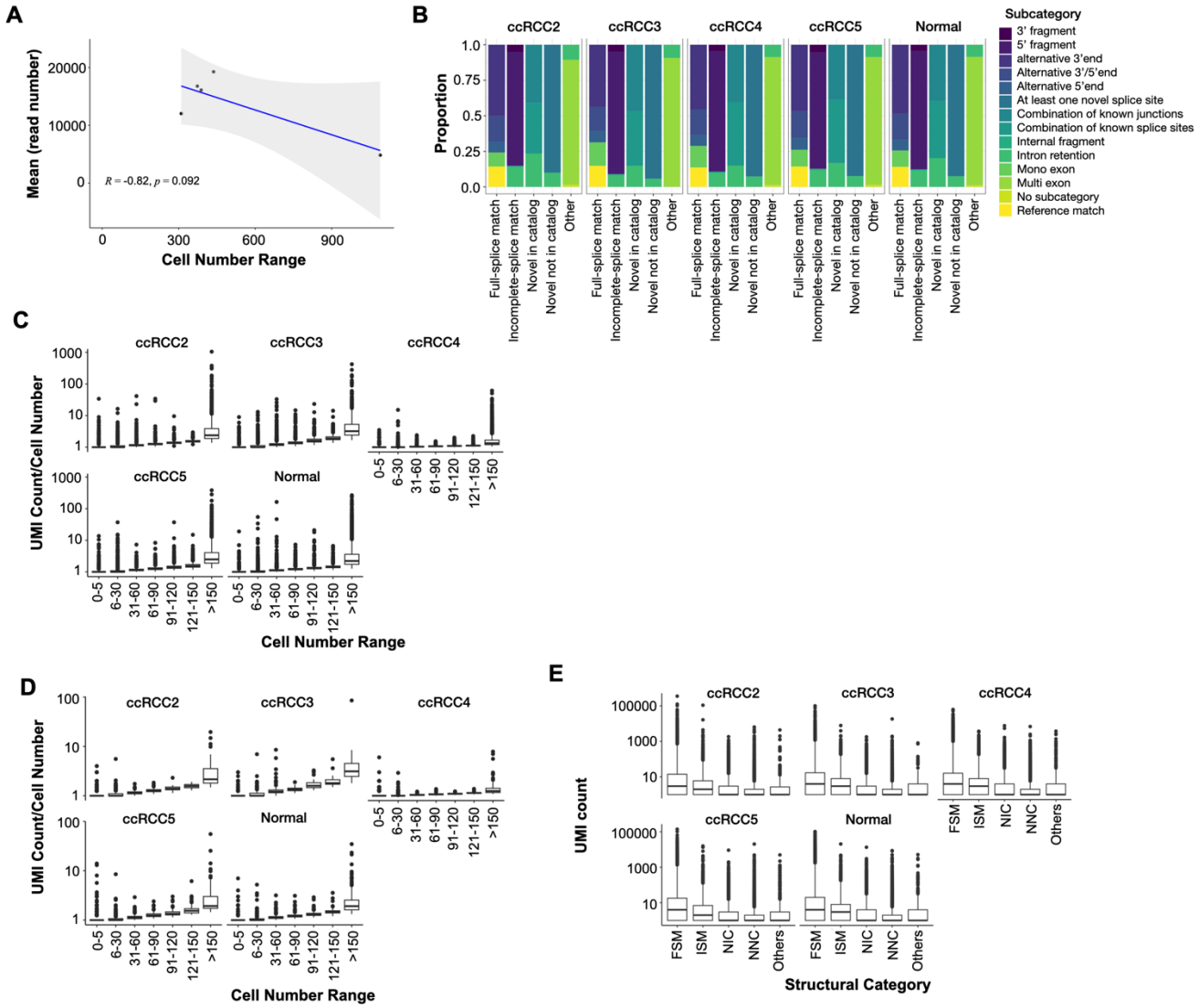
# Heterogeneous and novel transcript expression in single cells of patient-derived ccRCC organoids

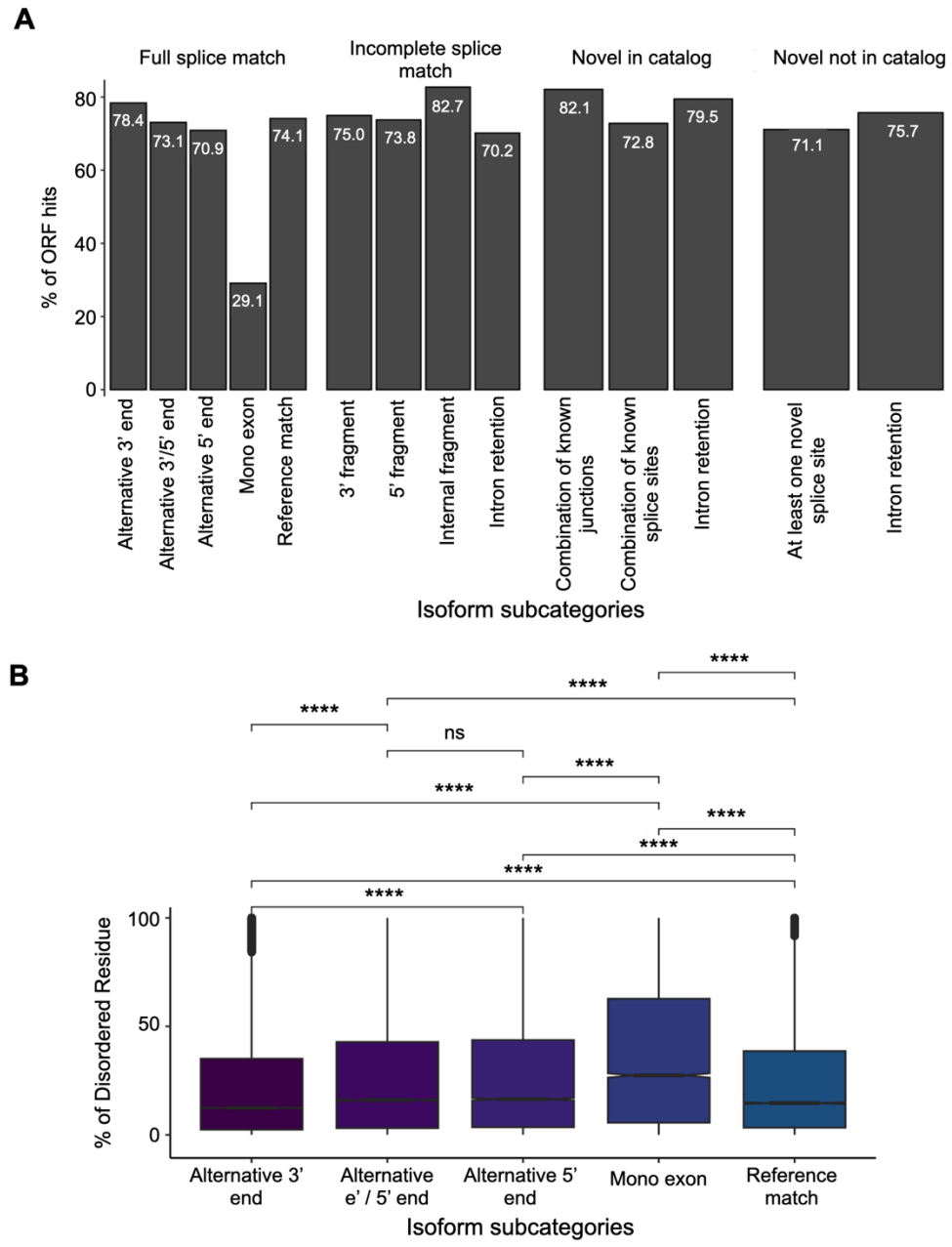
Tülay Karakulak[1,2,3,*], Natalia Zajac[3,4,*],Hella Anna Bolck[2,5], Anna Bratus-Neuenschwander[4], Qin Zhang[4], Weihong Qi[3,4], Debleena Basu[2], Tamara Carrasco Oltra[4,5], Hubert Rehrauer[3,4], Christian von Mering[1,3], Holger Moch[2], Abdullah Kahraman[3,6,**]

1 Department of Molecular Life Sciences, University of Zurich,
2 Department of Pathology and Molecular Pathology, University of Zurich and University Hospital Zurich
3 Swiss Institute of Bioinformatics, Lausanne
4 Functional Genomics Center Zurich, ETH, Zurich,
5 Centre for AI, School of Engineering, Zurich University of Applied Sciences (ZHAW), Technikumstrasse 71, 8400 Winterthur, Switzerland
6 School for Life Sciences, Institute for Chemistry and Bioanalytics, University of Applied Sciences Northwestern Switzerland, Muttenz

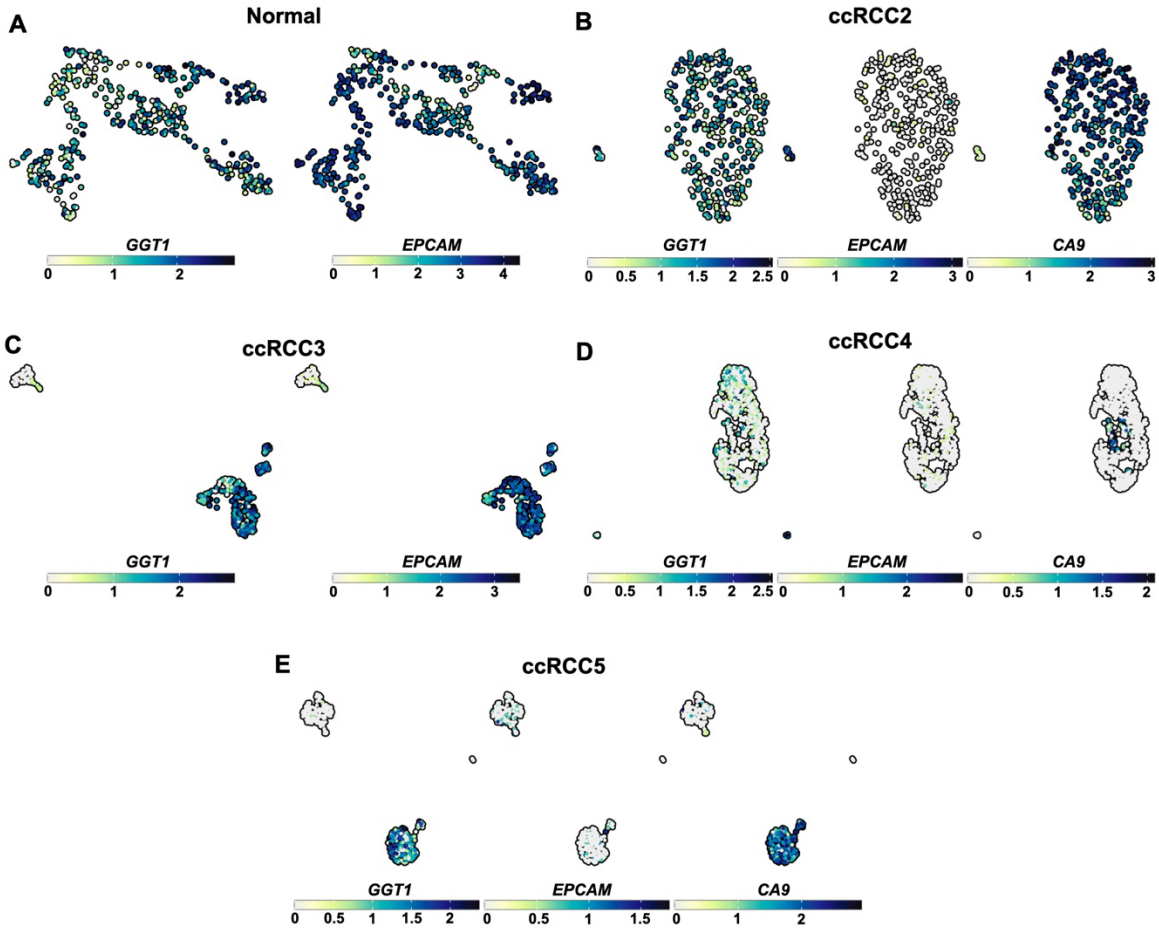
*shared first author
**corresponding author, abdullah.kahraman@fhnw.ch



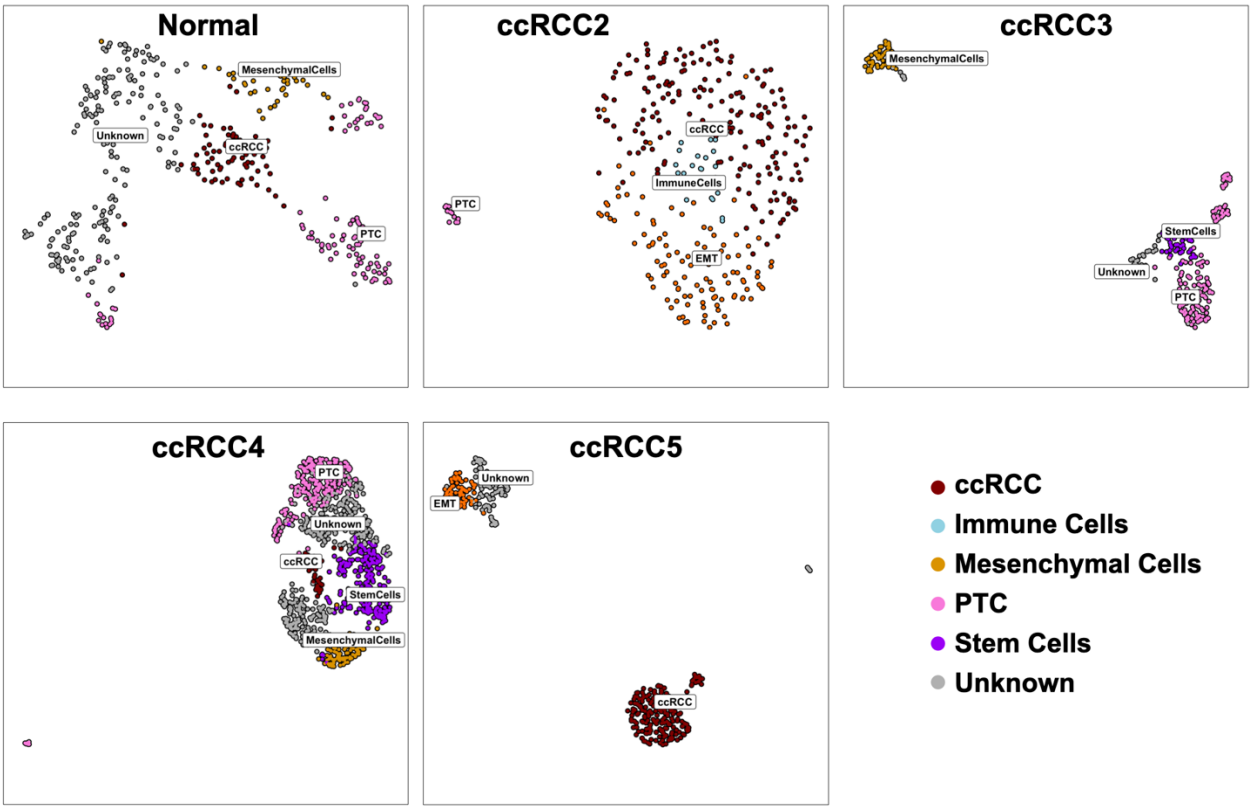

**Supplementary Figure 1: General statistics on transcripts: (A)** Pearson Correlation between number of cells and mean of read number on the log2 scale (R=-0.82, p=0.092). **(B)** Proportions of transcript subcategories within main structural categories for each sample. The x-axis represents the main structural categories as defined by SQANTI3 classification. Each bar graph depicts the fraction of the corresponding subcategories. **(C)** Distribution of UMI counts/Cell Number of known (FSM and ISM) transcripts across varying cell number ranges. **(D)** Distribution of UMI counts/Cell Number of novel transcripts (NIC and NNC) across varying cell number ranges. **(E)** Distribution of UMI counts across different structural categories of each sample. The Y-axis is on log10 scale in C, D, and normalised by the number of cells a transcript found in.



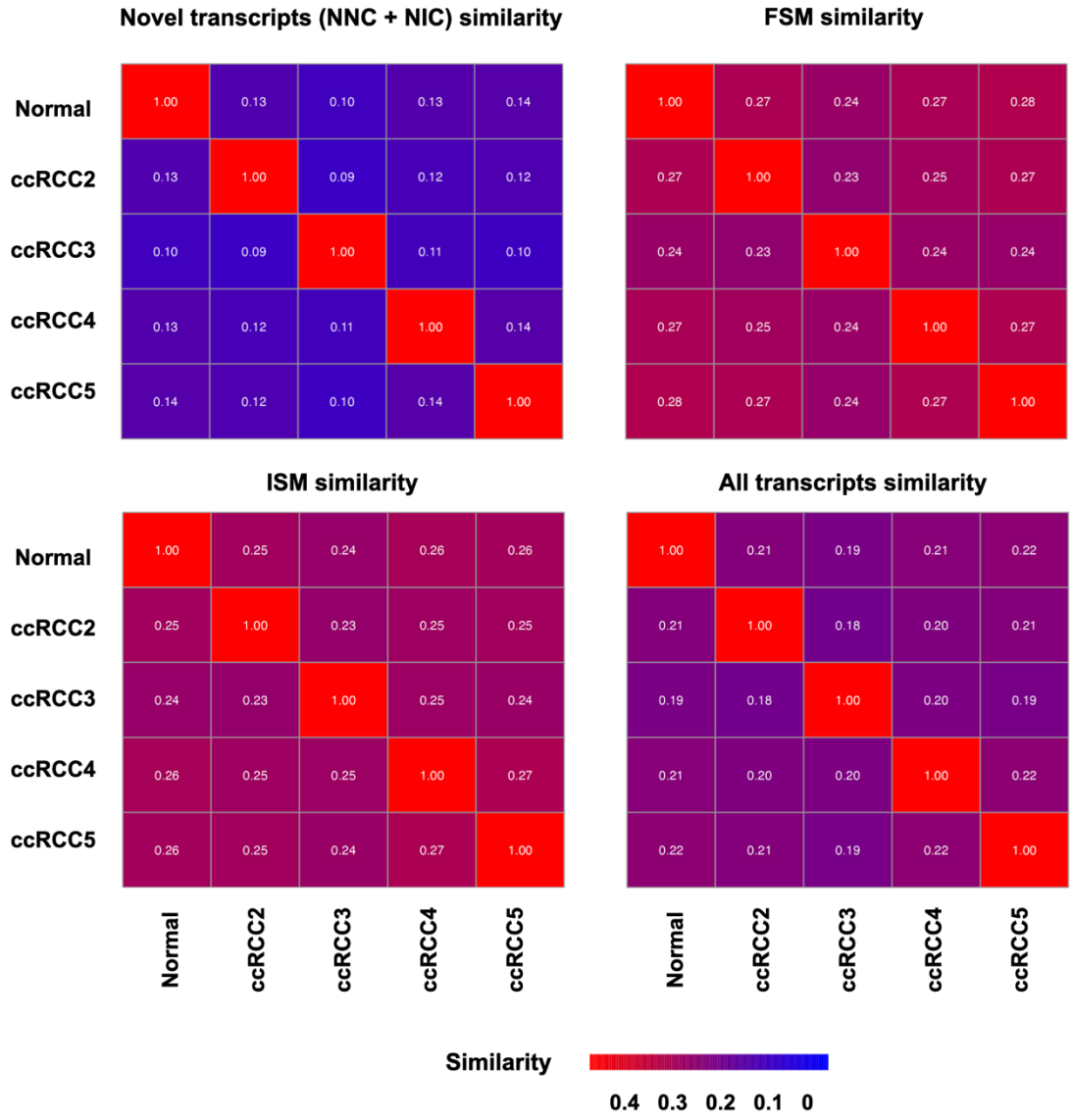

**Supplementary Figure 2: (A)** Percentage of ORF hit across isoform subcategories **(B)** Percentage of disordered regions across FSM subcategories.



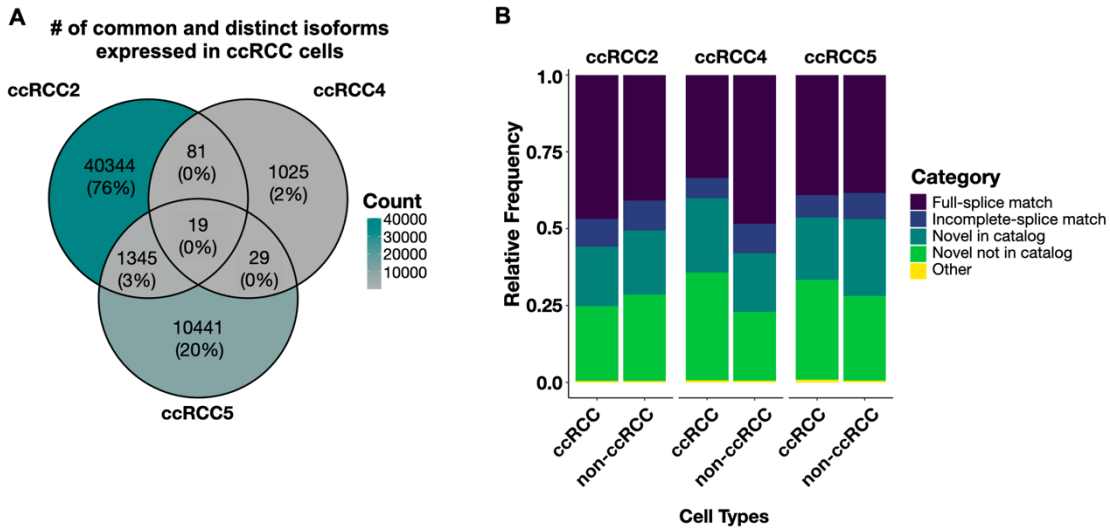

**Supplementary Figure 3:** Expression Pattern of *GGT1*, *EPCAM*, *CA9* in (A) Normal PDO, (B) ccRCC2 PDO, (C) ccRCC3 PDO, (D) ccRCC4 PDO and (E) ccRCC5 PDO.



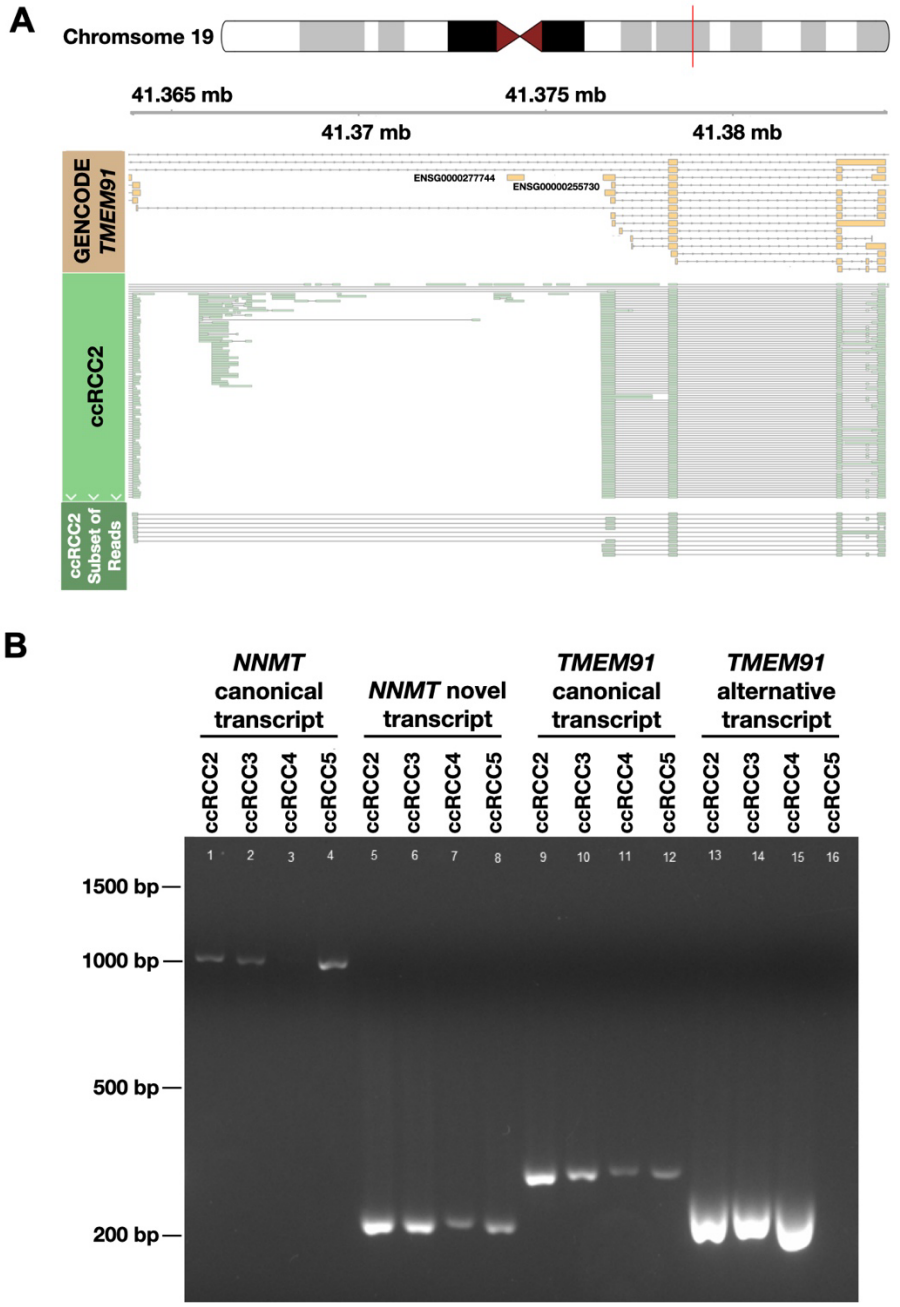

**Supplementary Figure 4:** Cell-type annotation of each sample with manually curated list of genes. PTC: Proximal tubule cells, EMT: Epithelial-Mesenchymal Transition.




**Supplementary Figure 5: Transcript Similarity and Cell Prevalence in Samples:** Heatmaps display the Jaccard similarity index among samples for different categories of transcripts. Top left: Novel In catalog and novel not in catalog transcripts (NIC and NNC, respectively) similarity; top right: Full Splice Match (FSM) similarity; bottom left: Incomplete Splice Match (ISM) similarity; bottom right: Overall similarity including all transcripts, with the criterion that transcripts must be present in at least three cells.




**Supplementary Figure 6: (A)** Number of overlapping isoforms explicitly in ccRCC cells. **(B)** Proportion of explicitly expressed isoforms either in ccRCC or non-ccRCC cells across ccRCC2, ccRCC4, ccRCC5.




**Supplementary Figure 7: (A)** *TMEM91* novel isoform reads. 520 PacBio reads were found for the transcripts. **(B)** PCR Validations of Isoforms in ccRCC Tumor Samples**:** Agarose gel (2%) electrophoresis image of PCR products amplified with common or isoform specific primers (described in supplementary table 1 & 2)**.** All lanes are marked with corresponding tumor sample names. Lane 1-4: *NNMT* Canonical Isoform amplified with *NNMT*_Canonical_Fp and *NNMT*_Common_Rp1, lane 5-6: *NNMT* Novel Isoform amplified with *NNMT*_Novel_Fp and *NNMT*_Common_Rp2, lane 9-12: *TMEM91* Canonical Isoform amplified with *TMEM91*_Common_Fp and *TMEM91*_Common_Rp, lane 13-16: *TMEM91* Alternative Isoform amplified with *TMEM91*_Novel_Fp and *TMEM91*_Common_Rp. The novel

transcript included five exons without exon 2. In both PCR and Sanger sequencing, we could confirm the exon 1 exon 3 junction of transcript ENST00000413014. Note that the target novel isoform was lost after filtering, which may explain its absence in the PCR results.

**Supplementary Table 1: Details of Primers used in PCR Validation**

| Target | Primer specification | Primer name | Sequence |
|---|---|---|---|
| *NNMT* | Specific forward primer against unique sequence of canonical isoform | *NNMT*_Canonical_Fp | TCAAGTGCTCCCTCTGGTCT |
| | Specific forward primer against unique sequence of novel isoform | *NNMT*_Novel_Fp | TGGTGTCTACTTCTTGGCTTTTG |
| | Reverse primer against sequence shared between canonical and novel isoforms | *NNMT*_Common_Rp1 | ACCGCCTGTCTCAACTTCTC |
| | Reverse primer against sequence shared between canonical and novel isoforms | *NNMT*_Common_Rp2 | AGTAGGTGGGGAGGTCTGG |
| *TMEM91* | Forward primer against shared sequence between canonical and novel isoforms | *TMEM91*_Common_Fp | AGACCCGCGTAGAGCAAAG |
| | Specific forward primer against unique sequence of novel isoform | *TMEM91*_Novel_Fp | GGGCTTGACTGCTTCTTTTC |
| | Reverse primer against sequence shared between canonical and novel isoforms | *TMEM91*_Common_Rp | CTCGGCAAAGGCTATCTCTC |

**Supplementary Table 2: Details of Anticipated and Observed Outcome of PCR Validations Using Different Primer Pairs**

| Target | | Primer Combination | Anticipated Outcome | Observed Outcome |
|---|---|---|---|---|
| *NNMT* | *NNMT* Canonical | *NNMT*_Canonical_Fp | Single PCR product corresponding to canonical isoform | Matched with anticipated outcome |
| | | *NNMT*_Common_Rp1 | | |
| | *NNMT* Novel | *NNMT*_Novel_Fp | Single PCR product corresponding to novel isoform | Matched with anticipated outcome |
| | | *NNMT*_Common_Rp2 | | |
| *TMEM91* | *TMEM91* Common | *TMEM91*_Common_Fp | Single PCR products corresponding to both canonical and novel isoforms | Matched with anticipated outcome |
| | | *TMEM91*_Common_Rp | | |
| | *TMEM91* Novel | *TMEM*_Novel_Fp | Single PCR product corresponding to novel isoform | Single PCR product but with shorter length |
| | | *TMEM*_Common_Rp | | |

**Supplementary Table 3: Comparison of cMDT calculations and Acorde results.** The table shows the overlap of upregulated isoforms in non-ccRCC cells of ccRCC2 found by Acorde and MDTs in non-ccRCC cells of ccRCC2. Isoforms found by two methods are highlighted in bold.

| cMDT Tama ID | Gene Name | Cell Number | MDT in Normal | Cohort | Acorde Result |
|---|---|---|---|---|---|
| G14020.26 | ENSG00000237550 | 4 | **G14020.22** | ccRCC2 | **G14020.22** |
| G16060.56 | *CYBA* | 1 | **G16060.25** | ccRCC2 | **G16060.25** |
| G42843.68 | *JPX* | 4 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.1 | *JPX* | 3 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.75 | *JPX* | 3 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.107 | *JPX* | 1 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.127 | *JPX* | 1 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.145 | *JPX* | 1 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.148 | *JPX* | 1 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.159 | *JPX* | 1 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |

| | | | | | |
|---|---|---|---|---|---|
| G42843.212 | *JPX* | 1 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.35 | *JPX* | 1 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.44 | *JPX* | 1 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.56 | *JPX* | 1 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.64 | *JPX* | 1 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |
| G42843.81 | *JPX* | 1 | G42843.13, G42843.130, **G42843.150**, G42843.23, G42843.25 | ccRCC2 | **G42843.150** |

# Supplementary Method

### Generation and Characterization of ccRCC Patient-Derived Organoid Samples

Surgically resected renal tissue was reviewed by a pathologist with specialization in uropathology (Holger Moch) and suitable specimens were stored at 4 °C in transport media (RPMI (Gibco) with 10 % fetal calf serum (FCS, Gibco) and Antibiotic-Antimycotic® (Gibco)). For organoid derivation, tissue specimens were processed within 24 hours by rinsing them once with PBS followed by finely cutting and digesting them in 0.025 mg/ml Liberase (Roche) for 15 min at 37 °C. The slurry was passed through a 100 µm cell strainer and centrifuged at 1000 rpm for 5 min. Cells were washed once with PBS and erythrocytes were lysed in ACK buffer (150 mM $NH_4Cl$, 10 mM $KHCO_3$, 100 mM EDTA) for 2 min at room temperature. After a final wash with PBS, appropriate amounts of cell suspension were resuspended in CK3D medium (Advanced DMEM/F12 (Gibco) with

- 1X Glutamax (Gibco)
- 10 mM HEPES (Sigma-Aldrich)
- 1.5X B27 supplement (Gibco)
- Antibiotic-Antimycotic (Gibco)
- 1 mM N-Acetylcysteine (Sigma-Aldrich)
- 50 ng/mL Human Recombinant EGF (Sigma-Aldrich)
- 100 ng/mL Human Recombinant FGF-10 (Peprotech)
- 1 mM A-83-01 (Sigma-Aldrich)
- 10 mM Nicotinamide (Sigma-Aldrich)
- 100 nM Hydrocortisone (HC, Sigma-Aldrich)
- 0.5 mg/ml epinephrine (Sigma-Aldrich)
- 4 pg/mL Triiodo-L-thyronine (T3, Promocell)
- R-Spondin (conditioned media, self-made)

The composition was mixed with two volumes of growth factor reduced Matrigel (Corning). Drops of cell suspension/Matrigel were distributed in a 6-well low attachment cell culture plate

(Sarstedt) and allowed to solidify for 30 min at 37 °C, upon which CK3D media was added to cover the drops. To evaluate the growth of PDOs, bright-field images were captured using a microscope. Organoids at approximately 100-500 um were passaged, and at least 10,000 cells were collected for cell model validation using targeted DNA sequencing of the *VHL* gene. To achieve this, DNA was isolated using the Maxwell® 16 DNA Purification Kit (Promega) and corresponding Maxwell instrument. PCR and sequencing of *VHL* were performed as previously described (Rechsteiner et al. 2011).

**Short Read sequencing**

The second part of the cDNA was used for Illumina sequencing library preparation, following the 10x Genomics Chromium Single Cell 3' Reagent (v3.1 Chemistry Dual Index) protocol described above. The cDNA was enzymatically sheared to a target size of 200-300 bp, and Illumina sequencing libraries were constructed. This process included end repair and A-tailing, adapter ligation, a sample index PCR, and SPRI bead clean-ups with double-sided size selection. The sample index PCR added a unique dual index for sample multiplexing during sequencing. The final libraries contained P5 and P7 primers used in Illumina bridge amplification. Sequencing was performed using paired-end 28-91 bp sequencing on an Illumina NovaSeq 6000 to achieve approximately 300,000 reads per cell. Our second manuscript mainly used the data to compare quality parameters between PacBio long-read and Illumina short-read data (Zajac et al. 2024).

**Cell Type Annotations**

The following markers were used for the annotation of cells:

- ccRCC Cells: *CA9*, *ANGPTL4*, *NDUFA4L2*, *LOX*, *VEGFA*, *VIM*, and *EGLN3.*
- Proximal Tubule Cells (PTC); *EPCAM*, *PAX8*, *GGT1*, and *RIDA.*
- Stromal Cells: *ACTA2*, *FAP*, *COL1A1*, and *COL1A2*.
- Endothelial Vascular Cells: *CDH5*, *FLT1*, *PECAM1*, and *KDR.*
- Immune Cells: *CD3*, *CD8A*, *PD1*, *CTLA4*, *CD68*, *CD163*, and *ITGAX*.

- Stem Cells: *ALDH1A1*, *SOX2*, and *CD44*.
- Mesenchymal Cell: *VIM*, *FN1*, *SNAI1*, *SNAI2*, *ZEB1*, and *ZEB2*.
- Epithelial-mesenchymal transition (EMT): *CDH2*, *TWIST1*, *MMP2*, and *MMP9*.

### PCR Validations

For *NNMT*: A forward primer was specifically designed against the unique sequence of the novel isoform at the end of exon 2. To detect the canonical isoform, another forward primer was designed to span the unique sequence of the canonical transcript at exon 1. Both the reverse primers were designed against different regions of exon 3.

For *TMEM91*: A forward primer specific to the novel isoform was designed to span exon 1 of the novel transcript. Additionally, a forward primer was designed against the sequence shared between canonical and novel isoforms, corresponding to exon 2 in novel and exon 1 in canonical (mapping to ENST00000392002.7) transcripts. Reverse primer was designed to target a shared region of exon 3 (novel isoform)/exon 2 (canonical isoform).

## References

Rechsteiner MP, Von Teichman A, Nowicka A, Sulser T, Schraml P, Moch H. 2011. *VHL* Gene Mutations and Their Effects on Hypoxia Inducible Factor HIFα: Identification of Potential Driver and Passenger Mutations. *Cancer Research* 71: 5500–5511.

Zajac N, Zhang Q, Bratus-Neuschwander A, Qi W, Bolck HA, Karakulak T, Oltra TC, Moch H, Kahraman A, Rehrauer H. 2024. Comparison of Single-cell Long-read and Short-read Transcriptome Sequencing of Patient-derived Organoid Cells of ccRCC: Quality Evaluation of the MAS-ISO-seq Approach. http://biorxiv.org/lookup/doi/10.1101/2024.03.14.584953 (Accessed January 8, 2025).