# Quantitative proteomics analysis of clear cell Renal Cell Carcinoma for the identification of diagnostic and prognostic biomarker panels

by

**Aydanur Şentürk**

A Dissertation Submitted to the Graduate
School of Sciences and Engineering in
Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
in

Molecular Biology and Genetics

KOÇ ÜNİVERSİTESİ

July 16, 2021

**Quantitative proteomics analysis of clear cell Renal Cell Carcinoma for the identification of diagnostic and prognostic biomarker panels**

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a doctoral dissertation by

**Aydanur Şentürk**

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Assoc. Prof. Nurhan Özlü (Advisor, Koç University)

_____

Assoc. Prof. Nurcan Tunçbağ (Co-Advisor, Koç University)

_____

Assoc. Prof. Tuğba Bağcı Önder (Koç University)

_____

Prof. Şermin Genç (İzmir Biomedicine and Genome Center)

_____

Assist. Prof. Zeynep Gümüş (Icahn School of Medicine at Mount Sinai)

Date: 16.07.2021

*This thesis is dedicated to everybody who contributed.*

# ABSTRACT


**Quantitative proteomics analysis of clear cell Renal Cell Carcinoma for the identification of diagnostic and prognostic biomarker panels**

**Aydanur Şentürk**
**Doctor of Philosophy in Molecular Biology and Genetics**
**July 16, 2021**

Clear cell Renal Cell Carcinoma (ccRCC) is the third most common and most malignant urological cancer, with a 5-year survival rate of 10% for patients with advanced tumors. Despite an increasing rate of early detections, one third of the patients already show metastasis at diagnosis. Inherent and acquired resistance to chemo- and radiotherapies complicate the treatment of the disease, leaving the current treatment option primarily to surgical resection of the tumor. Biomarkers can help to monitor and to target tumor progression and growth, and to make a better prognosis on patient outcome. However, no universal biomarkers are in clinical use for ccRCC.

Here, a rigorous quantitative dimethylation-based proteomics approach is described to identify biomarker panels for the diagnosis (part I) and for the stratification (part II) of ccRCC tumors, and to illuminate the driving phosphosignaling events in renal cancers (part III). The comprehensive characterization of the ccRCC global proteome and phosphoproteome revealed that the candidate marker proteins PLOD2, FERMT3, SPARC and SIPRα are overexpressed, and that diverse kinases of the groups CDK, PAK and MAPK are highly activated in the tumor tissues compared to normal adjacent tissues. The associated phosphosignaling cascades are linked to tumor growth and metastasis. Furthermore, our analysis suggested that due to interpatient heterogeneity, ccRCC tumors distinguish into two groups with distinct overall survival of patients and different enriched malignant pathways.

Overall, the suggested biomarkers can serve as targets for future treatment strategies of ccRCC tumors in combination with approved therapeutics.

# ÖZETÇE

**Diyagnostik ve prognostik biyobelirteç belirlemesi için şeffaf hücreli tip renal hücreli karsinomun kantitatif proteomik analizi**

**Aydanur Şentürk**
**Moleküler Biyoloji ve Genetik, Doktora**
**16 Temmuz 2021**

Şeffaf hücreli renal hücreli karsinom (clear cell Renal Cell Carcinoma, ccRCC) ürolojik kanserler arasında üçüncü en sık görülen ve aynı zamanda en kötü huylu olanıdır. İleri evre tümörlü hastaların sadece 10% teşhisten 5 sene sonra hayatta kalmaktadır. Erken tespit sayıların artmasına rağmen, hastaların üçte biri teşhis konulduğunda metastaz göstermektedir. Ancak RCC'de, kemoterapi veya radyoterapi gibi geleneksel tedaviler yetersiz kaldığı için, direkt cerrahi yöntem ile tümörün çıkarılması yoluna gidilmektedir. Biyobelirteçler kansere neden olan mekanizmaları tanımlamaya ve tümör heterojenliği sorununu aşmaya yardımcı olabilir, fakat şu anda RCC için klinikte kullanılabilecek evrensel biyobelirteçler bulunmamaktadır.

Bu çalışmada, ccRCC tümörleri tespit etmek için (ilk kısım), sınıflandırmak için (ikinci kısım) ve fosforilasyon sinyal yolaklarını anlamak için (üçüncü kısım) kantitatif proteomik analizi uygulanmıştır. Bulgular dört aday proteinin (PLOD2, FERMT3, SPARC ve SIRPα) normal doku ile karşılaştırıldığında tümörde önemli seviyede artmış olduğunu, ve CDK, PAK ve MAPK kinaz gruplarından farklı kinazların tümörde aktif olduğunu göstermektedir. Üstelik, tümör hücre büyümesini ve metastazı destekleyen MAPK ve VEGF fosforilasyon sinyal yolaklarının tümörde artmış olduğu tespit edildi. Ayrıca, ccRCC hastaların heterojen proteom profillerine göre bilişimsel bir yaklaşım ile iki ayrı grupta kümelendirilmesi öngörüldü. Hasta grupların farklı toplam sağkalımı ve farklı kötü huylu yolakların zenginleşmiş olduğu tespit edildi.

# ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my PhD advisor Dr. Nurhan Özlü for accepting me in her group, for assigning me this interesting project and for her support, guidance and scientific teaching during my doctoral studies. I would also like to thank my two thesis monitoring committee members Dr. Nurcan Tunçbağ and Dr. Tuğba Bağcı Önder for their interest in my work and for the insightful scientific comments and helpful advice during my doctoral studies. I would like to thank Prof. Şermin Genç and Dr. Zeynep Gümüş for accepting to be members of my thesis defense jury, for their interest in my research work and for the critical reading of my thesis.

Further, I would like to thank our collaborators from the clinics: From Koç University School of Medicine, Dr. Ömer Acar, Dr. Murat Can Kiremit, Dr. Ayşe Armutlu, Dr. Özgür Kurt and Dr. Tarık Esen, and from Istanbul University Istanbul Faculty of Medicine, Dr. Selçuk Erdem and Sidar Bağbudar. They made this research possible by their contributions and guidance on medical questions.

Many thanks belong to all members of the NOZLU lab, especially to Ayşe Tuğçe Şahin for her help and contributions in my PhD project, and for her friendship. I would also like to thank my friends Büşra Harmanda, Nazan Saner, Mohammad Haroon Qureshi, Ceyda Seren Ceyhan and Berfu Nur Yiğit for their support and for the joyful time during my PhD.

Furthermore, I would like to express my gratitude to Koç University for awarding me a full PhD scholarship. I would also like to acknowledge the KUPAM and KUTTAM proteomics facilities of Koç University for allowing me to use the mass spectrometry systems for my research.

Lastly, I would like to thank my family for supporting me during this PhD. My special thanks belong to Jan Fehmi Sayılgan for his endless support and encouragement and for always being there during my doctoral research.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABBREVIATIONS

| | |
|---|---|
| aa | amino acid |
| ABC | Ammonium bicarbonate |
| ACN | Acetonitrile |
| ACTB | β-actin |
| AGC | Automatic gain control |
| ATP | Adenosine triphosphate |
| BCA | Bicinchoninic acid |
| BH | Benjamini-Hochberg |
| BSA | Bovine serum albumin |
| ccRCC | clear cell Renal Cell Carcinoma |
| CDK | Cyclin-dependent kinase |
| CGC | Cancer Gene Census |
| chRCC | chromophobe Renal Cell Carcinoma |
| CI | Confidence interval |
| CPTAC | Clinical Proteomic Tumor Analysis Consortium |
| CT | Computed tomography |
| Da | Dalton |
| DAB | 3, 3′-diaminobenzidine |
| DDA | Data-dependent acquisition |
| dotp | dot product |
| DTT | Dithiothreitol |
| EDTA | Ethylenediaminetetraacetic acid |
| EGFR | Epidermal growth factor receptor |
| ELISA | Enzyme-linked immunosorbent assay |
| EMT | Epithelial-to-mesenchymal-transition |
| ER | Endoplasmic reticulum |
| FA | Formic acid |
| FDA | Food and Drug Administration |
| FDR | False discovery rate |
| FFPE | Formalin-fixed paraffin-embedded |
| FPKM | Fragments per kilobase of transcript per million mapped reads |
| FT-ICR | Fourier transform ion cyclotron resonance |
| GO | Gene ontology |
| GTP | Guanosine triphosphate |

| | |
|---|---|
| HCD | Higher-energy C-trap dissociation |
| HF | High fidelity |
| HIF | Hypoxia-inducible factor |
| H/L | Heavy-light ratio |
| HPLC | High-performance liquid chromatography |
| HR | Hazard ratio |
| HRP | Horseradish peroxidase |
| IAA | Iodoacetamide |
| IgG | Immunoglobulin G |
| IHC | Immunohistochemistry |
| IL | Interleukin |
| kDa | kilo Dalton |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| KIRC | Kidney Renal Clear Cell Carcinoma |
| LC | Liquid chromatography |
| LFQ | Label-free quantification |
| LH | Lysyl hydroxylase |
| MA | M (log ratio) and A (average) plot |
| miRNA | micro RNA |
| MOAC | Metal oxide affinity chromatography |
| mRCC | metastatic Renal Cell Carcinoma |
| mRNA | messenger RNA |
| MS | Mass spectrometry |
| MS/MS | Tandem mass spectrometry |
| mTOR | mammalian target of rapamycin |
| m/z | mass-to-charge ratio |
| NAT | Normal adjacent tissue |
| NCE | Normalized collision energy |
| ORA | Over-representation analysis |
| OS | Overall survival |
| OXPHOS | Oxidative phosphorylation |
| PAK | p21-activated kinase |
| PBS | Phosphate buffered saline |
| PCA | Principal component analysis |
| PD | Proteome Discoverer |
| PD-1 | Programmed cell death protein 1 |
| ppm | parts per million |

| | |
|---|---|
| pRCC | papillary Renal Cell Carcinoma |
| PRM | Parallel reaction monitoring |
| pSer/pS | phosphoserine |
| PSM | Peptide spectral match |
| pThr/pT | phosphothreonine |
| PTM | Post-translational modification |
| pTyr/pY | phosphotyrosine |
| pVHL | VHL protein |
| RCC | Renal Cell Carcinoma |
| RFS | Recurrence-free survival |
| RNA | Ribonucleic acid |
| RP | Reverse phase |
| RT | Room temperature |
| SCX | Strong cation exchange |
| SDS | Sodium dodecyl sulfate |
| shRNA | short hairpin RNA |
| SPE | Solid phase extraction |
| SRM | Selected reaction monitoring |
| TCA | Tricarboxylic acid |
| TCC | Transitional Cell Carcinoma |
| TCGA | The Cancer Genome Atlas |
| TFA | Trifluoroacetic acid |
| $TiO_2$ | Titanium dioxide |
| TK | Tyrosine kinase |
| TMT | Tandem mass tag |
| VEGFR | Vascular endothelial growth factor receptor |
| VHL | von-Hippel-Lindau |
| VIM | vimentin |
| WB | Western blot |

*Chapter 1*

# Introduction

Renal Cell Carcinoma (RCC) is among the 10 most common cancers in both men and women and causes more than 140,000 deaths worldwide every year (1). Clear cell RCC (ccRCC) is the most common histological subtype (70% of cases) and is characterized by the loss of the *VHL* tumor suppressor gene leading to increased angiogenesis and cell growth (2). Approximately one third of the patients already display metastasis at the time of diagnosis and are expected to have a median survival of around 26 months (2, 3). This is mostly due to the low success of traditional therapies such as chemotherapy or radiotherapy, limiting the treatment options for renal cancers primarily to surgical intervention in combination with targeted therapy and/or immunotherapy in suitable cases (1). This suggests that predominant molecular aberrations of RCC tumorigenesis are still not fully comprehended.

One strategy in cancer therapy for a better understanding of underlying molecular mechanisms is the identification of biomarkers, which can be used as indicators of tumor progression and patient outcome. In recent years, their discovery has largely benefitted from advances in high-throughput technologies such as mass spectrometry-based proteomics. However, despite their urgent need, no universal biomarkers of clinical utility are currently known for ccRCC (4). This is due to previous studies lacking in coverage depth and providing insufficient validation of the candidates by orthogonal methods or in independent cohorts (5-11). Also, tumor heterogeneity caused by genetic and epigenetic alterations, as well as by the diversity in the composition of the tumor microenvironment, is a major drawback in the efficient treatment of the malignancy (12). In recent years,

patient classification based on in-depth molecular expression profiles have grown in importance to minimize interpatient heterogeneity and to offer individualized treatment (13).

Another crucial factor driving tumorigenesis are post-translational modifications (PTM) such as phosphorylation or glycosylation of proteins that substantially influence the pathobiology of cancers and are difficult to be portrayed in a large-scale manner. However, only few undertakings have been conducted for the depiction of the signaling cascades mediated by phosphorylation in ccRCC tumors (14-16).

Here, a rigorous proteomics-based characterization of 13 ccRCC tumors for the identification of biomarker panels was performed. In the first part (Chapter 2), a global shotgun proteomics approach was conducted for the identification of a diagnostic, secreted biomarker group. A total of 10,160 unique proteins were identified, of which 955 proteins were determined to be significantly regulated between tumor and normal adjacent tissues. Furthermore, four putatively secreted candidates, namely PLOD2, FERMT3, SPARC and SIRPα, were verified by three orthogonal approaches as highly expressed proteins in the tumors compared to normal adjacent tissues (NATs) that were not affected by intra- and intertumor heterogeneity. Three of the proteins were found to be secreted into urine of ccRCC patients, with SPARC displaying a significant increase in ccRCC urine samples compared to a healthy control group, making it a promising marker for the detection of the disease in body fluids.

In the second part (Chapter 3), a two-cluster-based classification system for the risk stratification of ccRCC patients was developed using the generated global proteome profiles. The determined clustering signature was conserved in independent ccRCC transcriptome and proteome expression data and suggested almost three times higher risk of death for cluster 1 tumors compared to cluster 2 tumors. Furthermore, cluster 1 tumors were characterized by a prevalence of malignant processes such as epithelial-to-mesenchymal-transition (EMT), hypoxic signaling, angiogenesis and immune system response, while cluster 2 tumors indicated a low-risk profile and were mainly associated

with mTOR signaling. Among the most significant cluster-discriminative proteins were 13 targets of repurposed inhibitory FDA-approved drugs.

In the third part (Chapter 4), a shotgun phosphoproteomics analysis of the same tissue samples was performed. More than 16,000 phosphopeptides were identified, of which 9,431 were reliably quantified in the tissues. Furthermore, 1,336 phosphopeptides were identified to be significantly differentially abundant between tumor and NATs and to be associated with increased regulation of cytoskeletal elements and cell growth. Kinase activity predictions showed mostly cyclin-dependent (CDKs) and p21-activated (PAKs) kinases to be highly activated in the tumor tissues, with CDK1 and PAK2 as predicted main drivers of RCC pathobiology. Moreover, EGFR and VEGFR mediated MAPK and VEGF signaling were determined to be mainly responsible for tumor growth, proliferation and invasion in RCC.

Overall, our comprehensive quantitative global proteome and phosphoproteome profiling of ccRCC tumors provides a better understanding of underlying mechanisms of renal tumorigenesis and gives promise for future drug treatment approaches.

*Chapter 2*

# Global Proteome Analysis of ccRCC

## 2.1 LITERATURE REVIEW

### 2.1.1 Cancer

Cancer is the uncontrolled growth of cells. Cancerous cells can arise from all parts of the body and acquire features that enable them to sustain and survive by evading the control mechanisms and at the same time by hijacking the growth mechanisms of the organism (17). An overview of general hallmarks of cancer cells is given in **Fig 1**.

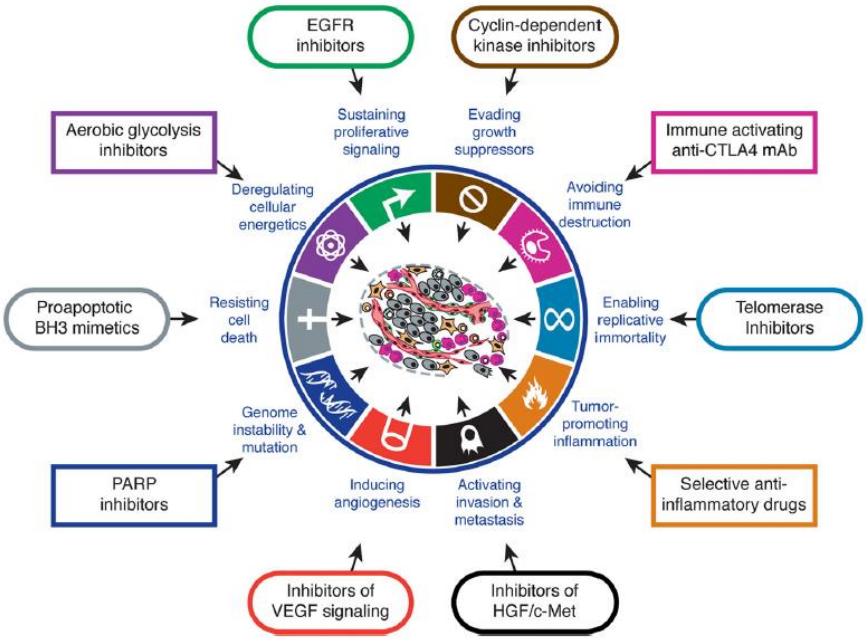


**Figure 1 | The hallmarks of cancer required for tumor growth and progression.** Indicated are also drug types targeting these features of cancers (adapted from (17)).

### 2.1.2 Epidemiology and survival of kidney cancer patients

Kidney cancer is among the 10 most common cancers in both men and women and accounts for 4% of all new cancer incidences worldwide (18, 19). More than 400,000 new individuals are diagnosed with this disease and more than 140,000 deaths are caused by renal cancer each year worldwide (1, 20). The median age of kidney cancer patients is 60 years and a 2:1 male-to-female ratio of occurrence has been reported (19). The male predominance might be due to mutations in the *KDM5C* gene (**Fig. 27**), which resides at the X chromosome and is the sole wild type allele (13), and also due to higher rates of smoking and obesity in the male population (see 2.1.3).

At the time of diagnosis, almost 30% of patients suffer from metastasis located in their lungs, bones, liver or brain, and about one third of these patients subsequently relapse after surgical treatment (1, 21). Furthermore, high-risk patients have an estimated 5-year survival rate of 8-10% and a median survival time of 26 months (3, 21).

### 2.1.3 Causes and risk factors of kidney cancer

Kidney cancer can occur due to following reasons:

#### (a) Familial kidney cancer

Approximately 2-3% of incidences are familial kidney cancer (hereditary) commonly expressed in patients with the von-Hippel-Lindau (VHL) disease (1 in 36,000 births) (19). Persons within this category, who inherit a defective allele of the *VHL* gene and acquire a defect in the other allele by somatic mutations, have an increased risk in developing vascular tumors in early years such as clear cell Renal Cell Carcinoma, haemanoglioblastomas (affects the central nervous system) or pheochromocytomas (arises in the adrenal gland of the kidney) (19, 22). Usually this type of disease is treated by monitoring the renal lesions and intervening when the size exceeds 3 cm.

**(b) Sporadic kidney cancer**

In the case of non-inherited sporadic kidney cancer, both alleles of the *VHL* gene acquire genetic defects (biallelic loss) leading to tumor development (19). Most commonly accepted risk factors for sporadic renal cancer development are smoking, obesity, diabetes, and high blood pressure (**Fig. 2**) (19, 23). Furthermore, substances such as cadmium, certain herbicides and organic solvents are suspected to increase the risk for this cancer type (22). However, for most patients a clearly identifiable cause is not found.

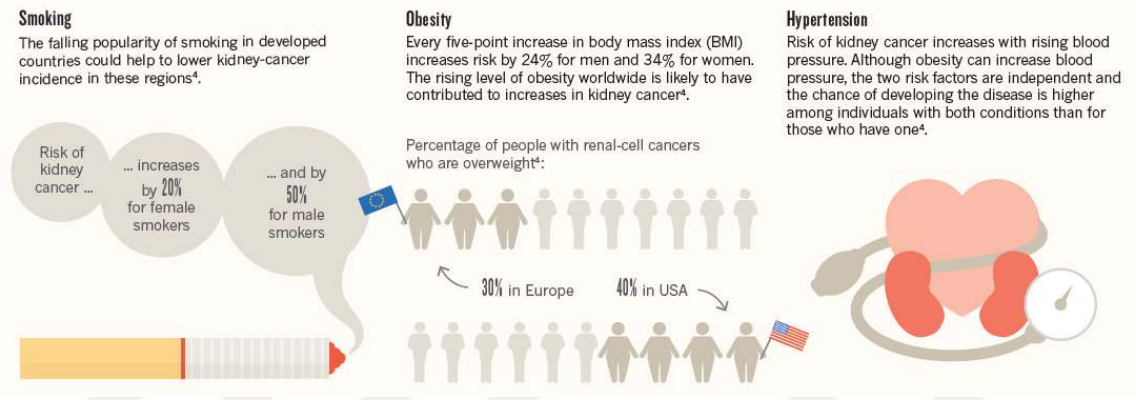


**Figure 2 | Risk factors for developing sporadic kidney cancer.** The three most common factors are displayed (adapted from (23)).

### 2.1.4 Subtypes of kidney cancer

The majority of kidney cancers are Renal Cell Carcinomas (RCC) accounting for approximately 90% of incidences and arising from the proximal tubules in the epithelial tissue, while the remaining cases are assigned as Transitional Cell Carcinomas (TCC) (~10%) originating from the uretopelvic junction (22).

The most common histological subtype of RCC is clear cell Renal Cell Carcinoma (ccRCC) (75-80%), which originates from the proximal convoluted tubule and is mostly associated with the loss of the *VHL* gene, a clear cytoplasm due to high lipid and glycogen

content and delicate vasculature (19) (**Fig. 3**). Similarly, the second most common subtype, papillary Renal Cell Carcinoma (pRCC), is also considered to arise from the proximal tubules, to account for approximately 15% of cases and to be characterized by cuboidal monocolumnar cells with a solid cytoplasm (24).

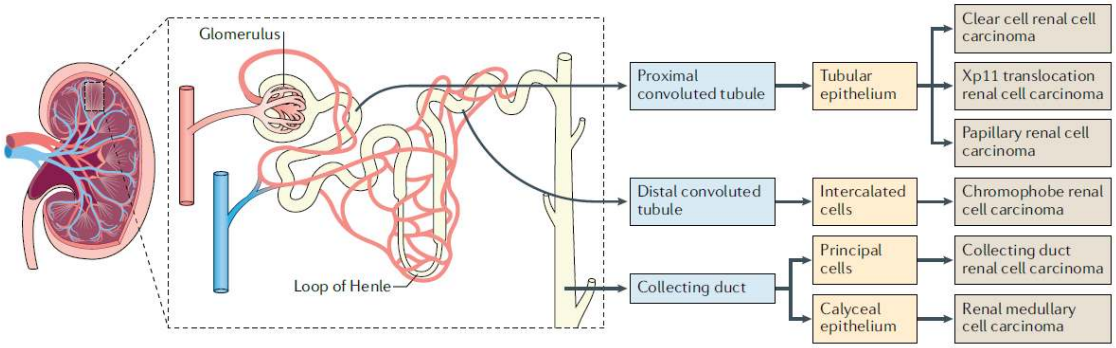


**Figure 3 | Origins of histological subtypes of Renal Cell Carcinoma in the renal epithelium.** While the two most common subtypes ccRCC (75-80%) and pRCC (15%) originate from the proximal convoluted tubule, the third common subtype chRCC (5%) is considered to arise from intercalated cells in the distal convoluted tubule. The remaining subtypes are rarely diagnosed and under debate (adapted from (24)).

The focus of this work is the clear cell-type of Renal Cell Carcinoma.

### 2.1.5 Altered pathways in ccRCC

The tumorigenesis of Renal Cell Carcinoma is dominated by two main signaling mechanisms, which have both the stabilization and accumulation of hypoxic factors in common.

#### 2.1.5.1 The VHL pathway

The loss of the *VHL* gene is assumed to be the main driver event in the development of RCC. *VHL* encodes the von-Hippel-Lindau protein (pVHL), a substrate recognition component of an E3 ubiquitin ligase complex, which is under normoxic conditions involved in the proteosomal degradation of hypoxia-inducible transcription factors (HIFs) (**Fig. 4** and **Fig. 6**) (25). HIFs are heterodimeric transcription factors that regulate the expression of genes for VEGF (vascular endothelial growth factor), PDGF (platelet derived growth factor), TNFα (transforming growth factor alpha), glycolysis related enzymes and basic fibroblast growth factors by binding to their hypoxia response elements (19, 26). Consequently, under hypoxic conditions or in pVHL depleted environments such as in renal tumors (pseudohypoxic condition) biological processes including cell proliferation, vascular tone, autocrine growth stimulation, cell cycle control, glucose transport, iron metabolism, energy production, cell migration and angiogenesis are dysregulated due to the aberrant stabilization and accumulation of HIFs (19, 26).

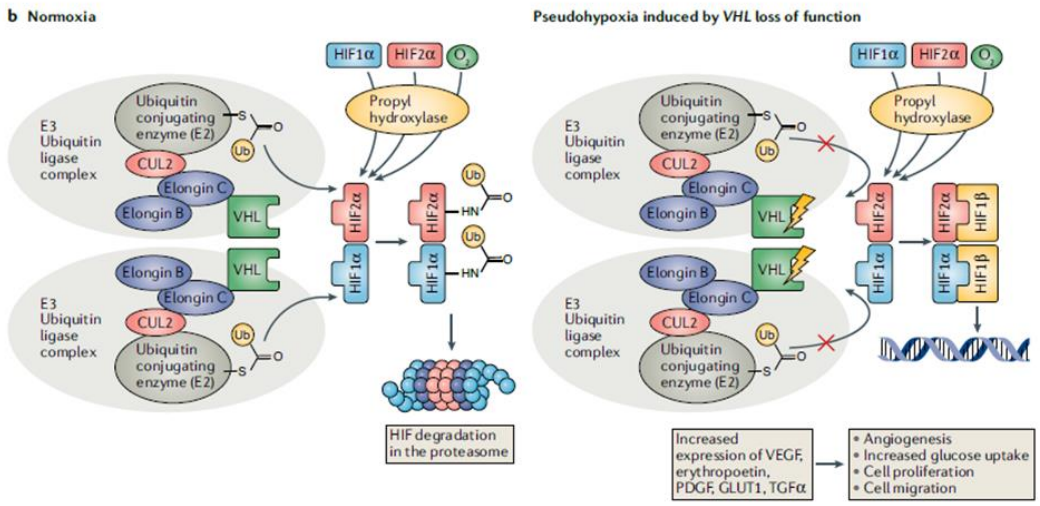


**Figure 4 | HIF1α and HIF2α stabilization and accumulation due to the loss of the *VHL* gene in ccRCC.** The impaired proteosomal degradation of HIF transcription factors by pVHL leads to hypoxia-like conditions (pseudohypoxia). Accumulated HIFs cause the enhanced expression of target genes responsible for glycolysis, angiogenesis, cell proliferation and migration (adapted from (24)).

#### 2.1.5.2 The mTOR pathway

The second most relevant pathway in ccRCC is the mTOR (mammalian target of rapamycin) mediated pathway also leading to the production of HIF as well as cell cycle regulators such as cyclin-D1 and c-Myc via activation of the p70S6 kinase (**Fig. 6**) (19). mTOR is a downstream target of the phosphoinositide 3-kinase (PI3K)/protein kinase B (AKT) signaling stimulated by extracellular signal molecules such as VEGF or PDGF (3, 19). The growth factors bind to the responsive tyrosine kinase receptors at the cell membrane of RCC tumor cells and trigger the formation of phosphatidylinositol-3,4,5-triphosphate (PIP3) by PI3K (3). Upon PIP3 generation, AKT is recruited to the membrane driving mTOR activation, and thereby cell proliferation, survival and angiogenesis (3). As in many cancers, the constitutive activation of AKT and consequently of mTOR is assumed in ccRCC to be the result of the inactivation of the

tumor suppressor gene *PTEN* (3).

### 2.1.6 Treatment of ccRCC

Renal cancers are more often detected at early stages due to more frequent incidental detections by routine abdominal CT imaging, however, the mortality has increased (19). In general, the treatment of renal cancer masses comprises surgical excision of the tumor by partial or radical nephrectomy. For small tumor masses (<7 cm diameter), partial nephrectomy is preferred since it is less invasive when performed by laparoscopic techniques (19). Radical nephrectomy (complete removal of the organ) is used to manage locally advanced tumors (>7 cm diameter), in case of multiple small tumors or when the tumor extends into the vasculature, but this intervention carries the risk of developing chronic kidney disease after surgery (19, 25).

For patients with inoperable or metastatic RCC (mRCC) the standard of care is systemic treatment with targeted therapies and/or immunotherapies (**Fig. 5**) (25). As initial treatment regimen, cytokines such as high-dose IL-2 and interferon-α have been used in the 1990s, but were soon replaced due to distressing toxicity by new targeted therapeutics mainly attacking the VHL (axitinib, sunitinib, sorafenib, pazopanib) and mTOR pathways (temsirolimus, everolimus) (25). Furthermore, novel antibody-based immunotherapeutics such as nivolumab, which targets the PD1-PDL1 axis, or ipilimumab, which targets the CTLA4 protein, have been approved for RCC treatment. Both aim to disrupt the evasive mechanisms of tumor cells to escape the immune system by inhibiting the immune checkpoints PD1 and/or CTLA4 and thereby promoting the activation of cytotoxic T cells to attack the cancer cells (**Fig. 6**) (25).

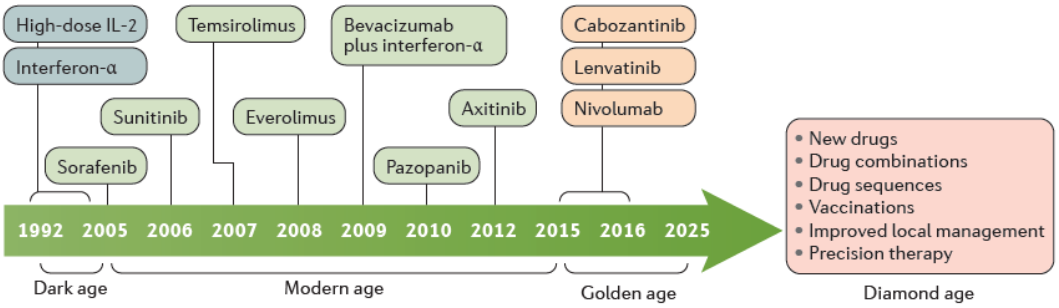


**Figure 5 | Development of targeted therapeutics for metastatic ccRCC.** Early phase of therapy consisted of cytokines which achieved only a median survival of ~15 months and showed high toxicity. Since 2005 novel approved therapeutics target the most prominent pathways (VEGF and mTOR) and doubled the expected survival time to ~30 months (adapted from (25)).

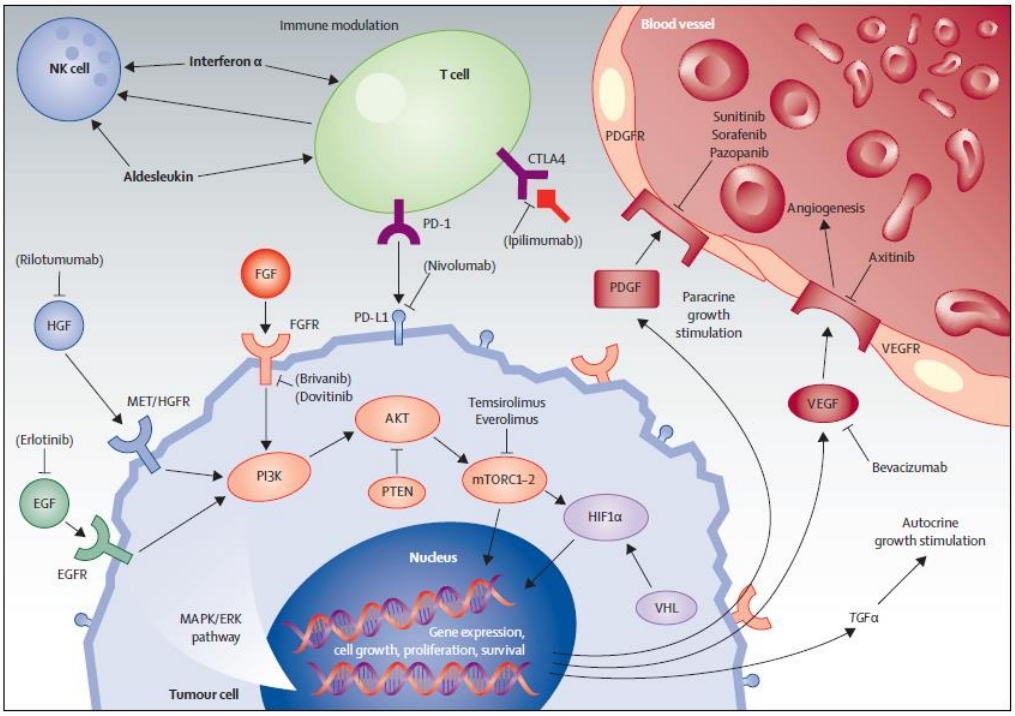


**Figure 6 | Inhibition mechanisms of targeted therapies approved for advanced ccRCC tumors.** NK=natural killer (adapted from (1)).

### 2.1.7 Biomarker discovery in cancer research

To overcome the challenge of tumor heterogeneity, individualizing clinical outcomes (so-called precision medicine) is emerging as a promising approach for the treatment of cancers. Biomarkers can support the assessment of a patients' tumor profile while not being restricted to any specific class. A biomarker can be a clinical parameter such as blood tension, genetic/epigenetic alterations (mutations, methylation) or molecular substances such as proteins (25). Furthermore, biomarkers are considered as measurable indicators of pathological or physiological processes of a condition or a disease (27). In general, biomedical markers are classified into three categories (27):

(a) diagnostic: used for risk stratification and early cancer detection
(b) prognostic: provide an indication of the likely progression of the disease
(c) predictive: used for predicting treatment measures to be taken on a patient

Moreover, there are three characteristics a robust biomarker should possess, namely it should be expressed homogenously within a specimen, it should be specific in an assay-like detection, and it should have biological relevance for tumor pathogenesis (28).

### 2.1.8 Proteomics in biomarker discovery

In recent years, the discovery of cancer biomarkers has largely benefitted from technological advances in high-throughput mass spectrometry-based proteomics. Given that cellular functions are carried out on the protein level, that a substantial discrepancy in regulation between mRNA and protein expression exists, that pharmaceutical targets are mostly proteins, and that post-translational modifications add another level of regulation to the protein level, proteomics offers an alternative methodology addressing the shortcomings of genomics- and transcriptomics-based landscaping of tumors (20).

Mass spectrometry analysis of biological samples constitutes of three main steps (**Fig. 7**):

1. Ionization: Peptides of a complex sample are separated using liquid chromatography prior to entry into the mass spectrometer where they are converted to gas-phase ions by electrospray ionization. By forcing the liquid sample through a narrow electrified needle, the generated molecule droplets evaporate and are accelerated through vacuum. The removal of electrons adds under low pH conditions multiple positive charges to the peptide molecules.
2. Mass analysis: The precursor ions are then accelerated through an electrical field and separated/deflected by their mass-to-charge-ratio in the mass analyzer, which can be a quadrupole, a quadrupole-Orbitrap hybrid system, a time-of-flight analyzer or a Fourier transform ion cyclotron resonance (FT-ICR) analyzer. For the identification of the peptide sequence, the precursor ions are fragmented by collision with neutral gas atoms in an ion trap, which creates singly charged product ions (fragment ions) that are also analyzed.
3. Detection: The precursor and product ions hit the detector, which amplifies the signal using electron multipliers. Finally, mass spectra are created.

While shotgun approaches as shown in **Fig. 7** are applied for the comprehensive profiling of biospecimens, targeted proteomics aims for the identification and detection of a list of peptides of interest, which are in the focus of the research topic. For targeted proteomics pre-existing information on the charge, mass-to-charge-ratio, retention time and tryptic features of the peptides of interest are used for their specific and sensitive monitoring (29). Both targeted acquisition methods Parallel Reaction Monitoring (PRM) and Selected Reaction Monitoring (SRM) apply quantification of tracked product ions on the MS2 level, but differ in their instrumentation and pre-required information about the peptides of interest. In contrast to SRM, PRM analysis does not require m/z value information of product ions, since a parallel acquisition is possible (29).

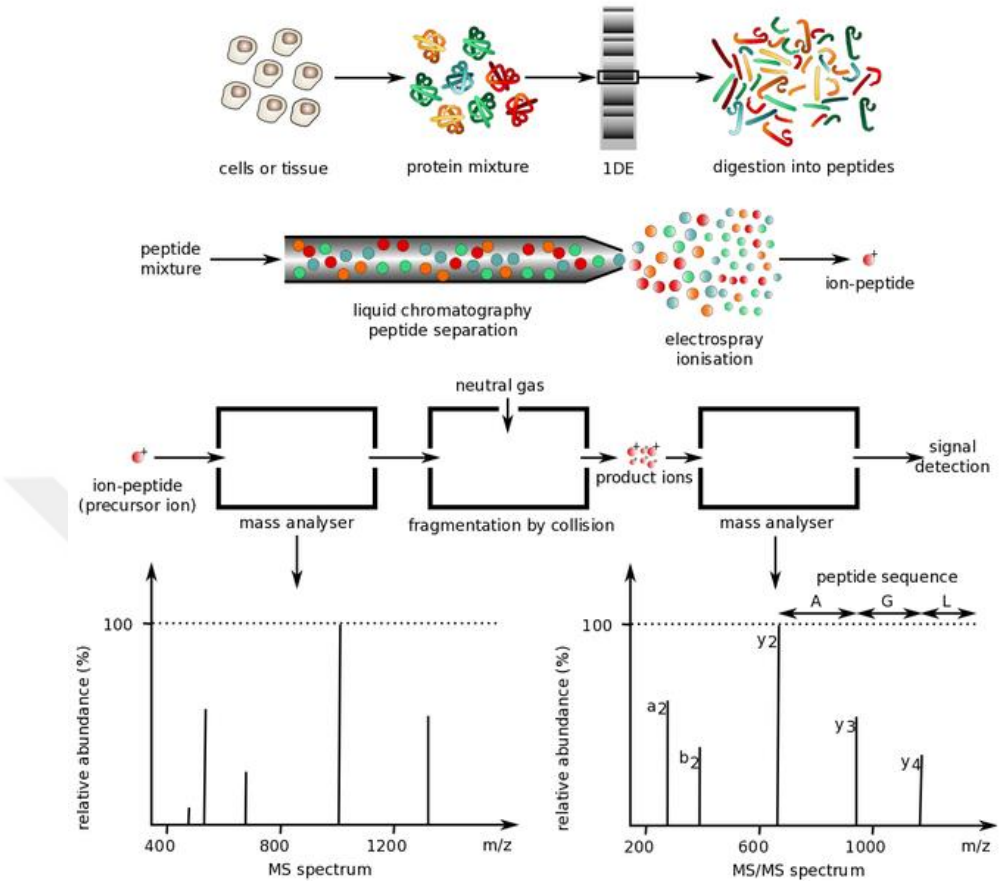


**Figure 7 | Scheme of bottom-up shotgun proteomics approach for the characterization of complex biological samples.** Full scan and corresponding fragment ion spectra are generated for the comprehensive characterization of peptide mixtures (adapted from https://en.wikipedia.org/wiki/Protein_mass_spectrometry).

### 2.1.9 ccRCC biomarkers

#### 2.1.9.1 Tissue biomarkers

Despite their urgent need, no universal biomarkers of clinical utility are currently known for ccRCC (4). This is due to previous studies lacking in coverage depth with generally less than 2,000 protein identifications and presenting insufficient validation of the candidates by multiple orthogonal methods or in independent cohorts (**Table 1**) (5-11).

**Table 1 | Overview of the most noticeable proteomics studies for the identification of ccRCC biomarkers.**

| Reference | #patients | Experimental details | #identified proteins | Suggested biomarker | Type of validation |
|---|---|---|---|---|---|
| **Atrih *et al.* (2014) (5)** | 8 | LFQ, frozen normal/tumor tissues used | 1,761 | CORO1A, ADFP | IHC |
| **Masui *et al.* (2013) (11)** | 6 | iTRAQ, frozen primary & metastatic tissues used | 1,256 | GAL-1, PFN-1, 14-3-3ζ | WB, IHC |
| **Perroud *et al.* (2009) (8)** | 50 | LFQ, grade-dependent, FFPE normal/tumor tissues used | 777 | ALDOA, ALDH1A1, PGK1, AIFM1, SERPINH1 | WB, IHC |
| **Raimondo *et al.* (2015) (6)** | 5 | LFQ, frozen normal/tumor tissues used, membrane microdomain focused analysis | 1,463 | CA2, CD13, ANXA2 | WB |
| **Teng *et al.* (2011) (30)** | 10 | LFQ, secreted proteins, tissue interstitial fluid used | 539 | NNMT, TSP1, ENO2, ANXA4, CD14, Gal-1, TBG | WB, SRM, ELISA |
| **Neely *et al.* (2016) (7)** | 84 | LFQ, frozen normal/tumor tissues used | 1,551 | CFL-1, PFN-1, NNMT, ALDOA | IHC |
| **Drendel *et al.* (2017) (9)** | 8 | Dimethyl labeling, FFPE normal/tumor tissues from VHL patients used | 1,716 | XPNPEP1, XPNPEP2 | IHC, WB, shRNA |
| **Zhang *et al.* (2017) (10)** | 135 | LFQ, frozen normal/tumor tissues used | 1,099 | PGRMC1 | WB, IHC, ELISA |

#### 2.1.9.2 Urine biomarkers

Most biomarker candidates are restricted to the tissue context, however, biomarkers for easily available and non-invasive biospecimens such as body fluids are of more practical utility in the clinical setting (27, 31). Urine has the advantage over blood to be collectable in large amounts, to be less prone to proteolytic degradation and to be less complex, however, reliable urine biomarkers for ccRCC are still scarce (31-34). Furthermore, it has a narrower dynamic range of abundance and allows the detection of low-abundant proteins, since urine does not contain extraordinarily high abundant proteins such as albumin or transferrin in blood (20).

Few undertakings attempted to find urinary biomarkers, such as Sandim and colleagues (32), who found the proteins CO3, FIBG, MGAT4A, and APO1A to be solely present in ccRCC urine samples, but not in healthy control samples. Also, extracellular exosomes were analyzed in urine samples of ccRCC and healthy controls, which confirmed the presence of several extracellular vesicle markers in the samples and revealed the proteins CP, MMP9, PODXL, CAIX, and DKK4 to be increased in tumor exosomes, while CD10, EMMPRIN, DPEP1, SDCB1, and AQP1 were decreased (20).

## 2.2 MATERIALS & METHODS

### 2.2.1 Ethics statement

The study proposal was approved by the Ethics Committee of Koç University in September 2017 (no. 2017.145.IRGB2.051). The approval was prolonged for 2 more years and was expanded for urine sample collection in 2019. Tissue and urine samples were collected with the patients' informed written consent following the guidelines of the Declaration of Helsinki.

### 2.2.2 Sample collection

Tumor and NAT samples were collected during partial or radical nephrectomy in two hospitals in Istanbul (Koç University School of Medicine and Istanbul University Istanbul Faculty of Medicine). Samples were immediately stored at -80°C after collection. The age of the patients varied between 28 and 74 and reflected a typical gender representation for RCC with a 2.6:1 male-to-female ratio. Urine samples were collected pre-operatively from ccRCC patients and from a tumor-free control group in the Urology Department of the Koç University School of Medicine. Samples were supplemented with complete EDTA-free protease inhibitor (0.5 tablet for 5 mL, Thermo Fisher Scientific) and immediately stored at -80°C. Detailed clinical information about the patient cohort is summarized in **Table A1** and **Table A2**.

### 2.2.3 In-solution protein digest of tissue samples

Normal adjacent and tumor tissue samples from 13 patients (discovery cohort) were subjected to protein isolation using 8 M urea, 1 mM sodium orthovanadate, complete EDTA-free protease inhibitor mixture, phosSTOP phosphatase inhibitor mixture (0.5 tablet for 5 mL, Roche) and 1% n-octylglucoside in 50 mM ammonium bicarbonate

(ABC, pH 8.0). For extensive protein extraction, 20-30 mg of tissue samples were homogenized with the Bullet Blender Tissue Homogenizer (Next Advance) using 0.5 mm zirconium oxide beads. The cells were further disrupted by passing the sample through a fine syringe needle (26G) 10 times on ice. The protein concentration was measured using the BCA assay (Pierce, Thermo Fisher Scientific), followed by reduction of disulfide bonds with 10 mM dithiothreitol (DTT) for 1 h at 56°C and cysteine alkylation with 20 mM iodoacetamide (IAA) for 45 min in the dark at room temperature (RT). The urea concentration was reduced to 1 M with 50 mM ABC and trypsin was added at 1:50 enzyme:protein ratio. The protein digest took place at 37°C overnight and was quenched by acidification using 10% formic acid (FA). The protein samples were desalted on SepPak C18 cartridges (Waters) using 0.1% FA for washing (RP A solution) and 0.1% FA in 80% acetonitrile (ACN) for elution (RP B solution).

### 2.2.4 Isotopic labeling of peptides

Peptide samples were labeled at their primary amines with stable dimethyl isotopes according to the protocol described by Boersema and colleagues (35). For light labeling 4% formaldehyde solution ($CH_2O$) and 0.6 M cyanoborohydride solution ($NaBH_3CN$) were mixed in 50 mM sodium phosphate buffer (pH 7.5), while for heavy labeling 4% $^{13}CD_2$-labeled formaldehyde solution ($^{13}CD_2O$) and 0.6 M cyanoborodeuteride solution ($NaBD_3CN$) were mixed. For seven patients, the NAT samples were light labeled while the tumor samples were heavy labeled. For the remaining six patients, the labeling was swapped in order to eliminate bias from the label choice. A total of 600-700 µg of peptides were loaded on 1cc SepPak C18 cartridges and flushed with the prepared light and heavy labeling solution, respectively. The labeled peptides were washed twice with RP A solution and eluted with RP B solution.

### 2.2.5 Strong Cation Exchange fractionation of peptides

The peptide mixtures were then fractionated on-column by Strong Cation Exchange (SCX) chromatography. For this purpose, the dried peptide samples were reconstituted in SCX loading buffer (7 mM $KH_2PO_4$, 30% ACN, pH 2.65) and light and heavy labeled peptides were mixed at 1:1 ratio based on their median peptide intensities. The SCX resin packed cartridge (HyperSep, Thermo Fisher Scientific) was prepared by sequential addition of following solutions: methanol, elution buffer F9, HPLC-grade water, equilibration buffer (50 mM $K_2HPO_4$, 500 mM NaCl, pH 7.5), HPLC-grade water and loading buffer. The bound peptides were eluted by step-wise addition of following elution buffers which consisted of the loading buffer with increasing KCl concentrations (F1: 30 mM; F2: 45 mM; F3: 60 mM; F4: 75 mM; F5: 90 mM; F6: 105 mM; F7: 120 mM; F8: 150 mM and F9: 500 mM). The obtained ten fractions per patient (flow-through included) were extensively desalted using 1cc SepPak columns.

### 2.2.6 Data acquisition

#### 2.2.6.1 Global proteomics

The collected SCX fractions were reconstituted in 5% FA and 5% ACN (MS analysis solution) and analyzed in duplicate with 120 min linear gradients on an UltiMate 3000 RSLCnano reversed phase chromatographic platform (Thermo Fisher Scientific) coupled to a Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). Approximately 300 ng of each fraction was loaded onto an in-house packed 100 µm i.d. × 17 cm C18 column (Reprosil-Gold C18, 5 µm, 200Å, Dr. Maisch) and run with a flow rate of 300 nL/min. The chromatographic separation of the peptides started at 4% of solution B (0.1% FA in ACN) and gradually increased to 25% in 69 min. The gradient continued from 25% to 40% of solution B in the next 20 min. Peptides in the mass range 400−1,500 m/z and with a positive polarity were allowed for detection in data-dependent acquisition mode (DDA). For the MS1 spectra acquisition the resolution was

set to 70,000, the automatic gain control (AGC) target to 1e6 and the maximum injection time to 32 ms. The top 15 most intense peptides per cycle were selected for fragmentation in the higher-energy collisional dissociation (HCD) cell with a normalized collision energy (NCE) of 26. MS2 spectra acquisition was conducted at a resolution of 17,500, an AGC target of 1e6, a maximum injection time of 85 ms and a fixed first mass of 120 m/z. Furthermore, the isolation window was set to 2.0 m/z, the dynamic exclusion was set to 35 s and the charge exclusion was set as unassigned, 1, 6-8, >8.

#### 2.2.6.2 Targeted proteomics

For the targeted proteomics approach, unlabeled digested tissue samples of patients were analyzed in PRM mode. Samples were reconstituted in MS analysis solution and run with a 90 min linear gradient on a similar system as described above except that a Q Exactive HF mass spectrometer was engaged. Approximately 600 ng were loaded onto an in-house packed 100 µm i.d. × 25 cm C18 column (Reprosil-Gold C18, 3 µm, 200Å, Dr. Maisch) and run with a flow rate of 300 nL/min. The gradient started at 4% of solution B, increased to 25% in 49 min and continued from 25% to 45% of solution B in the next 15 min. Peptides with a positive polarity were targeted. MS2 spectra acquisition was performed at a resolution of 45,000, an AGC target of 2e5, a maximum injection time of 100 ms and a fixed first mass of 100 m/z. Moreover, the isolation window was set to 1.2 m/z and the NCE to 28.

### 2.2.7 Data processing

Peptide identification and quantification was done with Proteome Discoverer (PD) (v1.4, Thermo Fisher Scientific) using the SequestHT search engine. Raw files were analyzed in the "MudPIT" processing mode via the Daemon utility of PD to merge the ten fraction files from a technical replicate. Peptide spectral matches (PSM) were searched against a Swissprot database containing 21,039 entries for *Homo sapiens* retrieved from Uniprot in March 2016. Trypsin was selected as hydrolytic enzyme with a maximum number of

allowed missed cleavages of 2. Regarding peptide identification, a mass tolerance of ±20 ppm for precursor masses and ±0.05 Da for fragment ions were selected. For quantitation, the 2plex dimethyl heavy/light method with a mass precision of 2 ppm for precursor measurements was used. Light and heavy dimethylation of peptide N termini and of lysine residues, as well as methionine oxidation, were set as dynamic modifications. Cysteine carbamidomethylation was set as fixed modification. The False Discovery Rates (FDR) for peptide and protein identifications were set to 1%. Only peptides with medium or high identification confidence, with a sequence length between 7 and 25 and with a peptide rank of minimum 1 were allowed.

### 2.2.8 Normalization and filtration of data

Quantification ratios of samples with labeling swap were converted to heavy/light (H/L) format. H/L ratios for biological replicates were average values from two technical replicates. The obtained data was filtered for proteins quantified in at least one of the 13 patients (5,799 proteins), which are hereafter denoted as "quantified proteins". The H/L values were sample-wise normalized by the sample median and $\log_2$ transformed.

### 2.2.9 Urine sample preparation

First, the pH value of urine samples was adjusted to pH 8.0 using 500 mM ABC supplemented with EDTA-free protease inhibitor and then, the biospecimen was filtered through an Amicon Ultra-0.5 mL Centrifugal Filter Device (10 kDa, Merck Millipore). Collection of the retentate (proteins in approximately 100 µL 50mM ABC) was performed by flipping the filter device upside-down and centrifugation at 1,000*g* for 2 min. The protein concentration was determined using the BCA assay.

### 2.2.10 Immunoblotting

For immunoblot analysis 40 mg of normal tissue and 70-80 mg of tumor tissue were lysed in 0.1% Triton-X in 1xPBS and EDTA-free protease inhibitor. In total, 40 µg of tissue samples and 20 µg of urine samples were run on 8-10% SDS gels and transferred to nitrocellulose membranes for 3-4 h at RT. Primary antibody incubation was done overnight at 4°C with following dilutions: anti-PLOD2 mouse at 1:500-1:1,000 (MAB4445, R&D Systems), anti-FERMT3 mouse at 1:500-1:1,500 (ab173416, Abcam), anti-SPARC mouse at 1:1,000 (33-5500, Invitrogen), anti-SIRPα rabbit at 1:1,500-1:3,500 (13379, Cell Signaling Technology) and anti-ACTB mouse at 1:5,000 (ab6276, Abcam). As secondary antibodies HRP-conjugated anti-mouse IgG at 1:2,000 (7076S, Cell Signaling Technologies) and HRP-conjugated anti-rabbit IgG at 1:2,000 (7074S, Cell Signaling Technologies) were used. Protein expression was visualized using Clarity Western ECL Substrate (Bio-Rad). Densitometric analysis was performed with ImageJ (v1.46r), and visualization and statistical analysis (Student's *t*-test *p*-value <0.05) with GraphPad PRISM v8, respectively.

### 2.2.11 Parallel Reaction Monitoring

Selected unlabeled peptides from digest experiments were targeted by PRM as quantifiable surrogates for the proteins of interest. Candidate peptides were selected as follows: uniqueness of peptides was ensured by BLASTp analysis, peptides with length between 7-18 aa were preferred, while peptides with unstable charge state across samples, with ragged ends, with missed cleavages and/or with post-translational modifications, respectively, were avoided. When no alternative peptides were available for a protein, exceptions to these exclusion criteria were made. No exception was made for the charge state criteria. MS/MS spectra were inspected extensively and peaks with a mass window <10 ppm were integrated by manually setting the borders in Skyline (v19.1). The spectral library consisted of 60,833 spectra from DDA-based data acquisition of tumor and normal digest samples of the discovery cohort. Doubly and triply charged peptides in the range

of 375-1,500 m/z with a library ion match tolerance of 0.5 m/z were accepted. Only spectra with a high correlation with the spectral library (dotp value >0.8) were considered. For each peptide, the summed peak area of its six fragment ions was calculated and then normalized by the summed peak area of the β-actin peptide GYSFTTTAER in the sample. Calculated tumor/normal peak area ratios (**Fig. 8**) were transformed to $\log_2$.

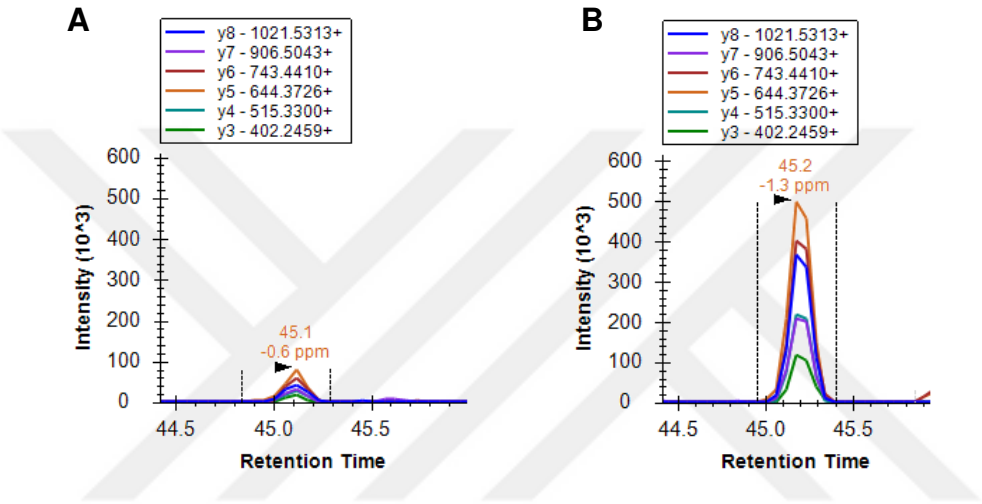


**Figure 8 | Targeted proteomics analysis of tissues using Parallel Reaction Monitoring.** Exemplary images illustrate peak area difference of same peptide in a normal adjacent tissue sample **(A)** and the corresponding tumor tissue sample **(B)** (shown peptide is SEDYVELVQR of PLOD3). The color labeling indicates the six most abundant fragment ions of the peptide.

### 2.2.12 Immunohistochemistry

Immunohistochemical analyses (IHC) were performed with the BenchMark Ultra autostaining platform (Ventana Medical Systems). Briefly, the tissue paraffin sections were cut to 3 µm thickness onto charged slides. Sections were deparaffinized and rehydrated through an alcohol series. Tissue sections were incubated with primary antibodies at 37°C with following dilutions: anti-SPARC mouse at 1:800 (33-5500, Invitrogen), anti-SIRPα rabbit at 1:25 (13379, Cell Signaling Technology), anti-FERMT3 mouse at 1:2,500 (ab173416, Abcam) and anti-PLOD2 mouse at 1:250 (MAB4445, R&D

Systems). The UltraView Detection kit (760-500, Ventana Medical Systems) was used for detection of protein abundance. The reaction product was visualized with 3, 3′-diaminobenzidine (DAB) chromogen and counterstained with hematoxylin. All stained slides were evaluated in a blinded fashion by a single pathologist. The intensity was scored as mild (score 1: any positivity that could be seen at high magnification), moderate (score 2: any staining between score 1 and 3), and marked (score 3: any staining that could be seen at low magnification). The prevalence of the staining was evaluated as score 0 (<5%), score 1 (5%-25%), score 2 (25%-75%) and score 3 (>75%).

### 2.2.13 Statistical analysis

Significantly dysregulated proteins between tumor and normal tissues were determined using the one-sample Wilcoxon signed-rank test (with 0 as reference $\log_2$ fold change value). For this purpose, the Python function scipy.stats.wilcoxon (v0.14.0) was extended by the feature to omit missing quantification values and was then applied gene-wise on the data. Benjamini-Hochberg adjusted *p*-values <0.05 were considered statistically significant.

### 2.2.14 Functional annotation

Significantly regulated proteins were annotated by their role in cancer according to the COSMIC Cancer Gene Census (CGC) database (v91). For Gene Ontology (GO) cellular component and for pathway annotations, over-representation analysis (ORA) was applied using the WebGestalt platform (v2019) (36) with a BH-adjusted FDR of 0.02 and 0.05, respectively. For GO biological process annotation of the significantly regulated proteins, the PANTHER (v14) platform (37) was used with an FDR setting of 0.05.

### 2.2.15 Network-based interpretation of proteomics data

Optimal subnetworks, which represent the list of significant proteins best, were reconstructed by integrating a reference human interactome with our proteomic data using the Forest module of Omics Integrator (v0.3.1) (38) and the visualization tool Cytoscape (v3.8.0) (39). The iRefWeb_2013 interactome file was used as global network reference. Nodes were annotated by GO cellular compartment using QuickGO (40), which was limited to Uniprot assignments. For multiple allocations, priority was given as follows: “extracellular”, “membrane”, “nucleus”, “cytoplasm”, “other” or “unknown”. Additionally, networks were annotated with GO biological process associations using the Cytoscape plug-in ClueGO (v2.5.7) (41).

## 2.3 RESULTS

### 2.3.1 Shotgun proteomics approach achieves deep coverage of ccRCC proteome

In this study, a comprehensive quantitative proteome analysis of ccRCC has been conducted to address the scarcity of reliable biomarker panels and to uncover underlying molecular phenotypes. To this end, dimethylation-based quantitative proteomics analysis with extensive homogenization of biospecimens has been performed with 13 ccRCC patient's frozen tumor and normal adjacent tissue samples (**Fig. 9**). The discovery cohort included patients of different age, grade, stage and gender (**Table A1**).

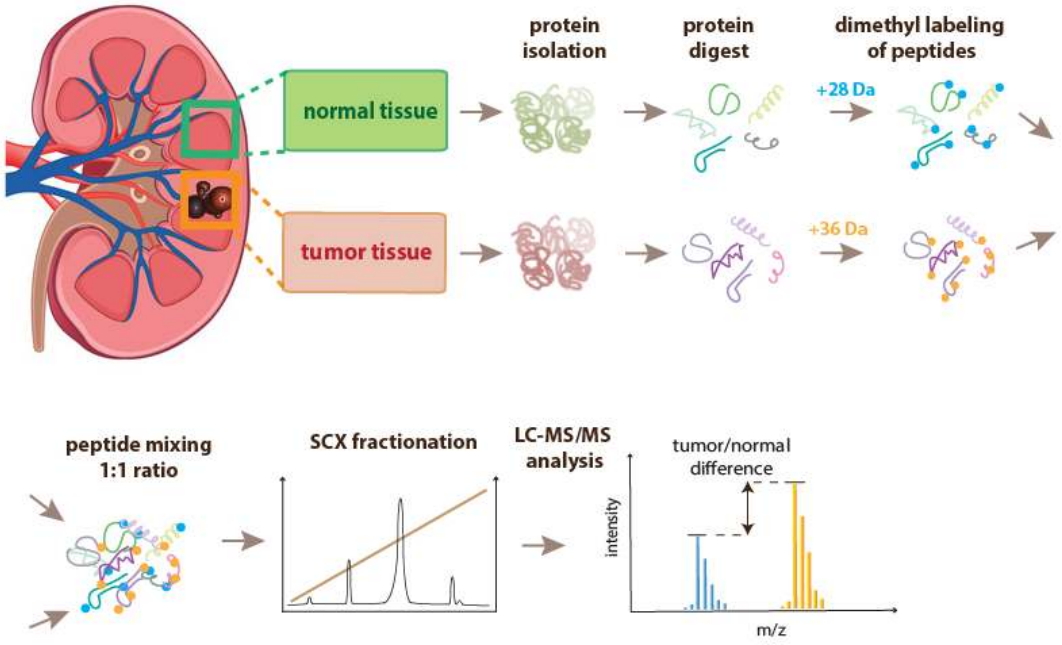


**Figure 9 | Experimental outline of shotgun proteomics approach.**

We identified a total of 10,160 unique proteins from all samples with an FDR of 0.01. Protein identification and quantification were robust across patients with a coefficient of variation of approximately 6.7% for the identifications and 8.5% for the quantifications, respectively (**Fig. 10**). On average, 4,453 proteins were identified and 3,118 proteins were quantified in the paired samples, respectively.

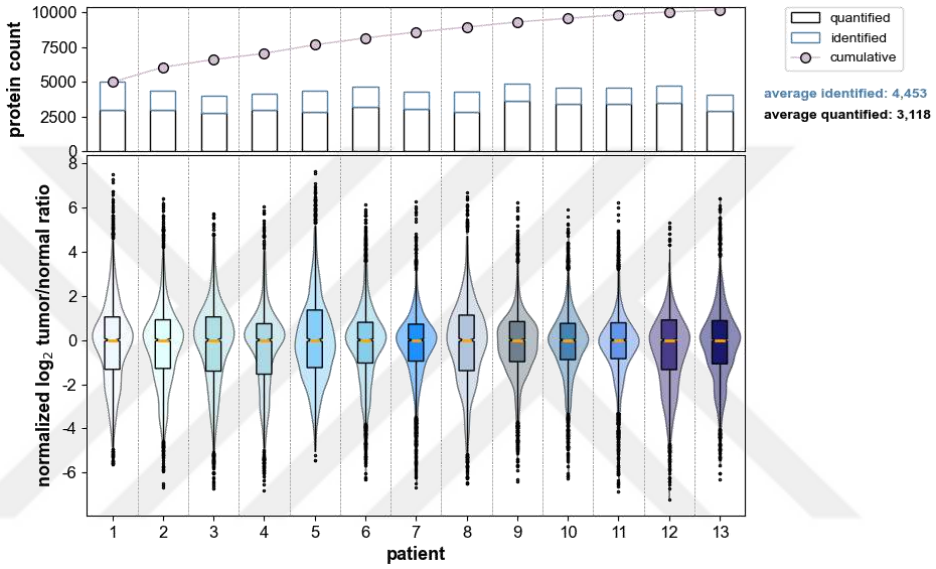


**Figure 10 | Number and distribution of identified and quantified proteins across all patients.** For cohort information see Table A1.

Abundance levels of the ccRCC proteome encompassed almost six orders of magnitude in dynamic range and correlated with diverse cellular compartments (**Fig. 11**). The most abundant proteins were associated with angiogenesis such as blood coagulation and platelet degranulation, and with translational processes such as chaperone activity and protein maturation in the ER. Notably, the low-abundance range was characterized by proteins predominantly localized to the mitochondrial matrix and mitochondrial inner membrane.

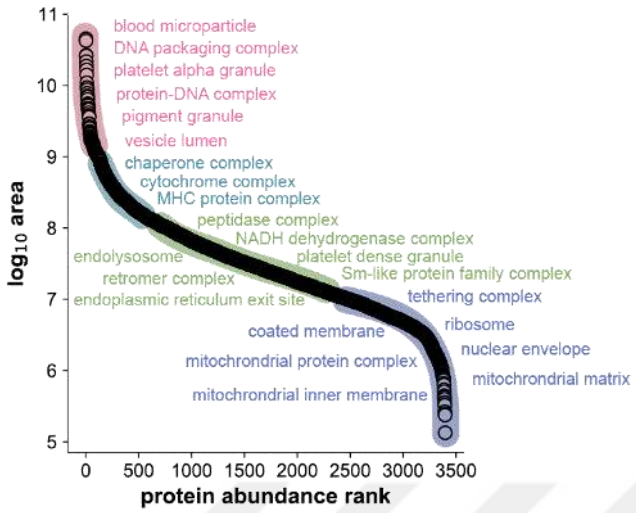


**Figure 11 | S curve showing associated cellular compartments of identified proteins, which were subdivided into four $log_{10}$ area groups.**

In general, the distribution of quantified proteins followed a near normal pattern (**Fig. 12A**, top histogram) over more than 14 orders of magnitude with a noticeable skewness towards positive fold change values suggesting more pronunciation of overexpressed processes in ccRCC tumors compared to normal tissues (**Fig. 12A** and **B**).

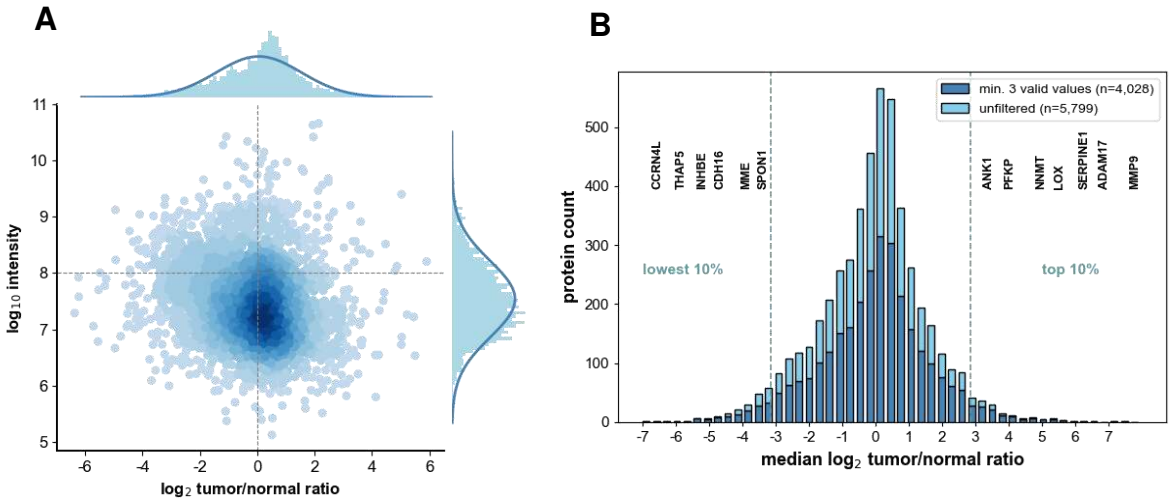


**Figure 12 | Distribution of quantified proteins. (A)** MA plot illustrating intensity and median

fold change distributions of quantified proteins. Top and right histograms show distribution of fold change values and intensities, respectively, with density lines representing theoretical normal distribution. **(B)** Histogram showing general distribution of median tumor/normal ratios of filtered and unfiltered quantified proteins.

### 2.3.2 Significantly regulated proteins mirror RCC-typical metabolic alterations and molecular characteristics

In total, 5,799 of the 10,160 identified proteins were quantified by dimethyl labeling (**Fig. 13**). We determined 955 proteins as statistically differentially expressed between tumors and NATs (BH-adjusted *p*-value <0.05) with 409 significantly up-regulated and 546 significantly down-regulated proteins.

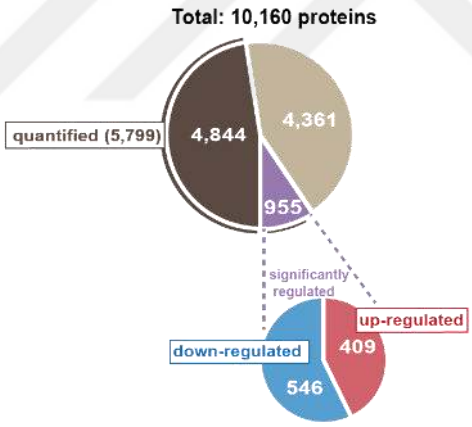


**Figure 13 | Significantly dysregulated proteins in ccRCC tumors compared to adjacent normal tissues.** Overview of the number of identified and quantified proteins. Inner pie chart represents number of significantly regulated proteins.

Among these were 44 genes associated with mutations that are causally implicated in cancer, such as oncogenes, tumor suppressor genes and fusion genes, according to the COSMIC CGC database (**Fig. 14A** and **B**).

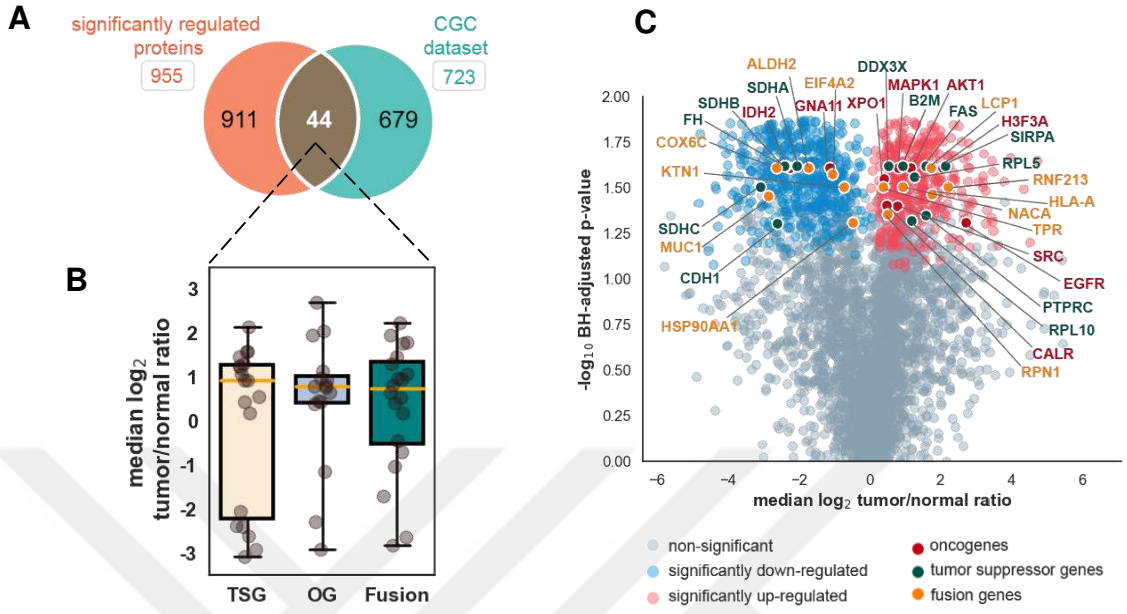


**Figure 14 | Causally implicated cancer genes in ccRCC tumors. (A)** Venn diagram depicting number of cancer genes among the significantly regulated proteins in ccRCC tumors retrieved from the COSMIC Cancer Gene Census database. **(B)** Distribution of significantly regulated cancer genes categorized by their function. Cancer genes might be allocated to multiple functional categories. **(C)** Volcano plot showing distribution of significantly regulated proteins with indication of gene functionality in cancer.

Significantly elevated oncogenes were EGFR, H3F3A, AKT1, MAPK1, SRC, CALR and XPO1 (**Fig. 14C**). The activation of the EGFR-MAPK1 axis has previously been reported to enhance cell proliferation in RCC (15, 42). Also, our data implicated that the EGFR-PI3K-AKT1 and AKT1 related mTOR signaling pathways might be activated in ccRCC, which are both associated with tumor growth and survival (43). Another significantly up-regulated oncogene is the tyrosine kinase SRC, which has previously been reported to be recruited by EGFR and to affect its downstream players RAS and AKT1 in RCC (44). Despite its categorization as a tumor suppressor, fatty acid synthase (FAS) was highly up-

regulated in the tumor tissues relative to NATs, which has previously been associated with tumor aggressiveness and poor patient outcome in RCC (45). We inferred that several oncogenes possibly mediate tumor proliferation and survival in ccRCC tumors, emphasizing the need for further investigation especially of the EGFR related molecular mechanisms in respect of alternative directed therapeutic intervention.

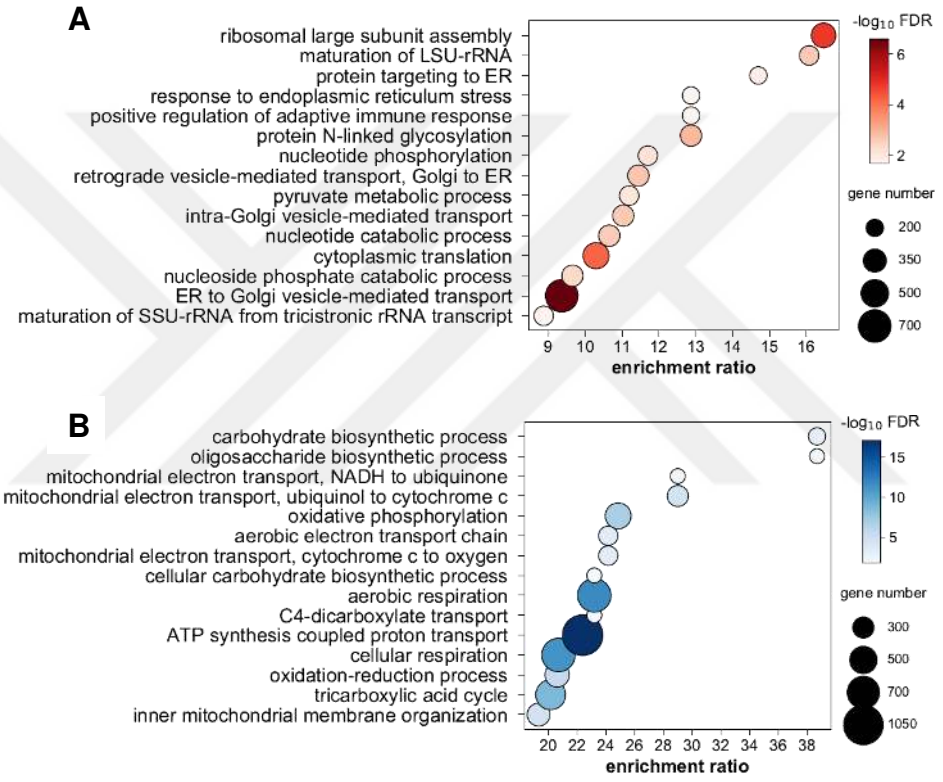


**Figure 15 | PANTHER GO-Slim biological process analyses of significantly up-regulated (A) and down-regulated (B) proteins, respectively.**

As many cancers, RCC undergoes substantial metabolic reprogramming by increasing glycolytic processes and diminishing mitochondrial ATP production to meet its excessive energy requirements for tumor growth and progression (17, 45). Our findings for the significantly altered biological processes were in agreement with the reported metabolic

pathobiology of ccRCC. The significantly up-regulated proteins were primarily associated with protein maturation and transport, PTM events such as glycosylation and phosphorylation of macromolecules, glycolysis-related processes and immune response (**Fig. 15A**). In contrast, the significantly down-regulated proteins were mainly involved in mitochondrial events such as oxidative phosphorylation, electron transport chain, proton transport and TCA cycle (**Fig. 15B**).

In order to unveil the most prominent cross-talks between significantly up-regulated proteins, we constructed a protein-protein interaction subnetwork using Omics Integrator (38), which confirmed that the highly elevated oncogenes EGFR and SRC were central players in ccRCC tumorigenesis (**Fig. 16** and **Fig. A3**). While most of the hubs were associated with intracellular transport of macromolecules (SRC, RPS4X, SUMO1 and YWHAZ) and organelle organization (UBQLN4, **Fig. A3**), EGFR was associated with immune system related processes and molecule secretion (**Fig. 16**). SUMOylation modifications by SUMO1 have previously been reported to be elevated in a subgroup of low-grade ccRCC tumors, which were particularly associated with angiogenesis and absence of infiltrating immune cells (15). Furthermore, the regulatory protein YWHAZ (14-3-3ζ) has previously been noted to play a promoting role in primary as well as metastatic ccRCC (7, 8, 11).

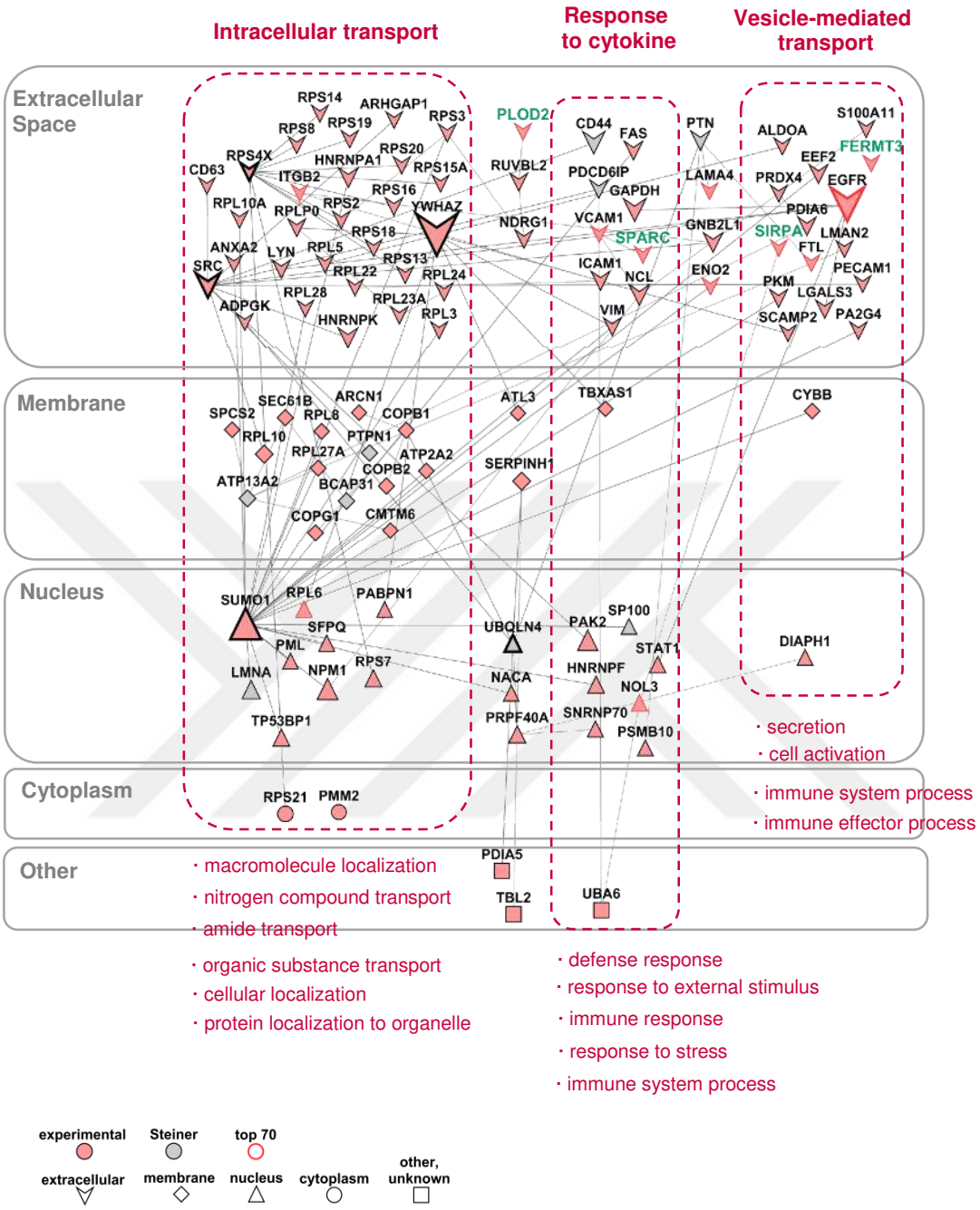


**Figure 16 | Protein-protein interaction network of significantly up-regulated proteins.** Edge annotations were created using Omics Integrator by integrating intermediate interactors (Steiner nodes). Associated cellular compartment and biological process information of proteins were retrieved using QuickGO and ClueGO, respectively. Size and border width of nodes represent BetweennessCentrality. Proteins selected as biomarker candidates are indicated with green label. Network is summary of most significant pathways, for full network see Fig. A3.

### 2.3.3 Biomarker candidates PLOD2, FERMT3, SPARC and SIRPα are highly expressed in ccRCC tumors

Given that secretion of macromolecules is essential for sculpting the tumor microenvironment to promote tumor growth and metastasis, and that vesicle-mediated transport and secretion were one of the significantly enriched biological processes in the ccRCC tumors (**Fig. 16**), we sought to identify and characterize novel secreted biomarkers. To this end, we selected four biomarkers among the top 70 highly up-regulated proteins predicted to be located in the "extracellular space". To the best of our knowledge, the selected biomarkers, namely PLOD2, FERMT3, SPARC and SIRPα, have previously not been suggested or validated as ccRCC biomarkers in proteomics-based studies. According to the CGC annotation, SIRPα has a tumor suppressor role in cancer (**Fig. 14C**), however, in this study, SIRPα was with a median $\log_2$ tumor/normal ratio of 2.15 clearly an up-regulated protein in the ccRCC tumor tissues.

Immunoblot analysis confirmed the elevated expression of all four biomarkers in the tumors compared to the adjacent normal tissues (**Fig. 17**). This was observed in grade 2, grade 3 as well as grade 4 patients.

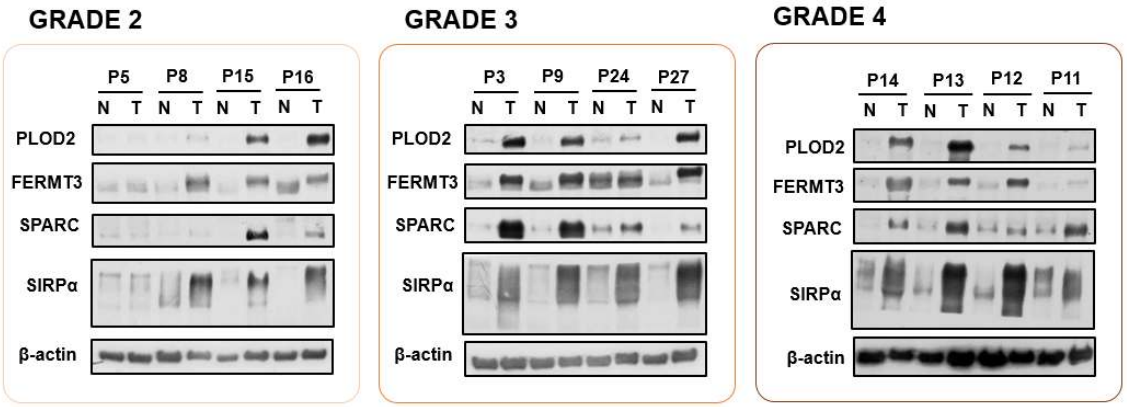


**Figure 17 | Immunoblot analysis of biomarkers in tissues of patients with different grades.**
N: normal adjacent tissue, T: tumor tissue. For cohort information see Table A1 and Table A2.

Next, we attempted to validate the biomarkers by targeted proteomics in digested unlabeled tissue samples. Parallel Reaction Monitoring allowed tracking of selected peptides, which served as surrogates for the proteins of interest. Besides the four selected biomarkers, six further putatively secreted proteins of interest were monitored via PRM (**Fig. 18** and **Fig. 19**). As expected, the selected 20 targets were more abundant in tumor samples compared to NATs and had comparable tumor/normal ratios across grades. We observed a significant positive Pearson correlation of 0.84 ($p$-value <0.001, **Fig. 20**) between the discovery cohort (**Fig. 18** and **Table A1**) and an independent cohort consisting of 11 additional ccRCC patients (**Fig. 19** and **Table A2**) of different grades, which supported the suitability of these proteins as potential biomarkers.

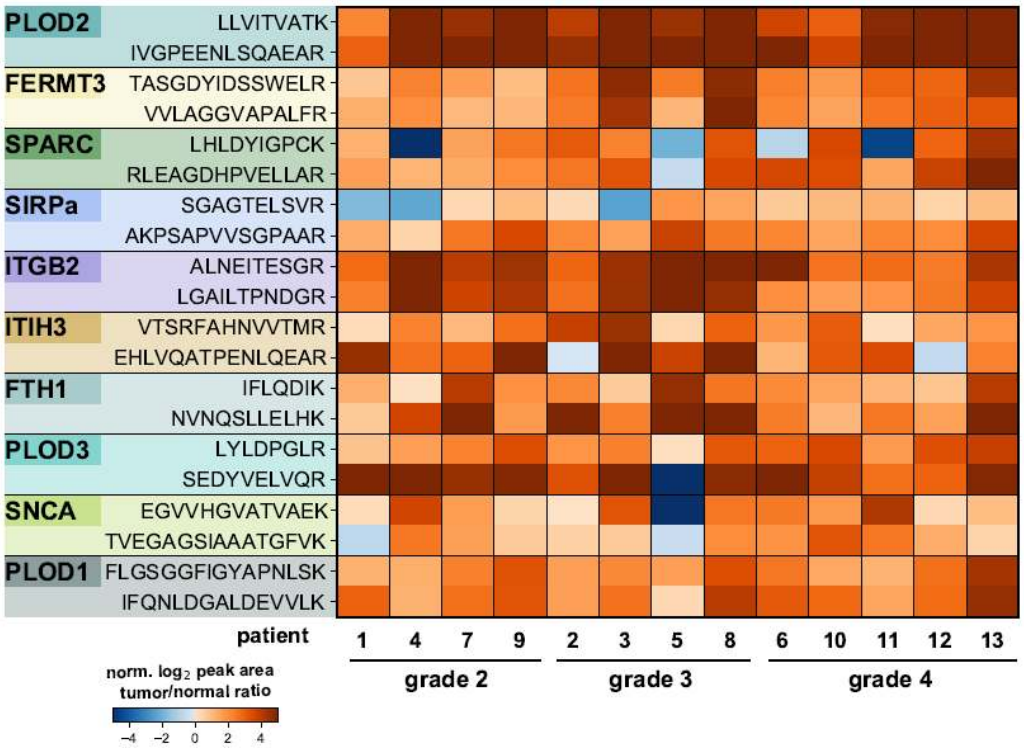


**Figure 18 | Heatmap showing $log_2$ tumor/normal ratios in PRM experiments of patient samples from discovery cohort.** For cohort information see Table A1.

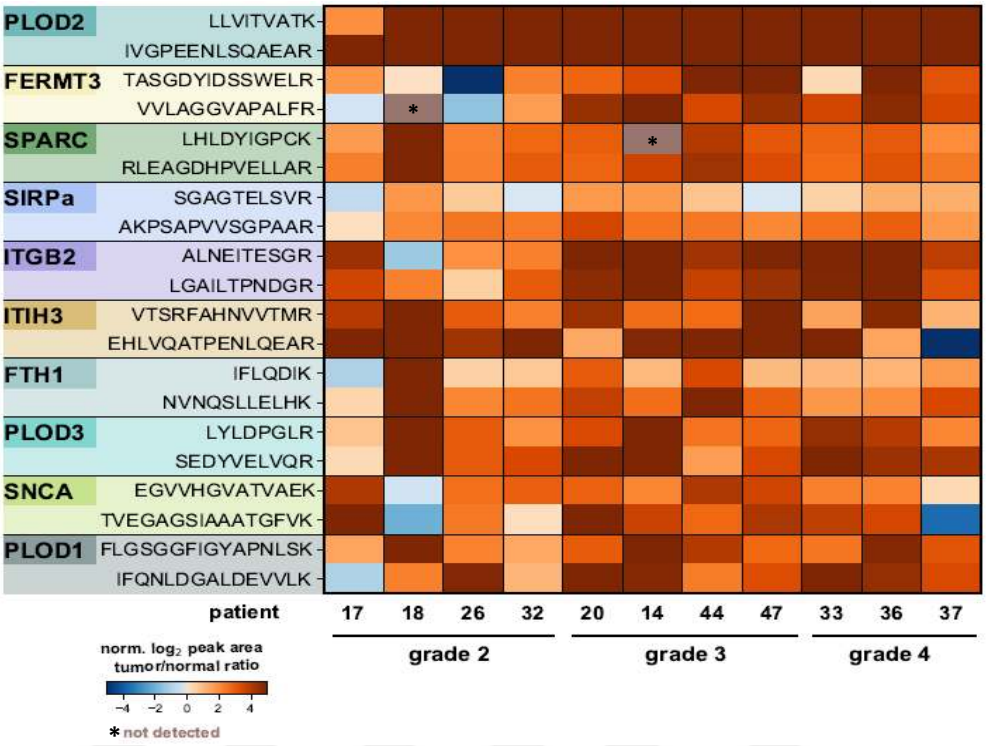


**Figure 19 | Heatmap showing $log_2$ tumor/normal ratios in PRM experiments of patient samples from independent validation cohort.** Fields with a star (*) indicate samples without peptide abundance in both tumor and NATs. For cohort information see Table A2.

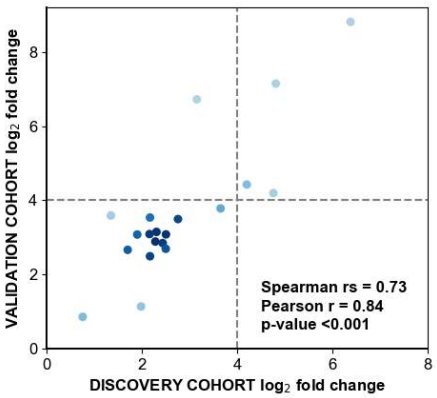


**Figure 20 | Correlation of PRM tumor/normal ratios between discovery and validation cohort.**

Next, we tested the impact of intratumoral heterogeneity on the expression of the biomarkers in random parts of tumor and adjacent normal tissues with morphological diversity from three selected patients (**Fig. 21** and **Fig. A4**).

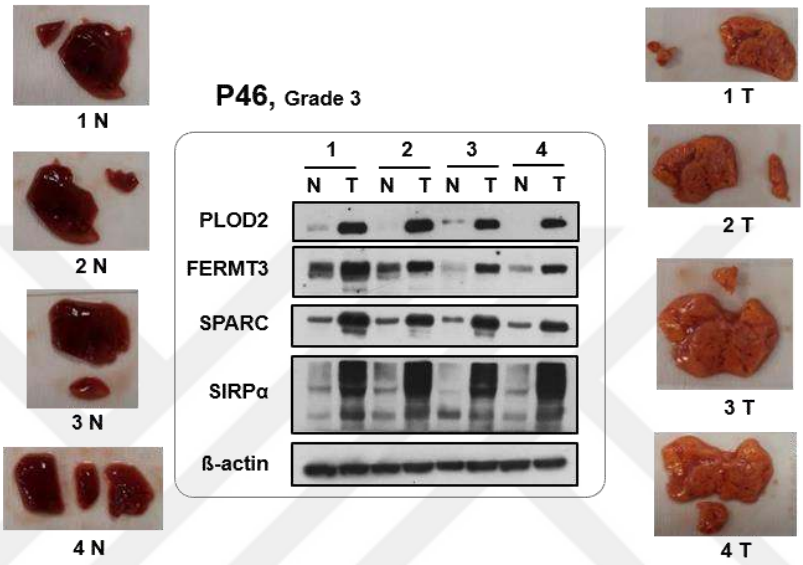


**Figure 21 | Impact of intratumoral heterogeneity on biomarkers.** Comparable expression of biomarkers in random sub regions of tissue samples from a single patient showing independence of biomarker levels from intratumor heterogeneity (experiments were performed by Ayşe Tuğçe Şahin, Koç University).

We could not observe any remarkable local distortion of abundance for the biomarkers as evidenced by a comparable expression detected in the random sections, which further qualifies them as highly expressed and robust potential biomarkers for ccRCC.

### 2.3.4 SPARC localizes to the intratumoral stroma and is significantly elevated in urine of ccRCC patients

We next investigated the expression of the four biomarkers by immunohistochemistry (**Fig. 22**). Although transitional regions illustrated that the expression of PLOD2 was more intense in tumor sections (**Fig. A1 A**), a similar prevalence (score 3, >75%) was detected in both tumor and NATs (**Fig. 22**). In contrast, SIRPα, FERMT3 and SPARC had a low prevalence score of 1 (5-25% spreading) and a low intensity score of 1 in the normal tissues, while the expression was largely distributed and markedly elevated in the tumor tissues with a prevalence score between 2 (25-75%) and 3 (>75%), and an intensity score of 3. Furthermore, for SIRPα, we determined a membranous localization in the tumor cells. Interestingly, for FERMT3 we observed almost no staining in large areas of the tumor tissues (**Fig. A1 B**), but high prevalence and intense staining in inflammatory regions (**Fig. 22**), coinciding with the previously reported role of FERMT3 in leukocyte transmigration, platelet degranulation and inflammation (46, 47).

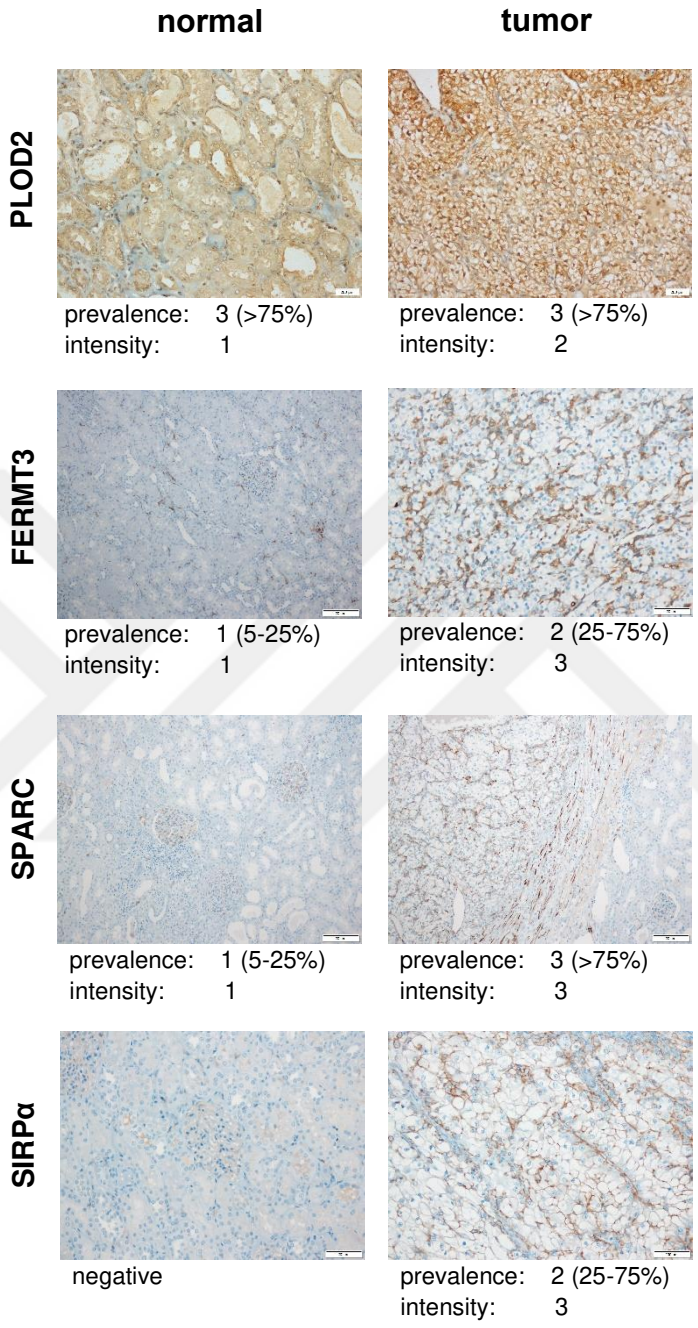


**Figure 22 | Immunohistochemistry staining of biomarker expression in normal and tumor tissues with prevalence and intensity scores.** For SPARC, tumor picture shows transitional region to adjacent normal tissue (experiments were performed by Dr. Ayşe Armutlu, Koç University School of Medicine).

We were intrigued by an intense staining (score 3) for SPARC in the intratumoral stroma, which suggested that SPARC might be a secreted protein and could possibly be detected in body fluids of ccRCC patients. To this end, we assessed the secretion of the four suggested proteins into urine, which is a readily available biospecimen and can be collected non-invasively (**Fig. 23**). We detected three of the four biomarkers in the urine samples of ccRCC patients of different grades, as well as the non-tumoral control group. While, FERMT3 and SIRPα did not display a significant expression difference, SPARC was significantly more abundant in urine samples collected from the cancer patients in comparison to the control group ($p$-value <0.05) (**Fig. 23A and B**), which supports our proteomics approach to find novel secreted biomarkers for ccRCC in urine.

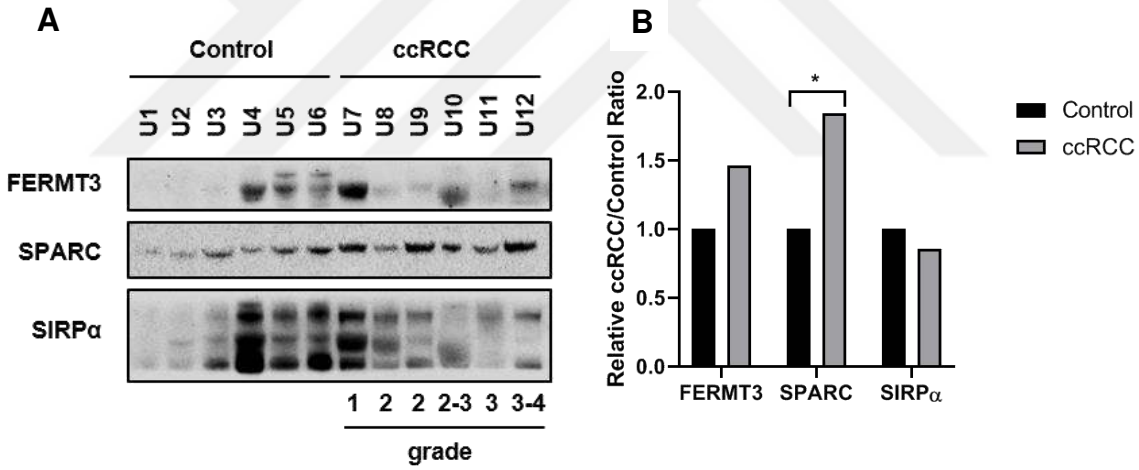


**Figure 23 | Expression of biomarkers in urine samples. (A)** Immunoblot of biomarkers. Control samples are from patients without tumors. **(B)** Quantification of biomarker expression in urine samples (experiments were performed by Ayşe Tuğçe Şahin, Koç University). For cohort information see Table A2.

## 2.4 DISCUSSION

In this study, we performed quantitative proteomics analysis of frozen ccRCC tissues and NATs and identified a total of 10,160 proteins, of which 5,799 were quantified by dimethylation (**Fig. 13**). Moreover, 955 of these proteins were determined as statistically significant between tumor and NATs. A remarkable undertaking for the comprehensive proteogenomic characterization of ccRCC was performed by the Clinical Proteomic Tumor Analysis Consortium (CPTAC), representing the deepest proteome coverage of ccRCC with a total of 11,355 proteins identified and 7,026 proteins quantified from 103 patients (15).

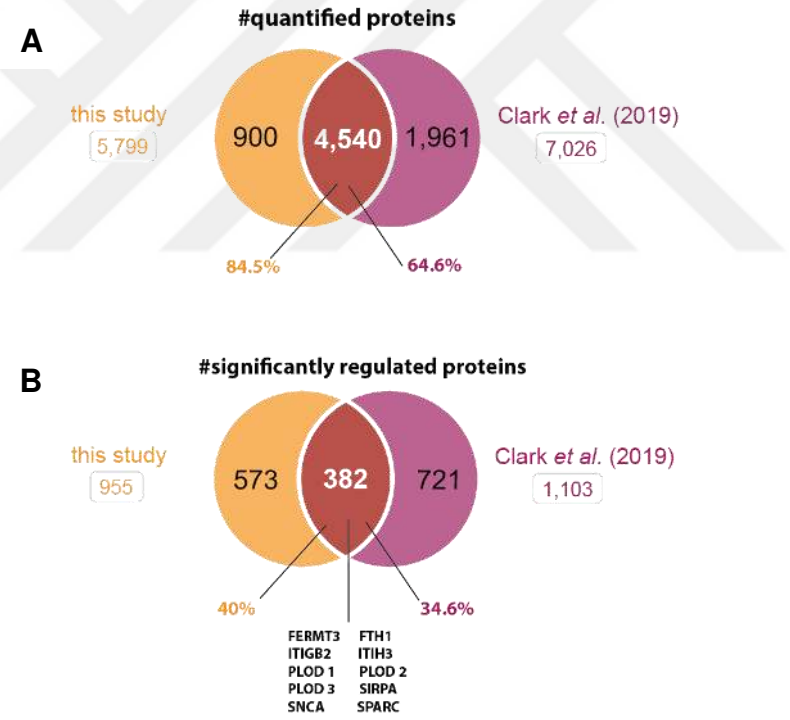


**Figure 24 | Comparison of proteome data with CPTAC study. (A)** Venn diagram showing number of common quantified proteins with CPTAC dataset. **(B)** Venn diagram showing number of common significantly regulated proteins with CPTAC dataset.

Comparing both proteome datasets revealed high conformity, with 4,540 of the quantified proteins of this study (84.5%) being shared with the CPTAC study (**Fig. 24A**) and displaying significant positive correlation ($p$-value<0.05) as indicated by each a Spearman and Pearson correlation of 0.68 (**Fig. 25**).

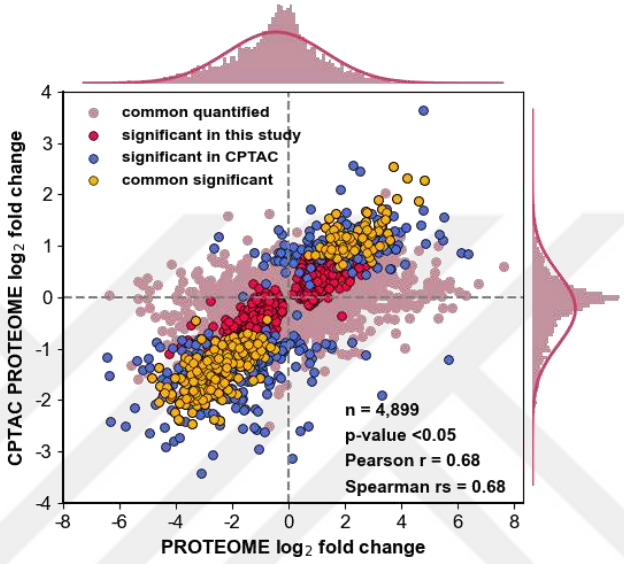


**Figure 25 | Comparison of proteome data with CPTAC study.** Correlation for all quantified proteins between proteome data and CPTAC data based on median $\log_2$ tumor/normal ratios.

A total of 382 proteins of the 1,103 significantly regulated proteins identified in the CPTAC study were also discovered in this study, validating our data (**Fig. 24B**). In addition, we identified 573 unique, differentially expressed proteins, promising a distinct proteome characterization of ccRCC compared to the CPTAC approach.

Out of the detected 409 significantly up-regulated proteins, we selected four potentially secreted biomarkers, namely PLOD2, FERMT3, SPARC and SIRPα, and confirmed their significantly elevated expression in ccRCC tumor tissues compared to NAT samples (**Fig. 17**). None of the biomarkers were significantly abundant in the proposed tumor clusters (see Chapter 3, **Fig. 34A**), which supports that these proteins are unaffected not

only by intratumor heterogeneity (**Fig. 21**), but also by interpatient tumor heterogeneity, supporting their eligibility as potential biomarkers for ccRCC. Although all four putative biomarkers are in the list of significantly up-regulated proteins in other ccRCC proteomics studies (9, 15, 48, 49), they have not been validated by other approaches orthogonal to proteomics yet.

PLOD2 (Lysyl hydroxylase 2, LH2) modifies telopeptidyl lysine residues of pro-collagen α chains on the endoplasmic reticulum and is generally regulated by HIF1α, TGF-ß and in RCC also by the tumor-suppressive microRNAs miR-26a-5p and miR-26b-5p (50-52). Aside from PLOD2, our PRM analysis also showed a higher abundance for the other known PLODs, PLOD1 (LH1) and PLOD3 (LH3) in ccRCC tumors (**Fig. 18** and **Fig. 19**). Despite its previously reported possible secretion in lung cancer (52), we could not detect PLOD2 in urine samples (**Fig. 23**), however, it is worthwhile to investigate its suitability in the clinical setting by using a larger cohort and by measuring the urinary level of its collagen degradation products.

FERMT3, also known as Kindlin-3 and URP2, functions in hemostasis and thrombosis and plays a pivotal role in integrin-mediated cell-to-cell crosstalk as well as cell-matrix junctions (53, 54). Although its role in cancer remains unclear, it has been reported to be highly expressed in different types of lymphoma and breast cancer, leading to increased tumor growth, angiogenesis and metastasis (53, 54). Interestingly, FERMT3 has been reported to act as a tumor suppressor in RCC, breast tumors and melanoma based on mRNA data (53). The ambiguous role of FERMT3 in cancer warrants the need for further investigations.

SIRPα is a transmembrane protein which interacts with the ubiquitously expressed transmembrane protein CD47 (“SIRPα-CD47 axis“) (55). This cross-talk contributes to the escape of tumor cells from the immune system response (55). Our data confirmed that SIRPα is located to the membrane in ccRCC tumor cells (**Fig. 22**), however, we could also detect SIRPα in urine samples of ccRCC patients as well as control patients, implicating that SIRPα is also a secreted protein. Given its role in immune evasion and

its high abundance in ccRCC tissues, the SIRPα-CD47 axis can be considered as another promising target for immunotherapy besides nivolumab against PD-1 or ipilimumab against CTLA-4 (56).

SPARC (Secreted Protein Acidic and Rich in Cysteine) is a multi-faceted matricellular glycoprotein that is involved in tissue remodeling, morphogenesis and bone mineralization (57). While it functions as a tumor suppressor in neuroblastomas, colorectal and ovarian cancer (57, 58), it has been reported to be elevated on the mRNA level in sarcomatoid RCC, as well as most of ccRCC tumors (70%) (59). SPARC has been reported to be expressed only by tumor-associated stromal cells such as vascular endothelial cells and fibroblasts, but not by cancer cells themselves (59). In line with this, we observed for SPARC an intense staining in the intratumor stroma (**Fig. 22**). Furthermore, we detected a significantly elevated abundance of SPARC in urine samples of ccRCC patients (**Fig. 23**), proving our concept to find secreted biomarkers from our comprehensive proteomics analysis.

Taken together, our comprehensive proteomics analysis identified four novel biomarkers for the detection of ccRCC. Future directions should entail the investigation of the role of these biomarkers in ccRCC tumorigenesis and metastatic spread, as well as the application of SPARC as an analytical tool in clinical assays.

*Chapter 3*

# Stratification of ccRCC tumors

## 3.1 LITERATURE REVIEW

### 3.1.1 Tumor heterogeneity

Kidney cancer is considered to be resistant to conventional chemo- and radiotherapy. This is particularly due to genetic and epigenetic intra- and interpatient tumor heterogeneity (**Fig. 26A**) caused by truncal (*VHL*, *PBRM1*) and branched (*SETD2*, *BAP1*, *KDM5C*, *TP53*, *PTEN*, *MTOR*) clonal evolution of the cancer cells (**Fig. 26B**) leading to poor treatment response or to treatment resistance (4, 18, 60). Over the past years, it became clear that a "one-size-fits-all" treatment strategy for RCC patients is insufficient for the effective management of the disease and the prevention of recurrence (18).

Another crucial factor causing tumor diversity is the composition of the tumor microenvironment, which consists of tumor microvesicles, fibroblasts, extracellular matrix, adipose cells, blood vessels, signaling molecules, and different types of stromal and immune cells (12, 18, 61). In recent years, single-cell transcriptomics approaches have gained in importance for the dissection of the tumor microenvironment of different cancer types (61).

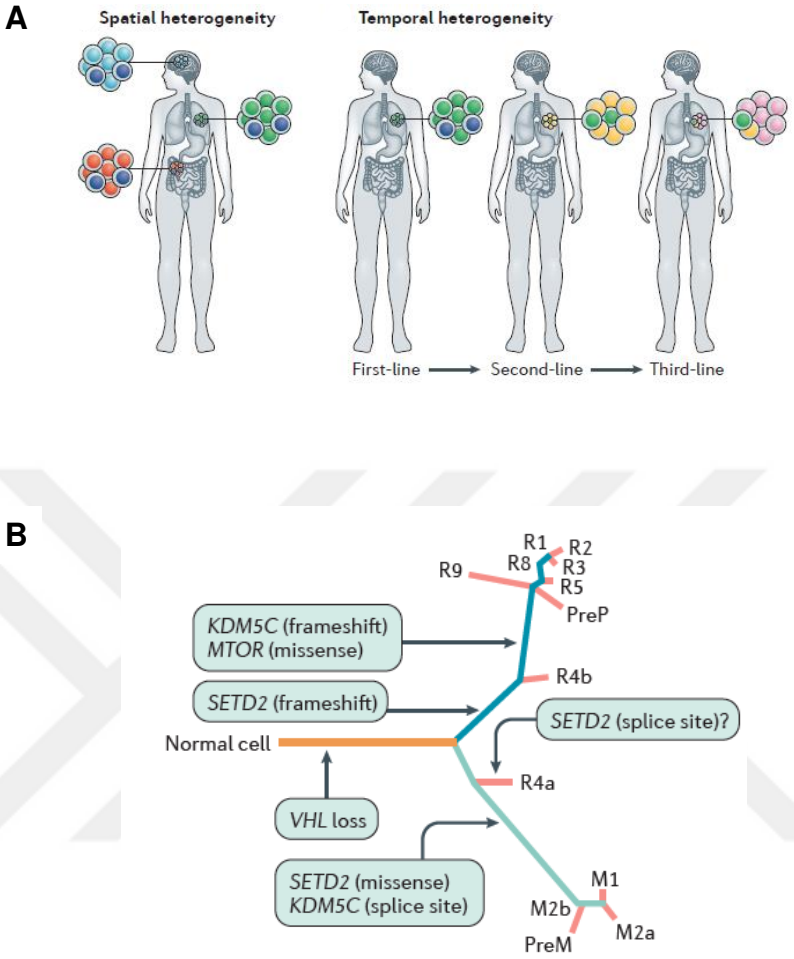


**Figure 26 | Tumor heterogeneity in RCC. (A)** Regional (spatial) and progressive (temporal) intratumoral heterogeneity (adapted from (62)). **(B)** Branched clonal evolution of ccRCC tumors causing genomic heterogeneity. Truncal mutations in the *VHL* gene are often accompanied by multiple parallel genetic alterations in other tumor suppressor genes. R: primary tumor, M: metastatic tumor, PreM: before metastasis (adapted from (25)).

### 3.1.2 Patient classification

One strategy to overcome tumor heterogeneity is the subtyping of patients for individualized tumor intervention. While classification schemes based on histological or morphological features such as the grade, stage or size of the tumor showed limited discriminative power, stratification based on in-depth molecular expression profiles has gained in importance (13). More importantly, drug treatment according to individual molecular aberrations opens new avenues for personalized medicine. Especially in the case of grade 2 and 3 tumors, the disease progress and outcome can vary dramatically (63), and therefore be difficult to predict, which further necessitates appropriate tumor risk stratification.

#### 3.1.2.1 Genomics-based classification

Initial attempts in the clinics to classify patients considered the genetic diversities across renal tumors. *VHL* mutations at the 3p25 arm are considered to be driver events in ccRCC tumorigenesis, however, genetic variations of other tumor suppressors in the 3p21 location have been found to be essential drivers (13) (**Fig. 27**). A set of genes including *VHL*, *PBRM1*, *BAP1*, *SETD2*, *KDM5C*, *TP53*, *PIK3CA*, *MTOR*, *TSC1* and *NF2* are frequently mutated in approximately 93% of ccRCC tumors, and are taken into consideration for classification of patients (13). While *BAP1* mutations are associated with larger tumors, higher grades and worse patient outcome due to resistance to targeted therapies, *SETD2* mutations are mostly associated with distant metastasis and intra-tumor heterogeneity due to convergent evolution (13). Moreover, mutations in the *KDM5C* gene are associated with male patients, with co-occurrence of *PBRM1* mutations and with good response in progression-free survival to sunitinib treatment (13).

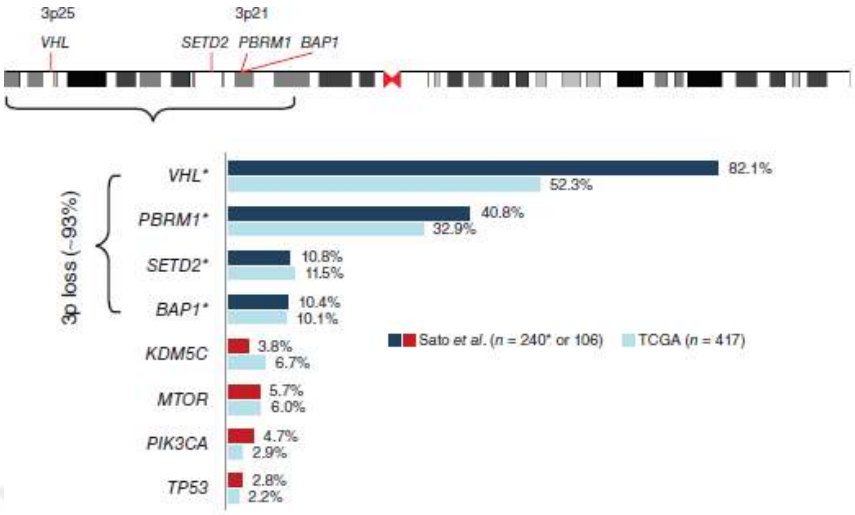


**Figure 27 | Frequently mutated genes in ccRCC.** The first five genes are tumor suppressors, which are all located at the short arm of chromosome 3 (adapted from (64)).

#### 3.1.2.2 Transcriptomics-based classification

Various subtyping systems based on transcriptomic profiles of ccRCC tumors have been proposed. One is the widely investigated microarray derived ccA/ccB clustering signature identified by unsupervised clustering of 48 tumors (**Fig. 28A**) (65). The approach revealed noticeable prognostic difference between the low-risk ccA tumors, characterized particularly by fatty acid metabolism and angiogenesis, and the high-risk ccB tumors, associated with EMT and Wnt signaling. The classifier has further been narrowed down to a panel of 34 cluster-discriminative genes that recapitulates the prognostic difference between the two tumor groups (63, 66).

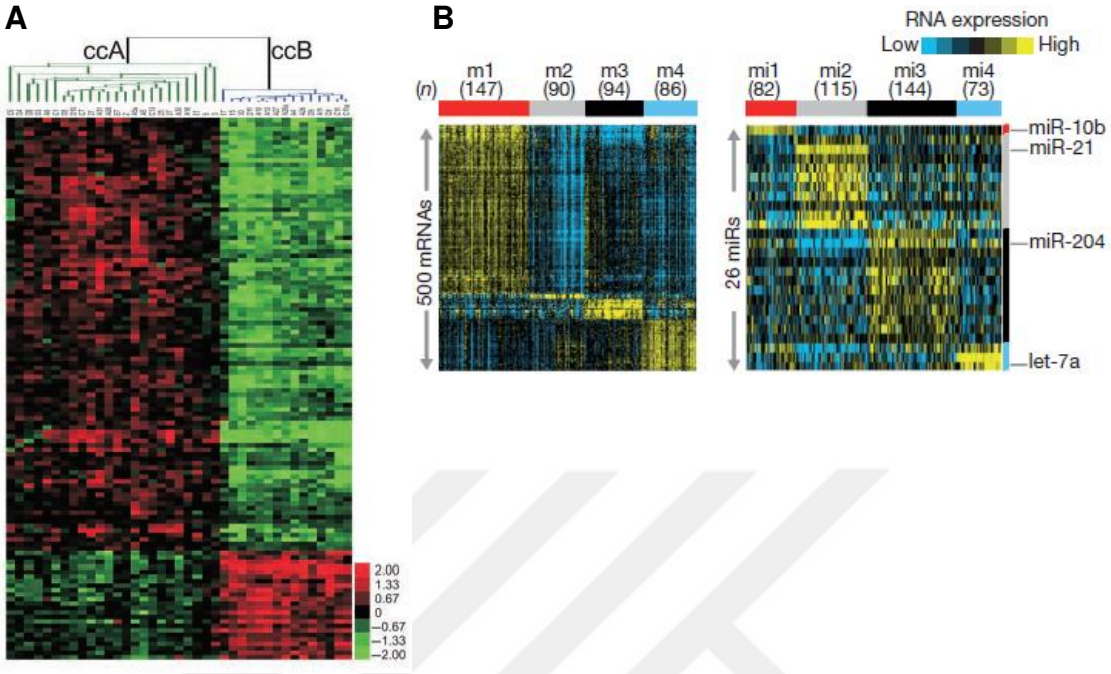


**Figure 28 | Stratification of ccRCC tumors based on transcriptomics expression profiles. (A)** Unsupervised clustering grouped 48 tumor samples based on microarray data into two distinct clusters with good (ccA) and poor (ccB) survival (adapted from (65)). **(B)** Unsupervised clustering of TCGA mRNA data of the KIRC (Kidney Renal Clear Cell Carcinoma) cohort (n=446) grouped the tumors into four distinct clusters (adapted from (67)).

Another unsupervised classification system is the m1-m4 classifier created by using mRNA and miRNA data of a cohort of 446 patients (67) (**Fig. 28B**). The characteristics of the suggested subgroups were described as follows:

- **m1 tumors:** higher frequency of PBRM1 mutations, more chromatin remodeling genes expressed; similarity to ccA cases
- **m2 tumors:** similarity to ccB cases
- **m3 tumors:** higher frequency of deletion of *CDKN2A* and mutations in *PTEN*; similarity to ccB cases

- **m4 tumors:** higher frequency of *BAP1* mutations and *MTOR* mutations, more base-excision repair, more ribosomal genes expressed; unclassified by ccA/ccB system

An overview of the three most established clustering signatures for ccRCC tumors based on transcriptomics data is given in **Table 2**:

**Table 2 | Overview of transcriptomics-based classification systems for ccRCC tumors.**

| | ccA/ccB (ClearCode34) (65), (66) | m1-m4 classifier (67) | CC-e.1,2,3 (13) |
|---|---|---|---|
| **Number of specimens** | n=48 | n=446 | n=894 |
| **Low grade tumors** | ccA | m1 | CC-e.2 |
| **High grade tumors** | ccB | m2 and m3 | CC-e.3 |
| **unclassified** | - | m4 | CC-e.1 |

Another suggested clustering approach is the mCluster 1-4 system, which showed that tumors with high levels of glutathione-related metabolites and dipeptides (mClusters 2 and 3) were associated with worse clinical outcome, while low levels of these metabolites correlated with better outcomes (mCluster 1 and 4) (13). Consequently, mClusters 1 and 4 might resemble ccA tumors, while mClusters 2 and 3 might resemble ccB tumors.

#### 3.1.2.3 Proteomics-based classification

In contrast, classification systems of RCC based on protein expression data have rarely been proposed (13, 15). Only one proteomics study describes a clustering signature which separates ccRCC tissues into three clusters (ccRCC1-3) with diverse malignancy and immune cell infiltration profiles (15) (**Fig. 29**).

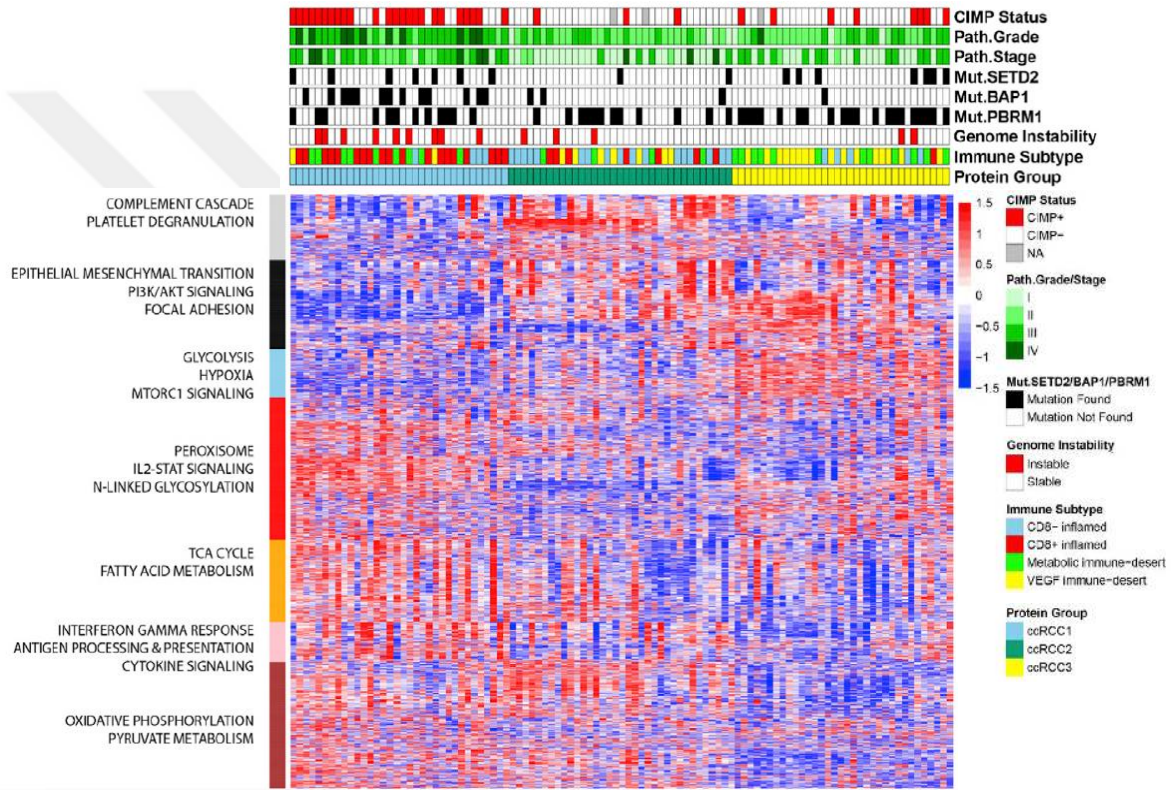


**Figure 29 | Stratification of ccRCC tumors based on proteomics expression profiles (CPTAC study).** A total of 103 tumors were clustered into three distinct groups (ccRCC1-3). Pathway enrichment annotations are Hallmark, KEGG and Reactome pathways (adapted from (15)).

The main characteristics of the suggested CPTAC clusters are summarized in **Table 3**:

**Table 3 | Characteristics of three ccRCC clusters based on CPTAC proteogenomics data.**

| ccRCC1 tumors | ccRCC2 tumors | ccRCC3 tumors |
|---|---|---|
| -higher grades & stages<br>-infiltrating CD8+ T cells<br>-BAP1 mutations<br>-genomic instability<br>-OXPHOS<br>-TCA cycle<br>-poor survival | - lower grades & stages<br>- infiltrating CD8- T cells<br>- EMT<br>- platelet degranulation<br>- PI3K-Akt signaling<br>- focal adhesion | -lower grades & stages<br>-no immune cell infiltration<br>-VEGF signaling<br>-glycolysis<br>-mTOR<br>-hypoxia<br>-EMT<br>-better survival |

## 3.2 MATERIALS & METHODS

### 3.2.1 Clustering and Principal Component Analysis

Unsupervised hierarchical clustering was applied sample-wise on the global proteome data. A distance matrix was prepared by calculating the square root of the sum of squared differences between samples (Euclidean distances) (**Fig. 30**) with an in-house generated Python script. The subsequent clustering was done using the linkage function of the scipy.hierarchy module with the weight method "ward" and the distance metric "Euclidean". Furthermore, k-means clustering of the 5,799 quantified proteins was applied using the tool ConsensusCluster with 300 iterations (68). Principal Component Analysis (PCA) was employed for the first two components using the PCA function of the sklearn.decomposition module.

$$d(\mathbf{p},\mathbf{q}) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$$

$p, q$ = two points in Euclidean n-space
$q_i, p_i$ = Euclidean vectors, starting from the origin of the space (initial point)
$n$ = n-space

**Figure 30 | Formula for the calculation of Euclidean distances used for the generation of distance matrices.**

### 3.2.2 Statistical analysis

For the determination of statistically significant cluster-discriminative proteins, the two-sample Student's *t*-test was applied. *P*-values were calculated gene-wise using the ttest_ind function of the scipy.stats module by omitting missing values (nan_policy="omit"). *P*-values <0.05 were considered statistically significant.

### 3.2.3 Functional annotation

For pathway annotations, over-representation analysis was applied using the WebGestalt platform (v2019) (36) with a BH-adjusted FDR of 0.05. Annotations were combined analysis of KEGG, Reactome, Hallmark50, and Wikipathway_cancer databases.

### 3.2.4 Comparison with other ccRCC expression data

Two independent expression datasets of ccRCC were used for comparison with our study. FPKM (Fragments per kilo base per million mapped reads) normalized RNASeq data of The Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma (TCGA-KIRC) cohort consisting of 539 tumor and 72 normal tissue samples was retrieved from the TCGAPortal (http://tcgaportal.org/download.html, access April 2020). Genes with more than 70% missing expression values were filtered out and tumor counts were gene-wise normalized to the upper quartile of respective normal counts and $\log_2$ transformed. Moreover, the CPTAC ccRCC proteome data was retrieved from the CPTAC portal (https://cptac-data-portal.georgetown.edu, access November 2019). Quality control samples and the reported contaminated sample ("C3N-00314") were excluded. $\text{Log}_2$ tumor/normal ratios were calculated for the 80 tumors with adjacent normal sample by eliminating the relation to the TMT reference sample. Pearson and Spearman correlations of the median $\log_2$ tumor/normal ratios between datasets were calculated using the respective Python scipy.stats functions.

### 3.2.5 Clustering validation and marker gene identification

To verify the observed clustering signature, the TCGA and CPTAC datasets were subjected to two-sample Student's *t*-test after clustering ($p$-value <0.05). Cluster assignment was then conducted based on highest overlap of the significantly different proteins between the clusters of the validation datasets with those between the clusters of our proteomics study. To identify the key discriminatory elements of the clustering

signature ("marker genes"), a cut-off value of ±2 for the *t*-test difference and a cut-off >2 for the -$\log_{10}$ *p*-value were set.

### 3.2.6 Survival analysis

For evaluation of the prognostic impact of the tumor clusters, the follow-up data of the TCGA-KIRC cohort for 534 patients was used (Firehouse Legacy, downloaded from cBioPortal, access August 2020). Survival analysis was performed using the KaplanMeierFitter and multivariate_logrank_test functions of the Python library lifelines (v0.25.0). Furthermore, statistical significance of the frequency of clinicopathological factors in the clusters was determined using the fisher_exact function of the Python scipy.stats library. To determine the impact of clinicopathological factors on survival, univariate and multivariate Cox regression analyses were performed using the CoxPHFitter function of the lifelines library. *P*-values <0.05 were considered statistically significant.

### 3.2.7 Network-based interpretation of proteomics data

Optimal subnetworks, which represent the list of cluster-significant proteins best, were reconstructed by integrating a reference human interactome with our proteomic data using the Forest module of Omics Integrator (v0.3.1) (38) and the visualization tool Cytoscape (v3.8.0) (39). The iRefWeb_2013 interactome file was used as global network reference. Nodes were annotated by GO cellular compartment using QuickGO (40), which was limited to Uniprot assignments. For multiple allocations, priority was given as follows: "extracellular", "membrane", "nucleus", "cytoplasm", "other" or "unknown". Additionally, networks were annotated with Reactome pathways using the Cytoscape plug-in ClueGO (v2.5.7) (41). The network was further complemented with associated inhibitory FDA-approved drugs retrieved from the Drugbank database (v5.1.7, access August 2020) (69). Chemical elements were excluded as drugs.

## 3.3 RESULTS

### 3.3.1 Molecular expression profiling stratifies ccRCC tumors into two clusters with distinct pathobiology

Our comprehensive proteomic analysis and rigorous comparison between normal and tumor tissues provided us with new biomarkers for the detection of ccRCC. Next, we focused on the interpatient comparison in an attempt to identify a novel biomarker panel for the classification of patients based on their individual molecular expression profiles. To this end, we performed unsupervised hierarchical clustering to unveil tumor clusters. Indeed, the tumors grouped into two distinct clusters based on the expression profile of the 5,799 quantified proteins (**Fig. 31**).

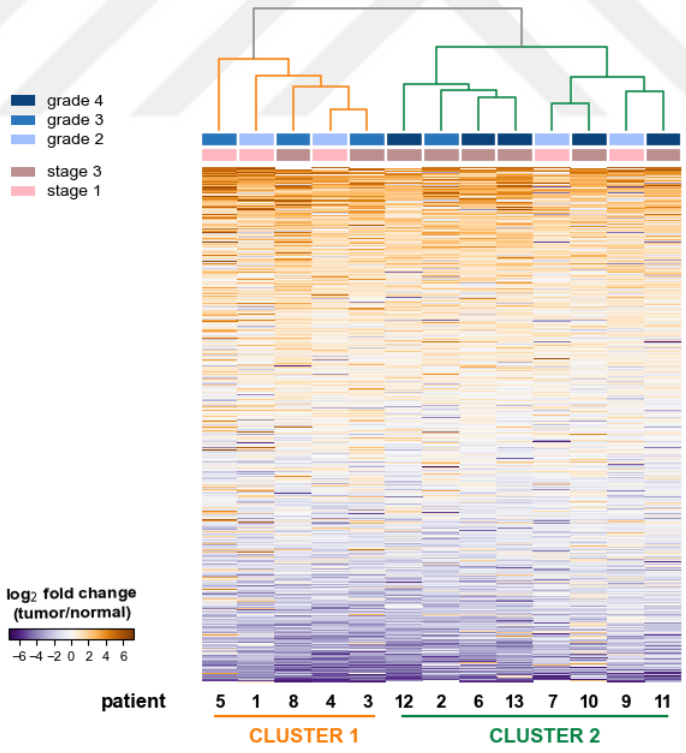


**Figure 31 | Hierarchical clustering of tumors into two main clusters using expression data of 5,799 quantified proteins.** Grade and stage information of tumors is given in column bars.

Further, we challenged the robustness of the clustering by applying k-means clustering on the data, which recapitulated the same classification of the patients (**Fig. 32**). Increasing the number of the k-value (degree of freedom) from two to four did not remarkably change the core clusters, which confirms that the observed two main clusters are stable.

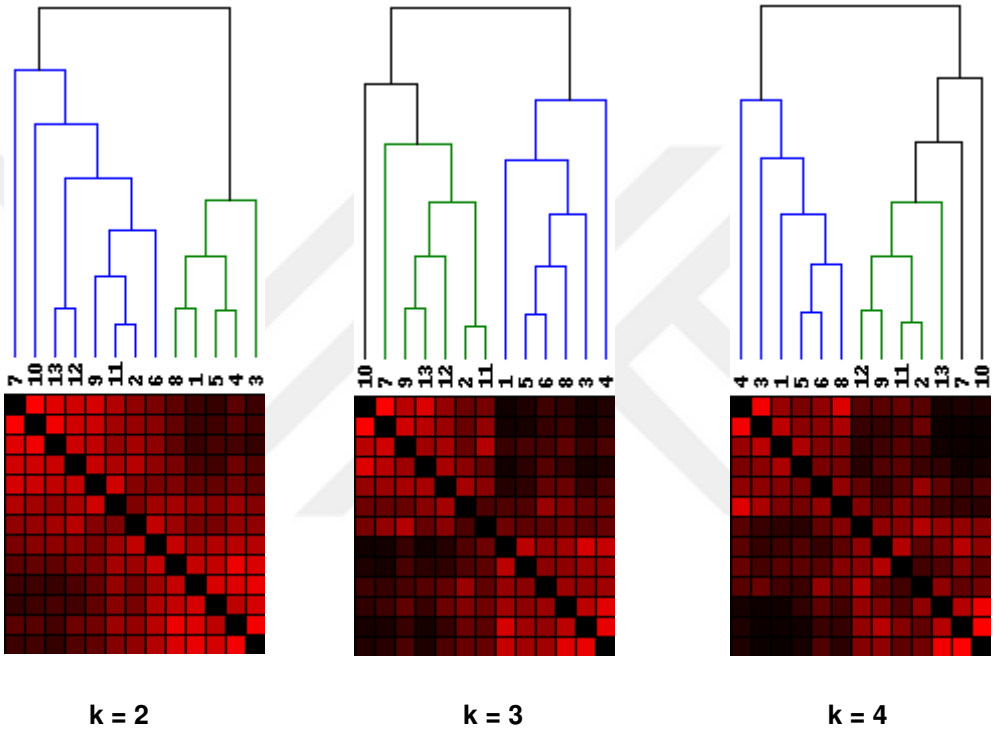


**Figure 32 | K-means clustering of same data with increasing degree of freedom (k-value) confirming two stable clusters.**

Next, we employed PCA, which showed concordance with the results of the hierarchical clustering and further confirmed the two main clusters (**Fig. 33**).

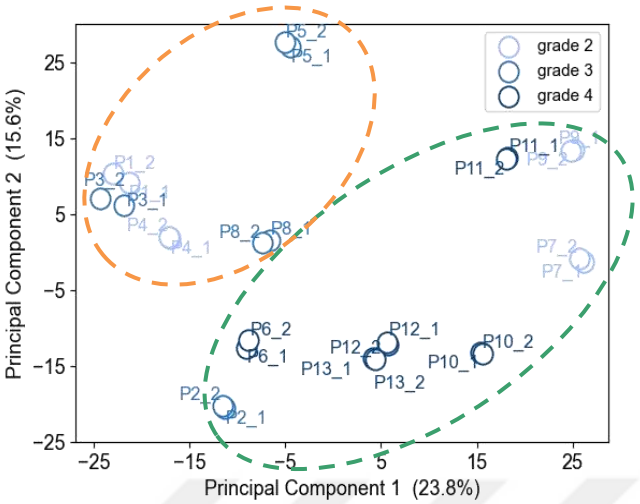


**Figure 33 | Principal component analysis of technical replicates of ccRCC samples.** Biological variance reflects observed cluster profiles.

Furthermore, statistical analysis revealed that a total of 599 proteins were differentially expressed between clusters (Student's *t*-test *p*-value <0.05), with 354 significantly higher abundant in cluster 1 ("cluster 1-specific") and 245 significantly higher abundant in cluster 2 ("cluster 2-specific"), respectively (**Fig. 34A**). Cluster 1-specific proteins were primarily involved in processes associated with the immune system, such as regulation of complement cascade and innate immune response (**Fig. 34B**), indicating that the observed enhanced immune cell infiltration in RCC (**Fig. 15** and **Fig. 16**) was primarily abundant in this subset of patients. Moreover, platelet activation and hemostasis implied neovascularization (**Fig. 15**), and therefore a concomitant angiogenic profile of these tumors. Also noteworthy is the enrichment of components associated with the regulation of the insulin-like growth factor IGF, which implied a primary dependence on this growth factor for tumor progression. In contrast, cluster 2 tumors displayed an enrichment of metabolic processes related to lipid, fatty acid and amino acid degradation (**Fig. 34C**). Moreover, proteins associated with mitochondrial processes such as oxidative phosphorylation and TCA cycle were higher abundant in cluster 2 tumors.

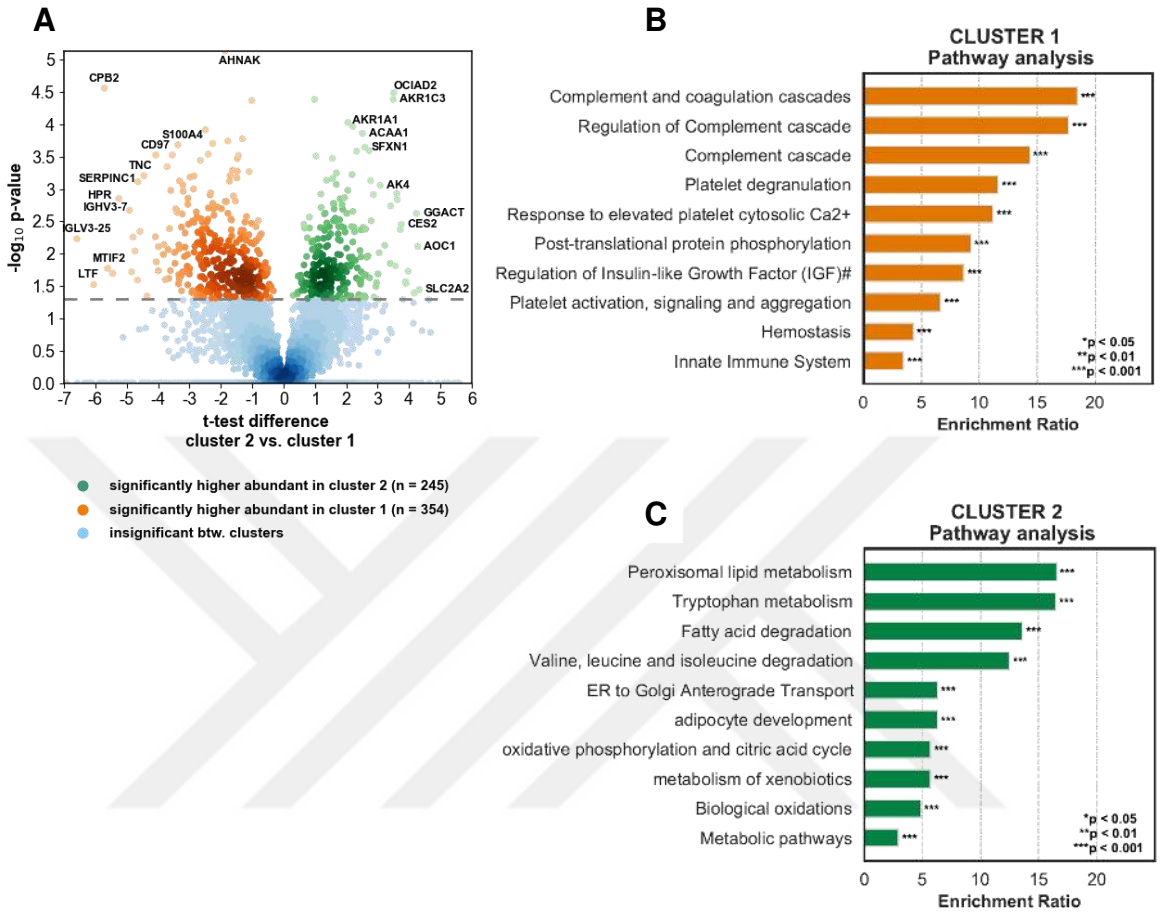


**Figure 34 | Significant cluster-specific proteins and pathways. (A)** Volcano plot representing proteins with significantly different expression between clusters. **(B-C)** Pathway analyses of cluster 1 **(B)** and cluster 2 **(C)** proteins, respectively, based on KEGG&Reactome&Hallmark50 annotations. Top 10 enriched annotations are represented. Full annotation of term marked with #: Regulation of Insulin-like Growth Factor (IGF) transport and uptake by insulin-like Growth Factor Binding Proteins (IGFBPs).

To obtain a predictive biomarker panel sufficient for cluster discrimination, we further reduced the 599 cluster-significant proteins to a subset of representative proteins by applying cut-off values on the statistics ($-\log_{10}$ *p*-value >2, *t*-test difference ±2). The resulting 73 discriminative proteins, hereafter denoted as "marker genes", comprised of 54 cluster 1-specific and 19 cluster 2-specific proteins, respectively (**Fig. 35**).

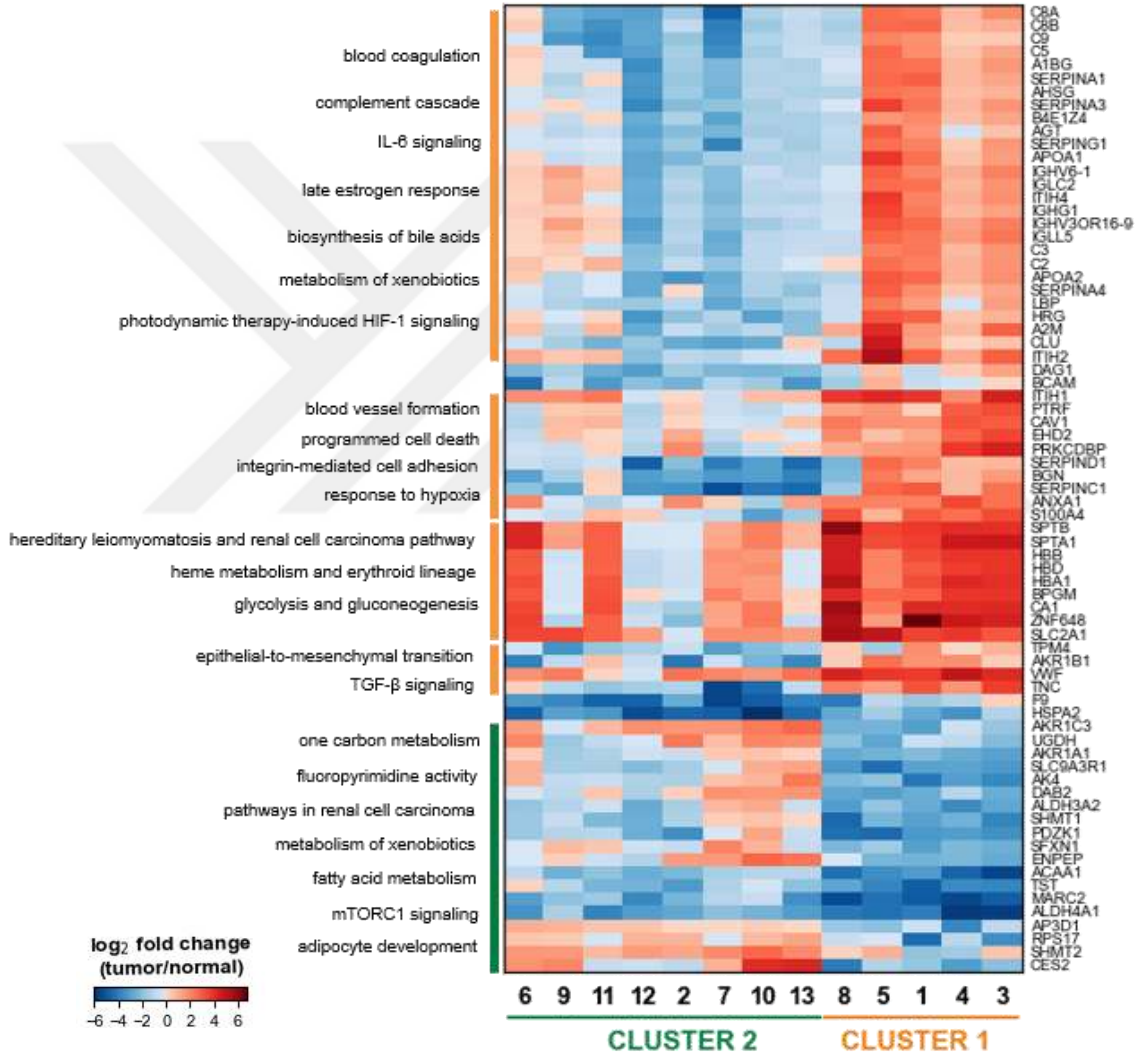


**Figure 35 | Heatmap representing 73 most discriminative clustering marker genes (*t*-test difference cut-off ±2, $-\log_{10}$ *p*-value >2).** Pathway analyses based on Reactome and Wikipathway_cancer annotations.

Besides their expected role in angiogenesis and immune response related processes (**Fig. 34B**), marker genes of cluster 1 were also associated with TGF-ß related metastasis, apoptosis and HIF1α-signaling (**Fig. 35**). This further implies a progressive and malignant clinical behavior of cluster 1 tumors. Contrarily, besides their expected primary involvement in metabolic processes (**Fig. 34C**), marker genes of cluster 2 were also associated with mTOR signaling and adipocyte development (**Fig. 35**), both of which have previously been described to promote RCC progression (3, 70).

### 3.3.2 Cluster signature is conserved in independent ccRCC mRNA and protein expression data

For the validation of the proposed clustering profile, two independent larger datasets were used. First, the TCGA-KIRC RNASeq dataset (67), consisting of 539 tumors (normalized to 72 normal tissue counts), was utilized. In general, more than 5,500 mRNA-protein pairs displayed a significant positive correlation (0.38-0.43, *p*-value <0.0001, **Fig. 36A**) similar to previous observations in proteotranscriptomics analysis of ccRCC (7, 15), warranting the suitability of this transcriptomics data for validating our proteome derived clustering signature. Furthermore, 565 cluster-significant proteins had an mRNA counterpart, which also demonstrated significant positive correlation (*p*-value <0.0001), with a Spearman correlation of 0.6 and a Pearson correlation of 0.42, respectively (**Fig. 36B**). Surprisingly, proteome-transcriptome regulation discrepancy was higher for cluster 1-specific pairs (Spearman 0.49, Pearson 0.38) compared to cluster 2-specific pairs (Spearman 0.74, Pearson 0.69). This may suggest that metabolic processes in cluster 2-tumors are less prone to discordant regulation between the transcriptome and proteome level than immune and angiogenesis related processes in cluster 1-tumors.

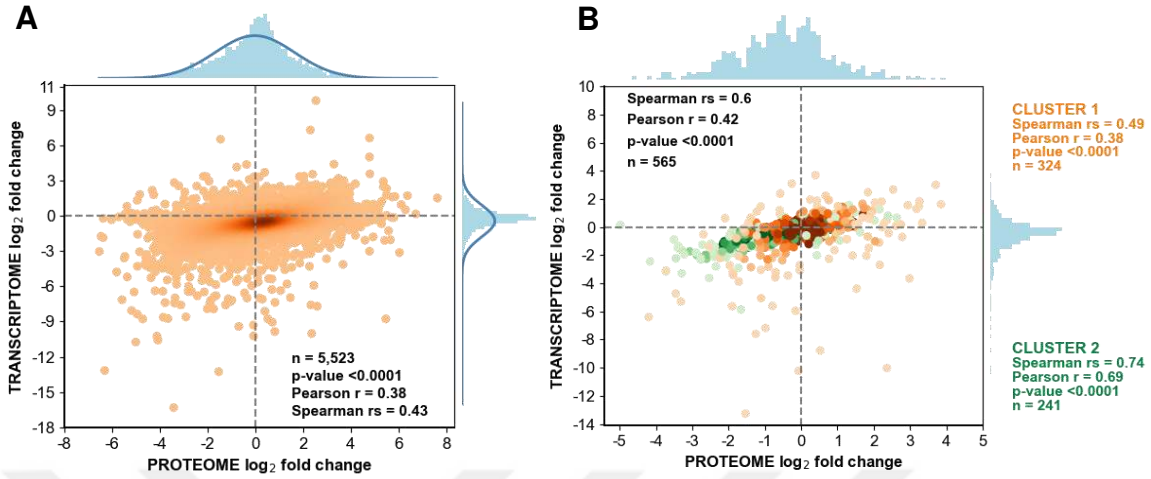


**Figure 36 | Proteotranscriptomics analysis of global proteome and TCGA-KIRC data by median $log_2$ tumor/normal ratios. (A)** Comparison of all 5,523 proteomics-TCGA pairs. **(B)** Comparison of 565 cluster-significant proteomics-TCGA pairs.

Unsupervised hierarchical clustering of the TCGA data based on the expression of these 565 mRNA/protein pairs classified the tumors into three clusters (TCGA-clusters), with two main clusters and one outlier cluster (n=19), which was excluded from the validation (**Fig. 37A**). For cluster assignment, the significantly differentially expressed genes between the TCGA-clusters were compared with those from the proteome-clusters of this study and assigned based on their overlap. TCGA-cluster 1 was assigned with more than 84% concordance with proteome-cluster1, and TCGA-cluster 2 was assigned with more than 62% overlap with proteome-cluster2 (**Fig. 37B**).

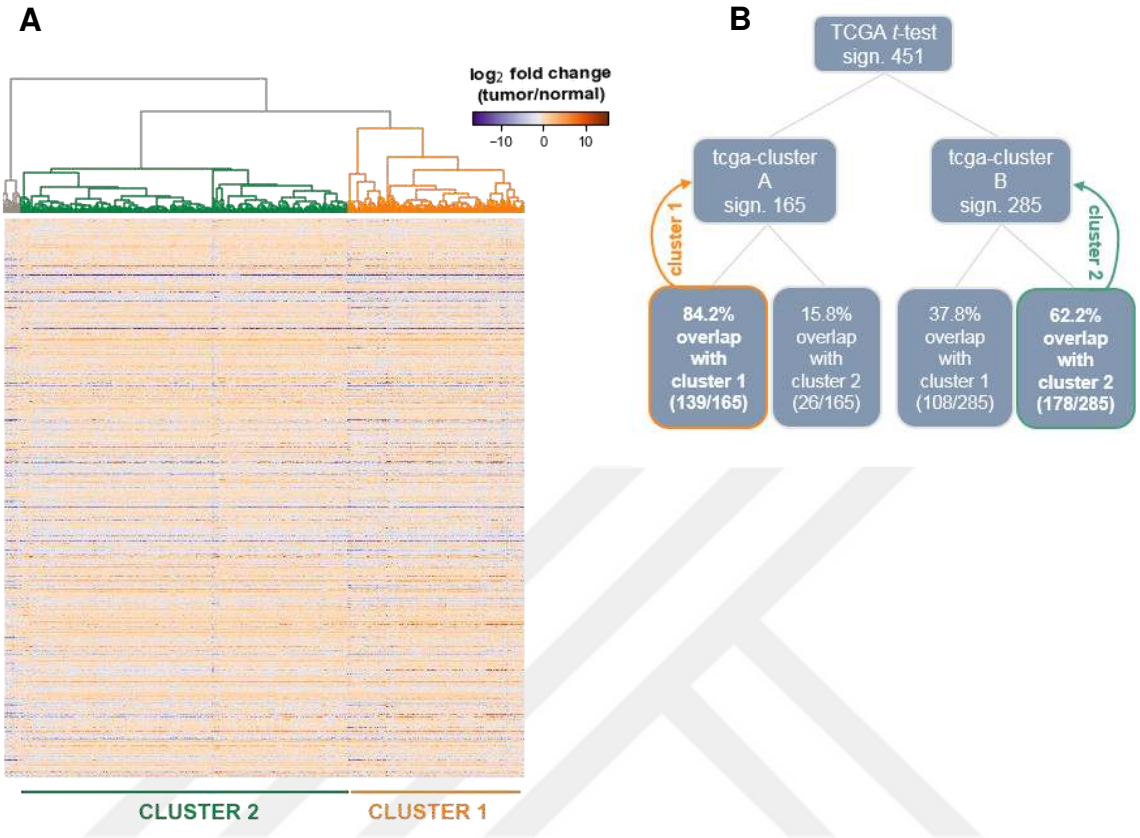


**Figure 37 | Hierarchical clustering of TCGA-KIRC samples (n=539) by 565 cluster-significant genes. (A)** Clustering indicates two main clusters and one outlier cluster (n=19), which was excluded from validation. **(B)** Clusters were assigned based on overlap of genes with proteome-based clusters of this study.

We further validated the clustering signature using the CPTAC data for ccRCC (15), which only included patients with tumor and NATs (n=80). Of the determined cluster-significant proteins in our study, 547 were detected in the CPTAC proteome (**Fig. 38**), which displayed significantly high concordance in regulation (correlations >0.82, *p*-value <0.0001).

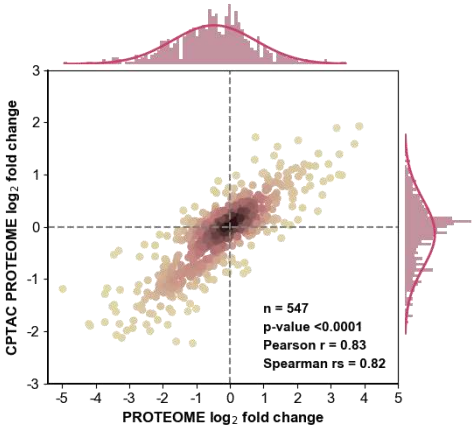


**Figure 38 | Correlation analysis between proteome data and CPTAC data for 547 common cluster-significant proteins based on median $log_2$ tumor/normal ratios.**

Unsupervised hierarchical clustering of the CPTAC data based on the expression of these common cluster-significant proteins also classified the tumors into two main clusters (CPTAC-clusters) (**Fig. 39A**). The subsequent cluster assignment resulted in an overlap of approximately 99% with cluster 1 proteins and 94% with cluster 2 proteins of this study, respectively (**Fig. 39B**).

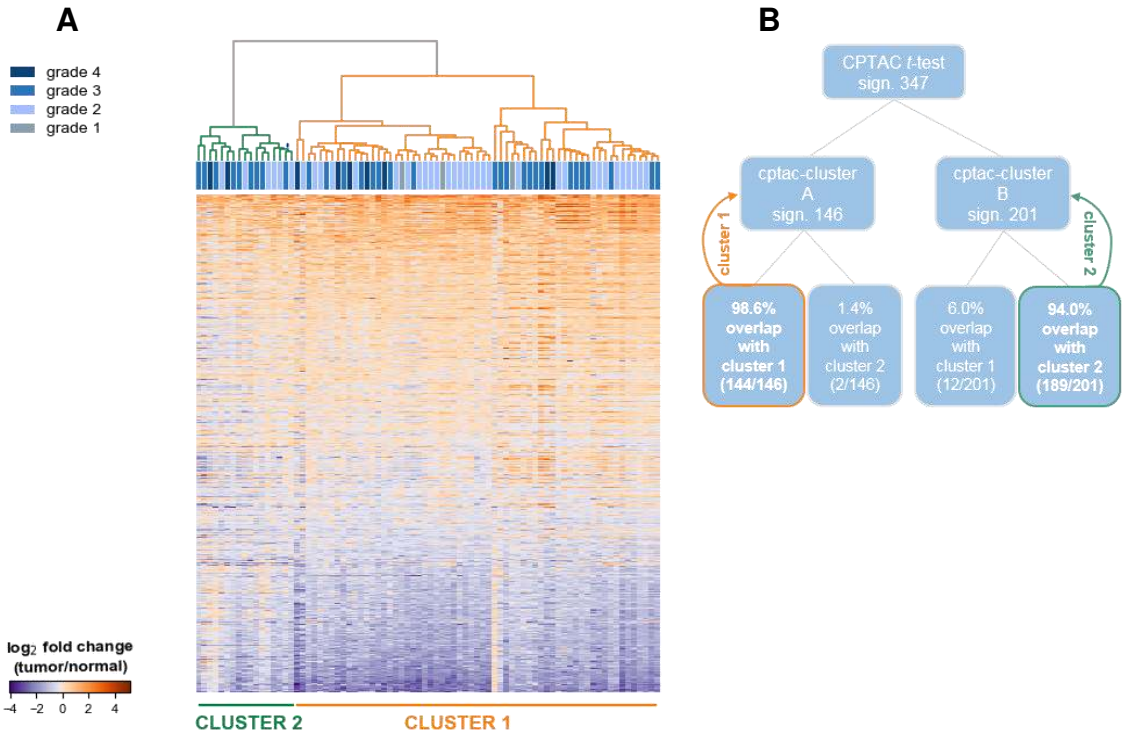

**Figure 39 | Hierarchical clustering of CPTAC samples (n=80) by cluster-significant proteins (n=547) into two clusters. (A)** Clustering indicates two main clusters. **(B)** Clusters were assigned based on overlap of proteins with proteome-based clusters of this study.

### 3.3.3 Cluster 1 patients have almost three times higher risk of death than cluster 2 patients

Next, we examined the prognostic difference between the clusters using the TCGA-KIRC follow-up data. Strikingly, Kaplan-Meier analysis of the overall survival (OS) revealed highly significant discrepancy in clinical outcome between the clusters (log-rank test *p*-value <0.001) (**Fig. 40A**). Patients of cluster 1 had an approximately 2.8 (95% confidence interval (CI) [2.03-3.74]) times higher risk of death (hazard ratio, HR) compared to patients of cluster 2. The HR value was even higher for the recurrence-free survival (RFS) of this cohort with more than four times higher risk of death (95% CI [2.59-6.71]) for patients of cluster 1 (**Fig. 40B**). Furthermore, the 5-year survival rate of the patients differed noticeably between clusters (**Fig. 40A**). While more than 75% (95% CI [69.6%, 80.8%]) of the patients in cluster 2 survived five years post-diagnosis, only approximately 43% (95% CI [34.4%, 51.1%]) of the patients in cluster 1 were alive. This significant difference in survival between the patient clusters was also recapitulated by the 73 marker genes, which proves that this biomarker panel is sufficient for clustering of ccRCC tumors (**Fig. 40C**).

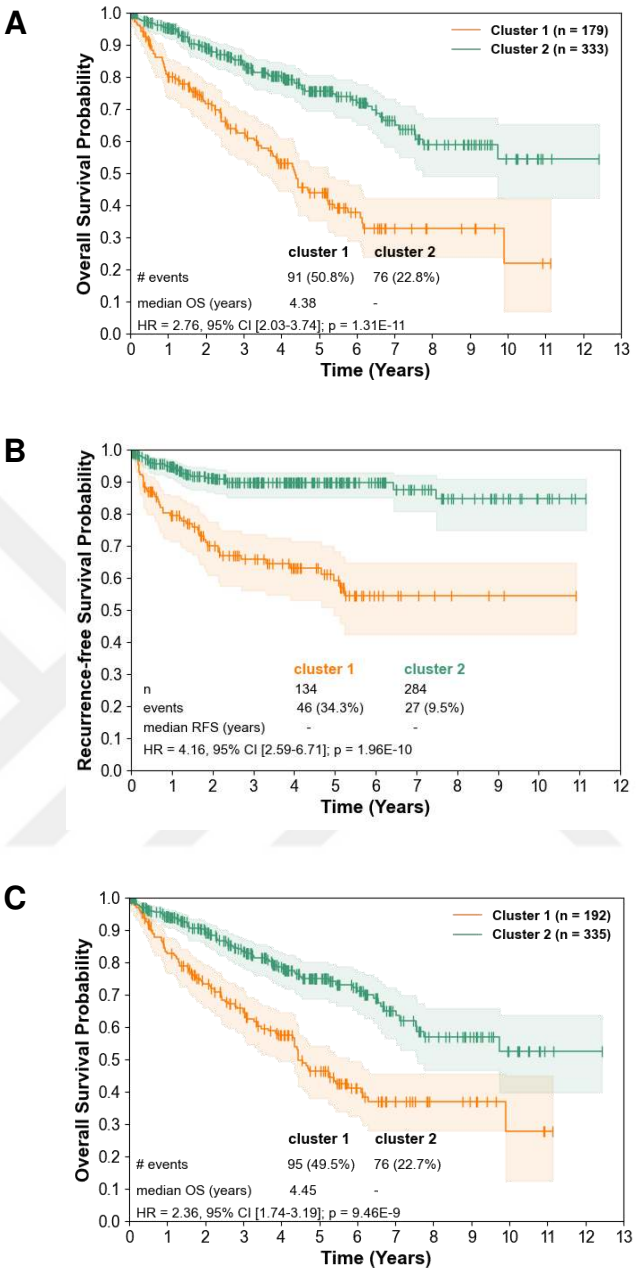


**Figure 40 | Kaplan-Meier analysis of TCGA cohort.** Log-rank test indicates significant survival discrepancy between patient clusters. **(A)** Overall survival and **(B)** recurrence-free survival using 565 cluster-discriminative genes for patient partitioning. **(C)** Overall survival based on clustering of patients solely by 73 cluster-discriminative marker genes recapitulates significant survival discrepancy between patient clusters.

The relatively higher mortality rate of cluster 1 patients was also supported by the evaluation of associated clinicopathological features (**Table 4**). Remarkably, cluster 1 patients demonstrated significantly higher correlation with unfavorable clinical factors such as the presence of lymph nodes or distant metastasis, higher stages and grades, and recurrent disease (*p*-values <0.005).

**Table 4 | Comparison of clinicopathological parameters in clusters of TCGA-KIRC cohort (n = 534).** For statistical analysis, the Python function ttest_ind was used for the diagnosis age, while the fisher_exact function of the scipy.stats library was used for the other parameters.

| | Cluster 1 | Cluster 2 | Fold frequency difference cluster 1 vs cluster 2 | *p*-values |
|---|---|---|---|---|
| **Size** | 181 | 333 | - | - |
| **Mean Diagnosis age** | 61.5 | 60.0 | - | 0.181 |
| **Metastasis stage M1** | 24.3% (44) | 9.9% (33) | 2.45 | 2.63E-05 |
| **Lymph node stage N1** | 6.7% (12) | 1.2% (4) | 5.58 | 2.02E-03 |
| **Disease stage I + II** | 44.2% (80) | 68.5% (228) | 0.64 | 1.13E-07 |
| **Disease stage III + IV** | 55.2% (100) | 30.9% (103) | 1.79 | 1.03E-07 |
| **Histologic grade 1 + 2** | 29.3% (53) | 54.7% (182) | 0.54 | 3.7E-08 |
| **Histologic grade 3 + 4** | 69.6% (126) | 44.7% (149) | 1.56 | 6.7E-08 |
| **Disease free** | 39.8% (72) | 66.9% (223) | 0.59 | 3.26E-09 |
| **Recurred/Progressed** | 35.4% (64) | 18.3% (61) | 1.93 | 2.49E-05 |
| **Fraction genome altered > 50%** | 6.1% (11) | 2.4% (8) | 2.54 | 0.048 |

Moreover, to find out whether the proposed patient clustering was among the confounding factors having a significant impact on the survival of ccRCC patients, univariate and multivariate Cox regression analyses were performed with the follow-up data of the TCGA cohort (**Table 5 and Table A3**). Indeed, besides the age at diagnosis (*p*-value <0.01), tumor stage (*p*-value <0.005), tumor grade (*p*-value <0.05) and the presence of

distant metastasis (*p*-value <0.0001) (also the presence of lymph node metastasis in univariate analysis, *p*-value <0.0001), the cluster membership was determined as a statistically significant predictor of survival in both analyses (*p*-value <0.0001). In consideration of all parameters, the assignment to cluster 1 tumors indicated a similarly high risk of mortality as the presence of distant metastasis (HR of 2.11 and 2.18, respectively) (**Table 5**).

**Table 5 | Summary of multivariate Cox Regression analysis of overall survival using clinical data of TCGA-KIRC cohort (n = 490).** Regression model was generated using CoxPHFitter function of Python lifelines library. "Not reported" entries for covariates were filtered out. CI: confidence interval.

| Parameter | Category setting | Hazard ratio | Lower 95% CI | Upper 95% CI | *p*-value |
|---|---|---|---|---|---|
| **Diagnosis age** | >60 vs <60 | 1.605 | 1.146 | 2.250 | 0.006 |
| **Prior cancer diagnosis** | Yes vs No | 1.040 | 0.657 | 1.648 | 0.866 |
| **Disease stage** | III+IV vs I+II | 1.888 | 1.242 | 2.869 | 0.003 |
| **Histologic grade** | G3+G4 vs G1+G2 | 1.493 | 1.005 | 2.216 | 0.047 |
| **Gender** | Male vs Female | 0.792 | 0.565 | 1.112 | 0.179 |
| **Cluster** | 1 vs 2 | 2.113 | 1.500 | 2.977 | 1.88E-05 |
| **Fraction genome altered** | >50% vs <50% | 0.963 | 0.468 | 1.982 | 0.919 |
| **Race category** | Asian vs Black vs White | 0.881 | 0.528 | 1.468 | 0.626 |
| **Cancer metastasis stage code** | M1 vs M0+MX | 2.177 | 1.482 | 3.196 | 7.22E-05 |
| **Lymph node stage** | N1 vs N0+NX | 1.478 | 0.750 | 2.914 | 0.259 |

### 3.3.4 Cluster-discriminative marker genes are druggable

To elucidate the cross-talk between marker genes of both clusters, we created a protein-protein interaction network complemented with Steiner nodes (**Fig. 41**), markers from cluster 1 are labelled yellow, markers from cluster 2 are labelled green). We observed that the most prominent hubs CAV1, F2R, FLOT2, CFTR, APOA1, CANX, GJA1, LGALS3,

ELANE and FN1 were highly interconnected (**Fig. 41B**, nodes with thick border). Considering that two of the hubs and most of their interactors were cluster 1-specific, we hypothesized that this underlying interconnectivity might be an attractive target for drug treatment of cluster 1 tumors. This prompted us to annotate the network with FDA-approved drugs with inhibitor function. A total of 13 out of the 87 cluster-discriminative marker genes and Steiner nodes were targets for 39 FDA-approved drugs (**Fig. 41A**). The majority of the pharmaceutical options that are used in RCC treatment target the VEGFR (sunitinib, pazopanib, sorafenib, axitinib) or mTOR (everolimus, temsirolimus) pathways (15). Although these annotated drugs for the marker genes are generally applied for diseases other than cancer such as hypertension, cardiovascular conditions, pulmonary embolism, anxiety disorders, osteoarthritis or inflammatory conditions, their repurposed application in combination with the FDA-approved drugs for ccRCC might be worthwhile as a future treatment perspective. For example, two marker gene hubs are targetable with drugs, namely ELANE with alpha-1-proteinase inhibitor and CFTR with ibuprofen and dexibuprofen. Also, as shown in our study, immune system related processes play a major role in the highly lethal cluster 1 tumors. Considering that five druggable marker genes (CA1, C5, MME, ELANE and ACAA1) were involved in the innate immune system response (**Fig. 41B**), stressing the importance of developing novel immunotherapy options is crucial for future treatment options.

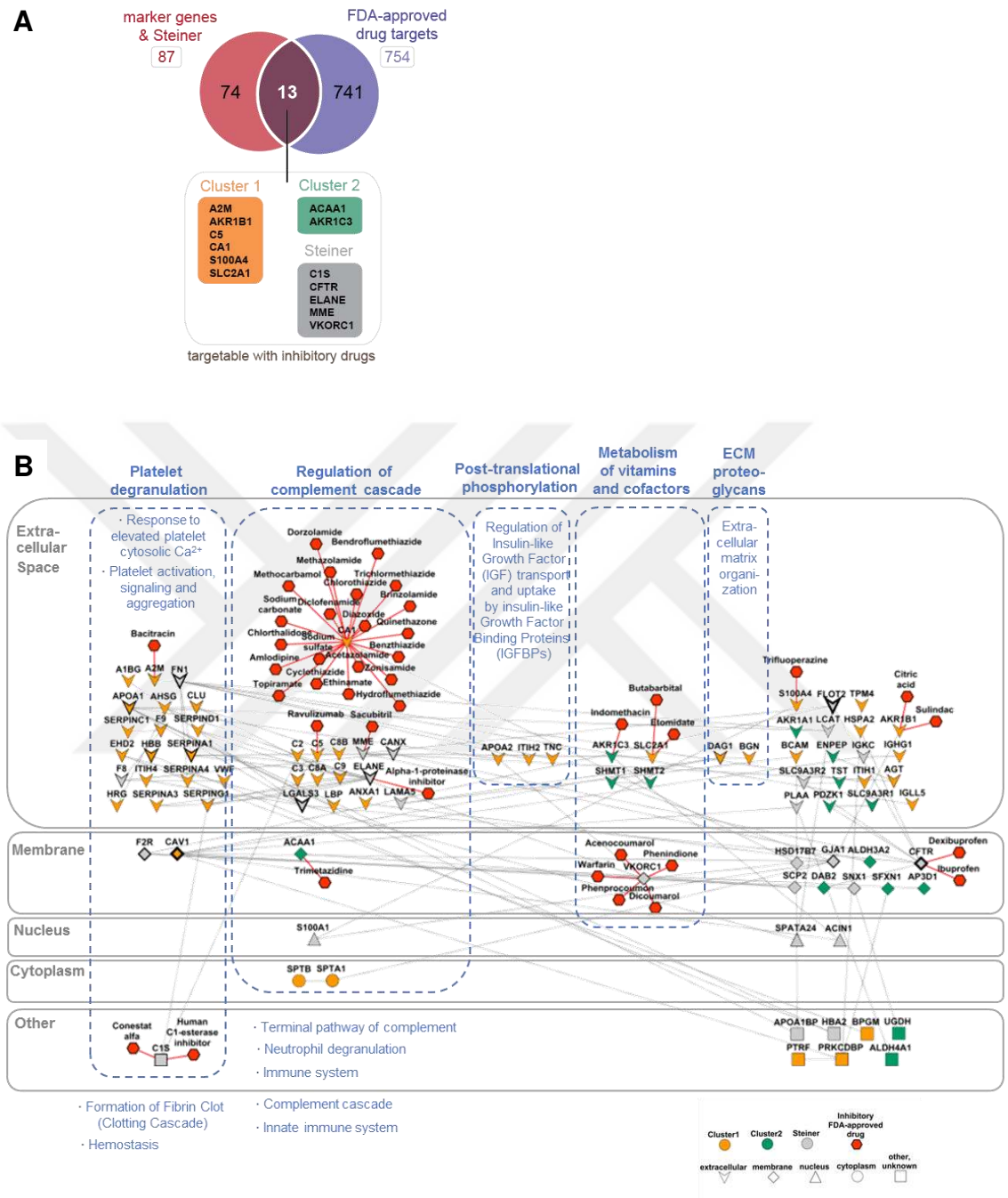


**Figure 41 | Reconstructed protein-protein interaction network with addition of Steiner nodes by using Omics Integrator.** **(A)** FDA-approved inhibitory drugs are linked to 13 drug targets. **(B)** Associated GO cellular component and Reactome pathways of proteins retrieved by QuickGO and ClueGO, respectively.

## 3.4 DISCUSSION

Here, we also propose a biomarker panel for the appropriate classification of ccRCC tumors based on the underlying molecular phenotypes, since classification solely based on histological or morphological features of tumors and subsequent treatment decisions can lead to varying treatment response (13, 65). In this study, we identified two ccRCC tumor clusters that differed in their molecular features. Our clustering signature is distinct from previously suggested ones such as the widely investigated transcriptomics based ClearCode34 grouping system, which has only five proteins in common (**Fig. 42**) (66).

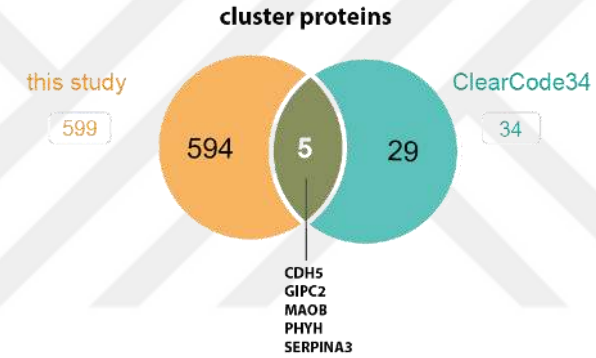


**Figure 42 | Common genes of our cluster-discriminative panel with ClearCode34 classification system for ccRCC tumors (66).**

Both the ClearCode34 as well as our classification system showed similar hazard ratios for the overall survival between the patient clusters (HR 2.4 and 2.2, respectively). However, regarding the recurrence-free survival, our clustering signature showed a 4.2 fold difference in the risk of death between the clusters while the difference was 2.3 fold between the ClearCode34 clusters. Furthermore, our clustering signature distinguishes from the proteomics based classification system ccRCC1-3 (15), which showed malignant features intermixed across the three proposed clusters, while in this study we showed that highly malignant processes such as EMT, HIF1α signaling, innate immune response and angiogenesis were condensed in cluster 1 tumors accounting for the relatively worse

survival outcome, while cluster 2 tumors had a low-risk profile and a more favorable prognosis (**Fig. 34**, **Fig. 35** and **Fig. 40**).

Advanced RCC is currently treated with monotherapies or combination therapies of immune checkpoint inhibitors and VEGF/mTOR inhibitors in first- and second-line strategies (15, 71). Hypothesizing from our data, the highly immunogenic and angiogenic cluster 1 tumors might be receptive to immunotherapy in combination with VEGF inhibitors (**Fig. 35**). In recent years, first-line approaches combining immunotherapeutics such as avelumab or pembrolizumab with the VEGF inhibitor axitinib have proven to be promising alternatives for the treatment of progressed RCC (71). However, using immunotherapies as first-line treatment excludes them as second-line therapy due to poor response. Thus, monotherapies with cabozantinib or axitinib are the preferred choices for second-line treatment (71). In contrast to cluster 1 tumors, mTOR signaling was elevated in cluster 2 tumors (**Fig. 35)**, which might be targetable by the RCC-approved mTOR inhibitors everolimus or temsirolimus.

Furthermore, we identified 13 cluster-discriminative marker genes targetable by repurposed FDA-approved drugs (**Fig. 41**) that can provide alternative treatment strategies to meet the need for novel anticancer therapeutics (72). One highly elevated EMT marker in cluster 1 tumors is the signaling protein S100A4, which is a target of the antipsychotic drug trifluoperazine (**Fig. 41B**). It has been shown that trifluoperazine can be repurposed as an anticancer drug and successfully prevent or minimize metastatic spread in different cancers by inhibiting S100A4 through protein oligomerization (73, 74). Targeting S100A4 could also be advantageous for reducing hypoxic signaling in cluster 1 tumors, since it was revealed as target of HIF1α. Another gene highly overexpressed in cluster 1 tumors upon hypoxic signaling is SLC2A1, also known as glucose transporter 1 (GLUT1). This marker gene can be targeted by the intravenous anesthetic etomidate (**Fig. 41B**), which has been reported to reduce GLUT1-mediated glucose transport and associated tumor growth in lung cancer (75, 76). SLC2A1 has also been revealed to be related to angiogenic processes such as coagulation and platelet degranulation, as well as the marker CA1, which participates in hypoxia-induced glucose

transport, angiogenesis and pH regulation (77). Multiple inhibitory FDA-approved sulfonamides such as acetazolamide, methazolamide, dichlorophenamide, dorzolamide and brinzolamide can target CA1 (**Fig. 41B**), and have been reported to counteract metastatic dissemination in different RCC cell lines (78).

Overall, we uncovered two distinct molecular phenotypes of ccRCC tumors and identified a cluster-discriminative biomarker panel targetable by several repurposed FDA-approved drugs. Future directions warrant the investigation of tumor-oriented intervention by combining these novel therapeutics with current medical treatment options for RCC.

*Chapter 4*

# Phosphoproteome Analysis of ccRCC

## 4.1 LITERATURE REVIEW

### 4.1.1 Phosphorylation

Phosphorylation is the most common and most important post-translational modification (79). It is assumed that around 75% of the human proteome is phosphorylated in order to regulate and maintain biological processes such as cell division, protein synthesis, cell growth, apoptosis and development (79, 80).

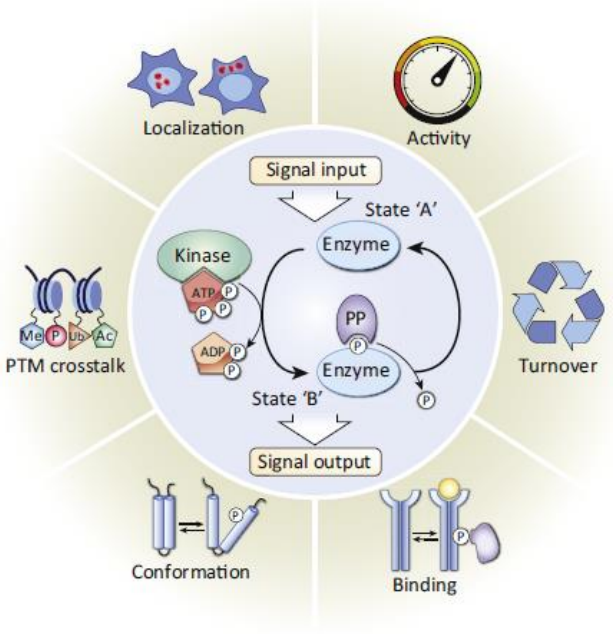


**Figure 43 | Protein phosphorylation acts as a molecular switch in cells that leads to different functional changes** (adapted from (80)).

Phosphorylation is considered to be a rapid response mechanism of cells (milliseconds to seconds) to react to internal and external signals (80). It acts as molecular switch that can alter enzymatic activity, localization, conformation, protein turnover and interactions of proteins (80). The reversible addition (from ATP) of a phosphate group to different residue types is catalyzed by kinases, while its removal is carried out by phosphatases, respectively (**Fig. 43**). In total, 518 human kinases are known while the number of phosphatases is considered to be 226 (79). In fact, kinases, which are in many cases enzymes or receptors, are attractive targets for numerous FDA-approved drugs due to their broad implications in cellular processes (79, 81).

### 4.1.2 Phosphoproteomics and its challenges

In recent years, high-throughput mass spectrometry-based phosphoproteomics analysis of biospecimens has emerged as a promising tool to comprehend underlying cellular signaling mechanisms. The more than 200,000 phosphosites deposited in databases such as PhosphoSitePlus and PhosphoELM were mainly identified by MS (~95%) (80). With the improved developments in phosphoproteomics pipelines, it is nowadays feasible to achieve a depth of 15,000-20,000 phosphosites for a meaningful coverage of the investigated biological profile (80). However, it is doubtful whether all identified phosphosites are truly biologically relevant or whether they are “noise” modifications that are more or less random events and detected due to the high sensitivity of modern mass spectrometers (80). One way to assess the functional impact of a phosphorylation event, is by determining the degree of phosphorylation for a specific site (stoichiometry), which represents the number of copies of a protein that are phosphorylated at a certain amino acid at a given time point compared to the total number of the protein (80).

In general, phosphoproteomics studies are impaired by further limitations (80, 82), which are as follows:

(a) the requirement of large sample amounts (>1 mg)

(b) analysis of peptides in the low-abundance dynamic range

(c) extensive fractionation

(d) low-throughput workflows

(e) the reduced ionization efficiency of phosphorylated peptides due to the negative charge

(f) phosphopeptide losses due to formation of unwanted metal-ion complexes during sample preparation and chromatography

(g) the labile nature of the phosphate group during fragmentation

(h) poor robustness and reproducibility of quantification

(i) and the correct localization of the phosphosite by search algorithms

(j) quantifications of phosphoproteins rely on single peptides due to the enrichment of phosphopeptides while global analyses benefit from many available peptides for protein quantification

Newer pipelines such as the EasyPhos platform provide a more streamlined application of phosphoproteome analysis by applying steps from digestion to phosphopeptide enrichment of low sample amounts in the 96-well-scale without the need for fractionation allowing rapid, high-throughput and automated in-depth proteomics analysis (80).

Another caveat of phosphoproteomics approaches is the compromise of *in silico* estimation of kinase activities due to the lack of technological opportunities for large-scale experimental validation. The prediction of kinase activities holds a strong bias towards well-studied kinases (83). One recently developed method tackling this limitation is the so-called QuantaKinome application, which probes kinase activities by tracking T-loop peptides using targeted proteomics after phosphopeptide enrichment (83). The

majority of kinases are activated by phosphorylation of their activation peptide (T-loop), which serves as a surrogate for the activation status of the kinase. However, this method is limited to 178 kinases (one third of the human kinome) with known T-loop phosphorylations.

### 4.1.3 Phosphorylation in cancer

Many kinases are involved in the oncogenesis of tumors due to aberrant activation or dysregulation caused by mutations or defective regulatory mechanisms (79). More than 1,000 variations in the expression of kinases in different tumor types have been described so far (79). The first kinase detected in 1976 to be involved in tumor cell proliferation was the tyrosine kinase v-Src, which is highly activated due to a mutation in the Y527 residue (79). To date, many different kinases have been exposed to be oncogenic biomarkers such as the Her2/neu kinase in breast cancer or the receptor tyrosine kinase EGFR in colon cancer (79). Due to their wide implication in cancers, tyrosine kinases are attractive targets of inhibitors, leading to 17 approved small molecules currently used for cancer treatment (79).

### 4.1.4 The phosphoproteome of RCC

There have only been a few attempts to portray the phosphorylation signature of ccRCC tumors. Deb and colleagues (84) report that the ccRCC phosphoproteome is in general distinct from that of five other cancer types as lower phosphorylation levels for 161 common phosphosites were determined by comparing the respective CPTAC datasets. In fact, besides the study of Peng and colleagues (16), the CPTAC phosphoproteome (15) represents one of the two large-scale phosphoproteomics profiles of ccRCC tissues with more than 40,000 identified phosphosites. However, while the CPTAC analysis uncovered druggable phosphorylation signaling events such as the EGFR-mediated MAPK and AKT1-mTOR pathways (**Fig. 44**), Peng and colleagues were rather interested

in the general identification of so-called "missing proteins" (evidenced only at the transcript level or predicted to exist at the protein level).

mTOR is known to be particularly active in RCC, where it increases the stabilization of HIF-1α and cyclin D driving the overexpression of VEGF and consequently angiogenesis (79).

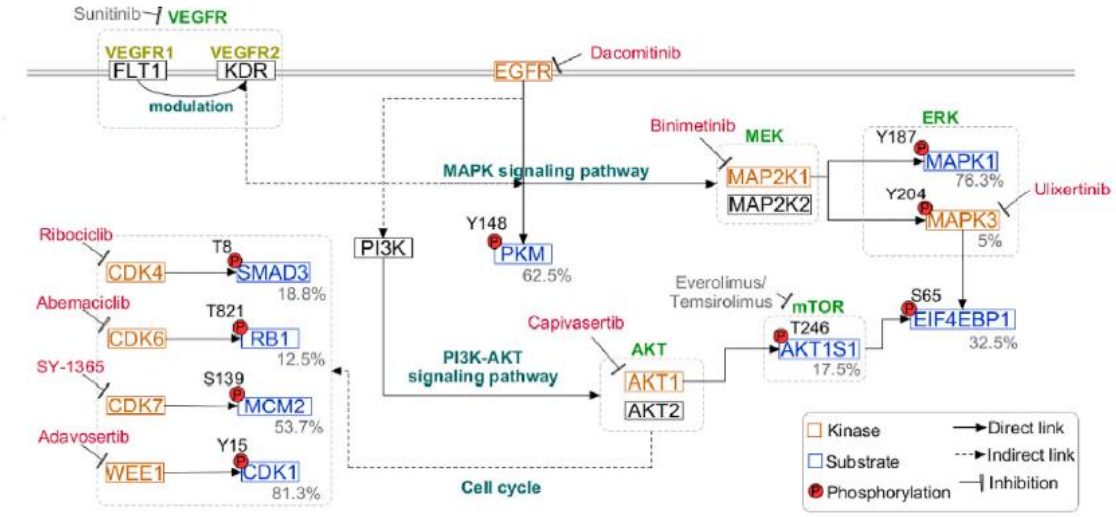


**Figure 44 | Suggested activated phospho pathways in ccRCC tumors.** Potential drugs are shown in red, FDA-approved drugs for ccRCC are shown in grey (adapted from (15)).

Another attempt described by Haake and colleagues (14) aimed to characterize tyrosine phosphorylations by pY-immunoprecipitation in ccRCC tissues and cell lines. Elevated tyrosine phosphorylations were determined for the EGFR, MAPK as well as the FAK pathways in the different biospecimens, which can be effectively targeted by various tyrosine-kinase inhibitors (14).

## 4.2 MATERIALS & METHODS

### 4.2.1 Phosphopeptide enrichment

For the enrichment of phosphorylated peptides, the same dimethyl-labeled SCX fractionated tumor/normal samples as described in Chapter 2.2 were used. The enrichment step was performed on-column using 500µg of titanium dioxide ($TiO_2$) beads (5µm, Sachtleben) packed into micro-columns (20µL GELoader, eppendorf), which were equilibrated with loading buffer (80% ACN and 6% TFA) (85). The dried SCX peptide fractions, which were stored at -80°C, were reconstituted in loading buffer to a final concentration of 1 µg/µL and slowly loaded onto the positively charged columns at 50$g$. The bound phosphopeptides were washed with 0.1% TFA in 50% ACN and then eluted with 20µL 10% ammonia solution into a new tube containing 35µL of 10% FA. The eluted peptides were further acidified by addition of 3µL FA and immediately analyzed in LC-MS/MS.

### 4.2.2 Data acquisition

The phospho enriched fractions were analyzed in triplicate with a 120 min linear gradient on an UltiMate 3000 RSLCnano reversed phase chromatographic platform (Thermo Fisher Scientific) coupled to a Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). The fractions were loaded onto an in-house packed 100 µm i.d. × 17 cm C18 column (Reprosil-Gold C18, 5 µm, 200Å, Dr. Maisch) and run with a flow rate of 300 nL/min. The chromatographic separation of the peptides started at 4% of solution B (ACN with 0.1% FA) and gradually increased to 25% in 67 min. The gradient continued from 25% to 45% of solution B in the next 20 min. Peptides in the mass range of 400−1,500 m/z and with a positive polarity were allowed for detection in data-dependent mode. For the MS1 spectra acquisition the resolution was set to 120,000, the AGC target to 1e6 and the maximum injection time to 60 ms. The top 20

most intense peptides per cycle (Top20) were selected for fragmentation in the HCD cell with a NCE of 27. MS2 spectra acquisition was conducted at a resolution of 30,000, an AGC target of 1e5, a maximum injection time of 60 ms and a fixed first mass of 100 m/z. Furthermore, the isolation window was set to 2.0 m/z, the dynamic exclusion to 20 s, the minimum AGC to 3e3 and the charge exclusion was set as unassigned, 1.

### 4.2.3 Data processing

Peptide identification and quantification from raw MS data files were done with PD (v1.4, Thermo Scientific) using the Mascot (v2.5.1, Matrix Science) search engine. Raw files were analyzed in the "MudPIT" processing mode via the Daemon utility of PD to merge the six fraction files from a technical replicate. The PSMs were searched against a Swissprot database containing 21,039 entries for Homo sapiens retrieved from Uniprot in March 2016. Trypsin was selected as hydrolytic enzyme with a maximum number of allowed missed cleavages of 2. For peptide identification a mass tolerance of ±20 ppm for precursor masses and ±0.05 Da for fragment ions was selected. For dimethyl labeling the 2plex dimethyl-based heavy/light quantitation method with a mass precision requirement of 2 ppm for precursor ions was used. Light and heavy dimethylation of peptide N termini and of lysine residues, phosphorylation at serine, threonine and tyrosine residues, as well as methionine oxidation were set as dynamic modifications. Cysteine carbamidomethylation was set as fixed modification. Determination of phosphosite localizations was done by using the implemented node phosphoRS (v3.0). The FDR for peptide and protein identifications were set to 1%. Furthermore, a filter was applied to allow only peptide identifications with medium and high confidence, with a sequence length between 7 and 25, a Mascot score >20 and a peptide rank of minimum 1.

### 4.2.4 Normalization and filtration

Quantification ratios of samples with swapped dimethyl labeling were converted to heavy/light (H/L) format. The data was filtered for peptides quantified in at least two out

of the 39 technical replicates. All quantification ratios were normalized to the median of the non-phosphorylated peptides in the sample and then $\log_2$ transformed. These peptides are hereafter denoted as "quantified peptides". The phosphorylated peptides were further filtered for ≥50% site probability, which led to a total of 9,431 quantified phosphopeptides.

### 4.2.5 Statistical analysis

To determine phosphopeptides significantly differentially abundant between tumor and normal tissues, we used the one-sample Wilcoxon signed-rank test. More specifically, the Python function scipy.stats.wilcoxon (v0.14.0) was extended by the feature to omit missing quantification values and was then applied peptide-wise on the technical replicates data. *P*-values <0.05 were considered statistically significant. A cut-off of ±1.5 was applied on the median $\log_2$ fold change values of the significant phosphopeptides.

For the determination of statistical difference between phospho- and non-phosphopeptide distributions, the Kolmogorov-Smirnov test was applied via the Python function scipy.stats.ks_2samp. *P*-values <0.05 were considered statistically significant.

### 4.2.6 Functional annotations

#### 4.2.6.1 Gene Ontology analysis

Significantly up-and down-regulated phosphorylated proteins were annotated by their GO biological process association using the PANTHER (v14) platform (37) with an FDR setting of 0.05. Furthermore, the proteins were complemented by their function in cancer using the COSMIC CGC database (v91) (86), and by associated FDA-approved drugs with inhibitor function retrieved from the Drugbank database (v5.1.7, access August 2020) (69). Chemical elements were excluded as drugs.

#### 4.2.6.2 EMT annotation of phosphosites

To further characterize the functionality of the phosphopeptides, a previously published study from our group on the epithelial-to-mesenchymal phosphoproteome (87) was used for sequence-based comparison with our significantly dysregulated phosphopeptides to especially find mesenchymal phosphorylations elevated in the tumor samples.

#### 4.2.6.3 Biological functionality of phosphosites

Moreover, we assessed the probability of biological relevance for the significant phosphopeptides by taking advantage of the approach of Ochoa and colleagues (88) and annotating the phosphopeptides by their reported functional score. More specifically, both phosphopeptide datasets were matched by the protein position of the phospho modification and the residue type. Only phosphopeptides with a score >0.5 were considered functional.

### 4.2.7 Kinase-substrate network recreation

To predict kinase activities in the tumor and NATs, we implemented the NetworKIN3.0 (89) algorithm on the significantly regulated phosphopeptides, which annotates kinases based on STRING interactions and phylogenetic links within kinase families. Only kinase-substrate predictions with a NetworKIN score >5 were further considered for network recreation using the visualization software Cytoscape (v3.8.0). The network components were annotated by their GO cellular compartment feature using QuickGO (40), limited only to Uniprot assignments. In case of multiple cellular allocations, priority was given as follows: “membrane”, “nucleus”, “cytoplasm”, “extracellular” and “other”, respectively. Additionally, networks were complemented by the enriched Reactome pathways, respectively, retrieved as over-representation analysis via the web tool WebGestalt (v2019) (36) at default settings.

The predicted kinases were mapped to the human kinome phylogenetic tree using the KinMap$_{beta}$ tool at www.kinhub.org (access March 2021) (90).

### 4.2.8 Motif analysis

To find conserved sequence motifs, the significantly regulated phosphopeptides were centered to their phosphosite and the surrounding ±7 amino acid sequence window was added. The web tool MoMo (v5.3.3) (91) applying the motif-x algorithm was used for prediction with 15 residue motif width, 20 occurrences and a *p*-value setting of <0.000001.

For the determination of the prognostic power of kinases, the online tool Kaplan-Meier Plotter (92) was taken advantage of using the TCGA-KIRC cohort dataset (n=530) and "auto select best cutoff" option to optimally partition the patients based on the lowest *p*-value and the highest hazard ratio.

### 4.2.9 Comparison with other phospho data

The generated phosphoproteomics data was compared to the corresponding global proteome of the same tissue samples described previously (Chapter 2) (93). The abundance of the common significantly regulated proteins in the phospho and global proteome were confronted by comparing the $\log_2$ tumor/normal ratios and PSM counts of the phosphopeptide with the average of all respective non-phosphopeptides of the protein of interest.

The identified phosphopeptides of this study were also compared with the PhosphoSitePlus database (access November 2020) (94) by matching the -/+7 aa sequence windows. We further compared our ccRCC phosphoproteome with that characterized by the CPTAC study (15) and matched peptides by their sequences.

## 4.3 RESULTS

### 4.3.1 Shotgun proteomics approach achieves deep coverage of ccRCC phosphoproteome

In this study, a comprehensive quantitative phosphoproteomics analysis of frozen ccRCC tissues has been conducted to illuminate signaling pathways mediated by protein phosphorylation. For this purpose, the dimethyl-labeled fractionated peptides of tumor and normal adjacent tissues from our previous global proteomics analysis (Chapter 2) (93) were subjected to phosphopeptide enrichment by metal oxide affinity chromatography (MOAC) using $TiO_2$ beads and were analyzed as technical triplicates (**Fig. 45**).

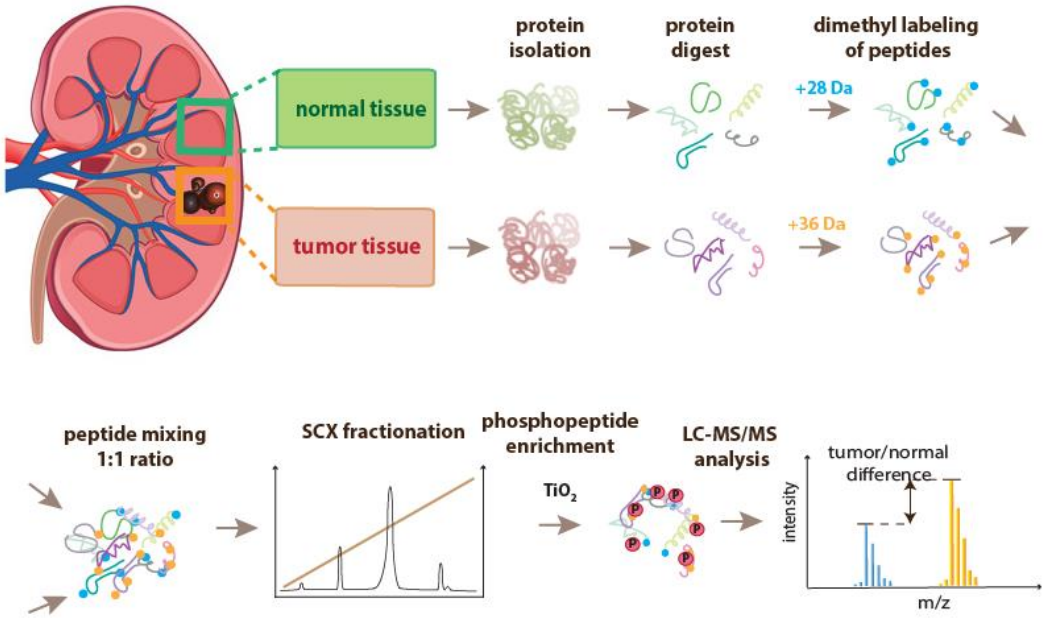


**Figure 45 | Phosphoproteome analysis of clear cell Renal Cell Carcinoma.** Experimental outline of dimethylation-based quantitative phosphoproteomics approach.

The discovery cohort consisted of 13 patients of different age, grade, stage and gender. In general, the identification and quantification of phosphorylated and non-phosphorylated peptides in the enriched samples were robust across patients and spanned approximately 12 orders of magnitude in the tumor/normal fold change dynamic range (**Fig. 46**).

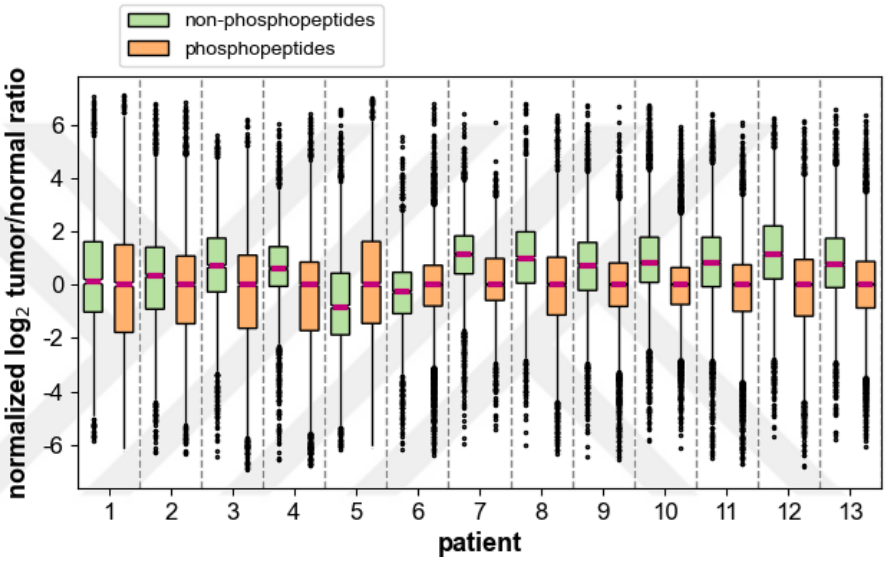


**Figure 46 | Distribution of quantified phospho- and non-phosphopeptides across all patients.** For cohort information see Table A1.

In total, we identified 16,253 phosphopeptides with an FDR of 0.01, which were assigned to 3,486 unique phosphoproteins (**Fig. 47**). The majority of the identified phosphosites (15,136) were class I sites with a location probability of ≥75%. Furthermore, 9,431 dimethyl-labeled phosphopeptides were quantified in both tumor and NATs, which corresponded to 2,443 unique phosphoproteins.

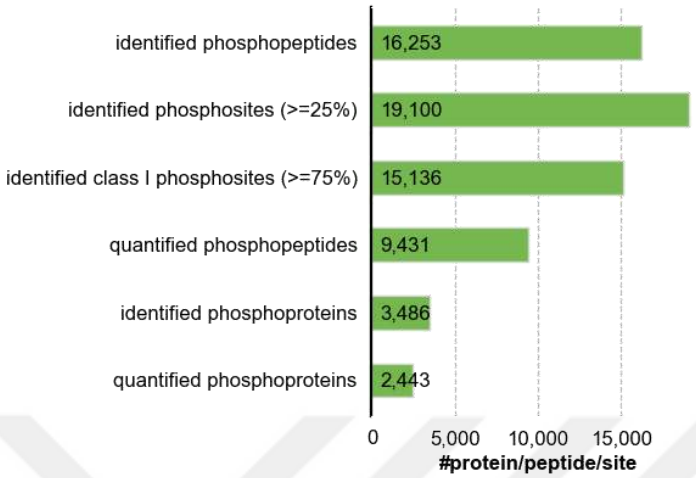


**Figure 47 | Number of identified and quantified phosphopeptides and phosphoproteins.**

In general, the abundance levels of quantified peptides varied over four orders of magnitude in dynamic range (**Fig. 48A**). Moreover, the fold change distributions of quantified phosphopeptides and non-phosphopeptides followed near normal patterns, however, a significant skewness of phosphopeptide ratios towards positive values was noticeable suggesting enhanced phosphorylation events in the tumor tissues compared to NATs (**Fig. 48B**).

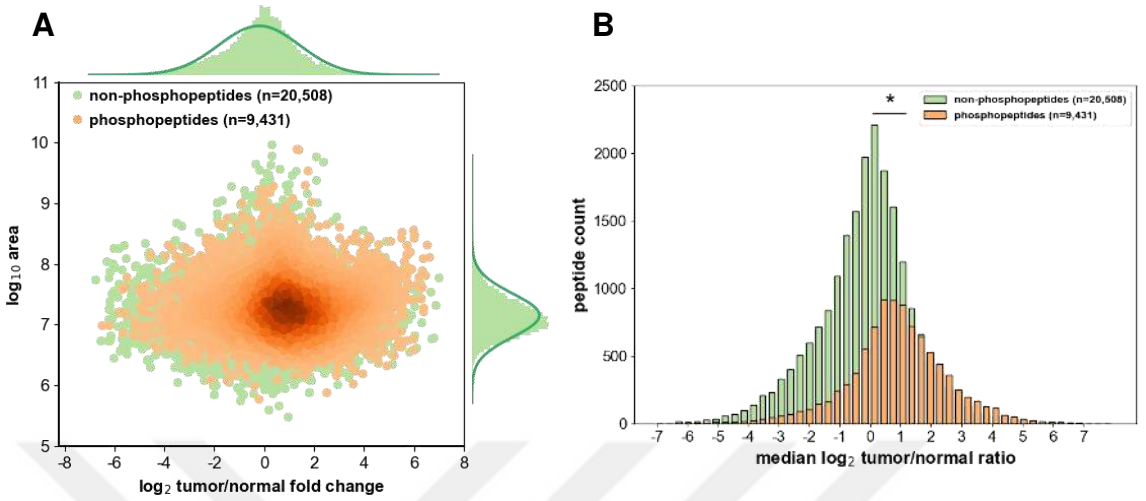


**Figure 48 | Distribution of quantified peptides. (A)** MA plot depicting intensity distribution of quantified phospho- and non-phosphopeptides. Top and right histograms show distribution of fold change values and intensities of non-phosphopeptides, respectively, with density lines representing theoretical normal distribution. **(B)** Histogram showing general distribution of median tumor/normal ratios of quantified phospho- and non-phosphopeptides. Kolmogorov-Smirnov test was used for statistical determination of skewness between distributions, *p*-value <0.05.

As expected, most phosphosites were serine residues (88.4%), followed by threonine (10.6%) and a few tyrosine residues (1.01%) coinciding with previously reported frequencies (79) (**Fig. 49A**). Furthermore, most peptides were singly phosphorylated (85.7%) while only a small portion of peptides harbored multiple phosphosites, which is also in line with previous observations (16) (**Fig. 49B**).

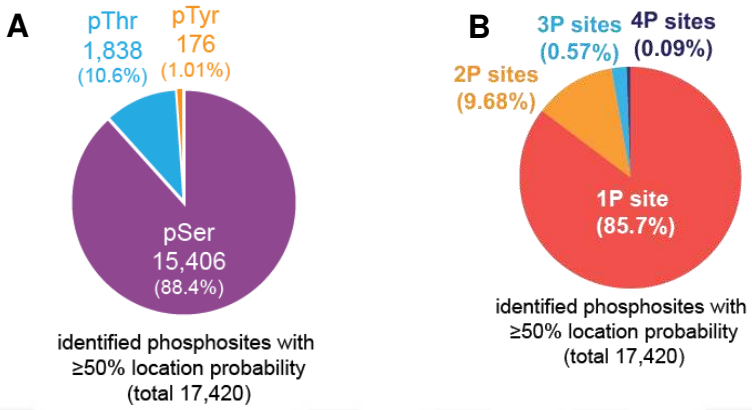


**Figure 49 | Residue and phosphosite number occurrences. (A)** Frequency distribution of identified serine (pSer), threonine (pThr) and tyrosine (pTyr) phosphosites with ≥50% location probability. **(B)** Frequency distribution of identified singly (1P site) and multiply (≥ 2P sites) phosphorylated peptides with ≥50% site probability.

### 4.3.2 Significantly regulated phosphorylations play a role in tumor invasion and are conserved kinase recognition sites

Out of more than 9,000 quantified phosphopeptides, we determined 1,336 as statistically differentially abundant between tumors and NATs (Wilcoxon signed-rank test with 0 as reference value, *p*-value <0.05, fold change cut-off ±1.5), which were assigned to 700 unique proteins (**Fig. 50**). The majority were up-regulated phosphopeptides supporting our observation of elevated phosphorylation events in ccRCC tumors (refers to **Fig. 48B**).

Interestingly, the 598 significantly up-regulated phosphoproteins were primarily associated with intermediate filament organization, cell growth and morphogenesis, while the 109 significantly down-regulated phosphoproteins were involved in cell-cell adhesion (**Fig. 51**). These aberrant biological processes suggest an increase in cell motility and invasiveness, and a reduction in cell-to-cell contact, which are indicators of malignant

tumors and have been previously linked to worse patient outcome (95).

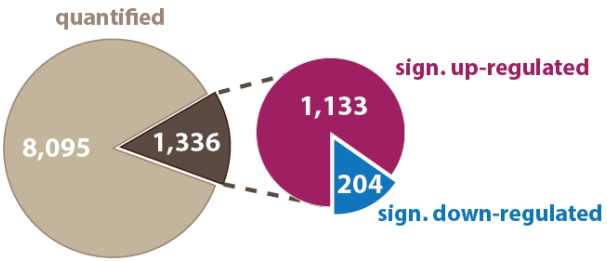


**Figure 50 | Overview of the number of quantified phosphopeptides.** Inner pie chart represents number of significantly regulated phosphopeptides with the majority being up-regulated in the tumor tissues.

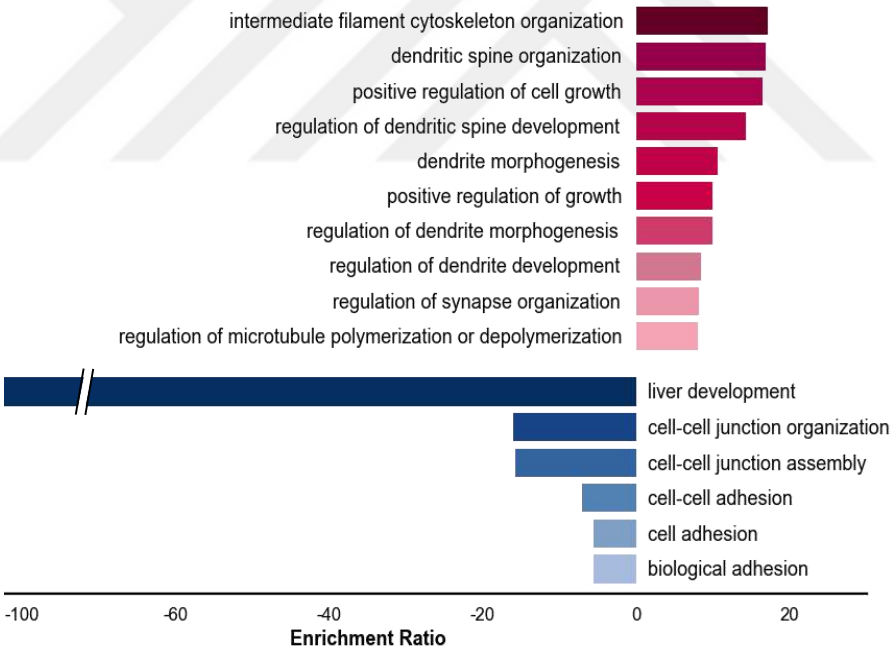


**Figure 51 | Significantly dysregulated biological processes in ccRCC tumors compared to adjacent normal tissues.** Top 10 PANTHER GO-Slim biological process annotations of significantly up-regulated (pink) and down-regulated (blue) phosphoproteins, respectively.

To assess whether the significantly differentially abundant phosphoproteins between tumor and NATs were truly dysregulated at the phosphorylation level rather than the general protein level, we compared the phosphoproteome with our global proteome data of the same ccRCC tissues (**Fig. 52**) (Chapter 2) (93). The majority of the 700 phosphoproteins (581 proteins) were insignificantly abundant between tumor and NATs in the global proteome analysis (83%), while 119 phosphoproteins (17%) were also significantly regulated on the global level (**Fig. 52A**). This implies that most of the significant phosphorylation events might not affect the abundance of a protein rather than other factors such as its activity, conformation or localization. To rule out whether the phosphorylation modification itself or the general abundance of the protein is elevated in the tumors, we confronted for the 119 common proteins the abundance levels of the phosphopeptide of interest with that of all non-phosphorylated peptides of the respective protein. Indeed, the $\log_2$ tumor/normal ratios and PSM counts of phosphorylated peptides were significantly higher abundant compared to the related non-phosphopeptides of the proteins of interest (Kolmogorov-Smirnov test, *p*-value <0.0001) (**Fig. 52B** and **C**), suggesting that truly the phosphorylation events are elevated in the tumors.

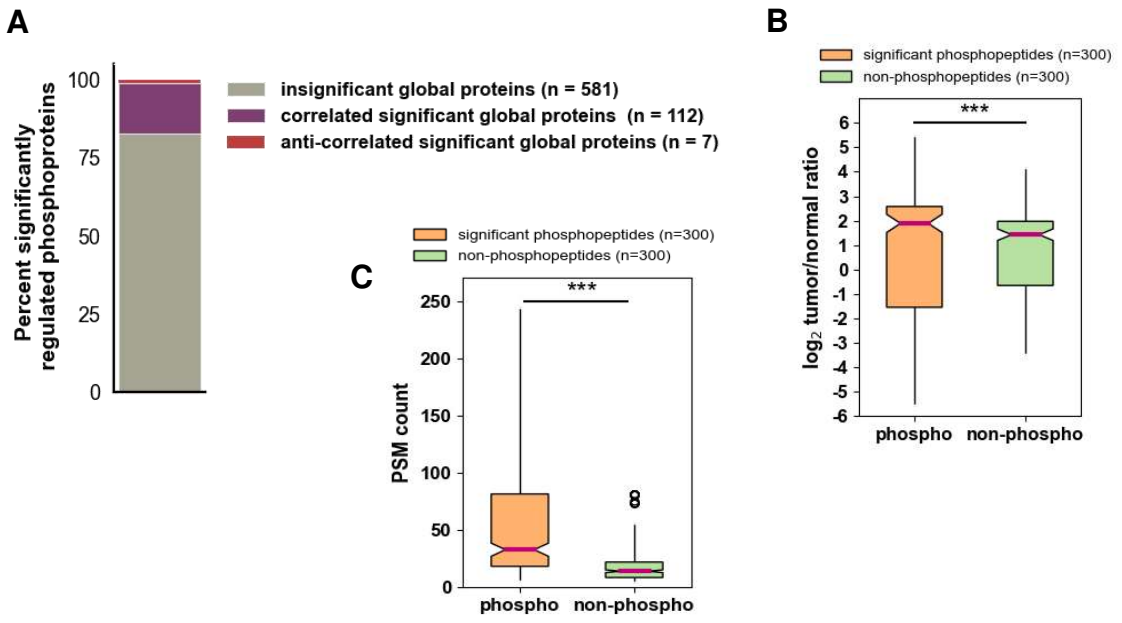

**Figure 52 | Regulation of phosphoproteins on global level. (A)** Corresponding regulation of significantly regulated phosphoproteins in global proteome of same ccRCC tissues as described in Chapter 2. **(B-C)** Tumor/normal fold change and PSM count distributions, respectively, of phospho- and non-phosphopeptides of 119 significantly regulated phosphoproteins that are also significantly dysregulated on global level (refers to **A**).

Furthermore, the 1,133 significantly up-regulated phosphopeptides were associated with conserved phosphorylation recognition sequences (**Fig. 53**) such as the proline-containing serine motifs xxxxxxxSPxxxxxx and xxxxxPxSPxxxxxx, which are known target sites of CDK1, CDK2 and CDK5, respectively (96). Another frequent kinase target motif was xxxxRxxSxxxxxxx, which is linked to several different kinase groups such as CaMK1, CaMK2, AKT, PKA and PKCα, respectively (96).

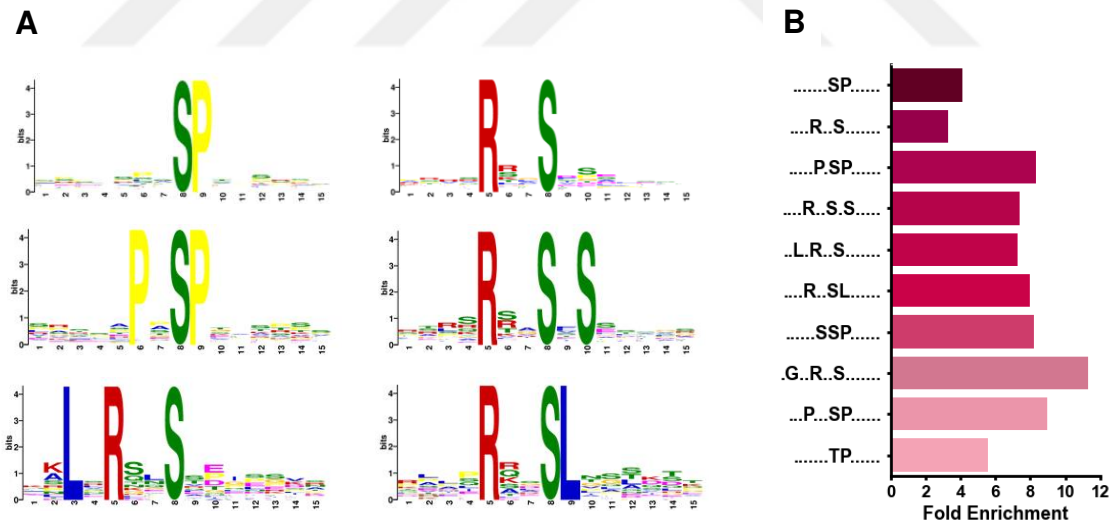


**Figure 53 | Recognition motifs of significantly up-regulated phosphosites. (A-B)** Enriched consensus motifs of significantly regulated serine and threonine phosphopeptides sorted by their ascending *p*-value. Phosphosites with ≥50% probability (centered to the phospho residue and enlarged by the ±7 aa sequence window) were used for the analysis with the motif-x algorithm.

### 4.3.3 CDK1 and PAK2 kinases are highly activated in ccRCC tumors

Kinases are the mediators of phosphorylation and belong to different kinase families. In order to unveil the most prominent kinase activities, we applied the kinase prediction tool NetworKIN on the determined significantly regulated phosphosites. We reduced the predictions to high-confident ones with a NetworKIN score >5 (329 predictions in total). In summary, out of the 700 significantly regulated phosphoproteins, 117 were substrates of a total of 73 different kinases. Among the most frequently activated kinases were different CDK kinases such as CDK1, CDK2 and CDK5 (**Fig. 54**), which confirms our previous finding that the conserved phosphosite recognition sequences of these kinases are highly enriched in our phospho data (**Fig. 53**). Furthermore, PAK2 is the second most activated kinase in the tumor tissues, followed by JNK1 and GSK3 beta (**Fig. 54**). Also, predictions for the Erk kinases Erk1 (MAPK3) and Erk2 (MAPK1) are highly enriched and suggest the activation of the MAPK pathway (15). In contrast, the activity for the kinases TGFbR2, PAK1, PKAC alpha as well as PDHK kinases was highly reduced in the tumor tissues compared to NATs.

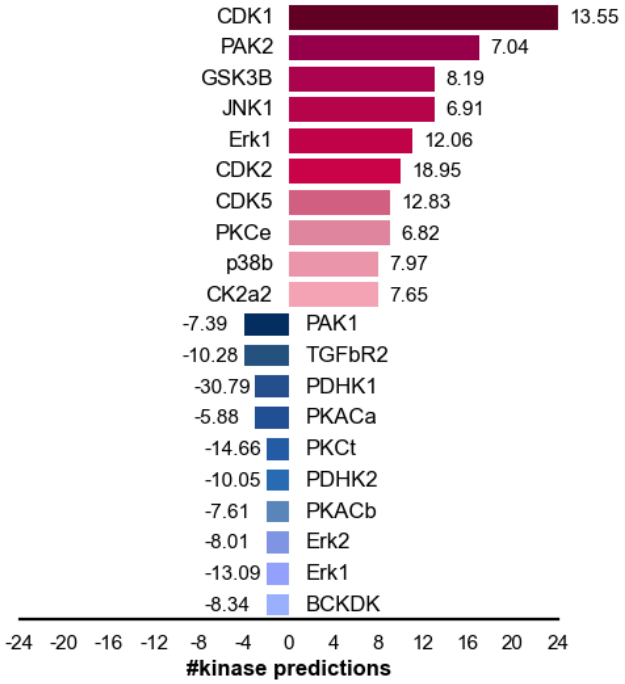


**Figure 54 | Kinase activities targeting significantly regulated phosphoproteins.** Predicted activated and reduced kinases in the tumor tissues compared to NATs according to the

NetworKIN3.0 tool. Top 10 kinases with highest number of predictions are shown. Average NetworKIN scores of predictions are shown as bar labels.

Almost all predicted kinases belonged to serine/threonine families, with only TYK2 and JAK2 belonging to the TK family of tyrosine kinases (**Fig. 55**). The frequency of activity across kinase families was comparable for both up- and down-regulated phosphopeptides with the majority of the kinases belonging to the CMGC and AGC families (**Fig. 56**).

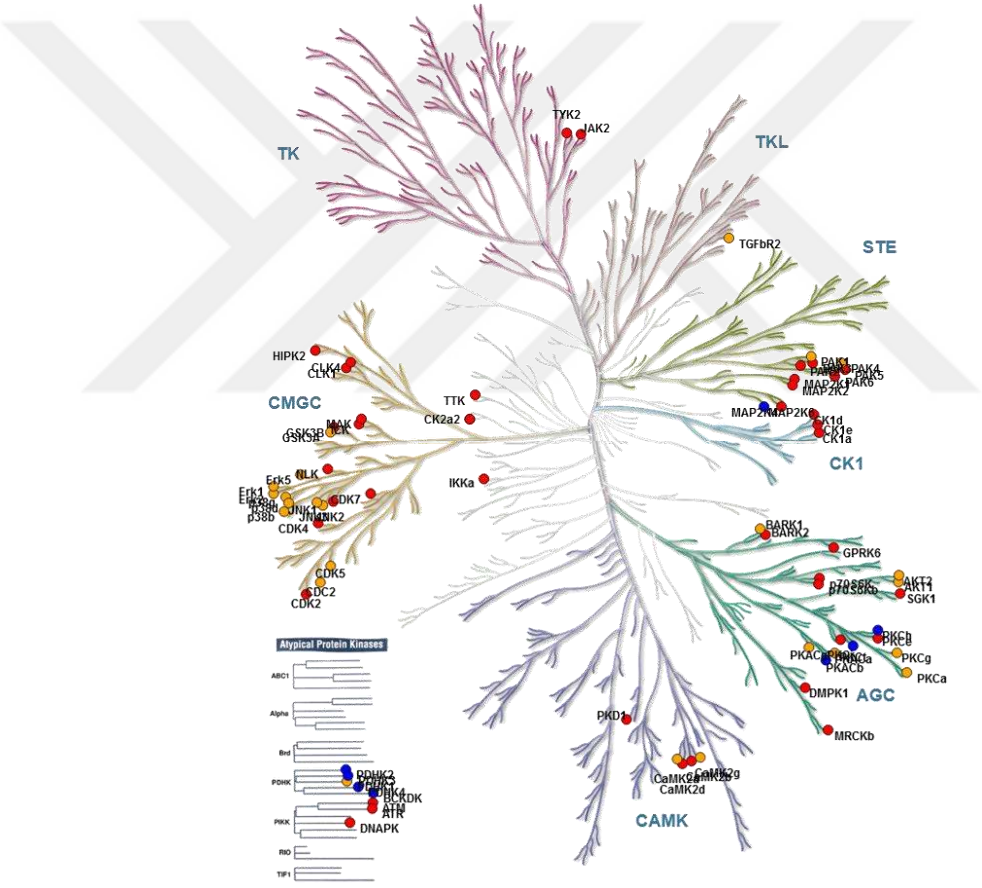


**Figure 55 | Human kinome tree of activated and reduced kinases in ccRCC tumors.** The web tool Kinhub was used with indications of predicted kinases associated with significantly up-regulated (red), up- and down-regulated (yellow), and down-regulated (blue) phosphopeptides, respectively.

Interestingly, rare kinase families such as PDHK were more associated with reduced phosphorylation events in the tumors, while rare PIKK-mediated phosphorylations were associated with up-regulated phosphosites (**Fig. 56**).

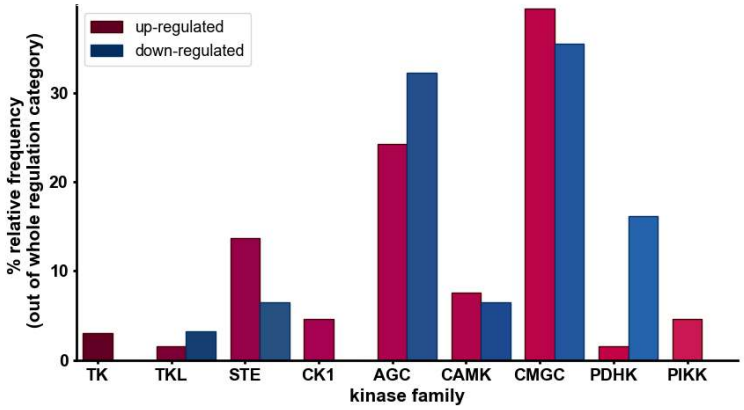


**Figure 56 | Frequency distribution of predicted kinases associated with significantly regulated phosphopeptides in the ccRCC tumors across kinase families.**

To illuminate the most prominent signaling pathways in the ccRCC tumors, the predicted kinase-substrate interactions was recreated as network and complemented with Reactome pathways (**Fig. 57**), which unveiled that certain malignant processes such as MAPK-related signaling, VEGF signaling and signaling by interleukins and second messengers shape the pathobiology of renal cancers. Moreover, the activation of RAC1 is the most enriched pathway in the tumors, with all known six PAK kinase isoforms predicted to be activated in the tumors (only PAK1 is associated with significantly up- as well as down-regulated phosphopeptides). We further annotated the genes in the network by their role in cancer (colored font) and noticed that several known oncogenes are substrates of the kinases such as EGFR, MAPK1, AKT1, AKT2, MAP2K1, MAP2K2, MAP2K4, MAP3K1, JAK2, STAT5B, CDK4 and NPM1. Furthermore, 17 substrates were identified to have a mesenchymal-associated phosphorylation (yellow border) such as PAK2, MAPK1 (Erk2), VIM, MAP1B, EIF4B, EIF4G1 and CLASP1, among others.

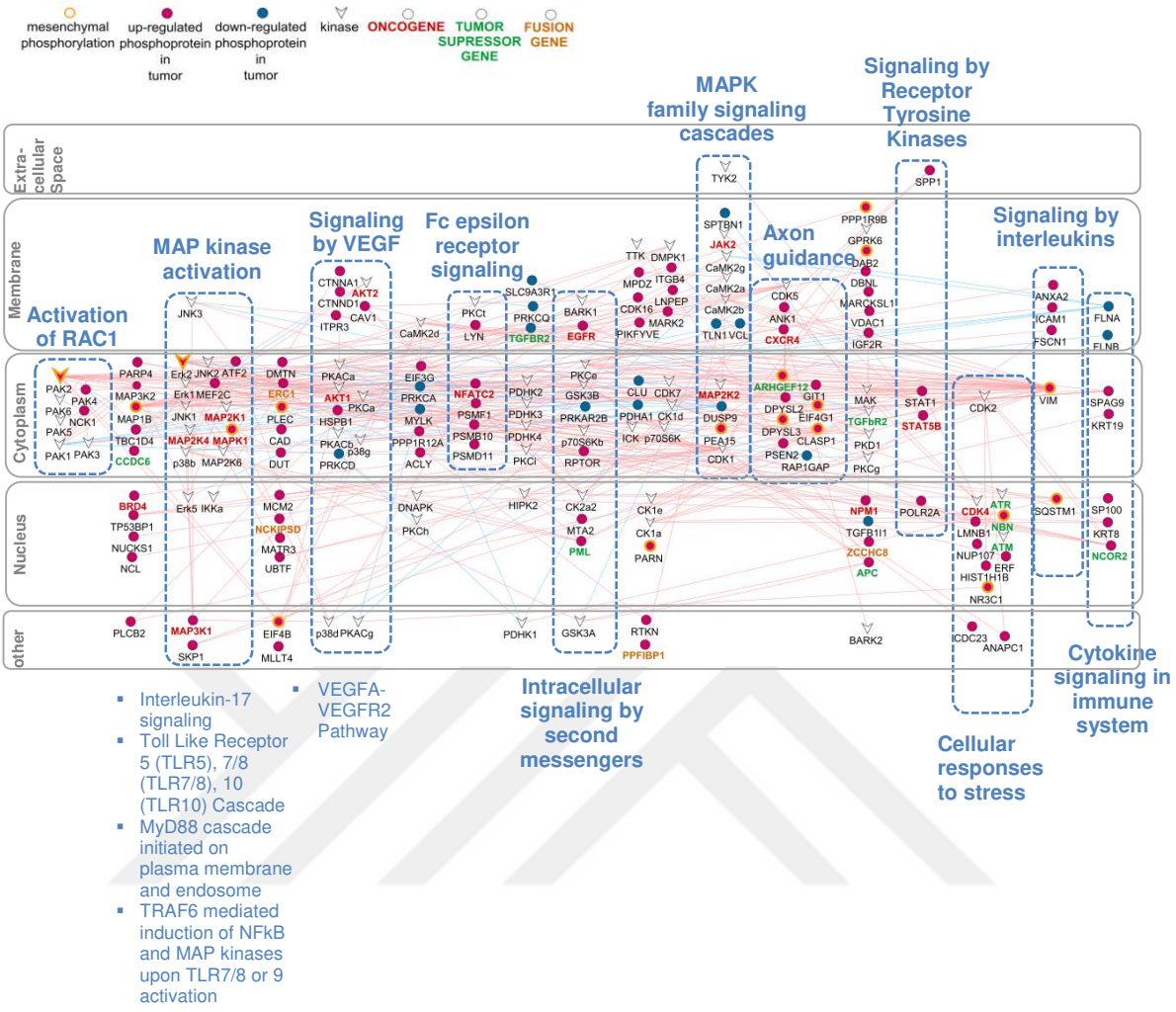


**Figure 57 | Kinase-substrate network depicting the most enriched Reactome pathways (from left to right).** Nodes are organized by their cellular compartment retrieved using the QuickGO web tool. Enrichment level of pathways decreases from left to right. Mesenchymal phosphorylations (orange border) as well as the role of substrates in cancer (node font color) are also indicated.

### 4.3.4 Functional phosphorylation events are targetable by FDA-approved drugs

Given that biologically relevant phosphorylations are redundant, which means that many modifications are non-functional (80, 88), we sought to unravel those most likely to contribute to tumor pathobiology. To this end, we compared our phosphopeptide data with the functionality analysis of Ochoa and colleagues (88), and determined that out of 1,336 significantly regulated phosphopeptides 434 are most likely functional (score >0.5) (**Fig. 58**).

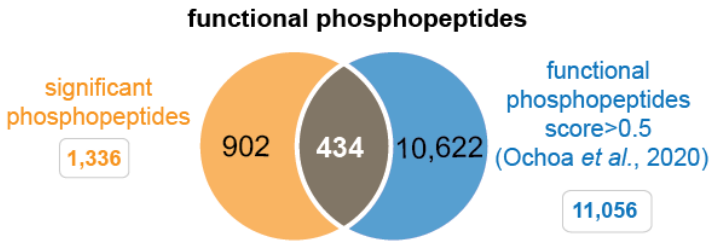


**Figure 58 | Number of significantly regulated phosphopeptides with predicted biological function (score >0.5) according to Ochoa *et al.* (2020).**

Next, we attempted to find the most relevant kinases and ranked the predicted kinases by their number of associated functional peptides (**Fig. 59**). This revealed that PAK2, CDK1, Erk1 (MAPK3) and Erk2 (MAPK1) are the kinases responsible for most of the functional phosphorylations in ccRCC tumors.

Since kinases are the major targets for cancer drugs (81), we further annotated the predicted kinases by their inhibitory FDA-approved drugs. Indeed, approximately 74% of the activated kinases (46 out of 62) are druggable (**Fig. 60**). A subnetwork for the three most active kinases with the indication of functional phosphorylations (black font) is shown in **Fig. 61**. While PAK2 and CDK1 can be targeted by the tyrosine kinase inhibitor

fostamatinib, Erk2 can be inhibited by the anti-inflammatory drug sulindac and by the antibiotic minocycline.

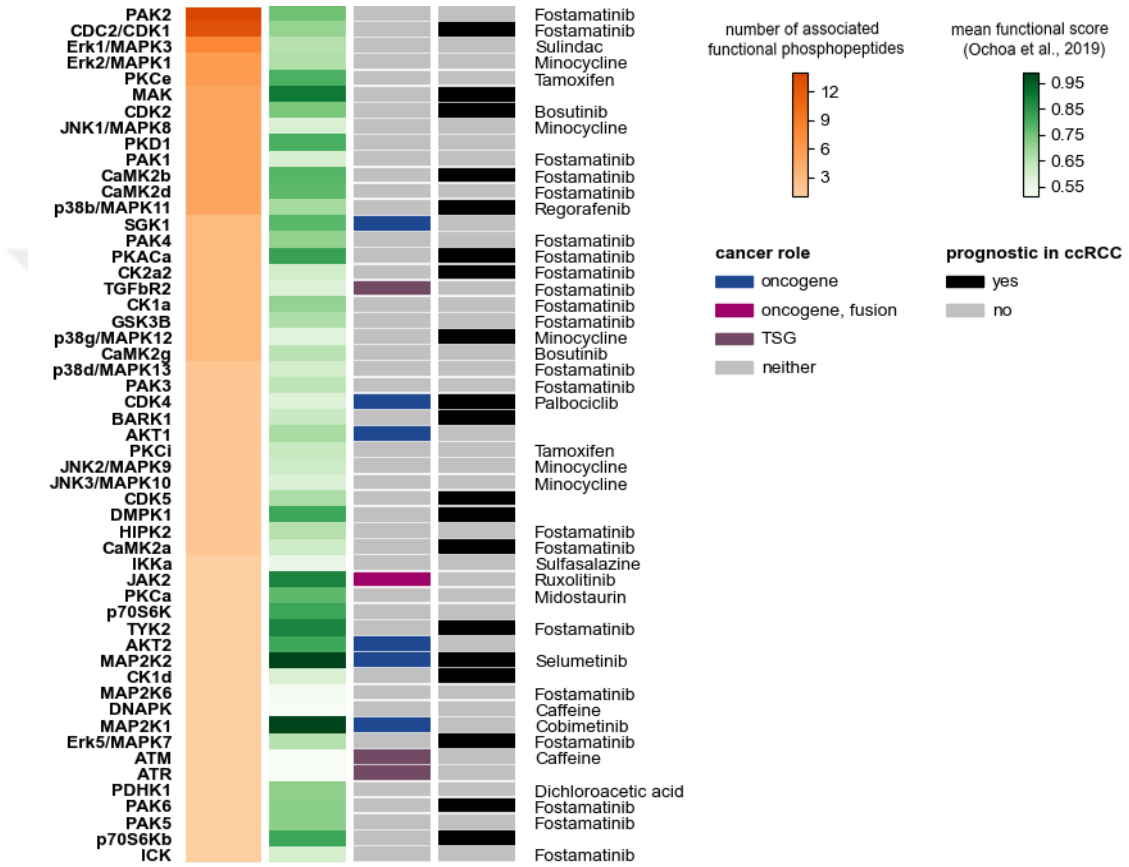


**Figure 59 | Druggable kinases associated with functional phosphorylations.** Predicted activated kinases associated with significantly up-regulated phosphopeptides with a functional score >0.5 according to Ochoa *et al*. (2020). Indicated drugs are associated FDA-approved therapeutics with inhibitor function. The prognostic feature of the kinases refers to Fig. 62.

Moreover, 20 out of the 53 kinases in **Fig. 59** have a prognostic value for ccRCC tumors based on TCGA-KIRC follow-up data (n=530 patients) including several CDKs such as

CDK1, CDK2, CDK4 and CDK5, as well as several MAPKs such as MAPK11, MAPK12, MAP2K2 and MAPK7 (**Fig. 62**), which stresses the importance to investigate cell proliferation related pathways.

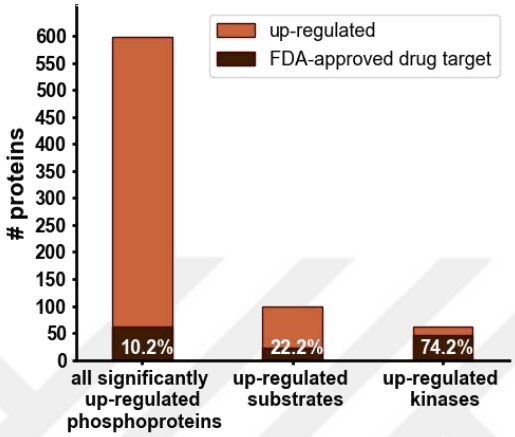


**Figure 60 | Percentage of druggable significantly up-regulated phosphoproteins and associated predicted kinases.**

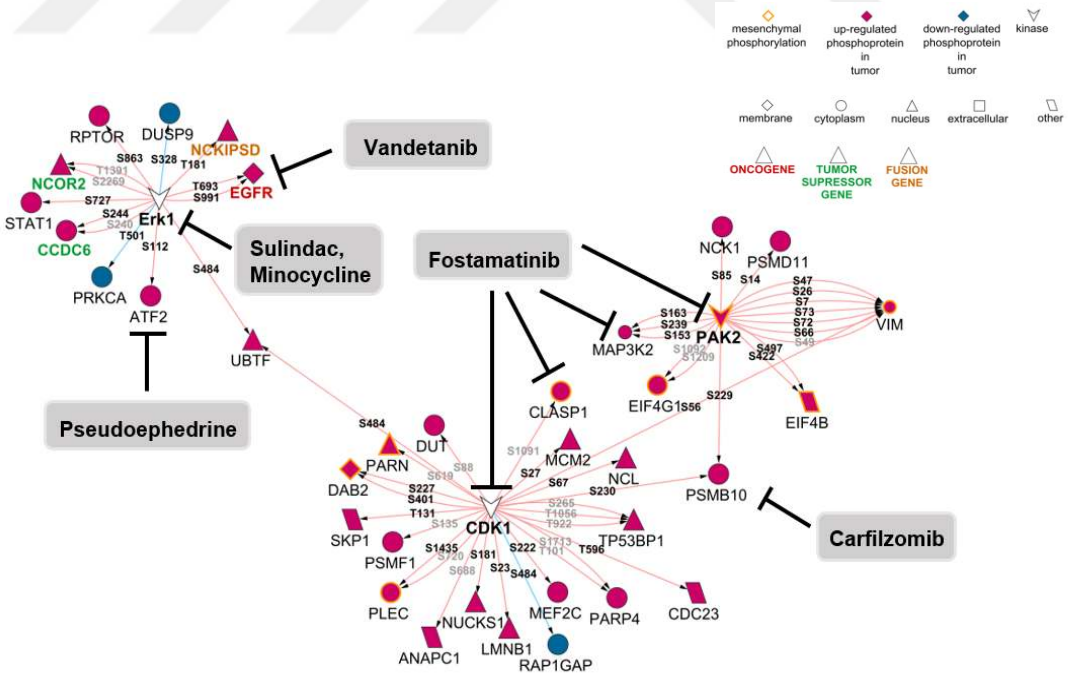


**Figure 61 | Subnetwork of top 3 activated kinases mostly associated with functional phosphorylations (refers to Fig. 59).** Non-functional phosphosites are indicated in grey.

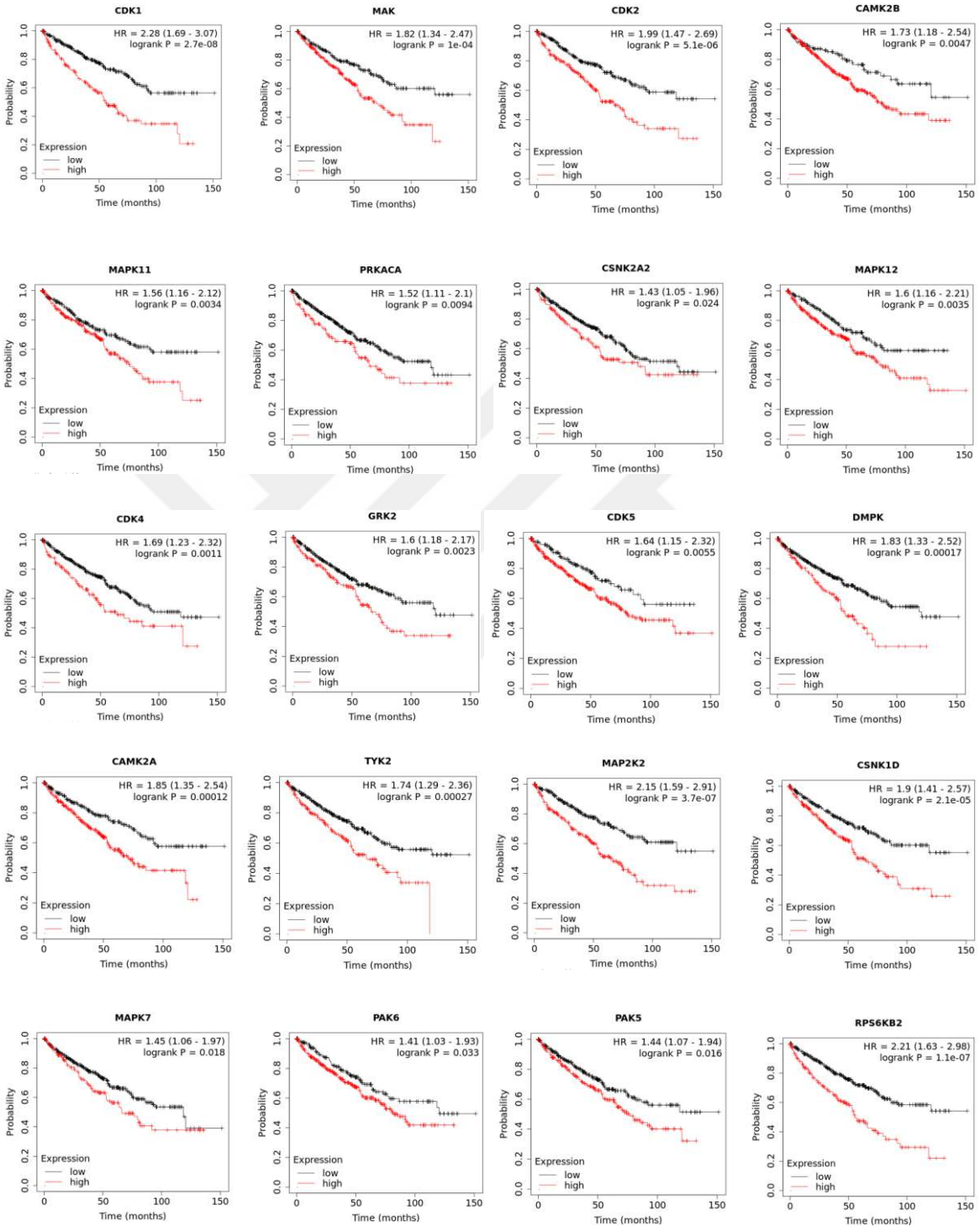
CDK1
HR = 2.28 (1.69 - 3.07)
logrank P = 2.7e-08
MAK
HR = 1.82 (1.34 - 2.47)
logrank P = 1e-04
CDK2
HR = 1.99 (1.47 - 2.69)
logrank P = 5.1e-06
CAMK2B
HR = 1.73 (1.18 - 2.54)
logrank P = 0.0047
MAPK11
HR = 1.56 (1.16 - 2.12)
logrank P = 0.0034
PRKACA
HR = 1.52 (1.11 - 2.1)
logrank P = 0.0094
CSNK2A2
HR = 1.43 (1.05 - 1.96)
logrank P = 0.024
MAPK12
HR = 1.6 (1.16 - 2.21)
logrank P = 0.0035
CDK4
HR = 1.69 (1.23 - 2.32)
logrank P = 0.0011
GRK2
HR = 1.6 (1.18 - 2.17)
logrank P = 0.0023
CDK5
HR = 1.64 (1.15 - 2.32)
logrank P = 0.0055
DMPK
HR = 1.83 (1.33 - 2.52)
logrank P = 0.00017
CAMK2A
HR = 1.85 (1.35 - 2.54)
logrank P = 0.00012
TYK2
HR = 1.74 (1.29 - 2.36)
logrank P = 0.00027
MAP2K2
HR = 2.15 (1.59 - 2.91)
logrank P = 3.7e-07
CSNK1D
HR = 1.9 (1.41 - 2.57)
logrank P = 2.1e-05
MAPK7
HR = 1.45 (1.06 - 1.97)
logrank P = 0.018
PAK6
HR = 1.41 (1.03 - 1.93)
logrank P = 0.033
PAK5
HR = 1.44 (1.07 - 1.94)
logrank P = 0.016
RPS6KB2
HR = 2.21 (1.63 - 2.98)
logrank P = 1.1e-07
Probability
Expression
low
high
Time (months)

**Figure 62 | Activated kinases with prognostic value associated with significantly up-regulated functional phosphorylations in ccRCC tumors.** Survival curves were retrieved from Kaplan-Meier-Plotter web tool for TCGA-KIRC cohort (n=530). Partitioning of patients into "low and high expression" of gene was conducted using "auto-select best cutoff" option. Only kinases with significant hazard ratios (HR) >1 for high expression of kinase are shown.

A summary of the most elevated signaling pathways orchestrated by phosphorylation in ccRCC tumors and relevant kinase inhibitors is given in **Fig. 63**.

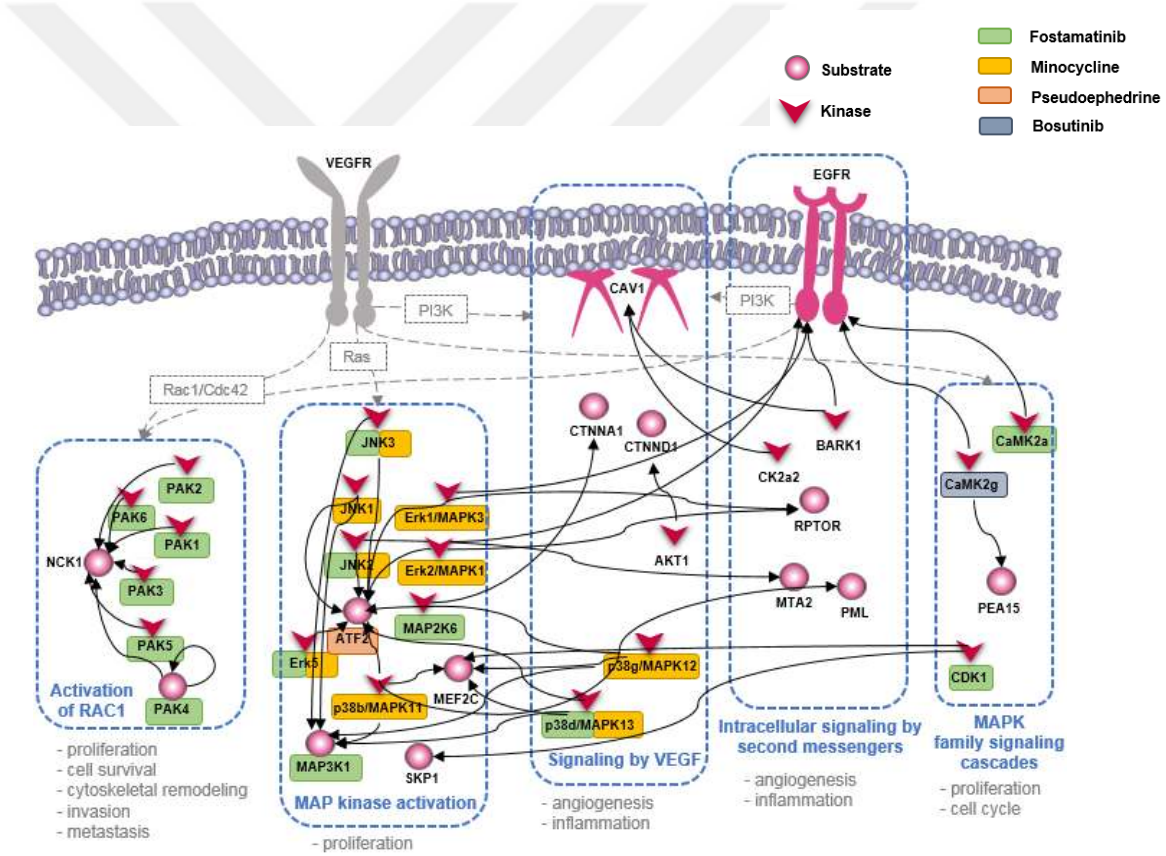


**Figure 63 | Summarized significantly enriched phosphorylated pathways in ccRCC tumor tissues targetable with FDA-approved drugs.** Enrichment level of pathways decreases from left to right. Signaling events that are predicted according to literature are shown in grey.

## 4.4 DISCUSSION

In this study, we identified 16,253 phosphopeptides corresponding to 3,486 unique proteins by quantitative phosphoproteomics analysis of ccRCC tissues and NATs (**Fig. 47**). Furthermore, we quantified more than 9,000 phosphopeptides, which were more prone to dysregulation than their non-phosphorylated counterparts in the tumors (**Fig. 48**). Moreover, we found a high level of concordance between our study and the CPTAC phosphoproteome (**Fig. 64A**). Approximately 66% of the identified phosphosites of this study were shared with the CPTAC dataset, however, more than 3,700 phosphosites are unique to our approach suggesting undescribed observations of phospho signaling in ccRCC. Comparing the identified phosphopeptides of this study with phosphorylations in the PhosphoSitePlus database revealed that approximately 93% were already reported, which implicates that the majority of phosphorylation modifications in kidney cancer are not tissue-specific (**Fig. 64B**).

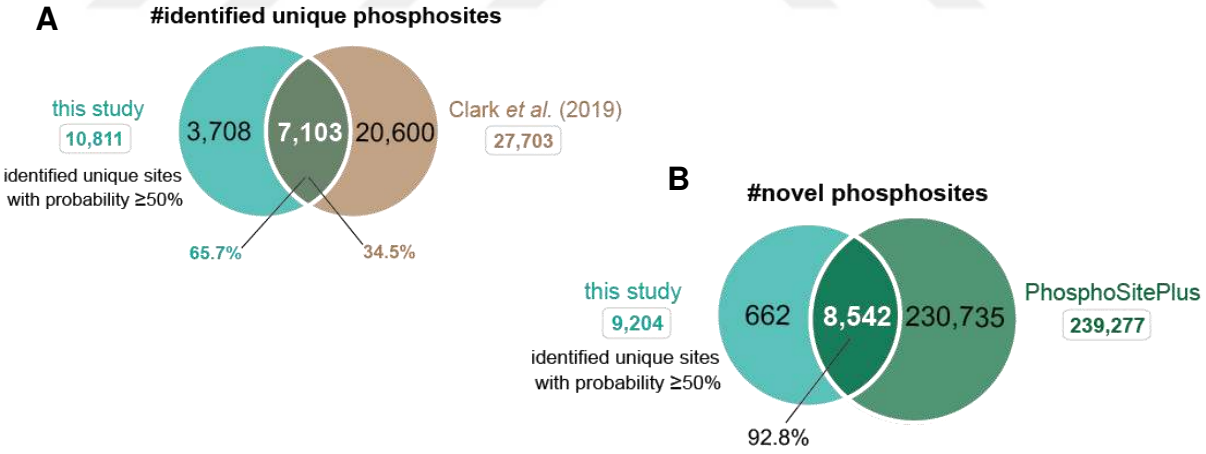


**Figure 64 | Comparison of phosphoproteome of this study with independent phospho datasets. (A)** Common identified phosphosites with CPTAC ccRCC phosphoproteome (Clark *et al*., 2019). **(B)** Comparison of identified phosphosites with ≥50% location probability with PhosphoSitePlus database. Peptides were centered to the phosphoresidue and ±7 amino acid sequence windows were compared.

Our in-depth proteomics analysis predicted all six members of the p21-activated kinase group (PAKs) to be highly activated in ccRCC tumors (**Fig. 57** and **Fig. 59**). PAKs are known to be switched on by the small GTPases Rac1 and Cdc42, and to be crucial for cytoskeletal dynamics, invasion, metastasis, cell death and proliferation (97). Previous phosphoproteomics studies have not described PAKs as relevant kinases in ccRCC tumorigenesis yet (14-16), however, our data strongly indicates that especially PAK2 is highly functional in the tumors as it is predicted to phosphorylate 14 functional peptides across 6 different substrates (**Fig. 61**). PAK2 shares similar substrate recognition preferences with PAK1, however, while the role of PAK1 is largely understood in different cancers, the function and targets of PAK2 remain to be elucidated (97, 98). It has been reported that PAKs play important roles in the cell cycle by interacting with different CDKs (98). PAK2 has previously been noted to bind to CDK12 and thereby to activate the MAPK pathway (98). Our study shows for CDK1, CDK2 and CDK5 high activity (**Fig. 53** and **Fig. 54**), which warrants further investigation to shed light into the synergistic functionality of PAKs and CDKs in RCC pathogenesis. Moreover, our analysis suggests that all PAKs can be inhibited by the tyrosine kinase inhibitor fostamatinib, which is currently not approved for RCC but is in clinical trial for other cancers such as ovarian cancer (www.clinicaltrials.gov). Also, it is worthwhile to check its efficacy along with inhibitors for other signaling components such as CDKs and MAPKs as combinatorial anti-cancer strategy for ccRCC treatment.

Interestingly, PAK2 is one of the few kinases that has also been experimentally detected by our proteomics approach. Furthermore, PAK2 harbors a significantly up-regulated phosphosite (S58) that is also annotated to be a mesenchymal phosphorylation in the tumors (87). Moreover, PAK2 itself is responsible for mesenchymal phosphorylations of its substrates VIM, EIF4B and EIF4G1 (**Fig. 59**). Vimentin (VIM) is an established mesenchymal marker known to be highly overexpressed in ccRCC tumors (84). Strikingly, it has been previously described that the levels of vimentin are four times higher in ccRCC, while the epithelial marker E-cadherin is expressed at half or only one third of the level compared to five other cancer types based on CPTAC phosphoproteome

data (84). Overall, our findings support the pronounced mesenchymal profile of renal cancers providing a possible explanation for the fact that around one third of ccRCC patients already display distant metastasis at diagnosis (84).

Another highly up-regulated phosphoprotein in the tumor tissues is the receptor tyrosine kinase EGFR. We have already described the potential involvement of the oncogenic EGFR-MAPK and EGFR-AKT1-mTOR axes in RCC pathobiology in our global proteome analysis of the same tissue samples (Chapter 2) (93). EGFR has been reported to be overexpressed in RCC tumors, however, only few cancer therapeutics target this membrane protein, and currently no clinically approved drug is in use for RCC treatment (14, 15). In contrast to other phosphoproteomics studies (14, 15), our data does not suggest that EGFR is a highly activated kinase rather than a substrate of different predicted kinases such as PDK1, MAPK1, MAPK3, CaMK2a and CaMK2g (**Fig. 65**).

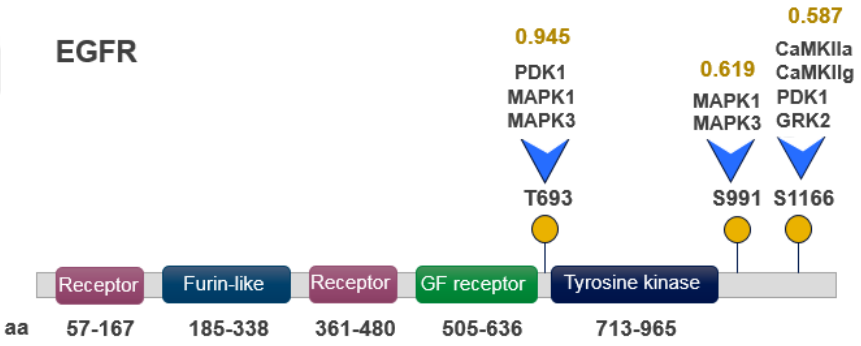


**Figure 65 | Protein domains of EGFR.** Significantly up-regulated functional phosphosites (functional scores in yellow) and predicted kinases targeting these sites are indicated.

Our quantitative approach identified three significantly up-regulated phosphosites for EGFR, namely T693, S991 and S1166, which are all three predicted to be functional modifications. EGFR is particularly known to be activated by tyrosine phosphorylations in its tyrosine kinase domain (713-965 aa) (79), however, we could not detect sites in this domain probably due to the intrinsic nature of our enrichment method to primarily catch

serine and threonine residues and due to the low abundance of tyrosine phosphorylations. Furthermore, EGFR is one of the central oncogenic signaling inducers in ccRCC tumors besides VEGFR and is probably responsible for many downstream effects such as proliferation, inflammation, survival and metastasis as previously suggested (15). Based on our findings, we propose that both the experimentally derived EGFR-mediated pathways, as well as the VEGFR-mediated signaling cascades are likely to synergistically govern malignant phospho signaling in ccRCC tumors (**Fig. 63**).

Overall, our comprehensive phosphoproteomics profiling of ccRCC tissues shed new light on the signaling cascades driving renal tumorigenesis. Future directions warrant the investigation of targeted therapies against PAK and CDK kinases as well as against the EGFR axis besides approved therapies for VEGFR such as sunitinib, pazopanib, axitinib and sorafenib, respectively. Also, the tyrosine kinase inhibitor fostamatinib is a promising drug alternative for the inactivation of the RAC1/Cdc42 axis and warrants further investigation in RCC treatment.

# BIBLIOGRAPHY

1. Capitanio U, Montorsi F. Renal cancer. The Lancet. 2016;387(10021):894-906.
2. Craven RA, Vasudev NS, Banks RE. Proteomics and the search for biomarkers for renal cancer. Clin Biochem. 2013;46(6):456-65.
3. Banumathy G, Cairns P. Signaling pathways in renal cell carcinoma. Cancer Biol Ther. 2010;10(7):658-64.
4. Raimondo F, Pitto M. Prognostic significance of proteomics and multi-omics studies in renal carcinoma. Expert Rev Proteomics. 2020;17(4):323-34.
5. Atrih A, Mudaliar MA, Zakikhani P, Lamont DJ, Huang JT, Bray SE, et al. Quantitative proteomics in resected renal cancer tissue for biomarker discovery and profiling. Br J Cancer. 2014;110(6):1622-33.
6. Raimondo F, Corbetta S, Savoia A, Chinello C, Cazzaniga M, Rocco F, et al. Comparative membrane proteomics: a technical advancement in the search of renal cell carcinoma biomarkers. Mol Biosyst. 2015;11(6):1708-16.
7. Neely BA, Wilkins CE, Marlow LA, Malyarenko D, Kim Y, Ignatchenko A, et al. Proteotranscriptomic Analysis Reveals Stage Specific Changes in the Molecular Landscape of Clear-Cell Renal Cell Carcinoma. PLoS One. 2016;11(4):e0154074.
8. Perroud B, Ishimaru T, Borowsky AD, Weiss RH. Grade-dependent proteomics characterization of kidney cancer. Mol Cell Proteomics. 2009;8(5):971-85.
9. Drendel V, Heckelmann B, Chen C-Y, Weisser J, Espadas G, Schell C, et al. Proteome profiling of clear cell renal cell carcinoma in von HippelLindau patients highlights upregulation of Xaa-Pro aminopeptidase-1, an anti-proliferative and anti-migratory exoprotease. Oncotarget. 2017;8(59):100066-78.
10. Zhang D, Xia X, Wang X, Zhang P, Lu W, Yu Y, et al. PGRMC1 Is a Novel Potential Tumor Biomarker of Human Renal Cell Carcinoma Based on Quantitative Proteomic and Integrative Biological Assessments. PLoS One. 2017;12(1):e0170453.
11. Masui O, White NM, DeSouza LV, Krakovska O, Matta A, Metias S, et al. Quantitative proteomic analysis in metastatic renal cell carcinoma reveals a unique set of proteins with potential prognostic significance. Mol Cell Proteomics. 2013;12(1):132-44.
12. Chen B, Chen W, Jin J, Wang X, Cao Y, He Y. Data Mining of Prognostic Microenvironment-Related Genes in Clear Cell Renal Cell Carcinoma: A Study with TCGA Database. Dis Markers. 2019;2019:8901649.
13. Casuscelli J, Vano YA, Fridman WH, Hsieh JJ. Molecular Classification of Renal Cell Carcinoma and Its Implication in Future Clinical Practice. Kidney

Cancer. 2017;1(1):3-13.
14. Haake SM, Li J, Bai Y, Kinose F, Fang B, Welsh EA, et al. Tyrosine Kinase Signaling in Clear Cell and Papillary Renal Cell Carcinoma Revealed by Mass Spectrometry-Based Phosphotyrosine Proteomics. Clin Cancer Res. 2016;22(22):5605-16.
15. Clark DJ, Dhanasekaran SM, Petralia F, Pan J, Song X, Hu Y, et al. Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. Cell. 2019;179(4):964-83 e31.
16. Peng X, Xu F, Liu S, Li S, Huang Q, Chang L, et al. Identification of Missing Proteins in the Phosphoproteome of Kidney Cancer. J Proteome Res. 2017;16(12):4364-73.
17. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
18. Beksac AT, Paulucci DJ, Blum KA, Yadav SS, Sfakianos JP, Badani KK. Heterogeneity in renal cell carcinoma. Urol Oncol. 2017;35(8):507-15.
19. Rini BI, Campbell SC, Escudier B. Renal cell carcinoma. The Lancet. 2009;373(9669):1119-32.
20. Clark DJ, Zhang H. Proteomic approaches for characterizing renal cell carcinoma. Clin Proteomics. 2020;17:28.
21. J. Cooper S, W. Tun H, M. Roper S, Kim Y. Current Status of Biomarker Discovery in Human clear Cell Renal Cell Carcinoma. Journal of Molecular Biomarkers & Diagnosis. 2012;01(S2).
22. ACS(AmericanCancerSociety). Kidney Cancer (Adult) - Renal Cell Carcinoma: American Cancer Society medical and editorial content team; 2016 [updated 01.08.2017.
23. Shaw G. The silent disease. Nature. 2016;537:S98-9.
24. Dizman N, Philip EJ, Pal SK. Genomic profiling in renal cell carcinoma. Nat Rev Nephrol. 2020;16(8):435-51.
25. Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, et al. Renal cell carcinoma. Nat Rev Dis Primers. 2017;3:17009.
26. Gleadle JM. Review article: How cells sense oxygen: lessons from and for the kidney. Nephrology (Carlton). 2009;14(1):86-93.
27. Bax C, Lotesoriere BJ, Sironi S, Capelli L. Review and Comparison of Cancer Biomarker Trends in Urine as a Basis for New Diagnostic Pathways. Cancers (Basel). 2019;11(9).
28. Stewart GD, O'Mahony FC, Powles T, Riddick AC, Harrison DJ, Faratian D. What can molecular pathology contribute to the management of renal cell carcinoma? Nat Rev Urol. 2011;8(5):255-65.
29. Borras E, Sabido E. What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. Proteomics. 2017;17(17-18).
30. Teng PN, Hood BL, Sun M, Dhir R, Conrads TP. Differential proteomic analysis of renal cell carcinoma tissue interstitial fluid. J Proteome Res. 2011;10(3):1333-42.
31. Macklin A, Khan S, Kislinger T. Recent advances in mass spectrometry

based clinical proteomics: applications to cancer research. Clin Proteomics. 2020;17:17.
32. Sandim V, Pereira Dde A, Kalume DE, Oliveira-Carvalho AL, Ornellas AA, Soares MR, et al. Proteomic analysis reveals differentially secreted proteins in the urine from patients with clear cell renal cell carcinoma. Urol Oncol. 2016;34(1):5 e11-25.
33. Santorelli L, Capitoli G, Chinello C, Piga I, Clerici F, Denti V, et al. In-Depth Mapping of the Urinary N-Glycoproteome: Distinct Signatures of ccRCC-related Progression. Cancers (Basel). 2020;12(1).
34. Wettersten HI, Weiss RH. Potential biofluid markers and treatment targets for renal cell carcinoma. Nat Rev Urol. 2013;10(6):336-44.
35. Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc. 2009;4(4):484-94.
36. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019;47(W1):W199-W205.
37. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 2019;47(D1):D419-D26.
38. Tuncbag N, Gosline SJ, Kedaigle A, Soltis AR, Gitter A, Fraenkel E. Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package. PLoS Comput Biol. 2016;12(4):e1004879.
39. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.
40. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics. 2009;25(22):3045-6.
41. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25(8):1091-3.
42. Cossu-Rocca P, Muroni MR, Sanges F, Sotgiu G, Asunis A, Tanca L, et al. EGFR kinase-dependent and kinase-independent roles in clear cell renal cell carcinoma. Am J Cancer Res. 2015;6(1):71-83.
43. Liu F, Shangli Z, Hu Z. CAV2 promotes the growth of renal cell carcinoma through the EGFR/PI3K/Akt pathway. Onco Targets Ther. 2018;11:6209-16.
44. Alonso-Gordoa T, Garcia-Bermejo ML, Grande E, Garrido P, Carrato A, Molina-Cerrillo J. Targeting Tyrosine kinases in Renal Cell Carcinoma: "New Bullets against Old Guys". Int J Mol Sci. 2019;20(8).
45. 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.

46. Lefort CT, Rossaint J, Moser M, Petrich BG, Zarbock A, Monkley SJ, et al. Distinct roles for talin-1 and kindlin-3 in LFA-1 extension and affinity regulation. Blood. 2012;119(18):4275-82.
47. Oksala N, Parssinen J, Seppala I, Klopp N, Illig T, Laaksonen R, et al. Kindlin 3 (FERMT3) is associated with unstable atherosclerotic plaques, anti-inflammatory type II macrophages and upregulation of beta-2 integrins in all major arterial beds. Atherosclerosis. 2015;242(1):145-54.
48. Zheng J, Wang L, Peng Z, Yang Y, Feng D, He J. Low level of PDZ domain containing 1 (PDZK1) predicts poor clinical outcome in patients with clear cell renal cell carcinoma. EBioMedicine. 2017;15:62-72.
49. Klatt MG, Kowalewski DJ, Schuster H, Di Marco M, Hennenlotter J, Stenzl A, et al. Carcinogenesis of renal cell carcinoma reflected in HLA ligands: A novel approach for synergistic peptide vaccination design. OncoImmunology. 2016;5(8).
50. Qi Y, Xu R. Roles of PLODs in Collagen Synthesis and Cancer Progression. Front Cell Dev Biol. 2018;6:66.
51. Kurozumi A, Kato M, Goto Y, Matsushita R, Nishikawa R, Okato A, et al. Regulation of the collagen cross-linking enzymes LOXL2 and PLOD2 by tumor-suppressive microRNA-26a/b in renal cell carcinoma. Int J Oncol. 2016;48(5):1837-46.
52. Chen Y, Guo H, Terajima M, Banerjee P, Liu X, Yu J, et al. Lysyl Hydroxylase 2 Is Secreted by Tumor Cells and Can Modify Collagen in the Extracellular Space. J Biol Chem. 2016;291(50):25799-808.
53. Djaafri I, Khayati F, Menashi S, Tost J, Podgorniak M-P, Sadoux A, et al. A novel tumor suppressor function of Kindlin-3 in solid cancer. Oncotarget. 2014;5(19):8970-85.
54. Rognoni E, Ruppert R, Fassler R. The kindlin family: functions, signaling properties and implications for human disease. J Cell Sci. 2016;129(1):17-27.
55. Murata Y, Saito Y, Kotani T, Matozaki T. CD47-signal regulatory protein alpha signaling system and its application to cancer immunotherapy. Cancer Sci. 2018;109(8):2349-57.
56. Yanagita T, Murata Y, Tanaka D, Motegi SI, Arai E, Daniwijaya EW, et al. Anti-SIRPalpha antibodies as a potential new tool for cancer immunotherapy. JCI Insight. 2017;2(1):e89140.
57. Tai IT, Tang MJ. SPARC in cancer biology: its role in cancer progression and potential for therapy. Drug Resist Updat. 2008;11(6):231-46.
58. Said N. Role of SPARC in Cancer; Friend or Foe. Annals of Carcinogenesis. 2016;1(1):11.
59. Sakai N, Baba M, Nagasima Y, Kato Y, Hirai K, Kondo K, et al. SPARC expression in primary human renal cell carcinoma: upregulation of SPARC in sarcomatoid renal carcinoma. Hum Pathol. 2001;32(10):1064-70.
60. Ricketts CJ, Linehan WM. Intratumoral heterogeneity in kidney cancer. Nat Genet. 2014;46(3):214-5.
61. Heidegger I, Pircher A, Pichler R. Targeting the Tumor Microenvironment in Renal Cell Cancer Biology and Therapy. Front Oncol.

2019;9:490.
62. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;15(2):81-94.
63. Ghatalia P, Rathmell WK. Systematic Review: ClearCode 34 - A Validated Prognostic Signature in Clear Cell Renal Cell Carcinoma (ccRCC). Kidney Cancer. 2018;2(1):23-9.
64. Hakimi AA, Pham CG, Hsieh JJ. A clear picture of renal cell carcinoma. Nat Genet. 2013;45(8):849-50.
65. Brannon AR, Reddy A, Seiler M, Arreola A, Moore DT, Pruthi RS, et al. Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns. Genes Cancer. 2010;1(2):152-63.
66. Brooks SA, Brannon AR, Parker JS, Fisher JC, Sen O, Kattan MW, et al. ClearCode34: A prognostic risk predictor for localized clear cell renal cell carcinoma. Eur Urol. 2014;66(1):77-84.
67. Cancer Genome Atlas Research N. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499(7456):43-9.
68. Seiler MH, C. C.; Szalma, S.; Bhanot, G. ConsensusCluster: A Software Tool for Unsupervised Cluster Discovery in Numerical Data. OMICS A Journal of Integrative Biology. 2010;14(1):109-13.
69. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074-D82.
70. Bruna FA, Romeo LR, Campo-Verde-Arbocco F, Contador D, Gómez S, Santiano F, et al. Human renal adipose tissue from normal and tumor kidney: its influence on renal cell carcinoma. Oncotarget. 2019;10(52):5454-67.
71. Chowdhury N, Drake CG. Kidney Cancer: An Overview of Current Therapeutic Approaches. Urol Clin North Am. 2020;47(4):419-31.
72. Zhang Z, Zhou L, Xie N, Nice EC, Zhang T, Cui Y, et al. Overcoming cancer therapeutic bottleneck by drug repurposing. Signal Transduct Target Ther. 2020;5(1):113.
73. Xia Y, Jia C, Xue Q, Jiang J, Xie Y, Wang R, et al. Antipsychotic Drug Trifluoperazine Suppresses Colorectal Cancer by Inducing G0/G1 Arrest and Apoptosis. Front Pharmacol. 2019;10:1029.
74. Pulkoski-Gross A, Li J, Zheng C, Li Y, Ouyang N, Rigas B, et al. Repurposing the antipsychotic trifluoperazine as an antimetastasis agent. Mol Pharmacol. 2015;87(3):501-12.
75. Stephenson KN, Croxen Rl Fau - El-Barbary A, El-Barbary A Fau - Fenstermacher JD, Fenstermacher Jd Fau - Haspel HC, Haspel HC. Inhibition of glucose transport and direct interactions with type 1 facilitative glucose transporter (GLUT-1) by etomidate, ketamine, and propofol: a comparison with barbiturates. (0006-2952 (Print)).
76. Liu J, Dong W, Wang T, Liu L, Zhan L, Shi Y, et al. Effects of etomidate and propofol on immune function in patients with lung adenocarcinoma. Am J Transl Res. 2016;8(12):5748-55.

77. Angeli AA-O, Carta FA-O, Nocentini AA-OX, Winum JA-O, Zalubovskis R, Akdemir AA-O, et al. Carbonic Anhydrase Inhibitors Targeting Metabolism and Tumor Microenvironment. LID - E412 [pii] LID - 10.3390/metabo10100412 [doi]. (2218-1989 (Print)).
78. Supuran CT. Carbonic anhydrase inhibitors. Bioorg Med Chem Lett. 2010;20(12):3467-74.
79. Ardito F, Giuliani M, Perrone D, Troiano G, Lo Muzio L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). Int J Mol Med. 2017;40(2):271-80.
80. Humphrey SJ, James DE, Mann M. Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. Trends Endocrinol Metab. 2015;26(12):676-87.
81. Harsha HC, Pandey A. Phosphoproteomics in cancer. Mol Oncol. 2010;4(6):482-95.
82. Solari FA, Dell'Aica M, Sickmann A, Zahedi RP. Why phosphoproteomics is still a challenge. Mol Biosyst. 2015;11(6):1487-93.
83. Schmidlin T, Debets DO, van Gelder C, Stecker KE, Rontogianni S, van den Eshof BL, et al. High-Throughput Assessment of Kinome-wide Activation States. Cell Syst. 2019;9(4):366-74 e5.
84. Deb B, Sengupta P, Sambath J, Kumar P. Bioinformatics Analysis of Global Proteomic and Phosphoproteomic Data Sets Revealed Activation of NEK2 and AURKA in Cancers. Biomolecules. 2020;10(2).
85. Zhou H, Ye M, Dong J, Corradini E, Cristobal A, Heck AJ, et al. Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. Nat Protoc. 2013;8(3):461-80.
86. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer. 2018;18(11):696-705.
87. Uretmen Kagiali ZC, Sanal E, Karayel O, Polat AN, Saatci O, Ersan PG, et al. Systems-level Analysis Reveals Multiple Modulators of Epithelial-mesenchymal Transition and Identifies DNAJB4 and CD81 as Novel Metastasis Inducers in Breast Cancer. Mol Cell Proteomics. 2019;18(9):1756-71.
88. Ochoa D, Jarnuczak AF, Vieitez C, Gehre M, Soucheray M, Mateus A, et al. The functional landscape of the human phosphoproteome. Nat Biotechnol. 2020;38(3):365-73.
89. Horn H, Schoof EM, Kim J, Robin X, Miller ML, Diella F, et al. KinomeXplorer: an integrated platform for kinome biology studies. Nat Methods. 2014;11(6):603-4.
90. Eid S, Turk S, Volkamer A, Rippmann F, Fulle S. KinMap: a web-based tool for interactive navigation through human kinome data. BMC Bioinformatics. 2017;18(1).
91. Cheng A, Grant CE, Noble WS, Bailey TL. MoMo: discovery of statistically significant post-translational modification motifs. Bioinformatics. 2019;35(16):2774-82.

92. Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res Treat. 2010;123(3):725-31.
93. Senturk A, Sahin A, Armutlu A, Kiremit M, Acar Ö, Erdem S, et al. Quantitative proteomics identifies secreted diagnostic biomarkers as well as tumor-dependent prognostic targets for clear cell Renal Cell Carcinoma2021.
94. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 2015;43(Database issue):D512-20.
95. Sharma P, Alsharif S, Fallatah A, Chung BM. Intermediate Filaments as Effectors of Cancer Development and Metastasis: A Focus on Keratins, Vimentin, and Nestin. Cells. 2019;8(5).
96. Rust HL, Thompson PR. Kinase consensus sequences: a breeding ground for crosstalk. ACS Chem Biol. 2011;6(9):881-92.
97. Ye DZ, Field J. PAK signaling in cancer. Cell Logist. 2012;2(2):105-16.
98. Liu H, Liu K, Dong Z. The Role of p21-Activated Kinases in Cancer and Beyond: Where Are We Heading? Front Cell Dev Biol. 2021;9:641381.

# VITA

Aydanur Şentürk was born in Turkey on 02.02.1989. She graduated from Wirtschaftsgymnasium Böblingen High School in Böblingen, Germany, in 2008. She received her Bachelor's degree in 2012 and her Master's degree in 2016, respectively, both in the Biotechnology program at University of Hohenheim in Stuttgart, Germany. From February 2016 to July 2021 she aspired a PhD in Molecular Biology and Genetics in the Cell Biology and Proteomics lab under the supervision of Dr. Nurhan Özlü, and worked during that time as a teaching assistant at Koç University, Istanbul, Turkey. Her doctoral research focused on the "Quantitative proteomics analysis of clear cell Renal Cell Carcinoma for the identification of diagnostic and prognostic biomarker panels".

## Publications

**Senturk, A.**, Sahin, A.T., Armutlu, A., Kiremit, M.C., Acar, O., Erdem, S., Bagbudar, S., Esen, T., Tuncbag, N. and Ozlu, N. (2021). Quantitative proteomics identifies secreted diagnostic biomarkers as well as tumor-dependent prognostic targets for clear cell Renal Cell Carcinoma. *Molecular Cancer Research* (accepted for publication)
DOI: 10.1158/1541-7786.MCR-21-0004

Üretmen Kagıalı, Z.C., **Şentürk, A.**, Özkan Küçük, N.E., Qureshi, M.H. and Özlü, N. (2017). Proteomics in Cell Division. *Proteomics* 17(20)
DOI: 10.1002/pmic.201600100

# APPENDIX

## Appendix A

**Table A 1 | List of patients in discovery cohort.**

| Patient number | Age | Gender | ISUP grade | Stage | Nephrectomy | Follow-up status | Patient cluster |
|---|---|---|---|---|---|---|---|
| **1** | 57 | Male | 2 | pT1a | Partial | Tumor free | **1** |
| **2** | 65 | Male | 3 | pT3a | Radical | Progression | **2** |
| **3** | 53 | Male | 3 | pT3a | Radical | Progression | **1** |
| **4** | 47 | Female | 2 | pT1a | Partial | Tumor free | **1** |
| **5** | 48 | Male | 3 | pT1b | Partial | Dead | **1** |
| **6** | 53 | Male | 4 | pT3a | Radical | Tumor free | **2** |
| **7** | 66 | Female | 2 | pT1b | Partial | Tumor free | **2** |
| **8** | 50 | Female | 3 | pT3a | Radical | Tumor free | **1** |
| **9** | 52 | Male | 2 | pT1b | Partial | Tumor free | **2** |
| **10** | 56 | Female | 4 | pT3a | Radical | Progression | **2** |
| **11** | 64 | Male | 4 | pT3a | Radical | Tumor free | **2** |
| **12** | 53 | Male | 4 | pT3a | Radical | Progression | **2** |
| **13** | 74 | Female | 4 | pT3a | Radical | Progression | **2** |

**Table A 2 | List of patients in validation cohort and urine cohort.**

| Patient number | Age | Gender | ISUP grade | Stage | Nephrectomy | Follow-up status | Cohort |
|---|---|---|---|---|---|---|---|
| **14** | 42 | Male | 3 | pT2a | Partial | Tumor free | Validation |
| **15** | 40 | Male | 2 | pT1a | Partial | Tumor free | Validation |
| **16** | 53 | Male | 2 | pT1b | Partial | Tumor free | Validation |
| **17** | 42 | Male | 2 | PT1a | Partial | Tumor free | Validation |
| **18** | 35 | Male | 2 | pT1a | Partial | Tumor free | Validation |
| **20** | 50 | Male | 3 | pT3a | Radical | Tumor free | Validation |
| **24** | 54 | Male | 3 | pT1a | Partial | Tumor free | Validation |
| **26** | 72 | Female | 2 | pT1a | Partial | Tumor free | Validation |
| **27** | 54 | Male | 3 | pT1a | Partial | Tumor free | Validation |
| **32** | 67 | Male | 2 | pT3a | Radical | Tumor free | Validation |
| **33** | 40 | Male | 4 | pT2a | Radical | Dead | Validation |
| **20** | 60 | Male | 3 | pT1a | Partial | Tumor free | Urine |
| **22** | 57 | Female | 1 | pT1a | Partial | Tumor free | Urine |
| **9** | 68 | Male | 2 | pT1a | Partial | Tumor free | Urine |
| **14** | 67 | Male | 3 | pT1a | Partial | Tumor free | Urine |
| **18** | 57 | Male | 2 | pT1a | Partial | Tumor free | Urine |
| **12** | 42 | Female | 3 | pT2a | Partial | Tumor free | Urine |

## Appendix B

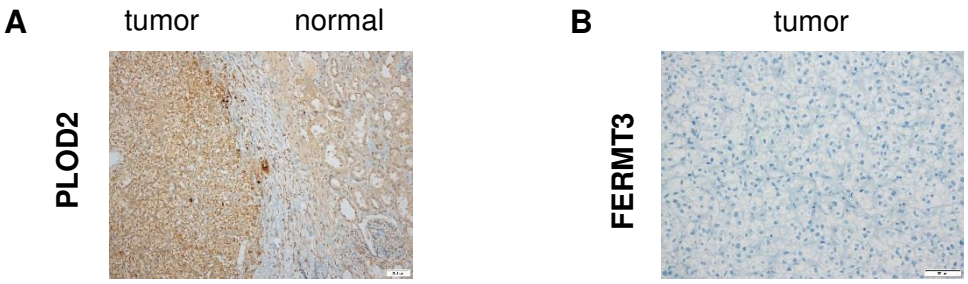


**Figure A 1 | Immunohistochemical staining of (A) PLOD2 and (B) FERMT3.** PLOD2 staining shows transitional region between tumor and adjacent normal tissue (experiments were performed by Dr. Ayşe Armutlu, Koç University School of Medicine).

## Appendix C

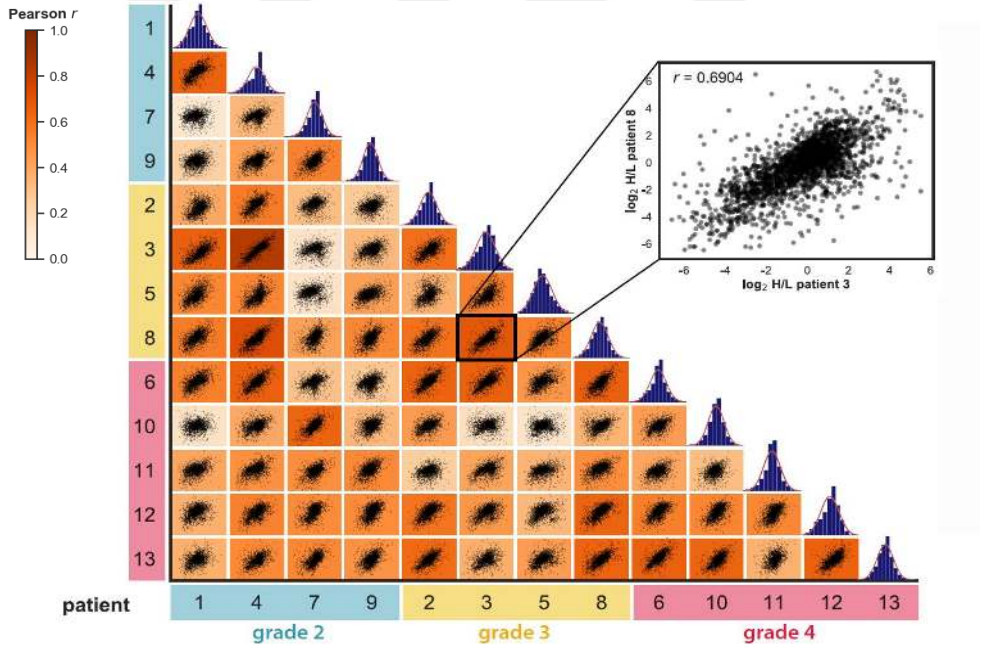


**Figure A 2 | Biological variance across patients of discovery cohort.** Pearson correlations of quantified common proteins are shown for patients of different grades. Miniature histograms depict distribution of fold change values with indication of theoretical normal distribution as red line.

## Appendix D

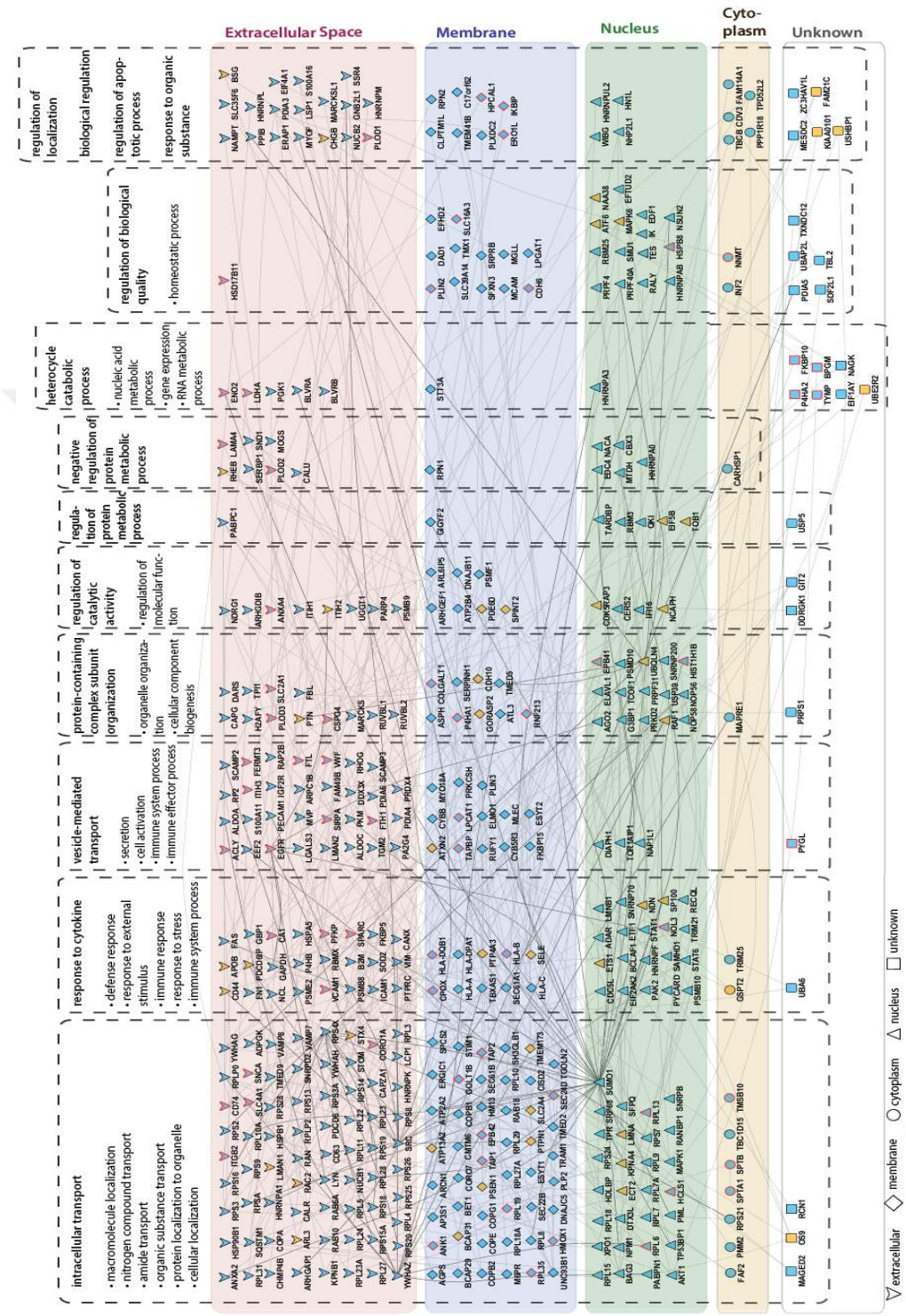


**Figure A 3 | Complete protein-protein interaction network of significantly up-regulated proteins in ccRCC tumors.**

## Appendix E

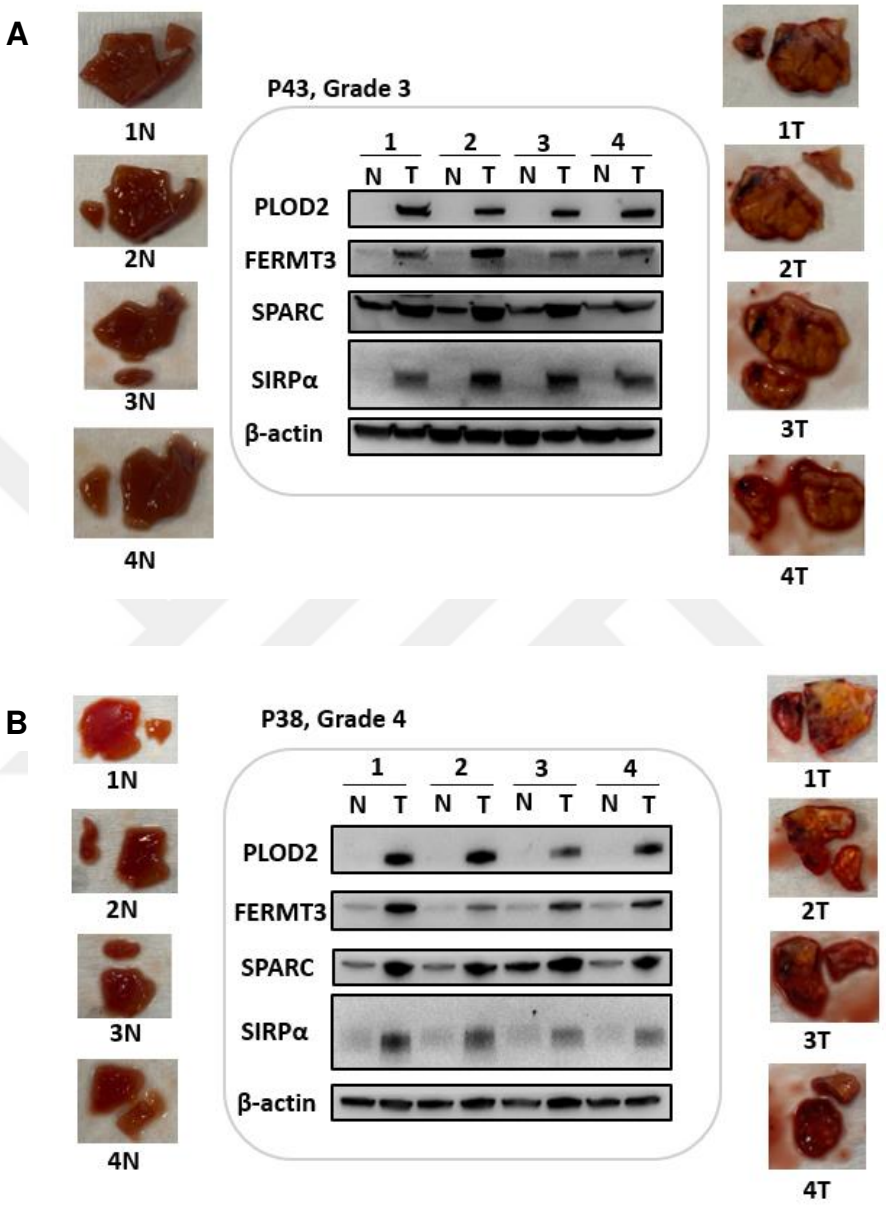


**Figure A 4 | Impact of intratumor heterogeneity on biomarker expression.** Comparable expression of biomarkers in random sub regions of tissue samples with morphological diversity from two selected patients (**A** and **B**) showing independence of biomarker levels from local distortions (experiments were performed by Ayşe Tuğçe Şahin, Koç University)

## Appendix F

**Table A 3 | Summary of univariate Cox Regression analysis of overall survival using clinical data of TCGA-KIRC cohort (n = 534).** Regression model was generated with the CoxPHFitter function of the Python lifelines library. "Not reported" entries for covariates were filtered out. CI: confidence interval.

| Parameter | Category setting | Hazard ratio | Lower 95% CI | Upper 95% CI | *p*-value | Cohort size |
|---|---|---|---|---|---|---|
| **Diagnosis age** | >60 vs <60 | 1.857 | 1.351 | 2.551 | 1.36E-04 | 531 |
| **Prior cancer diagnosis** | Yes vs No | 0.834 | 0.542 | 1.285 | 0.411 | 531 |
| **Disease stage** | III+IV vs I+II | 3.863 | 2.810 | 5.311 | 8.69E-17 | 528 |
| **Histologic grade** | G3+G4 vs G1+G2 | 2.700 | 1.911 | 3.816 | 1.81E-08 | 523 |
| **Gender** | Male vs Female | 0.939 | 0.688 | 1.281 | 0.691 | 531 |
| **Cluster** | 1 vs 2 | 2.756 | 2.029 | 3.741 | 8.37E-11 | 512 |
| **Fraction genome altered** | >50% vs <50% | 1.030 | 0.526 | 2.017 | 0.931 | 523 |
| **Race** | Asian vs Black vs White | 0.813 | 0.484 | 1.362 | 0.431 | 524 |
| **Distant metastasis stage** | M1 vs M0+MX | 4.483 | 3.287 | 6.114 | 2.6E-21 | 529 |
| **Lymph node stage** | N1 vs N0+NX | 3.909 | 2.116 | 7.223 | 1.35E-05 | 531 |