# Exploring potential roles of ccRCC cell populations in the TCGA cohort

Guangyuan Li, BS[1,2], Behrouz Shamsaei, PhD[3], Jarek Meller, PhD[1,2], Maria F. Czyzyk-Krzeska, MD,PhD[4]

[1]Department of Biomedical Informatics, University of Cincinnati College of Medicine, Cincinnati, OH, USA
[2]Division of Biomedical Informatics, University of Cincinnati College of Medicine, Cincinnati, OH, USA
[3]Division of Biostatistics and Bioinformatics, Department of Environental and public health Sciences, University of Cincinnati College of Medicine, Cincinnati, OH, USA
[4]Department of Cancer Biology, University of Cincinnati College of Medicine, Cincinnati, OH, USA

## Background

**ccRCC**

- clear-cell Renal Cell Carcinoma (ccRCC) is the eighth leading death of cancer-related disease in U.S.
- Almost all adjuvant agents for ccRCC patients have shown no benefits in clinical trials so far.
- Extensive heterogeneity within the ccRCC tumor micro-environment is still poorly understood.

**single-cell RNA Seq (scRNA)**

- scRNA Seq directly measures the gene expression within each single cell in a high-throughput manner
- It offers unprecedented high-resolution compared to traditional bulk RNA Seq
- Combining the cell populations information back to rich clinical data in TCGA cohort can potentially reveal the roles of each cell type in tumor progress.

## Methods

**scRNA dataset**

- Young et al. (Science 2018) scRNA datasets
- 3 Renal Cell Carcinoma (RCC) patients (RCC1, RCC2, RCC3), two patients (RCC1, RCC2) sequencing data retained via a rational QC control.
- 7,949 single cell in RCC1
- 17,187 single cell in RCC2

**Bulk RNA dataset**

- TCGA Kidney Clear Cell Carcinoma (KIRC) cohort
- Log(TPM+1) transformed expression matrix from UCSC Xena Web portal
- 20,530 gene mapped to transcriptome reference, 606 tumor samples sequenced
- DeconRNASeq R package to perform bulk deconvolution

**TCGA cohort clinical metadata**

- 606 corresponding clinical metadata for each tumor sample
- Stage information (I, II, III, IV)
- Overall survival (OV) time and status (deceased or alive)

## Results

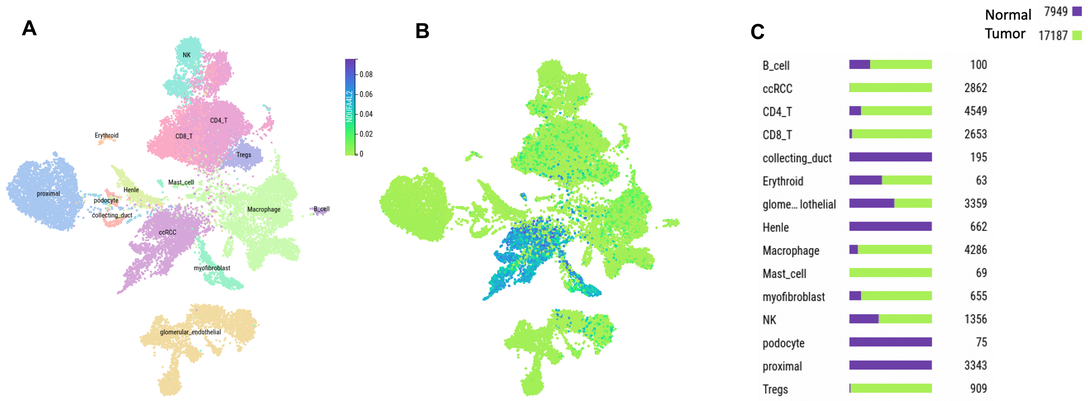


**Figure1**: Cell type characterization in batch-effect removed scRNA dataset (RCC1 and RCC2). (A) Manual cell type annotation using canonical marker gene expression curated from literatures. (B) An example marker gene NDUFA4L2 to depict the boundary of ccRCC tumor cells. (C) Abundances changes of each cell populations between normal and tumor tissues

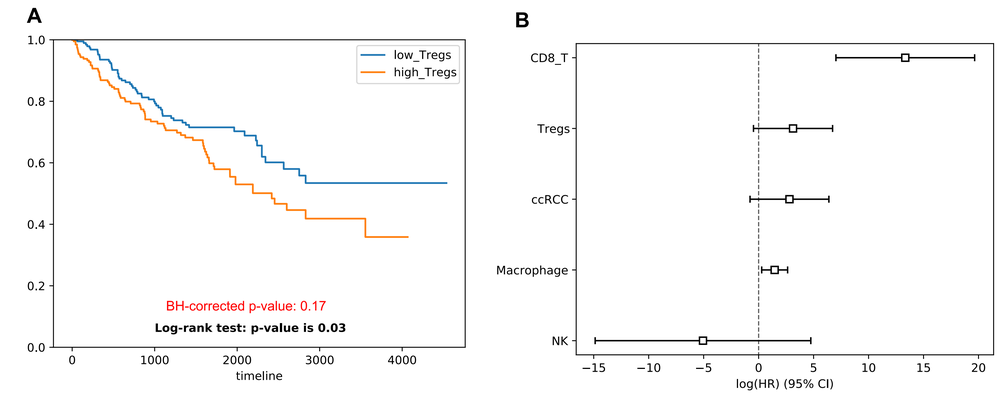


**Figure2**: The effects of different cell types on patients' overall survival. (A) Kaplan-Meier survival curve for cohorts with low abundance of Tregs versus high abundance of Tregs. (B) Cox proportional hazard ratio regression analysis on five immune cell types (CD8+ T, Tregs, ccRCC, Macrophages, NK cells)

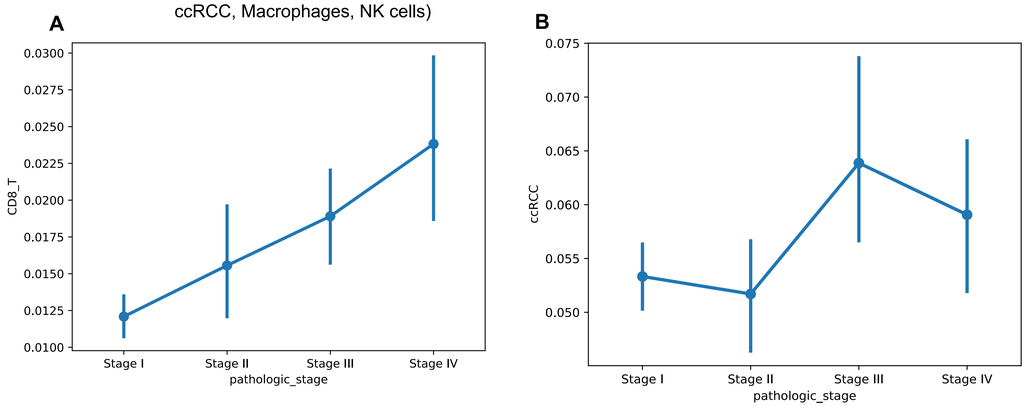


**Figure3**: The dynamic changes of cell type abundance as the tumor progress (stage I - stage IV). (A) CD8+ T cell (B) ccRCC tumor cells.

## Discussion

**Technical challenges:**

- The quality of Single cell RNA Sequencing data is not ideal, it suffers from relatively high percentage of mitochondrial genes, which indicates the broken cell membrane during droplet formation, along with high doublet rate.
- DeconRNASeq is sensitive to the selection of marker genes of each queried cell type, and data normalization techniques as well.
- The abundance of ccRCC cell type is not entirely consistent with the expected tumor tissue proportion in TCGA.
- Only 2 patients were considered for the abundance analysis, more patients scRNA sequencing data need to be leveraged for confirming the findings here.
- Multiple hypothesis testing correction might be needed for inferring statistically significantly survival differences in log-rank test.

## Conclusions

- We presented a generalizable approach to link single cell RNA-seq data with existing TCGA clinical information. It can be readily applicable to every cancer types in TCGA portal.
- More robust computational methods for deconvoluting bulk RNA-seq data are needed to accurately infer the abundance of cell types.
- Code Availability: ( https://github.com/frankligy/kidney_project/tree/main/frank_course )

## References

1. Young, M. D. et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science 361, 594–599 (2018).

2. Hu, J. et al. Single-Cell Transcriptome Analysis Reveals Intratumoral Heterogeneity in ccRCC, which Results in Different Clinical Outcomes. Mol. Ther. 28, 1658–1672 (2020).

3. Hakimi, A. A., Pham, C. G. & Hsieh, J. J. A clear picture of renal cell carcinoma. Nature genetics vol. 45 849–850 (2013).

4. Gong, T. & Szustakowski, J. D. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. Bioinformatics 29, 1083–1085 (2013).