Title: Mapping Immune Landscape in Clear Cell Renal Carcinoma by Single-Cell Genomics

**Authors:** Ajaykumar Vishwakarma?, Nick Bocherding7, Kenneth Nepple10, Aliasger Salem5, 6, Russell W. Jenkins1, 2, 3 \*, Weizhou Zhang 5, 11 \*, Yousef Zakharia5, 10, 12 \*

Also need to add Andrew Belizzi

**Affiliations:**

1 Department of Medicine, Massachusetts General Hospital Cancer Center, Harvard Medical School, Boston, MA

2 Laboratory of Systems Pharmacology, Harvard Program in Therapeutic Science, Harvard Medical School, Boston, MA

3 Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA

4 Cancer Biology Graduate Program, Carver College of Medicine, University of Iowa, Iowa City IA

5 Holden Comprehensive Cancer Center, University of Iowa, Iowa City, IA

6 Department of Pharmaceutics and Translational Therapeutics, College of Pharmacy, University of Iowa, Iowa City, IA

7 Department of Pathology and Immunology, Washington University School of Medicine, St Louis, MO

10 Department of Urology, University of Iowa Hospitals and Clinics, Iowa City, IA

11 Department of Pathology, Immunology, and Laboratory Medicine, University of Florida, Gainesville, FL

12 Department of Internal Medicine, University of Iowa Hospitals and Clinics, Iowa City, IA

\* Correspondence: [rwjenkins@partners.org](mailto:rwjenkins@partners.org) (R.W.J), [zhangw@ufl.edu](mailto:zhangw@ufl.edu) (W.Z) and [yousef-zakharia@uiowa.edu](mailto:yousef-zakharia@uiowa.edu) (Y.Z)

**Abstract**:

Human clear cell renal cell carcinoma (ccRCC) is one of the most immunologically distinct tumor types due to high levels of tumor-infiltrating immune cells including T cells, yet not every patient responds to immunotherapy. Interestingly, in contrast to other cancers, infiltration with cytotoxic CD8+ T cells is associated with poorer overall survival in ccRCC, suggesting that sub-populations of CD8+ and other immune cells may underlie this observation. To characterize the tumor immune microenvironment of ccRCC, we applied single-cell-RNA sequencing (SCRS) along with T-cell-receptor (TCR) sequencing to map the transcriptomic heterogeneity of 25,688 individual CD45+ lymphoid and myeloid cells in matched tumor and blood from patients with ccRCC. Will need to update based on new findings. This report represents the first such characterization of the ccRCC immune landscape using scRNA-seq. With further characterization and functional validation, these findings may identify novel sub-populations of immune cells amenable to therapeutic intervention.

**Introduction**

Clear cell renal cell carcinoma (ccRCC) is the most common type of renal cell carcinoma arising from epithelial cells of the proximal tubule of the kidney, comprising more than 70% of all renal cancers (1). ccRCC represents an immune sensitive tumor type well known for early advances in systemic immunotherapy using T cell proliferation cytokine IL-2 and interferon (IFN) -ɑ2b therapy (2). Recent novel immunotherapies targeting T cell checkpoints as standard of care has transformed the treatment paradigm of ccRCC (3,4). However, a substantial subset of renal cancer patients still do not respond to these therapies and patients who initially do can eventually progress (5,6). Cytotoxic tumor-infiltrating lymphocytes (TILs), in particular CD8+ T cells are key effectors of the adaptive anti-tumor immune response (7) and abundance of CD8+ T cells in solid cancers is generally associated with better survival in cancer patients (8–10). However, in RCC, immune cell abundance is inversely correlated with survival, specifically TILs (11–14). Biomarker analysis results from recent clinical trials comparing PD-1 blockade versus anti-angiogenic inhibitors and combination therapies in treatment-naïve ccRCC patients also supported the negative prognostic significance of T cell infiltrate in the absence of immunotherapy (15,16). Other abundant immune players in the ccRCC tumor microenvironment include monocytes, dendritic cells and tumor-associated macrophages (TAMs) (17) which are now being harnessed for discovery of novel gene programs but remain far less studied than T cells.

Quantifying and inferring immune cell abundance from transcriptional analysis of primary or metastasized bulk tumor samples is inadequate to provide a clear picture of the immune cell types (18,19). While these studies are suggestive, they lack single cell resolution for characterizing heterogeneous cell subpopulations that ultimately shape anti-tumor response, as has been demonstrated in breast cancer and melanoma (20,21). Single-cell methodologies including flow cytometry, immunohistochemistry, and mass cytometry (13,17,22) have revealed immune cell states in ccRCC but only as discrete phenotypes when in vivo they typically display diverse spectrum of differentiation or activation states. Also, these methods require use of antibody panels targeting known immune cell components, and by design are not capable of identifying novel sub-populations of cells. SCRS has enabled comprehensive characterization of heterogeneous lymphoid and myeloid immune cells in several cancers (23–26), providing an unbiased approach to profiling cells and enabling molecular classification of different subpopulations and identification of novel gene programs. Transcriptome mapping of T lymphocytes coupled with TCR sequencing allows additional measurement of clonal T cell response to cancer at an unprecedented depth (27,28).

Here, we report the single cell RNA profiling of the immune landscape in ccRCC mapping a total 25,688 of immune single cells (5’ gene expression and recombined V(D)J region of the T cell receptor) in matched samples of tumor and peripheral blood isolated from three treatment-naïve ccRCC patients. Will need to update with results. Analysis of myeloid cells revealed a complex mixture of pro- and anti-inflammatory polarized phenotypes across patients. This represents the first such report of the immune landscape of ccRCC using SCRS for gene expression.

**Results**

*Single-cell expression profiling of immune cells in ccRCC.*

In order to define the immune microenvironment of human ccRCC, we obtained tumor and peripheral blood specimen of three treatment-naïve ccRCC patients and subject flow-sorted lymphoid and myeloid cells to SCRS. The general workflow for isolation and sequencing is available in Supplemental Figure 1. To these samples, we added immune cells from peripheral blood and normal renal parenchyma (21) to obtain an integrated UMAP projection of 22 clusters across 37,055 primary immune cells (Figure 1A). Across the three tissues, peripheral blood (n=21,160), tumor (n=12,239) and normal kidney (n=3,556), we a number of clusters with similar gene expression notably in Clusters 0, 1, 3 and 7 (Figure 1B). Despite the integration of sequencing runs to reduce tissue-type divergence, each tissue type had enrichment for clusters: peripheral blood formed the majority of Cluster 2, tumor tissues were enriched for Clusters 14, 17, 18, and 19, and normal kidney was enriched for Cluster 11. Using gene expression, we assigned cell lineages to each cluster using a three method approach: 1) examining expressions of canonical markers for T cells (*CD3E*, *CD8A*, *CD4*, and *IL7R* or CD127), B cells (*CD19* and *MS4A1*), myeloid cells (*CD14* and *FCGR3A* or CD16) and natural killer (NK) cells (*KLRD1* and *NKG7*) (Figure 1C), 2) correlations with pure-cell gene signatures derived from ENCODE (29) (Figure 1D), and 3) assigning clonotypes based on the TCR sequencing (Supplemental Figure 2). Based on these approaches we annotated clusters as monocytes (Clusters 0, 5, 11, 12, and 16), CD4+ T Cells (Clusters 4, 6, 10, 13, 15, and 20), CD8+ T cells (Clusters 1, 8, 9 and 17), NK cells (Clusters 3 and 7), B cells (Cluster 2), macrophages (Mφ, Cluster 14) and dendritic cells (DC, Clusters 18 and 19) (Figure 1E). Based on these annotations, we also examined the relative proportion of cell types comprising the sequencing runs by tissue type (Figure 3E). We observed a decrease in CD4+ T cells and B cells comparing peripheral blood to normal kidney tissue or tumor (Figure 1E). Conversely and as expected, we also found an increase in CD8+ T cells and Mφ in ccRCC versus peripheral blood (Figure 1E).

*CD8+ T cells in ccRCC tumors exhibit a transcriptional continuum with distinct populations*

The presence of CD8+ T cells in the tumor microenvironment is recognized as key effectors in the anti-tumor immune response (30) and have been associated in with response to immunotherapy (31). Sub-clustering of CD8+ T cells revealed 8 distinct clusters (Figure 2A) with relative tissue-specific distribution (Figure 2B). To understand the distribution of these new CD8+ clusters along the UMAP, we first examined the relative percent of single-cells represented in each cluster by tissue type. Tissue-infiltrating CD8+ T cells (both tumor and normal kidney) comprised the majority of CD8\_0, CD8\_1, CD8\_3, CD8\_5, CD8\_6 and CD8\_7. Only Clusters CD8\_2 and CD8\_4 had increased relative levels of peripheral blood cells (Figure 2B). Going from right to left across the x-axis of the UMAP there is a change in tissue-specific contribution starting from peripheral blood (right) to kidney (middle) to increasing levels of ccRCC tumor CD8+ T cells (left) and may represent the process of tissue infiltration itself. Within SCRS literature, these is a concern for variations in cell cycle leading to increased heterogeneity or obscure subpopulations (32), however proliferation of CD8+ T cells is an important surrogate markers of anti-tumor immune response (cite). We next examined the variation in proliferative gene signatures, finding a similar distribution to the tissue-type with increasing cells in S or G2M phases from right to left, peaking with Cluster CD8\_6 (Figure 3C).

In order to better characterize the CD8+ clusters, we next used canonical and differential T cells markers to examine gene expression differences along the UMAP (Figure 2D) with several patterns. The first pattern was the discovery of a naïve *CCR7*+ *SELL*+ *TCF7*+ being seen in CD8\_4 (Figure 2D). Looking for effector CD8+ T cells, we next observed two populations of *IFNG*+ *PRF1*+ T cells, principally in CD8\_1 and CD8\_0 (Figure 2D). The latter also expressed negative immune checkpoint, such as *CTLA4*, *HAVCR2*, *PDCD1*, and *TIGIT* (Figure 2D). These negative immune checkpoints were expressed at more moderate levels in both CD8\_5 and CD8\_6, however, CD8\_6 exclusively expressed a number of proliferation markers, such as *CDK1*, *MKI67*, *STMN1*, and *TOP2A* (Figure 2D). In order to examine gene expression patterns above single or selected genes, we used slingshot (33) to building minimum spanning trees between clusters, generating curves based on the most varied genes (Figure 2E). We identified 5 distinct curves (labeled B1 to B5) with the origin in the *CCR7*+ *SELL*+ *TCF7*+ CD8\_4. With the exception of B1 extending into CD8\_2, the remining curves graphed along a similar trajectory with a common node of CD8\_1 and branching at distinct clusters with increased levels of tumor-infiltrating CD8+ T cells (Figure 2E). Although based on gene expression, these curves also varied by CD8+ clonotypes based on TCR sequencing, with the root having no clonal expansion and the B3, B4 and B5 curves terminating into regions with higher levels of clonal expansion compared to B1 or B3 (Figure 2E). This clonotype relationship was also observed in specific clonotype sequences (nucleotide and TCR genes) with overlapping clonotypes seen in subclusters CD8\_0, CD8\_6, CD8\_5 and CD8\_3 (Figure 2F). In contrast, CD8\_7 had minimal overlaping clonotypes with other subclusters (Figure 2F). This relationship was seen also independent of the individual patient sequenced (Supplemental Figure X). In order to assess possible functional differences based on these branching, we performed gene set enrichment analysis (Figure 2G). As expected based on the immune checkpoint inhibitors expression, Clusters CD8\_0 and CD8\_5 showed increased terminal differentiation and exhaustion (Figure 2G). Cytolytic gene enrichment was seen in CD8\_1, the PRF1+ IFNG+ population lacking immune checkpoints (Figure 2G). The highly proliferative CD8\_6 population was enriched for metabolic activity, such as the tricarboxylic acid cycle and glycolysis, and DNA repair (Figure 2G). The B2 curve termination cluster, CD8\_2, has preferential enrichment of cytokine signaling, such as IL-2/STAT5, TGFβ, and type 1 interferons (Figure 2H). With immune checkpoint inhibitor responsiveness associated with distinct CD8+ T cell populations (21), we next examined enrichment of signatures associated with response or nonresponse to anti-PD-1 therapies (Figure 2H). Using the ordinal construction of the trajectories, we created a pseudotime variable for cells, allowing us to see the difference in the enrichment along the curves. This approach allowed us to see an overall enrichment in responsiveness to anti-PD-1 at the terminal points of B2 and midpoints of B3, B4 and B5, corresponding to cells in CD8\_1 (Figure 2H, blue lines). Likewise, we observed an overall increase in gene expression associated with no response or progression was observed at the terminal points of B3, B4, and B5 (Figure 2H, red lines). Concluding thought.

*Prominent infiltration of macrophages in ccRCC have transcriptional divergence.*

With the previous observation of an overall increase in macrophages and decreased monocytes in integrated UMAP (Figure 1E), we next focused on differential analyses of the myeloid populations (Figure 3A). Across monocytes, macrophages and dendritic cells, sublustering found 20 distinct clusters (Figure 3A). Tissue-specific distribution was observed, with the majority tumor-infiltrating myeloids cells in subclusters 0, 3, 7, 8 and 15 (Figure 3B). In contrast, both normal kidney parenchyma and peripheral blood were comprised of a majority of monocytic subclusters (Figure 3B). Populations were assigned using canonical markers and in addition to the previously decribed singleR approach with macrophage subclusters (0, 3, 7, 13 and 14) identified using markers such as CD1C, CLECL10A (CD301), FCER1A, and MSR1 (Figure 3C). In total, we observed a decreased the proportion in CD14+ monocytes in tissue-infiltrating myeloid cells compared to peripheral blood and an increase in macrophages (Figure 3D). Normal renal parenchyma had a variable increased in CD14- CD16- monocytes, which was not significant. As previously seen, we found a small number of DC (subclusters 15, 16, and 19) with distinct expression profiles associated with conventional DC 1, plasmacytoid DC, and monocyte-derived DCs, respectively (Supplemental Figure X).

Next we isolated the five macrophage subclusters, relabeling them tumor-associated macrophage 1 (TAM\_1), TAM\_2, TAM\_3, resident macrophage (rM) and peripheral macrophage (pM) based on the relative percent of cells derived from the respective tissue (Figure 3E). Although similar expression, these five clusters had distinct expression patterns. For example, the X+ X+ TAM\_1 expressed increased levels of chemokines and cytokines, like *CCL3*, *CCL4*, *CXCL2*, and *IL10*; X+ X+ TAM\_2 subcluster expressed the apolipoporetin gene *APOE*, lysosomal lipase (*LIPA*), and ferroportin (*SLC40A1*); and X+ X TAM\_3 had high levels of *CD1C*, *IL1R2*, and Langerhin (*CD207*), a marker of the skin-resident Langerhan cells (Figure 3F). The pM subcluster had the highest level of the cell adhesion molecules *CLEC10A*, *SELL* (CD62L), and *ITGB7*, which can dimerize with *ITGA4* (CD49d) or *ITGAE* (CD103). Like the CD8+ T cells, we built cell trajectories based on varied genes find two distinct curves converging into TAM\_3 and pM (Figure 3G). In order to assess potential functional differences for the macrophages, we performed gene set enrichment analysis (Figure 3H). As previously observed in single-cell data (cite here), no subclusters was distinctly M1 or M2 polarized. For example, TAM\_1 had enrichment for gene sets commonly associated with the M2 macrophage compartment, such has angiogenesis and the production of lipid mediators, while also having the highest levels of TNFɑ signaling enrichment, a common M1 macrophage characteristic. Across the three TAM subclusters, modest enrichment in both type 1 and type 2 interferon signaling was observed (Figure 3H). The non-TAM subclusters, rM and pM had relatively lower levels of enrichment with the exception of proinflammatory signaling and glycolysis, respectively (Figure 3H). Concluding thought

**Discusion**

**The tumor microenvironment for ccRCC tumors with** ccRCC had the highest T cell infiltration score among tumor types within the TCGA (34). Unlike other tumors that respond to immune checkpoint blockade, ccRCC has a relatively low tumor mutational load (18,35) and is not associated with response to anti-PD-1 therapy (14). Similarly, therapeutic response to anti-PD-1 therapy have been correlated with HLA heterozygosity (36), which does not seem to be the case for ccRCC (14). Immunogenicity of ccRCC has been tied to antigen presenting machinery expression through MHC-I (34). This may be a result of apparent differential immune infiltration, which has been closely linked to tumor stage and histological grade (14,34).

CD8+ T cells with lower activation and exhausted expression patterns were associated with improved anti-PD-1 responses in melanoma. Unlike the more exhausted cells, these responsive T cells had minimal clonotypes shared between the more activated CD8+ subsets (21).

Immune-regulated subtype associated with infiltration of polyclonal exhausted T cells, Tregs, and dysfunctional DC associated with advanced histological features and increased risk of disease progression (34,37). This same exhausted CD8+ T cell phenotype have been associated with increased dysfunctional DC (6) and macrophage populations (17).

**Methods**

*Subject Details and Tissue Collection*

Fresh blood and primary clear cell renal cell carcinoma (ccRCC) samples were obtained from the University of Iowa Tissue Procurement Core and GUMER repository through the Holden Comprehensive Cancer Center from subjects providing written consent approved by the University of Iowa ethics board committee. The patients were males with an age range of 67 to 74 years old. Tumor grades were histologically determined by a pathologist. Primary tumor stage for Patient 1 and Patient 2 were reported as pT1b without extension, while Patient 3 was reported as pT3a with vascular invasion into renal vein. Three ccRCC tumor specimens paired with individual blood samples were used in the study. Will need IRB FOR publication

*Tumor Dissociation and Isolation of Mononuclear Cells*

Renal tumor samples were dissociated into single cells by a semi-automated combined mechanical/enzymatic process. The tumor tissue was cut into pieces of (2-3mm) in size and transferred to C Tubes (Miltenyi Biotech, Bergisch Gladbach, Germany) containing a mix of Enzymes H, R and A (Tumor Dissociation Kit, human; Miltenyi Biotech). Mechanical dissociation was accomplished by performing three consecutive automated steps on the gentleMACS Dissociator (h\_tumor\_01, h\_tumor\_02 and h\_tumor\_03). To allow for enzymatic digestion, the C tube was rotated continuously for 30 min at 37°C, after the first and second mechanical dissociation step (38). Cells from fresh tumor specimens were incubated with FcR blocking reagent (StemCell Technologies, Vancouver, Canada) for 10 min at 40C and labelled with 1ug/ml of the FITC anti-human CD45 antibody (BioLegend, San Diego, CA) per 107 cells for 20 min at 40C. CD45+ cells were isolated using the EasySepTM FITC Positive Selection Kit (StemCell Technologies). Alternatively, mononuclear cells (MNCs) from whole peripheral blood of paired subjects were isolated using SepMate Tubes (StemCell Technologies) by density gradient centrifugation. Cells were then viably frozen in 5% DMSO in RPMI complemented with 95% FBS. Cryopreserved cells were resuscitated for flow cytometry analyses by rapid thawing and slow dilution.

*Cell Sorting for Single-Cell RNA sequencing*

Viable immune (CD45+ Hoechst-) single cell suspensions generated from three ccRCC tumor samples and blood were FACS sorted on a FACS ARIA sorter (BD Biosciences) for lymphoid and myeloid cells (Ratio 3:1). The cells were sorted into ice cold Dulbecco’s PBS + 0.04% non-acetylated BSA (New England BioLabs, Ipswitch, MA). Sorted cells were then counted and assessed viability MoxiGoII counter (Orflo Technologies, Ketchum, ID) ensuring that cells were re-suspended at 1000 cells/ul with a viability >90%.

Library Preparation, Single-Cell 5’ and TCR Sequencing

Single-cell library preparation was carried out as per the 10X Genomics Chromium Single Cell 5' Library and Gel Bead Kit v2 #1000014 (10x Genomics, Pleasanton, CA). Cell suspensions were loaded onto a Chromium Single-Cell Chip along with the reverse transcription (RT) master mix and single cell 5′ gel beads, aiming for 7,500 cells per channel. Following generation of single-cell gel bead-in-emulsions (GEMs), reverse transcription was performed using a C1000 Touch Thermal Cycler (Bio-Rad Laboratories, Hercules, CA); 13 cycles were used for cDNA amplification. Amplified cDNA was purified using SPRIselect beads (Beckman Coulter, Lane Cove, NSW, Australia) as per the manufacturer’s recommended parameters. Post-cDNA amplification reaction QC and quantification was performed on the Agilent 2100 Bioanalyzer using the DNA High Sensitivity chip. For input into the gene expression library construction, 50ng cDNA and 14 cycles was used. To obtain TCR repertoire profile, VDJ enrichment was carried out as per the Chromium Single Cell V(D)J Enrichment Kit, Human T Cell #1000005 (10x Genomics) using the same input sample. Sequencing libraries were generated with unique sample indices (SI) for each sample and quantified. Libraries were sequenced on an Illumina HiSeq 4000 using a 150-pair-end sequencing kit. Gene expression FASTQ files were aligned to the human genome (GRCh38) using the CellRanger v2.2 pipeline, while clonotype sequencing was aligned to the vdj\_GRCh38\_alts\_ensembl genome build provided by the manufacturer.

*Incorporation of other SCRS data sets*

SCRS and TCR sequencing data processed using Cell Ranger v2.2 for healthy donor peripheral blood immune cells were acquired from the 10x Genomics website on 6/20/2020. Filtered gene matrix and contig annotations were used in the incorporation of the uniform manifold approximation and project (UMAP). Total number of cells from healthy peripheral blood control were 7,726. SCRS of normal immune populations in the kidney were derived previously published data (39). Gene expression matrices were downloaded from the EGAS00001002325 and filtered for normal renal parenchyma cells using the provided cell manifest for the samples RCC1, RCC2, and RCC3. These processed using the procedure as described above to form a UMAP. Immune cells were identified using canonical markers for lineage and were then isolated. Isolated immune cells for normal renal parenchyma were: RCC1 (n=1,011), RCC2 (n=888), and RCC3 (n=1,757).

*SCRS Integration*

Initial processing of cells isolated from ccRCC patients; Patient 1 (n=10,694), Patient 2 (n=5,174) and Patient 3 (n=9,805) were processed and integrated with the above samples using the Seurat R package (v3.0.2) (40,41). Samples were normalized using the *SCTtransform* approach (42) with default settings. Preparation for integration used 3,000 anchor features and *PrepSCTIntegration*. The integration of sequencing runs occurred with the SCT-transformed data. The dimensional reduction to form the UMAP utilized the top 30 calculated dimensions and a resolution of 0.7. Data characteristics by sequencing run can be found in Supplemental Table 1. Cell type subclustering used the SCTtransform approach as described above, but integrating the data across samples instead of individual sequencing runs. The adjusted dimensional inputs for the subclustering analysis can be found in Supplemental Table 2.

*SCRS Data Analysis and Visualizations*

The schex R package (v1.1.5) was used to visualize mRNA expression of lineage-specific or highly differential markers by converting the UMAP manifold into hexbin quantifications of the proportion of single-cells with the indicated gene expressed. Default bins across all cells was 80 and 40 for subcluster analyses, unless otherwise indicated in the figure legend. Differential gene expression utilized the Wilcoxon rank sum test on count-level mRNA data. For differential gene expression across clusters or subclusters, *FindAllMarkers* function in the Seurat package using the log-fold change threshold > 0.25, minimum group percentage = 10%, and the pseudocount = 0.1. Differential comparisons between condition utilized the *FindMarkers* function in Seurat, without filtering and a pseudocount = 0.1. Multiple hypothesis correction was reported using the Bonferroni method. Cell cycle regression was performed in Seurat using the *CellCycleScoring* function and genes derived from Nestorowa et alia (43). Genes were isolated by calling *cc.genes.updated.2019* in R.

Cell type identification utilized the SingleR (v1.0.1) R package (44) with correlations of the single-cell expression values with transcriptional profiles from pure cell populations in the ENCODE (29). In addition to correlations, canonical markers for cell lineages (Supplemental Table 3) and corresponding TCR sequences were used. Gene set enrichment analysis was performed using the escape R package (v0.99.0). Gene sets were derived from the Hallmark library of the Molecular Signature Database and from previous publications (21,24). Enrichment for anti-PD-1 therapy response was derived from Sade-Feldmen et alia to develop gene signatures for the CD8\_B (nonresponsive) and CD8\_G (responsive) single-cell populations(21). Differential enrichment analysis was performed using the *getSignificance* function in escape that is based on the limma R package linear fit model. TCR analysis utilized our previously described scRepertoire R package (v0.99.3) (45) with clonotype being defined as the combination of the gene components of the VDJ and the nucleotide sequence for both chains and assigned on the integrated Seurat object. Cell trajectory analysis used the slingshot (v1.6.0) R package (33) with default settings for the *slingshot* function and using the embedding from the subclustering for each cell type. Inferred start and end clusters were applied in the CD8+ T cell trajectory based on gene expression markers. Ranked importance of genes were calculated using the top 300 variable genes and rsample (v0.0.9) and tidymodels (v0.1.0) R packages were used to generate random forest model based on a training data set of 75% of the cells. The *rand\_forest* function in the parsnip (v0.1.1) R package was used, with mtry set to 200, trees to 1400, and minimum number of data points in a node equal to 15 across all cell types. The code for all analysis is available at <https://github.com/ncborcherding/ccRCC>.

Mass Cytometry Analysis

Flow cytometry standard files were downloaded for 78 samples utilizing a previously-defined tumor-associated macrophage panel (17). Subsequent loading and analyses of the data was based on the accompanying published methods (17). These files were loaded into R using the flowCore (v2.0.1) R package. Protein signal was arcsinh transformed using a cofactor of 5, filtered for previously identified myeloid cells, and up to 2,000 cells were randomly selected for each sample using the *sample* function with *set.seed* of 123. After the transformation and filtering, 127,880 myeloid cells were subjected to t-SNE based embedding using the Rtsne (v0.1.5) package using only the defined cell markers. Clusters were defined using the R implementation of the PhenoGraph algorithm (46) in the Rphenograph R package (v0.99.1) with the k set at 100. Marker expression visualization utilized a normalized intensity between 0 and 1. Further data visualization utilized ggplot2 (v3.3.1) and pheatmap (v1.0.12).

*Machine Learning Modeling*

The renal clear cell carcinoma (KIRC) log2 gene expression data from the downloaded from the University of California Santa Cruz Xena Browser and filtered for only primary tumor samples. To the expression data, updated clinical information was assigned by tumor barcode (47). Gene signatures from subcluster analysis were generated by filtering differential gene expression results for genes with > 0.5 log-fold change and 15% difference in cell expression. Training and testing sample cohorts were divided using the *sample* function with *set.seed* set to 123, splitting the data into 1:1 ratio. The caret (v6.0-86) R package was used to generate random forest models, using the repeated cross-validation method splitting the data into 10 groups and repeating a total of 16 times and recursive feature selection to optimize feature selection for 1, 5, 10, 15, and 20 features. Generated models were then used to predict survival in the testing cohort and testing parameters were then calculated. Survival analyses utilized the survival (3.1-12) and survMiner (v) R packages. Edit if changes.

*Statistical Analysis*

Statistical Analyses were performed in R (v4.0.1). Two-sample significance testing utilized Welch’s T test, with significance testing for more than three samples utilizing one-way analysis of variance (ANOVA) with Tukey honest significance determination for correcting multiple comparisons. Two-proportion Z-tests was performed using the total number of cells in each condition as the number of trials and without a prior for proportion.

**Acknowledgments**

We thank Michael Knudson, Rita Sigmund, Joe Galbraith, Janice Cook-Granroth, Bethany Kilburg and Celeste Charchalac from University of Iowa Carver College of Medicine, Tissue Procurement Core (TPC) and Genito-Urologic Tissue Repository (GUMER) for receiving biological samples and clinical data. We thank Justin Fishbaugh, Heath Vignes and Michael Shey from the University of Iowa Flow Cytometry Facility. We thank Kevin Knudtson, Mary Boes, Garry Hauser and Mari Eyestone from the Iowa Institute of Human Genetics (IIHG) Genomics Division for planning and assisting use of Next Gen Sequencing (NGS) platforms, Diana Kolb from the IIHG Bioinformatics Division and the University of Iowa High Performance Computing (HPC) facility.

**Funding**

Funding for this project was provided by Rock ‘N’ Ride Foundation (PI: Y.Z.) and from the National Institute of Health under the R01 CA200673 (PI: W.Z.), R01 CA203834 (PI: W.Z.), K08 CA226391 (PI: R.W.J) and F30 CA206255 (PI: N.B.). The flow cytometry and sequencing **facilities are funded in part, by the National Cancer Institute of the National Institutes of Health under Award Number P30CA086862. The** FACSAria Fusion high-speed cell sorter was supported with funds from **the National Center for Research Resources of the National Institutes of Health under Award Number 1 S10 OD016199-01A1.** The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Author contributions**

**Conception and design:** AV, YZ, WZ

**Development of methodology:** AV NB WZ

**Acquisition of data:** KN, YZ, AV

**Analysis and interpretation of data:** NB AV AS RWJ WZ YZ

**Writing, review, and/or revision of the manuscript:** AV NB AS RWJ WZ YZ

Supervision: YZ, WZ, RWJ

**Declaration of interests**

Dr. Russell W. Jenkins has a financial interest in XSphera Biosciences Inc., a company focused on using ex vivo profiling technology to deliver functional, precision immune-oncology solutions for patients, providers, and drug development companies. Dr. Jenkins’ interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies.

**Data and materials availability:**

Quantified gene expression counts and V(D)J T cell receptor sequences for single-cell RNA sequencing are available at the Gene Expression Omnibus (GEO) at [GSE121638](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121638). Code for the analysis and visualizations are available at <https://github.com/ncborcherding/ccRCC>

**References**

1. Saad AM, Gad MM, Al-Husseini MJ, Ruhban IA, Sonbol MB, Ho TH. Trends in Renal-Cell Carcinoma Incidence and Mortality in the United States in the Last 2 Decades: A SEER-Based Study. Clin Genitourin Cancer. 2019;17:46–75.

2. Koneru R, Hotte SJ. Role of cytokine therapy for renal cell carcinoma in the era of targeted agents. Curr Oncol. Multimed Inc.; 2009;16:S40.

3. Motzer RJ, Penkov K, Haanen J, Rini B, Albiges L, Campbell MT, et al. Avelumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. N Engl J Med. 2019;380:1103–15.

4. Dudani S, Graham J, Wells C, Pal SK, Dizman N, Donskov F, et al. First-line (1L) immuno-oncology (IO) combination therapies in metastatic renal cell carcinoma (mRCC): Preliminary results from the International Metastatic Renal Cell Carcinoma Database Consortium (IMDC). J Clin Oncol. American Society of Clinical Oncology; 2019;37:584–584.

5. Sharma P, Allison JP. Immune checkpoint targeting in cancer therapy: Toward combination strategies with curative potential. Cell. 2015.

6. Giraldo NA, Becht E, Pagès F, Skliris G, Verkarre V, Vano Y, et al. Orchestration and prognostic significance of immune checkpoints in the microenvironment of primary and metastatic renal cell cancer. Clin Cancer Res. 2015;

7. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature. 2014;

8. Galon J, Fox BA, Bifulco CB, Masucci G, Rau T, Botti G, et al. Immunoscore and Immunoprofiling in cancer: An update from the melanoma and immunotherapy bridge 2015. J Transl Med. 2016;

9. Ziai J, Gilbert HN, Foreman O, Eastham-Anderson J, Chu F, Huseni M, et al. CD8+ T cell infiltration in breast and colon cancer: A histologic and statistical analysis. PLoS One. 2018.

10. Shimizu S, Hiratsuka H, Koike K, Tsuchihashi K, Sonoda T, Ogi K, et al. Tumor-infiltrating CD8+ T-cell density is an independent prognostic marker for oral squamous cell carcinoma. Cancer Med. 2019;

11. Patel HD, Puligandla M, Shuch BM, Leibovich BC, Kapoor A, Master VA, et al. The future of perioperative therapy in advanced renal cell carcinoma: How can we PROSPER? Futur Oncol. 2019;

12. Nakano O, Naito Y, Nagura H, Ohtani H, Nakano O, Sato M, et al. Proliferative activity of intratumoral CD8+ T-lymphocytes as a prognostic factor in human renal cell carcinoma: Clinicopathologic demonstration of antitumor immunity. Cancer Res. 2001;

13. Baine MK, Turcu G, Zito CR, Adeniran AJ, Camp RL, Chen L, et al. Characterization of tumor infiltrating lymphocytes in paired primary and metastatic renal cell carcinoma specimens. Oncotarget. 2015;

14. Braun DA, Hou Y, Bakouny Z, Ficial M, Sant’ Angelo M, Forman J, et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. Nat Med. 2020;

15. Choueiri TK, Albiges L, Haanen JBAG, Larkin JMG, Uemura M, Pal SK, et al. Biomarker analyses from JAVELIN Renal 101: Avelumab + axitinib (A+Ax) versus sunitinib (S) in advanced renal cell carcinoma (aRCC). J Clin Oncol. 2019;

16. Choueiri TK, Larkin J, Oya M, Thistlethwaite F, Martignoni M, Nathan P, et al. Preliminary results for avelumab plus axitinib as first-line therapy in patients with advanced clear-cell renal-cell carcinoma (JAVELIN Renal 100): an open-label, dose-finding and dose-expansion, phase 1b trial. Lancet Oncol. 2018;

17. Chevrier S, Levine JH, Zanotelli VRT, Silina K, Schulz D, Bacac M, et al. An Immune Atlas of Clear Cell Renal Cell Carcinoma. Cell. 2017;

18. Creighton CJ, Morgan M, Gunaratne PH, Wheeler DA, Gibbs RA, Robertson G, et al. Comprehensivemolecular characterization of clear cell renal cell carcinoma. Nature. 2013;

19. Van Den Heuvel CNAM, Van Ewijk A, Zeelen C, De Bitter T, Huynen M, Mulders P, et al. Molecular profiling of druggable targets in clear cell renal cell carcinoma through targeted RNA sequencing. Front Oncol. 2019;

20. Savas P, Virassamy B, Ye C, Salim A, Mintoff CP, Caramia F, et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. Nat Med. 2018;

21. Sade-Feldman M, Yizhak K, Bjorgaard SL, Ray JP, de Boer CG, Jenkins RW, et al. Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. Cell. 2018;

22. Geissler K, Fornara P, Lautenschläger C, Holzhausen HJ, Seliger B, Riemann D. Immune signature of tumor infiltrating immune cells in renal cancer. Oncoimmunology. 2015;

23. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun. 2017;

24. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell. 2018;

25. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med. 2018;

26. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science (80- ). 2016;

27. Beausang JF, Wheeler AJ, Chan NH, Hanft VR, Dirbas FM, Jeffrey SS, et al. T cell receptor sequencing of early-stage breast cancer tumors identifies altered clonal structure of the T cell repertoire. Proc Natl Acad Sci U S A. 2017;

28. Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. Cell. 2017;

29. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;

30. Apetoh L, Smyth MJ, Drake CG, Abastado JP, Apte RN, Ayyoub M, et al. Consensus nomenclature for CD8+ T cell phenotypes in cancer. Oncoimmunology. 2015;

31. Borcherding N, Kolb R, Gullicksrud J, Vikas P, Zhu Y, Zhang W. Keeping Tumors in Check: A Mechanistic Review of Clinical Response and Resistance to Immune Checkpoint Blockade in Cancer. J Mol Biol. 2018;

32. Luecken MD, Theis FJ. Current best practices in single‐cell RNA‐seq analysis: a tutorial. Mol Syst Biol. 2019;

33. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics. 2018;

34. Şenbabaoğlu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. Genome Biol. 2016;

35. Miao D, Margolis CA, Gao W, Voss MH, Li W, Martini DJ, et al. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. Science (80- ). 2018;

36. Chowell D, Morris LGT, Grigg CM, Weber JK, Samstein RM, Makarov V, et al. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. Science (80- ). 2018;

37. Giraldo NA, Becht E, Vano Y, Petitprez F, Lacroix L, Validire P, et al. Tumor-infiltrating and peripheral blood T-cell immunophenotypes predict early relapse in localized clear cell renal cell carcinoma. Clin Cancer Res. 2017;

38. Baldan V, Griffiths R, Hawkins RE, Gilham DE. Efficient and reproducible generation of tumour-infiltrating lymphocytes for renal cell carcinoma. Br J Cancer. 2015;

39. 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 (80- ). 2018;

40. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;

41. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;

42. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;

43. Nestorowa S, Hamey FK, Pijuan Sala B, Diamanti E, Shepherd M, Laurenti E, et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood. 2016;

44. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;

45. Borcherding N, Bormann NL. scRepertoire: An R-based toolkit for single-cell immune receptor analysis. F1000Research. 2020;

46. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir EAD, Tadmor MD, et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell. 2015;

47. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell. 2018;

**Figures**

A picture containing sitting, screen, colorful, small

Description automatically generated

**Figure 1: Single-cell sequencing results for immune cells in ccRCC.** A. UMAP of 37,055 primary immune cells of peripheral blood, normal renal parenchyma and tumor-infiltrating ccRCC patients. **B**. Distribution of cells by tissue type, peripheral blood (blue), tumor (red), and kidney (light blue). Arrows indicated potential enriched or unique immune cells populations for tissue type. **C**. Percent of cells expressing canonical immune cell markers across the UMAP. **D**. Normalized correlation values for predicted immune cell phenotypes based on the SingleR R package for each cluster. **E**. UMAP demonstrating inferred immune cell types in ccRCC, clusters are colored by cell type and proportion of single-cell per sequencing run by tissue type. P values based on one-way ANOVA; lack of p-values equates to value > 0.05.

**A picture containing room

Description automatically generatedFigure 2: CD8+ T cells in ccRCC tumors exhibit a transcriptional continuum with distinct populations. A.** UMAP subclustering of CD8+ T cells (original clusters 1, 8, 9, and 17). **B**. UMAP distribution of single cells by tissue type with relative percent of cells by tissue in each cluster. **C**. Cell cycle regression assignments for CD8+ T cells by cluster assignment. **D**. Percent of cells expressing selected markers for T cell biology. **E**. CD8+ UMAP of clusters (upper panel) and clonotype frequency (lower panel) overlaid with slingshot-based (33) cell trajectory starting at CD8\_4 and proceeding into 5 distinct curves: branch 1 (B1), B2, B3, B4, and B5. **F**. Clonotype overlap coefficients between subclusters. **G**. Z-transformed normalized enrichment scores from ssGSEA for selected gene sets by subcluster. **H**. Normalized enrichment scores for therapeutic response or lack of response to anti-PD-1 therapy across the CD8+ T cells (upper panel) and by pseudotime of each branch (lower panel).

A picture containing indoor, sitting, colorful, small

Description automatically generated

**Figure 3: Single-cell CD4+ T cell characterization in ccRCC finds disparate intratumoral with common endpoints.** **A**. UMAP subclustering of CD4+ T cells (original clusters 4, 6, 10, 13, 15, and 20). **B**. UMAP distribution of single cells by tissue type with relative percent of cells by tissue in each cluster. **C**. Percent of cells expressing selected markers for myeloid and macrophage markers. **D**. Proportion of assigned cell types compared to total antigen presenting cells by tissue type. **E**. Macrophage subclusters: TAM\_1 (n=), TAM\_2 (n=), TAM\_3 (n=), pM (n=), and rM (n=) with relative percent of cells by tissue in each cluster. F. Differential expression markers for macrophage subclusters. **G**. Macrophage UMAP overlaid with slingshot-based (33) cell trajectories starting at rM and TAM\_2 and proceeding into pM. Smaller UMAP shows pseudotime created by the cell trajectories. **H**. Z-transformed normalized enrichment scores from ssGSEA for selected gene sets by subcluster.

CD4+ UMAP of subclusters (upper panel) and clonotype frequency (lower panel) overlaid with slingshot-based (33) cell trajectory starting from CD4\_1 (root 1) and CD4\_1 (root 2) with relative pseudotime for all curves calculated using slingshot. **E**. Percentage difference (∆ percent of cells) and log-fold change based on the Wilcoxon rank sum test results for differential gene expression comparing TI to PB CD4+ T cells in ccRCC patients (left panel), colored points indicate adjusted p-values < 0.05. Right panel includes top 10 markers for TI-predominant CD4+ subclusters. Size of points are relative percent of cells in cluster expressing the indicated mRNA species. **F**. Z-transformed normalized enrichment scores from ssGSEA for selected gene sets by subcluster.