Histo-molecular differentiation of renal cancer subtypes by imaging mass spectrometry and rapid proteome profiling

Uwe Möginger[1,§], Niels Marcussen[2], Ole N. Jensen[1*]

[1]Department of Biochemistry & Molecular Biology and VILLUM Center for Bioanalytical Sciences, University of Southern Denmark, DK-5230 Odense M, Denmark.

[2]Institute for Pathology, Odense University Hospital, DK-5000 Odense C, Denmark.

**Corresponding author:*

Professor Ole N. Jensen, PhD, email: jenseno@bmb.sdu.dk

Department of Biochemistry & Molecular Biology and VILLUM Center for Bioanalytical Sciences, University of Southern Denmark, DK-5230 Odense M, Denmark.

Present address:

[§] Global Research Technologies, Novo Nordisk A/S Novo Nordisk Park 2760 Måløv, Denmark

Summary

Background: *Pathology assessment and differentiation of renal cancer types is challenging due to overlapping histological features of benign and malignant tumors, necessitating high-level expertise. Mass spectrometry (MS) is an emerging tool for tumor classification of clinical tissue sections by spatial histo-molecular imaging or quantitative microproteomics profiling.*

Results*: We applied MALDI MS imaging (MSI) and LC-MS/MS-based microproteomics technologies to analyze and classify renal oncocytoma (RO, n=11), clear cell renal cell carcinoma (ccRCC, n=12) and chromophobe renal cell carcinoma (CRCC, n=5). Both methods distinguished ccRCC, RO and CCRC with high accuracy in cross-validation experiments (MSI: 93-95%, LC-MS/MS: 100%).*

Significance: *This integrated strategy combining MSI and rapid proteome profiling by LC-MS/MS reveals molecular features of tumor sections and enables cancer subtype classification. Mass spectrometry is a promising complementary approach to current pathological technologies for precise digitized diagnosis of diseases.*

Keywords:

Renal Cell Cancer, Clear Cell Renal Cell Carcinoma, Renal Oncocytoma, Chomophobe Renal Cell Carcinoma, MALDI Mass Spectrometry Imaging, Microproteomics, Proteomics, Classification, Liquid Chromatography, Mass Spectrometry

Introduction

Kidney cancer (renal cell carcinoma, RCC) accounts for 2.2% of all diagnosed cancers and it is the 13th most common cause of cancer deaths worldwide[1]. Clear cell renal carcinoma constitutes 70% of all kidney cancers[2] and exhibit the highest rate of metastasis among renal carcinomas. Two other common but less aggressive subtypes of renal carcinoma are chromophobe renal carcinoma (CRCC) and the essentially benign oncocytoma (OC) which account for 5% and 3-7 % of all cases, respectively[3, 4]. The ability to distinguish between the malignant cancer types ccRCC and CRCC and the benign RO is crucial for a patient in terms of prognosis, progression and intervention strategies as severe as total nephrectomy. Histopathological kidney cancer diagnostics faces many challenges in daily routine. Typically, test panels consisting of a combination of different chemical and immuno-histochemical staining methods are used to systematically obtain a diagnosis[5]. Overlapping histological features can make it difficult to differentiate tumor types. Analysis, interpretation and diagnosis/prognosis greatly rely on visual inspection and the experience of the involved clinical pathologists. Complementary techniques such as MRI and electron microscopy involve costly instrumentation. Moreover, specific antibodies for staining can be expensive or unavailable for certain molecular targets. Mass spectrometry is emerging as a promising new tool in translational research, from molecular imaging of tissue sections to deep protein profiling of tissue samples[6]. The digital data readout provided by high mass accuracy mass spectrometry and feasibility of molecular quantification makes it a very attractive technology in translational research for investigating human diseases and for diagnostics and prognostics purposes in the clinic. Improvements in mass spectrometry instrument performance and

computational analysis paved the way for applications in clinical microbiology[7] and clinical genetics analysis[8]. The fact that mass spectrometry can be applied to a variety of different bio-molecules such as peptides, lipids, nucleic acid makes it extremely versatile and expands the translational and diagnostic possibilities greatly[8-11].

Molecular imaging of tissue sections by MALDI mass spectrometry was introduced more than 20 years ago[12, 13] and it has been applied in translational research and clinical applications, to study injuries, diseases, or distinguish between different cancer types such as Pancreatic Ductal Adenocarcinoma or Epithelial Ovarian Cancer Histotypes[14-18].

Mass spectrometry-based proteomics relies on advanced LC-ESI-MS/MS technology, where peptide mixtures are separated by liquid chromatography (LC) prior analysis by electrospray ionization tandem mass spectrometry (ESI MS/MS) and protein identification by protein database searching[19, 20]. Current LC-MS/MS strategies enable comprehensive quantitative protein profiling from tissues and body fluids[21, 22]. While having been used to identify potential biomarkers or new candidate cancer targets and molecular signaling networks the relatively long LC gradients (hours) and extensive sample preparation protocols make it difficult to apply in a routine clinical setting. Modern mass spectrometers are steadily increasing in sensitivity and scanning speed[23]. In addition, improved chromatographic systems that enable rapid solid phase extraction integrated with reproducible separations are emerging[24-27], enabling fast (minutes) and sensitive (nanogram) analysis of complex biological samples.

We hypothesized that the combination of imaging MALDI MS, *in situ* protein digestion and LC-MS/MS applied to detailed characterization of 5 µm cancer FFPE tissue sections will

provide both spatial molecular maps and sufficiently deep proteome profiles to characterize and classify tumor subtypes. We investigated this by testing a series of malignant and benign renal carcinomas, including clear cell renal cell carcinoma (ccRCC), chromophobe renal cell carcinoma (CRCC) and renal oncocytoma (RO). We obtained molecular images at a resolution of 150µm x150µm that sufficed to spatially resolve features to distinguish tumor subtypes. Miniaturized sample preparation by *in situ* protein digestion was used to recover peptides from distinct areas of the tumor sections for rapid proteome profiling by LC-MS/MS.

Material and Methods:

Materials

Xylene (analytical grade), ammonium bicarbonate, Sodium citrate, trifluor-acetic acid (TFA), formic acid (FA), acetic acid (AcOH), acetonitrile (ACN), methanol and αCyano-4-hydroxycinnamic acid (CHCA) were purchased from Sigma (). Polyimide coated fused silica capillary (75 µm ID) was from PostNova, C_{18} Reprosil Pur reversed phase material was from Dr. Maisch (Ammerbuch-Entringe, Germany), recombinant Trypsin was purchased from Promega (WI. USA), Indium-tin-oxide (ITO) glass slides were purchased from Bruker (Bremen, Germany), water was Milli-Q filtered.

Formalin fixed paraffin embedded samples:

Patient samples were collected at Odense University Hospital, Denmark. All samples were obtained upon patient's consent. Formalin fixed paraffin embedded (FFPE) tissues from 11 OC patients, 12 cccRCC patients and 5 CRCC patients were used for LC-MSMS analysis (for CCRC due to the lower number of patients 2 subsequent slides were used from 2 patients

adding up to a total of 7 sections). Out of the patient cohort 9 OC, 9 ccRCC and 5 CRCC were used for imaging mass spectrometry analysis.

Tissue preparation:

Preparation of formalin fixed paraffin embedded samples

FFPE blocks were cut into 5 µm thick sections and mounted onto indium tin oxide (ITO) covered glass slides (for MSI) or regular microscopy glass slides (for LC-MS/MS). Before deparaffination slides were left on a heated block at 65° C for 1 hour to improve adhesion.

Deparaffination

FFPE section slides were incubated in Xylene for an initial 10 min. and then another 5 min. using fresh solution each time. Slides were shortly dipped into 96% EtOH before they were washed for 2 min in a mixture of chloroform/Ethanol/AcOH (3:6:1; v:v:v). The slides were then washed in 96% EtOH, 70% EtOH, 50% EtOH and Water for 30 sec. each.

Antigen retrieval

Tissue slides were heated in 10mM citric acid buffer pH 6 for 10 min in a microwave oven at 400 Watt (just below the boiling point) before left for further 60 min incubation at 98°C. Slides were cooled down to room temperature and incubated for 5 minutes in 25 mM ammonium bicarbonate (ABC) buffer. Slides were allowed to dry before application of trypsin protease.

Tryptic digest

For MALDI imaging:

20µg of Trypsin (Promega) was used per slide and was dissolved at a concentration of 100ng/µl in 25mM ABC /10% ACN before being deposited on the tissue using the iMatrixSpray device equipped with a heating bed (Tardo Gmbh, Subingen, Switzerland [28]) using the following settings: sprayer height = 70mm, speed = 70mm/s, density = 1µL/cm^2, line distance= 1 mm , gas pressure= 2.5 bar, heat bed temperature= 25°C . After trypsin deposition the slides were incubated in a humid chamber containing 10mM ABC/ 50% MeOH at 37°C over night.

For *on-tissue* digest intended for LC-MS/MS proteome profiling:

Droplets of 2µl Trypsin solution (50ng/µL in 25mM ABC /10%ACN, 0.02%SDS) were deposited using a gel loading pipet tip. Droplets were placed on 3-4 different tumor areas of each FFPE tissue section. The droplets were shortly allowed to dry in order to prevent spreading across the tissue. Slides were transferred to a humid chamber (10mM ABC /50% MeOH) for overnight digestion at 37°C. After digest the digestion spots were extracted twice with 2µL of 0.1% FA and twice with 1.5µL of 30%ACN.

Matrix application

Matrix solutions were freshly prepared from recrystallized α-cyano-4-hydroxycinnamic acid (CHCA) matrix (10mg/mL in 50% Acetonitrile 1% TFA). Matrix was sprayed using the iMatrixSpray (Tardo, Switzerland). Temperature of the heat bed was set at 25°C. The sprayer distance was set to 70mm. Spray speed was set to 100 mm/s. Matrix was sprayed in 3

rounds: 8 cycles with a flowrate of 0.5µl/cm^2 line distance of 1mm, 8 cycles of 1µl/cm^2 line distance of 1mm, 8 cycles of 1µl/cm^2 and a line distance of 2mm.

MALDI MS Imaging data acquisition

Optical images of the tissue were obtained before matrix application using a flatbed scanner (Epson) at resolutions of 2400dpi. The imaging data was acquired via FlexImaging software (Bruker, Daltonics, Bremen, version 3.1) with 500 shots/ pixel on a Ultraflextreme MALDI-TOF/TOF MS (Bruker Daltonics, Bremen) equipped with a SmartBeam laser (Nd:YAG 355 nm). External mass calibration was performed with a tryptic digest of bovine serum albumin (Sigma). Spatial resolution was set to 150µm in x- and y-direction. Mass spectra were acquired in positive ion reflector mode in the range *m/z* 600-3500.

LC-MS/MS analysis

LC-MS/MS data was acquired by an Orbitrap Q-Exactive HF-X (Thermo, Bremen) coupled to an Ultimate 3000 capillary flow LC-system. Peptide samples were loaded at 150µl/min (2% ACN, 0.05% TFA) for 30 sec onto a 5µm, 0.3 x 5 mm, Acclaim PepMap trapping cartridge (Thermo Scientific). Samples were then eluted onto a pulled emitter analytical column (75µm ID, 15cm). The analytical column was “flash-packed” [29] with C_{18} Reprosil Pur resin (3µm) and connected by Nanoviper fittings and a reducing metal union (Valco, Houston, TX). The flowrate of the 15 min gradient was 1.2 µL/min with solvent A: 0.1% formic acid (FA) and solvent B: 0.1% FA in 80% ACN. Gradient conditions for solvent B were as followed: 8% to 25% in 10 min, 25% to 45% in 1.7 min. The trapping cartridge and the analytical column were

washed for 1 min at 99%B before returning to initial conditions. The column was equilibrated for 2 min.

Data Processing of MALDI MS imaging data

The data was baseline subtracted, TIC normalized and statistically recalibrated and then exported into imzML format[30] using the export function of FlexImaging software (Bruker). The exported mass range was 600-3000 m/z with a binning size of 9600 data points. The imzML files were imported into the R environment (version: 3.4.1) and further processed and analyzed using the R MSI package: Cardinal (version: 2.0.3)[31]. In order to extract pixels of tumor tissue each sample was preprocessed as follows: peaklist was generated by peak picking in every 10th spectrum and subsequent peak alignment. The whole data was then resampled using the “height” option and the previous created peaklist as spectrum reference. PCA scores were plotted using car-package (version 3.0.6). Samples were clustered using spatial shrunken centroid clustering[32]. Subsequently, pixel coordinates of cluster containing tumor areas (HE-stain comparison, **Supplementary Figure S1**) were extracted and manually trimmed, if necessary, so that result files predominantly contained data from tumor areas. The obtained coordinates were then used to extract the corresponding pixel from the unprocessed imzML file. Each tumor type was assigned with a diagnosis factor (ccRCC, RO or CRCC), which was later used as y-argument in the cross-validation. All extracted imaging acquisition files were further restricted to a mass range of m/z 700-2500. Data was resampled with step size 0.25 Da and combined into one file for further processing. Classification and cross

validation were performed using partial least square discriminant analysis (PLS-DA)[33]. PLS components were tested for optimum with 30 components for ccRCC/RO comparison and 12 components for RO/CRCC comparison (**Supplementary Figure S4**).

LC-MS/MS data processing

The MaxQuant[34] software package (version 1.5.7.0) was used for protein identification and label-free protein quantitation. LC-MS/MS data was searched against the Swissprot human proteome database, using standard settings and “match between runs” enabled.

Data filtering, processing and statistical analysis of the MaxQuant output files was performed using the Perseus[35] framework (version 1.6.1.3). Data was filtered excluding the following hits: only identified by site, contaminants and reversed. The log-transformed data was filtered for proteins present in at least 70% of all experiments. Significance filtering was based on ANOVA testing, using FDR threshold of 0.01 with Benjamini Hochberg correction. In order to perform PCA analysis and classification missing values were imputed by normal distribution. Data shown in heatmap was Z-score normalized. Perseus output tables were transferred into ClustVis [36] for visualization of hierarchical clustering and principle compound analysis (PCA). Gene Ontology and functional analysis was performed via String DB (version 11.0.0)[37] and Panther DB (version 14.1)[38]. For Panther DB analysis background genome was both human genome and all identified proteins from the experiment (supplementary material 3-7). Feature optimization cross validation type was “n-fold” with n = 6. Kernel was either linear or RGF. All other settings were left on their default value.

Results:

In this study we investigated the utility of mass spectrometry-based methods for histo-molecular profiling applications in clinical renal cancer pathology. We analyzed thin tissue/tumor sections from three different renal cancer types (ccRCC, RO, CRCC) by imaging MALDI MS and by an optimized rapid LC-MS/MS workflow adjusted to suit the demands for clinical settings.

IMAGING MALDI mass spectrometry

All samples were prepared as 5 µm thin FFPE tissue/tumor sections. Nine ccRCC samples were prepared and compared with nine RO samples, and five CRCC samples with five RO samples. From a pathology viewpoint RO and CRCC are more difficult to distinguish than ccRCC and CRCC. As the sample holder for the imaging experiments can only hold 2 slides at a time, two conditions (ccRCC vs. RO and RO vs. CRCC) were compared respectively. The entire FFPE tissue section was analyzed by histo-molecular imaging MALDI MS and data subsequently processed by unsupervised clustering (spatial shrunken centroid clustering [32]). The clustering results **(Figure 1A and 1B)** illustrate the heterogeneity of the tissue sections coming from various tissue types such as stroma, fibrotic fatty or healthy tissue. Furthermore, when comparing the tumor area of the HE-stain/microscopy with the results from the mass spectrometry based clustering, spectral differences even within the tumor tissue itself can be observed (**Figure 1A and 1C**).

Guided by the unsupervised clustering outcome and the corresponding image obtained by HE-staining, pixels from non-relevant surrounding tissue were discarded and only pixel clusters containing actual tumor tissue were used for subsequent comparative analyses.

Variance and similarities within the image sample set were estimated by principal component analysis (PCA). We compared scores from the imaging MALDI MS data from ccRCC and RO **(Figure 2A)** as well as from RO with CRCC over the first 3 components **(Figure 2B)**.

ccRCC and RO are well separated by PCA by three principal components. Data from ccRCC showed a wide spread in practically all principle component even splitting into 2 sub-populations, suggesting significant ccRCC tumor heterogeneity. In contrast, the data from RO samples is clustering in a much tighter manner and is well separated from ccRCC samples. The RO and CRCC datasets cluster closer together with some overlap. Even though, RO and CCRC do not separate as well as RO and ccRCC, there are clear histo-molecular differences between these cancer types.

The wide spread of ccRCC data points in PCA, as compared to the RO and CRCC subtypes, suggests a greater heterogeneity among the ccRCC patient samples (also observed by LC-MS/MS, see below). The fact that there are some overlapping data points for RO/ccRCC and RO/CRCC datasets indicated histo-molecular spectral similarity in parts of the patient tumor tissues.

Next, we assessed the ability and performance of the MSI data to distinguish and classify renal cancer subtypes. We created a classifier based on partial least squares discriminant analysis (PLS-DA) and subsequently tested the classifier by cross-validation using the whole

sample set. In this approach a classifier is trained with imaging data from all samples, except for the one sample being tested for the given condition. This was repeated for all tumor samples in order to test the complete dataset. The optimized PLS-DA model resulted in an accuracy of cancer subtype prediction of 94% (based on all pixels of the sample test set). **Figure 3** shows the PLS-DA prediction results plotted onto the pixel data from tissue/tumor sections. The prediction scores and the resulting color-coding for each cancer type condition are overlaid in the image. The predominant color appearances indicate the respective prediction (ccRCC: red, RO: blue, CRCC: yellow). All 9 RO sample were predominantly blue and 7 out of 9 ccRCC sample appeared predominantly red (**Figure3 a-i**) thus reflecting the true cancer type of the patient sample. Two ccRCC samples were falsely predicted as RO **(Figure 3b and 3g).** Overall this resulted in a 89% correct patient prediction rate (16 out of 18). Sporadic pale or mixed-color appearance can be observed in some of the images. This is a result of the prediction score color for each cancer type being overlaid on the image. The two false predicted ccRCC samples as well as the slightly blue appearance of some ccRCC samples might indicate a small tendency of mis-interpretation of the spectra as RO.

PLS-DA classification of 5 RO and 5 CRCC tumor samples was achieved with an accuracy of 95% based on all pixels of the test set. Upon visualization of the cross-validated prediction rate all RO samples predominantly appeared blue (predicted RO) and all CRCC samples appear predominantly yellow (predicted CRCC) (**Figure 3 j-n)** reflecting the true cancer type of the sample.

We then investigated the relative importance of individual histo-molecular features. PLS coefficients were plotted as a function of m/z values to depict the relative importance of the

individual m/z features for the predictive PLS-DA model **(Figure 4)**. The six most influential features for each comparison are listed in Figure 4. The imaging mass spectrometry signal at m/z 944 was identified by direct MALDI tandem mass spectrometry sequencing as the peptide AGLQFPVGR derived from Histone H2A (**supplementary Figure S3**). It was not possible to obtain sequence information from other signals when using MALDI MS/MS.

In summary, imaging mass spectrometry provides histo-molecular tumor profiles that can distinguish renal cancer subtypes with high accuracy. However, it was challenging to identify the molecular origin of the most prominent m/z features that distinguished the renal cancer subtypes in the PLS-DA model.

LC-MS/MS based rapid proteome profiling of tumor sections.

Imaging MALDI MS provides spatial resolution that is helpful to address molecular heterogeneity in tissue sections. However, imaging MALDI MS lacks analytical depth due to the limited dynamic range of MALDI MS and the poor performance of MS/MS for protein identification directly from tissue sections. Deeper insight into the tissue and tumor histo-molecular profiles and their variance will provide more diagnostic features. We therefore adapted and optimized a microproteomics approach, combining *in situ* protein sample preparation with fast proteome profiling by solid-phase extraction LC-MS/MS using a novel platform. First a miniaturized *in situ* sample preparation method was applied where a small droplet of trypsin solution is placed directly onto the tumor area of interest within a thin tissue section. After overnight incubation the digested protein extract from the tumor area is

subsequently recovered and analyzed by mass spectrometry [39]. We reduced the LC-MS/MS analysis time from 90 minutes to 15 min by using short LC gradients and rapid MS/MS functions, allowing for a sample throughput of up to 80 samples per day. A total of 125 *in situ* extracted areas from renal tumor sections were analyzed. Two to four *in situ* extracts were taken from each renal tumor sample (11 RO patients: 47 extraction spots; 12 ccRCC patients: 49 extraction spots; 5 CRCC patients: 29 extraction spots). Fast LC-MS/MS based microproteomics analysis of all 125 *in situ* digested tumor areas resulted in a total of 2124 identified human proteins. We filtered the data for proteins that were present in at least 70 % of all samples thereby reducing the protein number to 412 proteins. Comparative data analysis was performed for proteins that were significantly altered (FDR=0.01) in any of the renal cancer subtypes resulting in a list of 346 differentially regulated proteins. We then used unsupervised hierarchical clustering and PCA to identify similarities and differences between the tumor samples. The x-axis dendrogram of the heatmap shows that the majority of the renal tumor samples grouped according to cancer subtype RO, ccRCC or CRCC **(Figure 5A)**. Several large clusters of “co-regulated” proteins are evident on the y-axis dendrogram and heatmap for the individual cancer subtypes. This clearly demonstrates that there are renal cancer subtype specific histo-molecular features and patterns in the microproteomics dataset.

The protein expression profiles of the three renal cancer subtypes are different based on the heatmap patterns. ccRCC is clearly very different from RO and CRCC (**Figure 5A**: Protein group 2 and 4). RO and CRCC display some differences but generally exhibit a more similar expression pattern (**Figure 5A**: Protein Group 2).

These differences and similarities were also revealed by PCA analysis of the microproteomics dataset. RO and CRCC separate clearly from ccRCC **(Figure 5B)**. RO and CRCC datapoints are located close together, indicating that differences between the RO and CRCC cancer subtypes are less dominant. When considering principal components exhibiting less variance (PC3 and PC4), separation of RO and CRCC sample data can be observed (**Figure 5B**).

We observed eight CRCC samples that separated clearly form the other CRCC samples, both in hierarchical clustering analysis (**Figure 5A**) and PCA (**Figure 5B**). The protein expression profile of these 8 samples exhibited some similarities to both CRCC and ccRCC. Interestingly, this data originated from a tumor from a single patient. Further pathology analysis revealed that these samples were sarcomatoid renal cancer, originating from CRCC and, thus, indeed different from the other CRCC samples.

Protein differences in cancer subtypes

Hierarchical clustering revealed major differences in relative protein abundance between the three cancer types. (**Figure 5A**). We investigated the nature of these histo-molecular differences by examining the correlation of these proteins to cellular structures, functions, or biochemical processes. Protein groups that exhibited distinctive abundances for the respective cancer type (**Figure 5:** ccRCC: group 1 & 4, RO: group 2, CRCC: group 3) were searched for their involvement in protein interaction networks (supplementary material 8) as well as for their functional roles by using gene ontology (GO) enrichment (**Figure 6**, supplementary material 2-7).

RO and CRCC exhibited a set of upregulated proteins (**Figure 5A**, *protein group 2)* that were enriched for mitochondria associated proteins (GO:0005739), including various ATP synthase subunits. Enriched protein functions comprised *oxidative phosphorylation* (hsa00190), *citrate cycle* (hsa00020), and *fatty acid beta oxidation* (GO:0006635).

CRCC-specific regulated proteins (**Figure 5A**, *protein group 3*) included cytoplasmic proteins (GO:0044444), and proteins associated with *cytoplasmic vesicles* (GO:0031982) and *ribonucleoprotein complexes* (GO:1990904).

Subtype-specific protein groups in ccRCC (**Figure 5A**, *protein group 1, 4*) were functionally enriched for *complement activation* (GO:0006956), *regulation of blood coagulation* (GO:0030193) and *platelet degranulation* (GO:0002576). Functions of *protein group 4* were linked with *extra cellular matrix organization* (GO:0043062) and *cytoskeletal binding* (GO:0008092) including proteins *collagen* and *laminin*. We also found several proteins such as *glyceraldehyde-3-phosphate dehydrogenase* associated with the *glycolytic process* (GO:0006096).

These functionally important findings can be correlated to known biochemical and morphological features of each of the renal cancer subtypes. It is an established fact that the number of mitochondria is increased dramatically in RO and CRCC tumors (e.g. increased oxidative phosphorylation). It is also known that ccRCC contains a highly vascularized stroma (complement, coagulation, etc.) and exhibits a strong Warburg effect (glycolysis). Large intracellular vesicles are found in CRCC (cytoplamic proteins, vesicle proteins).

Classification

Unsupervised data analysis demonstrated the presence of renal cancer subtype specific differences in the tumor protein profiles. Next, we investigated the feasibility of tumor classification by using the microproteomics data to train a prediction algorithm. We implemented the tumor classification model by using a support vector machine (SVM) approach. We chose the k-fold cross validation strategy [40] ("n-fold" in Perseus). Here the data is randomly distributed in k groups. The model was then trained with data from k-1 groups and the prediction was applied to the samples in the remaining group. This was repeated k times. Low k-values tend to overestimate error rates. In our study 3-5 extraction spots (samples) were derived from the same patient so too high k-values could underestimate the true error rate. We therefore tested the prediction rate error over several k-values (**Figure 7 A**) applying Radial Basis Function (RBF) and linear kernel functions [41]. The tested error rates were in the range of 3.2% (4 wrong predictions) at the highest (k=3, linear kernel) and 0 % at the lowest. However, k=3 is a very low k-value (excluding more than 41 samples from the training set). We argue that the error rate is most likely overestimated in this case. For more commonly used k-values (k=5-10) the error rate was between 0.8% (1 wrong prediction) and 0% for RBF and 1.6% (2 wrong prediction) and 0% for linear kernel function. False predicted samples included samples from one OC patient predicted as ccRCC. Generally, RBF performed slightly better than linear kernel. Figure 7B exemplifies the outcome of the cross-validation resulted for RBF and k=6 (around 20 samples per group equivalent to 4 patients excluded from the training set). Each sample was scored for the three tested conditions (ccRCC, RO, CRCC). The highest scoring condition was used to classify a given sample.

Results are shown in a radar plot (**Figure 7B**) and demonstrate 100% accuracy in prediction of renal cancer subtypes. Interestingly despite the similarity to ccRCC (**Figure 5B**) the sarcomatoid patient samples were predicted as CRCC when given the choice between only the three subcancer types. This is in coherence with their cancer type origin prior transformation.

We initially used all 346 differentially abundant histo-molecular features (proteins) to classify the tumor subtypes. Next, we sought to estimate the minimum number of features that suffice to correctly classify all the renal tumor samples (for k=6 and RBF). We used the feature optimization function in the Perseus software to rank the features and tests the error rate when decreasing the number of features (**Figure 7C**). The minimal number out of the 346 features was found to be 43 features (list of ranked proteins can be found in supplementary material 9). Further reducing the number to 21 features resulted in an error rate of 1.6% and as little as 6 features lead to an error rate of 9.6%. Conclusively only a portion of the dataset, i.e. at least 43 features would suffice to successfully classify all the kidney tumor samples. However, keeping an excess of quantified protein features would be beneficial as “safety margin” assuring a high enough number of quantified protein features for robust classification of tumors.

Discussion

The increasing incidence of renal cancer in western countries calls for improved technologies for detection, diagnosis, treatment and prognosis. Innovative mass spectrometry-based applications are beginning to address challenges in clinics and the healthcare sector, such as

the use of targeted proteomics to characterize noninvasive liquid biopsies[42] or the so called iKnife, enabling surgeons to identify cancerous tissue in real time[43, 44]. Mass spectrometry is becoming increasingly applicable in a clinical setting[45, 46]. FFPE sections are a valuable source for mass spectrometry-based diagnosis. As many of the sample preparation steps for MS analysis overlap with the preparation steps for (immuno)histochemical staining, they can be seamlessly fit into the high-throughput sample preparation pipeline for FFPE sections (deparaffination, antigen retrieval) already existing in many hospitals. Distinguishing numerous cancer types or subtypes and making decisions for treatment modalities are daily challenges in hospitals. Approaches that include “digital” large-scale data acquisition and computer-based machine learning algorithms provide deep molecular insight into the respective disease and provides valuable information for early detection, diagnostic and prognostic purposes.

Our proof of concept study demonstrates the potential and benefits of mass spectrometry techniques for detailed characterization of clinical specimen. Specifically, we demonstrate that mass spectrometry provides valuable results in the diagnosis of different renal cancer subtypes (ccRCC, RO and CRCC). The imaging mass spectrometry (MSI) approach allows to collect spatially resolved spectra without *a priori* knowledge of the tissue, thereby enabling the differentiation between cancerous and noncancerous tissue, as well as subtyping of tumors. In our study MALDI-MSI could diagnose 93% of the tested patients correctly (16 out of 18 and 10 out of 10) distinguish between ccRCC and RO with 93.75 % accuracy and between RO and CRCC with 95% accuracy. Despite the high accuracy the classification seemed biased towards RO diagnosis. Our PCA data showed that the patient-to-patient tumor variability is

significant for ccRCC, necessitating a detailed histo-molecular profile for robust MSI performance. The inclusion of many more patients (n>100) to increase the number of renal cancer tumor samples will likely provide even higher confidence and resolve this issue.

Microproteomics analysis based on an optimized LC-MS/MS approach correctly classified all tested renal tumor samples. The efficient peptide separation and sequencing capability of modern LC-MS/MS gave a deeper insight into the renal cancer proteome than possible by the MSI approach. The higher dimensionality of many protein features enhanced the SVM based predictor and achieved 100% accurate classification of renal cancer subtypes.

Notably, unsupervised clustering identified data inconsistencies and irregularities in the patient cohort. An unexpected feature pattern revealed a sarcomatoid transformation within the CRCC cohort, without *a priori* knowledge (**Figure 5A, 5B**). This goes to demonstrate that once the “digital” data is acquired then the computational and statistical applications can uncover relevant and important features of the patient datasets. This sensitivity, specificity and versatility will have major implications for future clinical practices, including histo-molecular pathology technologies.

The presented microproteomics method based on optimized, fast chromatographic separation and fast MS/MS sequencing of peptides identified more than 2100 proteins in thin renal tumor FFPE sections. Using short LC runs of only 15 min. we generated a list of 346 significantly altered proteins (p=0.01). The minimum number of proteins determined to be necessary for 100% accurate tumor classification was 43. This low number of features will enable a targeted proteomics approach aimed at quantifying a select panel of proteins. Using fewer features

would allow a further reduction of LC run time and increase overall sample throughput. Using our fast LC-MS/MS setup we analyzed a total of 125 samples in a series without experiencing blocking of the LC columns, glass capillaries or ESI needles. LC systems such as the EvoSep system[25] that are specifically dedicated for clinical applications and tailored to be used also by non LC-MS experts can add additional robustness to our approach. Furthermore, implementation of image pattern recognition guided pipetting robots may enhance reproducibility and throughput, e.g. using liquid extraction surface analysis (LESA) technology[47, 48]. The latter has been successfully applied in the study of traumatic brain injuries[49] as well as in mouse brain for the identification of proteins and peptides from MSI experiments[50].

Functional protein analysis using bioinformatics tools revealed molecular networks and biochemical processes consistent with previously known macroscopic, morphological and histological features of the renal cancer subtypes. Cancer-type specific proteome expression features correlated to morphological characteristics of the respective cancer type. RO and CRCC exhibited upregulation of mitochondrial associated proteins. Indeed, increased numbers of mitochondria are frequently observed in these cancer types by electron microscopy[51] and have been identified in previous proteomics studies [52]. As most cancer rely on glycolysis as major energy source (Warburg effect) this seems rather unusual. However, those mitochondria are malfunctioning and it has been speculated that increase in number of mitochondria is a cellular response to the presence of dysfunctional mitochondria [53].

In addition to mitochondria associated proteins increased intracytoplasmic associated proteins were detected in CRCC distinguishing it from the other cancer types. Microscopically,

CRCC is distinguished from other renal carcinomas by its pale cytoplasm resulting from large intracytoplasmic vesicles. This accounts for our detection of an increase of intracellular cytoplasm-associated proteins and vesicle proteins, distinguishing CRCC from the other two renal cancer subtypes.

Clear cell renal cell carcinoma frequently contains zones of hemorrhage that are most likely responsible for the increased levels of complement and coagulation cascade associated proteins, as determined by our microproteomics method. ccRCC is also characterized by hypervascular stroma[3], which may account for the strong enrichment of extracellular matrix proteins. Again, enhanced glycolysis is a hallmark of many cancer types including ccRCC [54] correlating well with our detection of upregulated glycolysis associated proteins by microproteomics.

For classification we applied PLS-DA to MSI data and support vector machine to the LC-MSMS data. These common classification methods have previously been applied to MSI for the differentiation of papillary and renal cell carcinoma based on lipidomics analysis [55] as well as for the classification of epithelial ovarian cancer subtypes [16]. There are, however, numerous other classification methods available. Mascini *et al.* used principal component linear discriminant analysis in order to predict treatment response in xenograft models of triple-negative breast cancer [56]. Recently, deep convolutional networks were proposed [57].

Both MSI and short gradient LC-MS/MS microproteomics methods come with their individual advantages. Combining both approaches for routine analysis is most beneficial to improve confidence in diagnosis. In order to create very robust classifiers for use in clinical settings the

promising results of this study need to be further supported in the future by analysis of larger patient cohorts.

With the enormous progress in instrument technology, machine learning [58] and the availability of new databases [59] mass spectrometry is on its way to become a versatile tool in the hospitals of the future.

Acknowledgements

Proteomics and mass spectrometry research at SDU are supported by generous grants to the VILLUM Center for Bioanalytical Sciences (VILLUM Foundation grant no. 7292 to O.N.J.) and PRO-MS: Danish National Mass Spectrometry Platform for Functional Proteomics (grant no. 5072-00007B to O.N.J.). We thank Veit Schwämmle and Livia Rosa Fernandes for advice and discussion on the project.

Authors Contributions:

U.M., O.N.J. and N.M. planned and outlined the project. N.M. provided the patient samples and patient diagnosis. U.M. performed all experiments and data analysis. U.M. and O.N.J. wrote the manuscript.

Competing Interests:

The Authors declare no competing interest

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
2. Jonasch E, Gao J, Rathmell WK. Renal cell carcinoma. BMJ. 2014;349:g4797.
3. Muglia VF, Prando A. Renal cell carcinoma: histological classification and correlation with imaging findings. Radiol Bras. 2015;48(3):166-74.
4. Ng KL, Rajandram R, Morais C, Yap NY, Samaratunga H, Gobe GC, Wood ST. Differentiation of oncocytoma from chromophobe renal cell carcinoma (RCC): can novel molecular biomarkers help solve an old problem? J Clin Pathol. 2014;67(2):97-104.
5. Carella R, Deleonardi G, D'Errico A, Salerno A, Egarter-Vigl E, Seebacher C, Donazzan G, Grigioni WF. Immunohistochemical panels for differentiating epithelial malignant mesothelioma from lung adenocarcinoma: a study with logistic regression analysis. Am J Surg Pathol. 2001;25(1):43-50.
6. Angelidis I, et al. An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics. Nat Commun. 2019;10(1):963.
7. Clark AE, Kaleta EJ, Arora A, Wolk DM. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. Clin Microbiol Rev. 2013;26(3):547-603.
8. Kriegsmann M, Wandernoth P, Lisenko K, Casadonte R, Longuespee R, Arens N, Kriegsmann J. Detection of HPV subtypes by mass spectrometry in FFPE tissue specimens: a reliable tool for routine diagnostics. J Clin Pathol. 2017;70(5):417-23.
9. Ellis SR, Paine MRL, Eijkel GB, Pauling JK, Husen P, Jervelund MW, Hermansson M, Ejsing CS, Heeren RMA. Automated, parallel mass spectrometry imaging and structural identification of lipids. Nat Methods. 2018;15(7):515-8.
10. Hannes Hinneburg PK, Falko Schirmeister, Slavko Gasparov, Peter, H. Seeberger VZ, Daniel Kolarich. Spatial glycomics of histopathological formalin-fixed and paraffin embedded (FFPE) tissue microdissections. Molecular & Cell Proteomics. 2016;in review.
11. Moginger U, et al. Alterations of the Human Skin N- and O-Glycome in Basal Cell Carcinoma and Squamous Cell Carcinoma. Front Oncol. 2018;8:70.
12. Caprioli RM, Farmer TB, Gile J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. Anal Chem. 1997;69(23):4751-60.
13. Stoeckli M, Farmer TB, Caprioli RM. Automated mass spectrometry imaging with a matrix-assisted laser desorption ionization time-of-flight instrument. J Am Soc Mass Spectrom. 1999;10(1):67-71.
14. Casadonte R, Kriegsmann M, Perren A, Baretton G, Deininger SO, Kriegsmann K, Welsch T, Pilarsky C, Kriegsmann J. Development of a Class Prediction Model to Discriminate Pancreatic Ductal Adenocarcinoma from Pancreatic Neuroendocrine Tumor by MALDI Mass Spectrometry Imaging. Proteomics Clin Appl. 2019;13(1):e1800046.
15. Mascini NE, Teunissen J, Noorlag R, Willems SM, Heeren RMA. Tumor classification with MALDI-MSI data of tissue microarrays: A case study. Methods. 2018;151:21-7.
16. Klein O, et al. MALDI-Imaging for Classification of Epithelial Ovarian Cancer Histotypes from a Tissue Microarray Using Machine Learning Methods. Proteomics Clin Appl. 2019;13(1):e1700181.
17. Mallah K, Quanico J, Trede D, Kobeissy F, Zibara K, Salzet M, Fournier I. Lipid Changes Associated with Traumatic Brain Injury Revealed by 3D MALDI-MSI. Anal Chem. 2018;90(17):10568-76.

18. Briggs MT, Kuliwaba JS, Muratovic D, Everest-Dass AV, Packer NH, Findlay DM, Hoffmann P. MALDI mass spectrometry imaging of N-glycans on tibial cartilage and subchondral bone proteins in knee osteoarthritis. Proteomics. 2016;16(11-12):1736-41.
19. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999;20(18):3551-67.
20. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom. 1994;5(11):976-89.
21. Zhao M, et al. A Comparative Proteomics Analysis of Five Body Fluids: Plasma, Urine, Cerebrospinal Fluid, Amniotic Fluid, and Saliva. Proteomics Clin Appl. 2018;12(6):e1800008.
22. Schmidt A, Aebersold R. High-accuracy proteome maps of human body fluids. Genome Biol. 2006;7(11):242.
23. Eliuk S, Makarov A. Evolution of Orbitrap Mass Spectrometry Instrumentation. Annu Rev Anal Chem (Palo Alto Calif). 2015;8:61-80.
24. Falkenby LG, Such-Sanmartin G, Larsen MR, Vorm O, Bache N, Jensen ON. Integrated solid-phase extraction-capillary liquid chromatography (speLC) interfaced to ESI-MS/MS for fast characterization and quantification of protein and proteomes. J Proteome Res. 2014;13(12):6169-75.
25. Krieger JR, Wybenga-Groot LE, Tong J, Bache N, Tsao MS, Moran MF. Evosep One Enables Robust Deep Proteome Coverage Using Tandem Mass Tags while Significantly Reducing Instrument Time. J Proteome Res. 2019;18(5):2346-53.
26. Horning OB, Kjeldsen F, Theodorsen S, Vorm O, Jensen ON. Isocratic solid phase extraction-liquid chromatography (SPE-LC) interfaced to high-performance tandem mass spectrometry for rapid protein identification. Journal of proteome research. 2008;7(8):3159-67.
27. Hørning O. B. TS, Vorm O., Jensen O. N. Solid phase extraction-liquid chromatography (SPE-LC) interface for automated peptide separation and identification by tandem mass spectrometry. International Journal of Mass Spectrometry. 2007;268,(2–3):147-57.

28. Stoeckli M, Staab D, Wetzel M, Brechbuehl M. iMatrixSpray: a free and open source sample preparation device for mass spectrometric imaging. Chimia (Aarau). 2014;68(3):146-9.
29. Kovalchuk SI, Jensen ON, Rogowska-Wrzesinska A. FlashPack: Fast and Simple Preparation of Ultrahigh-performance Capillary Columns for LC-MS. Molecular & cellular proteomics : MCP. 2019;18(2):383-90.
30. Rompp A, Schramm T, Hester A, Klinkert I, Both JP, Heeren RM, Stockli M, Spengler B. imzML: Imaging Mass Spectrometry Markup Language: A common data format for mass spectrometry imaging. Methods in molecular biology. 2011;696:205-24.
31. Bemis KD, Harry A, Eberlin LS, Ferreira C, van de Ven SM, Mallick P, Stolowitz M, Vitek O. Cardinal: an R package for statistical analysis of mass spectrometry-based imaging experiments. Bioinformatics. 2015;31(14):2418-20.
32. Bemis KD, Harry A, Eberlin LS, Ferreira CR, van de Ven SM, Mallick P, Stolowitz M, Vitek O. Probabilistic Segmentation of Mass Spectrometry (MS) Images Helps Select Important Ions and Characterize Confidence in the Resulting Segments. Molecular & cellular proteomics : MCP. 2016;15(5):1761-72.
33. Wold S, Sjostrom M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemometr Intell Lab. 2001;58(2):109-30.
34. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367-72.
35. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016;13(9):731-40.

36. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic acids research. 2015;43(W1):W566-70.
37. Szklarczyk D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research. 2019;47(D1):D607-D13.
38. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 2003;13(9):2129-41.
39. Wisztorski M, Fatou B, Franck J, Desmons A, Farre I, Leblanc E, Fournier I, Salzet M. Microproteomics by liquid extraction surface analysis: application to FFPE tissue to study the fimbria region of tubo-ovarian cancer. Proteomics Clin Appl. 2013;7(3-4):234-40.
40. Hastie T, Tibshirani R, Friedman J. Model Assessment and Selection. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY: Springer New York; 2009. p. 219-59.
41. Hastie T, Tibshirani R, Friedman J. Kernel Smoothing Methods. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY: Springer New York; 2009. p. 191-218.
42. Kim Y, et al. Targeted proteomics identifies liquid-biopsy signatures for extracapsular prostate cancer. Nature Communications. 2016;7.
43. Phelps DL, et al. The surgical intelligent knife distinguishes normal, borderline and malignant gynaecological tissues using rapid evaporative ionisation mass spectrometry (REIMS). Br J Cancer. 2018;118(10):1349-58.
44. Schafer KC, Denes J, Albrecht K, Szaniszlo T, Balog J, Skoumal R, Katona M, Toth M, Balogh L, Takats Z. In vivo, in situ tissue analysis using rapid evaporative ionization mass spectrometry. Angew Chem Int Ed Engl. 2009;48(44):8240-2.
45. Heaney LM, Jones DJ, Suzuki T. Mass spectrometry in medicine: a technology for the future? Future Sci OA. 2017;3(3):FSO213.
46. Longuespee R, Casadonte R, Schwamborn K, Kriegsmann M. Proteomics in Pathology: The Special Issue. Proteomics Clin Appl. 2019;13(1):e1800167.
47. Wisztorski M, Desmons A, Quanico J, Fatou B, Gimeno JP, Franck J, Salzet M, Fournier I. Spatially-resolved protein surface microsampling from tissue sections using liquid extraction surface analysis. Proteomics. 2016;16(11-12):1622-32.
48. Wisztorski M, Quanico J, Franck J, Fatou B, Salzet M, Fournier I. Droplet-Based Liquid Extraction for Spatially-Resolved Microproteomics Analysis of Tissue Sections. Methods Mol Biol. 2017;1618:49-63.
49. Mallah K, Quanico J, Raffo-Romero A, Cardon T, Aboulouard S, Devos D, Kobeissy F, Zibara K, Salzet M, Fournier I. Mapping Spatiotemporal Microproteomics Landscape in Experimental Model of Traumatic Brain Injury Unveils a link to Parkinson's Disease. Mol Cell Proteomics. 2019;18(8):1669-82.
50. Ryan DJ, Nei D, Prentice BM, Rose KL, Caprioli RM, Spraggins JM. Protein identification in imaging mass spectrometry through spatially targeted liquid micro-extractions. Rapid communications in mass spectrometry : RCM. 2018;32(5):442-50.
51. Thoenes W, Storkel S, Rumpelt HJ. Human chromophobe cell renal carcinoma. Virchows Arch B Cell Pathol Incl Mol Pathol. 1985;48(3):207-17.
52. Drendel V, Heckelmann B, Schell C, Kook L, Biniossek ML, Werner M, Jilg CA, Schilling O. Proteomic distinction of renal oncocytomas and chromophobe renal cell carcinomas. Clin Proteomics. 2018;15:25.
53. Gasparre G, Romeo G, Rugolo M, Porcelli AM. Learning from oncocytic tumors: Why choose inefficient mitochondria? Biochimica et biophysica acta. 2011;1807(6):633-42.
54. Courtney KD, et al. Isotope Tracing of Human Clear Cell Renal Cell Carcinomas Demonstrates Suppressed Glucose Oxidation In Vivo. Cell Metab. 2018;28(5):793-800 e2.

55. Dill AL, Eberlin LS, Zheng C, Costa AB, Ifa DR, Cheng L, Masterson TA, Koch MO, Vitek O, Cooks RG. Multivariate statistical differentiation of renal cell carcinomas based on lipidomic analysis by ambient ionization imaging mass spectrometry. Anal Bioanal Chem. 2010;398(7-8):2969-78.
56. Mascini NE, Eijkel GB, ter Brugge P, Jonkers J, Wesseling J, Heeren RM. The use of mass spectrometry imaging to predict treatment response of patient-derived xenograft models of triple-negative breast cancer. Journal of proteome research. 2015;14(2):1069-75.
57. Behrmann J, Etmann C, Boskamp T, Casadonte R, Kriegsmann J, Maass P. Deep learning for tumor classification in imaging mass spectrometry. Bioinformatics. 2018;34(7):1215-23.
58. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015;349(6245):255-60.
59. Gessulat S, et al. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. Nat Methods. 2019;16(6):509-18.

Figure Legends

Figure 1 Overview on sample heterogeneity by unsupervised clustering. A) Spatial Shrunken centroid clustering of ccRCC and RO data obtained by imaging mass spectrometry of ccRCC and RO tissue sections. Based on differences and similarities in the spectra each pixel was automatically assigned a certain cluster (indicated by a different color). B) Average MALDI mass spectra of the respective tumor cluster reveal distinct features and individual variations in the m/z signals. C) HE-stain of section from same FFPE block tumor area is indicated in red

Figure 2: 3D PCA score plot from imaging MALDI MS experiments on tumor tissues. Each plot contains the extracted pixel data from all patients of a given cancer type. Data from ccRCC (A) (magenta) and CRCC (B) (blue) are compared to RO (yellow) respectively. The graph displays 3 principle components (PC1, PC2, PC3) plotted against each other shown from 2 angles for each comparison. Separation of data points between ccRCC and RO can be observed whereas RO and CRCC exhibit a great number of overlapping features.

Figure 3: Cross-validated PLS-DA classification of imaging MALDI MS data. Classification of 9 ccRCC and 9 RO (a-i) sample as well as 5 RO sample and 5 CRCC sample (j-n). Pathology diagnosis of the patient sample is indicated on the left side of each sample (OC, ccRCC or CRCC). Each spectrum-containing pixel is predicted individually. Samples classified as RO appear in blue, ccRCC appear in red and CRCC appear in yellow. Each sample is

predominantly predicted in the correct diagnosis, achieving 94-95% accuracy (pixel based value).

Figure 4: PLS coefficients as a function of *m/z*. The diagram displays the impact of each detected *m/z* signal feature in imaging MS data in the respective PLS-DA prediction (spectra are binned to 0.25 m/z bins). Positive coefficient indicates presence or higher abundance in the respective condition. Negative coefficient indicates absence or lower abundance of the m/z value in the respective condition. Top-6 m/z values and their association to cancer subtype are shown in the table.

Figure 5: Unsupervised Renal cancer subtype classification by microproteomics. A) Heatmap and hierarchical clustering of differential relative protein abundances. Columns indicate samples and rows indicate proteins. The renal cancer subtype of the patient sample is indicated in colored bars on top. The graph shows the large similarities in protein expression profiles among patient samples with the same cancer subtype causing them to cluster together. Furthermore, hierarchical clustering of the protein abundances reveals protein cluster that are detected in a cancer subtype specific manner. Protein groups selected for subsequent network analysis are indicated by color blocks on the y-axis dendrogram (groups 1-4). B) Principal component analysis of the sample set. Dotted ellipses are such that with a probability of 95% a new observation from the same group will fall inside the area. The first (PC1) and second (PC2) component explain 17.6 % of the total variance whereas the other

components lie at 7.7% and 4.4% respectively. There is a clear separation of ccRCC and RO samples already in the first two principal components. Differences between RO and CRCC are finer and become only evident when looking at components that display lower variance (PC2:PC3 and PC3:PC4). The small group the CRCC originated sarcomatoid renal cancers samples cluster relatively far from the other CRCC samples. They can be found either together with the ccRCC or entirely separated.

Figure 6: Descriptive enriched terms (PantherDB) for the different protein groups for each of the 4 protein groups. Terms for groups are marked in the respective color. Proteins were run against both proteins identified in the experiment (blue) as well as the human genome as background (red). Fold enrichment (increase over expected value) as well as - log of the false discovery rate (FDR) are shown.

Figure 7: Accurate SVM model classification of renal tumor subtypes by rapid microproteomics. A) Development of the error rate in relation to increasing k-value. RBF performs slightly better than linear kernel function. From value k=4 and higher, error rates vary between 0% and 0.8% (1 wrong prediction out of 125) for RBF and between 0% - 1.6 % for linear kernel. B) Radar plot of the cross-validated classification (k=6, kernel=RBF) of proteome profiles obtained from each tumor section extraction spot. Scores for each of the three tumor types are plotted. Scores range from lowest (center) to highest (outer circle). The highest score indicates highest likelihood for the respective diagnosis. Scores for ccRCC are plotted in magenta, score for CRCC are plotted in yellow and scores for RO are plotted in

blue. The pathological diagnosis for each sample is indicated on the outside of the radar plot. The plot shows that all samples score highest in correlation with the respective cancer type indicating the high accuracy of the classification. C) Feature optimization. The error rate for linear kernel and RBF are plotted over the number of ranked features (proteins). Decreasing feature number results in increase of false predictions. Minimum number of features for 0% error rate is at 43 using RBF and 86 for linear kernel.

Figure 1

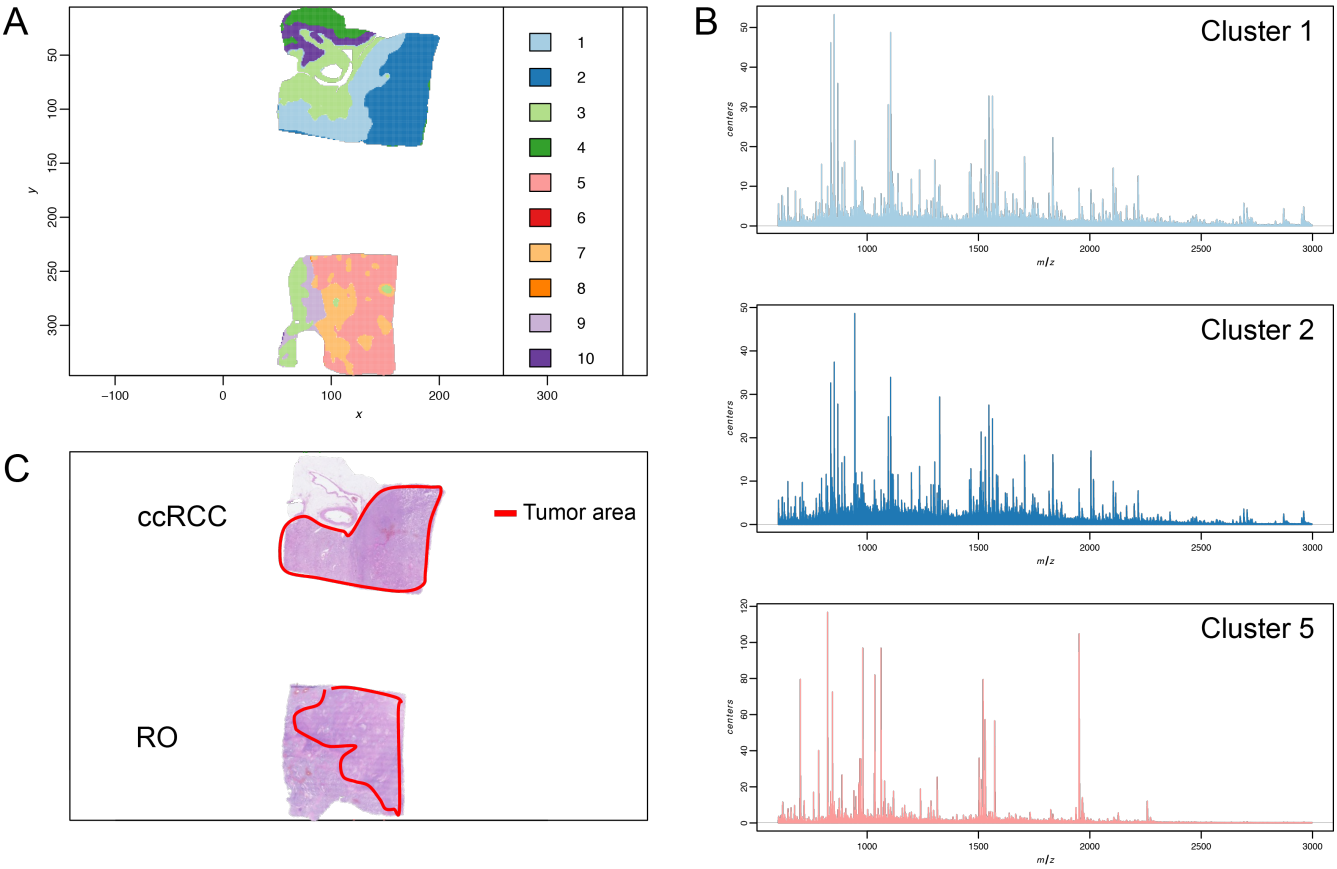


Figure 2

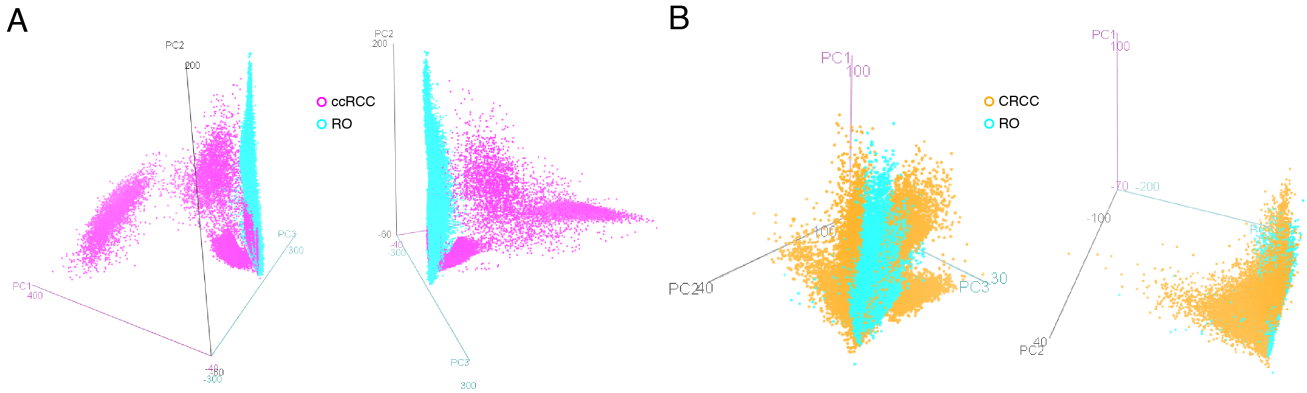


Figure 3

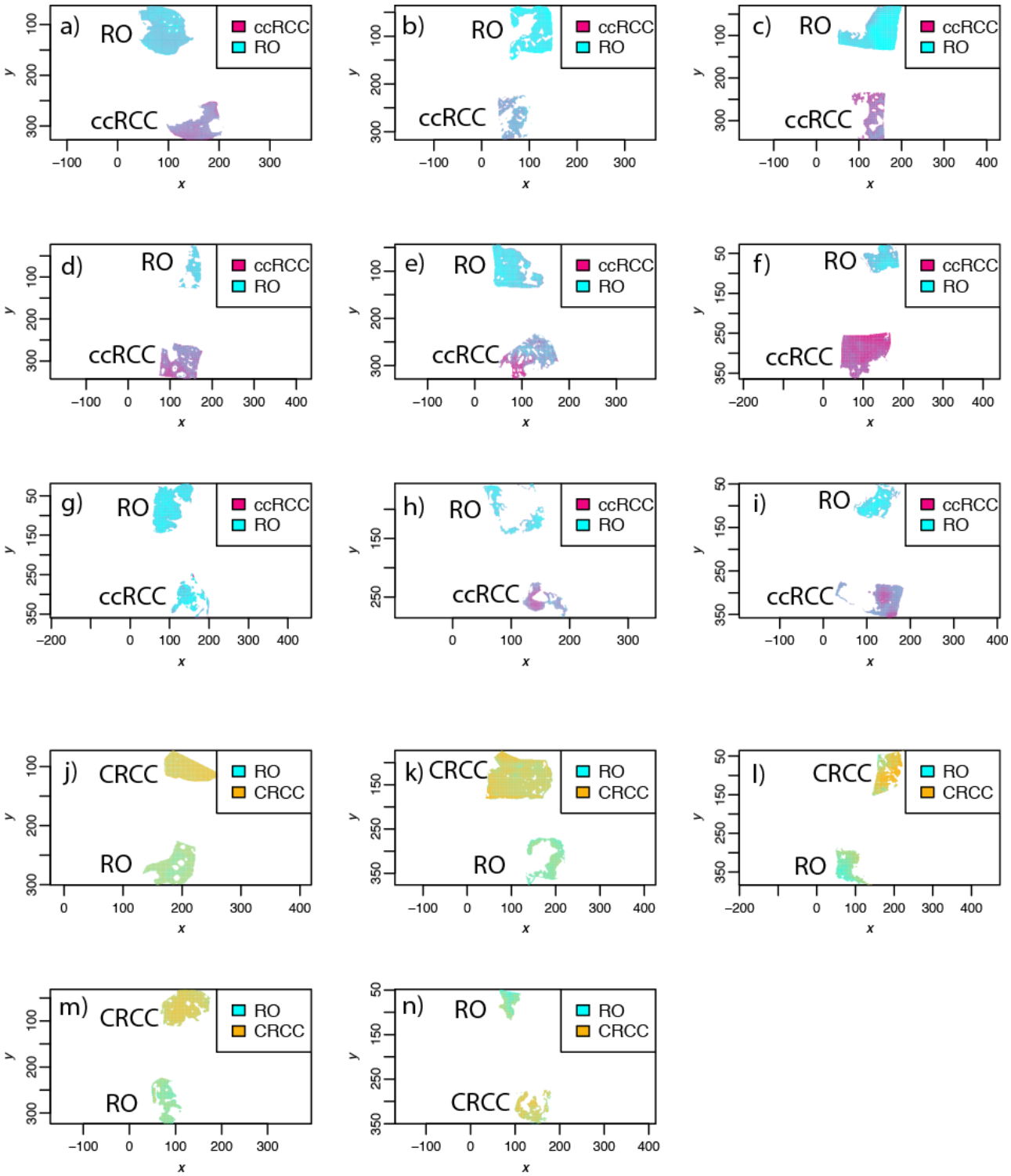
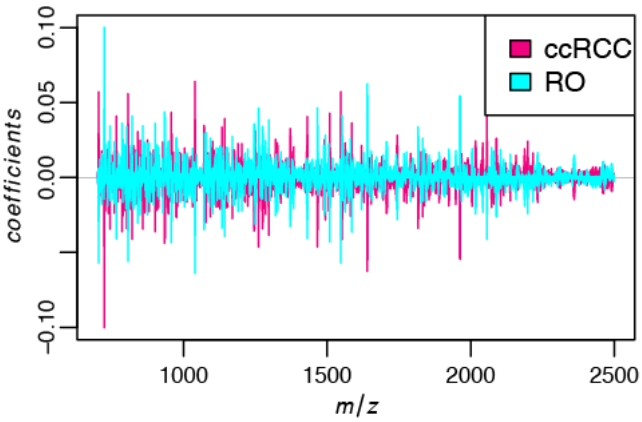
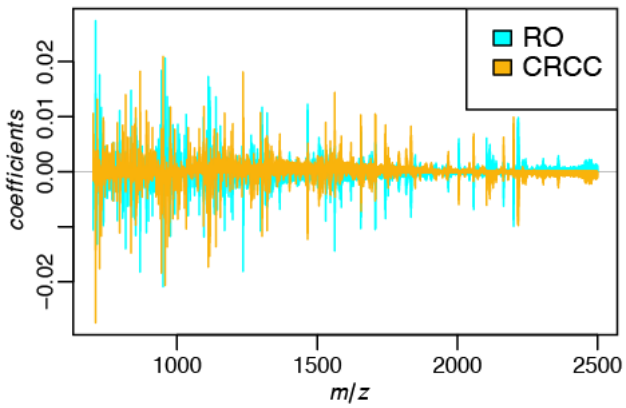

a)
RO
ccRCC
ccRCC
RO
b)
RO
ccRCC
ccRCC
RO
c)
RO
ccRCC
ccRCC
RO
d)
RO
ccRCC
ccRCC
RO
e)
RO
ccRCC
ccRCC
RO
f)
RO
ccRCC
ccRCC
RO
g)
RO
ccRCC
ccRCC
RO
h)
RO
ccRCC
ccRCC
RO
i)
RO
ccRCC
ccRCC
RO
j)
CRCC
RO
RO
CRCC
k)
CRCC
RO
RO
CRCC
l)
CRCC
RO
RO
CRCC
m)
CRCC
RO
RO
CRCC
n)
RO
RO
CRCC
CRCC
x
y

Figure 4:





	mz	ncomp	diagnosis	coefficients	loadings	weights
1	723.5	RO	0.07014631	0.02967771	-0.0016929
2	806.5	ccRCC	0.04898323	0.03053029	0.00239848
3	704.5	ccRCC	0.04190643	0.06228593	0.03978363
4	1039.5	ccRCC	0.03406684	0.0106639	0.00452998
5	1465.75	RO	0.03395778	-0.0395423	0.02236657
6	1640	RO	0.03275541	-0.0115516	0.01765775

	mz	ncomp	diagnosis	coefficients	loadings	weights
1	709.25	RO	0.02745346	-0.0279879	-0.0548115
2	950.5	CRCC	0.02091736	0.00175355	-0.0303994
3	957.5	RO	0.02063297	-0.0225405	-0.0462015
4	944.5	RO	0.01841526	-0.0428202	-0.1536461
5	868.5	CRCC	0.01823591	-0.0597823	0.00681281
6	1235.5	CRCC	0.01811409	0.05127861	0.06155005

Figure 5

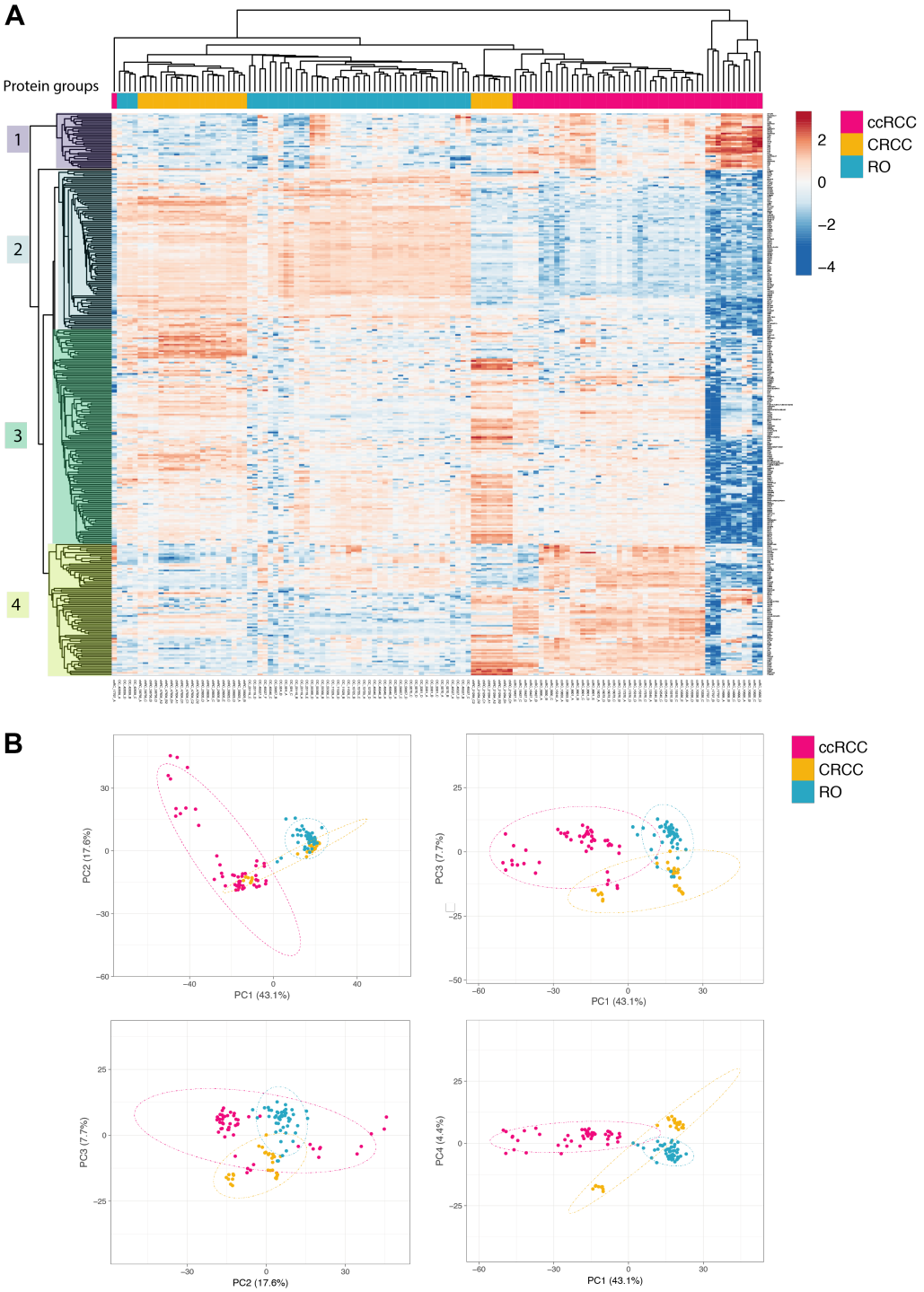

A
Protein groups
1
2
3
4
2
0
−2
−4
ccRCC
CRCC
RO
B
ccRCC
CRCC
RO
PC2 (17.6%)
PC1 (43.1%)
PC3 (7.7%)
PC1 (43.1%)
PC3 (7.7%)
PC2 (17.6%)
PC4 (4.4%)
PC1 (43.1%)

Figure 6

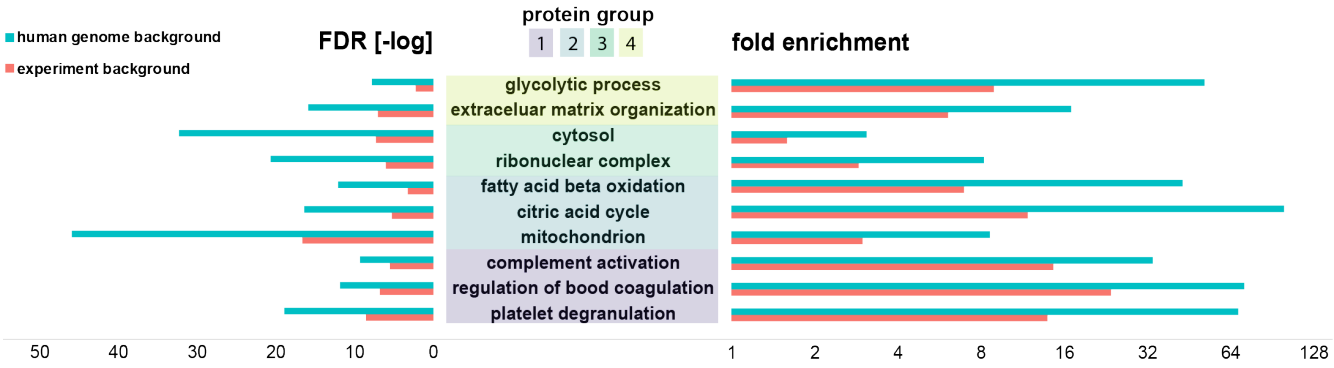
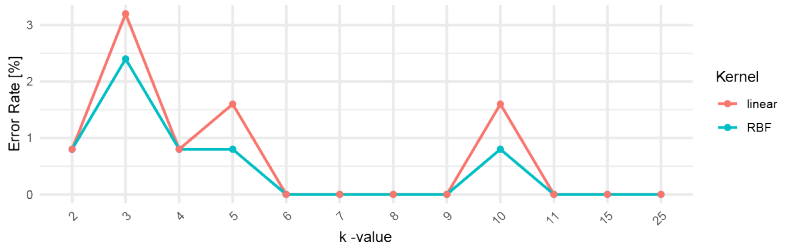
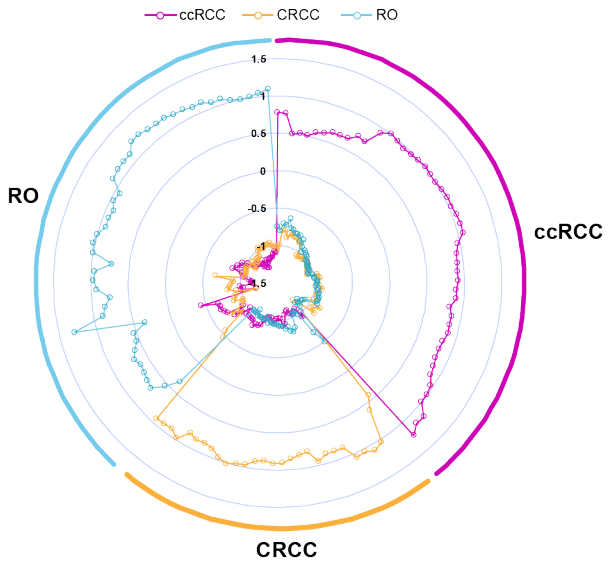

human genome background
experiment background
FDR [-log]
protein group
1 2 3 4
fold enrichment
glycolytic process
extraceluar matrix organization
cytosol
ribonuclear complex
fatty acid beta oxidation
citric acid cycle
mitochondrion
complement activation
regulation of bood coagulation
platelet degranulation
50 40 30 20 10 0
1 2 4 8 16 32 64 128

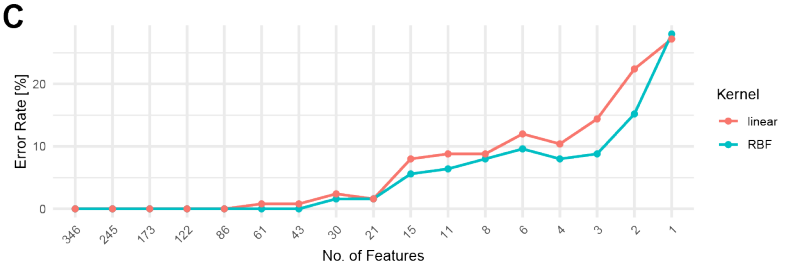
Figure 7

A



B



Predicted: ccRCC	Predicted: CRCC	Predicted: RO	T: True group
49	0	0	True: ccRCC
0	29	0	True: CRCC
0	0	47	True: RO

C