

# **Single-cell deconvolution of a specific malignant cell population as a poor prognostic biomarker in low-risk clear cell renal cell carcinoma patients**

Saout JR., Lecuyer G., Leonard S.<sup>1</sup>, Evrard B., Kammerer-Jacquet SF., Noel L., Khene ZE, Mathieu R., Brunot A., Rolland AD., Bensalah K., Rioux-Leclercq N., Lardenois A., Chalmel F.

## **Supplementary Materials**

### **Table of contents**

Supplementary methods .....	1
Supplementary results .....	11
Supplementary references .....	21

## 1 **Supplementary methods**

### 2 **Ethics statement**

3 Human samples were obtained through the *Centre de Ressources Biologiques (CRB) Santé* of Rennes  
4 (ID in the national biobank infrastructure: BB-0033-00056). The research protocol was conducted  
5 according to French legal guidelines and fulfilled the requirements of the local institutional ethics  
6 committee. Informed consent was signed by each patient.

7

### 8 **Histological staining**

9 Additional samples of each ccRCC were fixed for 24h in 4% paraformaldehyde-PBS and paraffin  
10 embedded following standard protocols. Microscopic H&E features were acquired on the CaseViewer  
11 v2.2 software.

12

### 13 **Tissue dissociation and tumor cell enrichment assessment**

14 For each tumor, punches were performed in locally distinct tumor areas without necrosis, fibrosis or  
15 edema. Punches were pooled and dissociated following a previously described protocol [1]. Briefly,  
16 samples were first subjected to enzymatic and mechanical disaggregation (tumor dissociation kit,  
17 Miltenyi Biotec) before undergoing red blood cell lysis and low-speed centrifugation steps to remove  
18 debris and dead cells while maintaining tumor cells (Fig. 1A, step1). Partial depletion of CD45 (an  
19 immune cell marker) positive cells was performed using the Miltenyi kit (CD45 MicroBeads, human)  
20 to enrich the samples in malignant/stromal cells and remove circulating and tumor-infiltrating immune  
21 cells (step2).

22

### 23 **Microscopic observations**

24 Cell suspension quality was assessed using an Olympus IX70 microscope from Olympus-Lifesciences.  
25 Cells were stained with Trypan blue to estimate mortality rate and cells were counted with a  
26 hemocytometer.

27

### 28 **Freezing and thawing of cells**

29 Approximately 300,000 cells were transferred in 1ml freezing solution consisting of 70% RPMI 1640,  
30 20 % fetal bovine serum (FBS), 10 % DMSO (Sigma) and immediately placed at -20°C for 1h before  
31 their temperature was progressively lowered (-1°C/min) to -80°C for storage. Cells were then  
32 transferred to liquid nitrogen for long-term storage. Frozen samples were hand-thawed (at 37°C) and  
33 immediately diluted with 14 ml RPMI 1640 in a 15 ml conical tube. With the exception of RCC4\_f,  
34 scRNA-Seq was performed on freshly harvested samples without prior freezing.

35

### 36 **Flow cytometry**

37 Cells were collected at two stages in the protocol: 1- after tissue disaggregation, RBC lysis, and low-  
38 speed centrifugation steps; and, 2- after partial CD45 depletion. Cells were preserved on ice for up to 2  
39 hours before co-staining with anti-human CD45-FITC/CD34-PE monoclonal antibodies and 7-AAD  
40 (BD Biosciences, San Jose, CA) for 15 minutes at -4°C in the dark. Cell suspensions were washed twice  
41 in PBS. Samples were acquired with a LSR Fortessa X-20 (Becton Dickinson) cytometer using BD  
42 FACSDiva software v8. The data were analyzed using FlowLogic v7.2.1 software.

43

### 44 **Estimation of tumor-infiltrating immune cells**

45 The proportion of infiltrating leukocytes was estimated in tumors from scanned images of H&E and  
46 LCA (Leukocyte Common Antigen which recognizes CD45; DAKO - ref: M0701)

47 immunohistochemical staining, using the ImageJ suite. Five counting zones were randomly selected  
48 within the tumor tissue, and an average proportion of stained cells was calculated for each tumor.  
49 Unstained cells were considered as non-immune and identified as blue nuclei with the following color  
50 parameters : Hue ranging from 128 to 192; Saturation ranging from 32 to 255; and Brightness ranging  
51 from 128 to 255. Stained cells appearing as intense brown cells were scored as Immune cells and  
52 identified with the following color parameters : Hue ranging from 0 to 255; Saturation ranging from 0  
53 to 255; and Brightness ranging from 0 to 128. For both counts, the image was changed to 8-bit.  
54 Watershed function was applied to distinguish cells within aggregates. Particles corresponding to either  
55 immune or non-immune cells were then automatically counted using the Analyze particle function,  
56 excluding particles under 35 and 50 pixels<sup>2</sup> for immune cells and non-immune cells, respectively.

57

## 58 **Data analysis of a single-cell atlas composed of 13 samples**

59 *Pre-processing, quality controls, integration and normalization.* Demultiplexed raw sequencing reads  
60 were processed, mapped to the GRCh38 human reference genome and quantified based on unique  
61 molecular identifiers (UMIs) with the Cell Ranger pipeline (v3.0.2, 10x Genomics). The resulting raw  
62 count matrices from the 7 individual samples were then merged. The cell-calling algorithm from Cell  
63 Ranger, which relies on the EmptyDrops method [2], was also used to establish a first set of potential  
64 valid cells. Subsequently, the scater R package. (v1.10.1) was used to remove outlier cells by using  
65 several cell features including the proportions of reads mapping to mitochondrial and ribosomal genes  
66 [3]. In parallel, doublets were filtered out independently in each individual matrix by using the  
67 DoubletFinder R package (v.2.0.2) [4]. Next, raw count matrix files from the scRNA-seq dataset  
68 published by Young and colleagues were downloaded from their online supplementary data [5]. Cells  
69 with fewer than 200 detected genes and genes detected in fewer than 10 cells were removed. Cells from  
70 both datasets were assigned a cell cycle phase by using both the Scran R package (v1.10.2) [6] and the  
71 Seurat v3.1.5 package [7]. Each dataset was individually normalized using the NormalizeData function  
72 implemented in Seurat. The two datasets were merged and batch-effect corrected using the CCA method  
73 implemented in the Seurat package to regress out technical covariates between datasets. This procedure

74 is based on the FindIntegrationAnchors and IntegrateData functions with the following parameters:  
75 anchor.features=3000 and dims=30.

76 ***Pseudobulk analysis.*** A “pseudobulk” count matrix resulting from the aggregation of counts of all  
77 selected cells of each sample was used to calculate pearson correlation coefficients and perform  
78 pairwise comparisons between samples. The “pseudobulk” count matrix was also used to perform a  
79 principal component analysis based on the FactoMineR package [8].

80 ***Reduction of dimensionality and clustering.*** The 3000 integration anchors identified by the  
81 FindIntegrationAnchors function were used to perform a principal component analysis with the  
82 RunPCA function implemented in Seurat. Cells were then clustered by using the FindNeighbors and  
83 FindClusters functions on the top-30 principal components, with default parameters. Finally, we used  
84 the uniform manifold approximation and projection (UMAP) method implemented in Seurat to project  
85 single cells in a reduced 2D space.

86 ***Cluster annotation and differential gene expression analysis.*** Cell clusters were annotated using a set  
87 of known marker genes complemented with genes that were significantly differentially expressed  
88 between cell clusters. Differentially expressed genes were identified with the FindAllMarkers function  
89 implemented in Seurat, with default parameters.

90

## 91 **Copy Number Variation analysis and phylogenetic tree construction**

92 The inferCNV R package (InferCNV of the Trinity CTAT Project:  
93 <https://github.com/broadinstitute/InferCNV>) was used to evaluate copy number variations (CNV) in  
94 malignant cells based on single-cell RNA-seq raw counts (InferCNV of the Trinity CTAT Project:  
95 <https://github.com/broadinstitute/InferCNV>). Inference was run separately on each individual sample,  
96 using nonmalignant cells as reference. The cutoff parameter was set to 0.1, and parameters denoise,  
97 cluster\_by\_groups and HMM were set to TRUE. The tool returns a matrix with a CNV score  
98 corresponding to the relative intensity of gene expression for a particular cytoband in a given cell. We

99 also calculated an overall CNV burden score for each cell corresponding to the weighted mean of the  
 100 absolute CNV scores taking into account the cytoband size, using the following formula:

$$CNV \text{ Burden score} = \frac{\sum_{i=1}^n |C_i| \cdot S_i}{\sum_{i=1}^n S_i}$$

101 with n being the number of cytobands, C the corresponding CNV score of cytoband i, and S the size of  
 102 cytoband i.

103 A differential CNV score analysis was performed by comparing the CNV scores of normal tubular  
 104 epithelial cells and malignant cells using a Wilcoxon test. The *p value* was then adjusted with the  
 105 Benjamin Hochberg method and significant CNV scores in malignant cells were considered for further  
 106 analysis. The phylogenetic tree was constructed from the outputs of inferCNV by using Uphyloplot2  
 107 [36] with parameter  $c = 1$ .

108

### 109 **Single-cell RNA data analysis of malignant cells and tubular epithelial cells**

110 ***Selection of malignant cells and tubular epithelial cells.*** Cell populations belonging to clusters  
 111 associated with “malignant cells” (c2, c3, c5, c10, c16, c20, c30, c31) and “tubular epithelial &  
 112 malignant cells” (c8, c12) were selected for further analysis. The Seurat object was reduced to those  
 113 selected cells by using the subset function implemented in Seurat. The ultimate aim of this analysis  
 114 being to infer a single-cell trajectory for each individual tumor sample, no batch-effect correction was  
 115 applied in order to avoid overcorrection [9]. Moreover, such batch-effect correction (CCA) was  
 116 performed but did not reveal major changes (especially regarding the identification of the five RCC2  
 117 malignant subpopulations) (data not shown).

118 ***Reduction of dimensionality and clustering.*** A multi-factorial analysis (MFA) was performed with the  
 119 single-cell transcriptomic data (based on the top-3000 most varying genes) and the single-cell CNV  
 120 matrix inferred by InferCNV (based on the cytobands with variance>0) (see **Copy Number Variation  
 121 analysis and phylogenetic tree construction**) by using the FactoMineR package [8]. Cells were then  
 122 clustered by using the FindNeighbors and FindClusters functions on the top-30 principal components

123 of the MFA, with default parameters. Finally, we used the uniform manifold approximation and  
124 projection (UMAP) method implemented in Seurat to project single cells in a reduced 2D space based  
125 on those top-30 MFA components.

126 ***Cluster annotation and differential gene expression analysis.*** Cell clusters were annotated using a set  
127 of known marker genes complemented with genes that were significantly differentially expressed  
128 between cell clusters. Differentially expressed genes were identified by pairwise comparisons between  
129 all clusters using the FindMarkers function implemented in Seurat, with default parameters.

130 ***Single-cell trajectory inference and pseudotime analysis.*** Malignant cell trajectories were inferred with  
131 Monocle3 [10] on the uncorrected scRNA-seq data. Each tumor sample was used as a distinct partition.  
132 Pseudotimes were estimated using the pseudotime function by indicating as starting cells those with the  
133 least chromosomal aberrations, *i.e.* the lowest CNV burden (see **Copy Number Variation analysis**  
134 **and phylogenetic tree construction**).

135

### 136 **Processing of the Cancer Genome Atlas Kidney Renal Cell Carcinoma (TCGA-KIRC) data**

137 ***Bulk RNA datasets of ccRCCs.*** The raw count expression matrix from The Cancer Genome Atlas  
138 Kidney Renal Cell Carcinoma (TCGA-KIRC) and the associated clinical data (tumor stage, survival,  
139 last day of follow-up etc.) were downloaded using the TGCABiolink R package and the Broad Institute  
140 TCGA-GDAC (<http://gdac.broadinstitute.org/>), respectively. The gene prognostic annotations were  
141 derived from the TCGA-KIRC database.

142 ***Survival analysis.*** Survival analysis was performed using the Survival R package [11] (version 3.2-13).  
143 First, clinical data were censored according to the last day of follow-up and whether the patient was  
144 still alive or not. For the gene prognostic analysis, patients were categorized in either low or high  
145 expression groups according to the lowest Log Rank Test p-value. As defined by the Human Protein  
146 Atlas, *p value* < 0.001 reflects significant differences in survival outcomes between the two groups and  
147 defines the gene as prognostic.

148 ***Selection of favorable and unfavorable sets of genes.*** Significantly favorable and unfavorable genes  
149 for survival were defined by comparing the survival probability at 2000 days between the two  
150 expression groups.

151

## 152 **Selection of gene sets of interest**

153 In addition to the sets of favorable and unfavorable (see previous section) genes, we also selected genes  
154 involved in the epithelial-mesenchymal transition (EMT), using a comprehensive collection of EMT-  
155 related genes via a web-accessible resource called dbEMT [12]. This database consists of a core dataset  
156 of 1,184 human genes involved in the EMT retrieved from 2,665 PubMed abstracts.

157

## 158 **RNA deconvolution and survival model**

159 Bulk RNA deconvolution was performed with the MuSiC R package v0.2.0 [13] on HTSEQ raw counts  
160 of TCGA bulk data, using single-cell RNA-seq data from either RCC2 malignant subpopulations or  
161 somatic cells across all patients. The proportion of each RCC2 malignant subpopulation and each  
162 somatic cell type was then inferred in each TCGA tumor sample. To avoid a bias related to the balance  
163 between cell populations in each TCGA sample, we used the relative proportion of each RCC2  
164 malignant subpopulation (such as “Relative c8”) defined as the proportion of the considered RCC2  
165 malignant subpopulation as compared to the proportion of all RCC2 malignant subpopulations.

166

## 167 **Survival analysis**

168 Survival analysis was performed using the Survival R package (version 3.2-13) as previously described  
169 (**The Cancer Genome Atlas Kidney Renal Cell Carcinoma (TCGA-KIRC) processing** section).  
170 Two statistical methods were implemented for the deconvolution-based survival analysis: the Cox  
171 Proportional-Hazards Model method, and the Log Rank Test in which patients were categorized as



172 having either a low or high proportion of a given cell type according to the lowest Log Rank Test p-  
173 value - method similar to that implemented in The Human Protein Atlas [14]. P-value < 0.01 reflects  
174 significant differences in survival outcomes between the two groups and defines the cell type as  
175 prognostic. Finally, survival analysis curves were produced using the Kaplan-Meier model.

176

### 177 **Genomic features of ccRCCs**

178 **Somatic mutations.** Genes with somatic mutations in TCGA-KIRC patients were downloaded from the  
179 supplementary data published by Chen *et al.* [15]. Those genes were complemented by two other genes  
180 (CSMD3, TSC2) described in Turajilic *et al.* [16] for which somatic mutations were inferred by using  
181 the TCGA biolinks R package [17]. Mutated genes were next reported in Table S10.

182 **VHL and CDKN2A methylation status.** Methylation status of VHL and CDKN2A (p16) were  
183 downloaded from the supplementary data provided by Chen *et al.* [15]. Methylation status of VHL and  
184 CDKN2A (p16) were reported in Table S10.

185 **Chromosomal abnormalities.** CNV data (Affymetrix SNP 6.0 arrays) provided by the TCGA  
186 consortium were downloaded with the TCGA biolinks R package [17]. Briefly, for each TCGA-KIRC  
187 sample and each chromosomal arm, we calculated the weighted mean value of all copy number  
188 segments belonging to a given chromosomal arm, with weights equal to the lengths of the copy number  
189 segments. Chromosomal arms showing a weighted mean >0.3 and <-0.3 were counted as “gained” and  
190 “lost”, respectively as reported in Table S10.

191 **Weighted genome instability index (wgII).** For each TCGA-KIRC sample, the weighted genome  
192 instability index (wgII) was calculated as described in previous studies [18,19]. Briefly, the ploidy of  
193 a given TCGA-KIRC tumor sample was determined as the weighted median integer copy number, with  
194 weights equal to the lengths of the copy number segments. For each autosomal chromosome of a given  
195 TCGA-KIRC sample, the number of bases deviating from the ploidy of the corresponding sample was  
196 added, and this sum was divided by the size of the chromosome in question. Finally, the wgII score of  
197 a given TCGA-KIRC sample is defined as the average of this percentage value over the 22 autosomal

198 chromosomes. Based on a survival analysis, we next categorized TCGA-samples as having either a low  
199 (<12.8%) or high (>12.8%) wGII score according to the lowest Log Rank Test p-value corresponding  
200 to  $p < 1.5e-3$  (Table S10, Fig. S10).

201 **Genomic subtypes of ccRCC.** The association between TCGA-KIRC samples with one of the nine RCC  
202 genomic subtypes (among which CC-e.1, CC-e.2, CC-e.3 which are enriched in ccRCC cases) described  
203 in Chen *et al.* [15] was downloaded and reported into Table S10.

204

### 205 **Functional analysis**

206 The AMEN suite of tools [20] was used to calculate the Fisher exact probability and the Gaussian  
207 hypergeometric test was used to identify significantly enriched terms from the Gene Ontology as well  
208 as from the selected sets of genes (see previous section). A specific term was considered enriched in a  
209 group of coexpressed genes if the adjusted (Benjamini–Hochberg procedure) *p value* was  $\leq 0.05$  and  
210 the number of genes bearing a given annotation was  $\geq 3$ .

211 The hypergeometric statistical test was also used to identify significantly enriched RCC genomics  
212 features in TCGA-KIRC patients which were classified according to their low or high proportion of  
213 each deconvolved subpopulation (as previously defined in the *Survival analysis* section). This analysis  
214 was independently performed on all patients, low-risk patients (T1 and T2) and high-risk patients (T3  
215 and T4) from the TCGA-KIRC dataset (Table S11, Fig. S12).

216

### 217 **Heatmap and spot plot**

218 Spot plots were generated using the FlexDotPlot package [21]. Heatmaps were generated with the  
219 pheatmap package v1.0.12 (<https://CRAN.R-project.org/package=pheatmap>).

220

221 **Data availability**

222 Raw data are available on the EGA (European Genome-Phenome Archive) portal under accession  
223 number EGAS00001006534. Processed single-cell RNA-seq and available on ArrayExpress portal  
224 under accession number E-MTAB-11310. All processed data are also available via the web-based portal  
225 UncoVer (The Uro-oncogenomics Viewer at <https://uncover.genouest.org>) (Lecuyer, Saout *et al.* in  
226 preparation).

## 227 **Supplementary results**

### 228 **Section 1 : Tumors displayed various clinical features**

229 Tumors were classified from pT1a to pT3b and from ISUP grade 2 to 4, according to the TNM and  
230 2017 WHO/ISUP classification systems, respectively (Table S1; Fig. S1). Four samples came from  
231 distinct primary ccRCC patients (RCC1 to RCC4) and two samples consisting of matched primary  
232 tumor and distant vena cava tumor thrombus came from the same patient (RCC5 and RCC5\_t,  
233 respectively). Macroscopic examinations revealed various appearances with heterogeneous patterns of  
234 vascularization, hemorrhagic and necrotic remodeling (Table S1; Fig. S2A). These features of  
235 heterogeneity led us to sample each tumor in up to 9 distinct macroscopic areas, depending on the size  
236 of the lesion (Fig. S2B-C).

237

### 238 **Section 2 : Experimental approach to enrich malignant cells in cell suspensions from ccRCC** 239 **tumors**

240 Tumor samples were processed following a preset protocol including tissue dissociation followed by  
241 low-speed centrifugation, referred to as step1; and depletion of CD45+ cells so as to remove circulating  
242 and infiltrating immune cells (IC) and enrich malignant cells (MCs), referred to as step-2 (Fig. 1A). In  
243 order to assess the impact of cryo-preservation, we froze part of the RCC4 cell suspension (RCC4\_f)  
244 and then processed it the same way as the fresh samples. As scRNA-seq is a sensitive technique  
245 requiring thorough cell dissociation with a high preservation of cell viability, we first evaluated cell  
246 viability and proportions of various cell subsets at the different steps of the protocol as illustrated with  
247 RCC4 in Fig. S3A-B. Microscopic examination at each step revealed well individualized and diverse  
248 cells, in terms of shape and size (Fig. S3A). To further confirm the distribution and integrity of different  
249 cell types , we used flow cytometry to analyze immune cells (CD45+), endothelial cells (CD34+),  
250 stromal and MCs (CD45-/CD3-) and viability (7AAD) (Fig. S3B). Cell counts decreased from step 1  
251 to step 2 in each sample, sometimes dramatically (Fig. S3C), while global cell viability slightly  
252 decreased (Fig. S3D). However, the proportion of viable cells remained the same in each of the

253 monitored cell types (Fig. S3F), suggesting the decrease in cell viability could be due to the removal of  
254 ICs given their relatively high rate of viability. ICs are globally predominant in the samples (Fig. S3B-  
255 E) and highly viable (above 90%) in each sample. Malignant, stromal and endothelial cells were  
256 efficiently enriched by the immune depletion (Fig. S3B-E). With the exception of RCC5, the most  
257 clinically advanced tumor (pT3b), the cell viability rate exceeded the 70% minimum viability threshold  
258 recommended for single-cell RNA sequencing experiments (Fig. S3D). Among the different cell types,  
259 endothelial cells were the most impacted by the tissue dissociation, with less than 50% cell viability at  
260 step 1 and 2, whereas malignant and stromal cells remained highly viable. To assess the impact of  
261 cryopreservation, we froze dissociated single cells from 6 ccRCCs, including RCC4\_f already frozen  
262 and thawed, and observed a slight decrease in their overall viability, suggesting that cells may be harmed  
263 by freezing (Fig. S3H). However, malignant and stromal cells still presented a viability rate above 75%.  
264 To better document the nature of the depleted ICs, we compared their proportion by flow cytometry at  
265 step 1 and immunohistochemistry prior to dissociation (Fig. S3D and Fig. S4). In comparison, the  
266 estimated proportion of ICs by flow cytometry was ~8-23 times higher than by IHC (ranging from 3–  
267 10%) (Table S2). This difference could result from the resistance and individualized state of ICs,  
268 facilitating their selection during dissociation, or contrasted sensitivities of the two techniques.  
269 However, given the large difference, these results suggest that a substantial proportion of ICs detected  
270 by flow cytometry may correspond to circulating ICs coming from blood circulation in the tumor  
271 sample.

272

### 273 **Section 3 : Single-cell RNA-seq uncovers the cellular composition of ccRCC tumors**

274 The final set of samples was composed of seven samples from our lab (RCC1-5, RCC4\_f and RCC5\_t)  
275 complemented with three ccRCC samples (RCC6-8) and three healthy samples (RCC[6-8]\_n) published  
276 by Young and colleagues [5]. After quality controls, the entire dataset contained the expression levels  
277 of 37,832 genes in 54,812 cells (Fig. 1B). The number of detected cells per sample ranged from 2,354  
278 and 4,549 in our dataset (sd = 796), and from 2,001 to 14,769 in the additional dataset (sd = 4,912)  
279 (Table S2). The median number of detected genes per sample ranged from 572 to 2,917, with more

280 genes in our dataset as compared to that of Young *et al.* (median of 1,719 versus 868), probably as a  
281 result of deeper sequencing.

282 In order to get an overview of the transcriptomic proximity between samples, we first aggregated their  
283 cells' transcriptomes and performed pairwise comparisons between the resulting pseudobulk samples  
284 (Fig. S5A). Samples sharing the same experimental and biological background were strongly correlated  
285 and distributed in three groups, consisting of samples from our dataset (RCC1 to RCC5), and tumor or  
286 normal samples from additional dataset (RCC6 to RCC8). From this "pseudobulk" expression matrix,  
287 we projected samples on the first two components of the PCA (Fig. S5B). Accordingly, the first  
288 component distinguishes samples depending on which dataset they belong to, suggesting differences in  
289 their experimental processing. Interestingly, strong correlations are observed between biologically  
290 related samples such as RCC4/RCC4\_f (0.99) and RCC5/RCC5\_t (0.96). Given the differences between  
291 tumor samples, the dataset was batch-effect corrected using *IntegrateData* function in Seurat.

292 The transcriptomic signal was then analyzed at the single cell resolution by performing dimensionality  
293 reduction on the single-cell count matrix and projecting single cells on a UMAP (Uniform Manifold  
294 Approximation and Projection) 2D space (Fig. 1C). The resulting single-cell transcriptomic atlas gives  
295 a general idea of the overall cell diversity of the ccRCC samples. We annotated single cells according  
296 to their sample of origin, which revealed various levels of intra-tumor heterogeneities, as each sample  
297 contained three or more broad cell populations. In accordance with the pseudobulk analysis, RCC5  
298 primary and thrombus derived cells, although distinguishable, appeared to be similarly distributed.  
299 More remarkably, fresh and frozen cells from RCC4 perfectly overlapped (Fig. 1D), demonstrating that  
300 the freezing step in our experimental protocol has little to no impact on single-cell transcriptomes. This  
301 indicates the feasibility of performing scRNA-Seq on frozen ccRCC dissociated cells, enabling the  
302 simultaneous analysis of samples that would normally be acquired at different time points, such as  
303 matched primary tumor and metastasis samples. Alternatively, single-nuclei RNA-seq techniques may  
304 be optimal to analyze frozen samples with only a slight decrease in sensitivity [22].

305 To identify cell types present in ccRCC, we performed clustering on the top-30 components of the PCA  
306 matrix, and identified 35 cell clusters (denominated c1-c35) (Fig. 2A) which were annotated based on

307 the expression of well known cellular markers (Fig. 2B-C). MCs (associated with eight cell clusters)  
308 were identified based on the expression of CA9 and NNMT, two well-known canonical markers [23,24].  
309 Fourteen clusters were classified as ICs (as they expressed canonical markers such as PTPRC) including  
310 T cells (CD8A, CD4; c6, c7, c22, c24), NK cells (KLRD1; c17, c19, c25), macrophages (CD68; c13,  
311 c14, c21), mast cells (MS4A2; c35) and B/plasma cells (MS4A1, MZB1; c9, c11, c35). Interestingly,  
312 the vast majority of B/plasma cells (c9, c11) belong to RCC4 but are PTPRC negative (Fig. 2B), which  
313 explains why they were not depleted in our protocol. The remaining clusters were classified as  
314 endothelial cells based on CD34 and PECAM1 (c1, c15, c27, c28, c29); pericytes based on ACTA2 and  
315 PDGFRB (c4, c23); myofibroblasts based on LUM, DCN and residual expression of pericytes markers  
316 (c34); collecting duct epithelial cells based on UMOD (c26, c33); tubular epithelial cells based on  
317 GATM (c8); erythrocytes based on HBB (c18); and finally cluster c12 exhibiting both expression of  
318 tubular epithelial cell markers (GATM, GPX3) and MC markers (CA9 and NNMT) that corresponds to  
319 an intermediate status between the two cell types.

320 To confirm the identity of MCs and further characterize them, we used inferCNV (inferCNV of the  
321 Trinity CTAT Project. <https://github.com/broadinstitute/inferCNV>) and checked the 3p chromosomal  
322 arm status across normal and MCs, as its deletion is a key initial event resulting in VHL loss of function  
323 in more than 90% of ccRCCs [25,26]. This approach confirmed the 3p deletion and revealed its  
324 occurrence in almost all of the cells identified as malignant (Fig. S6), although the sequencing depth  
325 and coverage were too weak to categorically confirm this observation in the RCC7 MCs. Interestingly,  
326 we also found some of the most common CNVs such as 5q and 5p gains (found in 60% and 26% of  
327 ccRCC, respectively) or 13q loss (24% of ccRCC) [26] (Fig. S7). Hence, the inference appeared to  
328 accurately reflect the CNV landscape of tumor cells and constitutes an additional marker of  
329 heterogeneity in the MCs. Interestingly, we also uncovered poorly described chromosomal aberrations  
330 with an heterogeneous distribution across the tumors, such as a very specific gain in chromosome 16  
331 for RCC5 primary MCs as compared to RCC5 thrombus MCs, or a significant loss in 18p for RCC2  
332 and RCC5.

333

#### 334 **Section 4: Generation of a single-cell atlas of malignant cells in ccRCCs**

335 We assembled a single-cell atlas of MCs by selecting all the cells belonging to clusters associated with  
336 MCs (c2, c3, c5, c10, c16, c20, c30, c31) and normal tubular epithelial cells (TECs) (c8, c12) and  
337 reanalyzed the resulting cells (15,471 MCs and 3,737 TECs). The dimensionality reduction was  
338 performed by using a multi-factorial analysis (MFA) based on the scRNA-seq data and the inferred  
339 CNV data. Finally cells were projected on a 2D space (UMAP) and partitioned into 24 cell clusters.

340

#### 341 **Section 5: CNV heterogeneity in RCC2 malignant cells**

342 Losses in 3p21-26 and 13q33-34, as well as gains of 1q21-22, 7q36, 10p14-15 and 5q were observed in  
343 all RCC2 MC clusters, designating them as early events in its initiation (Fig. S9A). Although the rate  
344 of gain was approximately the same, the losses increased with pseudotime, culminating in c8, in  
345 accordance with the global CNV burden. Based on this CNV accumulation we used Uphylot2 [27]  
346 to construct a phylogenetic tree (Fig. S9B), which revealed a branched evolution from a common  
347 ancestor (A) to distinct malignant clones (P, O, L, K, H, G, E and D). Interestingly, we observed that  
348 CNV-based clones overlap with transcriptomic clusters. In accordance with previous observations, the  
349 loss of 3p and the gain of 5q appeared to be common to all clones whereas some specific events were  
350 restricted, like the numerous specific aberrations in branch H, corresponding to c23. The phylogenetic  
351 tree, although probably incomplete, gives a broad idea of the diversity and the relationships between  
352 the MCs within RCC2, further supporting the idea of diverse behaviors.

353 The 3p arm, lost in all MCs, carries several genes involved in ccRCC carcinogenesis, such as SETD2,  
354 BAP1 and PBRM1, whose losses and their consequences are widely observed and well documented in  
355 ccRCC [25]. The 3q arm was also lost to various extents in certain MC clusters; larger proportions were  
356 affected in c23, c8 and c9 than in c3 and c10. In particular, band 3q26, lost at the end of the tumor  
357 lineage, carries the PIK3CA oncogene, which was found to be mutated in different ccRCC cohorts [28].  
358 Interestingly, PIK3CA is part of the same biological pathway as PTEN, a common tumor suppressor on  
359 band 10q23, which happens to also be lost specifically in c23, c9 and c8. Additionally, part of c8,



360 harbors a unique 9p loss (Fig. S9B), which carries CDKN2A (9p21), an important suppressor gene.  
361 Interestingly, the loss of 9p21 is observed in larger proportions of metastases than primary tumors which  
362 makes it a likely cause of metastasis in ccRCC [28]. Altogether, RCC2 MC clusters carry the most  
363 chromosomal alterations, in particular in c8, which could infer to it particularly aggressive properties.

364

### 365 **Section 6: Dynamics of the transcriptional landscape across the RCC2 tumor lineage**

366 In total we identified 2,229 genes differentially expressed between tubular epithelial cells (TECs) and  
367 RCC2 malignant cells (MCs) as well as along the RCC2 tumor lineage which were subsequently  
368 distributed into nine expression patterns (P1-9) (Fig. 4C, Table S5). Genes belonging to patterns P1 are  
369 preferentially expressed in TECs as compared to MCs. Patterns P2, P8 and P9 contain genes whose  
370 expression profiles are shared by several adjacent MC clusters: P2 contains genes expressed in all MCs;  
371 P8 includes genes expressed in c23 onwards; and, P9 regroups genes that are specifically down-  
372 regulated in c8. Patterns P3, P4, P5, P6 and P7 include genes that are preferentially expressed in c3,  
373 c10, c23, c9 and c8, respectively. The functional analysis revealed a progression towards an increasingly  
374 unfavorable score (corresponding to the significant over-representation of unfavorable genes within a  
375 given expression pattern) from P3 to P8, which reflects a gain in aggressive features along the tumor  
376 clonal lineage. Conversely the favorable score was globally higher in the first expression patterns (P2-  
377 6) and in P9 than in P7-8. With the exception of P5, all expression patterns corresponding to MC clusters  
378 displayed a significant enrichment of genes involved in the EMT. Gene ontology (GO) analysis revealed  
379 826 GO terms that were significantly enriched in at least one expression pattern (Table S6).

380 The pattern P1, comprising 107 DEGs preferentially expressed in TECs, was enriched in genes involved  
381 in metabolic processes such as ATP (13 genes, adjusted p value < 3.3e-7), carboxylic acid (26, 8.6e-  
382 11), and fatty acid metabolic processes (12, 5.5e-5). P2 comprised 400 genes associated to mRNA (47,  
383 1.3e-7), pyruvate (11, 2.9e-2), ATP (28, 3.9e-9), NADH (6, 2.6e-2) and nucleotide (29, 1.3e-4)  
384 metabolic processes, as well as intracellular protein transport (56, 2.3e-7), suggesting an overall  
385 energetic reprogramming in MCs as compared to TECs. No specific biological processes were found

386 to be enriched in P3 (81 genes). P4 (233 genes) was found to be enriched in genes associated with  
387 cellular response to cytokine stimuli (34, 1.5e-3) and response to type I interferon (9, 1.2e-2), indicating  
388 innate immune activity in c10; genes belonging to P5 (335 genes) were over-represented in organic acid  
389 (39, 3.4e-3) metabolic process, secretion (46, 1.8e-2) and programmed cell death (59, 1.3e-2); P6 genes  
390 (196) were significantly associated with cytokine-mediated signaling pathways (22, 6.5e-3), antigen  
391 processing and presentation (11, 8.2e-3), and more precisely T cell mediated cytotoxicity (5, 1.7e-2)  
392 suggesting the organization of an adaptive immune response in c9. Interestingly, P6 was also  
393 characterized by genes involved in programmed cell death (42, 3.9e-3), and is further linked to cell  
394 adhesion (34, 3.3e-3), secretion (30, 3.2e-2), and cellular responses to hypoxia (10, 1.0e-2), the initiation  
395 stage of angiogenesis [29]. As expected, P7 (354 DEGs) was found to be enriched in terms linked to  
396 aggressive tumor states, including inflammatory responses (33, 1.3e-4), blood vessel morphogenesis  
397 (31, 2.9e-5), cell cycle (51, 2.6e-2), cell migration (54, 1.0e-4), programmed cell death (79, 1.3e-8), cell  
398 adhesion (47, 3.7e-3), cytokine-mediated signaling pathway (40, 1.3e-6) and secretion (39, 4.9e-2); P8  
399 (391 DEGs) was over-represented in genes involved in programmed cell death (68, 3.0e-3) and cell  
400 migration (53, 8.0e-3) illustrating a globally more aggressive state from c23 cluster onwards. Finally,  
401 P9 (132 DEGs) was characterized by a significant association with genes involved in xenobiotic (7,  
402 1.2e-2), cellular amino acid (11, 7.6e-3) and fatty acid (15, 1.1e-4) metabolic processes, suggesting that  
403 c8 has an altered metabolic state and may respond differently to therapy.

404 Gene level analysis corroborates previous observations regarding the aggressiveness of patterns P7 and  
405 P8 (Table S7). Amongst the 2,229 DEGs, we found well-known oncogenes already described in renal  
406 cell carcinoma such as JUN/JUNB (P7) and AR (P8), the epithelial growth factor receptor (EGFR, P7)  
407 or CXCR4 (P8). A total of 228 genes were linked to EMT, including 92 in the P7 and P8 patterns. There  
408 were 628 genes associated with at least one GO term among angiogenesis, cell migration, cell adhesion  
409 or programmed cell death, including 253 for the P7 and P8 patterns. Twenty-seven of them were  
410 associated with all terms, including 16 genes in the P7 and P8 patterns, such as CCL2 [30] and CAV1  
411 [31] that have been reported in ccRCCs. Among genes that haven't been previously described in RCCs,  
412 108 were unfavorable in the TCGA cohort. For instance, RRAS and CALR (P8) were associated with

413 the regulation of cell migration and adhesion as well as with angiogenesis (RRAS) and programmed  
414 cell death (CARL), warranting deeper characterization of their impact in ccRCC. A total of 10 genes  
415 coded proteins that are targeted by FDA-approved drugs, including three gene products already targeted  
416 in kidney cancer (AXL, NTRK2 and MET). AXL (P8), in particular, showed peak expression in c23,  
417 c9 and c8, suggesting its inhibition with cabozantinib could counter the effects of the most aggressive  
418 MC clusters in RCC2. Similarly, 24 gene products, although not targeted by drugs currently used to  
419 treat ccRCCs, are reported to facilitate the action of FDA-approved drugs (“Tclin”), warranting further  
420 characterization of their status in ccRCCs. Among other notable genes, P8 contains MMP7, the  
421 metalloproteinase involved in extracellular matrix (ECM) degradation that promotes the invasion of  
422 MCs into the ECM [32], and PIK3R1 (Phosphatidylinositol 3-kinase regulatory subunit alpha), which  
423 is part of the highly activated PI3K/AKT pathway promoting cell proliferation in renal cell carcinomas  
424 [33]. P8 contains TGFBI (transforming growth factor beta induced), a cell adhesion inhibitor in the  
425 ECM [34] and MDK which is involved in cancer progression and EMT, notably modulating cell  
426 adhesion and migration [35]. Altogether, the functional analysis (Table S6) and the RCC2 DEGs  
427 annotations (Table S7) demonstrate the heterogeneous and aggressive states of the RCC2 tumor from  
428 the molecular standpoint. We also illustrate here a global method to identify relevant targets, warranting  
429 single-cell transcriptomics profiling efforts to support the design of novel targeted therapies.

## 430 **Section 7: Deconvolution of RCC2 malignant cell subpopulations are associated with patient** 431 **survival**

432 Based on log rank tests run on all the patients, we established groups having low versus high proportions  
433 of the deconvolved cell types by selecting the percentile displaying the most significant outcome  
434 difference (Table S9), for instance in c8 (percentile: 43.69; mean percentage in low group: 17.13%,  
435 CI[15.30-19.04] versus 70.10% [68.12-72.01] in high group), c9 (48.2; 23.70% [21.8,25.61] versus  
436 64.73% [63.42,66.04]), or endothelial cells (19.80; 12.72 [12.28,13.16 versus 37.83 [35.65,40.01]).  
437 Outcome was significantly unfavorable ( $p$  value < 0.01) among all the patients for those having high  
438 proportions of c8 among MCs (called “Relative c8”, see supplementary materials;  $p$  < 6.5e-7), and to a  
439 lesser extent myofibroblasts (<1.14e-5) and B cells (<1.5e-5) (Fig. 5B). Conversely, high proportions

440 of relative c9 ( $<7.1e-8$ ), endothelial cells ( $<4.55e-10$ ) as well as relative c10 ( $<7.5e-7$ ), despite its low  
441 levels, were significantly favorable factors. These associations with survival were also significant  
442 according to the Cox regression test. In high-risk patients (T3-T4), all previous cell types but B cells  
443 had a significant association with survival (data not shown). However, these patients most probably  
444 received therapies impacting their survival, which confounds the analysis. Very interestingly though,  
445 when considering low-risk patients (T1-T2), endothelial ( $p < 9.0e-3$ ), but most significantly relative c8  
446 ( $< 9.62e-4$ ) and c9 ( $< 9.42e-5$ ) remained predictive of survival, further discriminating them. For  
447 instance, patients displaying a high relative c8 had an approximately 25% lower 5-year survival  
448 probability than those with a low relative c8. Taking into account the confidence intervals, patients  
449 displaying more than 23.5% of relative c8 might be considered at higher risk than expected amongst  
450 T1-T2 patients. Overall the relative proportions of c8 and c9 among MCs constituted the two best poor  
451 and good prognostic factors, respectively, for survival in low-risk patients.

452

#### 453 **Section 8: Deconvolved RCC2 malignant subpopulations are associated with specific genomic** 454 **features**

455 To further characterize the deconvolved RCC2 malignant subpopulations and explore their aggressive  
456 potential, we next investigated their association with either recurrent genomic features described in  
457 ccRCC (CNVs, mutational status of 22 genes, methylation status of VHL and CDKN2A/p16) [15–17],  
458 RCC genomic subtypes as defined by Chen and colleagues, or genome instability as defined by Burrell  
459 and colleagues [18,19] (Table S10). RCC genomic subtypes include nine patient survival groups of  
460 TCGA samples among which CC-e.1 (intermediate prognosis), CC-e.2 (better) and CC-e.3 (worse)  
461 (Table S10). TCGA-KIRC samples were also classified into two categories, low-wGII and high-wGII  
462 samples, based on a survival analysis on their weighted genome instability index (wGII) ( $p < 1.5e-3$ )  
463 (c.f. supplementary methods). We performed an enrichment analysis (Table S11) which revealed  
464 significant associations of specific CNVs with TCGA-KIRC samples showing high or low proportion  
465 of RCC2 malignant subtypes (c8, c9 and c10) (high- or low-[c8,c9,c10] samples) (Fig. S11) including

466 recurrent CNVs (such as losses of 3p, 4q, 9p, and gains of 2q, 12p, 20q) [16]. When considering low-  
467 risk patients (Fig. S11, Table S11), only loss of 3p in low-[c8], high-[c9] TCGA-KIRC samples, gain  
468 of chromosome 2 in high-[c10] samples, and gain chromosome 12 in high-[c8] samples remained  
469 significantly enriched. Interestingly, gain of chromosome 12 has been shown to be associated with  
470 aggressiveness and lower survival rates in ccRCCs [36,37]. This analysis revealed no significant  
471 associations of the methylation and mutational status with low- or high-[c8,c9,c10] TCGA-KIRC  
472 samples. Examination of RCC genomic subtypes and genome instability revealed a significant  
473 association of high-[c9,c10] and low-[c8] with the good prognosis CC-e.2 subtype but also the low-  
474 wGII category, and, conversely, a significant association of low-[c9,c10] and high-[c8] with the poor  
475 prognosis CC-e3 subtype but also the high-wGII category (Fig. S11, Table S11). With the exception of  
476 low-[c10] and high-[c10] samples, it is noteworthy that these associations remained significant in low-  
477 risk TCGA-KIRC patients.

478 Based on these observations, we then performed a survival analysis in order to investigate whether the  
479 gain of chromosome 12, the RCC genomic subtypes, and the genome instability (wGII score) were  
480 predictive of survival in low-risk patients (Fig. S10, Fig. S12). Only the gain of 12p was found to be  
481 associated with survival in all TCGA-KIRC samples ( $p < 2.8e-3$ ) but not in low-risk samples ( $p < 0.23$ ).  
482 As expected, RCC genomic subtypes and wGII categories were significantly associated with patient  
483 survival ( $p < 1e-4$  and  $1.5e-3$ ). However, contrary to the relative proportions of c8, c9 and c10, these  
484 features could not further stratify low-risk patients into two prognostic subgroups ( $p < 2.6e-1$  and  $p <$   
485  $2.1e-2$ ) (Fig. S10, Fig. S12). Overall, these results corroborate the observation that the c8 clone is  
486 aggressive and associated with poor survival, and further support its consideration as a biomarker of  
487 interest in ccRCC low-grade tumors.

488 **Supplementary references**

- 489 [1] Bertrand E, Judikael S, Aurelie L, Solène-Florence K-J, Nathalie R-L, Frederic C. Human  
490 Kidney Tumor Dissociation for single-cell genomics. *Protocols.io* 2020.  
491 <https://www.protocols.io/view/human-kidney-tumor-dissociation-for-single-cell-g-2n2gdge>  
492 (accessed May 15, 2022).
- 493 [2] Lun ATL, Riesenfeld S, Andrews T, Dao TP, Gomes T, participants in the 1st Human Cell Atlas  
494 Jamboree, et al. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-  
495 cell RNA sequencing data. *Genome Biol* 2019;20:63.
- 496 [3] McCarthy DJ, Campbell KR, Lun ATL, Wills QF. Scater: pre-processing, quality control,  
497 normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 2017;btw777.  
498 <https://doi.org/10.1093/bioinformatics/btw777>.
- 499 [4] McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA  
500 Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* 2019;8:329–37.e4.
- 501 [5] Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR, et al. Single-  
502 cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science*  
503 2018;361:594–9.
- 504 [6] Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-  
505 cell RNA-seq data with Bioconductor. *F1000Res* 2016;5:2122.
- 506 [7] Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data  
507 across different conditions, technologies, and species. *Nat Biotechnol* 2018;36:411–20.
- 508 [8] Kostov B, Bécue-Bertaut M, Husson F. Multiple Factor Analysis for Contingency Tables in the  
509 FactoMineR Package. *The R Journal* 2013;5:29. <https://doi.org/10.32614/rj-2013-003>.
- 510 [9] Chen M, Zhan Q, Mu Z, Wang L, Zheng Z, Miao J, et al. Alignment of single-cell RNA-seq  
511 samples without overcorrection using kernel density matching. *Genome Res* 2021;31:698–712.
- 512 [10] Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional  
513 landscape of mammalian organogenesis. *Nature* 2019;566:496–502.
- 514 [11] Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. *Statistics for*  
515 *Biology and Health* 2000. <https://doi.org/10.1007/978-1-4757-3294-8>.
- 516 [12] Zhao M, Liu Y, Zheng C, Qu H. dbEMT 2.0: An updated database for epithelial-mesenchymal  
517 transition genes with experimentally verified information and precalculated regulation  
518 information for cancer metastasis. *Journal of Genetics and Genomics* 2019;46:595–7.  
519 <https://doi.org/10.1016/j.jgg.2019.11.010>.
- 520 [13] Wang X, Park J, Susztak K, Zhang NR, Li M. Bulk tissue cell type deconvolution with multi-  
521 subject single-cell expression reference. *Nat Commun* 2019;10:380.
- 522 [14] Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. *Protein*  
523 *Science* 2018;27:233–44. <https://doi.org/10.1002/pro.3307>.
- 524 [15] Chen F, Zhang Y, Şenbabaoğlu Y, Ciriello G, Yang L, Reznik E, et al. Multilevel Genomics-  
525 Based Taxonomy of Renal Cell Carcinoma. *Cell Rep* 2016;14:2476–89.
- 526 [16] Turajlic S, Xu H, Litchfield K, Rowan A, Horswell S, Chambers T, et al. Deterministic  
527 Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell*  
528 2018;173:595–610.e11.
- 529 [17] Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAAbiolinks: an  
530 R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016;44:e71.
- 531 [18] Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic  
532 heterogeneity in cancer evolution. *Nature* 2013;501:338–45.  
533 <https://doi.org/10.1038/nature12625>.
- 534 [19] Mengelbier LH, Karlsson J, Lindgren D, Valind A, Lilljebjörn H, Jansson C, et al. Intratumoral  
535 genome diversity parallels progression and predicts outcome in pediatric cancer. *Nat Commun*  
536 2015;6:6125.
- 537 [20] Chalmel F, Primig M. The Annotation, Mapping, Expression and Network (AMEN) suite of  
538 tools for molecular systems biology. *BMC Bioinformatics* 2008;9:86.
- 539 [21] Leonard S, Lardenois A, Tarte K, Rolland A, Chalmel F. FlexDotPlot: a universal and modular

- 540 dot plot visualization tool for complex multifaceted data n.d.  
541 <https://doi.org/10.1101/2020.04.03.023655>.
- 542 [22] Bakken TE, Hodge RD, Miller JA, Yao Z, Nguyen TN, Aevermann B, et al. Single-nucleus and  
543 single-cell transcriptomes compared in matched cortical cell types. *PLoS One*  
544 2018;13:e0209648.
- 545 [23] Baniak N, Flood TA, Buchanan M, Dal Cin P, Hirsch MS. Carbonic anhydrase IX (CA9)  
546 expression in multiple renal epithelial tumour subtypes. *Histopathology* 2020;77:659–66.
- 547 [24] Tang S-W, Yang T-C, Lin W-C, Chang W-H, Wang C-C, Lai M-K, et al. Nicotinamide N-  
548 methyltransferase induces cellular invasion through activating matrix metalloproteinase-2  
549 expression in clear cell renal cell carcinoma cells. *Carcinogenesis* 2011;32:138–45.
- 550 [25] Hakimi AA, Pham CG, Hsieh JJ. A clear picture of renal cell carcinoma. *Nat Genet*  
551 2013;45:849–50.
- 552 [26] Moore LE, Jaeger E, Nickerson ML, Brennan P, De Vries S, Roy R, et al. Genomic copy number  
553 alterations in clear cell renal carcinoma: associations with case characteristics and mechanisms of  
554 VHL gene inactivation. *Oncogenesis* 2012;1:e14.
- 555 [27] Kurtenbach S, Cruz AM, Rodriguez DA, Durante MA, Harbour JW. UPhyloplot2: visualizing  
556 phylogenetic trees from single-cell RNA-seq data. *BMC Genomics* 2021;22:419.
- 557 [28] Bui TO, Dao VT, Nguyen VT, Feugeas J-P, Pamoukdjian F, Bousquet G. Genomics of Clear-cell  
558 Renal Cell Carcinoma: A Systematic Review and Meta-analysis. *Eur Urol* 2022;81:349–61.
- 559 [29] Krock BL, Skuli N, Simon MC. Hypoxia-induced angiogenesis: good and evil. *Genes Cancer*  
560 2011;2:1117–33.
- 561 [30] Arakaki R, Yamasaki T, Kanno T, Shibasaki N, Sakamoto H, Utsunomiya N, et al. CCL2 as a  
562 potential therapeutic target for clear cell renal cell carcinoma. *Cancer Med* 2016;5:2920–33.
- 563 [31] Ruan H, Li X, Yang H, Song Z, Tong J, Cao Q, et al. Corrigendum to “Enhanced expression of  
564 caveolin-1 possesses diagnostic and prognostic value and promotes cell migration, invasion and  
565 sunitinib resistance in the clear cell renal cell carcinoma” [Exp. Cell Res. (2017) 269-278]. *Exp*  
566 *Cell Res* 2018;362:244.
- 567 [32] Cabral-Pacheco GA, Garza-Veloz I, la Rosa CC-D, Ramirez-Acuña JM, Perez-Romero BA,  
568 Guerrero-Rodriguez JF, et al. The Roles of Matrix Metalloproteinases and Their Inhibitors in  
569 Human Diseases. *International Journal of Molecular Sciences* 2020;21:9739.  
570 <https://doi.org/10.3390/ijms21249739>.
- 571 [33] Guo H, German P, Bai S, Barnes S, Guo W, Qi X, et al. The PI3K/AKT Pathway and Renal Cell  
572 Carcinoma. *J Genet Genomics* 2015;42:343–53.
- 573 [34] Corona A, Blobel GC. The role of the extracellular matrix protein TGFBI in cancer. *Cell Signal*  
574 2021;84:110028.
- 575 [35] Filippou PS, Karagiannis GS, Constantinidou A. Midkine (MDK) growth factor: a key player in  
576 cancer progression and a promising therapeutic target. *Oncogene* 2020;39:2040–54.  
577 <https://doi.org/10.1038/s41388-019-1124-8>.
- 578 [36] Young JR, Coy H, Douek M, Lo P, Sayre J, Pantuck AJ, et al. Clear cell renal cell carcinoma:  
579 identifying the gain of chromosome 12 on multiphasic MDCT. *Abdom Radiol (NY)*  
580 2017;42:236–41.
- 581 [37] Pan C-C, Yeh Y-C, Wang Y-C, Chang Y-H. Differential expression analysis of clear cell renal  
582 cell carcinomas in The Cancer Genome Atlas distinguishes an aggressive subset enriched with  
583 chromosomes 7 and 12 gains. *Histopathology* 2020;76:950–8.