C*) Relative tumor/normal mtDNA copy number ratio from WES reads for kidney tumors. *P* values are from a two-sided *t*-test to see if cohort mean (horizontal bars) was significantly different than 0.

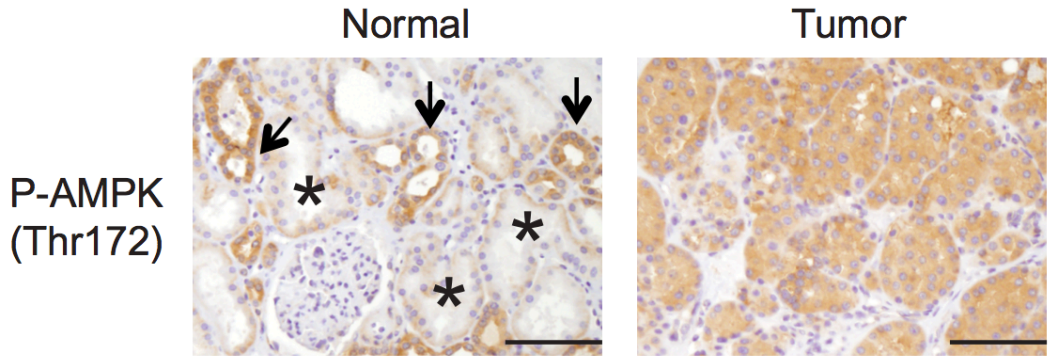

Fig. S6. Immunohistochemical analysis of normal kidney and renal oncocytoma (RO). Representative pictures from our RO^{MGH} cohort of normal kidney (N=19) and tumors (N=19) for P-AMPK Thr172 primary antibody. Each * indicates a proximal tubule and each arrow indicates a distal tubule (not all proximal and distal tubules in each image are labelled). Images were captured at 200x magnification with scale bar of 50 μM.

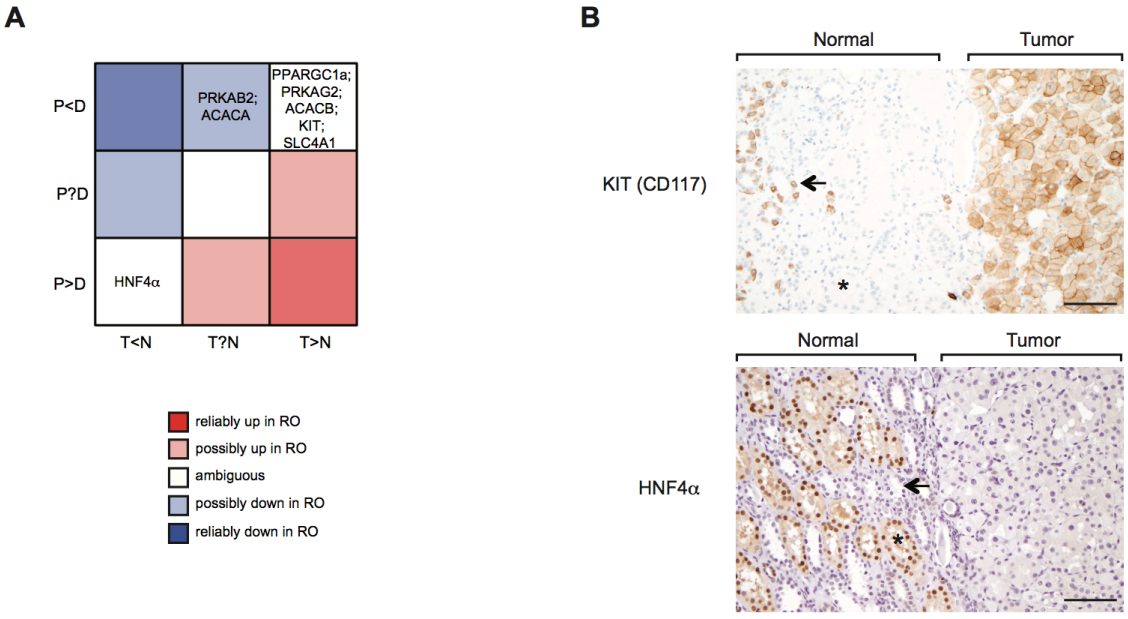

Fig. S7. Analysis of differentially expressed markers with respect to cell-of-origin. (*A*) Grid highlights genes previously reported to be differentially expressed by RNA-seq in RO versus normal kidney (*PPARGC1α*, *PRKAG2*, *PRKAB2*, *ACACA*, *ACACB*, *HNF4α*) or to be a classic marker of RO by IHC (*SLC4A1*, *KIT*). Previously reported genes are placed in their gene category with respect to cell-of-origin and differential expression in tumors versus normals. (*B*) IHC stains from representative samples from RO^{MGH} (N=19 normal kidney; N=19 tumors) showing tumor and adjacent normal kidney (indicated with brackets above each image). Primary antibodies used in the stain are indicated to the left of each picture. Each * indicates a representative proximal tubule and each arrow indicates a representative distal tubule in each image (not all proximal and distal tubules in normal kidney are highlighted). Images were captured at 200x magnification. Scale bar represents 50 μM.

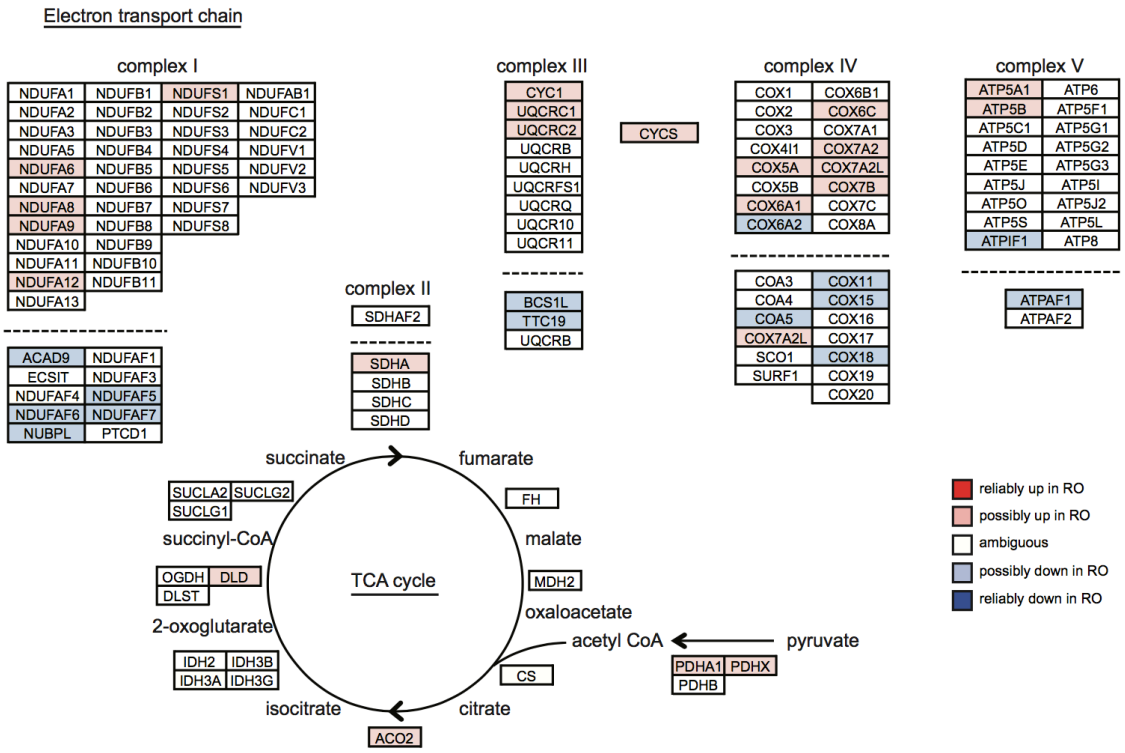

Fig. S8. Expression of OXPHOS and TCA cycle transcripts in tumors with consideration for cell-of-origin. OXPHOS subunits and assembly factors are shown by complex (separated by dashed lines), with background color indicating transcript expression based on our heuristic algorithm. TCA cycle enzymes and metabolites are also shown below complex II genes.

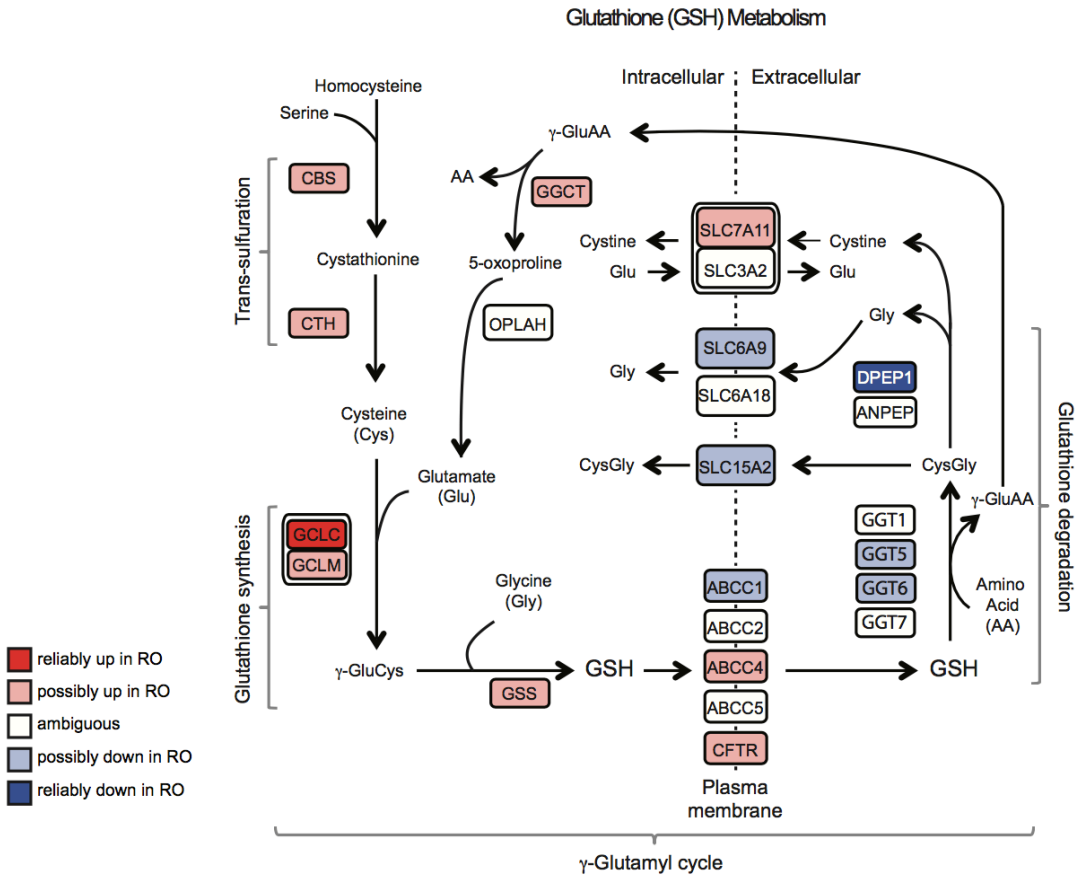

Fig. S9. Expression of transporters involved in the γ-glutamyl cycle with consideration for cell-of-origin. Pathway diagram for glutathione metabolism including transporters involved in glutathione export and substrate reuptake. Each gene is colored according to the figure legend. No transporter is drawn for γ-glutamyl-amino acid (γ-GluAA).

References

1. Liu L*, et al.* (2007) Immunohistochemical analysis of chromophobe renal cell carcinoma, renal oncocytoma, and clear cell carcinoma: an optimal and practical panel for differential diagnosis. *Archives of pathology & laboratory medicine* 131(8):1290-1297.
2. Joshi S*, et al.* (2015) The Genomic Landscape of Renal Oncocytoma Identifies a Metabolic Barrier to Tumorigenesis. *Cell reports* 13(9):1895-1908.
3. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15(12):550.
4. Tuominen VJ, Ruotoistenmaki S, Viitanen A, Jumppanen M, & Isola J (2010) ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast cancer research : BCR* 12(4):R56.