

Investigating gene expression profiles of therapeutic targets in renal cell carcinoma based on single-cell RNA sequencing data

37th OTDK Biology Section

Written by:

Zsófia Dobolyi

Eötvös Loránd University

Biology MSc, 1st year

Supervisors:

Dr. Otília Menyhart, Ph.D.

Assistant Professor

Oncology Biomarker Research Group, Institute of Enzymology, HUN-REN Research Centre for Natural Sciences

Prof. Dr. Balázs Gyórfy, MD, Ph.D., D.Sc

Professor

Oncology Biomarker Research Group, Institute of Enzymology, HUN-REN Research Centre for Natural Sciences

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF NATURAL SCIENCES
INSTITUTE OF BIOLOGY



Budapest, 2024

Table of contents

Table of contents	2
Abstract	3
List of abbreviations	4
Introduction	5
Objectives	7
Methods	9
Collecting data about anti-RCC drugs and their target genes	9
Collecting clinical trial data of targeted therapy drugs against RCC	10
Single-cell RNA sequencing datasets	11
Pre-processing data	13
Dimensionality reduction and clustering	13
Batch-effect correction and data integration	14
Cell type annotation	14
Results	15
Designed bioinformatic workflow in Python to analyze scRNA-seq data	15
Differentially expressed genes in annotated cancer cells	16
Genes expressed in most cancer cells	17
Expression of used anti-RCC drug target genes in cancer cells	18
Discussion	19
Potential biomarkers identified for RCC based on scRNA-seq data	19
Currently available drugs targeting ANGPTL4, VCAM1, and EDN1 are suggested to be used against RCC	19
Correlation between drug efficacy and target expression at a single-cell level	21
Acknowledgements	22
Author contribution	22
Code and data availability	22
References	22

Abstract

Renal cell carcinoma (RCC) is the most common type of kidney cancer. Treatment is difficult due to lack of drugs with proper efficacy. We hypothesize that inefficacy is connected to moderate expression of targeted genes in cancer cells. I investigated the expression profiles of 174,611 cells by integrating single-cell RNA sequencing data from RCC patients, downloaded from NCBI GEO. I designed a complete bioinformatic workflow in Python to cluster and annotate cancer cells in the tumor microenvironment. 32,977 cancer cells were annotated based on marker genes. I identified new biomarkers of RCC, such as CD24 (FC=4.42; p-value=0.0013); CRYAB (FC=4.31; p-value=0.00093) or TMEM176A (FC=3.98; p-value=0.0017). I also determined genes (ANGPTL4; VCAM1; EDN1) that were overexpressed in a high percent of annotated tumor cells (ATCs), but not in other cell types. Drug development on these targets is suggested to establish their effectiveness in treating RCC. I collected the drugs against RCC that are approved or in clinical trials from the DrugBank database and identified 51 of their target genes. Genes targeted by most drugs were only expressed in a small ratio of ATCs: IL2R – 4.5%; mTOR – 2.5% and PD-L1 – 0.8%. While VEGFA, the target of the currently most effective drugs was expressed in 85.1% of the ATCs and only 24.7% of other annotated cell types on average. Bevacizumab, a VEGF inhibitor has the highest progression-free survival time according to clinical trials (10.2 months), thus, single-cell gene expression profiles might correlate with targeted therapy effectiveness.

List of abbreviations

Abbreviation	Definition
RCC	renal cell carcinoma
NCBI	National Center for Biotechnology Information
GEO	Gene Expression Omnibus
ATC	annotated tumor cell
FC	fold change
mTOR	mammalian target of rapamycin
VHL	von Hippel-Lindau
HIF1A	hypoxia-inducible factor 1 subunit alpha
VEGFA	vascular endothelial growth factor A
mTORC1	mammalian target of rapamycin complex 1
mTORC2	mammalian target of rapamycin complex 2
NKC	natural killer cell
IFN- α	interferon alpha
IL-2	interleukin 2
PD-1	programmed cell death protein 1
PD-L1	ligand of programmed cell death protein 1
UMI	unique molecular identifier
scRNA-seq	single-cell RNA sequencing
PFS	progression-free survival
QC	quality control
MT	mitochondrial gene
UMAP	Uniform Manifold Approximation Projection
DGEA	differential gene expression analysis
DEGs	differentially expressed genes
AE	average expression value
F	fraction value

Introduction

Renal cell carcinoma (RCC) is the most common and lethal tumor of the kidney, approximately 175,000 deaths are caused by RCC in a year on average. It is responsible for around 5% of cancer diagnoses worldwide, predominantly affecting men [1]. RCC is an adenocarcinoma originating in the renal tubular epithelium. Treatment options include surgery, radiation therapy, and systematic treatments, but many patients face recurrence and resistance to these therapies. Therefore, novel targeted therapies and immunotherapies have emerged, and are still highly sought after [2]. Currently, the most frequent treatments for RCC include anti-angiogenic drugs, inhibitors of the mammalian target of rapamycin (mTOR) gene, and immunotherapeutic drugs combined with systemic chemotherapy. Anti-angiogenic agents are particularly effective against hereditary RCC with a loss-of-function mutation in the von Hippel-Lindau (VHL) gene, as this mutation activates the hypoxia-inducible factor 1 subunit alpha (HIF-1 α), leading to increased pro-angiogenic factors, such as the vascular endothelial growth factor (VEGF) [3]. Anti-VEGF drugs include Bevacizumab, Sorafenib, Sunitib, Pazopanib, Axitinib, Lenvatinib, and more recently, Cabozantinib (Figure 1). These drugs are multi-target receptor tyrosine kinase (RTK) inhibitors with wide-ranging cellular actions [4]. Inhibitors of mTOR include Everolimus and Temserolimus (Figure 1), which are rapamycin analogs. Rapamycin is a naturally derived immune suppressor. Second-generation mTOR inhibitors, such as Sapanisertib and Vistusertib, are undergoing clinical testing, with these drugs targeting mTORC1/2, proteins that control mTOR activity [5]. Immunotherapy agents enhance immunity by activating natural killer (NK) cells, boosting macrophage activity, and promoting dendritic cell differentiation. Before the appearance of new generation drugs, patients with RCC were treated with interferon alpha (IFN- α) and Aldesleukin, which is recombinant interleukin 2 (IL-2). Advanced immune checkpoint inhibitors that target programmed cell death protein 1 (PD-1) or its ligand PD-L1, reduce T cell inflammatory activity [6].

However, in general, the treatment of RCC remains extremely difficult, as even the most effective drugs, Pazopanib and Sunitinib provide a median overall survival of 28.4 and 29.3 months [7]. Different combinations of targeted therapies and immunotherapies resulted in an increased overall survival, but these are only extended with a few months. The inefficacy of therapies against RCC is also caused by the fact that most symptoms are only perceived after RCC has metastasized.

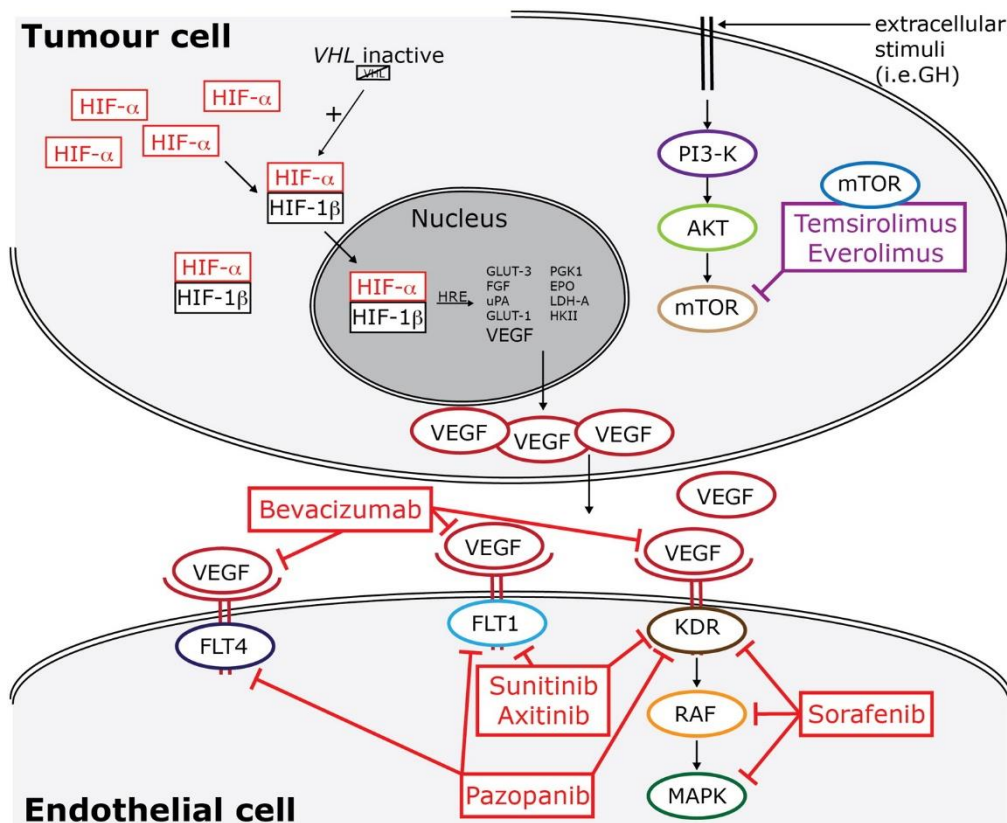


Figure 1. Most important target genes of targeted therapy drugs against RCC. This figure was published by Janowitz, in 2013 [8].

Single-cell RNA sequencing (scRNA-seq) is a pioneer technology that empowers scientists to determine the transcriptome of individual cells. Unlike traditional bulk RNA sequencing, which provides an average gene expression for all cell types within a sample, scRNA-seq offers the study of gene expression of thousands of cells in a single measurement. This approach fundamentally altered various research fields, including drug discovery and development, with its most significant impact exerted on target identification. In cancer research, scRNA-seq proves to be a potent tool, as it enables examination of significant variations of different cell types in the tumor and its microenvironment. For instance, scRNA-seq has been useful in identifying cells from which RCC originates [9]. Additionally, scRNA-seq has elucidated the role of tumor epithelia in promoting immune cell infiltration, potentially explaining why some RCCs respond favorably to immunotherapy [10].

A typical scRNA-seq workflow involves three main steps: library generation, pre-processing, and downstream analysis. The first step of library generation is the isolation of single cells from a tissue sample following tissue dissociation. These cells are then separated

into micro-droplet reaction chambers, typically using 10X Chromium technology. Chambers contain lysis buffer and reverse transcription enzymes within tiny oil droplets. After being separated from an aqueous flow of cells in a microfluidic device, single cells are added, to which barcoded primers carried in beads are attached [11]. RNA transcripts are then tagged with unique molecular identifiers (UMI) to separate cell transcripts and extraneous PCR amplicons (Figure 2). The resulting cDNA library is amplified for further analysis.

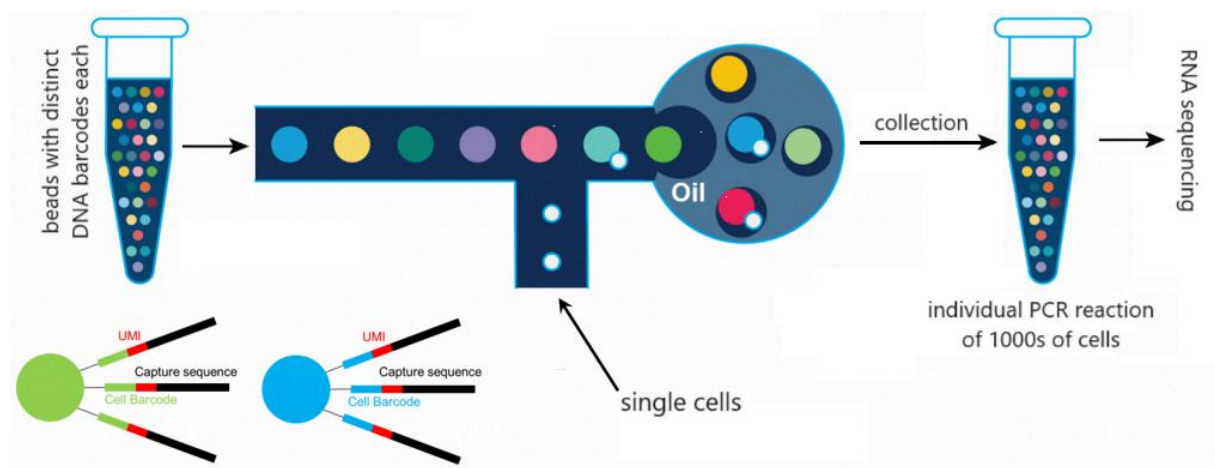


Figure 2. Schematic overview of a standard 10X Chromium workflow. This figure is a modified version of the one published on the website of the Lausanne Genomic Technologies Facility (<https://wp.unil.ch/gtf/single-cell-experiments-10x/>).

Pre-processing is the handling of highly multiplexed data to correctly assign UMI counts to cell barcodes and determine gene counts for each cell. This results in a cell-by-gene matrix, which is subsequently filtered and normalized before downstream analysis. Downstream analysis is a part of post-processing that allows scientists to answer relevant biological questions. It includes unsupervised clustering to group cells with similar expression profiles together and dimensionality reduction that allows 2D visualization of this clustering. Differential gene expression analysis (DGEA) is used to detect marker genes associated with clusters. Batch effects are corrected with integrative analysis and cell type annotation is performed based on marker gene databases [12].

Objectives

The central hypothesis of the study is that the analysis of combined, large databases of scRNA-seq data of surgically removed RCC samples provides unforeseen information on RCC treatment. Therefore, the main objective of the study was to collect and analyze scRNA-seq

data associated with target molecules of drugs against RCC. To investigate the hypothesis, the following specific objectives will be addressed:

1. Designing complete scRNA-seq data analysis workflow.

An integral scRNA-seq data processing workflow in Python will be created, as no such has been published. This will be used to integrate and analyze a large scRNA-seq database from several different sources.

2. Determination of biomarkers of RCC based on scRNA-seq data with the help of DGEA.

Biomarkers indicate the incidence of a disease, now RCC. These are typically genes that are significantly under- or overexpressed in one sample compared to another. Current biomarkers of RCC have been collected based on bulk sequencing data. Some recent studies have focused on identifying biomarkers of RCC from scRNA-seq data, however, this has not been done with an integrated dataset, with gene expression data from over 100,000 cells. The NCBI GEO database will be used to access the most reliable scRNA-seq datasets.

3. Identification of new potential drug target molecules in RCC.

The goal is to identify genes with high average expression levels in cancer cells, high ratios of cancer cells expressing these genes, and low ratios of other annotated cells doing so. As drug development is a costly process, to promote fast implementation of the targets, special focus will be placed on genes that are targets of drugs already approved in other types of diseases but not used currently against RCC.

4. Assembly of a database containing RCC drug target data and finding correlation between the ratio of tumor cells expressing the given target and the corresponding therapeutic efficacy.

The goal was to address potential underlying reasons of targeted therapy resistance. One reason could be that targeted genes are moderately expressed in cancer cells, despite them being overexpressed according to bulk RNA sequencing data. This lies on the assumption that the expression levels of drug target genes correlate with corresponding therapeutic efficacies.

Methods

Collecting data about anti-RCC drugs and their target genes

Drug data were collected from Drugbank (<https://go.drugbank.com/>), which is the most immense online database containing information about different therapeutic agents and their target molecules. I searched the disease “renal tumor”, and I did not only collect already approved drugs, but experimental agents currently tested in clinical trials, and novel preclinical compounds, as well. The database eventually contained 21 different drugs and 51 target molecules (Table 1), as most drugs are multi-target agents.

Drug name (Commercial name)	Drug type	Target genes
Axitinib (<i>Inlyta</i>)	Tyrosine-kinase inhibitor	FLT1, KDR, FLT4
Cabozantinib (<i>Cabometyx</i>)	Tyrosine-kinase inhibitor	MET, KDR, RET
Erlotinib (<i>Tarceva</i>)	Tyrosine-kinase inhibitor	EGFR, NR1I2
Lenvatinib (<i>Lenvima</i>)	Tyrosine-kinase inhibitor	FLT1, KDR, FLT4, FGFR1, FGFR2, FGFR3, FGFR4, PDGFRA, RET, KIT
Pazopanib (<i>Votrient</i>)	Tyrosine-kinase inhibitor	FLT1, KDR, FLT4, FGFR3, PDGFRA, KIT, PDGFRB, ITK, FGF1, SH2B3
Sorafenib (<i>Nexavar</i>)	Tyrosine-kinase inhibitor	BRAF, RAF1, FLT4, KDR, FLT1, FLT3, PDGFRB, KIT, FGFR1, RET
Sunitinib (<i>Sutent</i>)	Tyrosine-kinase inhibitor	PDGFRB, FLT1, KIT, KDR, FLT4, FLT3, CSF1R, PDGFRA, MET
Tivozanib (<i>Fotivda</i>)	Tyrosine-kinase inhibitor	FLT1, KDR, FLT4, KIT, PDGFRB, FLT3, PDGFRA, PTK6, TEK, FGFR1, MET
Everolimus (<i>Afinitor</i>)	mTOR inhibitor	MTOR
Temsirolimus (<i>Torisel</i>)	mTOR inhibitor	MTOR
Nivolumab (<i>Opdivo</i>)	Immune checkpoint inhibitor	PDCD1, CD274
Pembrolizumab (<i>Keytruda</i>)	Immune checkpoint inhibitor	PDCD1, CD274
Ipilimumab (<i>Yervoy</i>)	Immune checkpoint inhibitor	CTLA4
Bevacizumab (<i>Avastin</i>)	VEGF inhibitor	VEGFA, VEGFB, C1QA, C1QB, C1QC, FCGR3A, FCGR1A, FCGR2A, FCGR2B, FCGR2C
Belzutifan (<i>Welireg</i>)	HIF-2A inhibitor	EPAS1, HIF1A
Bemcentinib	AXL inhibitor	AXL
Olaparib (<i>Lynparza</i>)	PARP inhibitor	PARP1, PARP2, PARP3, AKR1C3

Aldesleukin (<i>Proleukin</i>)	Recombinant IL-2	IL2RA, IL2RB, IL2RG
Alemtuzumab (<i>Campath</i>)	Monoclonal antibody	CD52, FCGR3B, FCGR3A, FCGR1A, FCGR2A, FCGR2B, FCGR2C
Girentuximab	Monoclonal antibody	CA9, IL2
Relatlimab (<i>Opdualag</i>)	Monoclonal antibody	LAG3

Table 1. Anti-RCC drugs and their target genes. Data was collected with the DrugBank online database. This table contains the name of drug, the commercial name, the type and the corresponding target genes. In total 21 drugs and 51 target genes are listed.

Collecting clinical trial data of targeted therapy drugs against RCC

Anti-RCC drug data were collected from ClinicalTrials.gov (<https://clinicaltrials.gov/>), which is the most comprehensive online database containing information about privately and publicly funded clinical studies conducted all around the world. It is managed by the National Institute of Health (NIH). The website provides detailed information about clinical trials, including their purpose, participant eligibility and results. I searched for clinical trials connected to those, previously listed anti-RCC drugs that are targeted therapy agents (Table 2). I only observed trials with published results and where the number of enrolled patients was at least 30. I prioritized clinical trials where a monotherapy was applied, meaning that only one type of anti-RCC agent was given to the patients. In our research, we would like to correlate gene expression levels with a metric that defines therapy efficacy, which was median progression-free survival (PFS) in this study. PFS is the length of time during and after treatment that a patient lives with the disease, but it does not get worse. It is a measure of how well a treatment works in delaying disease progression. This value is shown on the “Results posted” page of each clinical trial Study Record.

Drug name	Clinical trial ID	Reference	Phase	Enrolled patients	Median PFS [months]
Axitinib	NCT00678392	BI Rini et al., Lancet, 2011 [13]	II	723	6.7
Cabozantinib	NCT01865747	TK Choueiri et al., Lancet Oncol, 2016 [14]	III	132	7.4
Lenvatinib	NCT01136733	Lee CH et al., Br J Cancer, 2021 [15]	II	52	7.4
Pazopanib	NCT00720941	Sheng X et al., BMC Cancer, 2020 [16]	III	557	8.4
Sorafenib	NCT01030783	RJ Motzer et al., J Clin Oncol, 2014 [17]	III	257	9.1
Sunitinib	NCT00720941	Sheng X et al., BMC Cancer, 2020 [16]	III	553	9.5

Tivozanib	NCT02627963	M Szarek et al., Clin Genitourin Cancer, 2021 [18]	III	175	5.6
Everolimus	NCT01206764	B Escudier et al., Annals of Oncol, 2012 [19]	II	92	7.3
Temsirolimus	NCT00494091	Y Sun et al., Jpn J Clin Oncol, 2012 [20]	II	76	7.2
Nivolumab	NCT01354431	RJ Motzer et al., J Clin Oncol, 2015 [21]	II	54	4.1
Pembrolizumab	NCT02853344	DF McDermott et al., J Clin Oncol, 2021 [22]	II	110	7.1
Ipilimumab	NCT02231749	RJ Motzer et al., Cancer, 2022 [23]	II	1390	8.4
Bevacizumab	NCT00738530	Wolfgang SS, Anticancer Res, 2019 [24]	III	649	10.2
Belzutifan	NCT04489771	MB Atkins et al., J Clin Oncol, 2022 [25]	II	78	9.1

Table 2. Clinical trial data connected to anti-RCC targeted therapy agents. Data was collected from ClinicalTrials.gov. The table contains the name of the drug, the trial ID, the reference article, the phase of the study, the number of enrolled patients in the corresponding cohort, and the median Progression-Free Survival (PFS) time in months. PFS values are not just relatively low, but also extremely similar to each other, proving that no current agent is truly effective in facing RCC. Highest PFS is connected to Bevacizumab, a VEGF inhibitor that targets VEGFA, VEGFB, C1QA, C1QB, C1QC, FCGR3A, FCGR1A, FCGR2A, FCGR2B and FCGR2C (based on Table 1). Notably, PFS does not reach one year even in this case.

Single-cell RNA sequencing datasets

Following the single-cell RNA sequencing protocol, post-processing starts with the analysis of a count matrix, which represents the number of transcripts observed for each gene and cell. I collected scRNA-seq data from a freely available online database, NCBI GEO. I used “*Kidney Cancer 10X*” as the search term and “*Homo sapiens*” as a species filter, thus I got 39 results. However, most of these results could not be used for further analysis in my research. The reason for exclusion was either the fact that there was no additional information provided, results had not been published in any scientific article, providers of the dataset did not use 10x Chromium technology, or they did not actually have patients with RCC. After manually filtering out non-optimal datasets for my project, I ended up with 3 GEO databases: GSE152938 [26], GSE213912 [27] and GSE207493 [28]. GSE152938 contained 2 RCC samples plus 1 normal tissue sample that I excluded. Each sample contained around 10,000 barcodes, which refer to cells (Table 3). GSE213912 contained 32 samples, from which I could use 3 tumor tissue samples that were derived from different RCC patients (Table 3). GSE207493 contained 19 scRNA-seq samples, from which I could use 15 (Table 3), as I only used datasets where the number of cells was larger than 1000. In summary, I used 20 different GEO datasets,

containing scRNA-seq information derived from RCC tumor tissue samples originating from different patients following nephrectomy, meaning that tissues were removed in the process of kidney-removing surgery.

Sample	GEO accession	Number of cells	Number of genes	Q1 TC	Q3 TC	Q1 %MT	Q3 %MT
19A4a	GSE213912	3322	24703	502	2542	0.5	5
19c7f	GSE213912	1463	24703	326	2822	0.5	4
19c26b	GSE213912	1859	24703	431	2420	0.5	5
RCC1	GSE152938	9762	32678	1312	17580	1	26
RCC2	GSE152938	9837	32678	1937	23482	1	15
RCC94	GSE207493	8070	33538	2356	24519	1	20
RCC96	GSE207493	8653	33538	2487	24783	1	21
RCC99	GSE207493	9920	33538	1934	18604	1	31
RCC100	GSE207493	11561	33538	2573	20340	1	11
RCC101	GSE207493	10004	33538	1254	10504	1	21
RCC103	GSE207493	10210	33538	1765	15319	1	15
RCC104	GSE207493	10009	33538	3555	40481	1	21
RCC106	GSE207493	10462	33538	3076	36048	1	43
RCC112	GSE207493	13126	33538	2854	27369	1	30
RCC113	GSE207493	9474	33538	2721	27092	1	17
RCC114	GSE207493	7803	33538	2435	20694	1	14
RCC115	GSE207493	9800	33538	1166	16091	1	25
RCC116	GSE207493	6884	33538	2684	32419	1	20
RCC119	GSE207493	12109	33538	2435	23923	1	30
RCC120	GSE207493	10243	33538	1973	22836	1	16

Table 3. Description of data used during my research. All raw datasets were downloaded from NCBI GEO. The first column refers to the sample names available on NCBI GEO, the second column contains GEO accession numbers, which are used for identification. I also represented the number of cells sequenced in each sample and the number of genes whose expression levels were addressed. Further columns represent statistic parameters for quality control (QC). QC metrics include the total number of read counts (TC) for each cell and the percentage of mitochondrial genes (MT genes). I calculated the 1st quantile and the 3rd quantile for these parameters that I included in the table above.

Pre-processing data

Data pre-processing in the case of scRNA-seq analysis is the computational process from raw count matrices to high-quality cellular data. I filtered out low-quality cells by using manually set threshold values for each sample. There are suggestions to find the proper values, however, no golden truths exist currently [29]. Low-quality cells have a low number of detected genes, low count depth, or high fraction of mitochondrial counts, which are related to dying cells, and thus should be eliminated. Genes can also be selected depending on the number of cells they are expressed in. The 20 samples that I used for analysis contained 174,611 cells and 18,694 genes in total before pre-processing. I filtered out cells in which less than 800 genes were expressed, and genes that were expressed in less than 3 cells. Expression is defined as the read count value being higher than zero. Cells followed a normal distribution considering the number of genes by counts, the number of total counts, and mitochondrial percentage. To identify outliers, I used the 1.5 IQR rule, meaning that data points were named outliers if they fell out of 1.5 times the interquartile range, the range between the third and first quantile (Table 3). Following pre-processing, I had 115,001 cells and 11,345 genes. Pre-processing also involved normalization of each sample database, which I did with the help of log normalization. This means applying a logarithmic function to all data points to stabilize variance and reduce skewness of the count matrices.

Dimensionality reduction and clustering

Dimensionality reduction techniques are used to extract relevant features from high dimensional data, reducing the complexity, thus allowing the computational possibility of subsequent analyses. the subsequent analysis results. In my research, I employed the Uniform Manifold Approximation and Projection (UMAP) algorithm for dimensionality reduction, and I used the Leiden algorithm for clustering. In brief, UMAP constructs a high-dimensional graph representation of the data, then optimizes a low-dimensional graph to be as structurally similar as possible [30]. UMAP is frequently used for its speed and capability to handle extensive datasets. Furthermore, its result can easily be visualized. The Leiden algorithm fine-tunes the community structure to identify the most optimal grouping of data points [31]. Following the steps of dimensionality reduction and clustering, genes are ranked according to their expression relative to other clusters, which is a part of differential gene expression

analysis (DGEA). To perform the DGEA, a *t-test* is employed, thus the calculated score is the *t*-statistic from the *t-test* comparing the mean expression levels between cancer cells, and all other cell types grouped together. The *Benjamini-Hochberg* method is used for correction.

Batch-effect correction and data integration

Effectively combining scRNA-seq datasets from various sources, laboratories, and experiments is essential for eliminating batch effects that could mask valuable biological insights. The best-performing algorithms for this task include scANVI, Scanorama, scVI and scGen, however, most of these are implemented in the R programming language [32]. During my research, I used Scanorama, an algorithm embedded in Python that identifies and merges the shared cell types among all pairs of datasets, ensuring the accurate integration of heterogeneous collections of scRNA-seq data [33].

Cell type annotation

After the processes of clustering and data integration, the functional type of each cell in the dataset should get labelled. Currently most methods use manual strategies to annotate cell types, however, such methods are extremely labor-intensive and heavily depend on the expertise of the user, which can result in inconsistent results. In my research, I employed SCSA, a tool designed to identify cell types through a scoring system, which combines differentially expressed genes (DEGs) and confidence levels of cell markers from the CellMarker2.0 and the CancerSEA databases, plus user-defined information [34]. I had to optimize the parameters of the algorithm for my current task. Cell annotation was based on the version two database that was updated in April 2023. The tissue type was human kidney cancer in the case of my project. I set the fold change threshold to 2, and the p-value parameter to 0.001. SCSA compares marker genes to the ranked gene expression data in my dataset to annotate different cell types.

Results

Designed bioinformatic workflow in Python to analyze scRNA-seq data

A crucial part of my research was the proper integration of different GEO datasets that all contain samples from tumor tissues of RCC patients sequenced with 10X Chromium technology, however, were collected in separate laboratories. Pre-processing was followed by data integration (Figure 3), for which I used the Scanorama Python package.

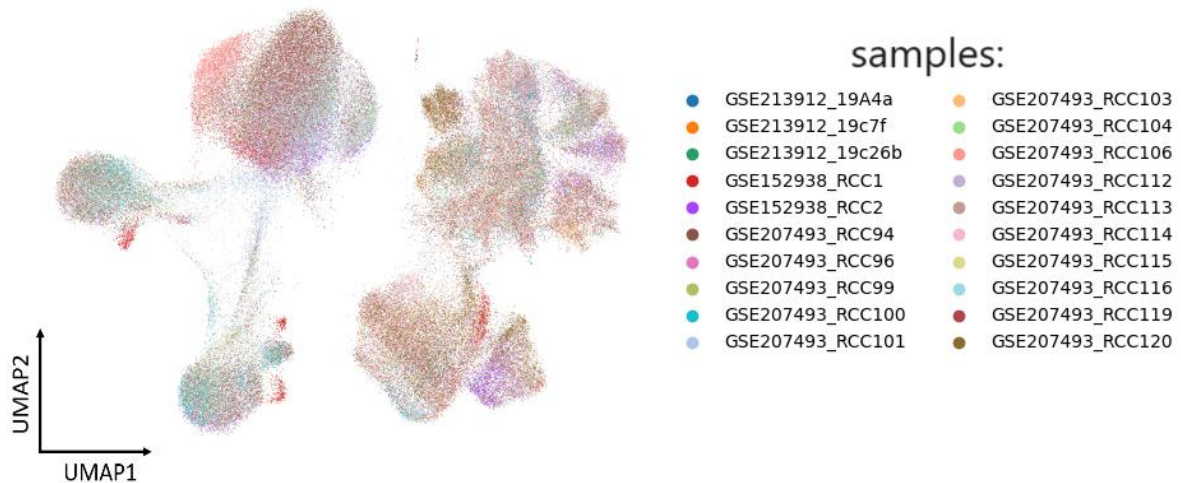


Figure 3. Visualization of the results of the Scanorama batch effect correction method. Each dot corresponds to a cell. Different colors represent different samples, the prefix refers to the original GEO accession number.

Our integrated database composed of 115,001 cells and 11,345 genes. I used UMAP and Leiden algorithms to reduce dimensionality and cluster the integrated scRNA-seq data, respectively, and identified 23 clusters. I performed DGEA on these clusters to identify marker genes that would later aid the SCSA cell type annotation algorithm (Figure 4). With SCSA, 32,977 cells were annotated as cancer cells. This means that 29% of all cells originating from RCC tissue were cancerous cells and the others are considered cells from the tumor microenvironment, mostly cells of the immune system. Other cell types in the integrated scRNA-seq dataset were B cells; CD8+ T cells; endothelial cells; fibroblasts; macrophages; neutrophils; plasma cells and a small portion of other types of cells, the number of annotated cells for each type was 3,222; 27,229; 11,659; 12,284; 14,829; 5,045 and 7,500, respectively. The algorithm could not classify few cells, these are marked as "NA" (Figure 4). This means that out of the originally identified 23 clusters, many were merged based on expressed genes, and eventually 9 groups (the previously listed cell types) could be formed.

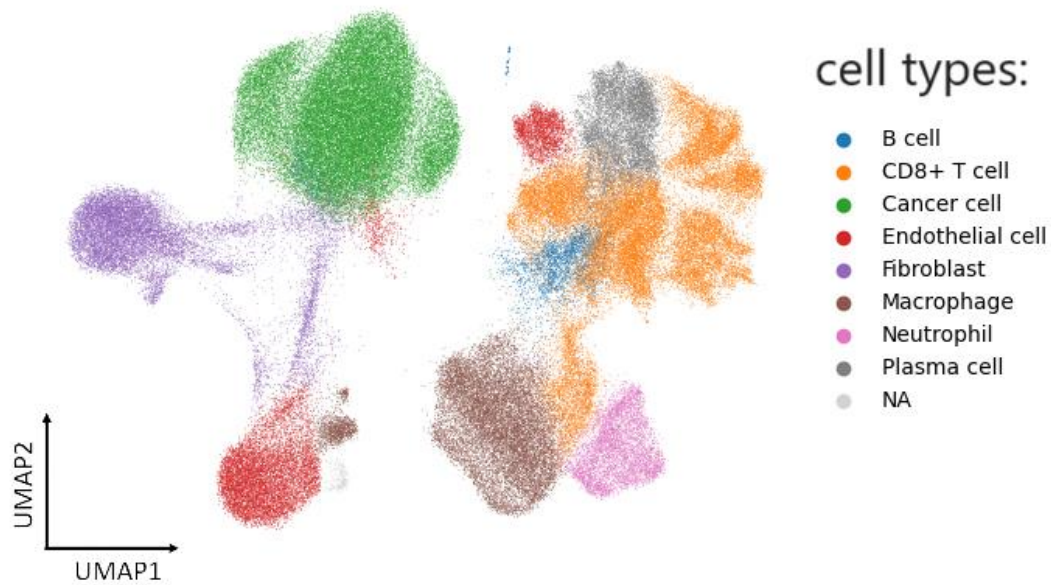


Figure 4. Visualization of annotated cell types using the SCSA algorithm. Different colors represent different cell types, annotated cancer cells are shown in green. Around 33,000 cancer cells were annotated from the original 115,001 cells. Other annotated cell types include B cells, CD8+ T cells, endothelial cells, fibroblasts, macrophages, neutrophils, and plasma cells.

Differentially expressed genes in annotated cancer cells

With the help of the annotated database (CellMarker2.0), I could do DGEA to identify markers of cancer cells, compared to other annotated cell types. To perform this, I grouped all non-cancer cells together, and used this batch as control of the DGEA. Scanpy's in-built function *rank_genes_by_group* identifies DEGs and creates a scoring system based on fold change and significance levels. Fold change is directly proportional to ranking, while lower p-values represent higher statistical significance, thus are ranked higher. DEGs could be considered biomarkers. Biomarkers are essential tools in research and medicine, as they enhance our ability to diagnose diseases accurately, personalize treatments and understand disease mechanisms. Cancer research often focuses on biomarkers, since disease progression can be monitored with their help, and treatment efficacy can also be evaluated. I found 30 of these potential biomarkers, from which the top 5 are: CD24 ($FC=4.42$; $p\text{-value}=0.0013$); CRYAB ($FC=4.31$; $p\text{-value}=0.00093$); TMEM176A ($FC=3.98$; $p\text{-value}=0.0017$); ATP1B1 ($FC=3.95$; $p\text{-value}=0.00087$) and CYB5A ($FC=3.81$; $p\text{-value}=0.0014$) (Figure 5).

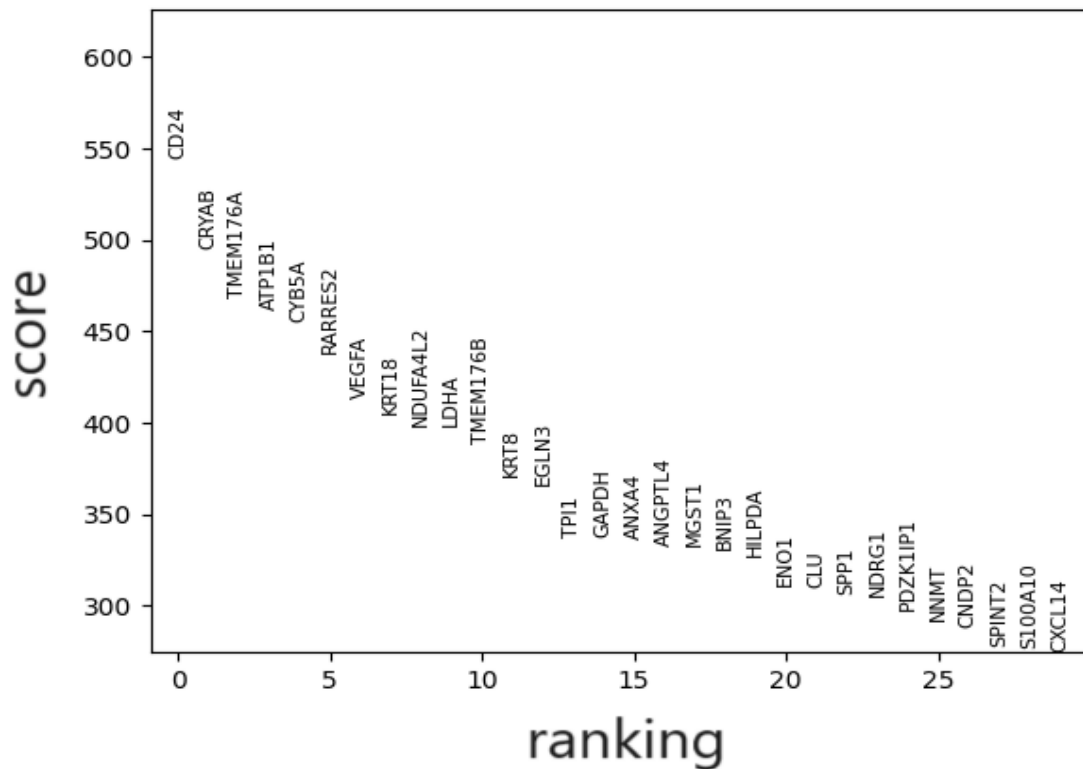


Figure 5. Marker genes of annotated cancer cells identified with DGEA. The x axis represents the ranks while the y axis shows the scores calculated based on fold change and p-values.

Genes expressed in most cancer cells

From here on, I only used gene expression data of annotated tumor cells. First, I observed all genes that were expressed in ATCs. I investigated average expression (AE) in cancer cells and the fraction (F) of cancer cells in which these genes are expressed.

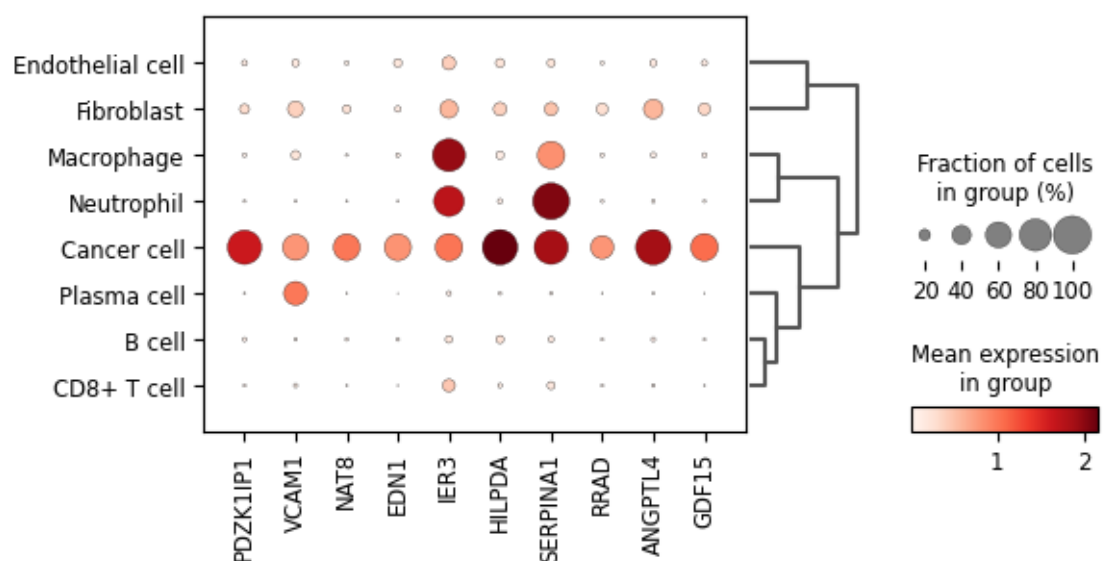


Figure 6. Complex dot plot showing each selected gene's mean expression in each group, plus the fraction of cells in the group in which the given gene is expressed.

I identified genes with the highest AE and F levels, the top 10 of these were: ANGPTL4 (AE=2.4; F=0.94); HILPDA (AE=2.4; F=0.92); SERPINA (AE=2.2; F=0.91); PDZK1IP1 (AE=2.1; F=0.89); GDF15 (AE=1.7; F=0.82); IER3 (AE=1.9; F=0.69); VCAM1 (AE=1.6; F=0.68); EDN1 (AE=1.7; F=0.67); NAT8 (AE=1.8; F=0.66) and RRAD (AE=1.5; F=0.58) (Figure 6). For this analysis, currently approved target genes were eliminated from the data. Highly expressed genes and biomarkers are not completely identical, since many genes that are highly expressed in cancer cells, can also be highly expressed in other cell types (for example IER3 on Figure 6).

Expression of used anti-RCC drug target genes in cancer cells

I observed the gene expression profiles of current target molecules. I found 23 target genes, which were expressed in more than 25% of cancer cells, and 11 target genes that were characterized with an average expression above 0.5, which means that the average read count value for that certain gene among annotated cancer cells was 0.5. I visualized the AE and F values of selected target genes for all annotated cell types with the help of a single, complex dot plot. The top 10 genes were: PPIA (AE = 2.21; F = 0.92); VEGFA (AE = 2.16; F = 0.84); UGT2B7 (AE = 1.00; F = 0.63); CA9 (AE = 0.97; F = 0.62); TUBB (AE = 0.87; F = 0.61); ATOX1 (AE = 0.81; F = 0.61); VEGFB (AE = 0.76; F = 0.55); EGFR (AE = 0.63; F = 0.46); CAMLG (AE=0.62; F = 0.49); TOP1 (AE= 0.53; F = 0.43) and SLC22A2 (AE = 0.53; F = 0.41) (Figure 7). However, many of these (for example PPIA and TUBB) are housekeeping genes, thus are irrelevant in my current research.

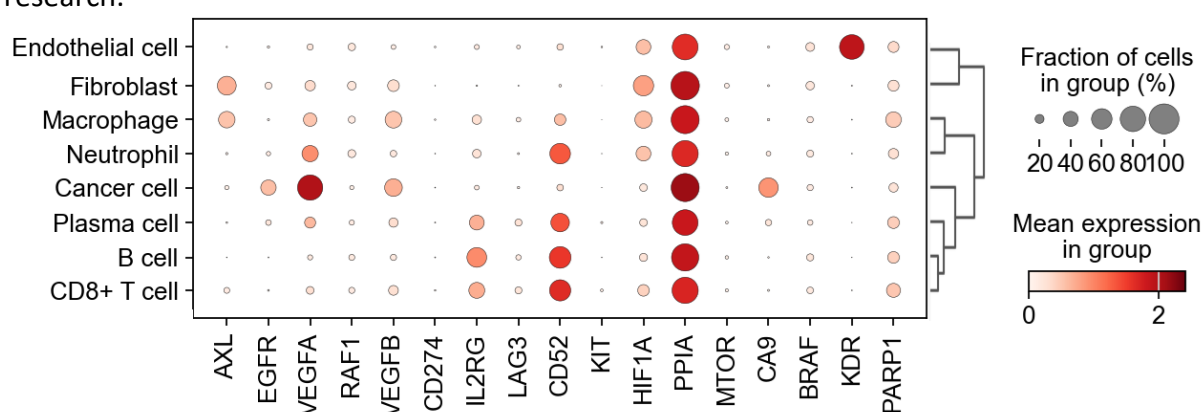


Figure 7. Complex dot plot showing the currently most commonly targeted genes' mean expression in each group, plus the fraction of cells in the group in which the given gene is expressed. The figure shows that EGFR; VEGFA; VEGFB; PPIA and CA9 have the highest mean expression in cancer cells, while other current drug targets like CD274; IL2RG and MTOR have a low expression in all cell types, including cancer cells.

Discussion

Potential biomarkers identified for RCC based on scRNA-seq data

Recent bioinformatic studies finding biomarkers mostly rely on the analysis of bulk RNA sequencing data. However, my analysis provided a much higher, cellular resolution. Thus, several new potential biomarkers for RCC were identified using differential gene expression analysis comparing the gene expression patterns of cancer cells to other cell types. The top 5 genes I identified as biomarkers are CD24, CRYAB, TMEM176A, ATP1B1 and CYB5A (Figure 5).

CD24 is a membrane protein whose increased levels have already been associated with poor prognosis in RCC cases [35]. CD24 is the ligand for P-selectin, an adhesive molecule present on activated endothelial cells. Crystallin Alpha B gene (CRYAB) has already been proven by scRNA-seq analysis to be overexpressed in RCC and have a significant role in tumor prognosis and immune infiltration [36]. The CRYAB protein is a molecular chaperone, it can prevent damages induced by heat shock, radiation, and oxidative stress. It might suppress apoptosis of tumor cells and regulate VEGF causing tumorigenesis [37]. Cation channel transmembrane protein 176 A is coded by the TMEM176A gene. It is a ubiquitous protein, whose low mRNA expression has been correlated with better overall survival in human gastric cancer [38]. The ATPase Na⁺/K⁺ Transporting Subunit Beta 1 (ATP1B1) gene encodes an integral membrane protein whose downregulation promotes cancer development in RCC [39]. Cytochrome B5 Type A (CYB5A) is a gene whose protein product is a membrane-bound hemoprotein functioning as an electron carrier for several membrane-bound oxygenases [40].

The findings suggest that the genes above are potentially new diagnostic biomarkers for RCC.

Currently available drugs targeting ANGPTL4, VCAM1, and EDN1 are suggested to be used against RCC

Finding marker genes of annotated RCC cells has great prognostic and diagnostic value, however, drug target identification is an even more powerful tool, especially if there already is a drug targeting the identified gene, which is not yet used against RCC. Target genes need to be expressed in a large portion of targeted cells and their expression must be relatively high in these. I observed the gene expression profiles of genes that were highly expressed in annotated cancer cells, and not in other cell types. The selected genes with the highest F

values and a relatively high AE level, plus available drugs against them were: ANGPTL4; VCAM1 and EDN1 (Figure 6).

Angiopoietin-related protein 4 (ANGPTL4) is encoded by a gene that has been reported to be involved in targeting cancer drug resistance in neck squamous cell carcinoma, however, the underlying mechanism of this process is currently unknown [41]. It has also been proved that ANGPTL4 is upregulated in a variety of tumors, and it promotes tumor proliferation and the process of epithelial-mesenchymal transition [42]. In general, ANGPTL4 controls lipoprotein catabolism and energy distribution across different tissues by inhibiting lipoprotein lipase. ANGPTL4 is a therapeutic target agent against dyslipidemia and cardiovascular diseases [43], however, it has never been used against RCC. Therefore, drugs such as Evinacumab, a monoclonal antibody targeting ANGPTL3 and ANGPTL4 could potentially be useful for RCC patients.

Vascular cell adhesion molecule 1 (VCAM1) is found on the surface of activated blood vessel cells and it mediates immune cell adhesion and extravasation. VCAM1 is overexpressed in RCC and is regulated by the von-Hippel Lindau (VHL) gene, which frequently gets mutated in RCC. Recent results suggest that differential effects on VCAM1 regulation depends on the type of mutation affecting VHL, VCAM1 expression levels are increased in RCC samples carrying VHL missense mutations [44]. Drugs targeting VCAM1, such as Carvedilol and Succinobucol already exist, they are currently used to treat atherosclerosis of the blood vessels for the heart, or coronary artery disease.

Endothelin 1 (EDN1) gene is an important regulator of renal cell growth and proliferation, and fluid and electrolyte excretion, as well. Patients with RCC tend to overexpress EDN1 and its receptors [45]. TBC-3711 is an investigational drug against heart failure and pulmonary hypertension. It is an endothelin antagonist, a potent mediator of blood vessel constriction and growth of smooth muscle in vascular walls.

The results suggest that these drugs already used against other diseases, should also be tested against RCC.

Correlation between drug efficacy and target expression at a single-cell level

In our study, we hypothesized that the mean expression levels of drug targets correlate with corresponding anti-RCC therapeutic efficacy. Single-cell RNA sequencing allowed scientists to observe gene expression at a much higher resolution. This technique enabled many studies that focus on exploring the tumor microenvironment, and some articles also focus on the connection between expression levels and therapeutic efficacy. The median expression levels of confirmed immune checkpoint genes have been shown to be associated with patient outcomes and hold potential for guiding immunotherapy decisions [46]. In our study, the most common target molecules had highly variable AE and F levels: VEGFA ($AE = 2.16$; $F = 0.85$); IL2R ($AE = 0.0711$; $F = 0.0445$); mTOR ($AE = 0.0271$; $F = 0.0251$) and PD-L1 ($AE = 0.0095$; $F = 0.0081$) (Figure 7). Out of these, only VEGFA was expressed in more than half of the annotated cancer cells with a relatively high average expression level, while other common targets such as IL2R, mTOR and PD-L1 were expressed in a small fraction of annotated cancer cells (4.45%; 2.51% and 0.81% respectively) and had a low AE level (0.0711; 0.0271 and 0.0095 respectively). Notably, VEGF-inhibitor Bevacizumab has the highest PFS (10.2 months), while Nivolumab, which targets PD-L1 promises the lowest PFS time (4.1 months). Other promising therapeutic agents' target genes also had high AE and F values, such as CA9 ($AE=1.12$; $F=0.62$). Drug efficacy is not easily measured, however, the current standard of care in treatment-naïve patients with RCC is VEGFR inhibitors combined with immune checkpoint inhibitors [47]. The results suggest a correlation between drug efficacy and target expression profiles explaining the low therapeutic potential of some of the standard drug treatments for RCC.

To summarize my findings, I developed an integral scRNA-seq analysis workflow in Python, used for different downstream analysis tasks, including data integration and cell type annotation. I was able to identify cancer cells in the tumor microenvironment and investigate gene expression profiles only in these cells. I identified biomarkers and potential target genes. I also proposed correlation between drug efficacy and target expression levels and ratios in annotated cancer cells.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Prof. Dr. Balázs Gyórfy and Dr. Otília Menyhárt, for the constant support during my TDK thesis work.

Author contribution

The whole presented workflow has been constructed by Zsófia Dobolyi, no previous or current results of the research group were used in this TDK thesis work.

Code and data availability

All data and codes used to perform the analyses presented in this study is available on GitHub at the following repository: https://github.com/dzsofi20/scRNAseq_TDK. This repository includes all raw data and Python scripts.

References

1. Padala, S. A., Barsouk, A., Thandra, K. C., Saginala, K., Mohammed, A., Vakiti, A., Rawla, P., and Barsouk, A. (2020) Epidemiology of renal cell carcinoma. *World journal of oncology* 11, 79.
2. Pontes, O., Oliveira-Pinto, S., Baltazar, F., and Costa, M. (2022) Renal cell carcinoma therapy: Current and new drug candidates. *Drug Discovery Today* 27, 304-314.
3. Tung-Wei, H., Jia-Hung, L., Kun-Tu, Y., Jen-Pi, T., Sheng-Wen, W., Hui-Chun, T., Wei-Tse, K., Shu-Hui, L., Ya-Wen, C., and Horng-Rong, C. (2013) Renal expression of hypoxia inducible factor-1 α in patients with chronic kidney disease: a clinicopathologic study from nephrectomized kidneys. *The Indian journal of medical research* 137, 102.
4. Huang, J. J., and Hsieh, J. J. (2020) The therapeutic landscape of renal cell carcinoma: from the dark age to the golden age. In *Seminars in nephrology* Vol. 40 pp. 28-41, Elsevier
5. Ghidini, M., Petrelli, F., Ghidini, A., Tomasello, G., Hahne, J. C., Passalacqua, R., and Barni, S. (2017) Clinical development of mTor inhibitors for renal cancer. *Expert opinion on investigational drugs* 26, 1229-1237.
6. Xu, W., Atkins, M. B., and McDermott, D. F. (2020) Checkpoint inhibitor immunotherapy in kidney cancer. *Nature Reviews Urology* 17, 137-150.

7. Motzer, R. J., Hutson, T. E., Cella, D., Reeves, J., Hawkins, R., Guo, J., Nathan, P., Staehler, M., de Souza, P., and Merchan, J. R. (2013) Pazopanib versus sunitinib in metastatic renal-cell carcinoma. *New England Journal of Medicine* 369, 722-731.
8. Janowitz, T., Welsh, S. J., Zaki, K., Mulders, P., and Eisen, T. (2013) Adjuvant therapy in renal cell carcinoma—past, present, and future. In *Seminars in oncology* Vol. 40 pp. 482-491, Elsevier
9. Lindgren, D., Eriksson, P., Krawczyk, K., Nilsson, H., Hansson, J., Veerla, S., Sjölund, J., Höglund, M., Johansson, M. E., and Axelson, H. (2017) Cell-type-specific gene programs of the normal human nephron define kidney cancer subtypes. *Cell reports* 20, 1476-1489.
10. Zhang, Y., Narayanan, S. P., Mannan, R., Raskind, G., Wang, X., Vats, P., Su, F., Hosseini, N., Cao, X., and Kumar-Sinha, C. (2021) Single-cell analyses of renal cell cancers reveal insights into tumor microenvironment, cell of origin, and therapy response. *Proceedings of the National Academy of Sciences* 118, e2103240118.
11. Baran-Gale, J., Chandra, T., and Kirschner, K. (2018) Experimental design for single-cell RNA sequencing. *Briefings in functional genomics* 17, 233-239.
12. Luecken, M. D., and Theis, F. J. (2019) Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular systems biology* 15, e8746.
13. Rini, B. I., Escudier, B., Tomczak, P., Kaprin, A., Szczyluk, C., Hutson, T. E. and Motzer, R. J. (2011) Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. *The Lancet*, 378(9807), 1931-1939.
14. Choueiri, T. K., Escudier, B., Powles, T., Tannir, N. M., Mainwaring, P. N., Rini, B. I. and Motzer, R. J. (2016). Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomised, open-label, phase 3 trial. *The lancet oncology*, 17(7), 917-927.
15. Lee, C. H., Shah, A. Y., Rasco, D., Rao, A., Taylor, M. H., Di Simone, C. and Motzer, R. J. (2021). Lenvatinib plus pembrolizumab in patients with either treatment-naïve or previously treated metastatic renal cell carcinoma (Study 111/KEYNOTE-146): a phase 1b/2 study. *The lancet oncology*, 22(7), 946-958.
16. Sheng, X., Jin, J., He, Z., Huang, Y., Zhou, A., Wang, J. and Guo, J. (2020). Pazopanib versus sunitinib in Chinese patients with locally advanced or metastatic renal cell carcinoma: pooled subgroup analysis from the randomized, COMPARZ studies. *BMC cancer*, 20, 1-8.

17. Motzer, R. J., Barrios, C. H., Kim, T. M., Falcon, S., Cosgriff, T., Harker, W. G. and Knox, J. J. (2014). Phase II randomized trial comparing sequential first-line everolimus and second-line sunitinib versus first-line sunitinib and second-line everolimus in patients with metastatic renal cell carcinoma. *Journal of Clinical Oncology*, 32(25), 2765-2772.
18. Szarek, M., Needle, M. N., Rini, B. I., Pal, S. K., McDermott, D. F., Atkins, M. B. and Escudier, B. J. (2021). Q-TWiST analysis of tivozanib versus sorafenib in patients with advanced renal cell carcinoma in the TIVO-3 study. *Clinical Genitourinary Cancer*, 19(5), 468-e1.
19. Escudier, B., Porta, C., Schmidinger, M., Algaba, F., Patard, J. J., Khoo, V. and Horwich, A. (2014). Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology*, 25, iii49-iii56.
20. Sun, Y., Rha, S., Lee, S. H., Guo, J., Ueda, T., Qin, S. and Akaza, H. (2012). Phase II study of the safety and efficacy of temsirolimus in East Asian patients with advanced renal cell carcinoma. *Japanese journal of clinical oncology*, 42(9), 836-844.
21. Motzer, R. J., Rini, B. I., McDermott, D. F., Redman, B. G., Kuzel, T. M., Harrison, M. R. and Hammers, H. J. (2015). Nivolumab for metastatic renal cell carcinoma: results of a randomized phase II trial. *Journal of clinical oncology*, 33(13), 1430-1437.
22. McDermott, D. F., Lee, J. L., Bjarnason, G. A., Larkin, J. M., Gafanov, R. A., Kochenderfer, M. D. and Atkins, M. B. (2021). Open-label, single-arm phase II study of pembrolizumab monotherapy as first-line therapy in patients with advanced clear cell renal cell carcinoma. *Journal of Clinical Oncology*, 39(9), 1020-1028.
23. Motzer, R. J., Martini, J. F., Mu, X. J., Staehler, M., George, D. J., Valota, O. and Ravaud, A. (2022). Molecular characterization of renal cell carcinoma tumors from a phase III anti-angiogenic adjuvant therapy trial. *Nature communications*, 13(1), 5959.
24. Schultze-Seemann, W., Schulz, H., Tschene, B. and Haeckl, M. (2019). Bevacizumab plus IFN-alpha-2a in first-line treatment of patients with advanced or metastatic renal cell carcinoma: A prospective german non-interventional study. *Anticancer Research*, 39(2), 875-882.
25. Atkins, M. B., Jegede, O. A., Haas, N. B., McDermott, D. F., Bilen, M. A., Stein, M., M. and Hammers, H. (2022). Phase II study of nivolumab and salvage nivolumab/ipilimumab in treatment-naïve patients with advanced clear cell renal cell carcinoma (HCRN GU16-260-Cohort A). *Journal of Clinical Oncology*, 40(25), 2913-2923.

26. Su, C., Lv, Y., Lu, W., Yu, Z., Ye, Y., Guo, B., Liu, D., Yan, H., Li, T., and Zhang, Q. (2021) Single-cell RNA sequencing in multiple pathologic types of renal cell carcinoma revealed novel potential tumor-specific markers. *Frontiers in Oncology* 11, 719564.
27. Szabo, P. M., Vajdi, A., Kumar, N., Tolstorukov, M. Y., Chen, B. J., Edwards, R., Ligon, K. L., Chasalow, S. D., Chow, K.-H., and Shetty, A. (2023) Cancer-associated fibroblasts are the main contributors to epithelial-to-mesenchymal signatures in the tumor microenvironment. *Scientific Reports* 13, 3051.
28. Yu, Z., Lv, Y., Su, C., Lu, W., Zhang, R., Li, J., Guo, B., Yan, H., Liu, D., and Yang, Z. (2023) Integrative single-cell analysis reveals transcriptional and epigenetic regulatory features of clear cell renal cell carcinoma. *Cancer Research* 83, 700-719.
29. Heumos, L., Schaar, A. C., Lance, C., Litinetskaya, A., Drost, F., Zappia, L., Lücken, M. D., Strobl, D. C., Henao, J., and Curion, F. (2023) Best practices for single-cell analysis across modalities. *Nature Reviews Genetics*, 1-23.
30. Pandit, V., Swain, A. K., and Yadav, P. (2022) Comparison of Dimensionality Reduction and Clustering Methods for Single-Cell Transcriptomics Data. *bioRxiv*, 2022.2010.2015.512334.
31. Yu, L., Cao, Y., Yang, J. Y., and Yang, P. (2022) Benchmarking clustering algorithms on estimating the number of cell types from single-cell RNA-sequencing data. *Genome biology* 23, 49.
32. Luecken, M. D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Müller, M. F., Strobl, D. C., Zappia, L., Dugas, M., and Colomé-Tatché, M. (2022) Benchmarking atlas-level data integration in single-cell genomics. *Nature methods* 19, 41-50.
33. Hie, B., Bryson, B., and Berger, B. (2019) Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nature biotechnology* 37, 685-691.
34. Cao, Y., Wang, X., and Peng, G. (2020) SCSA: a cell type annotation tool for single-cell RNA-seq data. *Frontiers in genetics* 11, 490.
35. Arik, D., Can, C., Dünder, E., Kabukçuoğlu, S., and Paşaoğlu, Ö. (2017) Prognostic significance of CD24 in clear cell renal cell carcinoma. *Pathology & Oncology Research* 23, 409-416.
36. Cheng, L., Zou, X., Wang, J., Zhang, J., Mo, Z., and Huang, H. (2023) The role of CRYAB in tumor prognosis and immune infiltration: A Pan-cancer analysis. *Frontiers in Surgery* 9, 1117307.

37. Chen, W., Lu, Q., Lu, L., and Guan, H. (2017) Increased levels of alphaB-crystallin in vitreous fluid of patients with proliferative diabetic retinopathy and correlation with vascular endothelial growth factor. *Clinical & experimental ophthalmology* 45, 379-384.
38. Sun, L., Zhang, Y., and Zhang, C. (2018) Distinct expression and prognostic value of MS4A in gastric cancer. *Open Medicine* 13, 178-188.
39. Zhang, D., Zhang, P., Yang, P., He, Y., Wang, X., Yang, Y., Zhu, H., Xu, N., and Liang, S. (2017) Downregulation of ATP1A1 promotes cancer development in renal cell carcinoma. *Clinical proteomics* 14, 1-12.
40. Huang, Y., Wang, Q., Tang, Y., Liu, Z., Sun, G., Lu, Z., and Chen, Y. (2022) Identification and validation of a cigarette smoke-related five-gene signature as a prognostic biomarker in kidney renal clear cell carcinoma. *Scientific Reports* 12, 2189.
41. Liao, Y., Chiang, K., Shieh, J., Huang, C., Shen, C., Huang, W., and Chen, B. (2017) Epidermal growth factor-induced ANGPTL4 enhances anoikis resistance and tumour metastasis in head and neck squamous cell carcinoma. *Oncogene* 36, 2228-2242.
42. Bao, Y., Yang, F., Liu, B., Zhao, T., Xu, Z., Xiong, Y., Sun, S., Qu, L., and Wang, L. (2018) Angiopoietin-like protein 3 blocks nuclear import of FAK and contributes to sorafenib response. *British journal of cancer* 119, 450-461.
43. Aryal, B., Price, N. L., Suarez, Y., and Fernández-Hernando, C. (2019) ANGPTL4 in metabolic and cardiovascular disease. *Trends in molecular medicine* 25, 723-734.
44. Labrousse-Arias, D., Martínez-Alonso, E., Corral-Escariz, M., Bienes-Martínez, R., Berridy, J., Serrano-Oviedo, L., Conde, E., García-Bermejo, M.-L., Giménez-Bachs, J. M., and Salinas-Sánchez, A. S. (2017) VHL promotes immune response against renal cell carcinoma via NF- κ B-dependent regulation of VCAM-1. *Journal of Cell Biology* 216, 835-847.
45. Schütte, U., Bisht, S., Heukamp, L. C., Kebschull, M., Florin, A., Haarmann, J., Hoffmann, P., Bendas, G., Buettner, R., and Brossart, P. (2014) Hippo signaling mediates proliferation, invasiveness, and metastatic potential of clear cell renal cell carcinoma. *Translational oncology* 7, 309-321.
46. Xia, Z. N., Wu, J. G., Yao, W. H., Meng, Y. Y., Jian, W. G., Wang, T. D. and Zhang, C. (2022). Identification of a differentiation-related prognostic nomogram based on single-cell RNA sequencing in clear cell renal cell carcinoma. *Scientific Reports*, 12(1), 10973.
46. Hirsch, L., Flippot, R., Escudier, B., and Albiges, L. (2020) Immunomodulatory roles of VEGF pathway inhibitors in renal cell carcinoma. *Drugs* 80, 1169-1181.