Aus der Klinik für Urologie
der Medizinischen Fakultät Charité - Universitätsmedizin Berlin

DISSERTATION

The role of cyclin-dependent kinase 18 (CDK18) in clear cell renal cell carcinoma

Die Rolle der Cyclin-abhängigen Kinase 18 (CDK18) im klarzelligen Nierenzellkarzinom

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Sinisa Simonovic

aus Serbien

Datum der Promotion:

04.06.2021

Table of contents

Tables

Figures

1. Abstract (English)

Introduction:

The most common and one of the more aggressive types of kidney cancer is clear cell renal cell carcinoma. It is characterized by sporadic occurrence, poor prognosis, and high resistance to therapies, which necessitates the discovery of new biomarkers for improving diagnostics and prognostics. The aim of the study was to identify a panel of genes whose mRNA is strongly upregulated in clear cell carcinoma tissue, in order to develop a qPCR detection assay based on their potential differential expression in the blood of cancer patients compared to healthy individuals. A further aim was to functionally characterize a novel gene in cell lines representing this cancer.

Methodology:

The construction of the gene panel was performed by a bioinformatic analysis of several databases containing tissue (tumor and normal) and blood expression (from healthy individuals) of all genes in the genome. The presence of selected genes was tested in tissue and blood of patients and healthy individuals by RT-qPCR. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) /Cas9 system enabled the generation of stable knockout clones for a loss-of-function analysis, and RNA sequencing allowed for the global transcriptome analysis of the knockout condition, revealing the possible mechanism of action of the investigated gene.

Results:

A ranked list of genes overexpressed in clear cell carcinoma tissue compared to adjacent normal kidney tissue was produced, among them CDK18 (cyclin-dependent kinase 18), CCND1 and LOX. Two genes, CDK18 and CCND1 were underexpressed in the blood of clear cell carcinoma patients, and LOX showed a tendency towards upregulation in metastatic compared to non-metastatic blood samples. CDK18 knockout in two renal cancer cell lines led to a reduced proliferation rate, possibly via effects on WDR77 and SOAT1, the former being downregulated, and the second showing a tendency towards downregulation in the knockout condition.

Conclusions:

This study exemplifies the difficulty of detecting tumor specific mRNAs in blood and revealed paradoxical underexpression of two genes in the blood of clear cell carcinoma patients contrary to tissue expression. It could establish the effect of CDK18 on tumor cell proliferation and suggest its possible mechanism of action, which should be further evaluated.

2. Abstract (Deutsch)

Hintergrund:

Die häufigste und eine der aggressiveren Formen von Nierentumoren ist das klarzellige Nierenzellkarzinom. Charakteristika sind sporadisches Auftreten, schlechte Prognose und hohe Therapieresistenz, und deswegen ist die Entdeckung neuer Biomarker zur Verbesserung von Diagnostik und Prognose erforderlich. Das Ziel der Studie war, eine Gruppe von Genen zu identifizieren, deren mRNA im klarzelligen Nierenzellkarzinom stark hochreguliert ist, um einen qPCR-Assay zu entwickeln, der auf ihrer potenziellen differenziellen Expression im Blut von Krebspatienten im Vergleich zu gesunden Personen basiert. Ferner sollte, ein neues Gen in Zelllinien, die diesen Tumor repräsentieren, funktionell charakterisiert werden.

Methoden:

Die Gruppe von Genen wurde durch bioinformatische Analyse mehrerer Datenbanken, die die Gewebe- und Blutexpression aller Gene enthalten, herausgefiltert. Die Expression von ausgewählten Genen wurde in Gewebe und Blut von Patienten und gesunden Personen durch RT-qPCR bestimmt. Das CRISPR/Cas9-System ermöglichte die Erzeugung von stabilen Knockout-Klonen für die Funktionsverlustanalyse, und die RNA-Sequenzierung ermöglichte die globale Transkriptomanalyse des Knockout-Zustands und die Aufdeckung des möglichen Wirkmechanismus des untersuchten Gens.

Ergebnisse:

Eine Rangliste von Genen, die in klarzelligem Nierenzellkarzinomgewebe im Vergleich zu benachbartem normalem Nierengewebe überexprimiert sind, wurde erstellt, darunter CDK18, CCND1 und LOX. Zwei Gene, CDK18 und CCND1, waren im Blut von Klarzellkarzinompatienten vermindert exprimiert, und LOX zeigt eine Tendenz zur

Hochregulation bei metastatischen im Vergleich zu nicht-metastatischen Blutproben. CDK18 Knockout in zwei Nierenkrebszelllinien führte zu einer reduzierten Proliferationsrate, möglicherweise durch Effekte auf WDR77 und SOAT1, wobei das erste herunterreguliert war und das zweite eine Tendenz zur Herunterregulierung im Knockout-Zustand zeigte.

Schlussfolgerungen:

Diese Studie veranschaulicht die Schwierigkeit, tumorspezifische mRNAs im Blut nachzuweisen, und zeigte paradoxerweise eine verminderte Expression von zwei Genen im Blut von Klarzellkarzinompatienten entgegen der Überexpression im Gewebe. Die Studie konnte den Einfluss von CDK18 auf die Tumorzellproliferation belegen und einen möglichen Mechanismus dafür aufzeigen, der noch näher erforscht werden sollte.

3. Introduction

3.1. Analysis of potential blood-based biomarkers for clear cell renal cell carcinoma

In the United States, it is estimated that 73,820 new cases and 14,770 deaths from kidney cancer will occur in 2019 (1). The most common renal cancer is clear cell renal cell carcinoma (ccRCC), itself responsible for approximately 80% of all cases of renal cancer (2), while together with papillary and chromophobe carcinoma, it comprises around 2% of all cancers in the world (3). Morphologic hallmarks of ccRCC are strong lipid and glycogen accumulation in the cytoplasm of tumor cells, while metabolically it features abundant reprogramming in glucose, lipid, and amino acid pathways, reflecting the characteristic "clear cell" phenotype (4). Genetically, ccRCC primarily harbors inactivation of the von Hippel-Lindau tumor suppressor gene (VHL), which negatively regulates hypoxia-inducible factor (HIF) proteins in an oxygen-sensitive manner (5). The loss of VHL enables constitutive activation of HIF-α subunits, and HIFs generally operate as activators of genes involved in glycolysis, angiogenesis, migration and metastasis of tumor cells (6). The incidence of ccRCC is twice higher for male compared to female individuals, and increases with age reaching the maximum at 50–70 years (7). Body weight, hypertension and cigarette smoking represent the major risk factors for RCC (8). Additionally, links have been made to various lifestyle, dietary, and environmental factors (9). RCC is asymptomatic until the late stages, and more than 50% of cases are discovered accidentally during imaging studies (10, 11). Hematuria, flank pain and weight loss are classical symptoms, present in only 10% of patients, while around 25% of RCC cases have already metastasized by the time of diagnosis (12). The main reason for inefficient RCC treatments is the high unresponsiveness of this cancer to conventional chemotherapy and radiation (13, 14). Nephrectomy is the gold standard for the treatment of renal masses, after which however approximately a third of patients develop recurrence or metastases (15, 16).

New biomarkers are urgently required for improved detection, diagnostics and the prediction of clinical outcomes of patients with RCC, as current models used for prognostics being based on conventional clinicopathology and imaging are insufficiently accurate (17-19). Biomarkers are characterized by their specificity, sensitivity, and reproducibility. As vehicles for cancer biomarker discovery, plasma, serum, and urine have gained interest, providing proteins, DNA, and various RNA species. Particular value in terms of kidney disease and low invasiveness

lies in blood. However, in spite of the progress in this field, none of the thus far identified ccRCC biomarkers have been clinically validated (20).

Blood-circulating RNA is normally less than 100 bp in length (21) due to degradation, and even though systems such as the PAXgene platform have made the stabilization and storage of whole blood mRNA possible, studies are usually limited to shorter RNA fragments or RNA molecules shielded from degradation due to their specific structure or linkage with proteins or membranous vesicular structures. Apart from blood, urine would be especially suitable as a source of ccRCC biomarkers. Nonetheless, this field is by far less productive in comparison with blood-oriented studies. As in the case of blood, RNA detection problems arise due to the presence of RNAses, but also because of PCR inhibition (22), so the focus is shifting toward the analysis of shorter RNA subspecies. Starting from plasma, liquid biopsy has spread to include other bodily fluids in an increasing number of cancers, making rapid advancement since 2008 (23). As for most cancers there is only sporadic progress in detecting tumor-derived mRNA in blood and making successful associations with cancer prognosis (24-26), utilization of
circulating tumor cells (CTCs), cell free DNA (cfDNA), and RNA species such as miRNA, circular RNA (circRNA) and long non-coding RNAs (lncRNA), are turning out to be more productive lines of approach. In addition, tumor-derived RNA detection is hindered by the uncertainty regarding the exact RNA source, as it can be derived from solid tumor or CTCs. Another question is what percentage of source cells are alive and actively secreting RNA as opposed to being in apoptosis (27).

The five proposed stages in biomarker development are: preclinical exploratory phase, clinical assay and validation, retrospective longitudinal phase, prospective screening, and finally cancer control (28). The first stage involves the comparison of tumor with non-tumor tissue, where techniques such as microarrays and more recently RNAseq are employed to evaluate gene expression; immunohistochemistry and mass spectroscopy are used to obtain information on protein expression with the end goal being the discovery of genes displaying dysregulation-normally overexpression in tumor compared to normal tissue. The second phase uses blood for non-invasive screening, and here blood levels of selected genes do not necessarily have to precisely mirror the expression in tissue. This could potentially stem from the specific rate of mRNA release from cancer tissue into blood. The approach conducted in this work, namely starting with mRNA expression of tumor tissue and analyzing the levels of respective transcripts in blood by quantitative real time PCR (RT-qPCR), has been used successfully, and led to promising assays meriting clinical validation. A notable recent study presented a successful validation of an RT-PCR assay based on prostate-specific RNA in whole blood from 97 patients with metastatic

castration-resistant prostate cancer (mCRPC) (29). Several databases were queried to select a panel of top 10 genes overexpressed in prostate tissue but which at the same time showed no detection in peripheral blood mononuclear cells (PBMC). Subsequently, blood samples of cancer patients and volunteers were analyzed by RT-PCR leading to the establishment of a 5-gene panel (KLK3, KLK2, HOXB13, GRHL2, FOXA2). This panel improved and could be used in combination with the previously established CTC enumeration assay, being prognostic for survival and assessment of patient risk. Likewise, another study dealt with the early detection of colorectal cancer (30), utilizing a meta-analysis of microarray data in order to identify RNAs with highest differential expression between cancer tissue and normal blood samples. The following RT-qPCR analysis demonstrated that blood expression of 3 specific genes shows promising sensitivity and specificity with regard to detection of colorectal cancer.

The first stage of this work was the identification of genes most highly overexpressed in the tissue of ccRCC patients, by using the data from the Cancer Genome Atlas database (TCGA, (31)). Subsequently, additional databases were used to arrive at a subset of genes that have no blood expression; this subset was tested by RT-qPCR in whole blood samples from ccRCC patients and healthy individuals (Figure 1).

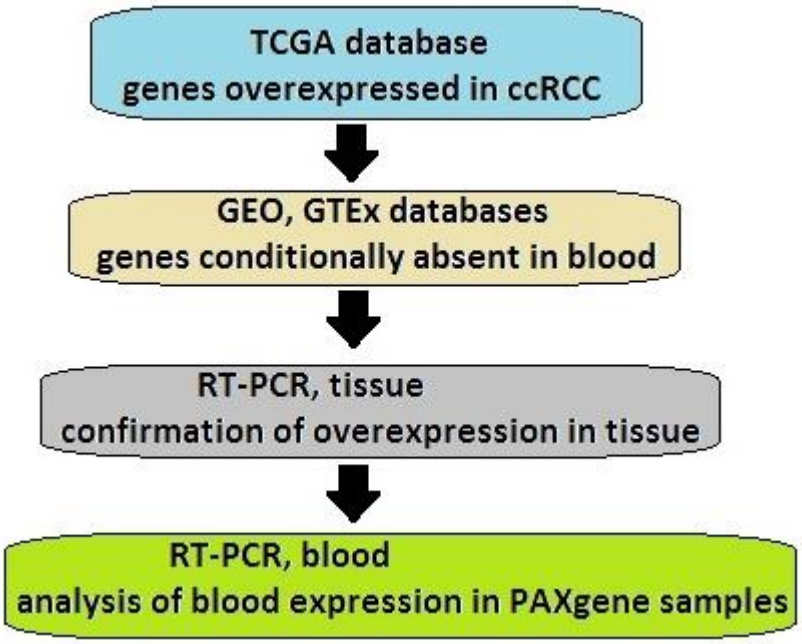


Figure 1. Workflow diagram.

Additionally, with the TCGA as the starting point, a separate selection workflow was used in order to choose one novel gene which would be investigated in renal cancer cell lines with respect to its potential functions in ccRCC.

3.2. Selection of genes overexpressed in ccRCC tissue for functional studies using CRISPR/Cas9-induced loss of function

A search of literature and consultation of the Protein Atlas was conducted for 75 top-ranked genes by ccRCC overexpression in TCGA while disregarding their blood expression, thereby arriving at a list of genes thus far uninvestigated in ccRCC and with potentially tumor-related roles. Owing to antibody availability, several genes were tested in ccRCC tissue, and after a consultation with a pathologist some of those were used for tissue microarray analysis, of the highest priority being CDK18, an incompletely researched cyclin-dependant kinase (Figure 2). Immunostainings were performed on ccRCC and adjacent normal tissue and external tissues were used as known positive controls, with reference to the Protein Atlas. Signal strength and presumable subcellular localization of the proteins were studied, followed by performing test cuts to determine optimal antibody dilutions and finally full tissue microarray to investigate if any of the genes show significant dysregulation. In order to confirm enhanced RNA levels in ccRCC compared to matched normal tissue, in accordance with the TCGA data, qPCR was performed on 8 patient samples. RNA and protein levels were investigated in 4 renal cancer cell lines (ACHN, A498, 786-O, CAKI-1) and a normal renal line (HK2), providing information about the suitability of each line for the downstream loss of function analysis.

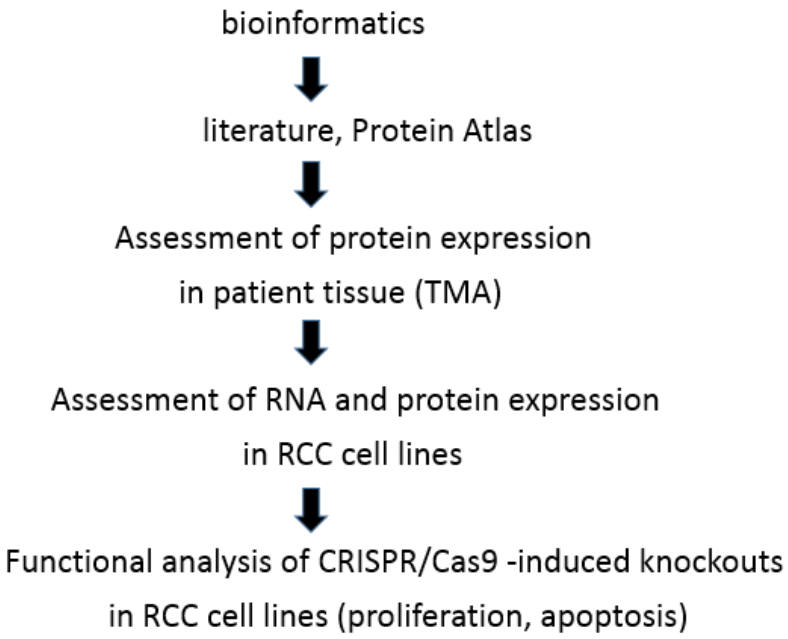


Figure 2. Gene selection workflow for CRISPR/Cas9 knockout generation.

3.3. CRISPR/Cas9 system

CRISPR/Cas9 system has become the predominant tool for gene editing and generation of knockouts (32, 33) and this approach was used here to elucidate functional roles of CDK18 in renal carcinoma cell lines. In comparison with previously dominant genetic engineering systems, such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), CRISPR is faster, less expensive, more efficient and versatile, as well as adaptable to any model system in use. It circumvents the creation of large guiding proteins, and multiple gRNAs can be used to target different loci at the same time (multiplexing) (34). It is recognized for its growing applications in biotechnology and medicine and advances in cancer research.

CRISPR systems are adaptive immune systems against bacteriophages and mobile genetic elements, found in most archaeal and half of bacterial genomes. Cas (CRISPR associated) proteins recognize and cut foreign DNA into short fragments ~30 bp in length, thereby adding them as spacers to the CRISPR array, which acts as a memory reservoir of past infections. Cas proteins are themselves encoded by a group of genes adjacent to the CRISPR array. Besides spacer regions, the CRISPR array contains direct repeats, which are indispensable for RNA processing. It is initially transcribed into a long RNA that is subsequently processed into mature CRISPR RNAs (crRNAs), which can direct the Cas complex to the foreign DNA based on sequence specificity, after which cleavage occurs (35). (Figure 3). Phages themselves combat this system in various

ways, their sophistication ranging from simply random mutations of PAM (Protospacer Adjacent Motif) or protospacer sequences, through deployment of small proteins that inhibit the interference machinery (e.g. via interaction with different subunits of the cascade and thereby preventing the binding to the target DNA) to turning CRISPR/Cas itself against their host (33). Another necessary element is the trans-activating crRNA (tracrRNA), which acts as a scaffold linking the crRNA to Cas, and facilitates the processing of pre-crRNAs.

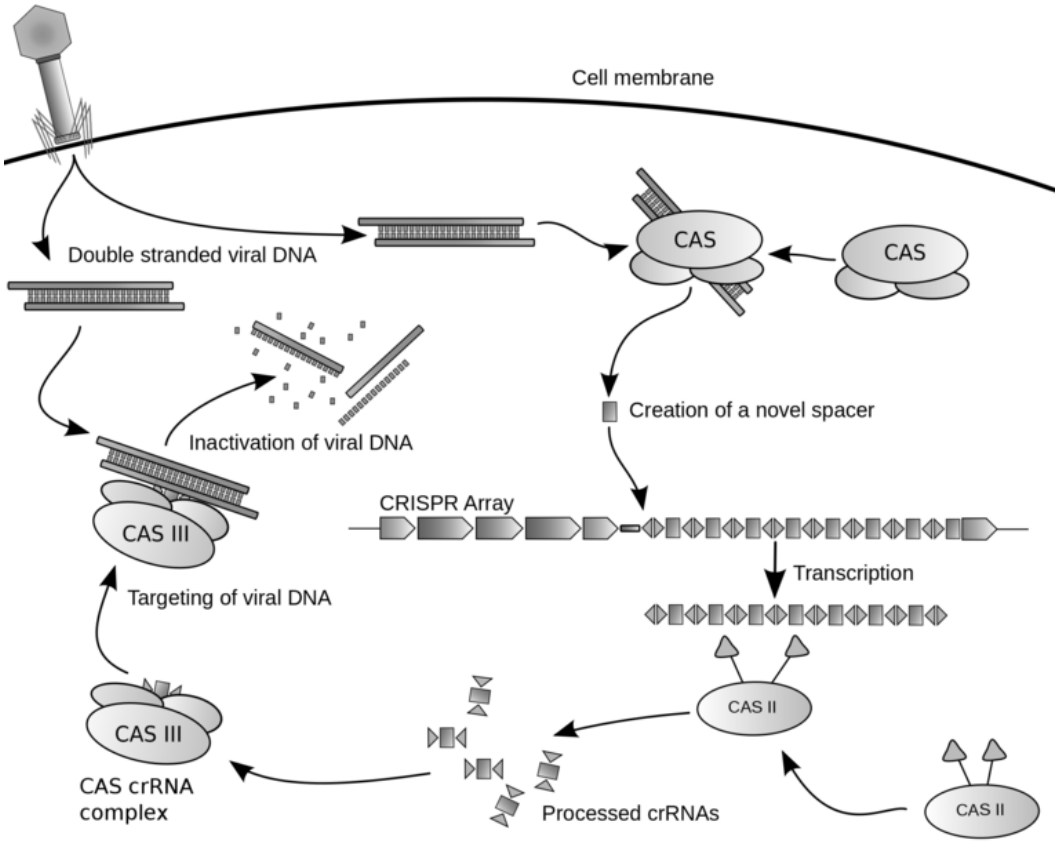


Figure 3. Prokaryotic CRISPR/Cas systems keep memory of previous infections in the CRISPR array, which upon infection is used to activate RNA-guided nucleases that perform sequence-specific cutting of the invading genetic material (adapted from Horvath et al, 2010 (35)).

The synthetic hybrid of crRNA and tracrRNA is called the guide RNA (gRNA), and in 2012 its experimental use within the CRISPR/Cas system for programmable targeted DNA cleavage was demonstrated (36). In the experimental setup, the only enzyme required is Cas9, a nuclease derived from Streptococcus pyogenes (37), which binds to both gRNA and target DNA which it cleaves. Therefore, Cas9 is a programmable nuclease that can be guided to any PAM-adjacent site inside the genome. Importantly, in order for Cas9 endonuclease to bind the target sequence and

perform a double-strand break, it requires both the guide RNA as well as a 3-base pair sequence known as the PAM, which must be located ~3-4 base pairs downstream of the cut site. The requirement of a suitable PAM next to the target sequence constitutes the major limitation of the CRISPR/Cas9 system in comparison with previous systems. Protospacer itself represents the foreign sequence of an invading microbe which is identical to the spacer. The absence of PAM in the CRISPR array of the host prevents possible auto-immune activation (33). Finally, when double-strand breaks are repaired through non-homologous end joining, insertions/ deletions are produced, potentially causing a frameshift mutation, and therefore enabling a loss-of-function study of a particular gene (38). (Figure 4).

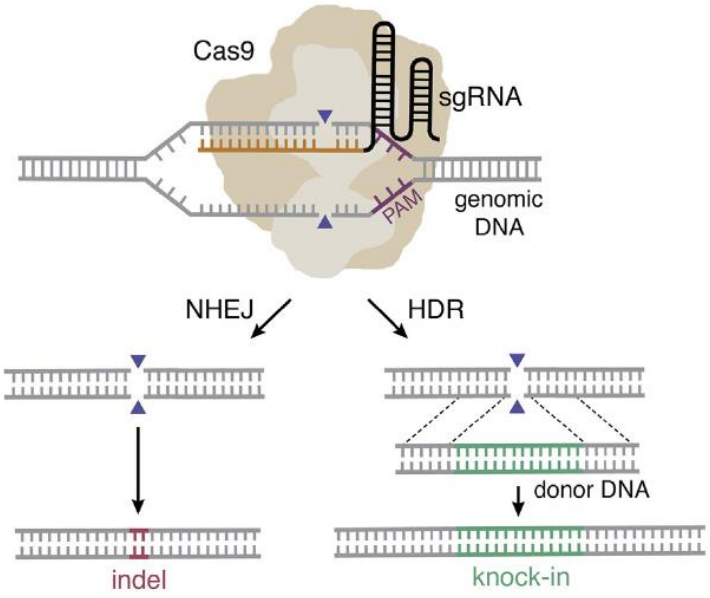


Figure 4. gRNA leads Cas9 to the target site resulting in a double-strand break; if homologous sequences are available the break is repaired by homology-directed repair, while in their absence the outcome is non-homologous end joining, which may result in insertion/deletion mutations (adapted from Moses et al, 2018 (38)).

3.4. Cyclin-dependent kinase 18

Cyclin-dependent kinases are a family of serine-threonine kinases initially discovered for their evolutionarily conserved role in the regulation of the cell cycle. The progression of cell cycle is driven by consecutive rise and fall of CDK activity. They are also involved in regulating transcription, mRNA processing, and the differentiation of nerve cells. Although CDKs are

traditionally divided into cell-cycle and transcriptional CDKs, these functions coexist in many members. CDKs become active upon binding a cyclin, which plays the role of their regulatory subunit, and whose protein levels are closely controlled during the cell cycle (39).

The PCTAIRE protein kinases (PCTAIRE are one letter abbreviations for amino acids) are a subfamily of cdc2-related serine/threonine protein kinases which have a single amino acid (cysteine instead of serine) substitution in the PCTAIRE motif. Their serine/threonine kinase domain has high homology to cdc2, while N and C-terminal domains are unique (40). It is hypothesized that the unique domains may replace the function of regulatory cyclins, although cyclin Y has been implicated as a potential binding activator for PCTAIRE-1 (41, 42) and cyclin A for PCTAIRE-3. PCTAIRE kinase subfamily comprises PCTK1/CDK16, PCTK2/CDK17, and PCTK3/CDK18. Apart from their expression in brain tissue, PCTAIRE 1 and 3 have been noted in postmeiotic germ cells suggesting they may be important in processes such as division, gametogenesis and differentiation. However, they may also have a more general regulatory function as they are also present in non-proliferating types of cells (brain and kidney) (43, 44). The best characterized of the PCTAIRE family is PCTAIRE-1, which has been shown to be regulated by protein kinase A (PKA) and was found to have a role in neurite outgrowth (45) and exert an effect on membrane trafficking via the early secretory pathway (46). PKA phosphorylation depresses the kinase activity of PCTAIRE-1 and appears to be a significant point of regulation of the PCTAIRE kinases (45). PCTAIRE-1 binds the p35 regulatory subunit of the CDK5 kinase and furthermore, its activity is increased following CDK5-dependent phosphorylation (47). It was shown to be highly expressed in prostate tumor lesions compared to adjacent normal tissues and comparison of PCTAIRE1 immunostaining with Gleason grade showed low expression levels in highly differentiated tumors relative to the less- differentiated ones (48). In another study, RNAi-mediated silencing of PCTAIRE1 provoked the inhibition of growth including irregular with abnormal mitosis resulting from centrosome behavior defects in prostate cancer cells, and its association with p27 was further analyzed as well (49).

CDK18 can apparently be activated by binding cyclin A2 and PKA (cAMP (cyclic adenosine monophosphate)-dependent protein kinase) as well as cyclin E1 (50). It can affect cell migration and adhesion in HEK293T cells, which is performed through negative regulation of FAK (focal adhesion kinase) and reorganizational changes of actin cytoskeleton. Its overexpression has been shown to lead to the formation of filopodia during the early stages of cell adhesion in HeLa cells (50, 51). An important role for CDK18 in replication stress and enhancement of genome stability has been revealed, via association with RAD9, a member of the 9-1-1 replication stress signalling complex (52). After experimental amplification of CDK18 in breast cancer cells using a dCRISPR

approach, cells were more likely to amass DNA damage, which was visualized by staining with a γ-H2AX, a marker of double strand breaks. The activation of γ-H2AX was wide across the nucleus, signifying a diffuse interference with DNA replication. These cells became especially sensitive to replication stress-inducing chemotherapeutic agents, because of the affected replication stress signalling. In addition, it was found that CDK18 protein expression may predict breast cancer disease progression and response to chemotherapy (53). Recently, CDK18 was found to accelerate oligodendrocyte precursor cell differentiation, while their proliferation and apoptosis were unchanged; this was performed by CDK18 activating the RAS/mitogen-activated protein kinase 1 (54). PCTAIRE-3, as well as PCTAIRE-1 have been implicated in vesicular transport by interacting with Sec23Ap. Disruption of PCTAIRE kinases results in massive alterations of the early secretory pathway, which implies that these kinases are important for the regulation of COPII function and ER-to-Golgi traffic (46). PCTAIRE-3 and PCTAIRE-2 have been linked to Alzheimer's disease (55, 56).

With further respect to oncological relevance, CTS-1 (Chimeric tumor suppressor-1, p53-derived synthetic tumor suppressor) was shown to be able to induce CDK18 expression which subsequently effected growth arrest as well as apoptosis in glioma cells (57). It was also found to have the potential to phosphorylate retinoblastoma tumorsupressor protein (Rb) *in vitro* (50). PCTAIRE-1 has been recently researched in various cancer cell lines using siRNA experiments. It has been shown to regulate p27 (cdk inhibitor, a cell cycle inhibitor protein) stability, and its knockdown lowers cancer cell proliferation and favours cell death in prostate, breast, cervical cancers, and melanoma cell lines, so that PCTAIRE1 controlling p27 can potentially be a common mechanism in cancers. Artificial inhibition of PCTAIRE1 in cancers which overexpress it may be a sensible line of treatment. In the same study, using Oncomine (mRNA expression database) PCTAIRE1 was revealed to be one of the most upregulated genes in various cancers in comparison with corresponding normal tissues (48).

Ultimately, CDK18 was selected as the most promising candidate gene based on its overexpression in ccRCC tissue, antibody quality and published physiological functions. The combination of its functional versatility and documented oncological relevance of its family members made it an attractive target for CRISPR/Cas9-induced knockout generation in renal cancer cell lines. CDK18 knockouts were used in functional assays where their properties and behaviour were compared to the respective wild type clones. Conversely, the use of CRISPR/Cas9 system facilitated the analysis of the potential roles of CDK18 in cell lines, overcoming the transient nature of the siRNA approach. Optimization of the knockout generation workflow may

provide increased efficiency and speed while lowering labor intensity for eventual future applications in A498 and 786-O lines.

4. Methodology

4.1. Patients and samples

The study was conducted in compliance with the declaration of Helsinki and written informed consent was obtained. Tumor samples were staged and graded according to the 2002 TNM Classification of Malignant Tumors (TNM) classification and the Fuhrman grading system (58, 59). The ccRCC tissue samples were obtained at the time of partial or radical nephrectomy at the University Hospital Charité in Berlin in 2011. They were frozen in liquid nitrogen directly after surgical resection and stored at -80°C pending RNA extraction. Their source was tumor and matched normal tissue of 3 male patients without diagnosed metastasis (ages: 47-71; tumor stages: 2 x pT1, and pT3; grading: G1, G2, G3). Regarding PAXgene blood samples, they were acquired between 2010 and 2016; they came from 27 individuals and included in total 16 ccRCC samples. Out of those, 10 were non-metastatic (8 male and 2 female patients; median age 70, range 47-84 years; tumor staging: 1x pT1, 2x pT2, 7x pT3; grading: 2x G1, 7x G2, 1x G3) and 6 metastatic: (5 male and 1 female patients; median age 67, range 47-72 years; tumor staging: 6x pT3; grading: 5x G2, 1x G3). On the other side, there were 11 samples without diagnosed cancer, 4 coming from patients suffering from non-cancer kidney illnesses, and 7 healthy volunteers (7 male and 4 female; median age 47, range 29-80 years).

4.2. Bioinformatics analysis

The first stage in gene selection was the analysis of ccRCC expression in TCGA database, followed by the use of Gene Expression Omnibus database (GEO, (60)) and Genotype Tissue Expression database (GTEx, (61)) databases to remove genes present in blood of healthy donors. In order to assess the blood biomarker potential of the candidate genes, meaning their ability to discern ccRCC from normal patients, their expression was evaluated in two phases. Firstly, RT-qPCR was done in ccRCC and normal tissues, and secondly in blood samples of cancer patients compared with non-cancer patients and healthy donors. If the bioinformatics analysis was to be confirmed, higher expression in ccRCC compared to normal tissue would have to be shown; subsequently, at the time of the testing of PAXgene blood samples potentially at least some of the candidate genes would be more highly expressed in PAX blood from cancer patients compared to the healthy. The first database used was the TCGA, which is the largest public resource giving

somatic and germline mutation, gene expression, gene methylation and copy number variation (CNV) data sets, for several thousands of tumor samples. From this database
RNA seq based expression profiles in ccRCC were retrieved and compared with the respective normal tissue profiles, as well as later with blood. In total, data was retrieved coming from 470 ccRCC patients, including 68 samples from matched normal tissue. In certain cases, multiple samples would corresponded to a single patient, and average expression values were calculated.
Out of the total number of 20533 genes in TCGA, blood expression profiles from sources described below were found for 20466 genes. As in the ideal case candidate genes shouldn't
possess wide expression domains, in order to provide a measure of kidney specificity for a gene, Tissue-specific Gene Expression and Regulation database (TiGER, (62)) was used. This database is based on the NCBI EST database (63) for 30 human tissues and features tissue-specific expression profiles for 20,000 UniGenes. Out of 458 genes enriched in kidney, those also expressed in blood and certain organs (liver, prostate and bladder) were deducted, leaving a list of 95 genes conditionally named 'kidney specific'.

The crucial step was the acquiring of normal blood expression profiles, which would enable the consideration of only those genes not present in the blood of healthy individuals. For this purpose, a detailed search for RNA seq expression data from healthy individuals was made in literature and online databases. GEO database holds microarray, next-generation sequencing, and other forms of high-throughput functional genomics data. It was queried using variations of 'blood[Sample Source] AND homo sapiens[Organism] AND high throughput sequencing [Platform Technology Type]' yielding a total of 7 usable datasets together comprising 91 individual blood samples. Additional 376 blood samples were retrieved from GTEx database and one blood sample pooled from five individuals was kindly provided by Dr. Zhao and Dr. Zhang of Pfizer.

RNA seq datasets from normal tissue were also considered in the analysis. This ensured that expression profiles in important organs and those associated with the urological system were taken into account. Nine and 11 samples were obtained for normal liver and bladder respectively from TCGA database; a similar GEO search yielded a small number of samples for kidney, liver and bladder. Finally, RNA seq Atlas (64) provided additional samples for kidney and liver (pooled from multiple donors) (Table 1).

Python was used to process the data from different databases, and also to calculate rpkm values (reads per kilobase million) where necessary, translate gene names and do statistics.
The formula used to calculate rpkm values was: raw count x 1000000/(gene length x library size). BioMart (65) was used to translate gene names, as different name variations for the same genes were used depending on the database. In order to distinguish cancer from matched normal samples

Mann-Whitey U test was used, with statistical significance defined as $p<0.05$. As in some cases genes had multiple isoforms, replicate samples, or duplicate gene names, absolute highest values were taken, with the goal of not underestimating the possible presence in blood.

Databases with respective sample numbers

DATABASE	TISSUE/BLOOD	SAMPLE NUMBER	SOURCE
GEO/GSE53655	blood	6	whole blood/PAXgene
GEO/GSE72509	blood	18	whole blood/PAXgene
GEO/GSE51799	blood	6	whole blood/PAXgene
GEO/GSE51799	blood	16	peripheral blood mononuclear cells
GEO/GSM833103	blood	16	whole blood/PAXgene
GEO/GSM1647922, Personal correspondence	blood	12	whole blood/EDTA, Tempus
Personal correspondence, Pfizer (66)	blood	1, pooled from 5	whole blood/PAXgene
GTEx	blood	376	whole blood
TCGA	liver, normal matched	9	cancer patients
TCGA	bladder, normal matched	11	cancer patients
GEO/GSE69360	kidney	2	adult normal tissue
GEO/GSE69360	liver	2	adult normal tissue
RNA seq Atlas	kidney	1, pooled	normal tissue
RNA seq Atlas	liver	1, pooled	normal tissue
GEO/GSE35178	bladder	1	adult normal tissue

Table 1. Sources of expression profile datasets.

4.3. RNA isolation and RT-qPCR analysis of blood, tissue and cell lines

For homogenization of tissue TissueLyser II (Qiagen, Hilden, Germany) was used. Purification of total tissue RNA (1µg) was done using miRNeasy Kit (Qiagen). As for PAXgene blood samples, total RNA was purified using the PAXgene Blood miRNA Kit (Qiagen). Total RNA concentration was measured by NanoDrop 1000 Spectrometer (Thermo Fisher Scientific Inc., Wilmington, USA) using absorbance assessment at 260 nm and RNA purity was obtained from A260/280 ratios.

Bioanalyzer (Agilent RNA 6000 Nano Kit, Santa Clara, USA) was used to analyze the integrity and size distribution of RNA from tissue and blood, and only samples which had RNA integrity number values equal or above 7 were considered. As the pooling of RNA samples from normal tissues and respectively for cancer ones was done, one normal and one cancer pool were produced. Transcriptor First Strand cDNA Synthesis Kit (Roche Applied Science, Mannheim, Germany) was used to synthesize complementary DNA

with a mix of random hexamer and oligo (dT) primers. In order to assess the quality of cDNA from tissue and PAXgene blood samples, RNA was also isolated and transcribed from the renal cell carcinoma cell line 786-0. Peptidylproline isomerase A (PPIA) (67) was used to normalize RT-qPCR data.

NCBI´s PrimerBlast and Primer3 (Table 2) were used to design primers and this was done in such a way as to cover the maximum number of isoforms. QuantiTect SYBR Green (Qiagen) was used for detection.

Primers were designed within the following criteria: amplicon length 60-150 nt, primer length 18-30 nt, intron spanning (with intron length > 1000 nt), GC content 40-60%. As in case of some genes UPL probes were used, here primers were automatically suggested with a given probe by the online tool Universal Probe Library (UPL, 68) (Roche); for genes which had multiple isoforms common assays were selected.

GENE NAME	PRIMER SEQUENCE 5'-3'	UPL* PROBE NUMBER
ANGPT2-F ANGPT2-R	atcagccaaccaggaaatga aggaccacatgcatcaaacc	58
CCND1-F CCND1-R	gctgtgcatctacaccgaca ttgagcttgttcaccaggag	17
CA9-F CA9-R	cttggaagaaatcgctgagg ttggaagtagcggctgaagt	51
DGCR5-F DGCR5-R	tcttcaaaccacctgaagaaaaa cagggtgtcgctttcacc	18
NDUFA4L2-F NDUFA4L2-R	ccagactgggaaaacaacg catgcccaggcagattaag	51
STC2-F STC2-R	tacctcaagcacgacctgtg gaggtccacgtagggttcg	2
BARX2-F BARX2-R	gcaggatgaaatggaagaaaat cttcagcttcaatctcttctgatg	58
CDK18-F CDK18-R	caccagcttgaagacactgc cctgttcttccctcgtcct	11
CP-F CP-R	gggattattccccacaaagg tgagcctatgtaaaactctccctta	17
CYP2J2-F	gcgcccaaagaactaccc	81

CYP2J2-R	aaaggttcccatatttcttcacaa	
PPP1R3C-F PPP1R3C-R	tctctgcctaatgagctgcac caaagcctcatggccacatc	N/A
NPTX2-F NPTX2-R	aaggacactatgggcgacct cccagcattggacacgtttg	N/A
TMEM45A-F TMEM45A-R	gcggtcaagtctcattctgc ggggatacaggacaaatccaa	N/A
CAV2-F CAV2-R	gcgggaattctctttgccac tgcactgaaggcagaaccat	N/A
FABP6-F FABP6-R	gagagctgtgttgtctgcgt tctccatctcgaacttgccg	N/A
LOX-F LOX-R	ggcggaggaaaactgtctgg cttggtcggctgggtaagaa	N/A
MET-F MET-R	tccgagaatggtcataaatgt tctctgaattagagcgatgttga	4
ESM1-F ESM1-R	acttgctaccgcacagtctc ctgcaatccatcccgaaggt	N/A
FABP7-F FABP7-R	gctacctggaagctgcaccaa acatttcccacctgcctagtg	N/A
FBX017-F FBX017-R	ggagtcatgatctgggtcagca gcccatctccagtagccagag	N/A
GAL3ST1-F GAL3ST1-R	gggctccctgcttcactttga cacggcataggagtacaccag	N/A
NOL3-F NOL3-R	aggagctgctacgctgtgc tagctgcggtcccggtag	N/A

Table 2. Primer sequences. A subset of genes were detected using UPL probes in which case probe numbers are given.

For relative quantification of transcripts, Light Cycler 480 (Roche) was used, with QuantiTec SYBR Green PCR Kit (Qiagen) as previously described (66). For the genes detected with UPL probes, LightCycler 480 Probes Master Kit (Roche) was used. PCR was performed on 96-well plates, and kidney cancer cell line 786-0 and ccRCC tissues served as positive controls.
The optimization of PCR conditions was done where necessary, and the size of PCR products was confirmed by electrophoresis with Bioanalyzer (DNA 1000 Kit, Agilent). The analysis of PCR data was done by qBasePLUS software (Biogazelle NV, Gent, Belgium). Regarding the processing with the qBasePLUS, a division of samples was done in 2 or 3 groups: normal vs. all cancer samples (cancer and metastatic cancer samples in a single group), normal vs. non-metastatic cancer, normal vs. metastatic cancer, and finally non-metastatic cancer vs. metastatic. The calculation of results was done for 100% PCR efficiency and ´unpaired´ experimental design.

For the statistical analysis, GraphPad Prism 6.07 (GraphPad Software, San Diego, USA) and qBasePLUS were used, with the Mann-Whitney U-test. P values <0.05 were taken as statistically significant. Graphs were made in GraphPad Prism using the Mann-Whitney U-test.

4.4. RNA isolation and RT-qPCR analysis of single cell derived clones

Total RNA was isolated from wild type and knockout clones using RNeasy Mini Kit (Qiagen). Cells were lysed after pelleting. RNA concentration and quality were determined by NanoDrop 1000 Spectrometer (Thermo Fisher Scientific, Waltham, U.S.) and bioanalyzer 2100 instrument (Agilent) using RNA 6000 Nano Kit (Agilent). All samples had RIN values higher or equal to 9.6. Five hundred nanograms of RNA were used for cDNA synthesis with RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) using a mix of random hexamer and oligo (dT) primers.

Primers were designed for SYBR Green using NCBI´s PrimerBlast and Primer3 (Table 3). The criteria for primer design were: amplicon length 60-150 nt, primer length 20-21 nt, intron spanning, intron length >1000 nt, GC content 40-60%. Primer pairs for all four genes cover all of their isoforms, have no predicted unintended products and are intron-spanning. SRSF2 gene structure allowed for the intron length of 334 nt, and amplicon length of 270 nt. Primers were synthesized by Biotez (Berlin, Germany).

GENE NAME	PRIMER SEQUENCE 5'-3'
SRSF2-F	CTATGGATGCCATGGACGGG
SRSF2-R	AGGTCGACCGAGATCGAGAA
WDR77-F	TACTCTGGGATACCCGCTGT
WDR77-R	TGGTGTCCACAAGGGAGACT
LTV1-F	TGAACAGCTGACCCTACATGA
LTV1-R	GCGATTGCTGTCCACTTGAA
SOAT1-F	GGTGGTCCATGACTGGCTAT
SOAT1-R	CAGCCAAGGCATATTCGTGT

Table 3. Primer sequences.

The relative quantification of transcripts was performed on the 7500 Real-Time PCR System Detector (Applied Biosystems, Waltham, USA) using the FastStart Universal SYBR Green Master kit (Roche). Normalization of the RT-qPCR data was done using the reference gene glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with primers: FW 5'-AGCCACATCGCTCAGACAC-3', RV- GCCCAATACGACCAAATCC-3'. The predicted size

of PCR products was confirmed on the agarose gel (Figure 5). Graphs were made in Excel and T test was used for statistics.

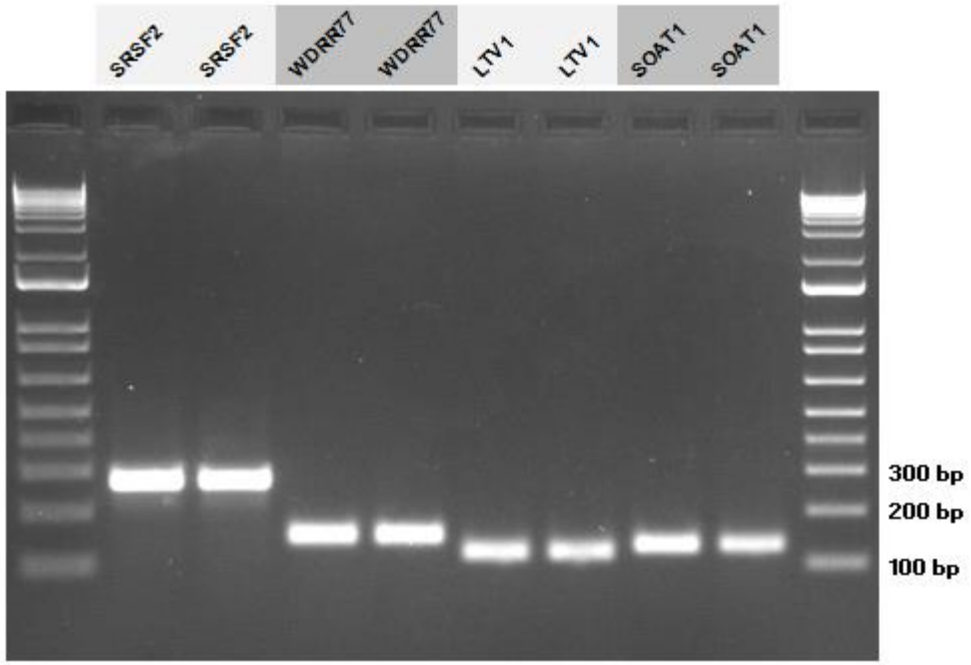


Figure 5. Confirmation of the predicted size of PCR products: SRSF2- 270 bp; WDR77- 150 bp; LTV1- 121 bp; SOAT1- 131 bp.

4.5. Cell culture

Human kidney cancer cell lines 786-O, ACHN, A498, CAKI-1 and a normal line HK-2 (ATCC, Manassas, USA) were used. 786-O cells were maintained in RPMI 1640 (Invitrogen, Waltham, USA), ACHN and A498 in Eagle's Minimum Essential Medium (Biochrom, Berlin, Germany), CAKI-1 in McCoy's 5a Medium Modified and HK-2 in Keratinocyte Serum Free Medium (Thermo Fisher Scientific) supplemented with bovine pituitary extract (BPE) and human recombinant epidermal growth factor (EGF). The media for 786-O, ACHN, A498, CAKI-1 were supplemented with 10% FCS (PAA laboratories, Pasching, Austria), and for all lines with 1% penicillin-streptomycin (PAA Laboratories). Cell lines were grown in a humidified 5% CO2 incubator at 37°C.

4.6. Western blot

For protein extraction, cells were lysed with RIPA buffer (0.5 mM Tris pH 6.8, 1% SDS, 1mM EDTA, 1 mM PMSF, 100 mig/ml Trypsin Inhibitor, 10mig/ml Aprotinin). Sonication was also used to disrupt cellular membranes and release the cell contents. Protein determination was done using Microplate BCA Protein Assay Kit (Thermo Fisher Scientific). Fifteen µg protein samples were separated by 12% SDS-PAGE and transferred onto a polyvinylidene fluoride membrane. Nonspecific binding sites of the nitrocellulose membrane were blocked with 5% non-fat milk in Tris-buffered saline containing 0.1% Tween. Primary antibodies used were BARX2 (sc-53177, Santa Cruz Biotechnology, USA), CDK18 (HPA045429, Atlas Antibodies, Stockholm, Sweden), P4HA1 (HPA007599, Atlas Antibodies), Vinculin (V9131, Sigma-Aldrich), alpha-Tubulin (T6074, Sigma-Aldrich). Membrane was incubated with primary antibodies in following dilutions: CDK18 (1:100), BARX2 (1:200), P4HA1 (1:200). Secondary antibody was HRP-conjugated polyclonal goat anti-rabbit (1:2000, Dako, Glostrup, Denmark). Images were acquired by Odyssey infrared imaging system (LI-COR Biosciences, Lincoln, USA).

4.7. CRISPR/Cas9 approach

The experimental design involved inducing a loss of function by targeting the 1st common translated exon of CDK18, resulting in a high probability of a non-functional protein following frame shift mutations, as the remainder of its sequence containing 15 exons and the catalytic site is located downstream of the Cas9-induced double-strand break (Figure 6). For vectorless delivery of Cas9 enzyme and CDK18-targeting guide RNA, Alt-R CRISPR-Cas9 kit (IDT, Coralville, United States) was used. Ribonucleoprotein complex formation and its lipofection into 786-O and A498 cell lines was performed according to the company protocol and the screening procedure comprized restriction digestion testing, western blot and sequencing (Figure 7). For the design of efficient guide RNA sequences, off-target prediction and knockout screening using restriction digestion, CRISPOR program (http://crispor.org) was consulted. Selected guide RNAs were: sgRNA1 5’ CATTCCGCCGGTTGTGGAGC 3’; sgRNA2 5’ TCAACCAGCTCCACAACCGG 3’ (IDT) and had specificity scores of 92 with none of the off-targets with less than 5 mismatches being adjacent to a PAM site. Selected enzymes for restriction digestion screening were Alu1 and BsrF1 (New England BioLabs, Ipswich, USA) with the latter preferentially used owing to a higher complexity of its corresponding restriction site (RCCGGY). The exact positions of the restriction

sites relative to the PAM sequences were such that the theoretical locations of Cas9-induced cuts fell within them (Figure 8). For PCR amplification of the targeted region, suitable primers (LGC Genomics, Berlin, Germany) were designed using NCBI´s PrimerBlast and Primer3 (FW 5'-TCATGTCCCAAGGGTGTTGG-3', RV 5'- ACCATGGTGCACAGAGGTTG-3'), such that restriction digestion would give clearly distinguishable products on gel.

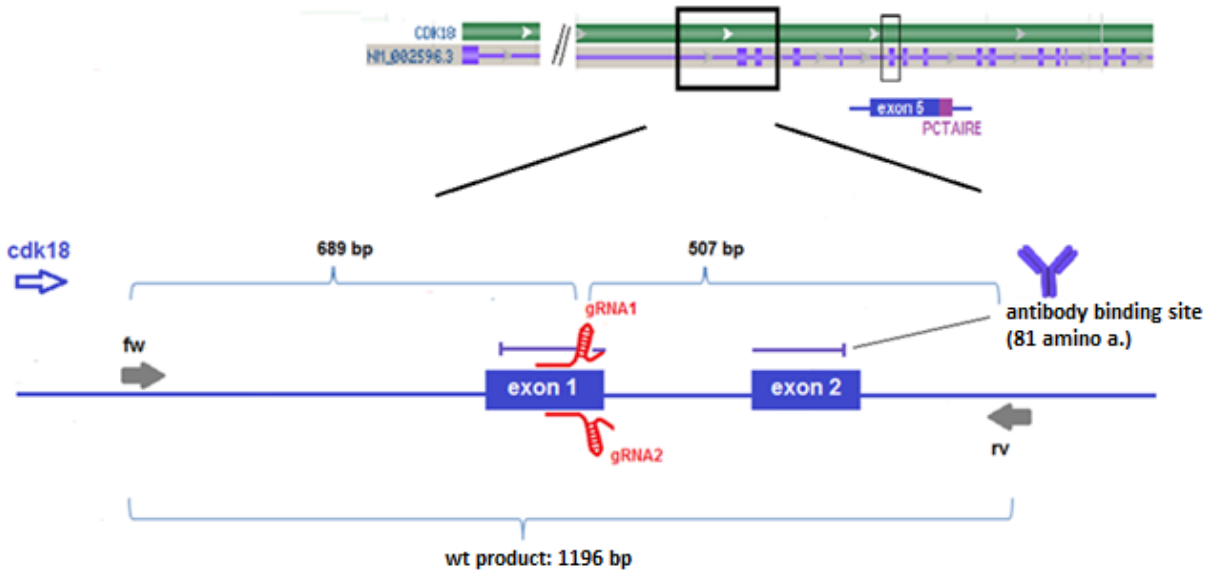


Figure 6. Locations of guide RNA binding sites on the first common translated exon with respect to PCTAIRE sequence, antibody binding site and primer positions.

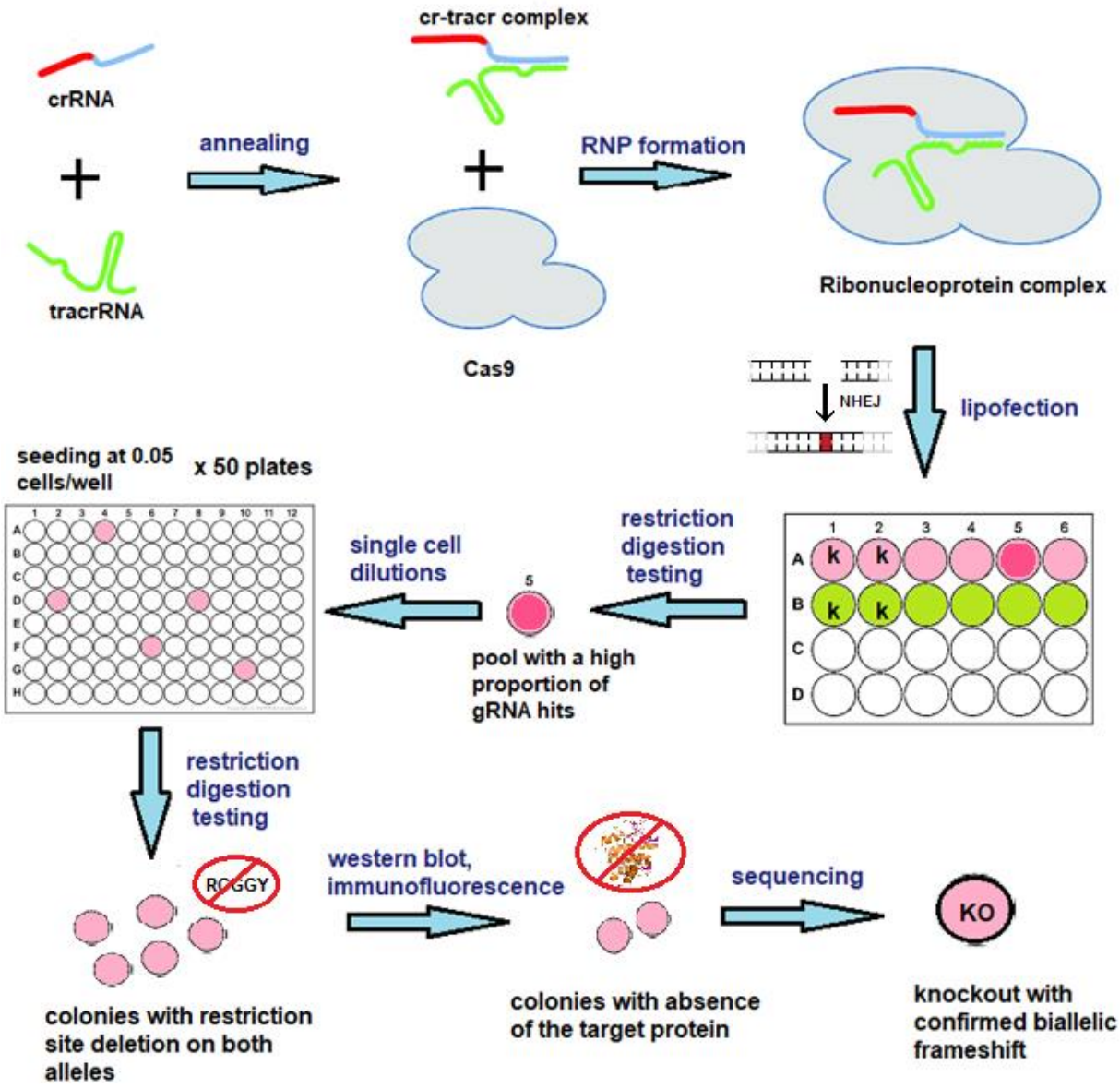


Figure 7. CRISPR/Cas9 knockout approach and screening. crRNA- crispr RNA, tracr - trans-activating crRNA, RNP- ribonucleoprotein, gRNA- guide RNA.

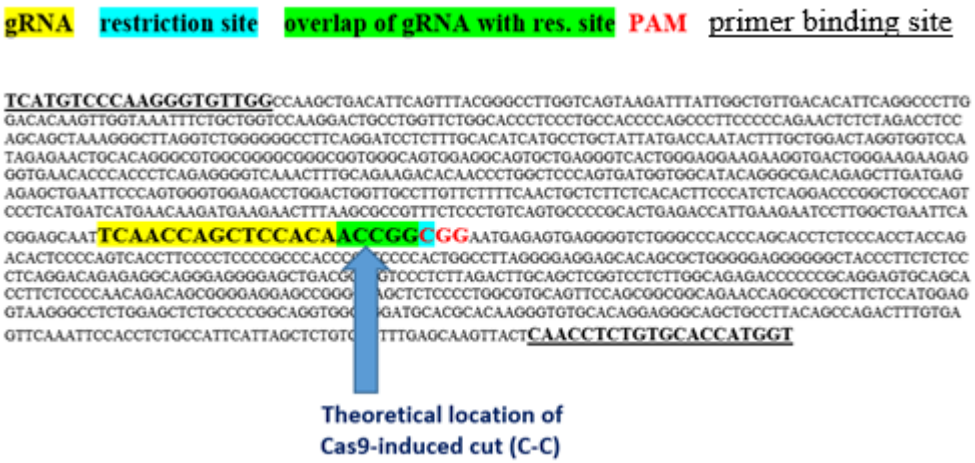


Figure 8. The combination of selected guide RNA and restriction enzyme is such as to ensure a high probability of restriction site alteration.

4.7.1. Transfection

Transfection was performed in 24-well plates using Lipofectamin CRISPRMAX Transfection Reagent (Thermo Fisher) and equimolar concentrations of gRNA and Cas9 enzyme dissolved in 50µl OptiMEM. Treatment was done on attached cells in 100µl OptiMEM and alternatively by resuspending the pellet in transfection solution, for a total of 10-20min after which 100µl of normal medium was added. Cell density was varied between 10^3-10^5, as well as the duration of the transfection treatment and gRNA-Cas9 concentration. Control wells were treated with Lipofectamin CRISPRMAX Transfection Reagent in OptiMEM without Cas9-gRNA complex. 24 hours post-transfection, cells were trypsinized and seeded to a 6-well plate while DNA was extracted and tested for restriction site mutation.

4.7.2. Single cell dilutions

From the pools of cells that had undergone transfection treatments, single cell dilutions were made and subsequently seeded in 50 96-well plates per cell line. To minimize the possibility of multiple cells being seeded in a single well of a 96-well plate, dilutions were made at the ratio of 0.05 cells/well, theoretically yielding ~5 single cell colonies per 1 96-well plate. Colonies were allowed around 3 weeks to grow before being transferred to 24-well plates, at which point the fastest growing colonies were disregarded, presumably stemming from pairs or clumps of cells

and therefore without use for downstream analysis. A-498 line exhibited extremely retarded growth and cell death after a small number of divisions, possibly reflecting its poor tolerance of the transfection treatment or limited propensity for single cell growth. The same dilution approach was followed to generate single cell clones of wild type lines for comparison with knockouts.

4.7.3. Subcloning

After DNA extraction and PCR amplification (Phire Animal Tissue Direct PCR Kit, Thermo Fisher Scientific), PCR products were purified using MSB Spin PCRapace kit (STRATEC Molecular, Berlin, Germany) and subsequently cloned into competent cells (One Shot Top10) using Zero Blunt TOPO PCR cloning kit (Thermo Fisher Scientific) according to the manual. Briefly, 2µl of the cloning reaction was added to a vial of competent cells, and following the 20min incubation on ice (corresponding to the length of the PCR product) the cells were heatshocked for 30s at 42°C. After the addition of S.O.C. medium and 1 hour shaking at 37°C, three volumes from each transformation (10µl, 30µl, 50µl) were spread on prewarmed LB selective plates containing 50µg/mL kanamycin and incubated overnight at 37°C. Depending on transfection efficiency, 10-50 colonies from each cloning reaction were picked and cultured overnight in LB medium. After pelleting, plasmids were isolated using QIAprep Spin Miniprep Kit (Qiagen) and eluted in 50µl water. Plasmid inserts were sequenced (LGC Genomics) using M13FW 5'-TGTAAAACGACGGCCAGT-3' and M13RV 5'-CAGGAAACAGCTATGAC-3'primers and the sequences compared with wild type using Chromas software.

4.8. Functional assays

Cell proliferation assay was performed by measuring metabolic activity of living cells with the Cell Counting Kit-8 (CCK8) (Dojindo, Kumamoto, Japan) according to the manufacturer's protocol. eight hundred cells of wild type stock and each clone were seeded on 96 well plate, in total volume of 100µl, in 4 well replicates for wild type and 8 for CKD18 knockout clones. PBS was put in all outer wells to reduce evaporation. Cells were cultured for 24, 48, and 72 hours. 10µl of reagent was added and after 2h of incubation time at 37°C, absorbance for each well was measured with a plate reader (Berthold Technologies, Oak Ridge, USA) at 450 nm. Three technical repetitions were made. For statistics, 72h absorbance values i.e. 5 data points for wild type clones versus 3 data points for CKD18 knockout clones were compared using the T test. Proliferation curves were made in GraphPad Prism.

Apoptosis was assessed using the Cell Death Detection ELISAPLUS kit (Roche) which quantifies the presence of histone-associated DNA fragments in the cell cytosol, according to the manual supplied by the manufacturer. Briefly 1500 cells of 786-O and A498 original stocks, 5 wild type and 3 CKD18 knockout clones of each line, were seeded in duplicates on 96-well plate and incubated for 24h. After centrifugation to spin down the released histone-associated DNA fragments in the supernatant, supernatant was removed and lysis buffer added. After incubation, 20µl volumes from the supernatant fraction of cell lysates were placed in a streptavidin-coated microplate. A mixture of biotinylated anti-histone antibody and peroxidase-conjugated anti-DNA antibody was added and incubated for 2 h. After removal of unbound antibodies by washing steps, and a 20 min incubation period, the peroxidase activity retained in the immunocomplex was determined photometrically (absorbance at 405 nm with a reference wavelength at 490 nm) with ABTS as substrate. Positive control (DNA-histone complex) produced the absorbance value of 2.1 and background control (incubation buffer) 0.08, which was deducted from all measurements. Three technical repetitions were performed. Absorbance values for wild type and CKD18 knockout clones were compared using the T test, i.e. 5 values for wild type clones versus 3 values for CKD18 knockout clones. Graphs were made in GraphPad Prism.

4.9. RNA sequencing

RNA was isolated from 5 wild type and 3 CKD18 knockout clones of each cell line using RNeasy Mini Kit (Qiagen). Cells were lysed directly on the culture plates. RNA concentration and quality were determined by NanoDrop 1000 Spectrometer (Thermo Fisher Scientific Inc.) and bioanalyzer instrument (Agilent 2100) using Agilent RNA 6000 Nano Kit. All samples had RIN values higher or equal to 9.6. 2µg of RNA of each sample was sent for sequencing to Novogene Company, Hong Kong. Raw counts were used to calculate TPM (transcripts per million) values for each gene, according to the formula: TPM= (length normalized raw count/sum of length normalized raw counts for all genes) x 1million; length normalized raw count= raw count/gene length. For each gene, fold change was calculated as TPM average of CKD18 knockout clones divided by TPM average of wt clones.

5. Results

5.1. Selection of candidate genes

In order to produce a list of genes with biomarker potential, only genes with presumably no blood expression, favorable statistical distance between distributions of cancer and normal values and high expression in cancer were taken considered. Importantly, with respect to blood expression, values below 1 rpkm (Reads Per Kilobase of transcript, per Million mapped reads) were considered low enough as to signify possible non-expression, regarding the sensitivity of detection. The ratio of 5th percentile of cancer distribution to 95th percentile from normal tissue distribution was taken as a measure of distance of one distribution to the other, so that values above 0.5 were considered favorable. A separate measure of distance was calculated where the score represents the multiplication of probabilities that patients from each distribution fall within the overlap interval (score = Xprob x Yprob). Individual probabilities were calculated as the ratio of the number of patients whose rpkm values fall within the overlap interval, and the total number of patients in the distribution (Xprob= patients within the overlap interval/total number of patients). Score was assigned 0 in the case where the distributions do not overlap, and 1 for identical distributions. In cases when one distribution was inside of the other, but there were no patients from the larger one falling into the overlap interval (they were distributed on both sides of it) score was assigned 1, as those genes were not valuable for further analysis. This second method of calculating statistical distance was stricter than the percentile ratio, with favorable distance represented by values less than 0.3.

For genes of interest, the expression levels in liver, bladder, prostate and kidney in healthy individuals were also considered, so that preferential ranking was given to genes with lower rpkm values. In order to obtain information regarding gene function and expression domains of selected genes, literature, the Human Protein Atlas (69), and OMIM (70) were consulted. The following aspects were taken as favorable regarding gene ranking: gene functions related to metabolic pathways in kidney or implicated in cancer (especially genes linked to ccRCC and hypoxia-inducible factors HIF1α and HIF2α), absence of expression in bone marrow and immune system, low or no expression in most tissues, and finally enrichment in kidney.

In total, 20 genes were found to strictly fulfill expression criteria, defined as: blood expression GEO sources 95^{th} percentile <1rpkm, GTEx 95^{th} percentile <=1; fold change TCGA cancer median/matched normal tissue median >1; distribution distance 5^{th} percentile TCGA cancer/95^{th}

percentile matched normal tissue >0.5, TCGA cancer median >5rpkm) (Table 4, first 20 genes). The first 13 genes were characterized by median cancer values above 10. Furthermore, when it is considered that the rate of release of RNA from ccRCC into blood can possibly be much higher than from normal kidney, as well as the potential presence of circulating tumor cells, then the fold change median cancer/matched normal tissue, as well as the percentile ratio distribution distance measure become less relevant and were relaxed in terms of gene selection. A similar argument could be applied for blood expression considering that individual blood sources may not have been fully reliable and false outliers may have been present. This allowed certain genes which did not satisfy all the criteria fully, but have excelled in some (last 11 genes in the table) to be included.

GENE	MEDIAN RPKM in ccRCC (based on expression data from TCGA consortium, (31))	FOLD DIFFERENCE, median rpkm in ccRCC vs. median rpkm in normal kidney (based on expression data from TCGA consortium, (31))	RPKM DISTRIBUTION DISTANCE, 5^{th} percentile ccRCC/ 95^{th} percentile normal kidney (based on expression data from TCGA consortium, (31))	BLOOD RPKM value of 95^{th} percentile (based on GEO database (60))	BLOOD RPKM value of 95^{th} percentile (based on GTEx database (61))
NDUFA4L2	701	145	1.06	0.16	0.15
EGLN3	174	23.2	0.93	0.72	0.62
CA9	117	1218	3.74	1	0.13
CCND1	138	4.34	0.58	0.58	0.37
CAV2	110	3.98	0.69	0.43	0.43
ESM1	92.5	12.4	0.6	0.77	0.2
PPP1R3C	20.1	3.66	0.54	0.03	0.67
STC2	19.1	22.3	1.33	0	0.07
NPTX2	18.4	150	1.48	0.07	0.13
ANGPT2	17.1	10.4	0.54	0.11	0.08
DGCR5	15.2	25.9	0.99	0.3	0.17
DOCK6	13.2	2.08	0.51	0.26	0.68
FABP6	11.9	91.3	1.18	0.62	0.85
TMEM133	8.9	2.67	0.52	0.60	0.24
LZTS1	8.81	5.54	0.52	0.26	0.25
COX4I2	6.85	4.37	0.52	0.06	0.15
KIAA1274	6.71	4.11	0.58	0.78	0.41
LPIN3	6.14	2.29	0.54	0.12	0.17
FKBP9L	6.11	1.32	0.51	0.71	0.14
RAB42	5.90	5.47	1.02	0.17	0.49
MET	112	2.08	0.36	0.07	0.1
CDK18	85.4	5.29	0.28	1.21	0.85
CP	79.7	21.3	0.11	0.89	0.4
TMEM45A	62.9	2.43	0.71	0.13	2.31
LOX	51.0	10.8	0.22	0.22	0.29

GAL3ST1	49.4	8.14	0.3	0.15	0.11
CYP2J2	49	39.3	0.27	0.65	0.2
NOL3	44.1	10.6	1.36	1.98	2.47
FBXO17	38.7	2.77	0.5	2	0.12
FABP7	19.9	974	0.22	0	0.05
BARX2	19.1	6.01	0.27	0.42	0.04

Table 4. Candidate genes. While the first 20 genes strictly fulfil the selection criteria, the last 11 were conditionally selected.

Many of the genes in the table have previously been implicated in ccRCC, largely in micro-array studies (71-78). This analysis identified this group of genes as having zero or low RNA blood presence, suggesting that they could have a potential application as ccRCC biomarkers in blood.

5.2. Expression analysis of candidate genes

In order to get an approximate overview of the tissue levels of selected genes, the expression was analyzed by RT-qPCR in cancer versus normal tissue for 15 genes of highest interest, using PPIA as the reference gene. This confirmed the bioinformatics analysis, as all these genes showed increased levels in cancer, notably so CA9 and NDUFA4L2 (Figure 9). Certain genes had to be excluded from the analysis because of detection issues (multiple isoforms, etc.).

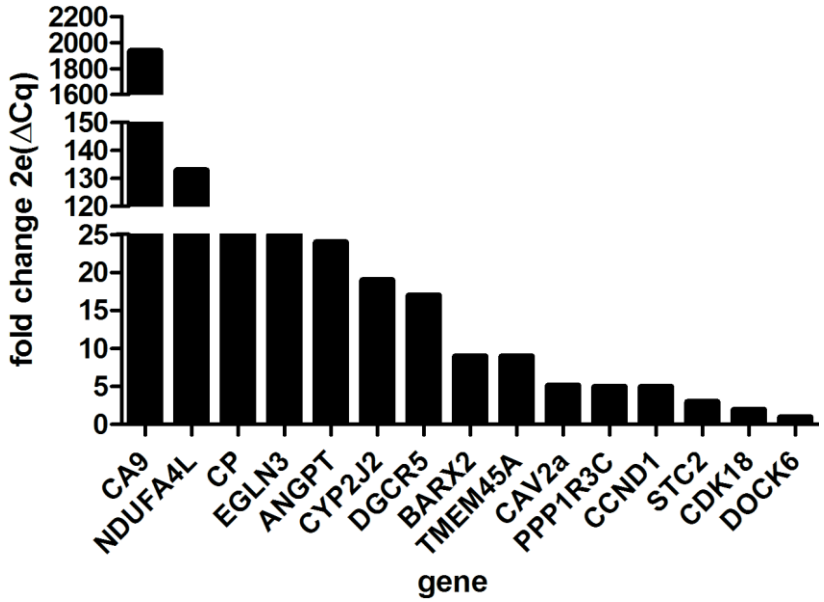


Figure 9. Confirmation of TCGA data by RT-qPCR: Candidate genes were overexpressed in ccRCC compared to normal tissue (all values are above 1). Fold change was calculated as $2^{(Cq\ normal-Cq\ cancer)}$. PPIA was used as the reference gene.

Blood testing was done in two stages: in the first one 3 PAXgene cancer samples were used (Table 5), and only genes with good (Cq <33) detectability were selected for the second stage of blood testing with 24 more PAXgene samples (13 cancer and 11 normal/healthy), so as to broadly evaluate their expression. All samples had RNA integrity number values equal or above 7. Especially good detectability was revealed for following genes: CDK18 (Cq=27), EGLN3 (Cq=26), TMEM45A (Cq=28), CAV2 (Cq=26). A larger number of genes were undetectable or with excessively high Cq values. The genes which supposedly had highest potential based on bioinformatics and tissue qPCR analysis (NDUFA4L2 and CA9) displayed low detectability with very high Cq values (~34). Nevertheless, one of them (NDUFA4L2) was tested on all 27 samples and was confirmed to be undetectable. In the end, 9 genes were selected for the second stage of testing (CAV2a, FABP7, ESM1, NOL3, LOX, CDK18, EGLN3, TMEM45A, CCND1). In this stage, the expression levels were revealed to be similar in cancer versus normal blood for most genes, except for CDK18 and CCND1, which paradoxically turned out to be downregulated in cancer blood (Table 6). Further testing, using 10 plasma samples indicated non-measurable expression (data not shown). No correlation was found between the expression levels in blood for CDK18, CCND1 and LOX, and patient data such as age, tumor grade and stage (data not shown).

CANDIDATE GENES	CONFIRMED HIGHER EXPRESSION IN ccRCC VS NORMAL TISSUE; fold change cancer/normal		DETECTABILITY IN 3 PAX BLOOD ccRCC SAMPLES, mean Cq value
NDUFA4L2	yes	133	>33
EGLN3	yes	25	26.1
CA9	yes	1938	>33
CCND1	yes	5	31.5
CAV2	not tested		26.1
ESM1	not tested		30.6
PPP1R3C	yes	5	>33
STC2	yes	3	>33
NPTX2	yes	24	>33
ANGPT2	yes	24	>33
DOCK6	yes	1	>33
FABP6	not tested		>33
MET	not tested		>33
CDK18	yes	2	27.6

CP	yes	40	>33
TMEM45A	yes	9	28.8
LOX	not tested		30.7
GAL3ST1	not tested		>33
CYP2J2	yes	19	>33
NOL3	not tested		29.8
FABP7	not tested		32.5
BARX2	yes	9	>33

Table 5. Evaluation of candidate genes by RT-qPCR in tissue and 3 blood samples. PPIA was used as the reference gene for tissue.

CANDIDATE GENES	MEAN Cq CANCER	MEAN Cq NORMAL	FOLD CHANGE CANCER/NORMAL	P VALUE	SIGNIFICANTLY DIFFERENT EXPRESSION OF ccRCC vs. NORMAL IN 27 PAX BLOOD SAMPLES
EGLN3	26.5	26.2	-1.20	0.215	no
CCND1	30.6	30	-1.36	0.039	downregulated in ccRCC
CAV2	26.8	27	1.3	0.497	no
ESM1	29.7	29.4	-1.04	0.668	no
CDK18	27.7	26.6	-1.87	0.001	downregulated in ccRCC
TMEM45A	28.5	28.4	-1.07	0.734	no
LOX	30.5	30.1	-1.13	0.668	no, tendency towards upregulation in metastatic ccRCC
NOL3	29.5	29.4	-1.07	0.641	no
FABP7	32.4	32.2	-1.01	0.671	no

Table 6. Evaluation of expression in the second stage of blood testing with 27 samples. P value was calculated using the Mann-Whitney U-test.

The downregulation of mRNA of CDK18 mRNA in ccRCC blood (metastatic and non-metastatic grouped together) compared to normal blood was significant with P value= 0.001 (Figure 10a), whereas CCND1 mRNA was downregulated with P= 0.039 (Figure 10c). For both genes, no significant difference in levels was found when non-metastatic and metastatic samples were compared to each other (Figure 10b,d). In addition, the results showed a tendency towards upregulation for LOX when non-metastatic were compared to metastatic cancer samples, with the P value being very close to significant (P= 0.058) (Figure 10f).

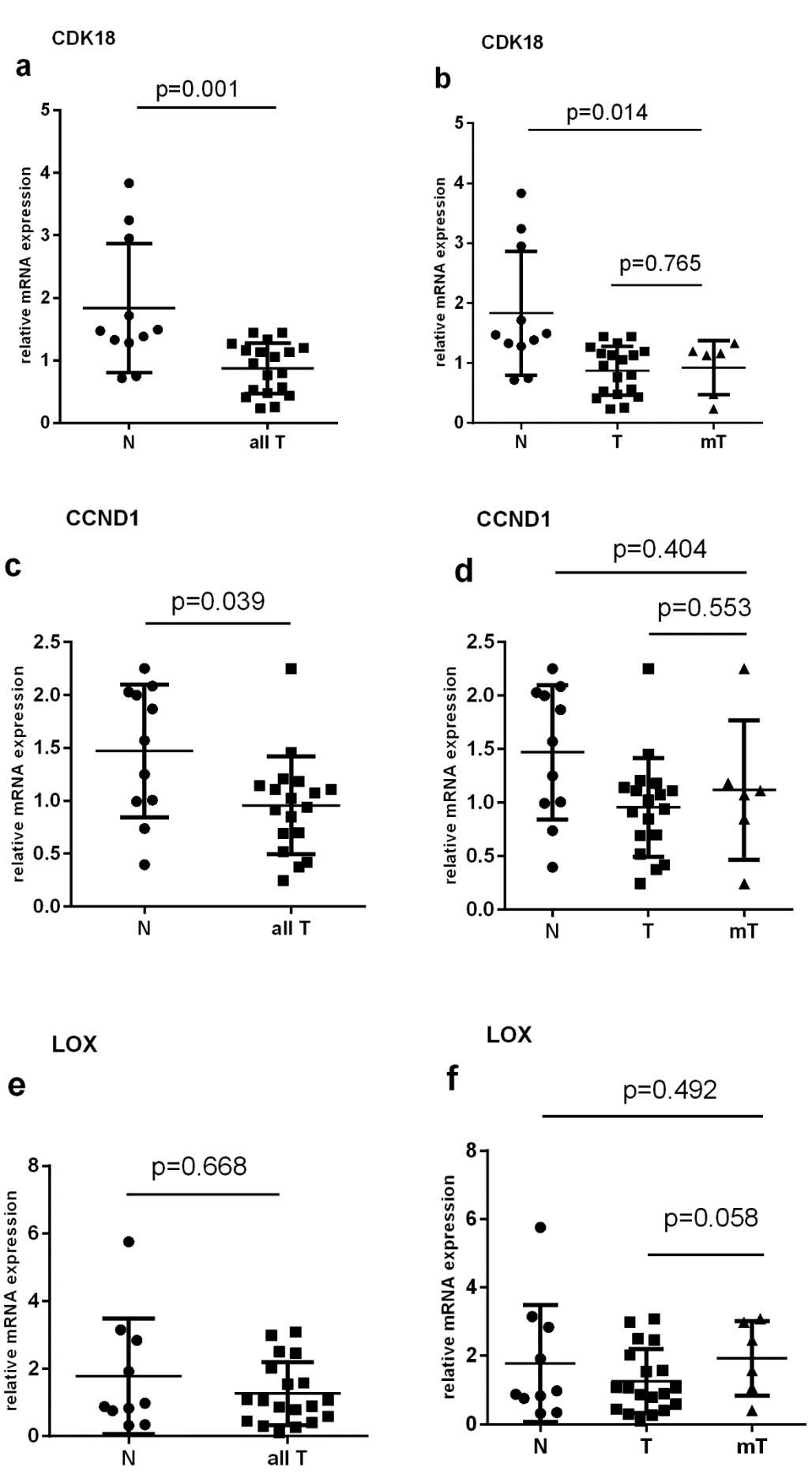

CDK18
a
p=0.001
relative mRNA expression
0
1
2
3
4
5
N
all T
CDK18
b
p=0.014
p=0.765
relative mRNA expression
0
1
2
3
4
5
N
T
mT
CCND1
c
p=0.039
relative mRNA expression
0.0
0.5
1.0
1.5
2.0
2.5
N
all T
CCND1
d
p=0.404
p=0.553
relative mRNA expression
0.0
0.5
1.0
1.5
2.0
2.5
N
T
mT
LOX
e
p=0.668
relative mRNA expression
0
2
4
6
8
N
all T
LOX
f
p=0.492
p=0.058
relative mRNA expression
0
2
4
6
8
N
T
mT

Figure 10. Blood relative mRNA expression of CDK18, CCND1, and LOX based on the qBase exported relative quantity (RQ) values, calculated from Cq values, according to the formula: RQ= $2^{(meanCq-Cq)}$; results from qBase (RQ values) were processed in GraphPad Prism in order to generate graphs using the Mann-Whitney U-test. N- normal patient samples; all T- tumor patient samples (metastatic and non-metastatic ccRCC); T- non-metastatic ccRCC; mT- metastatic ccRCC. a) CDK18 was underexpressed in PAX blood tumor samples compared to normal PAX blood; b) There was no significant difference in expression of CDK18 between tumor and metastatic tumor PAX blood samples; c) CCND1 was underexpressed in PAX blood tumor samples compared to normal PAX blood; d) There was no significant difference in expression of CCND1 between tumor and metastatic tumor PAX blood samples; e) There was no significant difference in expression of LOX in PAX blood tumor samples compared to normal PAX blood; f) LOX showed a tendency towards upregulation in metastatic compared to non-metastatic tumor PAX blood samples.

5.3. Gene selection for functional analysis

Starting from the TCGA, a search of literature and consultation of the Protein Atlas produced 11 uninvestigated and potentially relevant genes, out of which 5 (BARX2, CDK18, P4HA1, EHD2 and FBXO17) were tested in ccRCC tissue, and 2 (CDK18 and BARX2) were used for the tissue microarray analysis in order to investigate the level of protein expression in ccRCC compared with adjacent normal kidney tissue. The TMA analysis, which was done by another doctoral student (Ramon Greuter), was not part of this thesis, but is quoted here to substantiate the rationale for the following analyses. The TMA comprized tissue samples from a total of 443 patients, and CDK18 was found to be significantly overexpressed in tumor (consensus value 1.98) compared to normal (consensus value 1.66) with $p< 0.001$ according to Wilcoxon paired T test; 442 samples were found to be suitable for BARX2 which was found to be significantly overexpressed in normal (consensus value 2.48) compared to tumor (consensus value 2.04) with $p< 0.001$ according to the same test (unpublished data from A. Rabien).

5.4. RNA expression in tissue; RNA and protein expression in cell lines

Although an earlier qPCR analysis of 3 pooled ccRCC and matched normal tissues had confirmed tumor overexpression for CDK18, BARX2 and P4HA1, similar assessment of 8

matched samples failed to show the same, as for approximately half of the sample pairs RNA of all three genes was elevated in normal tissue (Figure 12).

Four renal cancer cell lines were used for the analysis of mRNA and protein expression (ACHN, 786-0, A-498 and CAKI-1), the first three being the most highly cited, as well as HK-2 cell line. A-498 cell line corresponds to the clear cell subtype, having mutated VHL (79) and giving xenografts characterized by compact areas of tumour cells with clear cytoplasm, analogous to the classical appearance of ccRCC. 786-0 cell line produces tumors in nude mice featuring poorly differentiated sarcomatoid cells. Both 786-0 and A498 cluster with ccRCC with respect to copy numbers and VHL mutations (80). Surface receptors and high levels of VEGF in these cells also point to the ccRCC phenotype of 786-O cells (81). Caki-1 models metastatic ccRCC, has wild-type VHL and high VEGF levels and gives tumors of clear cell histology in nude mice (81, 82). Histologically, xenografts from ACHN cells seem to be a poorly differentiated mainly sarcomatoid carcinoma. Mutationally, this line clusters with pRCC (80), it expresses VHL and has a c-met polymorphism that is specific for papillary RCC (83, 79). HK-2 is an immortalized proximal tubular cell (PTC) line derived from normal kidney. HK-2 cells keep many functionalities of renal PTCs, and also have the phenotype that corresponds to well differentiated PTCs (84).

CDK18 mRNA was found to be highly expressed in A498 and 786-O renal cancer cell lines, and even more so in HK-2. BARX2 mRNA was found to be poorly expressed or undetectable in most cell lines except A498. P4HA1 mRNA appeared to be highly expressed in CAKI-1 and HK-2 (Figure 13). At the protein level, all 3 genes appeared to be expressed in all cell lines; particularly strong expression was for CDK18 in A498, BARX2 in CAKI-1, and P4HA1 in HK-2 (Figure 14).

a) **b)** **c)**

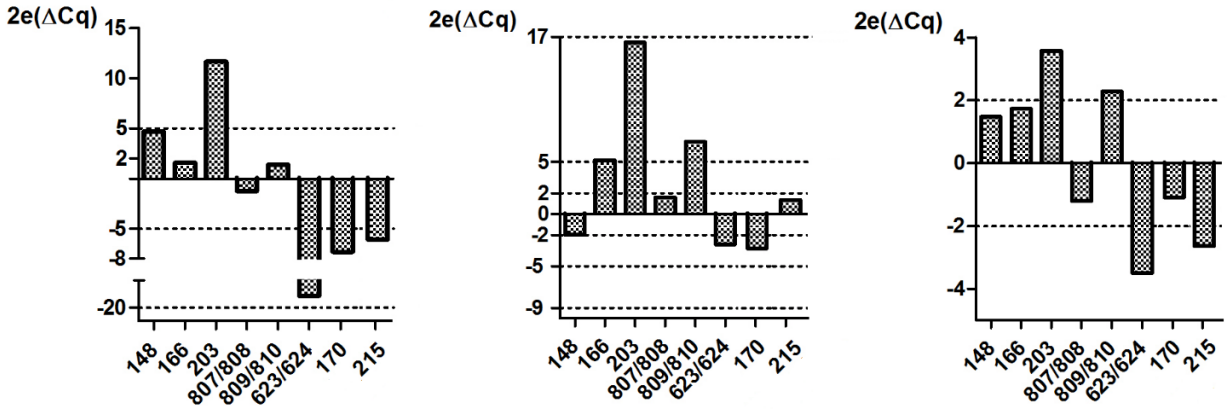


Figure 11. RNA levels (normalized fold change) in tumor compared to matched normal tissue, 8 patient samples: a) CDK18, b) BARX2, c) P4HA1. Fold change was calculated as $2^{(Cq\ normal-Cq\ tumor)}$. Values above X axis signify overexpression in tumor.

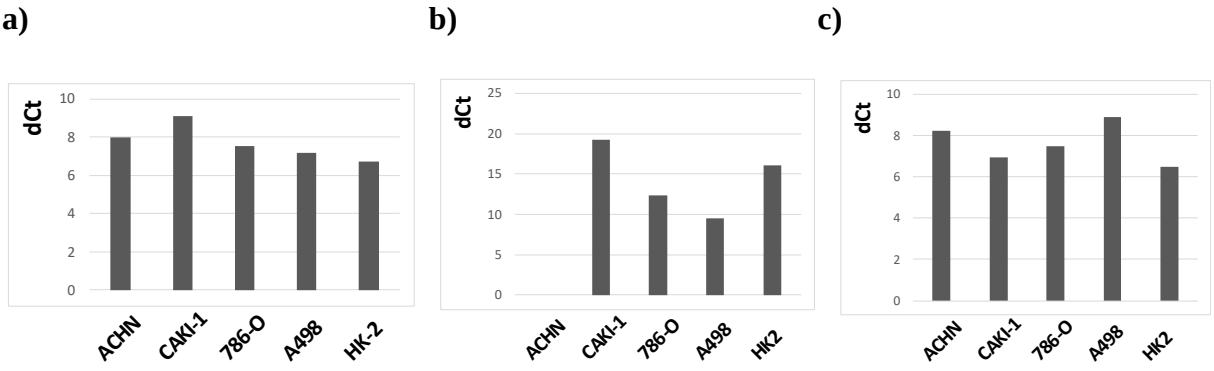


Figure 12. RNA levels (dCt values) of the three genes of interest measured by RT-qPCR in cell lines: a) CDK18, b) BARX2, c) P4HA1. dCt values were calculated as Ct (target gene)- Ct (reference gene); lower dCt levels signify higher expression. PPIA was used as the reference gene.

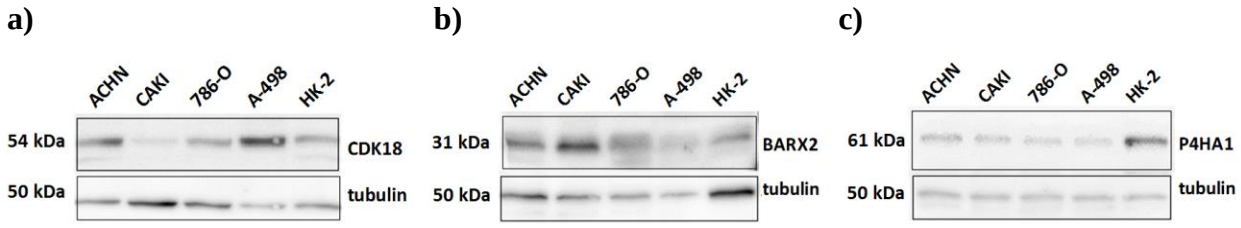


Figure 13. Western blot analysis of protein expression in cell lines: a) CDK18, b) BARX2, c) P4HA1. Bands representing the proteins of interest were detectable in all cell lines tested. CDK18 was particularly highly expressed in the A498 cell line, BARX2 in CAKI-1, and P4HA1 in HK2 cell line. Tubulin was used as the loading control (50 kDa).

5.5. Generation of CKD18 knockout clones

Transfection treatments of the two cell lines 786-O and A488 with two selected guide RNAs yielded high proportions of cells with restriction site deletions, as indicated by the ratio of undigested band to digestion products on agarose gel, in comparison with control (Figure 15). Transfection efficiency did not appear to vary with gRNA concentration, cell density, type of

gRNA, or whether transfection was done in solution by pellet resuspension or on attached cells in OptiMEM (data not shown). Following single cell dilutions, colonies carrying biallelic restriction site deletions (Figure 16) were analyzed by Western blot, and compared with original wild type stocks and single-cell derived wild type clones (Figure 17). Three CKD18 knockout clones were selected from five preselected potential knockout clones depending on the genetic profile and appearance on Western blot.

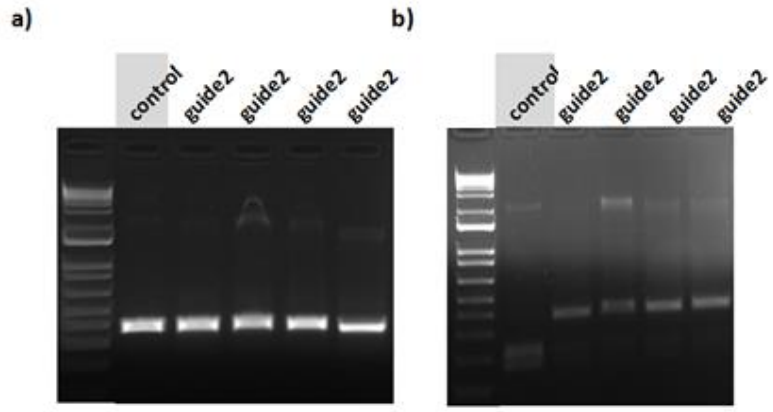


Figure 14. Removal of BsrF1 restriction site after a guide RNA treatment of the original stock of 786-O cells in a high-efficiency transfection:
a) PCR, b) BsrF1 digestion of PCR products. PCR product length is 349 bp and digested products are 198 bp and 151 bp.

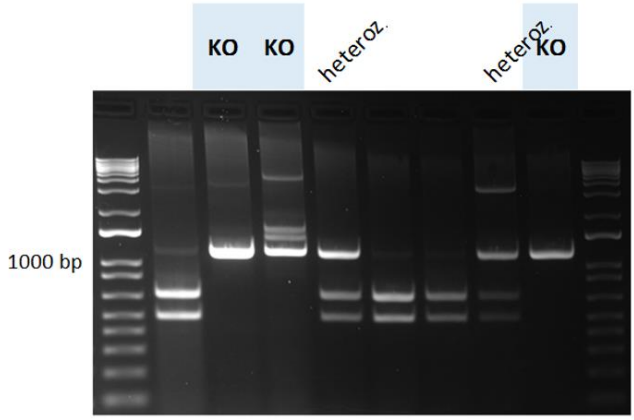


Figure 15. Restriction digestion analysis of single cell originating colonies following transfection reveals biallelic guide RNA hits, marked with 'KO'. Colonies marked 'heterozygous' have one wild type allele (which produces two digestion products) while the restriction site on the other is mutated by the gRNA treatment (undigested band). PCR product length is 1196 bp and digested products are 689 bp and 507 bp (different primers compared with Figure 15).

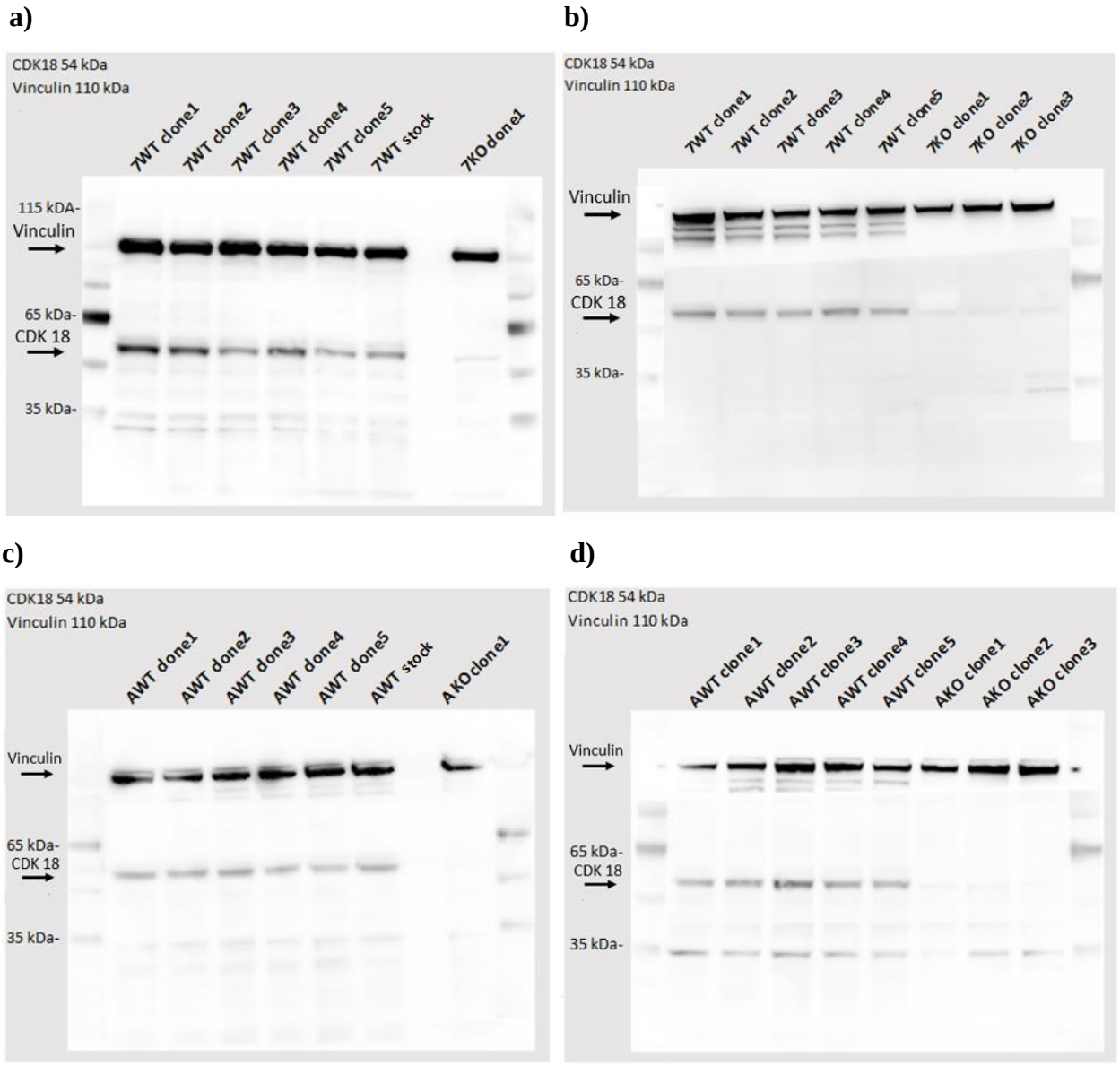
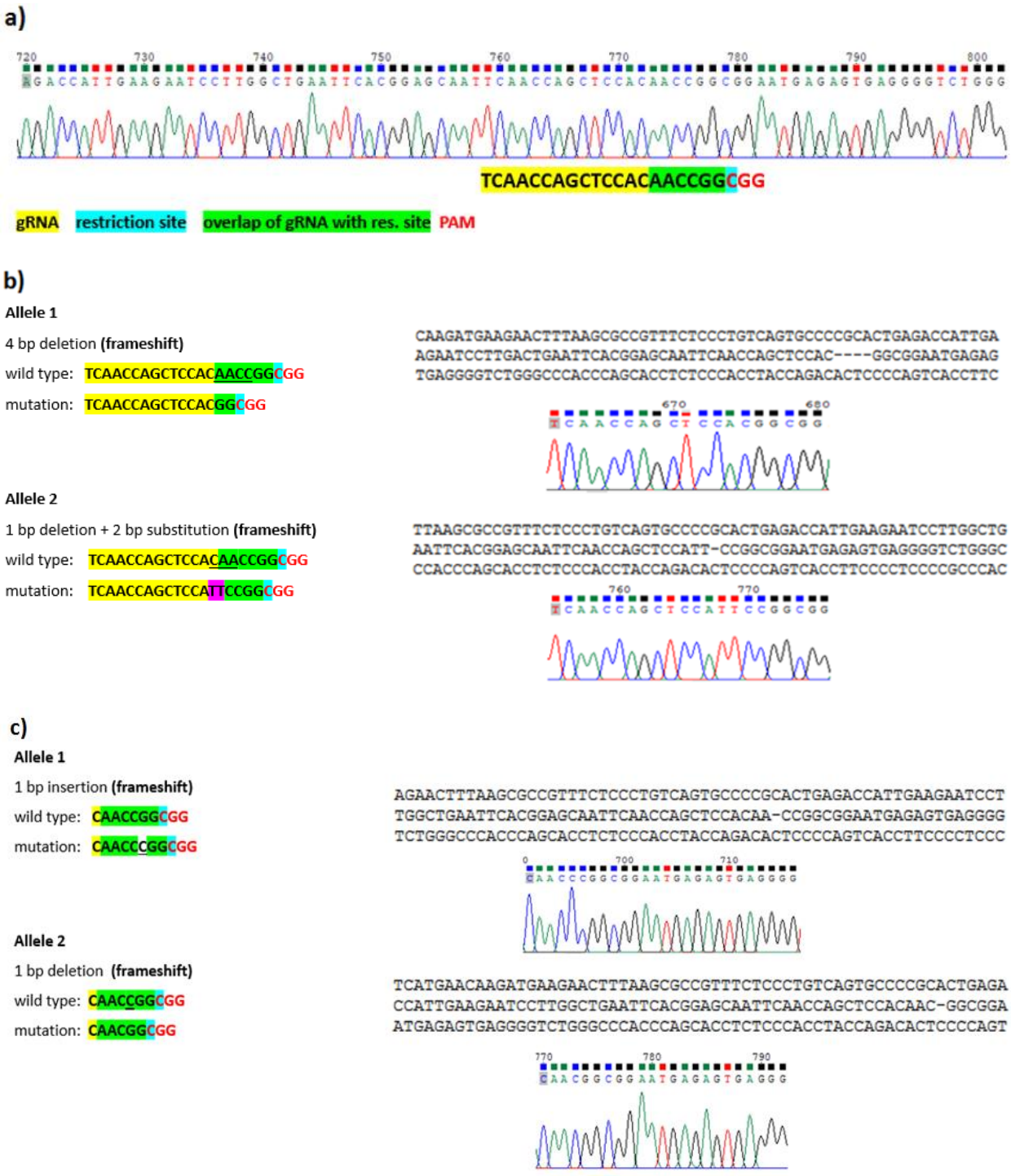


Figure 16. Western blot comparison of CKD18 knockout clones to wild type clones and the original stock of 786-O cell line- a, b; and A498 cell line- c, d. 7WT- wild type clone of 786-O; 7WT pool- original 786-O stock; 7KO- CKD18 knockout clone of A498 line; AWT- wild type clone of A498 line; AWT pool- original A498 stock; AKO- CKD18 knockout clone of A498 line. Vinculin was used as the loading control (110 kDa).

In total, five wild type clones and 3 CKD18 knockout clones of each cell line were generated. Several colonies exhibited presumable deletions of ~200 bp. Such deletions along with other large unintended rearrangements following Cas9-induced double-strand breaks have been documented in various systems (85, 86). Clones carrying these deletions were also selected for downstream analysis. Sequencing of plasmid inserts revealed short deletions and insertions leading to

frameshift, or alternatively larger deletions causing the removal of the splicing donor site. Figure 18 gives a visual presentation of mutated sequences for one of the clones of each line, compared to the wild type sequence, as well as an example of a large deletion. A summary of all Cas9-induced mutations carried by each of the clones is given in Table 7.



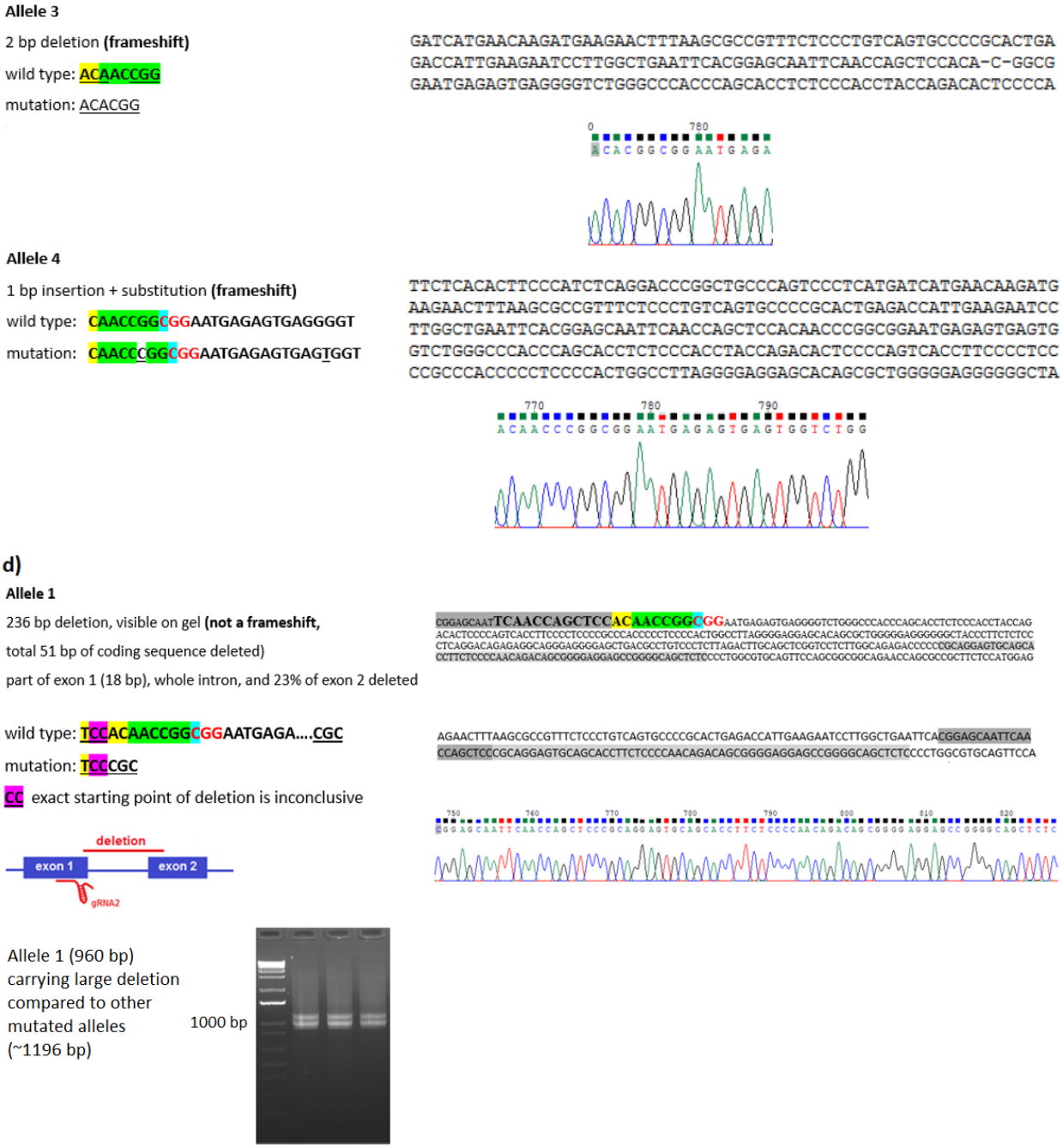


Figure 17. Cas9- induced mutations for one of the CKD18 knockout clones of each cell line as revealed by the DNA sequencing of subcloned DNA fragments, and an example of a large deletion. The mutations were subsequently confirmed by RNA sequencing. On the left, mutated sequences are given in comparison with wild type with an indication of the mutation size and presence of frameshift. On the right, a wider segment of a mutated allele is given, denoting the location and size of the mutation when compared with wild type sequence; on the right below, a snapshot of a short segment of the mutated allele containing the mutation as seen using the sequence viewing Chromas software. a) wild type sequence with the location of the gRNA binding region, PAM

sequence and the restriction site. A successful gRNA hit leads to the eliminaton of the restriction site, which facilitates screening using restriction digestion; a knockout clone contains frameshift mutations on both alleles; b) a bi-allelic frameshift CKD18 knockout clone of the 786-O cell line (7KOclone1) containing a 4 bp deletion on one and 1 bp deletion on the other allele; c) a CKD18 knockout clone of the A498 line carrying frameshift mutations (two 1 bp insertions, a 1 bp deletion, and a 2 bp deletion) on all 4 alleles (ApKOclone3); d) an example of a large deletion removing the entire intron carried by one of the alleles of an A498 clone (ApKOclone4). The allele carrying deletion is distinguishable on the gel from mutated alleles without a large deletion. A graphical representation provides the location of the large deletion with respect to the gene region targeted by the gRNA.

a)

DNA sequencing analysis of 786-O clones

clone	**7KOclone1***	**7pKOclone2***	**7pKOclone3**	**7pKOclone4**	**7KOclone5***
provisional clone status	bi-allelic frameshift	mono- or bi-allelic frameshift	mono-allelic frameshift	potential mix of 2 bi-allelic frameshifted clones	bi-allelic frameshift
total number of sequenced plasmid inserts	8	5	5	7	8
number of empty plasmids	1	0	0	1	2
number of discovered alleles	2	2	2	4	2
mutation 1 (allele 1)	4 bp deletion	35 bp deletion	9 bp deletion	1 bp deletion	26 bp deletion
allele coverage (number of identical sequenced inserts)	5	2	4	1	1
sequence change	wild type: tcaaccagctccacaaccggcgg mutation: tcaaccagctccacggcgg	wild type: gctccacaaccggcggaa…ggccca mutation: gctcca	wild type: caaccggcggaatga mutation: caatga	wild type: aaccggcggaa mutation: aacagcagaa	wild type: cgggagcaattcaaccagctccacaaccggcgc mutation: cggcgg
frameshift	yes	inconclusive	no	yes	yes
intron splicing affected	no	yes	no	no	no
mutation 2 (allele 2)	1 bp deletion + 2 bp substitution	9 bp deletion + substitutions	2 bp deletion + substitutions	4 bp insertion + substitutions	2 bp deletion

allele coverage	2	1	1	1	5
sequence change	wild type: tcaaccagctccacaaccggcgg mutation: tcaaccagctccattccggcgg	wild type: caaccggcggaatgagagtga mutation: caatgagaatga	wild type: tcaaccagctccacaaccg mutation: tcagccagctcattccg	wild type: cagctccasaaccgg mutation: cagcactcacattcacgg	wild type: aaccggcg mutation: aacgcg
frameshift	yes	inconclusive	yes	yes	yes
intron splicing affected	no	yes	no	no	no
mutation 3 (allele 3)	NA	NA	NA	440 bp insertion	NA
allele coverage				2	
sequence change				wild type: caaccggcg mutation: caa/443bp/gcg	
frameshift				yes	
intron splicing affected				no	
mutation 4 (allele 4)	NA	NA	NA	1 bp deletion + substitutions	NA
allele coverage				1	
sequence change				wild type: cagctccacaaccggcggaa mutation: cagatacacaacagcagaa	
frameshift				yes	
intron splicing affected				no	

b)

DNA sequencing analysis of A498 clones

clone	**ApKOclone1***	**ApKOclone2**	**ApKOclone3***	**ApKOclone4***	**ApKOclone5**
provisional clone status	potential mix of 2 clones (2 frameshifted alleles, 2 possibly frameshifted alleles	potential mix of 2 clones (2 frameshifted alleles, 1 possibly frameshifted, 1 not frameshifted)	potential mix of 2 clones (4 frameshifted alleles)	potential mix of 2 possible KOs (1 frameshifted allele, 2 with large deletions)	1 frameshifted allele, others unclear
total number of sequenced plasmid inserts	5	5	46	35	26
number of empty plasmids	1	0	17	17	22
number of discovered alleles	4	4	4	4	1
mutation 1 (allele 1)	1 bp deletion	164 bp deletion	1 bp insertion	236 bp deletion	1 bp deletion

allele coverage (number of identical sequenced inserts)	1	1	9	7	4
sequence change	wild type: caaccg mutation: caacg	wild type: ccacaaccggcggaatga…gag mutation: ccagag	wild type: caaccggcgg mutation: caacccggcgg	wild type: tccacaaccggcggaatgaga…cgc mutation: tcccgc	wild type: aaccggc mutation: aacggc
frameshift	yes	inconclusive	yes	no	yes
intron splicing affected	no	yes, donor splice site and most of intron deleted	no	whole intron deleted	no
mutation 2 (allele 2)	34 bp deletion	2 bp deletion	1 bp deletion	5 bp deletion	NA
allele coverage	1	1	12	3	
sequence change	wild type: accagctccacaaccggcggaatgagagtgaggggtctgg mutation: acctgg	wild type: aaccggcgg mutation: aacgcg	wild type: caaccggcgg mutation: caacggcgg	wild type: tccacaaccgg mutation: tcccgg	
frameshift	inconclusive	yes	yes	yes	
intron splicing affected	yes	no	no	no	
mutation 3 (allele 3)	26 bp deletion	38 bp deletion	2 bp deletion	166 bp deletion	NA
allele coverage	1	2	9	3	
sequence change	wild type: cggagcaattcaaccagctccacaaccggcgg mutation: cggcgg	wild type: gctgaattcacggagcaattcaaccagctccacaaccggcggaa mutation: gctgaa	wild type: acaaccgg mutation: acacgg	wild type: cagctccacaaccggcggaat…aggaca mutation: cag/166bp/aca	
frameshift	yes	yes	yes	inconclusive	
intron splicing affected	no	no	no	yes	
mutation 4 (allele 4)	19 bp deletion	33 bp deletion	1 bp insertion + substitution	4 bp deletion	NA
allele coverage	1	1	1	2	
sequence change	wild type: aaccggcggaatgagagtgaggggt mutation: aacggt	wild type: gagcaattcaaccagctccacaaccggcggaatgagagt mutation: gagagt	wild type: caaccggcggaatgagagtgaggggt mutation: caacccggcggaatgagagtgagtggt	wild type: cacaaccggc mutation: cacggc	
frameshift	inconclusive	no	yes	yes	
intron splicing affected	yes	no	no	no	

Table 7. DNA sequencing analysis of clones carrying restriction site mutations. Clones selected as CKD18 knockouts are marked with an asterisk. a) 786-O cell line; b) A498 cell line. 7KOclone1- CKD18 knockout clone of 786-O cell line; 7pKOclone2- potential CKD18 knockout clone of 786-

O cell line; ApKOclone1- potential CKD18 knockout clone of A498 cell line. Clone name legend for comparison with Figure 17: 7pKOclone2- 7KOclone2; 7KOclone5- 7KOclone3; ApKOclone1- AKOclone1; ApKOclone3- AKOclone2; ApKOclone4- AKOclone3.

5.6. Functional assays

In the proliferation assay, the CKD18 knockout clones of both cell lines showed significantly decreased proliferation over 3 days; the phenotype was stronger for the A498 line with P= 0.0025, while the 786-O line showed P= 0.0387 according to the T test (Figure 19).

In the apoptosis assay using the Cell Death Detection ELISAPLUS kit, for both cell lines, CKD18 knockout clones did not show any significant change in apoptosis over a period of 24 hours compared to wild type clones; for the 786O line P= 0.2558, and for the A498 line P= 0.5881 according to the T test (Figure 20).

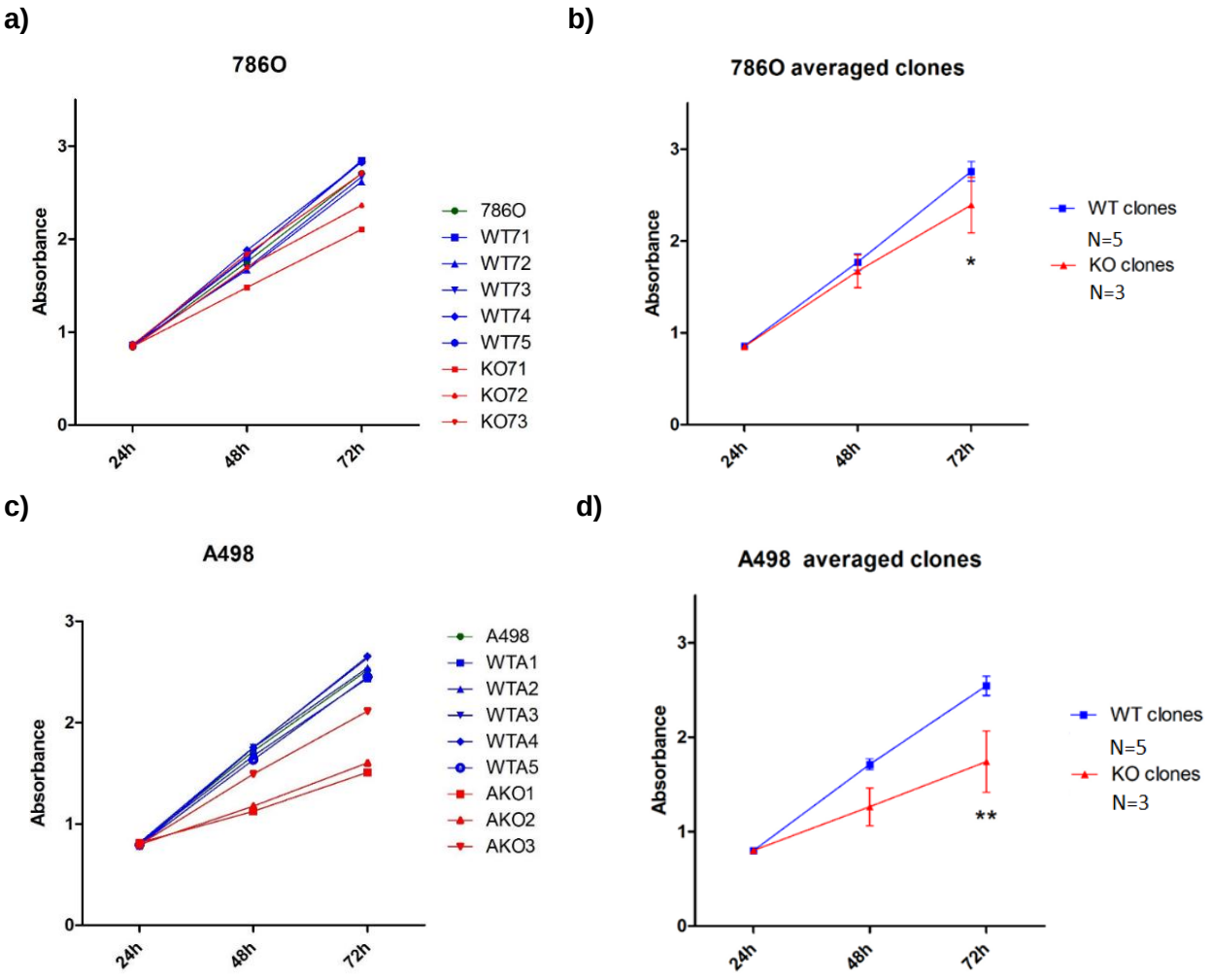


Figure 18. Proliferation of wild type (blue) compared to CKD18 knockout clones (red) measured in the CCK8 assay. CKD18 knockout clones of both cell lines showed significantly decreased proliferation compared with wild type clones, with a stronger phenotype for the A498 line. a) individual clones of 786-O cell line; b) averaged clones of 786-O cell line; c) individual clones of A498 cell line; d) averaged clones of A498 cell line. Graphs a) and c) show averaged values from three replicates for each of the clones separately; graphs b) and d) show averaged values from wild type clones and averaged values from CKD18 knockout clones. Wild type stocks are in green. Statistics was done using the T test, for the 72h time point; *P<0.05, **P<0.01. Error bars signify SD.

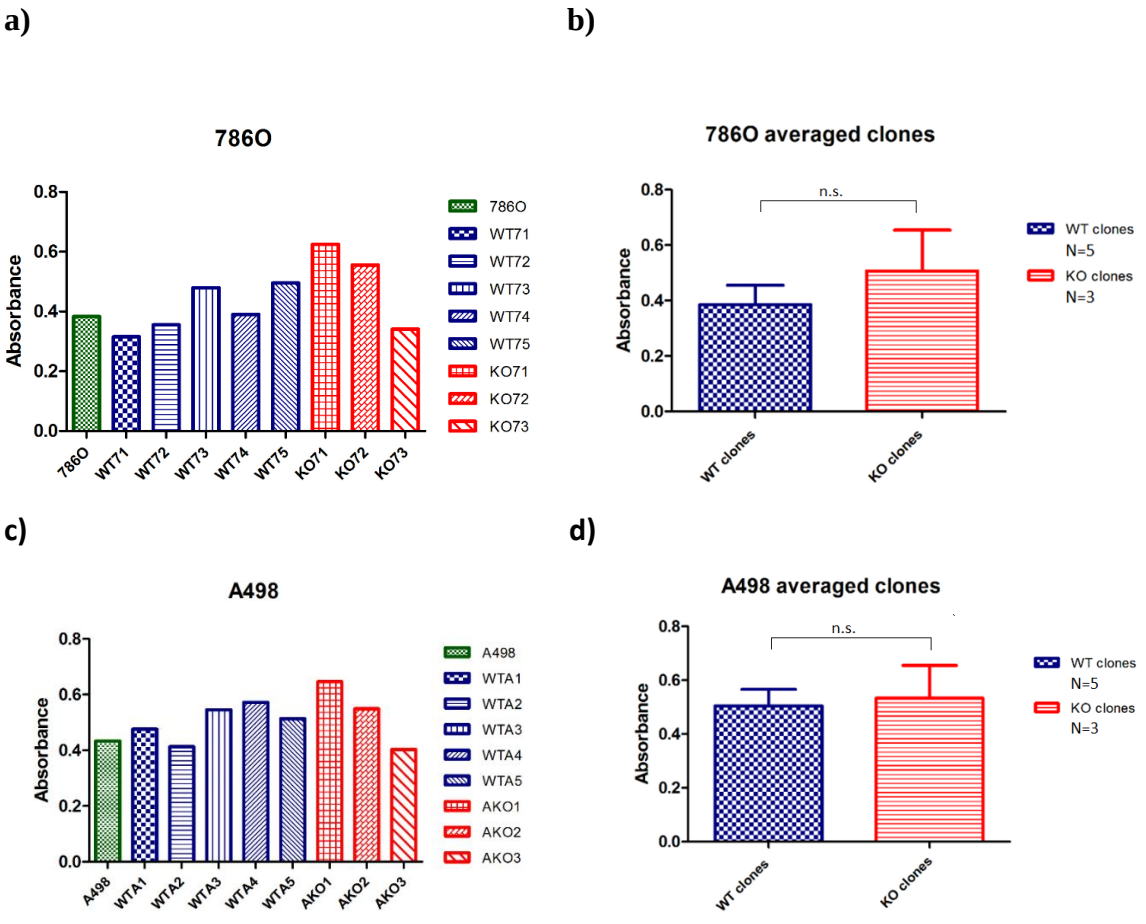


Figure 19. Apoptosis of wild type (blue) compared to CKD18 knockout clones (red) measured using the Cell Death Detection ELISAPLUS kit following the incubation period of 24h. CKD18 knockout clones did not show a significant change in apoptosis compared to wild type clones, for both cell lines. a) individual clones of 786-O cell line; b) averaged clones of 786-O cell line; c) individual clones of A498 cell line; d) averaged clones of A498 cell line. Graphs a) and c) show averaged values from three replicates for each of the clones separately; graphs b) and d) show

averaged values from wild type clones and averaged values from CKD18 knockout clones. Wild type stocks are in green. Statistics was done using the T test, n.s. = not significant. Error bars signify SD.

5.7. Gene selection and qPCR analysis of potentially dysregulated genes based on RNA sequencing

At Novogene, after the reassessment of RNA quality, mRNA was randomly fragmented and cDNA synthesized using random hexamer primers, followed by the second strand synthesis. After adaptor ligation, precise quantification of the concentration of the double-stranded cDNA transcriptome library was done using qPCR, and subsequently libraries were fed into HiSeq Illumina sequencers. Raw data were recorded into a FASTQ file which was the finally supplied by the company.

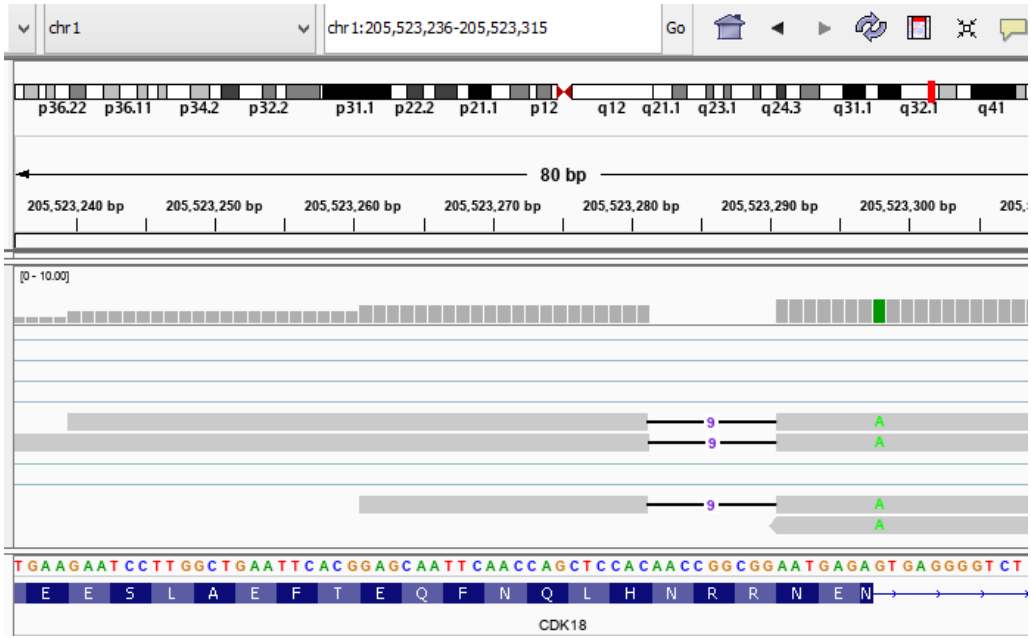
Raw paired-end sequenced reads averaging 150 bp in length were aligned to the reference genome using STAR mapping in bash script. Reads were visually inspected using the sequence data visualization tool Integrative Genomics Viewer (IGV) with comparison to reference genome Human hg38. Average quality score per read (Illumina 1.9 encoding) was around 36 ('F'). The number of reads per clone was between 21.1 and 37.3 million, with alignment quality represented by >92% uniquely mapped reads. The knockouts and wild type clones were compared by the two-tailed T test.

RNA sequencing ultimately revealed Cas9-induced CDK18 mutations, thereby confirming those initially discovered by DNA sequencing after subcloning (Figure 21). Based on the RNA sequencing analysis, preselected were genes with a maximal TPM >10, a variability (maxTPM/minTPM) of more than 1.4 fold, and a significant < -1.25 fold downregulation (for upregulated genes > 1.25 upregulation) in CKD18 knockout vs WT in both cell lines. Out of 33 preselected genes, 23 were downregulated and 10 upregulated (Figure 22). For all preselected genes, a literature search and functional annotation analysis using DAVID (87) was done, focusing on possible functional roles related to cell proliferation, cancer, and ccRCC. Four genes with following signed fold changes CKD18 knockout/wild type were selected for qPCR validation: SRSF2 (-1.38), WDR77 (-1.28), SOAT1 (-1.38), LTV1 (-1.28) (Figure 23).

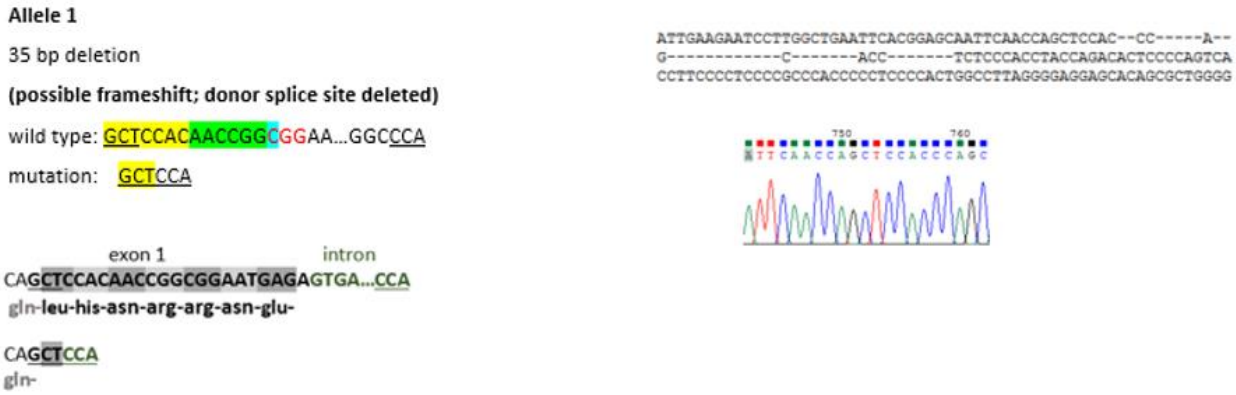
The level of RNA expression of these 4 genes was tested by RT-qPCR in wild type and CKD18 knockout clones of both cell lines. The results confirmed a significant (P= 0.0273), 1.409 fold downregulation of WDR77 in A498 line, whereas SOAT1 showed a tendency (P = 0.0959)

towards 1.268 fold downregulation in the same line, according to the T test (Figure 24). There was no significant dysregulation for the remaining genes.

a)



b)



Allele 2

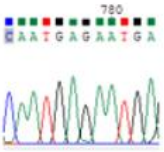
substitutions + 9 bp deletion

(possible frameshift, donor splice site mutated)

wild type: CAACCGGCGGAATGAGAGTGA

mutation: CAATGAGAATGA

```
GATCATGAACAAGATGAAGAACTTTAAGCGCCGTTTCTCCCTGTCAGTGCCCCGCACTGA
GACCATTGAAGAATCCTTGGCTGAATTCACGGAGCAATTCAACCAGCTCCACAAT-G---
-A--GA-A-TGAGGGGTCTGGGCCCACCCAGCACCTCTCCCACCTACCAGACACTCCCCA
```



exon 1 intron

CACAACCGGCGGAATGAGAGTGA

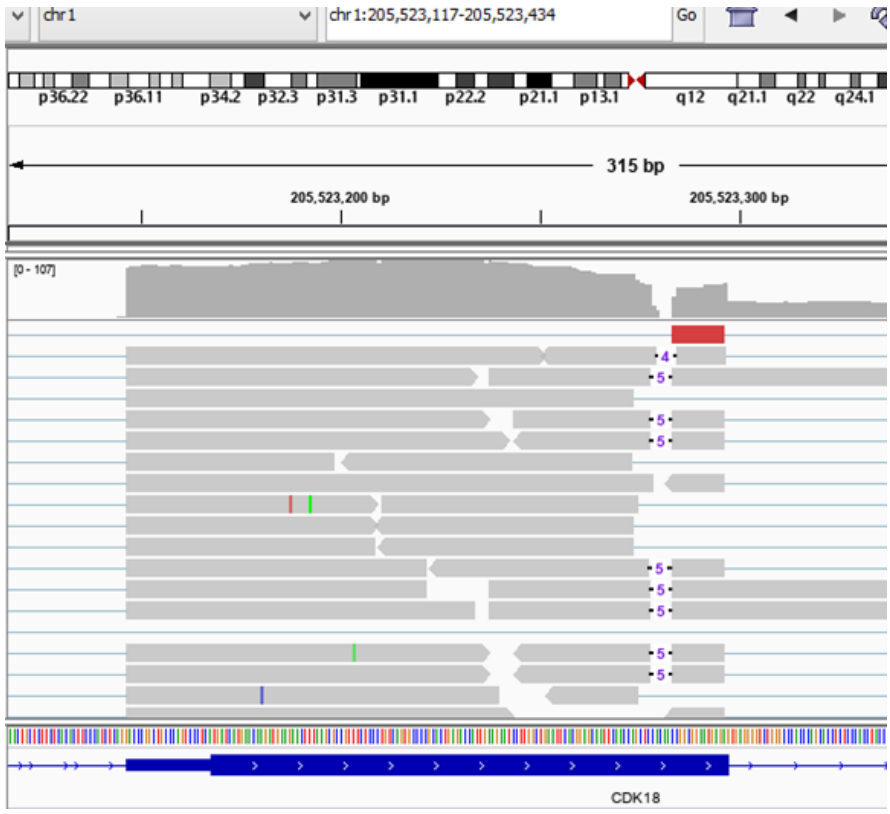
his-asn-arg-arg-asn-glu-

CACAATGAGAATGA

his-asn-glu-

gRNA restriction site overlap of gRNA with res. site PAM

c)



d)

Allele 1

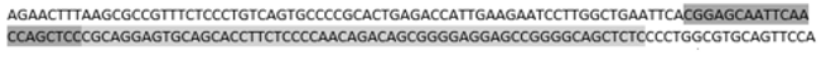
236 bp deletion, visible on gel **(not a frameshift,**

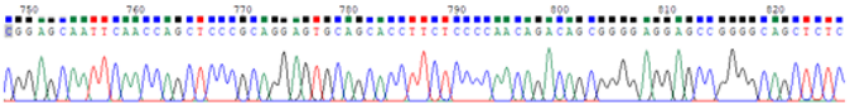
total 51 bp of coding sequence deleted)

wild type: TCCACAACCGGCGGAATGAGA….CGC

mutation: TCCCGC

CC exact starting point of deletion is inconclusive



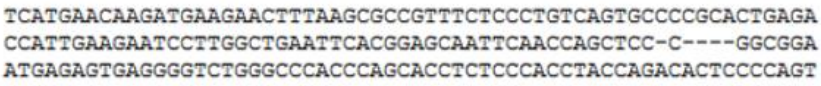


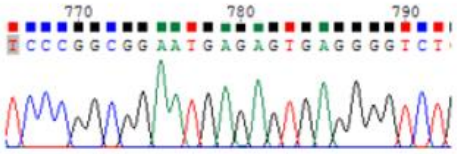
Allele 2

5 bp deletion **(frameshift)**

wild type: TCCACAACCGG

mutation: TCCCGG





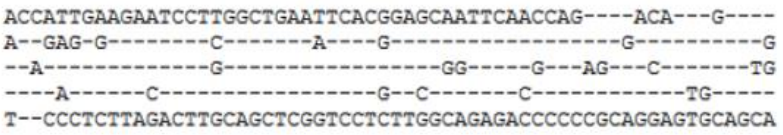
Allele 3

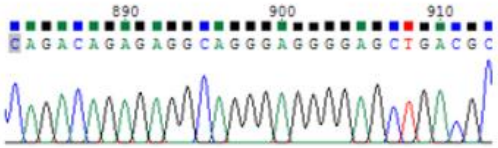
166 bp deletion, almost entire intron deleted

(possible frameshift)

wild type: CAGCTCCACAACCGGCGGAAT…AGGACA

mutation: CAG /166 bp/ ACA



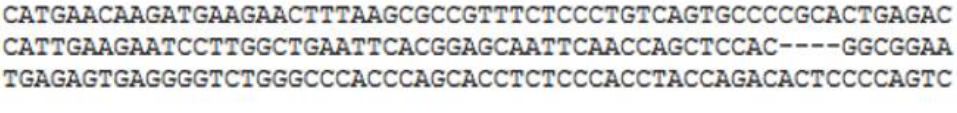


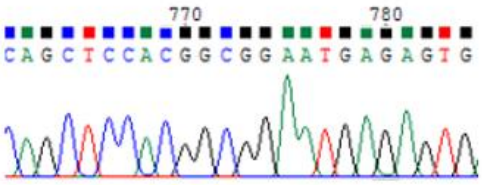
Allele 4

4 bp deletion **(frameshift)**

wild type: CACAACCGGC

mutation: CACGGC





gRNA restriction site overlap of gRNA with res. site PAM

Figure 20. Comparison of RNA sequencing of CDK18 knockout clones with previously found alleles using DNA sequencing: a) CDK18 reads of 7pKOclone2 (786-O cell line); b) DNA-sequenced alleles of 7pKOclone2; c) CDK18 reads of ApKOclone4 (A498 cell line); d) DNA-sequenced alleles of ApKOclone4.

a)

GENE NAME	FOLD CHANGE KO/WT 786-O	FOLD CHANGE KO/WT A498	P VALUE 786-O	P VALUE A498
AL390195.1	0.797	0.737	0.027	0.003
WDR77	0.793	0.757	0.008	0.002
DUSP12	0.794	0.778	0.033	0.009
SOAT1	0.650	0.792	0.014	0.009
TSEN15	0.699	0.765	0.025	0.006
AL359921.1	0.589	0.634	0.043	0.039
HEATR1	0.737	0.736	0.039	0.023
ANKRD39	0.711	0.791	0.047	0.010
MIR1226	0.444	0.522	0.007	0.028
LTV1	0.786	0.760	0.016	0.006
AF241726.2	0.557	0.616	0.044	0.008
AL353763.2	0.573	0.706	0.045	0.008
NUDT15	0.792	0.778	0.040	0.005
AC007610.1	0.593	0.669	0.014	0.026
HEATR3	0.778	0.703	0.010	0.001
MT1E	0.724	0.778	0.047	0.018
SRSF2	0.795	0.639	0.023	0.006
AC011499.1	0.756	0.646	0.013	0.001
RF02202	0.167	0.777	0.011	0.018
LINC01835	0.266	0.777	0.006	0.018
VN1R81P	0.741	0.777	0.025	0.018
ZNF587B	0.711	0.682	0.005	0.030
MIR155HG	0.653	0.433	0.032	0.032

b)

GENE NAME	FOLD CHANGE KO/WT 786-O	FOLD CHANGE KO/WT A498	P VALUE 786-O	P VALUE A498
GREB1	3.427	1.629	0.049	0.030
HIST1H2BC	2.047	2.281	0.003	0.009
CCL26	3.294	1.558	0.022	0.012
EPHB6	2.596	2.172	0.023	0.008
RF02219	2.021	2.143	0.019	0.034
ADAMTS7	7.854	1.592	0.017	0.025
CDH13	2.746	2.551	0.021	0.019
AC113189.2	2.133	1.710	0.017	0.011
SNTA1	1.573	2.264	0.045	0.000
PPM1N	2.163	1.580	0.034	0.002

Table 8. Preselected genes based on the RNA sequencing analysis. a) genes with lower expression levels in the CKD18 knockout condition; b) genes with higher expression levels in the CKD18 knockout condition. P value was calculated using the T test. 'FOLD CHANGE KO/WT 786-O'- calculated fold change based on 3 CKD18 knockout and 5 wild-type clones of 786-O cell line; 'FOLD CHANGE KO/WT A498'- calculated fold change based on 3 CKD18 knockout and 5 wild-type clones of A498 cell line.

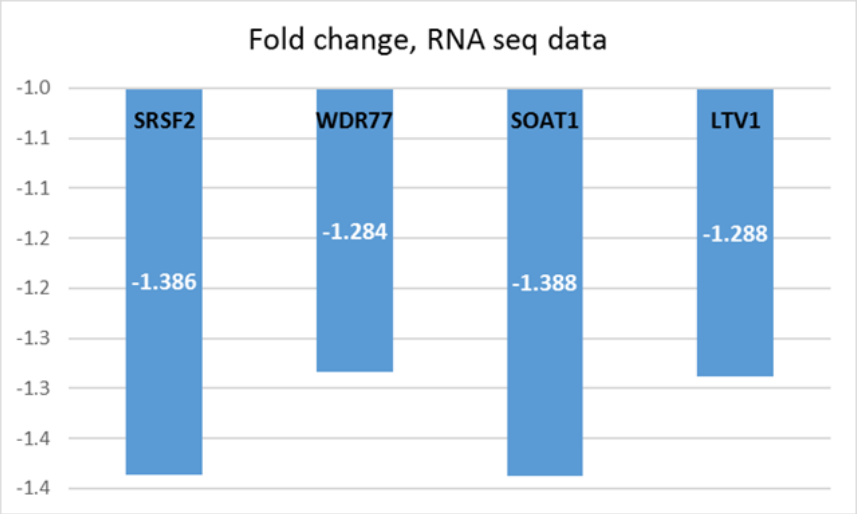
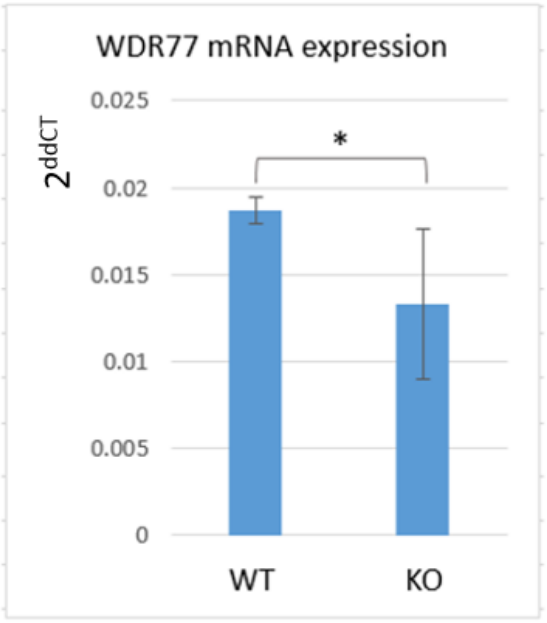


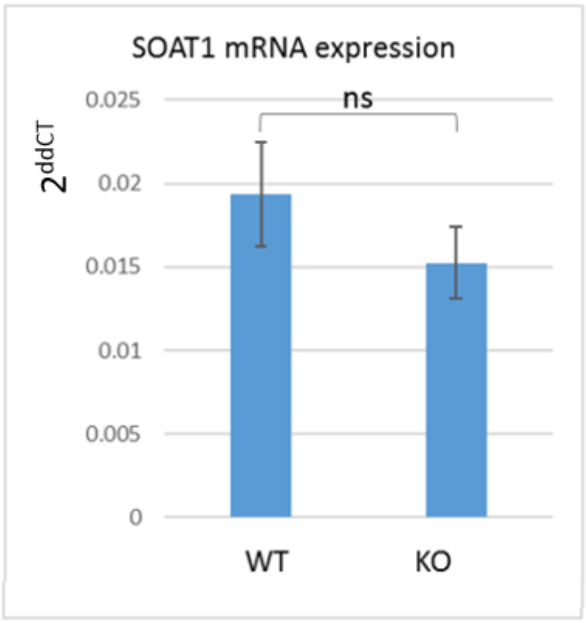
Figure 21. Fold changes of potentially dysregulated genes according to the RNA seq data of Novogene Company. Fold change for a gene was calculated as the average of fold changes for each cell line, which were in turn calculated as TPM average of CKD18 knockout clones divided by TPM average of wild type clones.

a)



N = 5 wt vs 3 ko clones
A498 cell line
Fold change -1.409
P= 0.0273

b)



N = 5 wt vs 3 ko clones
A498 cell line
Fold change -1.268
P= 0.0959

Figure 22. RT-qPCR analysis revealed: a) a significant downregulation of WDR77 in the CKD18 knockout clones of A498 cell line compared to wild type clones, while b) SOAT1 showed a trend towards down-regulation in the CKD18 knockout clones of A498 cell line compared to wild type clones. GAPDH was used as the reference gene. Statistics was done with the T test, *P<0.05, ns = not significant. Error bars signify SD.

6. Discussion

In this study, a gene panel was made comprising the most highly overexpressed genes in ccRCC tissue whose mRNA also had the potential of being absent in the blood of healthy individuals. The first stage in the construction of this panel was the TCGA database- to select a panel of the most overexpressed genes in ccRCC, followed by the GEO and GTEx databases- to deduct from this panel genes showing measurable expression in the blood of healthy individuals. After confirming the tissue overexpression on samples from ccRCC patients in the next step, RT-qPCR analysis was done to evaluate mRNA levels in whole blood of ccRCC patients vs. patients without ccRCC and healthy donors. While RNA transcripts of some of these genes had good detectability in blood, none of the genes were significantly up-regulated in blood from ccRCC patients and two genes paradoxically displayed downregulation (88).

Regarding CDK18 expression in patient tissue, although according to the TMA analysis its protein level was found to be overexpressed in tumor compared to normal adjacent tissue in the sample size of 443 patients, higher levels of its RNA were just found in 4 tumor samples out of the total of 8 tumor and matched normal tissues. This is in clear discrepancy with the TCGA data where for 63 tumor and matched normal tissues, only a small percentage shows lower expression in tumor compared to normal tissue (CDK18- 4.8%; BARX2- 6.4%; P4HA1- 3.2%), whereas the majority shows overexpression. This may be explained by low sample size of the 8 used patient samples.

CRISPR/Cas9 system enabled a loss-of-function analysis of CDK18 using stable knockout clones of two renal cancer cell lines, which revealed its effect on cell proliferation, while RNA sequencing of the CKD18 knockouts pointed to it plausible effectors. CDK18 represents an attractive research topic, being a cyclin-dependent kinase, most of which have well-investigated functions in cancers due to their important cell-cycle related roles, coupled with its functional versatility somewhat owing to its cytoplasmic location. Furthermore, according to TCGA database, out of all CDKs, it is the most highly expressed one and the only one overexpressed in ccRCC, while according to the Protein Atlas, it has by far the highest overexpression in ccRCC compared to other cancers. The recent elucidation of its role in breast cancer (53), as well as the emerging relevance for various tumors of its closely related subfamily member CDK16, adds to its potential importance.

In the following, a detailed discussion of the results as well as limitations of the project are given, firstly for the analysis of the gene panel in blood, and subsequently regarding the functional investigation of CDK18 by the CRISPR/Cas9 system to which the focus of the project was shifted.

6.1. The investigation of expression of selected genes in whole blood does not confirm increased mRNA levels

According to the bioinformatics results, the genes with the highest prospective were NDUFA4L2 and CA9. Based on the TCGA data, NDUFA4L2 has a very high median expression in ccRCC tissue (701 rpkm), while CA9 has the highest fold change in ccRCC compared to normal tissue (1218). Nevertheless, they were both revealed to be undetectable in whole blood by qPCR. Most of the genes that were detectable (EGLN3, CAV2, ESM1, TMEM45A, NOL3, FABP7) did not show significant expression dysregulation when cancer and healthy PAXgene samples were compared. A possible way to overcome this was to test these genes in plasma, as the mRNA levels (which presumably originate from the expression in blood cells) might turn out to be significantly lower in healthy samples compared to cancer ones once the blood cells are not present, thereby showing the real effect of the tumor derived RNA. Nevertheless, the examination of 10 plasma samples revealed that plasma gene expression was below the measurable limit. The PAXgene system was created for stabilization and isolation of mRNA and other nucleic acid species, e.g. genomic DNA or miRNA. For use in this system, blood specimens are collected in tubes containing a stabilization reagent which averts nuclease degradation and transcriptional alterations in anticoagulated whole blood, and is able to keep RNA intact for up to 3 days at room temperature, so that its expression can be analyzed (89). The handling of RNA was done with due care, and even though it can be assumed that for many or all of the candidate genes RNA was degraded by blood RNAses, the RNA integrity of whole PAXgene samples was confirmed to be satisfactory, as shown by their high RIN values. Other than the issues of RNA stability and blood cell expression interference, possible limitations of this study may come from the bioinformatics phase. GEO datasets which were downloaded and used to screen for genes with supposedly no blood expression may not be 100% reliable; they originated from different sources and were not perfectly mutually consistent. In addition, a separate issue is the taken cutoff value of <1 rpkm to mean the absence of blood presence of a gene. Most authors somewhat arbitrarily place the expression threshold at 1 rpkm (and more generally anywhere between 0.3 rpkm and 1 rpkm), below which the sensitivity of RNA sequencing is not high enough to confirm expression and discern it from the background

(90, 91). The cutoff used in this study may potentially have enabled genes with minute expression in blood cells to later be included in the wet lab part of the analysis.

mRNAs upregulated in cancer tissues have been detected in the peripheral blood of patients in case of many different cancers (92). The basis of this part of this thesis was the mentioned mRNA RT-PCR detection assay (29) for prostate cancer, which was designed and validated based on the levels of 5 genes in whole blood. The study comprised 97 men with metastatic prostate cancer, with the assay being prognostic for survival and with increased predictive accuracy when used in combination with CellSearch CTC enumeration. A few other studies with successful tumor-derived mRNA detection carrying biomarker potential can be mentioned. Beta-catenin mRNA, which is linked to the colorectal cancer carcinogenesis, was analyzed in the plasma of cancer patients, correlated to tumor stage and ascertained to have biomarker potential (93). In another study, plasma levels of thymidylate synthase displayed significant difference between cancer and control patients and were related with lymph node metastases as well as advanced colon cancer stages (94). Telomerase RNA was successfully detected in the serum of breast cancer and plasma of prostate cancer, which bears biomarker implications (95, 96) and in a later study by the same group plasma telomerase reverse transcriptase (hTERT) mRNA was linked to poor prostate cancer prognosis (97). More recently, in pancreatic cancer, the serum levels of alternatively spliced type IV collagen mRNA were shown to provide satisfactory discrimination between cancer and benign masses (98). In breast cancer, the plasma presence of cyclin D1 mRNA in patients treated with tamoxifen was able to be correlated with poor prognosis (99). In hepatocellular carcinoma (HCC), several mRNAs have been successfully investigated as potential biomarkers, such as TGF-beta1, LMNB1, and most recently MCM6 (100- 103). More recently, a panel consisting of four circulating mRNAs (FAM129A, MME, KRT7, SOD2) was shown to be differentially expressed in whole blood samples from prostate cancer patients, while the tissue protein expression of individual genes did not always mimic the direction of the dysregulation of the corresponding mRNAs' (104).

Our study was the first to show the downregulation of the two genes CDK18 and CCND1, in ccRCC blood compared to healthy samples, and a tendency towards upregulation for LOX in metastatic compared to non-metastatic ccRCC. These results may be suitable for an additional wider analysis in a larger patient cohort. Cyclin D1 (CCND1) regulates CDK4 or CDK6, whose activity is necessary for G1/S cell cycle transition. It has been investigated using microarray and TMA in ccRCC, and shown to be upregulated and may also serve as a potential therapeutic target (105). In another study, CCND1 was found to be beneficial as an immunohistochemical marker for the purpose of distinguishing between chromophobe renal cell carcinoma and renal

oncocytoma (106). Lysyl oxidase (LOX) does covalent cross linking in elastin and collagen molecules by oxidizing lysine residues, maintaining extracellular matrix integrity (107). LOX is a HIF target (108) and is highly overexpressed in ccRCC in comprison with normal tissue, as well as having prognostic importance for the overall survival of patients with ccRCC (109). Additionally, it has been found to function in a positive loop with HIF-1α in RCC cell cultures, and to alter ccRCC progression by modification of cellular migration and adhesion, as well as alteration of collagen matrix rigidity (110).

6.2. CRISPR/Cas9- induced CKD18 knockout clones

The CRISPR/Cas9 system has been used in cell lines, organoids, and patient-derived xenografts to produce loss-of-function mutations by NHEJ, gain of function (GOF) by HDR and chromosomal rearrangements by the induction of cuts at two different loci (111, 112). In this work, the intention was to use this system in order to create permanent knockout clones of CDK18, and thereby overcome the transient nature of RNAi approach.

Two renal cancer cell lines were selected for this purpose (A498 and 786-O) which are most representative of clear cell renal cell carcinoma according to their genetic profile, surface receptors, and the type of tumors they form in mice (80, 81), as well as having high expression levels of CDK18 according to the Cancer Cell Line Encyclopedia (www.broadinstitute.org/ccle) and own analysis. This work demonstrates the ease and expediency of CRISPR/Cas9 knockout generation in these cell lines and highlights the 786-O line as being more practical with respect to the survival of the single cell dilution treatment, and possibly polyploidy as well. A498 line exhibited difficulty in single cell survival, a high colony death rate, and frequent phenotype loss. It is plausible that these cells survived only in doublets, as very low numbers of colonies were being found following single cell dilution treatments, and ultimately four different alleles were discovered in all of its knockout clones. Alternatively, this can be explained by the specific polyploid locus.

The gene structure of CDK18 allowed for convenient and favorable targeting of the 1st common translated exon, so that the largest part of the protein, containing the functionally crucial catalytic site and PCTAIRE sequence, as well as most of the epitope for the used antibody, were located downstream of the induced double-strand break. The gRNAs were characterized by excellent specificity and no off-targets. These can be chosen using online software with the goal of minimizing the likelihood of target binding and editing (113, 114). For their selection in this work, the online tool CRISPOR was used, whose various aspects such as gRNA activity prediction

(originally based on zebrafish data) and off-target detection algorithms were reviewed here (115) along with other gRNA selection tools. Inclusive to the transfection treatments, factors such as cell density, treatment duration and gRNA-Cas9 concentration were varied, but did not have a considerable effect on transfection efficiency, with apparently same efficiency even for higher cell densities (~10^5/well); nor was there variations between the cell lines.

The specific location of the Cas9-induced break at the 3' end of the targeted exon effected that even shorter deletions led to the removal of the splicing donor site, providing another source of protein disruption and complicating the analysis of frameshift. Clones containing alleles which do not have a frameshift on DNA level, but have a deletion of the splicing donor site were included in the analysis, provided their appearance on Western blot was identical to bi-allelic frameshift clones (Figure 25).

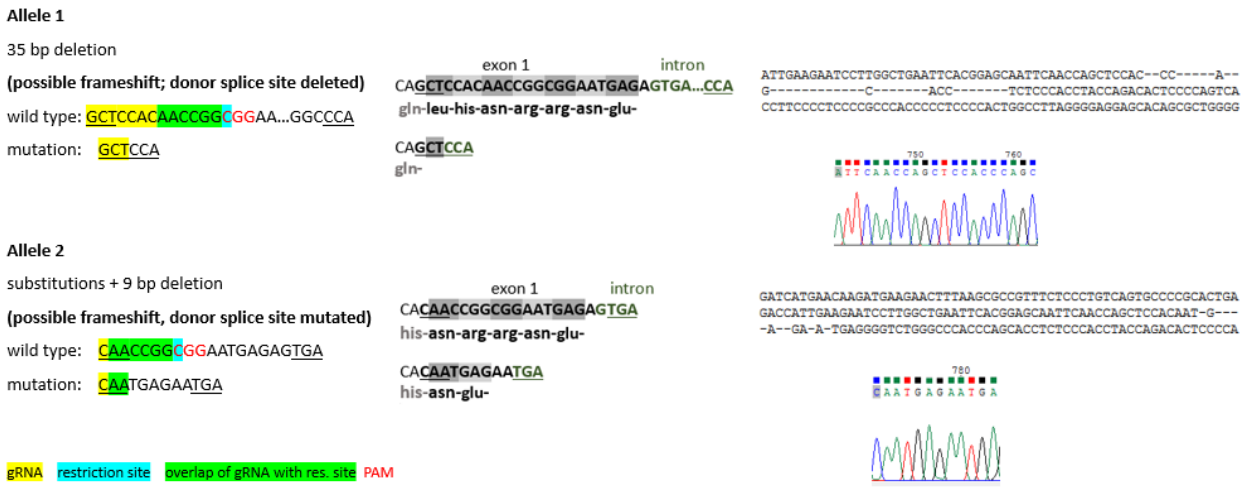


Figure 23. Two alleles of 7pKOclone2 with inframe mutations on DNA level but with possible frameshift mutations on RNA level due to aberrant splicing

The delivery of purified Cas9 and gRNA as ribonucleoproteins into cells is replacing the use of plasmid constructs, due to higher speed and cleavage efficiency, as well as the avoidance of the risk of plasmid fragment being integrated at on- and off-target sites (116, 117). As a method of delivery of Cas9/gRNA ribonucleoprotein complexes, lipid-mediated transfection was found to be appropriate. Although electroporation can be used with satisfactory efficiency, lipid-mediated transfection remains popular due to low cost and straightforward use, but also high throughput adaptation. Specifically, here lipofectamine CRISPRMAX was used, which had been revealed in

a screen of 60 transfection reagents as the best lipid nanoparticle system for delivery into various mammalian cell lines (118), and also has low cell toxicity.

A number of different approaches have been developed in order to perform screening for mutants following Cas9 induced breaks (119- 122), each with their own advantages and limitations. Here, for speed and convenience, restriction enzyme assay was used for genotyping, itself limited by the availability of restriction endonuclease sites, and inability to identify homozygous mutants. The selected combination of gRNA and restriction enzyme was such that the restriction site was deleted as a consequence of Cas9 cleavage. This enables screening by restriction digestion of the target region PCR product. Conversely, transfection treatments expectedly produced dissimilar indels in the two alleles, resulting in scrambled areas upon DNA sequencing. Therefore, subcloning was done for all potential clones to resolve allele sequences and eventually confirm bi-allelic frameshifts by sequencing single-allele originating plasmid inserts.

6.3 RNA-seq and potential function of CDK18 in ccRCC

All wild type and CKD18 knockout clones from both cell lines were used for transcriptome profiling in order to detect any dysregulated genes in the knockout condition. High-throughput (next-generation) technologies have revolutionized genomic research, leading to the possibility of sequencing the whole human genome in a single day (123). RNA-seq is generally viewed as superior in comparison with older technologies, such as microarray hybridization, especially as it has low background, it is more precise, quantifiable, and not limited to existing genomic sequences (124). It has become widely used for studying gene expression and various aspects of RNA metabolism. RNA-seq has also been shown to be highly reproducible and accurate for quantifying exact levels of expression, and represents the first method based on sequencing that permits the whole transcriptome to be analyzed in a high-throughput and quantitative way (125, 126). It also facilitates the study of alternative splicing, post-transcriptional modifications, different mutations and gene expression variations over time (127). New improved RNA-seq methods are expected to arise in the future, especially concerning aspects such as the length of reads, sequencing sensitivity, and the problem of homopolymers (poly(A) tail regions) (128).

Relating to this study, application of the selection criteria across both cell lines to RNA-seq results did not produce any genes with drastic dysregulation, with the highest fold changes (both lines averaged values) for overexpressed genes being ~ 2.7, and for underexpressed ones ~ -1.28. Out of these 33 preselected genes, only 6 genes have been associated with cell proliferation, with

SRSF2, WDR77, SOAT1, LTV1 having strong links to cell proliferation and various cancers (129-140); SRSF2 and SOAT1 have also been previously linked to ccRCC (141- 144).

One limiting factor in the analysis may have been the small number of generated CKD18 knockout clones. With regard to mild dysregulation levels of downstream genes, another limitation of this study is the absence of the data on their protein presence, and here it is presumed that their protein levels are in accordance with the mRNA expression. CDK18 may also be able to influence cell proliferation directly, rather than via regulation of WDR77 and/or SOAT1. Its aforementioned links to DNA replication and breast cancer (53, 54) may be indicative, where it mechanistically interacts with RAD and other proteins, and CDK18-depleted cells seem to show retarded transit through S-phase. Alternatively, direct proliferation effect may also come from CDK18's predominant cytoplasmic localization (51, 53), via effects on cell division machinery components.

This study presents the first investigation of CDK18 in ccRCC, and reveals its effect on cell proliferation and potential relevance for this tumor. Thus far, only limited research on CDK18 has been done, and this contributes to the growing body of information on the importance of this gene for cancer; it also suggests that promotion of proliferation may possibly be a general mechanism of action of CDK18 in cancer. Future prospects of this work would include the analysis of protein expression of WDR77 and SOAT1 in order to confirm their protein downregulation in the CDK18 knockout condition, as well as their manipulation in wild type clones in order to observe any direct effects on cell proliferation. Finally, the specific mechanism of their proliferation phenotype would be elucidated. The follow-up to the proliferation assay that revealed the CDK18 effect would be to perform additional assays (such as using propidium iodide with FACS) in order to investigate which specific phase/phases of cell cycle were affected in the knockout condition. Furthermore, invasion and migration assays may reveal additional differences in behavior of CDK18 knockout clones compared to wild type ones, and thereby additional relevance of CDK18 for ccRCC.

The following gives a description of WDR77 and SOAT1 based on the literature with a special focus on their association with cancer, and potential mechanisms of their effect on proliferation in ccRCC.

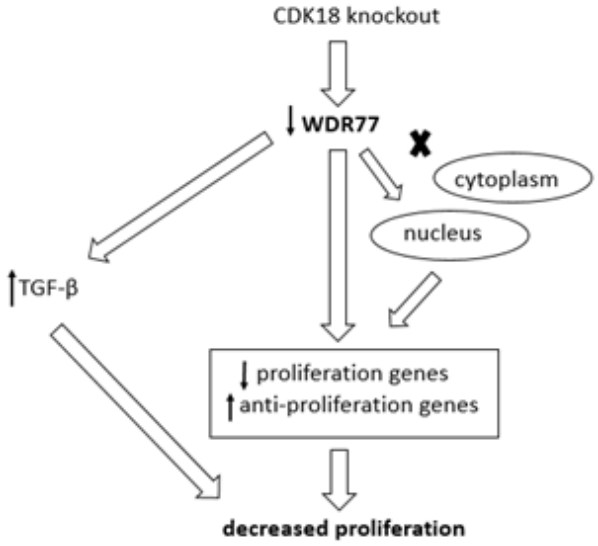
6.3.1. WDR77 and SOAT1

WDR77 (WD repeat domain 77) is an androgen receptor coactivator that forms a stoichiometric complex with PRMT5 (Protein arginine methyltransferase 5), which dimethylates arginine residues, especially in spliceosomal Sm proteins. It also functions as a coactivator of AR mediated transcription via WDR77-AR-Smad1 complex, and affects the expression of certain AR-

target genes in the prostate. It hinders growth in prostate cancer. WDR77 is strongly expressed in the nuclei and weakly in the cytoplasm in benign epithelial cells, whereas in tumor nuclear expression becomes lower and cytoplasmic levels higher. Nuclearly localized WDR77 inhibits PCa growth, whereas cytoplasmic WDR77 favors it and nuclear exclusion of WDR77 is linked to androgen-independent PCa growth (132). WDR77 repositioning from the nucleus to the cytoplasm in prostate cancer cells, as well as the absence of one allele in mouse leads to unrestrained prostate EC division (133). The PRMT5/ WDR77 complex is indispensable for the proliferation of lung and prostate epithelial cells in the early developmental stages (in time of rapid divisions) and is then reactivated during tumorigenesis in the lung and prostate. In lung cancer, PRMT5 and WDR77 enhance cell proliferation via the suppression of genes coding for anti-growth factors, and expression of genes responsible for growth factors that support cell growth (134). WDR77 possibly downregulates the expression of the TGFβ receptor and ligand, therefore effecting a non-response to TGFβ signalling. Silencing WDR77 increased responsiveness to TGFβ signalling which was inversely correlated with lower cell proliferation, while the expression of WDR77 was correlated with higher proliferation and lower TGFβ signalling during tumorigenesis in the lung (135). Circular RNA of this gene (CircWDR77) is up-regulated in high glucose induced VSMCs (vascular smooth muscle cells). CircWDR77 can target FGF-β and in such a way regulate VSMC migration and division by sponging miR-124; when CircWDR77 is artificially silenced it reduces proliferation and migration of VSMCs (145).

In the CDK18 knockout, absence of CDK18, leading to decreased expression of WDR77 may suppress genes involved in proliferation and/or enhance anti-proliferation genes, via transcriptional control or post translational modifications. Also, absence of CDK18 may potentially cause an increase in the nuclear localization of WDR77 (via translocation or transcriptional control) leading to anti-proliferative effects. In addition, downregulation of WDR77 may also lead to increased cellular sensitivity to TGFβ signalling and therefore decreased cell proliferation (Figure 26).

Figure 24. Potential links between CDK18, WDR77 and cell proliferation in CDK18 knockout condition in RCC cell lines.

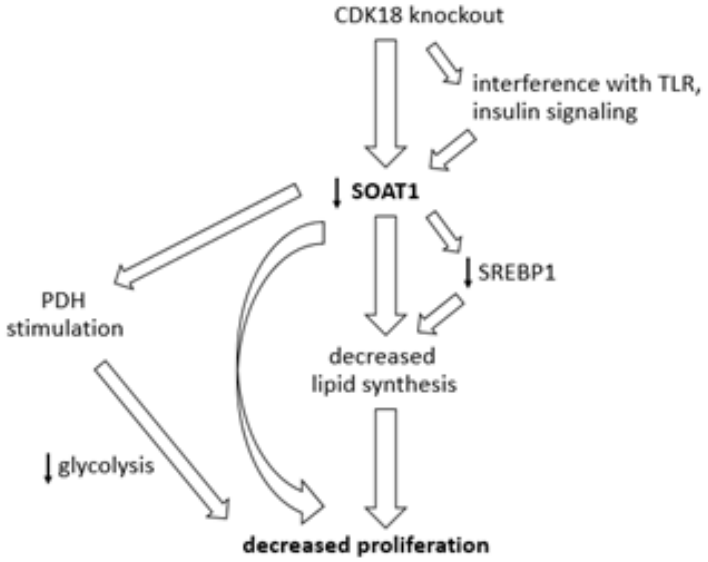


SOAT1 (sterol O-acyltransferase 1) is an enzyme located in the endoplasmic reticulum, which catalyses the formation of fatty acid-cholesterol esters for storage in lipid droplets. It is highly expressed in glioblastoma, and when it is inhibited glioblastoma growth is suppressed and survival extended in xenograft models, which is effected via the downregulation of SREBP1 activation (146). In another study of SOAT1 cholesterol esterification in glioblastoma, it was found that K604, which selectively inhibits SOAT1, suppresses the proliferation of U251 MG cells (glioblastoma cell line) and downregulates Akt and extracellular signal regulated kinase in glioblastoma cells undergoing division (136). SOAT1 is also an upstream mitochondrial lysine acetyl transferase for pyruvate dehydrogenase (PDH). PDH represents a key decision point between glycolysis and oxidative phosphorylation. As the result of oncogenic changes of SOAT1 glycolysis and tumor growth are promoted, whereas targeting SOAT1 with inhibitors derepresses its inhibitory effect on PDH, resulting in anti-cancer effects (147). SOAT1 was found to be dysregulated in a mass spectrometry analysis using 12 matched pairs of ccRCC and adjacent kidney tissues (143). Although the data in TCGA database did not show a dysregulation of SOAT1 RNA in tissue, a study showed that SOAT1 protein expression was 2.9-fold higher in ccRCC than in normal kidney, and its enzymatic activity 5.7-fold higher in ccRCC. Upregulation of SOAT1 and its enhanced enzymatic activity increase cholesterol ester deposition in ccRCC (144). In breast cancer, SOAT1 inhibition reduced proliferation of cancer cells (148). In colorectal cancer cells treated with TLR4 siRNA, cell proliferation, migration and invasion were inhibited (138). Lastly, several studies showed that insulin promotes SOAT1 gene expression at the transcriptional level;

insulin-effected proliferation and metastatic changes of colorectal cancer cells have shown to be mediated by SOAT1 (149).

In the CDK18 knockout, absence of CDK18, causing decreased expression of SOAT1, possibly via interference of TLR or insulin signalling, may lead to decreased lipid synthesis which could negatively affect cell proliferation. Lower SOAT1 levels may decrease lipid synthesis through the downregulation of transcription factor SREBP1. SOAT1 may also influence proliferation in alternative ways, bypassing lipid metabolism, and potentially via stimulation of pyruvate dehydrogenase, causing glycolysis inhibition (Figure 27).

Figure 25. Potential links between CDK18, SOAT1 and cell proliferation in the CDK18 knockout condition in RCC cell lines



In summary, the follow-up to this study would include on the one side the evaluation of the potential of CDK18, CCND1 and LOX mRNAs as ccRCC biomarkers in a larger patient cohort. On the other, as the generation of CDK18 knockout clones in renal cancer cell lines pointed to the effect of this gene on cell proliferation, the next step would be to analyse the specific mechanism of this effect, and the parts that WDR77 and SOAT1 potentially play in it.

7. References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7-34.

2. Srigley JR, Delahunt B, Eble JN, Egevad L, Epstein JI, Grignon D, Hes O, Moch H, Montironi R, Tickoo SK, Zhou M, Argani P, ISUP Renal Tumor Panel. The International Society of Urological Pathology (ISUP) Vancouver Classification of Renal Neoplasia. Am J Surg Pathol. 2013;37(10):1469-89.

3. Rydzanicz M, Wrzesinski T, Bluyssen HA, Wesoly J. Genomics and epigenomics of clear cell renal cell carcinoma: recent developments and potential applications. Cancer Lett. 2013;341(2):111-26.

4. Wettersten HI, Aboud OA, Lara PN, Jr., Weiss RH. Metabolic reprogramming in clear cell renal cell carcinoma. Nat Rev Nephrol. 2017;13(7):410-9.

5. Sato Y, Yoshizato T, Shiraishi Y, Maekawa S, Okuno Y, Kamura T, Shimamura T, Sato-Otsubo A, Nagae G, Suzuki H, Nagata Y, Yoshida K, Kon A, Suzuki Y, Chiba K, Tanaka H, Niida A, Fujimoto A, Tsunoda T, Morikawa T, Maeda D, Kume H, Sugano S, Fukayama M, Aburatani H, Sanada M, Miyano S, Homma Y, Ogawa S. Integrated molecular analysis of clear-cell renal cell carcinoma. Nat Genet. 2013;45(8):860-7.

6. Mucaj V, Shay JE, Simon MC. Effects of hypoxia and HIFs on cancer metabolism. Int J Hematol. 2012;95(5):464-70.

7. Martel CL, Lara PN. Renal cell carcinoma: current status and future directions. Crit Rev Oncol Hematol. 2003;45(2):177-90.

8. McLaughlin J, Lipworth L, Tarone R, Blot W. Cancer Epidemiology and Prevention. Oxford: Oxford Univ. Press; 2006.

9. Chow WH, Dong LM, Devesa SS. Epidemiology and risk factors for kidney cancer. Nat Rev Urol. 2010;7(5):245-57.

10. Gospodarowicz MK, Miller D, Groome PA, Greene FL, Logan PA, Sobin LH. The process for continuous improvement of the TNM classification. Cancer. 2004;100(1):1-5.

11. Powles T, Staehler M, Ljungberg B, Bensalah K, Canfield SE, Dabestani S, Giles RH, Hofmann F, Hora M, Kuczyk MA, Lam T, Marconi L, Merseburger AS, Volpe A, Bex A. European Association of Urology Guidelines for Clear Cell Renal Cancers That Are Resistant to Vascular Endothelial Growth Factor Receptor-Targeted Therapy. Eur Urol. 2016;70(5):705-6.

12. Jonasch E, Gao J, Rathmell WK. Renal cell carcinoma. Bmj. 2014;349:g4797.
13. Ljungberg B, Bensalah K, Canfield S, Dabestani S, Hofmann F, Hora M, Kuczyk MA, Lam T, Marconi L, Merseburger AS, Mulders P, Powles T, Staehler M, Volpe A, Bex A. EAU guidelines on renal cell carcinoma: 2014 update. Eur Urol. 2015;67(5):913-24.
14. Singer EA, Gupta GN, Srinivasan R. Update on targeted therapies for clear cell renal cell carcinoma. Curr Opin Oncol. 2011;23(3):283-9.
15. MacLennan S, Imamura M, Lapitan MC, Omar MI, Lam TB, Hilvano-Cabungcal AM, Royle P, Stewart F, MacLennan G, MacLennan SJ, Dahm P, Canfield SE, McClinton S, Griffiths TRL, Ljungberg B, N'Dow J, UCAN Systematic Review Reference Group; EAU Renal Cancer Guideline Panel. Systematic review of perioperative and quality-of-life outcomes following surgical management of localised renal cancer. Eur Urol. 2012;62(6):1097-117.
16. Janzen NK, Kim HL, Figlin RA, Belldegrun AS. Surveillance after radical or partial nephrectomy for localized renal cell carcinoma and management of recurrent disease. Urol Clin North Am. 2003;30(4):843-52.
17. Sun M, Shariat SF, Cheng C, Ficarra V, Murai M, Oudard S, Pantuck AJ, Zigeuner R, Karakiewicz PI. Prognostic factors and predictive models in renal cell carcinoma: a contemporary review. Eur Urol. 2011;60(4):644-61.
18. Atzpodien J, Royston P, Wandert T, Reitz M. Metastatic renal carcinoma comprehensive prognostic system. Br J Cancer. 2003;88(3):348-53.
19. Kim HL, Seligson D, Liu X, Janzen N, Bui MH, Yu H, Shi T, Belldegrun AS, Horvath S, Figlin RA. Using tumor markers to predict the survival of patients with metastatic renal cell carcinoma. J Urol. 2005;173(5):1496-501.
20. Majer W, Kluzek K, Bluyssen H, Wesoly J. Potential Approaches and Recent Advances in Biomarker Discovery in Clear-Cell Renal Cell Carcinoma. J Cancer. 2015;6(11):1105-13.
21. Reddi KK, Holland JF. Elevated serum ribonuclease in patients with pancreatic cancer. Proc Natl Acad Sci U S A. 1976;73(7):2308-10.
22. Huggett JF, Novak T, Garson JA, Green C, Morris-Jones SD, Miller RF, Zumla A. Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon. BMC Res Notes. 2008;1:70.
23. Connolly ID, Li Y, Gephart MH, Nagpal S. The "Liquid Biopsy": the Role of Circulating DNA and RNA in Central Nervous System Tumors. Curr Neurol Neurosci Rep. 2016;16(3):25.
24. Yoruker EE, Holdenrieder S, Gezer U. Blood-based biomarkers for diagnosis, prognosis and treatment of colorectal cancer. Clin Chim Acta. 2016;455:26-32.

25. Kalnina Z, Meistere I, Kikuste I, Tolmanis I, Zayakin P, Line A. Emerging blood-based biomarkers for detection of gastric cancer. World J Gastroenterol. 2015;21(41):11636-53.
26. Nandagopal L, Sonpavde G. Circulating Biomarkers in Bladder Cancer. Bladder Cancer. 2016;2(4):369-79.
27. Stroun M, Lyautey J, Lederrey C, Olson-Sand A, Anker P. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. Clin Chim Acta. 2001;313(1-2):139-42.
28. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst. 2001;93(14):1054-61.
29. Danila DC, Anand A, Schultz N, Heller G, Wan M, Sung CC, Dai C, Khanin R, Fleisher M, Lilja H, Scher HI. Analytic and clinical validation of a prostate cancer-enhanced messenger RNA detection assay in whole blood as a prognostic biomarker for survival. Eur Urol. 2014;65(6):1191-7.
30. Rodia MT, Ugolini G, Mattei G, Montroni I, Zattoni D, Ghignone F, Veronese G, Marisi G, Lauriola M, Strippoli P, Solmi R. Systematic large-scale meta-analysis identifies a panel of two mRNAs as blood biomarkers for colorectal cancer detection. Oncotarget. 2016;7(21):30295-306.
31. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499(7456):43-9.
32. Hess GT, Tycko J, Yao D, Bassik MC. Methods and Applications of CRISPR-Mediated Base Editing in Eukaryotic Genomes. Mol Cell. 2017;68(1):26-43.
33. Hille F, Richter H, Wong SP, Bratovic M, Ressel S, Charpentier E. The Biology of CRISPR-Cas: Backward and Forward. Cell. 2018;172(6):1239-59.
34. Khan S, Mahmood MS, Rahman SU, Zafar H, Habibullah S, Khan Z, Ahmad A. CRISPR/Cas9: the Jedi against the dark empire of diseases. J Biomed Sci. 2018;25(1):29.
35. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science. 2010;327(5962):167-70.
36. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337(6096):816-21.
37. Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D, Marraffini LA. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature. 2015;519(7542):199-202.

38. Moses C, Garcia-Bloj B, Harvey AR, Blancafort P. Hallmarks of cancer: The CRISPR generation. Eur J Cancer. 2018;93:10-8.
39. Malumbres M. Cyclin-dependent kinases. Genome Biol. 2014;15(6):122.
40. Meyerson M, Enders GH, Wu CL, Su LK, Gorka C, Nelson C, Harlow E, Tsai LH. A family of human cdc2-related protein kinases. Embo j. 1992;11(8):2909-17.
41. S Charrasse S, Carena I, Hagmann J, Woods-Cook K, Ferrari S. PCTAIRE-1: characterization, subcellular distribution, and cell cycle-dependent kinase activity. Cell Growth Differ. 1999;10(9):611-20.
42. Shehata SN, Deak M, Morrice NA, Ohta E, Hunter RW, Kalscheuer VM, Sakamoto K. Cyclin Y phosphorylation- and 14-3-3-binding-dependent activation of PCTAIRE-1/CDK16. Biochem J. 2015;469(3):409-20.
43. Herskovits AZ, Davies P. Cloning and expression analysis of two novel PCTAIRE 3 transcripts from human brain. Gene. 2004;328:59-67.
44. Herskovits AZ, Davies P. Generation and characterization of monoclonal antibodies to human PCTAIRE 3. Hybridoma (Larchmt). 2005;24(2):98-105.
45. Graeser R, Gannon J, Poon RY, Dubois T, Aitken A, Hunt T. Regulation of the CDK-related protein kinase PCTAIRE-1 and its possible role in neurite outgrowth in Neuro-2A cells. J Cell Sci. 2002;115(Pt 17):3479-90.
46. Palmer KJ, Konkel JE, Stephens DJ. PCTAIRE protein kinases interact directly with the COPII complex and modulate secretory cargo transport. J Cell Sci. 2005;118(Pt 17):3839-47.
47. Cheng K, Li Z, Fu WY, Wang JH, Fu AK, Ip NY. Pctaire1 interacts with p35 and is a novel substrate for Cdk5/p35. J Biol Chem. 2002;277(35):31988-93.
48. Yanagi T, Matsuzawa S. PCTAIRE1/PCTK1/CDK16: a new oncotarget? Cell Cycle. 2015;14(4):463-4.
49. Yanagi T, Krajewska M, Matsuzawa S, Reed JC. PCTAIRE1 phosphorylates p27 and regulates mitosis in cancer cells. Cancer Res. 2014;74(20):5795-807.
50. Matsuda S, Kominato K, Koide-Yoshida S, Miyamoto K, Isshiki K, Tsuji A, Yuasa K. PCTAIRE kinase 3/cyclin-dependent kinase 18 is activated through association with cyclin A and/or phosphorylation by protein kinase A. J Biol Chem. 2014;289(26):18387-400.
51. Matsuda S, Kawamoto K, Miyamoto K, Tsuji A, Yuasa K. PCTK3/CDK18 regulates cell migration and adhesion by negatively modulating FAK activity. Sci Rep. 2017;7:45545.
52. Barone G, Staples CJ, Ganesh A, Patterson KW, Bryne DP, Myers KN, Patil AA, Eyers CE, Maslen S, Skehel JM, Eyers PA, Collis SJ. Human CDK18 promotes replication stress signaling and genome stability. Nucleic Acids Res. 2016;44(18):8772-85.

53. Barone G, Arora A, Ganesh A, Abdel-Fatah T, Moseley P, Ali R, Chan SYT, Savva C, Schiavone K, Carmell N, Myers KN, Rakha EA, Madhusudan S, Collis SJ. The relationship of CDK18 expression in breast cancer to clinicopathological parameters and therapeutic response. Oncotarget. 2018;9(50):29508-24.
54. Pan Y, Jiang Z, Sun D, Li Z, Pu Y, Wang D, Huang A, He C, Cao L. Cyclin-dependent Kinase 18 Promotes Oligodendrocyte Precursor Cell Differentiation through Activating the Extracellular Signal-Regulated Kinase Signaling Pathway. Neurosci Bull. 2019;35(5):802-14.
55. Herskovits AZ, Davies P. The regulation of tau phosphorylation by PCTAIRE 3: implications for the pathogenesis of Alzheimer's disease. Neurobiol Dis. 2006;23(2):398-408.
56. Chaput D, Kirouac L, Stevens SM, Jr., Padmanabhan J. Potential role of PCTAIRE-2, PCTAIRE-3 and P-Histone H4 in amyloid precursor protein-dependent Alzheimer pathology. Oncotarget. 2016;7(8):8481-97.
57. Naumann U, Huang H, Wolburg H, Wischhusen J, Weit S, Ohgaki H, Weller M. PCTAIRE3: a putative mediator of growth arrest and death induced by CTS-1, a dominant-positive p53-derived synthetic tumor suppressor, in human malignant glioma cells. Cancer Gene Ther. 2006;13(5):469-78.
58. Fuhrman SA, Lasky LC, Limas C. Prognostic significance of morphologic parameters in renal cell carcinoma. Am J Surg Pathol. 1982;6(7):655-63.
59. Sobin LH, Wittekind C. TNM classification of malignant tumours. New York: Wiley-Liss; 2002. p.193–195.
60. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-10.
61. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5.
62. Liu X, Yu X, Zack DJ, Zhu H, Qian J. TiGER: a database for tissue-specific gene expression and regulation. BMC Bioinformatics. 2008;9:271.
63. Boguski MS, Lowe TM, Tolstoshev CM. dbEST--database for "expressed sequence tags". Nat Genet. 1993;4(4):332-3.
64. Krupp M, Marquardt JU, Sahin U, Galle PR, Castle J, Teufel A. RNA-Seq Atlas--a reference database for gene expression profiling in normal tissue by next-generation sequencing. Bioinformatics. 2012;28(8):1184-5.
65. BioMart. http://www.ensembl.org. Accessed 20 April 2016.
66. Zhao S, Zhang Y, Gordon W, Quan J, Xi H, Du S, von Schack D, Zhang B. Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. BMC Genomics. 2015;16:675.

67. Jung M, Ramankulov A, Roigas J, Johannsen M, Ringsdorf M, Kristiansen G, Jung K. In search of suitable reference genes for gene expression studies of human renal cell carcinoma by real-time PCR. BMC Mol Biol. 2007;8:47.
68. Universal Probe Library. www.universalprobelibrary.com. Accessed 5 May 2016.
69. Uhlen M, Bjorling E, Agaton C, Szigyarto CA, Amini B, Andersen E, Andersson A, Angelidou P, Asplund A, Asplund C, Berglund L, Bergström K, Brumer H, Cerjan D, Ekström M, Elobeid A, Eriksson C, Fagerberg L, Falk R, Fall J, Forsberg M, Björklund MG, Gumbel K, Halimi A, Hallin I, Hamsten C, Hansson M, Hedhammar M, Hercules G, Kampf C, Larsson K, Lindskog M, Lodewyckx W, Lund J, Lundeberg J, Magnusson K, Malm E, Nilsson P, Odling J, Oksvold P, Olsson I, Oster E, Ottosson J, Paavilainen L, Persson A, Rimini R, Rockberg J, Runeson M, Sivertsson A, Sköllermo A, Steen J, Stenvall M, Sterky F, Strömberg S, Sundberg M, Tegel H, Tourle S, Wahlund E, Waldén A, Wan J, Wernérus H, Westberg J, Wester K, Wrethagen U, Xu LL, Hober S, Pontén F. A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics. 2005;4(12):1920-32.
70. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005;33(Database issue):D514-7.
71. Meyer HA, Tolle A, Jung M, Fritzsche FR, Haendler B, Kristiansen I, Gaspert A, Johannsen M, Jung K, Kristiansen G. Identification of stanniocalcin 2 as prognostic marker in renal cell carcinoma. Eur Urol. 2009;55(3):669-78.
72. Thibodeau BJ, Fulton M, Fortier LE, Geddes TJ, Pruetz BL, Ahmed S, Banes-Berceli A, Zhang PL, Wilson GD, Hafron J. Characterization of clear cell renal cell carcinoma by gene expression profiling. Urol Oncol. 2016;34(4):168.e1-9.
73. Feng JY, Diao XW, Fan MQ, Wang PX, Xiao Y, Zhong X, Wu RH, Huang CB. Screening of feature genes of the renal cell carcinoma with DNA microarray. Eur Rev Med Pharmacol Sci. 2013;17(22):2994-3001.
74. Fu L, Minton DR, Zhang T, Nanus DM, Gudas LJ. Genome-Wide Profiling of TRACK Kidneys Shows Similarity to the Human ccRCC Transcriptome. Mol Cancer Res. 2015;13(5):870-8.
75. Schrodter S, Braun M, Syring I, Klumper N, Deng M, Schmidt D, Perner S, Müller SC, Ellinger J. Identification of the dopamine transporter SLC6A3 as a biomarker for patients with renal cell carcinoma. Mol Cancer. 2016;15:10.

76. Minton DR, Fu L, Mongan NP, Shevchuk MM, Nanus DM, Gudas LJ. Role of NADH Dehydrogenase (Ubiquinone) 1 Alpha Subcomplex 4-Like 2 in Clear Cell Renal Cell Carcinoma. Clin Cancer Res. 2016;22(11):2791-801.
77. Zhao Z, Liao G, Li Y, Zhou S, Zou H, Fernando S. Prognostic value of carbonic anhydrase IX immunohistochemical expression in renal cell carcinoma: a meta-analysis of the literature. PLoS One. 2014;9(11):e114096.
78. Bruick RK, McKnight SL. A conserved family of prolyl-4-hydroxylases that modify HIF. Science. 2001;294(5545):1337-40.
79. Shinojima T, Oya M, Takayanagi A, Mizuno R, Shimizu N, Murai M. Renal cancer cells lacking hypoxia inducible factor (HIF)-1alpha expression maintain vascular endothelial growth factor expression through HIF-2alpha. Carcinogenesis. 2007;28(3):529-36.
80. Sinha R, Winer AG, Chevinsky M, Jakubowski C, Chen YB, Dong Y, Tickoo SK, Reuter VE, Russo P, Coleman JA, Sander C, Hsieh JJ, Hakimi AA. Analysis of renal cancer cell lines from two major resources enables genomics-guided cell line selection. Nat Commun. 2017;8:15165.
81. Brodaczewska KK, Szczylik C, Fiedorowicz M, Porta C, Czarnecka AM. Choosing the right cell line for renal cell cancer research. Mol Cancer. 2016;15(1):83.
82. Liu YH, Lin CY, Lin WC, Tang SW, Lai MK, Lin JY. Up-regulation of vascular endothelial growth factor-D expression in clear cell renal cell carcinoma by CD74: a critical role in cancer cell tumorigenesis. J Immunol. 2008;181(9):6584-94.
83. Schmidt L, Junker K, Nakaigawa N, Kinjerski T, Weirich G, Miller M, Lubensky I, Neumann HP, Brauch H, Decker J, Vocke C, Brown JA, Jenkins R, Richard S, Bergerheim U, Gerrard B, Dean M, Linehan WM, Zbar B. Novel mutations of the MET proto-oncogene in papillary renal carcinomas. Oncogene. 1999;18(14):2343-50.
84. Gildea JJ, Shah I, Weiss R, Casscells ND, McGrath HE, Zhang J, Jones JE, Felder RA. HK-2 human renal proximal tubule cells as a model for G protein-coupled receptor kinase type 4-mediated dopamine 1 receptor uncoupling. Hypertension. 2010;56(3):505-11.
85. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol. 2018;36(8):765-71.
86. Shin HY, Wang C, Lee HK, Yoo KH, Zeng X, Kuhns T, Yang CM, Mohr T, Liu C, Hennighausen L. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. Nat Commun. 2017;8:15464.
87. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57.

88. Simonovic S, Hinze C, Schmidt-Ott KM, Busch J, Jung M, Jung K, Rabien A. Limited utility of qPCR-based detection of tumor-specific circulating mRNAs in whole blood from clear cell renal cell carcinoma patients. BMC Urol. 2020;20(1):7.
89. Rainen L, Oelmueller U, Jurgensen S, Wyrich R, Ballas C, Schram J, Herdman C, Bankaitis-Davis D, Nicholls N, Trollinger D, Tryon V. Stabilization of mRNA expression in whole blood samples. Clin Chem. 2002;48(11):1883-90.
90. Ramskold D, Wang ET, Burge CB, Sandberg R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol. 2009;5(12):e1000598.
91. Magee R, Loher P, Londin E, Rigoutsos I. Threshold-seq: a tool for determining the threshold in short RNA-seq datasets. Bioinformatics. 2017;33(13):2034-6.
92. Kishikawa T, Otsuka M, Ohno M, Yoshikawa T, Takata A, Koike K. Circulating RNAs as new biomarkers for detecting pancreatic cancer. World J Gastroenterol. 2015;21(28):8527-40.
93. Wong SC, Lo SF, Cheung MT, Ng KO, Tse CW, Lai BS, Lee KC, Lo YMD. Quantification of plasma beta-catenin mRNA in colorectal cancer and adenoma patients. Clin Cancer Res. 2004;10(5):1613-7.
94. Garcia V, Garcia JM, Pena C, Silva J, Dominguez G, Hurtado A, Alonso I, Rodriguez R, Provencio M, Bonilla F. Thymidylate synthase messenger RNA expression in plasma from patients with colon cancer: prognostic potential. Clin Cancer Res. 2006;12(7 Pt 1):2095-100.
95. Chen XQ, Bonnefoi H, Pelte MF, Lyautey J, Lederrey C, Movarekhi S, Schaeffer P, Mulcahy HE, Meyer P, Stroun M, Anker P. Telomerase RNA as a detection marker in the serum of breast cancer patients. Clin Cancer Res. 2000;6(10):3823-6.
96. Dasi F, Martinez-Rodes P, March JA, Santamaria J, Martinez-Javaloyas JM, Gil M, Aliño SF. Real-time quantification of human telomerase reverse transcriptase mRNA in the plasma of patients with prostate cancer. Ann N Y Acad Sci. 2006;1075:204-10.
97. March-Villalba JA, Martinez-Jabaloyas JM, Herrero MJ, Santamaria J, Alino SF, Dasi F. Cell-free circulating plasma hTERT mRNA is a useful marker for prostate cancer diagnosis and is associated with poor prognosis tumor characteristics. PLoS One. 2012;7(8):e43470.
98. Kang CY, Wang J, Axell-House D, Soni P, Chu ML, Chipitsyna G, Sarosiek K, Sendecki J, Hyslop T, Al-Zoubi M, Yeo CJ, Arafat HA. Clinical significance of serum COL6A3 in pancreatic ductal adenocarcinoma. J Gastrointest Surg. 2014;18(1):7-15.
99. Garcia V, Garcia JM, Pena C, Silva J, Dominguez G, Lorenzo Y, Diaz R, Espinosa P, García de Sola P, Cantos B, Bonilla F. Free circulating mRNA in plasma from breast cancer patients and clinical outcome. Cancer Lett. 2008;263(2):312-20.

100. Dong ZZ, Yao DF, Yao M, Qiu LW, Zong L, Wu W, Wu X, Yao D, Meng X. Clinical impact of plasma TGF-beta1 and circulating TGF-beta1 mRNA in diagnosis of hepatocellular carcinoma. Hepatobiliary Pancreat Dis Int. 2008;7(3):288-95.
101. Sun S, Xu MZ, Poon RT, Day PJ, Luk JM. Circulating Lamin B1 (LMNB1) biomarker detects early stages of liver cancer in patients. J Proteome Res. 2010;9(1):70-8.
102. Zheng T, Chen M, Han S, Zhang L, Bai Y, Fang X, Ding S, Yang Y. Plasma minichromosome maintenance complex component 6 is a novel biomarker for hepatocellular carcinoma patients. Hepatol Res. 2014;44(13):1347-56.
103. Garcia V, Garcia JM, Silva J, Martin P, Pena C, Dominguez G, Diaz R, Herrera M, Maximiano C, Sabin P, Rueda A, Cruz MA, Rodriguez J, Canales MA, Bonilla F, Provencio M. Extracellular tumor-related mRNA in plasma of lymphoma patients and survival implications. PLoS One. 2009;4(12):e8173.
104. Groskopf J, Aubin SM, Deras IL, Blase A, Bodrug S, Clark C, Brentano S, Mathis J, Pham J, Meyer T, Cass M, Hodge P, Macairan ML, Marks LS, Rittenhouse H. APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer. Clin Chem. 2006;52(6):1089-95.
105. Karim S, Al-Maghrabi JA, Farsi HM, Al-Sayyad AJ, Schulten HJ, Buhmeida A, Mirza Z, Al-Boogmi AA, Ashgan FT, Shabaad MM, NourEldin HF, Al-Ghamdi KBM, Abuzenadah A, Chaudhary AGA, Al-Qahtani MH. Cyclin D1 as a therapeutic target of renal cell carcinoma- a combined transcriptomics, tissue microarray and molecular docking study from the Kingdom of Saudi Arabia. BMC Cancer. 2016;16(Suppl 2):741.
106. Zhao W, Tian B, Wu C, Peng Y, Wang H, Gu WL, Gao F. DOG1, cyclin D1, CK7, CD117 and vimentin are useful immunohistochemical markers in distinguishing chromophobe renal cell carcinoma from clear cell renal cell carcinoma and renal oncocytoma. Pathol Res Pract. 2015;211(4):303-7.
107. Kumari S, Panda TK, Pradhan T. Lysyl Oxidase: Its Diversity in Health and Diseases. Indian J Clin Biochem. 2017;32(2):134-41.
108. Tanaka T, Yamaguchi J, Shoji K, Nangaku M. Anthracycline inhibits recruitment of hypoxia-inducible transcription factors and suppresses tumor cell migration and cardiac angiogenic response in the host. J Biol Chem. 2012;287(42):34866-82.
109. Liep J, Kilic E, Meyer HA, Busch J, Jung K, Rabien A. Cooperative Effect of miR-141-3p and miR-145-5p in the Regulation of Targets in Clear Cell Renal Cell Carcinoma. PLoS One. 2016;11(6):e0157801.

110. Di Stefano V, Torsello B, Bianchi C, Cifola I, Mangano E, Bovo G, Cassina V, De Marco S, Corti R, Meregalli C, Bombelli S, Viganò P, Battaglia C, Strada G, Perego RA. Major Action of Endogenous Lysyl Oxidase in Clear Cell Renal Cell Carcinoma Progression and Collagen Stiffness Revealed by Primary Cell Cultures. Am J Pathol. 2016;186(9):2473-85.
111. Schwank G, Koo BK, Sasselli V, Dekkers JF, Heo I, Demircan T, Sasaki N, Boymans S, Cuppen E, van der Ent CK, Nieuwenhuis EES, Beekman JM, Clevers H. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. Cell Stem Cell. 2013;13(6):653-8.
112. Torres-Ruiz R, Rodriguez-Perales S. CRISPR-Cas9: A Revolutionary Tool for Cancer Modelling. Int J Mol Sci. 2015;16(9):22151-68.
113. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J, Root DE. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016;34(2):184-91.
114. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol. 2014;32(12):1262-7.
115. Cui Y, Xu J, Cheng M, Liao X, Peng S. Review of CRISPR/Cas9 sgRNA Design Tools. Interdiscip Sci. 2018;10(2):455-65.
116. Kim S, Kim D, Cho SW, Kim J, Kim JS. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. 2014;24(6):1012-9.
117. Liang X, Potter J, Kumar S, Zou Y, Quintanilla R, Sridharan M, Carte J, Chen W, Roark N, Ranganathan S, Ravinder N, Chesnut JD. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. J Biotechnol. 2015;208:44-53.
118. Yu X, Liang X, Xie H, Kumar S, Ravinder N, Potter J, du Jeu X, Chesnut JD. Improved delivery of Cas9 protein/gRNA complexes using lipofectamine CRISPRMAX. Biotechnol Lett. 2016;38(6):919-29.
119. Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res. 2009;19(7):1279-88.
120. Zhu X, Xu Y, Yu S, Lu L, Ding M, Cheng J, Song G, Gao X, Yao L, Fan D, Meng S, Zhang X, Hu S, Tian Y. An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. Sci Rep. 2014;4:6420.
121. Ramlee MK, Yan T, Cheung AM, Chuah CT, Li S. High-throughput genotyping of CRISPR/Cas9-mediated mutants using fluorescent PCR-capillary gel electrophoresis. Sci Rep. 2015;5:15587.

122. Hua Y, Wang C, Huang J, Wang K. A simple and efficient method for CRISPR/Cas9-induced mutant screening. J Genet Genomics. 2017;44(4):207-13.
123. Straiton J, Free T, Sawyer A, Martin J. From Sanger sequencing to genome databases and beyond. Biotechniques. 2019;66(2):60-3.
124. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One. 2014;9(1):e78644.
125. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63.
126. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science. 2008;320(5881):1344-9.
127. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy P, Chinnaiyan AM. Transcriptome sequencing to detect gene fusions in cancer. Nature. 2009;458(7234):97-101.
128. Hrdlickova R, Toloue M, Tian B. RNA-Seq methods for transcriptome analysis. Wiley Interdiscip Rev RNA. 2017;8(1).
129. Ahn EY, DeKelver RC, Lo MC, Nguyen TA, Matsuura S, Boyapati A, Pandit S, Fu X, Zhang D. SON controls cell-cycle progression by coordinated regulation of RNA splicing. Mol Cell. 2011;42(2):185-98.
130. Xiao R, Sun Y, Ding JH, Lin S, Rose DW, Rosenfeld MG, Fu X, Li X. Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. Mol Cell Biol. 2007;27(15):5393-402.
131. Wang P, Guo L, Li K, Ning S, Shi W, Liu Z, Chen Y. Serine/arginine rich splicing factor 2 expression and clinic pathological features indicating a prognostic factor in human hepatocellular carcinoma patients. Cancer Biomark. 2018;21(3):681-7.
132. Daniels G, Jha R, Shen Y, Logan SK, Lee P. Androgen receptor coactivators that inhibit prostate cancer growth. Am J Clin Exp Urol. 2014;2(1):62-70.
133. Zhou L, Wu H, Lee P, Wang Z. Roles of the androgen receptor cofactor p44 in the growth of prostate epithelial cells. J Mol Endocrinol. 2006;37(2):283-300.
134. Sheng X, Wang Z. Protein arginine methyltransferase 5 regulates multiple signaling pathways to promote lung cancer cell proliferation. BMC Cancer. 2016;16:567.
135. Yi P, Gao S, Gu Z, Huang T, Wang Z. P44/WDR77 restricts the sensitivity of proliferating cells to TGFbeta signaling. Biochem Biophys Res Commun. 2014;450(1):409-15.

136. Ohmoto T, Nishitsuji K, Yoshitani N, Mizuguchi M, Yanagisawa Y, Saito H, Sakashita N. K604, a specific acylCoA:cholesterol acyltransferase 1 inhibitor, suppresses proliferation of U251MG glioblastoma cells. Mol Med Rep. 2015;12(4):6037-42.
137. Saraon P, Trudel D, Kron K, Dmitromanolakis A, Trachtenberg J, Bapat B, van der Kwast T, Jarvi KA, Diamandis EP. Evaluation and prognostic significance of ACAT1 as a marker of prostate cancer progression. Prostate. 2014;74(4):372-80.
138. Ye K, Wu Y, Sun Y, Lin J, Xu J. TLR4 siRNA inhibits proliferation and invasion in colorectal cancer cells by downregulating ACAT1 expression. Life Sci. 2016;155:133-9.
139. Ghalei H, Schaub FX, Doherty JR, Noguchi Y, Roush WR, Cleveland JL, Stroupe ME, Karbstein K. Hrr25/CK1delta-directed release of Ltv1 from pre-40S ribosomes is necessary for ribosome assembly and cell growth. J Cell Biol. 2015;208(6):745-59.
140. Collins JC, Ghalei H, Doherty JR, Huang H, Culver RN, Karbstein K. Ribosome biogenesis factor Ltv1 chaperones the assembly of the small subunit head. J Cell Biol. 2018;217(12):4141-54.
141. Kedzierska H, Poplawski P, Hoser G, Rybicka B, Rodzik K, Sokol E, Bogusławska J, Tański Z, Fogtman A, Koblowska M, Piekiełko-Witkowska A. Decreased Expression of SRSF2 Splicing Factor Inhibits Apoptotic Pathways in Renal Cancer. Int J Mol Sci. 2016;17(10).
142. Piekielko-Witkowska A, Wiszomirska H, Wojcicka A, Poplawski P, Boguslawska J, Tanski Z, Nauman A. Disturbed expression of splicing factors in renal cancer affects alternative splicing of apoptosis regulators, oncogenes, and tumor suppressors. PLoS One. 2010;5(10):e13690.
143. Zhao Z, Wu F, Ding S, Sun L, Liu Z, Ding K, Lu J. Label-free quantitative proteomic analysis reveals potential biomarkers and pathways in renal cell carcinoma. Tumour Biol. 2015;36(2):939-51.
144. Matsumoto K, Fujiwara Y, Nagai R, Yoshida M, Ueda S. Expression of two isozymes of acyl-coenzyme A: cholesterol acyltransferase-1 and -2 in clear cell type renal cell carcinoma. Int J Urol. 2008;15(2):166-70.
145. Chen J, Cui L, Yuan J, Zhang Y, Sang H. Circular RNA WDR77 target FGF-2 to regulate vascular smooth muscle cells proliferation and migration by sponging miR-124. Biochem Biophys Res Commun. 2017;494(1-2):126-32.
146. Geng F, Cheng X, Wu X, Yoo JY, Cheng C, Guo JY, Mo X, Ru P, Hurwitz B, Kim S, Otero J, Puduvalli V, Lefai E, Ma J, Nakano I, Horbinski C, Kaur B, Chakravarti A, Guo D. Inhibition of SOAT1 Suppresses Glioblastoma Growth via Blocking SREBP-1-Mediated Lipogenesis. Clin Cancer Res. 2016;22(21):5337-48.

147. Garcia-Bermudez J, Birsoy K. Drugging ACAT1 for Cancer Therapy. Mol Cell. 2016;64(5):856-7.
148. Antalis CJ, Arnold T, Rasool T, Lee B, Buhman KK, Siddiqui RA. High ACAT1 expression in estrogen receptor negative basal-like breast cancer cells is associated with LDL-induced proliferation. Breast Cancer Res Treat. 2010;122(3):661-70.
149. Chen X, Liang H, Song Q, Xu X, Cao D. Insulin promotes progression of colon cancer by upregulation of ACAT1. Lipids Health Dis. 2018;17(1):122.

Abbreviations

ccRCC: clear cell renal cell carcinoma; RT-qPCR: quantitative real time PCR; RCC: renal cell carcinoma; VHL: von Hippel–Lindau tumor suppressor gene; HIF: hypoxia-inducible factor; FAK: focal adhesion kinase; PSA: prostate-specific antigen; CEA: carcinoembryonic antigen; TCGA: Cancer Genome Atlas database; CNV: copy number variation; GTEx: Genotype Tissue Expression database; TiGER: Tissue-specific Gene Expression and Regulation database; GEO: Gene Expression Omnibus database; miRNAs: microRNAs; CTCs: circulating tumor cells; cfDNA: circulating cell-free DNA; circRNA: circular RNA; lncRNA: long non-coding RNAs; HIF: hypoxia inducible factor; HCC: hepatocellular carcinoma; CDK18: cyclin-dependent kinase 18; LOX: lysyl oxidase; CCND1: cyclin D1; CRISPR: Clustered Regularly Interspaced Short Palindromic Repeat; PAM: Protospacer Adjacent Motif; gRNA: guide RNA; crRNAs: CRISPR RNAs; tracrRNA: trans-activating crRNA; GOF: gain of function; IGV: Integrative Genomics Viewer; rpkm: Reads Per Kilobase Million; TPM: Transcripts Per Million; TNM: TNM Classification of Malignant Tumors; BPE: bovine pituitary extract; EGF: human recombinant epidermal growth factor

Statutory declaration / Declaration of own contribution

I, Sinisa Simonovic, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic ‘The role of cyclin-dependent kinase 18 (CDK18) in clear cell renal cell carcinoma’, (‘Die Rolle der Cyclin-abhängigen Kinase 18 (CDK18) im klarzelligen Nierenzellkarzinom’) independently and without the support of third parties, and that I used no other sources and aids than those stated.
All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.
Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons (cf. declaration of contribution). I have correctly marked all texts or parts of texts that were generated in collaboration with other persons.

My contributions to any publications to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; www.icmje.org) on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty.

The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me.”

Date Signature

Declaration of own contribution to publications

Sinisa Simonovic contributed the following to the below listed publication:

Publication 1: Limited utility of qPCR-based detection of tumor-specific circulating mRNAs in whole blood from clear cell renal cell carcinoma patients. Sinisa Simonovic, Christian Hinze, Kai M. Schmidt-Ott, Jonas Busch, Monika Jung, Klaus Jung and Anja Rabien. BMC Urol. 2020 Feb 4;20(1):7.

Contribution in detail:

The candidate has independently applied all methods designated by the plan of the study. He has made suggestions to the study design, was solely responsible for the data collection and analysis, and contributed to the critical interpretation of the data. The candidate has independently written the first draft of the publication and contributed to its revision.

_____

Signature, date and stamp of first supervising university professor / lecturer

_____

Signature of doctoral candidate

Curriculum vitae

My curriculum vitae does not appear in the electronic version of my paper for reasons of data protection.

List of publications

1. Limited utility of qPCR-based detection of tumor-specific circulating mRNAs in whole blood from clear cell renal cell carcinoma patients. Sinisa Simonovic, Christian Hinze, Kai M. Schmidt-Ott, Jonas Busch, Monika Jung, Klaus Jung and Anja Rabien. BMC Urol. 2020 Feb 4;20(1):7.

Acknowledgments

I am very grateful to my supervisors Prof. Dr. Kai Schmidt-Ott and PD. Anja Rabien; Dr. Christian Hinze and Tatjana Luganskaja at the Max Delbrück Center, as well as scientific and technical staff at the Urology Laboratory at Charité Mitte: Silke Rabenhorst, Sabine Becker, Siegrun Blauhut, Bettina Ergün, Sabine Weickmann, Dr. Monika Jung and Prof. Klaus Jung.