Single-cell Deconvolution of a Specific Malignant Cell Population as a Poor Prognostic Biomarker in Low-risk Clear Cell Renal Cell Carcinoma Patients

Judikael R Saout, Gwendoline Lecuyer, Simon Leonard, Bertrand Evrard, S.-F. Kammerer-Jacquet, Laurence Noël, Zine-Eddine Khene, Romain Mathieu, Angelique Brunot, Antoine D. Rolland, et al.

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

Judikael R. Saout[1], Gwendoline Lecuyer[1], Simon Léonard[1,2,3], Bertrand Evrard[1], Solène-Florence Kammerer-Jacquet[1,4], Laurence Noel[1], Zine-Eddine Khene[4], Romain Mathieu[1,5], Angélique Brunot[6], Antoine D. Rolland[1], Karim Bensalah[5], Nathalie Rioux-Leclercq[1,4,¤], Aurélie Lardenois[1,¤], Frédéric Chalmel[1,*]

[1] Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) - UMR_S 1085, F-35000 Rennes, France

[2] LabEx IGO "Immunotherapy, Graft, Oncology", F-35043 Nantes, France

[3] Univ Rennes, INSERM, EFS, UMR S1236, F-35000, Rennes, France

[4] Pathology Department, University Hospital of Rennes, Rennes, France

[5] Urology Department, University Hospital of Rennes, Rennes, France

[6] Department of Medical Oncology, Centre Eugène Marquis, Unicancer, Rennes, France.

* To whom correspondence should be addressed

¤ These authors contributed to the manuscript equally

Correspondence: frederic.chalmel@inserm.fr

Keywords: ccRCC, deconvolution, low-risk progressors, malignancy atlas, single-cell RNA-sequencing, tumor progression

Word count of the abstract: 304

Word count of the text: 2899

Abstract

BACKGROUND: Intra-tumor heterogeneity (ITH) is a key feature in clear-cell renal cell carcinomas (ccRCCs) which impacts outcomes such as aggressiveness, response to treatments or recurrence. In particular, it may explain tumor relapse after surgery in clinically low-risk patients who did not benefit from adjuvant therapy. Recently, single-cell RNA-seq (scRNA-seq) has emerged as a powerful tool to unravel expression intra-tumor heterogeneity (eITH) and might enable better assessment of clinical outcomes in ccRCC.

OBJECTIVE: To explore eITH in ccRCC with a focus on malignant cells (MCs), and assess its relevance to improve prognosis for low-risk patients.

DESIGN, SETTING AND PARTICIPANTS: We performed scRNA-seq on tumor samples from five untreated ccRCC patients ranging from pT1a to pT3b. Data was complemented with a published dataset composed of pairs of matched normal and ccRCC samples.

INTERVENTION: Radical or partial nephrectomy on untreated ccRCC patients.

OUTCOME MEASUREMENTS AND STATISTICAL ANALYSIS: Viability and cell type proportions were determined by flow cytometry. Following scRNA-seq, functional analysis was performed and tumor progression trajectories were inferred. A deconvolution approach was applied on an external cohort and Kaplan-Meier survival curves were estimated with respect to malignant clusters prevalence.

RESULTS AND LIMITATIONS:

We analyzed 54,812 cells and identified 35 cell subpopulations. The eITH analysis revealed that each tumor contained various degrees of clonal diversity. The transcriptomic signatures of MCs in one particularly heterogeneous sample were used to design a deconvolution-based strategy which allowed to risk-stratify 310 low-risk ccRCC patients.

CONCLUSIONS: We described eITH in ccRCCs and used this information to establish significant cell population-based prognostic signatures and better discriminate ccRCC patients. This approach has the potential to improve the stratification of clinically low-risk patients and their therapeutic management.

PATIENT SUMMARY: We sequenced the RNA content of individual cell subpopulations composing ccRCCs and identified specific malignant cells whose genetic information can be used to predict tumor progression.

Introduction

Clear cell renal cell carcinoma (ccRCC) is the most frequent and lethal subtype of kidney cancer which represents 2.2% of all cancers and 180,000 deaths annually, worldwide [1]. It has a poor prognosis, as 45% of diagnosed patients are or become metastatic [2], and 30% of those treated with surgery eventually relapse, including 10-25% of clinically low-risk patients with progressing disease (called low-risk progressors) [3–6]. Recent targeted therapies have provided significant benefits for patients in terms of progression-free and overall survival [7]. However, their efficacy is limited with 30% of responsive patients, contrasting responses within tumors, and long-term relapses in most cases [8–10]. These pitfalls are not only due to inter- but also, importantly, to intra-tumor heterogeneity (ITH) which underlies the emergence of novel malignant clones, allowing the tumor to proliferate, escape treatments or relapse after surgery [11–14]. Single-cell RNA sequencing (scRNA-seq) has become the method of choice to decipher expression intra-tumor heterogeneity (eITH) [15] by simultaneously profiling the individual transcriptome of thousands of cells from a dissociated tissue. It has contributed to demonstrate that eITH is critical for tumor maintenance and progression [15]. In ccRCCs, a pioneering single-cell study provided the first evidence of eITH and revealed differences in targetable signaling pathways between a primary tumor and its metastasis [16]. Other studies have used scRNA-seq to explore renal carcinomas [17–29], providing new insights into tumor subtypes [27], origin and differentiation of RCC cells [26,29], or characterization of the immune landscape [17,23,28]. Most of the datasets used by these studies comprise a large proportion of immune cells (ICs) and only a small fraction of malignant cells (MCs). Hence, there is still a need for single-cell atlases focusing on MC heterogeneity which might provide new insights into the ability of ccRCC to proliferate, escape treatments, metastasize, or relapse. Here we designed and validated a scRNA-seq protocol to characterize the MCs in ccRCC tumors from five patients ranging from pT1a to pT3b, which was complemented with a scRNA-seq dataset composed of three pairs of ccRCC and adjacent normal kidney samples [18]. The subsequent analysis revealed contrasted degrees of heterogeneity within tumors. One particular tumor displayed extensive clonal diversity, recapitulating aggressive features linked to epithelial-mesenchymal transition, cell migration, angiogenesis and cell adhesion. By applying an *in silico* deconvolution approach to a bulk RNA-seq

dataset of ccRCCs, we found that the predicted proportion of a specific MC cluster was a poor prognosis marker in low-risk progressors. Our results provide a valuable resource for the community that has been made available via the Uro-oncogenomics Viewer database (https://uncover.genouest.org) (Lecuyer, Saout *et al.*, in preparation).

Patients and methods

Tumor specimens

Fresh tumor samples, ranging from pT1a to pT3b and ISUP 2 to 4, were collected from a total of five untreated patients who underwent either radical or partial nephrectomy for ccRCC at the urology department of the Rennes university-hospital center. Patients met the requirements of the local institutional ethics committee (see Supplementary methods). Detailed clinical information about each patient is provided in Table S1.

Single-cell preparation, library preparation, sequencing and analysis

For each tumor, ~2-9 punches were pooled and dissociated, and a partial depletion of ICs was performed to enrich the samples in MCs (Fig. 1A; Supplementary methods). Construction of single-cell libraries (Chromium™ Single Cell 3' Library & Gel Bead Kit v2) was carried out according to the manufacturer's instructions (10x Genomics). Single-cell libraries were quality-controlled using the Agilent 2100 Bioanalyzer system and sequenced on an Illumina HiSeq 4000 instrument using a 2x100 paired-end sequencing protocol (Illumina, San Diego, CA). Data processing was performed in R as extensively detailed in Supplementary methods.

Results

Validation of the experimental procedure

We studied five samples (designated RCC1-5) each from separate untreated ccRCC patients with various clinical features, as well as one sample from a vena cava tumor thrombus from the RCC5 patient (RCC5_t) and a sample of RCC4 dissociated cells that were frozen (RCC4_f) (Fig. S1; Fig. S2; Table S1; Supplementary results [SR], Section 1). As our primary objective was to build a single-cell atlas of predominantly MCs, we designed and evaluated, by flow cytometry, a dedicated protocol that showed good dissociation, viability and cell type representativeness (Fig. 1A; Fig. S3; SR, SR-Section 2). Importantly, it allowed us to enrich the MC content of the samples by dramatically reducing the proportion of immune cells, and proved to only slightly affect viability in frozen dissociated cells (Fig. S3; Fig. S4).

A single-cell atlas of ccRCCs composed of ~55,000 cells

After confirming our ability to efficiently dissociate the tumor and enrich MCs, we sequenced the transcriptome of individual cells of seven samples (RCC1-5, RCC4_f and RCC5_t). Our dataset was complemented with three pairs of matched healthy renal and ccRCC tumor tissues (termed RCC[6-8]_n and RCC[6-8], respectively) [18] (Table S1; Table S2). Following data processing, a pseudobulk analysis revealed differences between both datasets (Fig. S5; SR, Section 3), prompting us to perform a batch effect correction. The final scRNA-seq dataset was composed of 54,812 cells (Fig. 1B; SR, Section 3) which were projected on 2D UMAP spaces (Fig. 1C-D).

Single-cell RNA-seq uncovers the cellular composition in ccRCC

Cells were next partitioned into 35 cell clusters (termed c1-c35) (Fig. 2A) which were associated with eight broad cell types based on specific marker genes such as CA9 and NNMT for MCs [30,31] (Table S3; Fig. 2B-C; SR, Section 3). The balance between ICs and MCs across samples clearly distinguished

both datasets (Fig. 2D), stressing the relevance of our approach to enrich MCs. Remaining ICs in RCC4 and RCC4_f were CD45- (PTPRC-) B/plasma cells, which explains their insensitivity to IC depletion (Fig. 2B). To further characterize MCs, we inferred their copy number variations (CNV) which confirmed the expected 3p loss in all samples and revealed sample-specific CNVs (Fig. S6; Fig. S7; SR, Section 3). Overall, these results confirm that ccRCC exhibits both intra- and inter-tumor heterogeneity (SR, Section 3).

Malignant cells display clonal diversity at chromosomal and gene expression levels

To deepen the analysis of MC diversity, we next focused the analysis on the 15,471 MCs and the 3,737 tubular epithelial cells (or TECs; the normal cells from which MCs originate) which were subsequently partitioned into 24 clusters (termed c1-c24) (Fig. 3A; Table S4; SR, Section 4). With the exception of TECs (c1, c12 and c15), the vast majority of cell clusters did not mix across samples suggesting that each tumor had its own evolutionary history (Fig. 3A-B), as highlighted by the inter-tumor heterogeneity in chromosomal aberrations (Fig. S6; SR, section 3). Tumors also harbored various degrees of eITH (Fig. 3B-C). In particular in RCC2, MCs were distributed in five well delineated clusters; c3 (shared with RCC1, RCC3 and RCC4; Fig. 3D), c10, c23, c9 and c8, whereas other samples had a more homogeneous composition. Interestingly, the CNV analysis also revealed that the RCC2 sample carried higher intra-tumor CNV heterogeneity than the other tumors, in particular regarding 3q, 6p, 9q, 14q and 18p, that were lost in c8 but had contrasted status in other RCC2 malignant clusters (Fig. S7). With the exception of c16 and c24, the cell cycle analysis revealed that all clusters were composed of a vast majority of cells in G1 and 9-36% in G2M (Fig. 3E), in accordance with typical observation in solid tumors [32]. Along with CA9 and NNMT, we uncovered specific markers of MCs such as EGLN3, CP, MT3, NPTX2 and TRIB3 (Fig. 3F-G). Expression of cell cycle markers (CKS2, CKAP2L and TOP2A) confirmed the cycling status of c16. As expected, we also found high expression of TEC markers such as ALDOB, MIOX, QDPR and GATM in c1, c12 and c15. We also noticed the expression of PTPRC and CD34 in two small cell clusters (c24 and c22, respectively) corresponding to residual contaminants by immune cells and endothelial cells, respectively.

The RCC2 tumor clonal lineage recapitulates important steps of carcinogenesis

Given the variable degree of eITH observed in each tumor, we next tried to reconstruct tumor lineages within each sample (Fig. S8). RCC2 displayed the longest cellular trajectory going from c3 through c10, c23 and c9, up to c8 (Fig. S8). The chromosomal imbalance analysis revealed that the CNV burden increased along the pseudotime, indicating progressive and cumulative chromosomal alterations, which provides additional elements in favor of a c3 to c8 lineage progression (Fig. 4B). Fifty-two regions were found to be significantly gained or lost along the RCC2 pseudotime, which was further supported by a phylogenetic analysis (SR, Section 5; Fig. S9). We also found 2,229 differentially expressed genes (DEGs) across RCC2 MC clusters and tubular epithelial cells which were subsequently partitioned into nine expression patterns (termed P1-9) (Fig. 4C; Table S5). The functional analysis demonstrated the accumulation of aggressive features along the RCC2 lineage, with a notable peak in c8, related to epithelial-mesenchymal transition (EMT, a crucial initial stage of the metastatic process), angiogenesis, cell migration and cell adhesion (Table S6-S7; SR, Section 6). Overall these results suggest that MCs from RCC2 are linked together through a common tumor clonal lineage from indolent (c3) to aggressive (c8) MC clusters showing the sequential acquisition of typical molecular features corresponding to important steps of tumor progression.

Deconvolution of specific malignant clusters is associated with tumor stages and survival in low-risk patients

We next hypothesized that the RCC2 lineage could represent an *in vivo* model of ccRCC carcinogenesis which could be used as a reference. To verify this, we deconvolved the proportions of somatic cell types and RCC2 MC clusters on the TCGA-KIRC bulk RNA-seq dataset (525 samples) of ccRCCs (Fig. 5A, Table S8). Substantial proportions of clusters c8 and c9 as well as non-immune microenvironment cells (endothelial cells, pericytes, myofibroblasts) were significantly detected, covering all four T stages. Despite their cellular proximity with MCs, TECs were not detected in ccRCC tumors, as expected.

Concerning MCs, the main deconvolved populations were c9 and c8 with opposite trends (Fig. 5A, Table S8), by respectively decreasing and increasing with tumor stages. We then evaluated the ability of deconvolved cell types to predict survival in either all or in T1-T2 low-risk ccRCC patients of the TCGA-KIRC cohort (SR, section 7, Fig. 5B, Table S9). When considering all patients, we found a significant association with poor prognosis of c8, and, to a lesser extent, of myofibroblasts and B cells. Conversely, c9, c10 and endothelial cells were three excellent predictors of good prognosis. When looking more specifically at low-risk patients, strikingly, only c8, c9 and c10, and, to a lesser extent, endothelial cells were still associated with survival. In addition, enrichment analysis of recurrent genomic features in TCGA-KIRC cohort (SR, section 8) revealed that the proportion of c8 was significantly associated with unfavorable tumors (CC-e.3 subtype and high genome instability) and the proportions of c9/c10 with favorable tumors (CC-e.2 and low genome instability), including in low-risk patients (Fig. S10, Fig. S11, Fig. S12, Table S10, Table S11). However, contrary to the relative proportions of c8, c9 and c10, previous RCC genomic prognosis subtypes and genome instability could not further stratify low-risk patients into two prognostic subgroups.

Discussion

On the basis of conventional diagnosis, a significant proportion of patients considered to be at low-risk on the basis of conventional diagnosis will eventually relapse [6,33]. A major issue in the management of ccRCCs hence consists in improving patient stratification through early detection of those who will eventually relapse and could have benefited from specific surveillance and potential adjuvant treatments. Biomarkers can be used to stratify these patients, but they remain insufficient. Moreover, the differences between gene expression signatures and corresponding protein detection constitute an important gap for clinical diagnosis [34,35], prompting the use of alternative and complementary approaches, such as transcriptomics in future diagnostic tools. In this context, it remains important to improve our understanding of tumor progression in ccRCCs in order to enhance patient care and consider personalized therapeutic strategies. To address these issues, we used scRNA-seq to decipher eITH in ccRCCs, with a specific focus on the malignant compartment, hypothesizing that this information could notably contribute to improve the detection of low-risk progressors.

Since published scRNA-seq datasets in ccRCCs are often focused on the characterization of the immune compartment [17,19,23,28], we established an experimental procedure to enrich the fraction of MCs by reducing the fraction of ICs. We also examined several punches per tumor to better capture eITH as suggested by Gerlinger and colleagues [14]. Samples from five ccRCC patients ranging from pT1a to pT3b enabled us to assemble an atlas composed of ~25,826 cells (including ~48.3% of MCs) which we complemented with 28,986 cells (including ~10.3% of MCs) published by Young *et al*. [18]. Despite common molecular events such as the truncal 3p loss, the analysis of MCs revealed an important heterogeneity across samples. Various degrees of eITH in the distinct samples were observed, from low (RCC1, 6 and 7) to very high (RCC2 and 5 to a lower extent). We also noticed that both the inter- and intra-tumor heterogeneities in the samples of Young *et al* were lower than in our dataset, possibly due to the difference in sequencing depths, the number of punches per tumor and/or the higher proportion of MCs in our samples. Those differences appear as potentially critical features to properly capture eITH. While the overall CNV detection also revealed differences between datasets and turns out to be

informative in our analysis, it should be considered with caution as CNVs were inferred from gene expression data and not from gold standard DNA analysis techniques (FISH, aCGH). Together these results suggest that ITH is not necessarily the rule, as pointed out in hepatocellular carcinoma [36]. Further works will be required to draw a complete picture of ccRCC heterogeneities notably by including more tumors along with their paired metastases but also by combining single-cell multiomics with spatial transcriptomics technologies.

Among the different patients, we could identify a unique tumor sample (RCC2), displaying a yet never observed eITH consisting of five distinct MC subpopulations. Those were connected through a common lineage characterized by the progressive acquisition of highly aggressive features related to EMT, angiogenesis, cell migration and cell adhesion, as well as specific chromosomal abnormalities. RCC2 was a localized and apparently low-grade tumor (pT2b) without clinical/histological markers of aggressiveness. Yet, the RCC2 patient is currently under surveillance after having developed a suspicious pancreatic nodule, a preferred site for ccRCC metastasis. We believe that the RCC2 MC lineage can therefore represent a valuable resource for the community as an *in vivo* and representative model of tumor progression.

A landmark study used cell type deconvolution from TCGA data to successfully determine three groups of ccRCC patients according to lymphocyte T cell states and demonstrated their significant association with prognosis [21]. Here we used a similar approach based on MC subpopulations composing the RCC2 tumor. Our results proved useful to predict survival and confirmed previous findings encouraging the consideration of eITH to identify predictive biomarkers [37]. The ability of this method to stratify ccRCC patients from bulk RNA-seq is remarkable, especially the capacity of the predicted proportion of c8 and c9 MC clusters to serve as poor and good prognostic factors, respectively. In c9, despite the expression of aggressive features, the marked organization of adaptive immunity seems to confer an overall good clinical outcome. Importantly, some deconvolved cell types constituting the tumor-microenvironment are already known to be associated with survival [38–40]. Here, not only do we confirm these results, but we go further by obtaining refined and resolutive malignant signatures with superior predictive power, especially in low-risk patients. Furthermore, although significantly associated

with relevant RCC genomic subtypes defined by Chen *et al*. [41], we demonstrated that our MC-deconvolved approach better stratify low-risk patients. Interestingly, the proportion of T1-T2 patients (~10%) with a relatively high proportion of c8 is similar to the proportions of clinically low-risk progressors (10-25%), suggesting a potential overlap between the predicted and actual populations with poor outcomes. Our results suggest that low-grade tumors exceeding ~25% of c8 cells warrant reinforced surveillance. Hence, we present here the first proof of concept for using MC expression signatures to improve clinical management of ccRCC patients. Importantly a tumor that better recapitulates the stages of carcinogenesis could also complement and/or update our current deconvolution-based strategy. In the future it will be interesting to develop nomograms including relevant MC clusters and compare their accuracy with other conventional nomograms such as one recently developed in ccRCCs, based on scRNA-seq data [42]. Finally, it will be necessary to explore, in dedicated cohorts, to what extent these deconvolution-based approaches can predict other features such as relapse, metastasis and treatment response. Therefore, we strongly encourage the use of transcriptomics in future clinical routine, especially since low-cost technologies such as BRB-seq [43] are now available and easily deployable, in order to enable the use of such predictive strategies and develop a personalized medicine. In this study we have assembled a valuable atlas of ccRCC MCs; we are committed to expanding this dataset and to making it available online via the UncoVer resource. Thanks to the single-cell profiling of a rare tumor showing a high eITH, we have established a reliable stratification approach for low-risk patients based on the deconvolution of specific MC subpopulations which could contribute to improving patient care.

Conclusion

We described eITH in ccRCCs and used this information to establish significant cell population-based prognostic signatures and better discriminate ccRCC patients. This approach has the potential to improve the stratification of clinically low-risk patients and their therapeutic management.

Author's contributions

FC designed and supervised the study. JS, BE, SFKJ, RM, AB, ADR, ZEK, KB, AL and NRL acquired the data. JS, GL, SL, ADR, AL and FC analyzed and interpreted the data. JS and FC drafted the manuscript. GL, SL, BE, FC, AL, ADR and NRL participated in the critical revision of the manuscript for important intellectual content. JS, GL, SL, AL and FC performed the statistical analyses. FC, AB and NRL obtained fundings. LN provided administrative and technical support.

Funding

This study was funded by La ligue contre le cancer committees 22, 35 and 81 and by the clinical and translational research committee (CORECT) of the Rennes University Hospital Center. S.L. was supported by the chair "Cancer & Innovation" of Rennes 1 Foundation (https://fondation.univ-rennes1.fr/) and a specific grant from the LabEx IGO program (n° ANR-11-LABX-0016) funded by the «Investment into the Future» French Government program, managed by the National Research Agency (ANR).

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgements

We thank all the staff of the Pathology and the Urology Departments of the Rennes University Hospital for their expert assistance, and the participating patients. We thank the GenOuest bioinformatics facility, especially Olivier Collin, Olivier Sallou and Matéo Boudet, for hosting the software. We thank the SciLicium company, and especially Thomas Darde, for their assistance in developing the UncoVer website. We thank the H2P2 and the CytomeTri platforms from Biosit (SFR UMS CNRS 3480 - INSERM 018), especially Laurent Deleurme and Marion Mandon for their precious help and services. For their valuable help, our special thanks also go to Catherine Lavau, William Taylor, Stéphane Boutet (10X Genomics), Marjorie Gournay, Mikael Roussel and Frédéric Dugay. The results published here

are in part based upon data generated by the Cancer Genome Atlas pilot project established by the NCI and NHGRI (information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at hhtp://cancergenome.nih.gov/). We thank all members of the Research Institute for Environmental and Occupational Health (IRSET) for stimulating discussions. We also thank the Institut national de la santé et de la recherche médicale (Inserm), the Université de Rennes 1, and the French School of Public Health (EHESP) for supporting this work.

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians 2021;71:209–49. https://doi.org/10.3322/caac.21660.
[2] Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, et al. Renal cell carcinoma. Nat Rev Dis Primers 2017;3:17009.
[3] Nerich V, Hugues M, Nai T, Stein U, Hon TNT, Montcuquet P, et al. Clinical impact of targeted therapies in patients with metastatic clear-cell renal cell carcinoma. OncoTargets and Therapy 2014:365. https://doi.org/10.2147/ott.s56370.
[4] Rini BI. Metastatic renal cell carcinoma: many treatment options, one patient. J Clin Oncol 2009;27:3225–34.
[5] Sorbellini M, Kattan MW, Snyder ME, Reuter V, Motzer R, Goetzl M, et al. A postoperative prognostic nomogram predicting recurrence for patients with conventional clear cell renal cell carcinoma. J Urol 2005;173:48–51.
[6] Makhov P, Joshi S, Ghatalia P, Kutikov A, Uzzo RG, Kolenko VM. Resistance to Systemic Therapies in Clear Cell Renal Cell Carcinoma: Mechanisms and Management Strategies. Mol Cancer Ther 2018;17:1355–64.
[7] Ross K, Jones RJ. Immune checkpoint inhibitors in renal cell carcinoma. Clin Sci 2017;131:2627–42.
[8] Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. N Engl J Med 2007;356:115–24.
[9] Molina AM, Lin X, Korytowsky B, Matczak E, Lechuga MJ, Wiltshire R, et al. Sunitinib objective response in metastatic renal cell carcinoma: analysis of 1059 patients treated on clinical trials. Eur J Cancer 2014;50:351–8.
[10] Crusz SM, Tang YZ, Sarker S-J, Prevoo W, Kiyani I, Beltran L, et al. Heterogeneous response and progression patterns reveal phenotypic heterogeneity of tyrosine kinase inhibitor response in metastatic renal cell carcinoma. BMC Med 2016;14:185.
[11] López JI, Angulo JC. Pathological Bases and Clinical Impact of Intratumor Heterogeneity in Clear Cell Renal Cell Carcinoma. Curr Urol Rep 2018;19:3.
[12] McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. Cancer Cell 2015;27:15–26.
[13] Raatz M, Shah S, Chitadze G, Brüggemann M, Traulsen A. The impact of phenotypic heterogeneity of tumour cells on treatment and relapse dynamics. PLoS Comput Biol 2021;17:e1008702.
[14] Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366:883–92.
[15] Suvà ML, Tirosh I. Single-Cell RNA Sequencing in Cancer: Lessons Learned and Emerging

Challenges. Mol Cell 2019;75:7–12.
[16] Kim K-T, Lee HW, Lee H-O, Song HJ, Jeong DE, Shin S, et al. Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. Genome Biol 2016;17:80.
[17] Bi K, He MX, Bakouny Z, Kanodia A, Napolitano S, Wu J, et al. Tumor and immune reprogramming during immunotherapy in advanced renal cell carcinoma. Cancer Cell 2021;39:649–61.e5.
[18] Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, Ferdinand JR, et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science 2018;361:594–9.
[19] Borcherding N, Vishwakarma A, Voigt AP, Bellizzi A, Kaplan J, Nepple K, et al. Mapping the immune environment in clear cell renal carcinoma by single-cell genomics. Commun Biol 2021;4:122.
[20] Fendler A, Bauer D, Busch J, Jung K, Wulf-Goldenberg A, Kunz S, et al. Inhibiting WNT and NOTCH in renal cancer stem cells and the implications for human patients. Nat Commun 2020;11:929.
[21] Hu J, Chen Z, Bao L, Zhou L, Hou Y, Liu L, et al. Single-Cell Transcriptome Analysis Reveals Intratumoral Heterogeneity in ccRCC, which Results in Different Clinical Outcomes. Mol Ther 2020;28:1658–72.
[22] Argüello RJ, Combes AJ, Char R, Gigan J-P, Baaziz AI, Bousiquot E, et al. SCENITH: A Flow Cytometry-Based Method to Functionally Profile Energy Metabolism with Single-Cell Resolution. Cell Metab 2020;32:1063–75.e7.
[23] Braun DA, Street K, Burke KP, Cookmeyer DL, Denize T, Pedersen CB, et al. Progressive immune dysfunction with advancing disease stage in renal cell carcinoma. Cancer Cell 2021;39:632–48.e8.
[24] Liu K, Gao R, Wu H, Wang Z, Han G. Single-cell analysis reveals metastatic cell heterogeneity in clear cell renal cell carcinoma. J Cell Mol Med 2021;25:4260–74.
[25] Yu Z, Lu W, Su C, Lv Y, Ye Y, Guo B, et al. Single-Cell RNA-seq Identification of the Cellular Molecular Characteristics of Sporadic Bilateral Clear Cell Renal Cell Carcinoma. Front Oncol 2021;11:659251.
[26] Zhang Y, Narayanan SP, Mannan R, Raskind G, Wang X, Vats P, et al. Single-cell analyses of renal cell cancers reveal insights into tumor microenvironment, cell of origin, and therapy response. Proc Natl Acad Sci U S A 2021;118. https://doi.org/10.1073/pnas.2103240118.
[27] Su C, Lv Y, Lu W, Yu Z, Ye Y, Guo B, et al. Single-Cell RNA Sequencing in Multiple Pathologic Types of Renal Cell Carcinoma Revealed Novel Potential Tumor-Specific Markers. Front Oncol 2021;11:719564.
[28] Krishna C, DiNatale RG, Kuo F, Srivastava RM, Vuong L, Chowell D, et al. Single-cell sequencing links multiregional immune landscapes and tissue-resident T cells in ccRCC to tumor topology and therapy efficacy. Cancer Cell 2021;39:662–77.e6.
[29] Young MD, Mitchell TJ, Custers L, Margaritis T, Morales-Rodriguez F, Kwakwa K, et al. Single cell derived mRNA signals across human kidney tumors. Nat Commun 2021;12:3896.
[30] Baniak N, Flood TA, Buchanan M, Dal Cin P, Hirsch MS. Carbonic anhydrase IX (CA9) expression in multiple renal epithelial tumour subtypes. Histopathology 2020;77:659–66.
[31] Tang S-W, Yang T-C, Lin W-C, Chang W-H, Wang C-C, Lai M-K, et al. Nicotinamide N-methyltransferase induces cellular invasion through activating matrix metalloproteinase-2 expression in clear cell renal cell carcinoma cells. Carcinogenesis 2011;32:138–45.
[32] Terz JJ, Lawrence W Jr, Cox B. Analysis of the cycling and noncycling cell population of human solid tumors. Cancer 1977;40:1462–70.
[33] Parasramka M, Serie DJ, Asmann YW, Eckel-Passow JE, Castle EP, Stanton ML, et al. Validation of Gene Expression Signatures to Identify Low-risk Clear-cell Renal Cell Carcinoma Patients at Higher Risk for Disease-related Death. Eur Urol Focus 2016;2:608–15.
[34] Buccitelli C, Selbach M. mRNAs, proteins and the emerging principles of gene expression control. Nat Rev Genet 2020;21:630–44.
[35] Gygi SP, Rochon Y, Robert Franza B, Aebersold R. Correlation between Protein and mRNA Abundance in Yeast. Molecular and Cellular Biology 1999;19:1720–30.

https://doi.org/10.1128/mcb.19.3.1720.
[36] Xue R, Li R, Guo H, Guo L, Su Z, Ni X, et al. Variable Intra-Tumor Genomic Heterogeneity of Multiple Lesions in Patients With Hepatocellular Carcinoma. Gastroenterology 2016;150:998–1008.
[37] Gulati S, Martinez P, Joshi T, Birkbak NJ, Santos CR, Rowan AJ, et al. Systematic evaluation of the prognostic impact and intratumour heterogeneity of clear cell renal cell carcinoma biomarkers. Eur Urol 2014;66:936–48.
[38] Edeline J, Mottier S, Vigneau C, Jouan F, Perrin C, Zerrouki S, et al. Description of 2 angiogenic phenotypes in clear cell renal cell carcinoma. Hum Pathol 2012;43:1982–90.
[39] Ambrosetti D, Coutts M, Paoli C, Durand M, Borchiellini D, Montemagno C, et al. Cancer-associated fibroblasts in renal cell carcinoma: implication in prognosis and resistance to anti-angiogenic therapy. BJU Int 2022;129:80–92.
[40] Li S, Huang C, Hu G, Ma J, Chen Y, Zhang J, et al. Correction: Tumor-educated B cells promote renal cancer metastasis via inducing the IL-1β/HIF-2α/Notch1 signals. Cell Death Dis 2022;13:415.
[41] Chen F, Zhang Y, Şenbabaoğlu Y, Ciriello G, Yang L, Reznik E, et al. Multilevel Genomics-Based Taxonomy of Renal Cell Carcinoma. Cell Rep 2016;14:2476–89.
[42] Xia Z-N, Wu J-G, Yao W-H, Meng Y-Y, Jian W-G, Wang T-D, et al. Identification of a differentiation-related prognostic nomogram based on single-cell RNA sequencing in clear cell renal cell carcinoma. Sci Rep 2022;12:10973.
[43] Alpern D, Gardeux V, Russeil J, Mangeat B, Meireles-Filho ACA, Breysse R, et al. BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. Genome Biol 2019;20:71.

Figure legends

Figure 1. Single-cell analysis of ccRCC. (A) Flowchart depicting the experimental design for each ccRCC sample. Single-cell suspensions were assessed by flow cytometry at two steps in protocol: after tumor dissociation (step 1), and, after low centrifugations and immune cell removal (step 2). (B) Bar plots representing the number of cells and the median number of detected genes for each sample. Each sample corresponds to a distinct color, with normal samples shown in shades of gray. _f = frozen sample; _t = thrombus sample; _n = normal kidney. (C) UMAP plot of all cells from pooled samples colored according to panel B, in which nearby points represent transcriptionally similar cells. (D) Deconvolution of the global UMAP plot (panel C) per patient. In each UMAP plot, the cells are colored by sample according to panel B.

Figure 2. Cellular diversity in ccRCC. (A) UMAP plot representing the cell type distribution (main colors) and clustering (shades of color) of all cells. (B) Expression levels of cell population markers. (C) Dot plot representing the expression of known cell type markers in each cell cluster. MCs = Malignant cells; TECs = Tubular epithelial cells; CDEC = Collecting duct epithelial cells. (D) Stacked bar plot illustrating the distribution of the main cell types in each sample. RCC4_f = frozen sample; _t = thrombus sample; _n = normal kidney sample.

Figure 3. Characterization of ccRCC malignant cells and normal tubular epithelial cells. (A) UMAP plots and clustering of all malignant cells. (B) Deconvolution of the global UMAP plot (panel A) per patient. Cells are colored by sample as indicated previously (Figure 1B). MCs = Malignant cells ; TECs = Tubular epithelial cells; TCs = Thrombus cells. (C) Distribution of clusters per sample. _t = thrombus sample; _n = normal kidney. (D) Distribution of samples per cluster. (E) Distribution of cells with S, G1 or G2M status per cluster. (F) Dot plot representing the expression of differentially expressed genes (DEGs) in malignant cell clusters. (G) Expression of DEGs as UMAP plots. Colors correspond to the scaled expression of panel F.

Figure 4. Intra-tumor heterogeneity of RCC2 malignant cells. (A) UMAP plot showing pseudotime values as computed by Monocle3 in shades of red and the lineage trajectory in RCC2 malignant cells as

a black line. Here, the pseudotime is an abstract value describing the distance of a cell from the beginning of the lineage (green dot) in terms of transcriptomic variations, hence partially depicting tumor progression. (B) Violin plots of proportions of RCC2 MC clusters (left) and of the CNV burden (right) along the pseudotime. (C) Heatmap representation of the scaled expression values of 2,229 genes that are differentially expressed across RCC2 malignant cell and tubular epithelial cell clusters. These were ordered according to the pseudotime and were further colored given their CNV burden. DEGs are organized in seven patterns according to their expression in RCC2 malignant clusters. Enriched GO terms associated with each gene expression pattern are presented along with their associated adjusted *p values*. Shades of gray in annotation boxes on the left of heatmap correspond to different levels of significance for enrichment. Selection of genes for EMT and outcome (Favorable and Unfavorable) annotations is detailed in supplementary methods.

Figure 5. Deconvolution and survival analysis. (A) Density plots showing the distribution of percentages of the most prevalent deconvolved cell types and RCC2 malignant cells in ccRCC TCGA patients divided into T stages. Malignant clusters c8, c9 and c10, correspond to relative malignant clusters as defined in supplementary methods. MCs = Malignant cells; TECs = Tubular epithelial cells. (B) Kaplan-Meier survival curves of patients according to their low (blue curves) or high (red curves) prevalence of the deconvolved cell types. The analysis was performed for either all ccRCC patients (upper plots) or on lower risk (T1-T2) patients (lower plots). Cell types with significant association (log rank p value inferior to 0.01, in bold) in all patients are presented.

Supplementary figure legends

Fig. S1. Histology of the six primary ccRCC specimens used in the study. Sections were obtained from formalin-fixed, paraffin-embedded tissue stained with hematoxylin and eosin. ISUP = International Society of Urological Pathology ; pT = Pathological tumor.

Fig. S2. Processing of macroscopic tumor samples. (A) Macroscopic view of tumor samples. (B) Up to 9 distinct macroscopic areas were sampled in each tumor specimen. (C) Each punch was separated into three parts : a quarter for snap freezing, a quarter for tissue fixation, and one half which was pooled with other half punches for tissue dissociation.

Fig. S3. Assessment of quality at different dissociation steps as illustrated with the RCC4 sample. (A) Microscopic pictures (x10) of single cells after dissociation. (B) Flow cytometry cytograms of single cell suspensions. From left to right are the side scatter versus forward scatter gating, the live cell gating (side scatter versus 7AAD staining), and the proportions of different cell types based on the expression of CD34 (endothelial marker) and CD45 (immune cell marker). Malignant and stromal cells are negative for CD34 and CD45, hence located in the bottom left section. (C-D) Cell count and cell viability in each sample. (E-F) Proportions and cell viability of each monitored cell population before and after immune cell depletion. (G) Proportions of immune cells measured by flow cytometry after tumor dissociation or estimated by immunohistochemistry (IHC, prior to dissociation). (H) Cell viability of fresh versus frozen single cells (n=6).

Fig. S4. Illustration of the automatic counting method performed on IHC sections (LCA immune cell labeling) using ImageJ software. (A) Raw picture. (B) Picture with detection (red) of labeled immune cells. (C) Picture with detection (red) of unlabeled non-immune cells.

Fig. S5. Pseudo-Bulk analysis of scRNA-seq data. (A) Correlation matrix comparing all samples. _f = frozen sample; _t = thrombus sample; _n = normal kidney. (B) PCA space projection of all samples. _f = frozen sample; _t = thrombus sample; _n = normal kidney.

Fig. S6. Heatmap representation of chromosome gains or losses estimated by inferCNV in all tumors. (A) CNV in malignant cells. f = frozen sample; _t = thrombus sample; _n = normal kidney. (B)

CNV in normal cells. f = frozen sample; _t = thrombus sample; _n = normal kidney.

Fig. S7. CNVs in malignant and tubular epithelial cells. (A) UMAP plot and clustering of all malignant cells. Gray cells correspond to tubular epithelial cells whereas colored cells correspond to malignant cells. (B) Dot plot showing computed gains (red) or losses (blue) of chromosome arms in malignant cell clusters. (C) UMAP plots presenting the distribution of CNVs across malignants cells.

Fig. S8. Lineage trajectories in malignant and tubular epithelial cells of each tumor, as computed with Monocle3. Lineages begin at the green dot corresponding to the cells with the least chromosomal aberrations. In the resulting trajectories, the degrees of branching match the levels of heterogeneity.

Fig. S9. RCC2 malignant cells hierarchy. (A) Heatmap of CNV in RCC2 malignant cell clusters. (B) Phylogenetic tree of RCC2 malignant cells. The phylogenetic tree was constructed using the Uphyloplot2 R package and is based on imputed CNVs from inferCNV. Percentages in brackets indicate the overlap between malignant cells defined with Seurat and Uphyloplot2. (+) : chromosomal gain; (-) : chromosomal loss.

Figure S10. Association of weighted genome instability index (wGII) and survival. Kaplan-Meier survival curves of TCGA-KIRC patients according to their weighted genome instability index (wGII).

Figure S11. Enrichment of recurrent ccRCC genomic features inTGCA patients according to RCC2 deconvolved malignant subpopulations. Dot plot representing significantly enriched CNVs and RCC genomic subtypes (as defined by Chen *et al.* 2021) in the TCGA patients according to their low or high prevalence of deconvolved c8, c9 and c10 malignant subpopulations. The analysis was performed for either all ccRCC patients (All) or lower risk (T1-T2) patients (Low risk). Red stars indicate recurrent CNVs in ccRCC as described in landmark studies (Beroukhim et al., 2019; Cancer Genome Atlas Research Network, 2013).

Figure S12. Association of enriched genomic features and survival. Kaplan-Meier survival curves of patients according to either their CNV profiles on chromosome 12, or RCC genomic subtypes (as defined by Chen *et al.* 2021). The analysis was performed for either all ccRCC patients (plots on the left) or on lower risk (T1-T2) patients (plots on the right).

Supplementary table legends

Table S1. Clinical features of the patients' tumors. A total of 8 tumors were included in this study, from current (Saout et al.) and additional (Young et al.) datasets. ccRCC = clear cell renal cell carcinoma ; ISUP = International Society of Urological Pathology ; pTNM = Tumor-Node-Metastasis.

Table S2. Characteristics of single-cell RNA-seq samples. Both experimental and sequencing data are presented for each sample. RCC4_f = RCC4 frozen cells ; RCC5_t = RCC5 tumor thrombus ; RCC[6-8]_n = RCC[6-8] adjacent normal tissue ; IC = Immune Cells ; LCA = Leukocyte Common Antigen ; IHC = ImmunoHistoChemical. Step 1 and step 2 refer to the malignant cell enrichment protocol steps described in the Patients and Methods section. Step 1 is after tumor dissociation and step 2 is after low speed centrifugations and immune cells removal.

Table S3. Top differentially expressed genes (DEGs) between the 34 clusters in the total dataset. DEGs were identified using the FindAllMarkers command from R package Seurat with default parameters. Pct.1 = percentage.

Table S4. Top differentially expressed genes (DEGs) in the 24 clusters of malignant and tubular epithelial cells. DEGs were identified using the FindAllMarkers command from R package Seurat with default parameters.

Table S5. Top differentially expressed genes (DEGs) in the 5 clusters of malignant cells from RCC2. DEGs were identified by pairwise comparisons between clusters using the FindMarkers command from R package Seurat.

Table S6. Top enriched gene ontology (GO) terms associated with gene expression patterns in RCC2 malignant cell clusters. Gene expression patterns were found to be enriched in 884 biological processes terms from the gene ontology. Custom terms related to TCGA prognostic and EMT involvement categories were tested for enrichment (see supplementary methods). r is the number of genes related to the GO term in a given pattern whereas R is the total number of genes related to the GO term.

Table S7. Details of annotations of RCC2 differentially expressed genes (DEGs). Genes are presented as being associated with different gene ontology (GO) terms and annotations. Genes in the EMT column are either tumor suppressors ("Tsup"), oncogenes ("Onco"), both ("Dual"), related to EMT ("other") or not ("unrelated"). The number of pubmed references was determined with the search "renal cell carcinoma" and the gene name in PubMed. The TDL category designates genes expressing surface proteins (GESPs) with their level of Targetability development ("TDL" or "Target Development Level"). These genes encode proteins through which approved drugs act ("Tclin") (i.e., mode-of-action drug targets), or are either known to bind small molecules with high potency ("Tchem"), have well-studied biology ("Tbio"), or are understudied ("Tdark"), as described elsewhere (PMID: 31898878). Genes were identified as favorable or unfavorable according to the processing of the Cancer Genome Atlas Kidney Renal Cell Carcinoma (TCGA-KIRC) data described in Supplementary Material.

Table S8. Results of deconvolution of cell types in TCGA-KIRC cohort bulk RNA-seq data, using the Music R package. Cell type estimations are indicated in percentages. Among malignant cell clusters, "Others" correspond to few RCC2 malignant cells that were in clusters other than c3, c10, c23, c9 or c8. Relative malignant cell clusters are initial percentages of the considered malignant cells divided by the total number of RCC2 malignant cells.

Table S9. Survival analysis results. Both the Cox Proportional-Hazards Model and Log Rank test methods were used to evaluate the significance of associations of deconvolved cell types with survival. An association was considered significant when the p value was inferior to 0.01.

Table S10: Genomic features of TCGA-KIRC samples. Genomic features used in the study are listed in the table below. Details of their obtaining and calculations are provided in supplementary materials.

Table S11. Enrichment analysis of genomic features in TCGA-KIRC data for malignant populations significantly associated with patient survival. Enrichment results for genomics features (mutations, CNVs and methylations), RCC subtypes and weighted genome instability index (wGII) in patients grouped based on the low or high prevalence of c8, c9 and c10 malignant population. Details of enrichment method and calculation are provided in supplementary methods.

Figures

Fig. 1

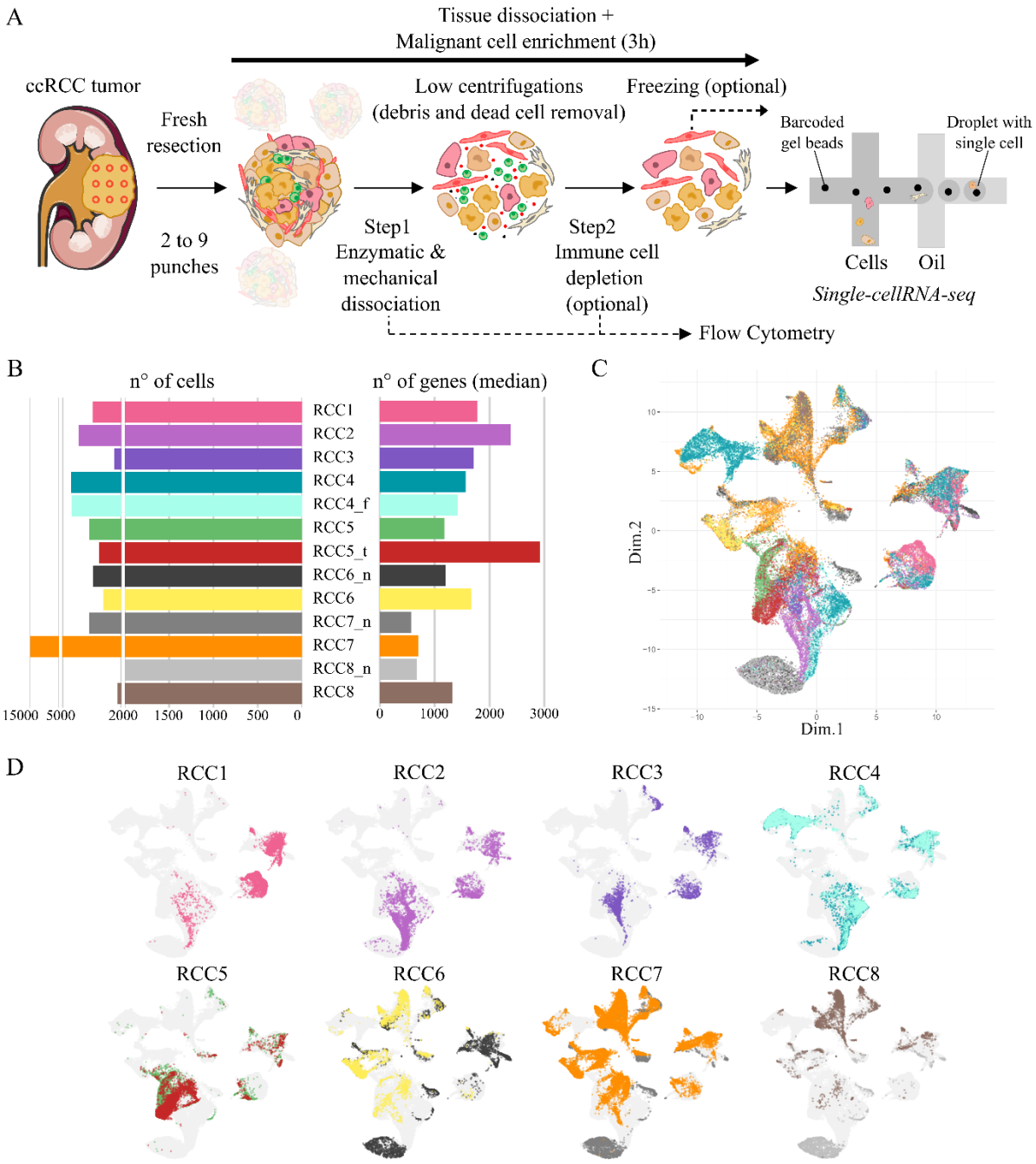

Fig. 2

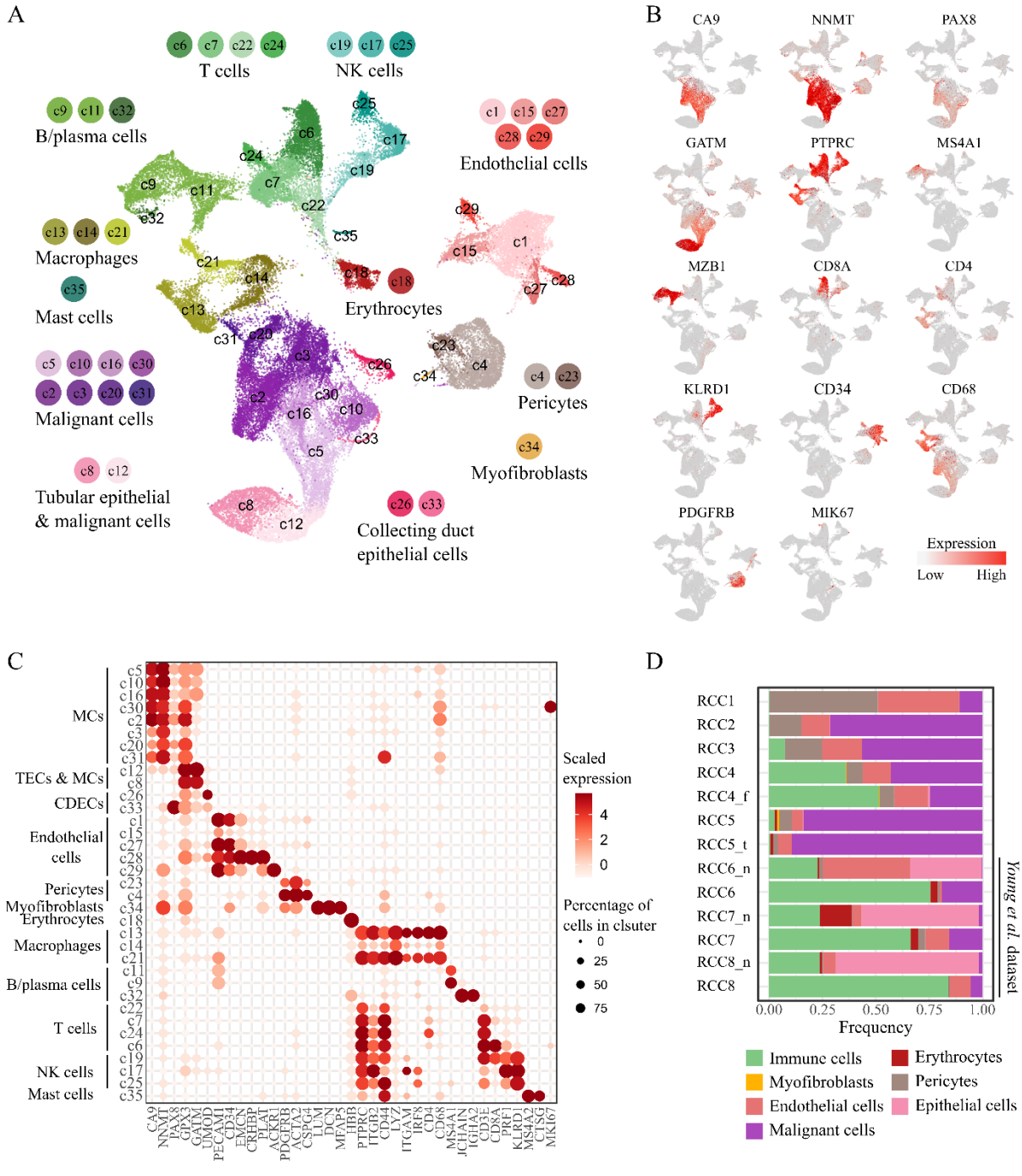

Fig. 3

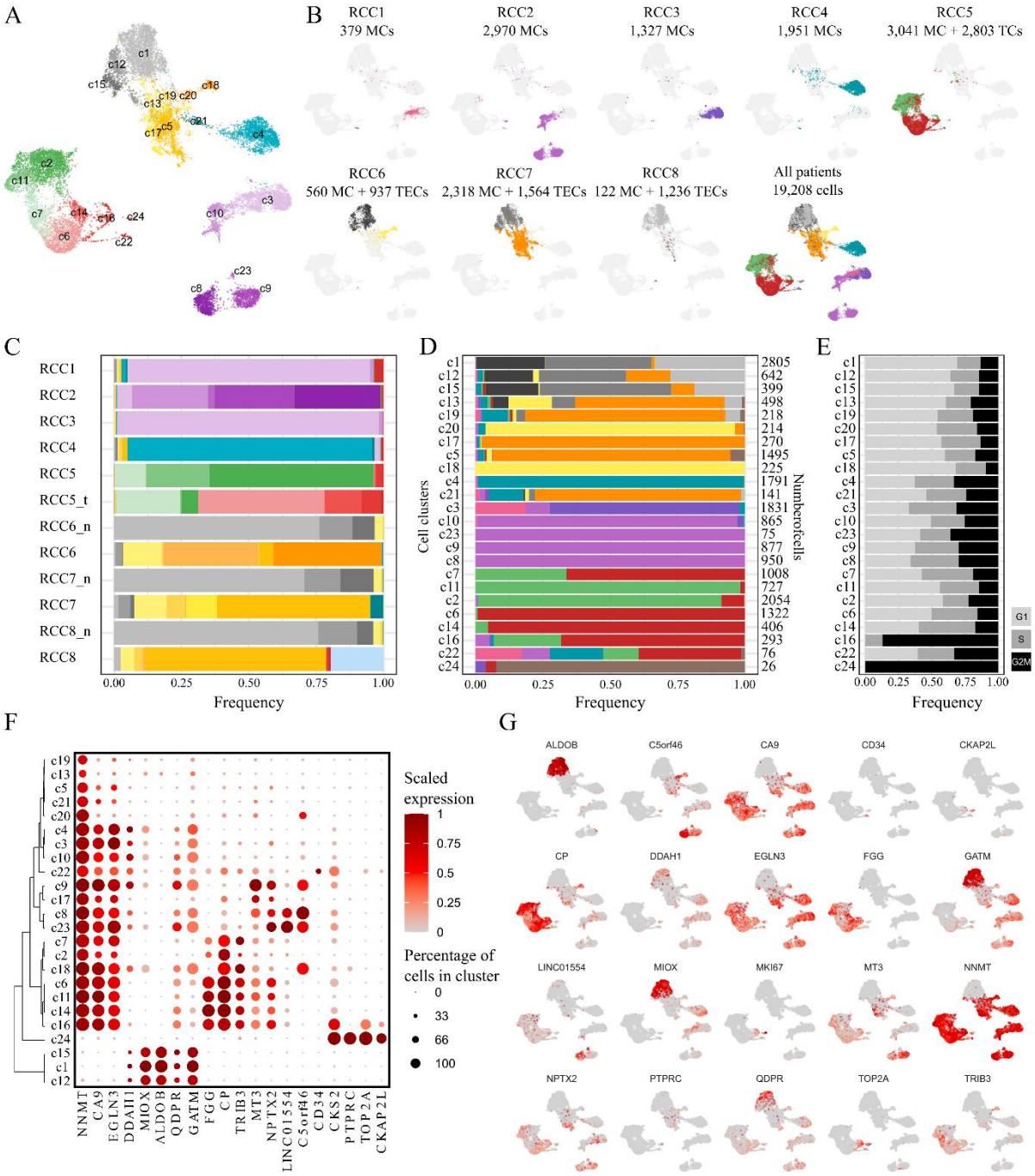

Fig. 4

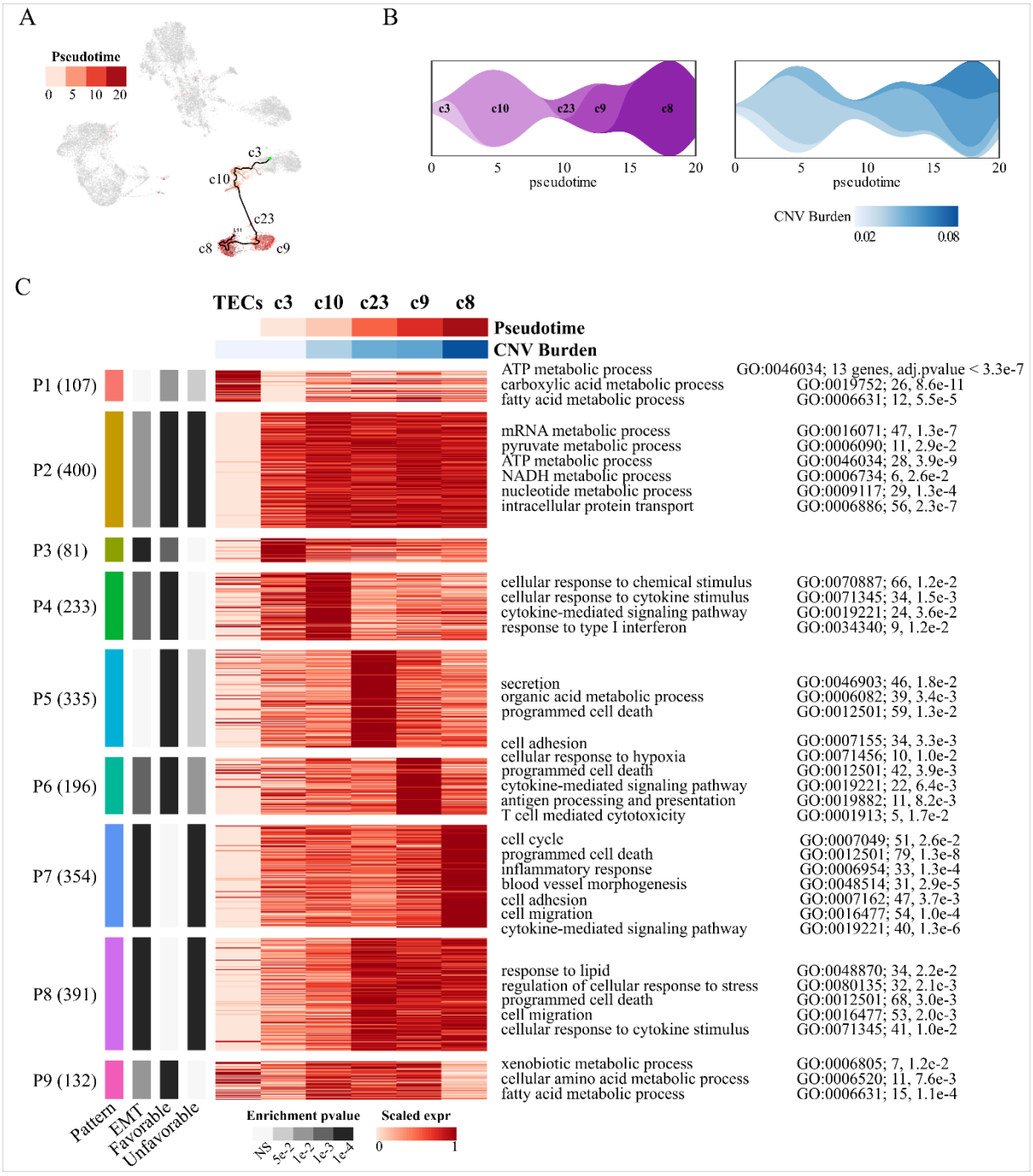

Fig. 5