Histo-molecular differentiation of renal cancer subtypes by mass spectrometry imaging and rapid proteome profiling of formalin-fixed paraffin-embedded tumor tissue sections

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, DK-2760 Måløv, Denmark

Keywords:

Renal Cell Cancer, MALDI Mass Spectrometry Imaging (MALDI MSI), Microproteomics, Statistical Classification, Liquid Chromatography Mass Spectrometry (LC-MS)

Abstract

Pathology differentiation of renal cancer types is challenging due to tissue similarities or overlapping histological features of various tumor (sub)types. As assessment is often manually conducted outcomes can be prone to human error and therefore require high-level expertise and experience. Mass spectrometry can provide detailed histo-molecular information on tissue and is becoming increasingly popular in clinical settings. Spatially resolving technologies such as mass spectrometry imaging and quantitative microproteomics profiling in combination with machine learning approaches provide promising tools for automated tumor classification of clinical tissue sections.

In this proof of concept study we used MALDI-MS imaging (MSI) and rapid LC-MS/MS-based microproteomics technologies (15 min/sample) to analyze formalin-fixed paraffin embedded (FFPE) tissue sections and classify renal oncocytoma (RO, n=11), clear cell renal cell carcinoma (ccRCC, n=12) and chromophobe renal cell carcinoma (ChRCC, n=5). Both methods were able to distinguish ccRCC, RO and ChRCC in cross-validation experiments. MSI correctly classified 87% of the patients whereas the rapid LC-MS/MS-based microproteomics approach correctly classified 100% of the patients.

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

41

42 **Introduction**

43 Kidney cancer (renal cell carcinoma, RCC) accounts for 2.2% of all diagnosed cancers and is
44 the 13th most common cause of cancer deaths worldwide [1]. Clear cell renal cell carcinoma
45 (ccRCC) constitutes 70% of all kidney cancers [2] and exhibits the highest rate of metastasis
46 among renal carcinomas. Two other common but less aggressive subtypes of renal
47 carcinoma are chromophobe renal cell carcinoma (ChRCC) and the essentially benign renal
48 oncocytoma (RO), which account for 5% and 3-7 % of all cases, respectively [3,4]. The ability to
49 distinguish between the malignant cancer types ccRCC and ChRCC and the benign RO is
50 crucial for a patient in terms of prognosis, progression and intervention strategies as severe
51 as total nephrectomy. Histopathological kidney cancer diagnostics faces many challenges in
52 daily routine. Typically, test panels consisting of a combination of different chemical and
53 immuno-histochemical staining methods are used to systematically obtain a diagnosis [5].
54 Overlapping histological features can make it difficult to differentiate tumor types. Analysis,
55 interpretation and diagnosis/prognosis greatly rely on visual inspection and the experience of
56 the involved clinical pathologists. Complementary techniques such as MRI and electron
57 microscopy involve costly instrumentation. Moreover, specific antibodies for staining can be
58 expensive or unavailable for certain molecular targets. Mass spectrometry is emerging as a
59 promising new tool in translational research, from molecular imaging of tissue sections to
60 deep protein profiling of tissue samples [6]. The digital data readout provided by high mass
61 accuracy mass spectrometry and feasibility of molecular quantification makes it a very
62 attractive technology in translational research for investigating human diseases and for
63 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 (MSI) 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 histo-molecular information from both MALDI MS imaging (MSI), *in situ* protein digestion and LC-MS/MS applied to detailed characterization of 5 µm cancer FFPE

85 tissue sections will provide spatial molecular maps and sufficiently deep proteome profiles to
86 characterize and classify tumor subtypes. We investigated this by testing a series of
87 malignant and benign renal carcinomas, including clear cell renal cell carcinoma (ccRCC),
88 chromophobe renal cell carcinoma (ChRCC) and renal oncocytoma (RO). We obtained histo-
89 molecular images at a resolution of 150µm x 150µm that sufficed to spatially resolve features
90 to distinguish tumor subtype areas from surrounding tissue. Miniaturized sample preparation
91 by *in situ* protein digestion was used to recover peptides from distinct areas of the FFPE
92 tumor sections for rapid proteome profiling by LC-MS/MS.

93

94 **Material and Methods:**

95 **Materials**

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

103 **Formalin fixed paraffin embedded samples:**

104 Patient samples were collected at Odense University Hospital, Denmark. All samples were
105 obtained upon patient's consent. Formalin fixed paraffin embedded (FFPE) tissues from 11
106 RO patients, 12 ccRCC patients and 5 ChRCC patients were used for LC-MSMS analysis (for
107 ChRCC due to the lower number of patients 2 subsequent slides were used from 2 patients
108 adding up to a total of 7 sections). Out of the patient cohort 9 RO, 9 ccRCC and 5 ChRCC
109 were used for mass spectrometry imaging analysis.

110 **Tissue preparation:**

111 *Preparation of formalin fixed paraffin embedded samples*

112 FFPE blocks were cut into 5 µm thick sections and mounted onto indium tin oxide (ITO)
113 covered glass slides (for MSI) or regular microscopy glass slides (for LC-MS/MS). Before
114 deparaffination slides were left on a heated block at 65° C for 1 hour to improve adhesion (an
115 overview on the used FFPE samples can be found in supplementary table S1 and S2).

116 *Deparaffination*

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

121 *Antigen retrieval*

122 Tissue slides were heated in 10mM citric acid buffer pH 6 for 10 min in a microwave oven at
123 400 Watt (just below the boiling point) before left for further 60 min incubation at 98°C on a

124 heating plate. Slides were cooled down to room temperature and incubated for 5 minutes in
125 25 mM ammonium bicarbonate (ABC) buffer. Slides were allowed to dry before application of
126 trypsin protease.

127 *Tryptic digest*

128 For MALDI MS imaging:

129 20µg of Trypsin (Promega) was used per slide and was dissolved at a concentration of
130 100ng/µl in 25mM ABC /10% ACN before being deposited on the tissue using the
131 iMatrixSpray [28] device equipped with a heating bed (Tardo Gmbh, Subingen, Switzerland)
132 using the following settings: sprayer height = 70mm, speed = 70mm/s, density = 1µL/cm^2,
133 line distance= 1 mm , gas pressure= 2.5 bar, heat bed temperature= 25°C . After trypsin
134 deposition the slides were incubated in a humid chamber containing 10mM ABC/ 50% MeOH
135 at 37°C over night.

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

137 Droplets of 2µl Trypsin solution (50ng/µL in 25mM ABC /10%ACN, 0.02%SDS) were
138 deposited using a gel loading pipet tip. Droplets were placed on 3-4 different tumor areas of
139 each FFPE tissue section. The droplets were shortly allowed to dry in order to prevent
140 spreading across the tissue. Slides were transferred to a humid chamber (10mM ABC /50%
141 MeOH) for overnight digestion at 37°C. After digest the digestion spots were extracted twice
142 with 2µL of 0.1% FA and twice with 1.5µL of 30%ACN. Samples were collected, speedvac
143 dried. Before injection samples were reconstituted in 0.05%TFA and shortly spun down.

144 *Matrix application*

145 Matrix solutions were freshly prepared from recrystallized α-cyano-4-hydroxycinnamic acid
146 (CHCA) matrix (10mg/mL in 50% Acetonitrile 1% TFA). Matrix was sprayed using the
147 iMatrixSpray (Tardo, Switzerland). Temperature of the heatbed was set at 25°C. The sprayer
148 distance was set to 70mm. Spray speed was set to 100 mm/s. Matrix was sprayed in 3
149 rounds: 8 cycles with a flowrate of 0.5µl/cm^2 line distance of 1mm, 8 cycles of 1µl/cm^2 line
150 distance of 1mm, 8 cycles of 1µl/cm^2 and a line distance of 2mm.

151 *MALDI MS Imaging data acquisition*

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

159

160 *LC-MS/MS analysis*

161 LC-MS/MS data was acquired by an Orbitrap Q-Exactive HF-X (Thermo, Bremen) coupled to
162 an Ultimate 3000 capillary flow LC-system. Setup was adopted and modified from Thermo
163 Scientific Technical note: 72827. Peptide samples were loaded at 150µl/min (2% ACN, 0.05%
164 TFA) for 30 sec onto a 5µm, 0.3 x 5 mm, Acclaim PepMap trapping cartridge (Thermo

165 Scientific). Samples were then eluted onto a pulled emitter analytical column (75µm ID,
166 15cm). The analytical column was "flash-packed" [29] with C_{18} Reprosil Pur resin (3µm) and
167 connected by Nanoviper fittings and a reducing metal union (Valco, Houston, TX). The
168 flowrate of the 15 min gradient was 1.2 µL/min with solvent A: 0.1% formic acid (FA) and
169 solvent B: 0.1% FA in 80% ACN. Gradient conditions for solvent B were as followed: 8% to
170 25% in 10 min, 25% to 45% in 1.7 min. The trapping cartridge and the analytical column were
171 washed for 1 min at 99%B before returning to initial conditions. The column was equilibrated
172 for 2 min.

173

174 *Data Processing of MALDI MS imaging data*

175 The data was baseline subtracted, TIC normalized and statistically recalibrated and then
176 exported into imzML format [30] using the export function of FlexImaging software (Bruker). The
177 exported mass range was 600-3000 m/z with a binning size of 9600 data points. The imzML
178 files were imported into the R environment (version: 3.4.1) and further processed and
179 analyzed using the R MSI package: Cardinal (version: 2.0.3 & 2.4) [31]. In order to extract pixels
180 of tumor tissue each sample was preprocessed as follows: peaklist was generated by peak
181 picking in every 10th spectrum and subsequent peak alignment. The whole data was then
182 resampled using the "height" option and the previous created peaklist as spectrum reference.
183 PCA scores were plotted using car-package (version 3.0.6). Samples were clustered using
184 spatial shrunken centroid clustering [32]. Subsequently, pixel coordinates of cluster containing
185 tumor areas (HE-stain comparison, supplementary material 1: Figure S3) were extracted and

186 if necessary manually trimmed, so that result files predominantly contained data from tumor
187 areas. The obtained coordinates were then used to extract the corresponding pixel from the
188 unprocessed imzML file. Each tumor type was assigned with a diagnosis factor (ccRCC, RO
189 or ChRCC), which was later used as y-argument in the cross-validation. All extracted imaging
190 acquisition files were further restricted to a mass range of m/z 700-2500. Data was resampled
191 with step size 0.25 Da to allow combining them into one file for further processing.
192 Classification and cross validation were performed using partial least square discriminant
193 analysis (PLS-DA) [33]. PLS components were tested for optimum with 34 components
194 (supplementary material 1: Figure S1).

195

196 *LC-MS/MS data processing*

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

200 Data filtering, processing and statistical analysis of the MaxQuant output files was performed
201 using the Perseus [35] framework (version 1.6.1.3). Data was filtered excluding the following
202 hits: only identified by site, contaminants and reversed. The log-transformed data was filtered
203 for proteins present in at least 70% of all experiments. Significance filtering was based on
204 ANOVA testing, using FDR threshold of 0.01 with Benjamini Hochberg correction. In order to
205 perform PCA analysis and classification missing values were imputed by normal distribution.
206 Data shown in heatmap was Z-score normalized. Perseus output tables were transferred into

207 ClustVis [36] for visualization of hierarchical clustering and principle compound analysis (PCA).
208 Gene Ontology and functional analysis was performed via String DB (version 11.0.0) [37] and
209 Panther DB (version 14.1) [38]. For Panther DB analysis background genome was the human
210 genome and the total of identified proteins from all LC-MSMS runs in the experiments
211 (supplementary material 4-8). Feature optimization cross validation type was “n-fold” with n =
212 6. Kernel was either linear or RGF. All other settings were left on their default value.

213

214 **Results:**

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

220

221 *Imaging by MALDI mass spectrometry*

222 All samples were prepared as 5 µm thin FFPE tissue/tumor sections. The entire FFPE tissue
223 section was analyzed by imaging MALDI MS imaging (MSI). The data was subsequently
224 processed by unsupervised clustering (spatial shrunken centroid clustering [32]). The clustering
225 results **(Figure 1A and 1B)** illustrate the heterogeneity of the tissue sections coming from
226 various tissue types such as stroma, fibrotic fatty or healthy tissue and the capabilities of
227 imaging MSI for the delineation of cancerous and non-cancerous tissue. Furthermore, when
228 comparing the tumor area of the HE-stain/microscopy with the results from the mass
229 spectrometry imaging based clustering, spectral differences even within the tumor tissue itself
230 can be observed (**Figure 1A and 1C**).

231 Guided by the unsupervised clustering outcome and the corresponding image obtained by
232 HE-staining, pixels from non-relevant surrounding tissue were discarded and only pixel
233 clusters containing actual tumor tissue were used for subsequent comparative analyses
234 (schematic workflow overview can be found in supplementary material: Figure S2).

235 In mass spectrometry imaging, principal component analysis (PCA) is often used for initial
236 analysis of a given data set. Variance and similarities within the image sample set were
237 estimated by PCA over the first 3 components. From a pathology viewpoint RO and ChRCC
238 are more difficult to distinguish than ccRCC and ChRCC. As the sample holder for the
239 imaging experiments can only hold 2 slides at a time, we first compared two conditions in a
240 pairwise manner: 9 ccRCC vs. 9 RO **(Figure 2A)** and 5 RO vs. 5 ChRCC **(Figure 2B)**. Then
241 the data set was combined to compare all three cancer conditions (9 ccRCC, 9 RO, 5
242 ChRCC) to each other **(Figure 2C).** PCA using the first three principle components separate
243 ccRCC well from RO and ChRCC (**Figure 3A, 3C).** Data points from ccRCC showed a wide
244 spread and were splitting into 2 sub-populations. In contrast, the data from RO and ChRCC
245 samples cluster in a much tighter manner and with some overlap **(Figure 3A, 3B, 3C)**. This is
246 particularly the case when considering all three cancer types together (**Figure 3C)**. When
247 compared in a pairwise manner RO and ChRCC show slight separation **(Figure 3B)**
248 suggesting at least some degree of histo-molecular differences between these cancer types.
249 Some overlapping data points in the different cancer type datasets can be observed indicating
250 histo-molecular spectral similarity in parts of the patient tumor tissues. The spread of ccRCC
251 data points in PCA, as compared to the RO and ChRCC subtypes, suggests a greater
252 heterogeneity among the ccRCC patient samples (also observed by LC-MS/MS, see later
253 section).

254 Next, we assessed the ability of the MSI data to distinguish and classify renal cancer
255 subtypes. We generated a classifier based on partial least squares discriminant analysis
256 (PLS-DA) that can then be applied to a given MSI sample set. Due to the limited number of

257 FFPE kidney tumor samples we chose to use a cross-validation strategy that maximizes the
258 use of a sample set for model generation and testing. In this approach a classifier is trained
259 with imaging data from all samples, except for one sample that is set aside. As this sample is
260 not part of the classifier model it can then be used for testing purposes. This was repeated as
261 many times as there are samples ultimately allowing for testing the complete dataset (for n
262 samples we obtain n classifiers and n tested samples).

263 The optimized PLS-DA model resulted in an accuracy of cancer subtype prediction of 93% for
264 ccRCC and 88% for RO and for ChRCC. Results of the cross-validation study using PLS-DA
265 to classify 23 kidney tumor samples are depicted in **Figure 3**. The PLS-DA prediction scores
266 for each of the three possible tumor type outcomes are shown, i.e. ccRCC, RO and ChRCC.
267 The scores obtained for each pixel are presented by intensity scaled colors plotted over the
268 respective x-y-coordinate of the tissue/tumor sections.

269 Twenty of the 23 patient tumor samples were correctly assigned by the PLS-DA model
270 showing highest intensity for the respective cancer condition **(Figure 3)**. Eight out of nine
271 ccRCC samples were correctly assigned **(Figure 3a-i).** The PLS-DA classification provided
272 high scores for ccRCC samples and clearly distinguished the ccRCC samples from the other
273 two kidney tumor types (Figure 3, left panels). This is in accordance with the PCA results.
274 Likewise, the PLS-DA model provided low scores for ccRCC in the cases of RO and ChRCC
275 samples (Figure 3, middle and right panel). One ccRCC sample was incorrectly classified by
276 PLS-DA as RO **(Figure 3g**).

277 All 5 ChRCC samples (**Figure 3 j-n**) were correctly assigned having the highest score for the
278 ChRCC condition (right panel). PCA indicated mass spectral similarities between the RO and
279 ChRCC samples. Likewise, the PLS-DA model reflects such similarities in the classification
280 outcome. Two ChRCC patient samples received highest scores for ChRCC but only slightly
281 lower scores for RO (**Figure 3k, 3n**). Furthermore, in the case of two RO sample the
282 classification could exclude ccRCC as diagnosis but was not able to further provide clear
283 information on if the cancer type in question is RO or ChRCC **(Figure 3a, 3e)**.

284 One kidney tumor sample (Figure 3l) exhibited a rather unusual scoring pattern as compared
285 to the other tumor samples. This particular sample received high scores for both ccRCC and
286 ChRCC classification (ChRCC being the highest). As mentioned above, we typically observed
287 clear distinction between ccRCC and ChRCC in all the other cases. Upon further pathology
288 and microproteomics analysis this tumor section was re-classified as a sarcomatoid
289 transformation (see below), i.e. a tumor type not included in the PLS-DA model used for
290 classification.

291 The relative importance of individual histo-molecular features of the classifier can be
292 visualized by plotting the PLS coefficients for each condition as a function of m/z values
293 **(Figure 4)**. A positive coefficient indicates presence or higher abundance of the m/z value in
294 the respective cancer model. A negative coefficient indicates absence or lower abundance in
295 the respective condition. For ccRCC the two highest-ranking m/z values were m/z=723.5 and
296 m/z=704.5. The two highest values for RO were m/z=806.5 and m/z=1640.0 whereas the
297 most influential signals for ChRCC comprised m/z= 1169.5 and m/z=1039.5 (A top 100 list of
298 the features can be found in **supplementary material 2).** Unfortunately, we were not able to

299 obtain informative MALDI MS/MS fragment ion spectra in order to reveal the identity of these
300 peptide ion signals. Nevertheless, for classification purposes the knowledge of distinct
301 protein/peptide identities (m/z values) is not necessary as long as the signal is characteristic
302 for the tested condition.

303 In conclusion mass spectrometry imaging provided histo-molecular tumor profiles that can be
304 used to distinguish renal cancer subtypes. However, the misclassification of one ccRCC
305 patient and uncertainty of two additional diagnosis outcomes suggested that additional
306 independent test methods would be beneficial for confident classification of renal cancer
307 tumor types.

308

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

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

320 recovered and analyzed by mass spectrometry [39]. We reduced the LC-MS/MS analysis time
321 from 90 minutes to 15 min by using short LC gradients and rapid MS/MS functions, allowing
322 for a sample throughput of up to 80 samples per day. A total of 125 *in situ* extracted areas
323 from renal tumor sections were analyzed. Two to six *in situ* extracts were taken from each
324 renal tumor sample (11 RO sections from 11 patients: 47 extraction spots; 12 ccRCC sections
325 from 12 patients: 49 extraction spots; 7 ChRCC sections from 5 patients: 29 extraction spots).
326 Fast LC-MS/MS based microproteomics analysis of all 125 *in situ* digested tumor areas
327 resulted in a total of 2124 identified human proteins. We filtered the data for proteins that
328 were present in at least 70 % of all samples thereby reducing the protein number to 412
329 proteins. Comparative data analysis was performed for proteins that were significantly altered
330 (FDR=0.01) in any of the renal cancer subtypes resulting in a list of 346 differentially
331 regulated proteins. We then used unsupervised hierarchical clustering and PCA to identify
332 similarities and differences between the tumor samples. The x-axis dendrogram of the
333 heatmap shows that the majority of the renal tumor samples grouped according to cancer
334 subtype RO, ccRCC or ChRCC **(Figure 5A)**. Several large clusters of "co-regulated" proteins
335 are evident on the y-axis dendrogram and heatmap for the individual cancer subtypes. This
336 clearly demonstrates that there are renal cancer subtype specific histo-molecular features and
337 patterns in the microproteomics dataset.

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

342 These differences and similarities were also revealed by PCA analysis of the microproteomics
343 dataset. RO and ChRCC separate clearly from ccRCC **(Figure 5B)**. RO and ChRCC
344 datapoints are located close together, indicating that differences between the RO and ChRCC
345 cancer subtypes are less dominant. When considering principal components exhibiting less
346 variance (PC3 and PC4), separation of RO and ChRCC sample data is observed (**Figure**
347 **5B**).

348 We observed eight ChRCC proteomics datasets that separated clearly form the other ChRCC
349 datasets, both in hierarchical clustering analysis (**Figure 5A**) and PCA (**Figure 5B**). The
350 protein expression profile of these 8 samples exhibited some similarities to both ChRCC and
351 ccRCC. Interestingly, this data originated from a tumor from a single patient. This was the
352 same patient that also exhibited outlier MSI data with similarities to both ChRCC and ccRCC
353 tumor types, as discussed above (Figure 3l). Further pathology analysis revealed that these
354 samples were sarcomatoid renal cancer, originating from ChRCC and, thus, indeed different
355 from the other ChRCC samples.

356

357 *Protein differences in cancer subtypes*

358 Hierarchical clustering of the proteomics datasets revealed major differences in relative
359 protein abundance between the three renal cancer tumor types. (**Figure 5A**). We investigated
360 the nature of these histo-molecular differences by examining the correlation of these proteins
361 to cellular structures, functions, or biochemical processes. Protein groups that exhibited
362 distinctive abundances for the respective cancer type (**Figure 5:** ccRCC: group 1 & 4, RO:

363 group 2, ChRCC: group 3) were searched for their involvement in protein interaction networks
364 (supplementary material 9) as well as for their functional roles by using gene ontology (GO)
365 enrichment (**Figure 6**, supplementary material 3-8). We compared GO enrichment relative to
366 the experimental gene background comprising all genes corresponding to all 2124 identified
367 proteins in the LC-MSMS experiments (Figure 6, experimental background: blue) and the
368 complete human genome (Figure 6, human genome background: red).

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

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

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

383 These functionally important findings can be correlated to known biochemical and
384 morphological features of each of the renal cancer subtypes. It is an established fact that the
385 number of mitochondria is increased dramatically in RO and ChRCC tumors (e.g. increased
386 oxidative phosphorylation) [40]. It is also known that ccRCC contains a highly vascularized
387 stroma (complement, coagulation, etc.) and exhibits a strong Warburg effect (glycolysis) [41].
388 Large intracellular vesicles are found in ChRCC (cytoplasmic proteins, vesicle proteins) [40].

389

390 *Classification*

391 Unsupervised data analysis demonstrated the presence of renal cancer subtype specific
392 differences in the tumor protein profiles. Next, we investigated the feasibility of tumor
393 classification by using the microproteomics data to train a prediction algorithm. We
394 implemented the tumor classification model by using a support vector machine (SVM)
395 approach. We chose the k-fold cross validation strategy [42] ("n-fold" in Perseus). Here the data
396 is randomly distributed in k groups. The model was then trained with data from k-1 groups
397 and the prediction was applied to the samples in the remaining group. This was repeated k
398 times. Low k-values tend to overestimate error rates. In our study 2-6 extraction spots
399 (samples) were derived from an FFPE section from each patient so too high k-values could
400 underestimate the true error rate. We therefore tested the prediction rate error over several k-
401 values (**Figure 7 A**) applying Radial Basis Function (RBF) and linear kernel functions [43]. The
402 tested error rates were in the range of 3.2% (4 wrong predictions) at the highest (k=3, linear
403 kernel) and 0 % at the lowest. However, k=3 is a very low k-value (excluding more than 41

404 samples from the training set). We argue that the error rate is most likely overestimated in this
405 case. For more commonly used k-values (k=5-10) the error rate was between 0.8% (1 wrong
406 prediction) and 0% for RBF and 1.6% (2 wrong prediction) and 0% for linear kernel function.
407 Incorrectly predicted outcomes included samples from one RO patient predicted as ccRCC.
408 Generally, RBF performed slightly better than linear kernel. Figure 7B exemplifies the
409 outcome of the cross-validation resulted for RBF and k=6 (around 20 samples per group
410 equivalent to 4 patients excluded from the training set). Each sample was scored for the three
411 tested conditions (ccRCC, RO, ChRCC). The highest scoring condition was used to classify a
412 given sample. Results are shown in a radar plot (**Figure 7B**) and demonstrate 100% accuracy
413 in prediction of renal cancer subtypes.

414 As mentioned previously, one renal tumor was initially incorrectly classified and was indeed a
415 ChRCC derived sarcomatoid renal tumor. The computational classification methods used
416 here cannot classify anything, which has not been previously "taught" to the system. In the
417 particular case of the sarcomatoid patient samples this means that they will be classified as
418 one of the three cancer types ccRCC, ChRCC or RO. Interestingly despite the similarity to
419 ccRCC (**Figure 5B**) the sarcomatoid patient samples were predicted by our SVM as ChRCC.
420 This is in accordance with their cancer type origin prior transformation.

421 We initially used all 346 differentially abundant histo-molecular features (proteins) to classify
422 the tumor subtypes. Next, we sought to estimate the minimum number of features that suffice
423 to correctly classify all the renal tumor samples (for k=6 and RBF). We used the feature
424 optimization function in the Perseus software, which first ranks the features and then tests the
425 error rate for a decreasing number of features (**Figure 7C**). The minimal number out of the

426 346 features was found to be 43 features (list of the ranked proteins can be found in
427 supplementary material 9). Further reducing the number to 21 features resulted in an error
428 rate of 1.6% and as little as 6 features lead to an error rate of 9.6%. Conclusively only a
429 portion of the dataset, i.e. at least 43 features would suffice to successfully classify all the
430 kidney tumor samples. However, keeping an excess of quantified protein features would be
431 beneficial as “safety margin” assuring a high enough number of quantified protein features for
432 robust classification of tumors.

433 *Data integration from MSI and rapid proteome profiling*

434 Having both MSI and microproteomics sets of data at hand provides several advantages for
435 classification of cancer tumor FFPE samples. Using the MSI approach for tumor classification
436 we observed a higher error rate than with the rapid micro-proteomics approach. In two cases
437 MSI could exclude one cancer type but was not providing clear results towards an RO or a
438 ChRCC diagnosis. Another case where one ccRCC sample was misclassified as RO is
439 particularly problematic as ccRCC might need surgery whereas RO does not. Therefore, by
440 integrating the micro-proteomics classification data the outcome of the MSI classification can
441 be further confirmed, clarified or rejected (**Table 1**). This allows more confidence in diagnosis
442 or could possibly even provide further information on cancer stage or treatment strategies. In
443 a case where the classification model does not cover the cancer condition such as in the
444 patient sample with sarcomatoid transformation we have demonstrated how irregularities and
445 inconsistencies are detected by both MSI and rapid LC-MS/MS based microproteomics
446 (**Figure 3I, 5A, 5B**). This provides an opportunity to further investigate, refine and expand the
447 range of computational and statistical classification models.

448 An additional application making use of the combined data set includes the investigation of
449 histo-molecular properties observed in MSI (e.g. intra- tumor heterogeneity) by correlation to
450 information from the rapid microproteomics approach. Usually a detailed investigation of MSI
451 feature data is achieved by either microproteomics "in situ protein digestion" or laser
452 microdissection based approaches [44] using LC-MS/MS based proteomics analysis with long
453 LC gradient times (1-4 hours). Despite the shorter gradient times and thus lower protein
454 coverage in the here presented approach the information can nevertheless be used to
455 investigate histo-molecular properties observed in MSI to a certain degree (e.g. intra- tumor
456 heterogeneity). We exemplified this in Figure 8, using the RO MSI data set previously shown
457 (Figure 1, top). Unsupervised clustering of MSI data revealed two distinct regions within the
458 tumor area (Figure 8: cluster 1 and cluster 2). Correlating the LC-MSMS data from the
459 respective extraction spots within these distinct regions in deed reveals significant differential
460 abundances of in 80 proteins (the protein list can be found in supplementary material 11).
461 Hierarchical clustering of these 80 proteins with regard to their extraction position correlates
462 well with the distinct regions depicted by the MSI clustering (MSI Cluster 2 correlates with
463 extraction spots E-F, MSI Cluster 2 correlates with extraction spots A-D; Figure 8). The
464 proteomics data suggests a lower abundance of mitochondrial associated proteins and a
465 higher abundance in some cytoskeletal protein binding proteins in cluster 2. The area
466 comprising Cluster 2 located on the edge of the tumor and might indicate the differences that
467 can be encountered between the inner and outer tumor regions [45].

468

469 **Discussion**

470 The increasing incidence of renal cancer in western countries calls for improved technologies
471 for detection, diagnosis, treatment and prognosis. Innovative mass spectrometry-based
472 applications are beginning to address challenges in clinics and the healthcare sector, such as
473 the use of targeted proteomics to characterize noninvasive liquid biopsies [46] or the so called
474 iKnife, enabling surgeons to identify cancerous tissue in real time [47,48]. Mass spectrometry is
475 becoming increasingly applicable in a clinical setting [49,50]. FFPE sections are a valuable
476 source for mass spectrometry-based diagnosis. As many of the sample preparation steps for
477 MS analysis overlap with the preparation steps for (immuno)histochemical staining, they can
478 be seamlessly fit into the high-throughput sample preparation pipeline for FFPE sections
479 (deparaffination, antigen retrieval) already existing in many hospitals. Distinguishing
480 numerous cancer types or subtypes and making decisions for treatment modalities are daily
481 challenges in hospitals. Approaches that include "digital" large-scale data acquisition and
482 computer-based machine learning algorithms provide deep molecular insight into the
483 respective disease and provides valuable information for early detection, diagnostic and
484 prognostic purposes.

485 Our proof of concept study demonstrates the potential and benefits of mass spectrometry
486 techniques for detailed characterization of clinical specimen. Specifically, we demonstrate
487 that mass spectrometry provides valuable results in the diagnosis of different renal cancer
488 subtypes (ccRCC, RO and ChRCC). The imaging mass spectrometry (MSI) approach allows
489 to collect spatially resolved spectra without *a priori* knowledge of the tissue, thereby enabling
490 the differentiation between cancerous and noncancerous tissue, as well as subtyping of
491 tumors. In our study MALDI-MSI could correctly diagnose 20 out of 23 of the tested patients.

492 In two of the 3 misclassified cases it was however possible to narrow down the diagnosis to
493 either RO or ChRCC. Despite the promising results the misclassification of a ccRCC sample
494 as RO might be problematic as RO may not require surgery but ccRCC does. Both cases
495 stress how using rapid proteome profiling data in parallel provides additional confidence and
496 can help avoid a false negative prognosis.

497 The MSI PCA data showed that the patient-to-patient tumor variability is significant for
498 ccRCC, necessitating a detailed histo-molecular profile for robust MSI performance. The
499 inclusion of many more patients (n>100) to increase the number of renal cancer tumor
500 samples will likely provide higher confidence and could resolve this issue.

501 Microproteomics analysis by in situ protein digestion in FFPE sections combined with
502 optimized and rapid LC-MS/MS based protein identification and quantification correctly
503 classified all the tested renal tumor samples in cross validation experiments. The efficient
504 peptide separation and sequencing capability of modern LC-MS/MS provided deeper insight
505 into the renal cancer proteome than possible by the MSI approach alone. The higher
506 dimensionality of the many protein features revealed by microproteomics improved the SVM
507 based prediction and achieved 100% accurate classification of renal cancer subtypes in cross
508 validation experiments.

509 Notably, unsupervised clustering identified data inconsistencies and irregularities in the
510 patient cohort. An unexpected feature pattern revealed a sarcomatoid transformation within
511 the ChRCC cohort, without *a priori* knowledge (**Figure 5A, 5B**). This goes to demonstrate that
512 once the “digital” data is acquired then the computational and statistical applications can

513 uncover relevant and important features of the patient datasets. This sensitivity, specificity
514 and versatility will have major implications for future clinical practices, including histo-
515 molecular pathology technologies.

516 The presented microproteomics method based on optimized, fast chromatographic separation
517 and fast MS/MS sequencing of peptides identified more than 2100 proteins in thin renal tumor
518 FFPE sections. Using short LC runs of only 15 min. we generated a list of 346 significantly
519 altered proteins (p=0.01). The minimum number of proteins determined to be necessary for
520 100% accurate tumor classification was 43. This low number of features enables a targeted
521 proteomics approach aimed at quantifying a select panel of proteins. Using fewer features
522 would also allow a further reduction of LC run time and increase overall sample throughput.
523 Using our fast LC-MS/MS setup we analyzed a total of 125 samples in a series without
524 experiencing blocking of the LC columns, glass capillaries or ESI needles. LC systems such
525 as the EvoSep system [25] that are specifically dedicated for clinical applications and tailored to
526 be used also by non LC-MS experts can add additional robustness to our approach.
527 Furthermore, implementation of image pattern recognition guided pipetting robots may
528 enhance reproducibility and throughput, e.g. using liquid extraction surface analysis (LESA)
529 technology [51,52]. The latter has been successfully applied in the study of traumatic brain
530 injuries [53] as well as in mouse brain for the identification of proteins and peptides from MSI
531 experiments [54]. The missing value problem is still a common problem in label free quantitative
532 proteomics. Successful implementation of protein identification on MS1 level only has been
533 presented recently [55] and could be interesting to following up in the here presented context in
534 future experiments.

535 Functional protein analysis using bioinformatics tools revealed molecular networks and
536 biochemical processes consistent with previously known macroscopic, morphological and
537 histological features of the renal cancer subtypes. Cancer-type specific proteome expression
538 features correlated to morphological characteristics of the respective cancer type. RO and
539 ChRCC exhibited upregulation of mitochondrial associated proteins. Indeed, increased
540 numbers of mitochondria are frequently observed in these cancer types by electron
541 microscopy [56] and have been identified in previous proteomics studies [57]. As most cancer rely
542 on glycolysis as major energy source (Warburg effect) this seems rather unusual. However,
543 those mitochondria are malfunctioning and it has been speculated that increase in number of
544 mitochondria is a cellular response to the presence of dysfunctional mitochondria [58].

545 In addition to mitochondria associated proteins increased intracytoplasmic associated
546 proteins were detected in ChRCC distinguishing it from the other cancer types.
547 Microscopically, ChRCC is distinguished from other renal carcinomas by its pale cytoplasm
548 resulting from large intracytoplasmic vesicles. This accounts for our detection of an increase
549 of intracellular cytoplasm-associated proteins and vesicle proteins, distinguishing ChRCC
550 from the other two renal tumor subtypes RO and ccRCC.

551 Clear cell renal cell carcinoma frequently contains zones of hemorrhage that are most likely
552 responsible for the increased levels of complement and coagulation cascade associated
553 proteins, as determined by our microproteomics method. ccRCC is also characterized by
554 hypervascular stroma [3], which may account for the enrichment of extracellular matrix proteins.
555 Again, enhanced glycolysis is a hallmark of many cancer types including ccRCC [41] correlating
556 well with our detection of upregulated glycolysis associated proteins by microproteomics.

557 For classification we applied PLS-DA to MSI data and support vector machine to the LC-
558 MSMS data. These common classification methods have previously been applied to MSI for
559 the differentiation of papillary and renal cell carcinoma based on lipidomics analysis [59] as well
560 as for the classification of epithelial ovarian cancer subtypes [16]. There are, however,
561 numerous other classification methods available. Mascini *et al.* used principal component
562 linear discriminant analysis in order to predict treatment response in xenograft models of
563 triple-negative breast cancer [60]. Recently, deep convolutional networks were proposed [61].
564 Both MSI and short gradient LC-MS/MS microproteomics methods come with their individual
565 advantages. Applying both approaches in parallel for routine analysis is most beneficial to
566 improve confidence in diagnosis and identify irregularities. In order to create very robust
567 classifiers for use in clinical settings the promising results of this study need to be further
568 supported in the future by analysis of larger patient cohorts.

569 With the enormous progress in sample handling and instrument technology, machine learning
570 [62] and the availability of new databases [63] mass spectrometry is on its way to become a
571 versatile tool in the hospital clinics of the future.

572

573 **Acknowledgements**

574 Proteomics and mass spectrometry research at SDU are supported by generous grants to the
575 VILLUM Center for Bioanalytical Sciences (VILLUM Foundation grant no. 7292 to O.N.J.) and
576 PRO-MS: Danish National Mass Spectrometry Platform for Functional Proteomics (grant no.

577 5072-00007B to O.N.J.). We thank Veit Schwämmle and Livia Rosa Fernandes for advice
578 and discussion on the project.

579 **Authors Contributions:**

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

583 **Competing Interests:**

584 The Authors declare no competing interest

585

586 ## References

587

588 1 Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide
589 for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394-424, doi:10.3322/caac.21492 (2018).
590 2 Jonasch, E., Gao, J. & Rathmell, W. K. Renal cell carcinoma. *BMJ* **349**, g4797, doi:10.1136/bmj.g4797
591 (2014).
592 3 Muglia, V. F. & Prando, A. Renal cell carcinoma: histological classification and correlation with imaging
593 findings. *Radiol Bras* **48**, 166-174, doi:10.1590/0100-3984.2013.1927 (2015).
594 4 Ng, K. L. *et al.* Differentiation of oncocytoma from chromophobe renal cell carcinoma (RCC): can novel
595 molecular biomarkers help solve an old problem? *J Clin Pathol* **67**, 97-104, doi:10.1136/jclinpath-2013-
596 201895 (2014).
597 5 Carella, R. *et al.* Immunohistochemical panels for differentiating epithelial malignant mesothelioma
598 from lung adenocarcinoma: a study with logistic regression analysis. *Am J Surg Pathol* **25**, 43-50 (2001).
599 6 Angelidis, I. *et al.* An atlas of the aging lung mapped by single cell transcriptomics and deep tissue
600 proteomics. *Nat Commun* **10**, 963, doi:10.1038/s41467-019-08831-9 (2019).
601 7 Clark, A. E., Kaleta, E. J., Arora, A. & Wolk, D. M. Matrix-assisted laser desorption ionization-time of
602 flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin*
603 *Microbiol Rev* **26**, 547-603, doi:10.1128/CMR.00072-12 (2013).
604 8 Kriegsmann, M. *et al.* Detection of HPV subtypes by mass spectrometry in FFPE tissue specimens: a
605 reliable tool for routine diagnostics. *J Clin Pathol* **70**, 417-423, doi:10.1136/jclinpath-2016-204017
606 (2017).
607 9 Ellis, S. R. *et al.* Automated, parallel mass spectrometry imaging and structural identification of lipids.
608 *Nat Methods* **15**, 515-518, doi:10.1038/s41592-018-0010-6 (2018).
609 10 Hannes Hinneburg, P. K., Falko Schirmeister, Slavko Gasparov, Peter & H. Seeberger, V. Z., Daniel
610 Kolarich. Spatial glycomics of histopathological formalin-fixed and paraffin embedded (FFPE) tissue
611 microdissections. *Molecular & Cell Proteomics* **in review** (2016).
612 11 Moginger, U. *et al.* Alterations of the Human Skin N- and O-Glycome in Basal Cell Carcinoma and
613 Squamous Cell Carcinoma. *Front Oncol* **8**, 70, doi:10.3389/fonc.2018.00070 (2018).
614 12 Caprioli, R. M., Farmer, T. B. & Gile, J. Molecular imaging of biological samples: localization of peptides
615 and proteins using MALDI-TOF MS. *Anal Chem* **69**, 4751-4760, doi:10.1021/ac970888i (1997).
616 13 Stoeckli, M., Farmer, T. B. & Caprioli, R. M. Automated mass spectrometry imaging with a matrix-
617 assisted laser desorption ionization time-of-flight instrument. *J Am Soc Mass Spectrom* **10**, 67-71,
618 doi:10.1016/S1044-0305(98)00126-3 (1999).
619 14 Casadonte, R. *et al.* Development of a Class Prediction Model to Discriminate Pancreatic Ductal
620 Adenocarcinoma from Pancreatic Neuroendocrine Tumor by MALDI Mass Spectrometry Imaging.
621 *Proteomics Clin Appl* **13**, e1800046, doi:10.1002/prca.201800046 (2019).
622 15 Mascini, N. E., Teunissen, J., Noorlag, R., Willems, S. M. & Heeren, R. M. A. Tumor classification with
623 MALDI-MSI data of tissue microarrays: A case study. *Methods* **151**, 21-27,
624 doi:10.1016/j.ymeth.2018.04.004 (2018).
625 16 Klein, O. *et al.* MALDI-Imaging for Classification of Epithelial Ovarian Cancer Histotypes from a Tissue
626 Microarray Using Machine Learning Methods. *Proteomics Clin Appl* **13**, e1700181,
627 doi:10.1002/prca.201700181 (2019).
628 17 Mallah, K. *et al.* Lipid Changes Associated with Traumatic Brain Injury Revealed by 3D MALDI-MSI. *Anal*
629 *Chem* **90**, 10568-10576, doi:10.1021/acs.analchem.8b02682 (2018).

18 Briggs, M. T. *et al.* MALDI mass spectrometry imaging of N-glycans on tibial cartilage and subchondral bone proteins in knee osteoarthritis. *Proteomics* **16**, 1736-1741, doi:10.1002/pmic.201500461 (2016).
19 Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567, doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2 (1999).
20 Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976-989, doi:10.1016/1044-0305(94)80016-2 (1994).
21 Zhao, M. *et al.* A Comparative Proteomics Analysis of Five Body Fluids: Plasma, Urine, Cerebrospinal Fluid, Amniotic Fluid, and Saliva. *Proteomics Clin Appl* **12**, e1800008, doi:10.1002/prca.201800008 (2018).
22 Schmidt, A. & Aebersold, R. High-accuracy proteome maps of human body fluids. *Genome Biol* **7**, 242, doi:10.1186/gb-2006-7-11-242 (2006).
23 Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu Rev Anal Chem (Palo Alto Calif)* **8**, 61-80, doi:10.1146/annurev-anchem-071114-040325 (2015).
24 Falkenby, L. G. *et al.* 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* **13**, 6169-6175, doi:10.1021/pr5008575 (2014).
25 Krieger, J. R. *et al.* Evosep One Enables Robust Deep Proteome Coverage Using Tandem Mass Tags while Significantly Reducing Instrument Time. *J Proteome Res* **18**, 2346-2353, doi:10.1021/acs.jproteome.9b00082 (2019).
26 Horning, O. B., Kjeldsen, F., Theodorsen, S., Vorm, O. & Jensen, O. N. Isocratic solid phase extraction-liquid chromatography (SPE-LC) interfaced to high-performance tandem mass spectrometry for rapid protein identification. *Journal of proteome research* **7**, 3159-3167, doi:10.1021/pr700865c (2008).
27 Hørning O. B. , T. S., 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* **268,**, 147-157, doi:10.1016/j.ijms.2007.06.017. (2007).
28 Stoeckli, M., Staab, D., Wetzel, M. & Brechbuehl, M. iMatrixSpray: a free and open source sample preparation device for mass spectrometric imaging. *Chimia (Aarau)* **68**, 146-149, doi:10.2533/chimia.2014.146 (2014).
29 Kovalchuk, S. I., Jensen, O. N. & Rogowska-Wrzesinska, A. FlashPack: Fast and Simple Preparation of Ultrahigh-performance Capillary Columns for LC-MS. *Molecular & cellular proteomics : MCP* **18**, 383-390, doi:10.1074/mcp.TIR118.000953 (2019).
30 Rompp, A. *et al.* imzML: Imaging Mass Spectrometry Markup Language: A common data format for mass spectrometry imaging. *Methods in molecular biology* **696**, 205-224, doi:10.1007/978-1-60761-987-1_12 (2011).
31 Bemis, K. D. *et al.* Cardinal: an R package for statistical analysis of mass spectrometry-based imaging experiments. *Bioinformatics* **31**, 2418-2420, doi:10.1093/bioinformatics/btv146 (2015).
32 Bemis, K. D. *et al.* Probabilistic Segmentation of Mass Spectrometry (MS) Images Helps Select Important Ions and Characterize Confidence in the Resulting Segments. *Molecular & cellular proteomics : MCP* **15**, 1761-1772, doi:10.1074/mcp.O115.053918 (2016).
33 Wold, S., Sjostrom, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab* **58**, 109-130, doi:Doi 10.1016/S0169-7439(01)00155-1 (2001).
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* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).

676 35 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data.
677 *Nat Methods* **13**, 731-740, doi:10.1038/nmeth.3901 (2016).
678 36 Metsalu, T. & Vilo, J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal
679 Component Analysis and heatmap. *Nucleic acids research* **43**, W566-570, doi:10.1093/nar/gkv468
680 (2015).
681 37 Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage,
682 supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* **47**,
683 D607-D613, doi:10.1093/nar/gky1131 (2019).
684 38 Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function.
685 *Genome Res* **13**, 2129-2141, doi:10.1101/gr.772403 (2003).
686 39 Wisztorski, M. *et al.* Microproteomics by liquid extraction surface analysis: application to FFPE tissue to
687 study the fimbria region of tubo-ovarian cancer. *Proteomics Clin Appl* **7**, 234-240,
688 doi:10.1002/prca.201200070 (2013).
689 40 **John. Eble, L. C.** in *Modern Surgical Pathology (Second Edition)* (ed Richard J. Cote Noel Weidner, Saul
690 Suster, Lawrence M. Weiss) 1043-1078 (W.B. Saunders, 2009).
691 41 Courtney, K. D. *et al.* Isotope Tracing of Human Clear Cell Renal Cell Carcinomas Demonstrates
692 Suppressed Glucose Oxidation In Vivo. *Cell Metab* **28**, 793-800 e792, doi:10.1016/j.cmet.2018.07.020
693 (2018).
694 42 Hastie, T., Tibshirani, R. & Friedman, J. in *The Elements of Statistical Learning: Data Mining, Inference,*
695 *and Prediction* 219-259 (Springer New York, 2009).
696 43 Hastie, T., Tibshirani, R. & Friedman, J. in *The Elements of Statistical Learning: Data Mining, Inference,*
697 *and Prediction* 191-218 (Springer New York, 2009).
698 44 Dewez, F. *et al.* Precise co-registration of mass spectrometry imaging, histology, and laser
699 microdissection-based omics. *Analytical and bioanalytical chemistry* **411**, 5647-5653,
700 doi:10.1007/s00216-019-01983-z (2019).
701 45 Oppenheimer, S. R., Mi, D., Sanders, M. E. & Caprioli, R. M. Molecular analysis of tumor margins by
702 MALDI mass spectrometry in renal carcinoma. *Journal of proteome research* **9**, 2182-2190,
703 doi:10.1021/pr900936z (2010).
704 46 Kim, Y. *et al.* Targeted proteomics identifies liquid-biopsy signatures for extracapsular prostate cancer.
705 *Nature Communications* **7**, doi:ARTN 11906

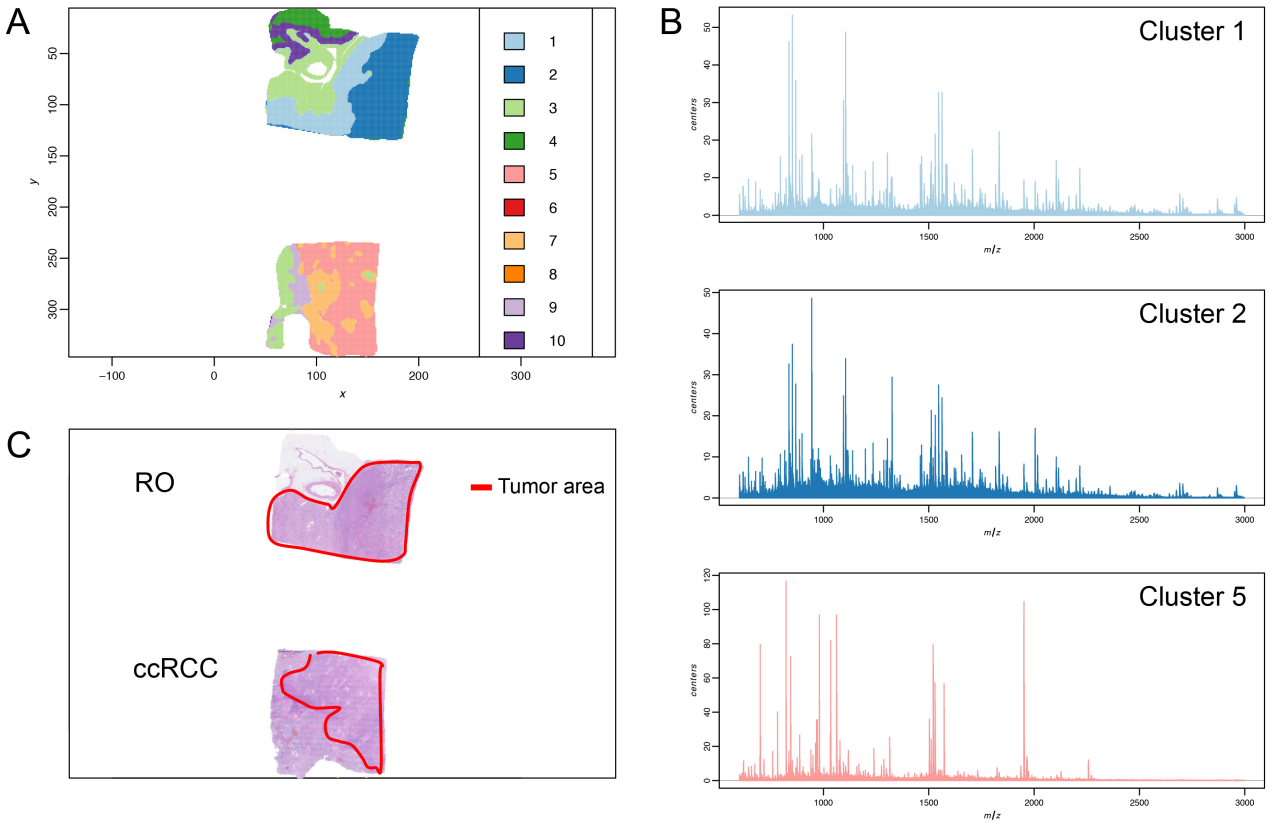
706 10.1038/ncomms11906 (2016).
707 47 Phelps, D. L. *et al.* The surgical intelligent knife distinguishes normal, borderline and malignant
708 gynaecological tissues using rapid evaporative ionisation mass spectrometry (REIMS). *Br J Cancer* **118**,
709 1349-1358, doi:10.1038/s41416-018-0048-3 (2018).
710 48 Schafer, K. C. *et al.* In vivo, in situ tissue analysis using rapid evaporative ionization mass spectrometry.
711 *Angew Chem Int Ed Engl* **48**, 8240-8242, doi:10.1002/anie.200902546 (2009).
712 49 Heaney, L. M., Jones, D. J. & Suzuki, T. Mass spectrometry in medicine: a technology for the future?
713 *Future Sci OA* **3**, FSO213, doi:10.4155/fsoa-2017-0053 (2017).
714 50 Longuespee, R., Casadonte, R., Schwamborn, K. & Kriegsmann, M. Proteomics in Pathology: The Special
715 Issue. *Proteomics Clin Appl* **13**, e1800167, doi:10.1002/prca.201800167 (2019).
716 51 Wisztorski, M. *et al.* Spatially-resolved protein surface microsampling from tissue sections using liquid
717 extraction surface analysis. *Proteomics* **16**, 1622-1632, doi:10.1002/pmic.201500508 (2016).
718 52 Wisztorski, M. *et al.* Droplet-Based Liquid Extraction for Spatially-Resolved Microproteomics Analysis of
719 Tissue Sections. *Methods Mol Biol* **1618**, 49-63, doi:10.1007/978-1-4939-7051-3_6 (2017).

720 53 Mallah, K. *et al.* Mapping Spatiotemporal Microproteomics Landscape in Experimental Model of
721 Traumatic Brain Injury Unveils a link to Parkinson's Disease. *Mol Cell Proteomics* **18**, 1669-1682,
722 doi:10.1074/mcp.RA119.001604 (2019).
723 54 Ryan, D. J. *et al.* Protein identification in imaging mass spectrometry through spatially targeted liquid
724 micro-extractions. *Rapid communications in mass spectrometry : RCM* **32**, 442-450,
725 doi:10.1002/rcm.8042 (2018).
726 55 Ivanov, M. V. *et al.* DirectMS1: MS/MS-Free Identification of 1000 Proteins of Cellular Proteomes in 5
727 Minutes. *Analytical chemistry* **92**, 4326-4333, doi:10.1021/acs.analchem.9b05095 (2020).
728 56 Thoenes, W., Storkel, S. & Rumpelt, H. J. Human chromophobe cell renal carcinoma. *Virchows Arch B*
729 *Cell Pathol Incl Mol Pathol* **48**, 207-217, doi:10.1007/bf02890129 (1985).
730 57 Drendel, V. *et al.* Proteomic distinction of renal oncocytomas and chromophobe renal cell carcinomas.
731 *Clin Proteomics* **15**, 25, doi:10.1186/s12014-018-9200-6 (2018).
732 58 Gasparre, G., Romeo, G., Rugolo, M. & Porcelli, A. M. Learning from oncocytic tumors: Why choose
733 inefficient mitochondria? *Biochimica et biophysica acta* **1807**, 633-642,
734 doi:10.1016/j.bbabio.2010.08.006 (2011).
735 59 Dill, A. L. *et al.* Multivariate statistical differentiation of renal cell carcinomas based on lipidomic
736 analysis by ambient ionization imaging mass spectrometry. *Anal Bioanal Chem* **398**, 2969-2978,
737 doi:10.1007/s00216-010-4259-6 (2010).
738 60 Mascini, N. E. *et al.* The use of mass spectrometry imaging to predict treatment response of patient-
739 derived xenograft models of triple-negative breast cancer. *Journal of proteome research* **14**, 1069-
740 1075, doi:10.1021/pr501067z (2015).
741 61 Behrmann, J. *et al.* Deep learning for tumor classification in imaging mass spectrometry. *Bioinformatics*
742 **34**, 1215-1223, doi:10.1093/bioinformatics/btx724 (2018).
743 62 Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **349**, 255-
744 260, doi:10.1126/science.aaa8415 (2015).
745 63 Gessulat, S. *et al.* Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning.
746 *Nat Methods* **16**, 509-518, doi:10.1038/s41592-019-0426-7 (2019).

747

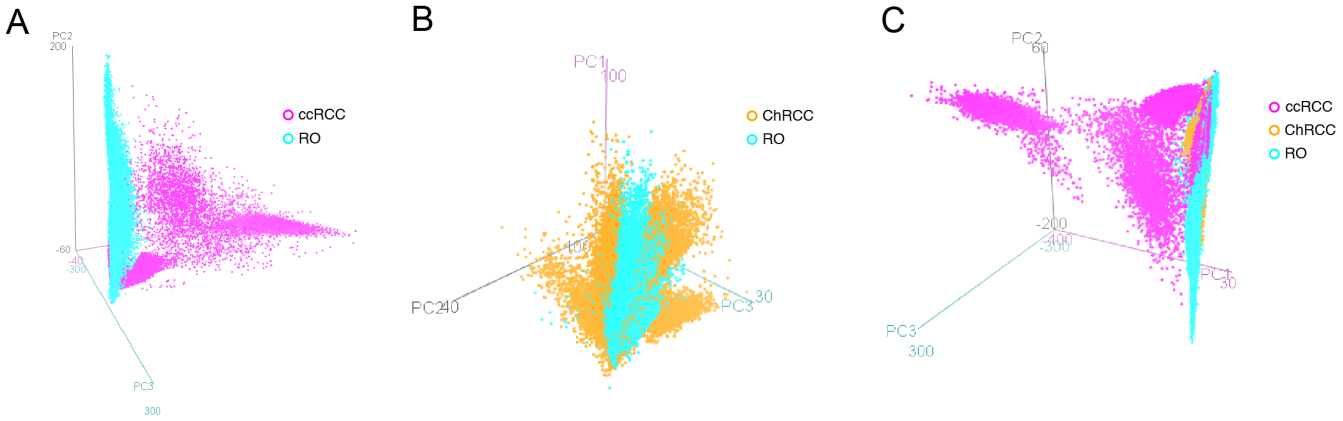
748

749 **Figure 1**



750

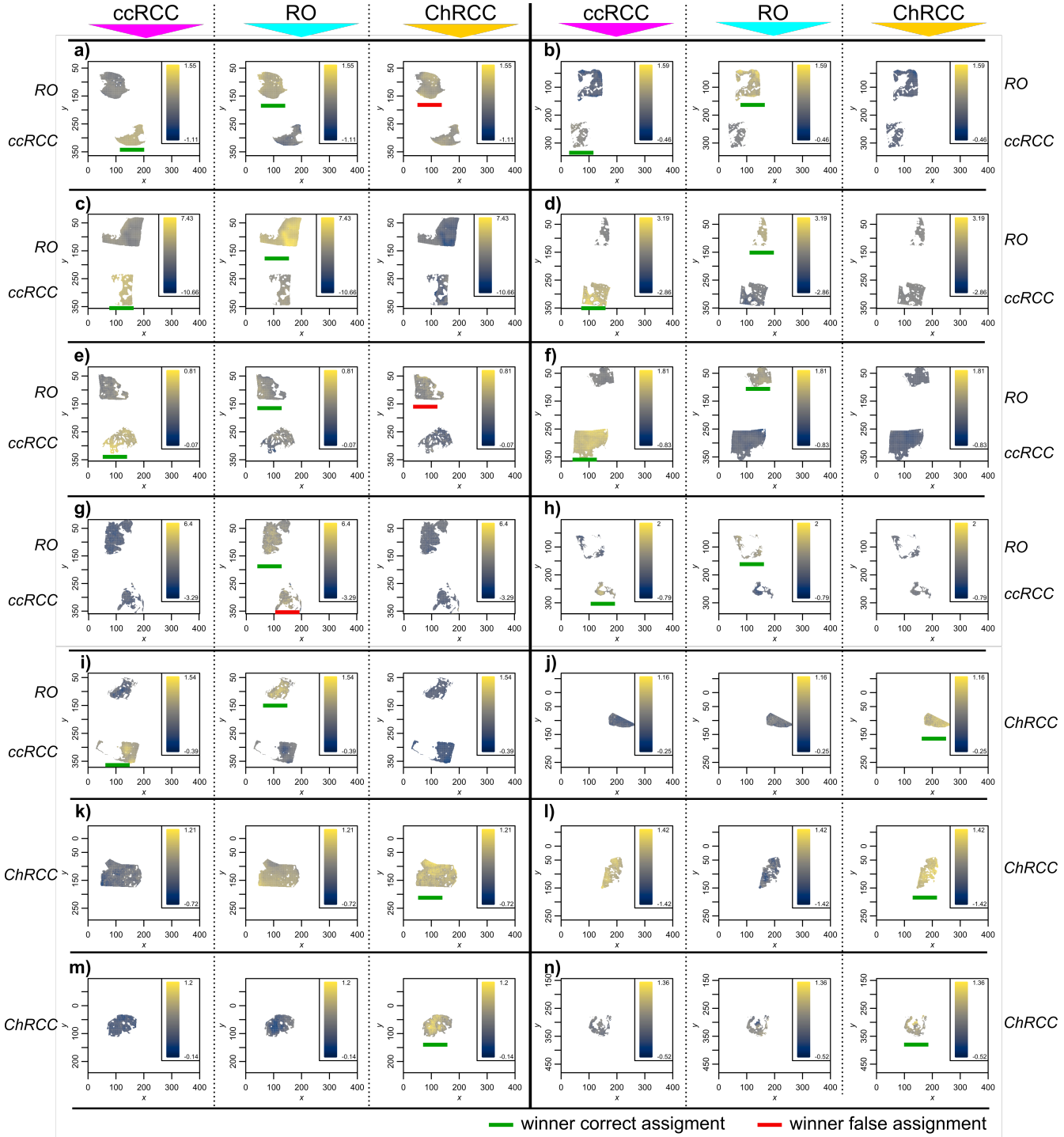
751 **Figure 2**



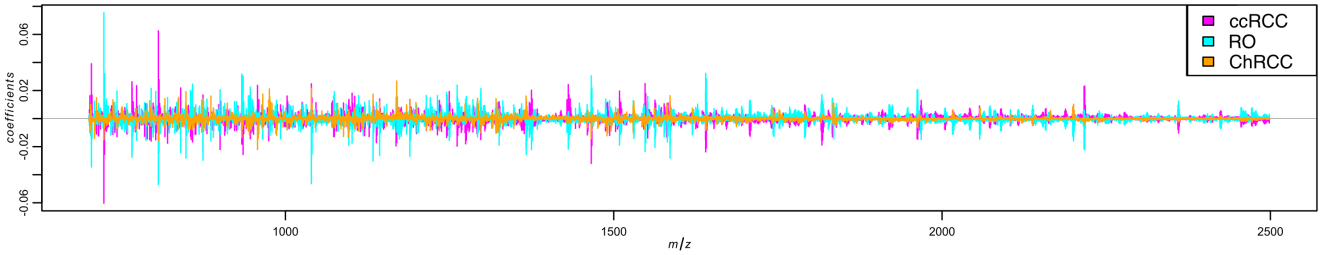
752

753

Figure 3


ccRCC
RO
ChRCC
ccRCC
RO
ChRCC
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
m)
n)
RO
ccRCC
ChRCC
x
y
winner correct assigment
winner false assignment

757 **Figure 4:**

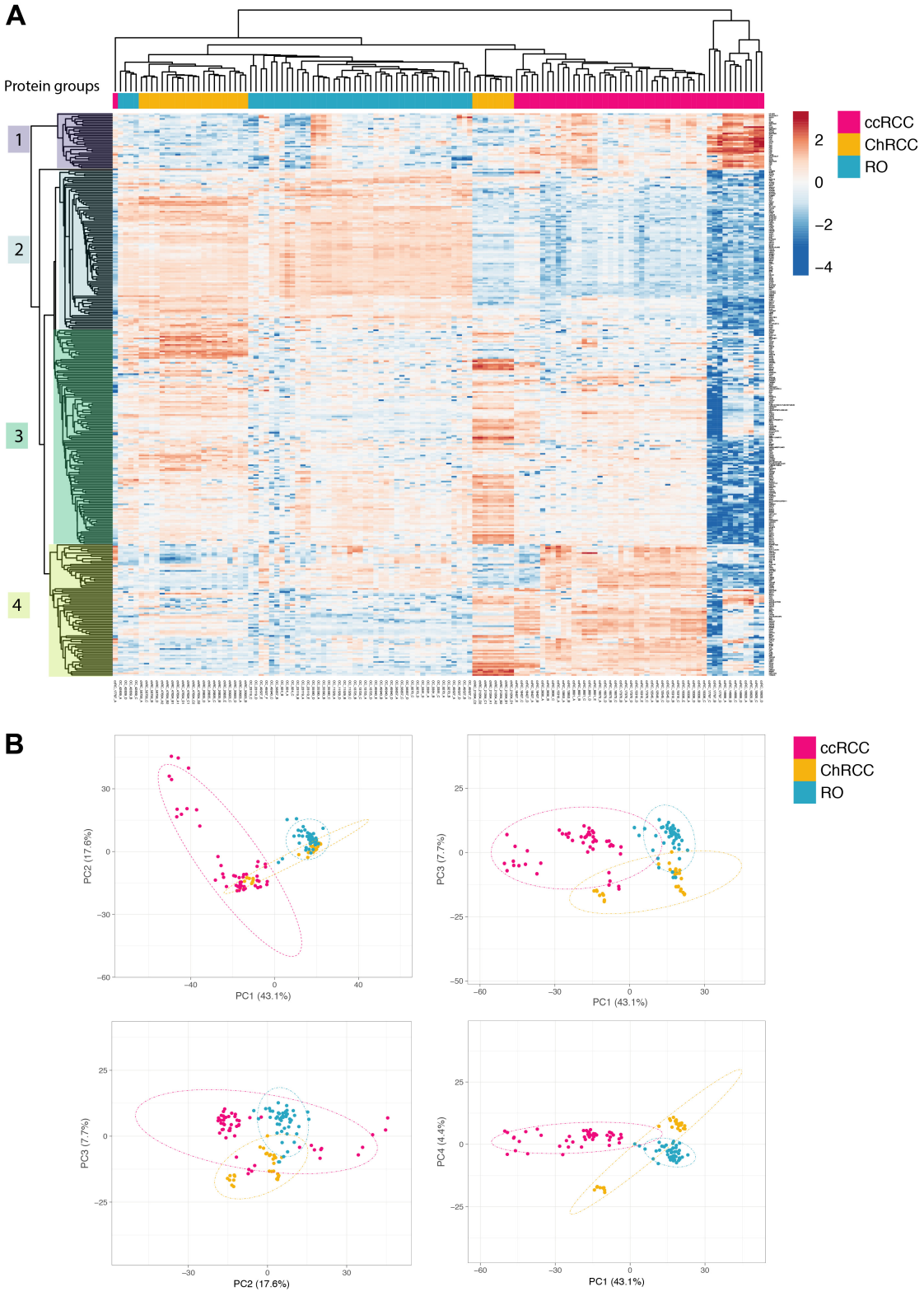


758

759

760

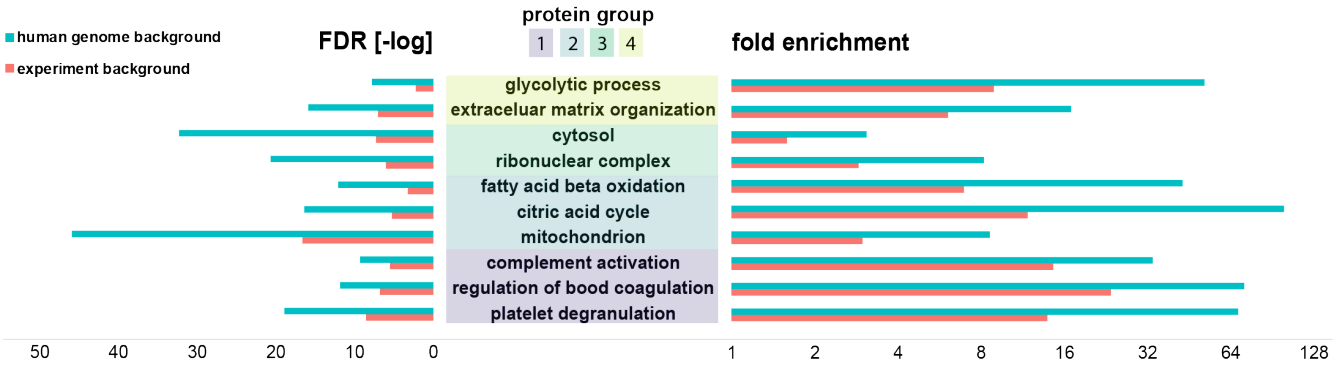
761 **Figure 5**



762

763 **Figure 6**

764

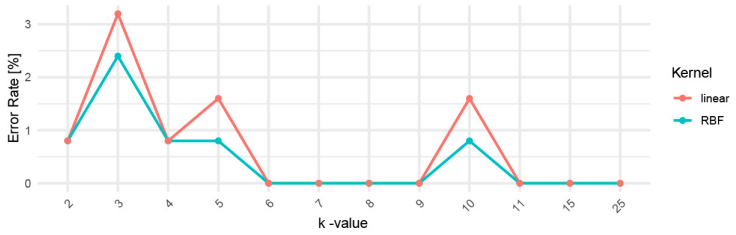


765

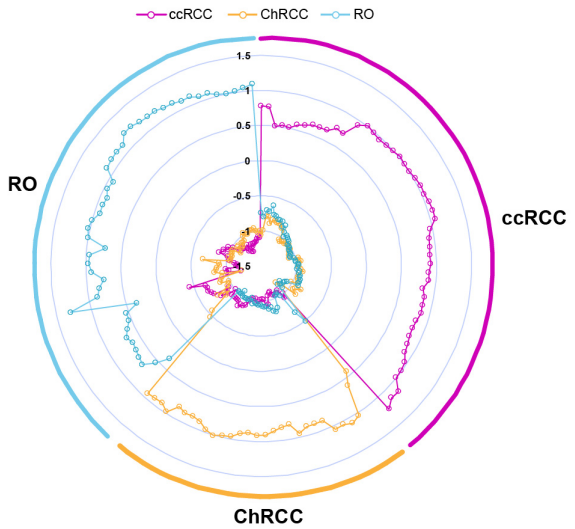
766

Figure 7

A

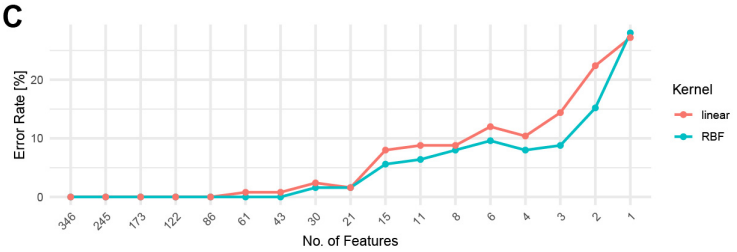


B



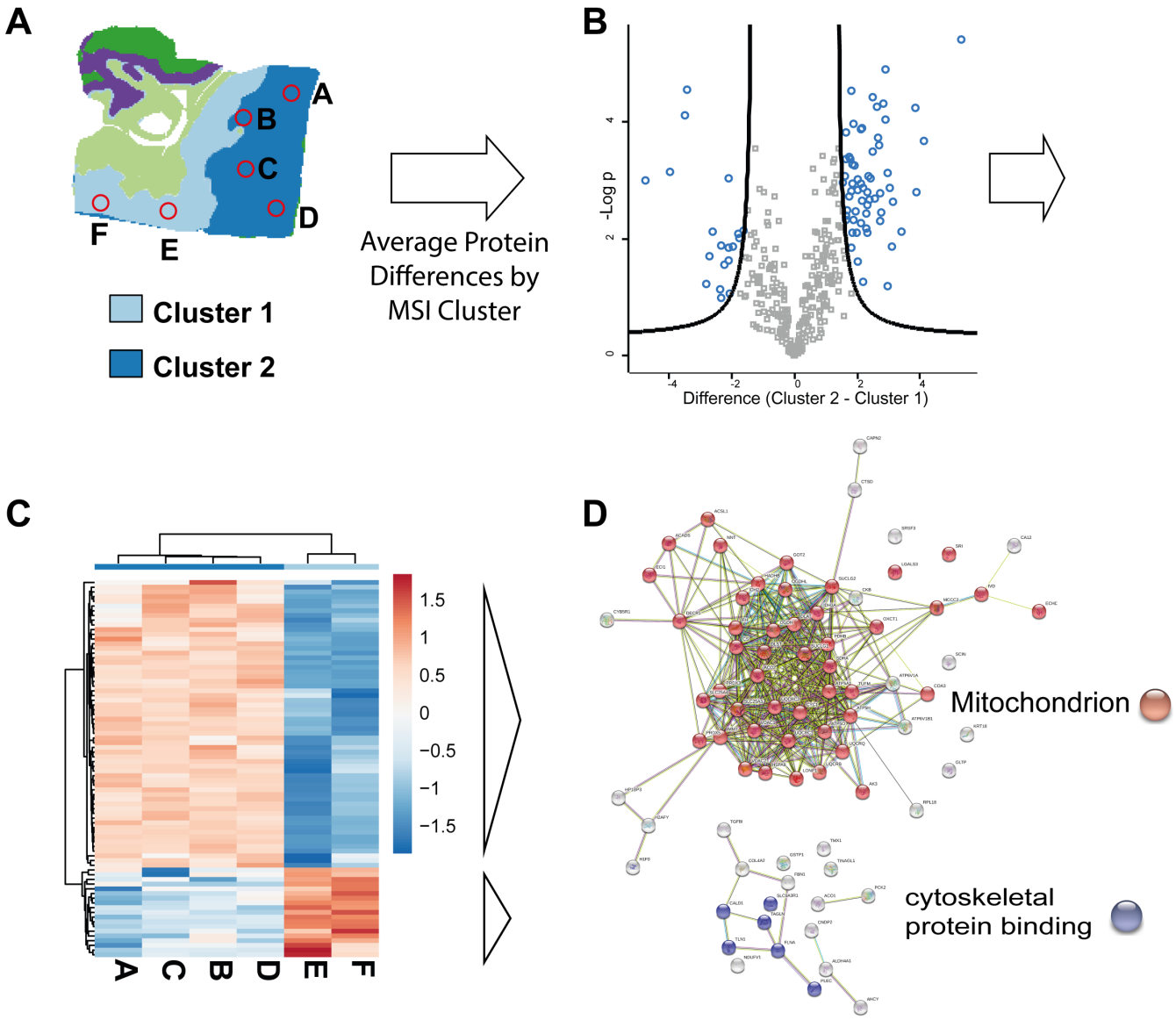
Predicted: ccRCC	Predicted: ChRCC	Predicted: RO	T: True group
49	0	0	True: ccRCC
0	29	0	True: ChRCC
0	0	47	True: RO

C



770 **Figure 8**

771



772

773

774

775

776 **Table 1**

Pathologist diagnosis	Patient	MSI diagnosis	Rapid LC-MSMS diagnosis	Conclusion
RO	42839	RO/ChRCC	RO	RO
RO	23119	RO	RO	RO
RO	49527	RO	RO	RO
RO	9270	RO	RO	RO
RO	11529	RO/ChRCC	RO	RO
RO	55560	RO	RO	RO
RO	49940	RO	RO	RO
RO	56857	RO	RO	RO
RO	3381	RO	RO	RO
ccRCC	18427	ccRCC	ccRCC	ccRCC
ccRCC	17370	ccRCC	ccRCC	ccRCC
ccRCC	10620	ccRCC	ccRCC	ccRCC
ccRCC	16073	ccRCC	ccRCC	ccRCC
ccRCC	12545	ccRCC	ccRCC	ccRCC
ccRCC	14999	ccRCC	ccRCC	ccRCC
ccRCC	17797	RO	ccRCC	ccRCC/further validation
ccRCC	8601	ccRCC	ccRCC	ccRCC
ccRCC	10336	ccRCC	ccRCC	ccRCC
ChRCC	47638	ChRCC	ChRCC	ChRCC
ChRCC	6835	ChRCC	ChRCC	ChRCC
ChRCC	21264*	ChRCC with irregularities	ChRCC with irregularities	further validation -> sarcomatoid transformation
ChRCC	58756	ChRCC	ChRCC	ChRCC
ChRCC	39925	ChRCC	ChRCC	ChRCC

777

778 **Table 1**: integrated testing strategy for classification of renal cancer types. Initial pathologist
779 diagnosis and patient number are indicated in the first 2 columns. *Patient sample showed
780 irregularities and after reassessment could be diagnosed as sarcomatoid transformation.
781 Concluding contradictory results would either necessitate further validation or the outcome of
782 the more reliable method (LC-MSMS) could be favored

783

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

792

793 **Figure 2:** 3D PCA score plot from imaging MALDI MS experiments of kidney tumor tissues.
794 Each plot contains the extracted pixel data from all patients of a given cancer type. Data from
795 ccRCC (A) (magenta) and ChRCC (B) (blue) are compared to RO (yellow). (C) Data from all
796 three cancer types are compared to each other. The graph displays the first 3 principle
797 components (PC1, PC2, PC3) plotted against each other. Clear separation of data points
798 between ccRCC and RO can be observed by pairwise comparison but also in the combined

799 comparison to both RO and ChRCC. In pairwise comparison (B) RO and ChRCC show slight
800 separation but exhibit a great number of overlapping features. ccRCC exhibits the largest
801 differences to RO and ChRCC. RO and ChRCC appear to share more spectral similarities.

802

803 **Figure 3:** Tumor classification by MALDI MS imaging and cross-validation using PLS-DA
804 classification. Pixels/spectra from tumor areas only, were extracted for processing.
805 Classification of 9 ccRCC and 9 RO (a-i) sample as well as 5 RO sample and 5 ChRCC
806 sample (j-n). Pathology diagnosis of the respective patient samples are indicated to the left
807 and right of the images (*RO, ccRCC* or *ChRCC*). Each spectrum-containing pixel is predicted
808 individually. The prediction scores are represented by a color scale. Each cancer condition
809 was tested and scored so that each sample set is represented in 3 panels. Each of the panels
810 displays scores for ccRCC (first panel) RO (second panel) and ChRCC (third panel) The
811 respective testing condition is indicated on top above the panels. (Cardinal´s smooth.image-
812 function was used for better visibility. Unprocessed image can be found in supplementary
813 material 1 Figure S4)

814 Each sample is predominantly predicted in the correct diagnosis, achieving accuracies (pixel-
815 based value) of 93% (ccRCC), 88% (RO), 88% (ChRCC). Winner of the classification is
816 marked with a green bar for correct classification and a red bar for incorrect classification.

817

818 **Figure 4:** PLS coefficients as a function of *m/z*. The diagram displays the impact of each
819 detected *m/z* signal feature in imaging MS data from PLS-DA prediction (spectra are binned

820 to 0.25 m/z bins). Positive coefficient indicates presence or higher abundance in the
821 respective condition. Negative coefficient indicates absence or lower abundance of the m/z
822 value in the respective condition (a list with the 100 most influential features can be found in
823 supplementary material 2)

824

825 Figure 5: Unsupervised renal cancer subtype classification by microproteomics using rapid
826 LC-MS/MS protein profiling. A) Heatmap and hierarchical clustering of differential relative
827 protein abundances. Columns indicate samples and rows indicate proteins. The renal cancer
828 subtype of the patient sample is indicated in colored bars on top. The graph shows the large
829 similarities in protein expression profiles among patient samples with the same cancer
830 subtype causing them to cluster together. Furthermore, hierarchical clustering of the protein
831 abundances reveals protein cluster that are detected in a cancer subtype specific manner.
832 Protein groups selected for subsequent network analysis are indicated by color blocks on the
833 y-axis dendrogram (groups 1-4). B) Principal component analysis of the sample set. Dotted
834 ellipses are such that with a probability of 95% a new observation from the same group will
835 fall inside the area. The first (PC1) and second (PC2) component explain 17.6 % of the total
836 variance whereas the other components lie at 7.7% and 4.4% respectively. There is a clear
837 separation of ccRCC and RO samples already in the first two principal components.
838 Differences between RO and ChRCC are subtle and are only evident when considering
839 components that display lower variance (PC2:PC3 and PC3:PC4). The small group of the
840 eight ChRCC-derived sarcomatoid renal cancers samples cluster relatively far from the other
841 ChRCC samples, thereby identifying these as clear “outliers” that require further attention.

842 **Figure 6:** Bioinformatics analysis (PantherDB) identified enriched biochemical functions in
843 renal tumors. Protein groups were compared against a background of all the 2124 proteins
844 identified in the experiment (blue) and against the background of the human genome (red).
845 Fold enrichment (increase over expected value) as well as -log of the false discovery rate
846 (FDR) are shown.

847

848 **Figure 7:** Microproteomics and SVM model correctly classifies all renal tumor subtypes. A)
849 Dependency of the classification error rate in relation to increasing k-value. RBF performs
850 slightly better than the linear kernel function. From value k=4 and higher, error rates vary
851 between 0% and 0.8% (1 wrong prediction out of 125) for RBF and between 0% - 1.6 % for
852 linear kernel. B) Radar plot of the cross-validated classification (k=6, kernel=RBF) of
853 proteome profiles obtained from each tumor section extraction spot. Scores for each of the
854 three tumor types are displayed. Scores range from lowest (center) to highest (outer circle).
855 The highest score indicates highest likelihood for the respective diagnosis. (Scores for
856 ccRCC: magenta, ChRCC: yellow, RO: blue). The pathological diagnosis for each sample is
857 indicated on the outside of the radar plot. The plot shows that all samples score highest in
858 correlation with the respective cancer type indicating the high accuracy of the classification.
859 C) Feature optimization. The error rate for linear kernel and RBF are plotted over the number
860 of ranked features (proteins). Decreasing feature number results in increase of false
861 predictions. Minimum number of features for 0% error rate is at 43 using RBF and 86 for
862 linear kernel (list of ranked proteins can be found in supplementary material 10).

863

864 **Figure 8:** Combined use of MS imaging and rapid LC-MSMS microproteomics provides histo-
865 molecular details of tumor heterogeneity. A) MSI based unsupervised clustering analysis of a
866 RO patient sample. Two clusters (cluster 1 and cluster 2) are detected within the tumor area.
867 Positions used for extraction of LC-MSMS samples are indicated by red circles extractions A-
868 F (2 extractions for cluster 1, 4 extractions in cluster 2).

869 B) Volcano plot of LC-MSMS data derived ratio of protein abundance (Cluster 2 / Cluster1).
870 Proteins with significant (t-test: p=0.01, 2-fold difference) different abundances are colored in
871 blue.

872 C) Heatmap display of the significantly different proteins from extraction spots A-F. On the x-
873 axis extraction spots A-D and E-F group together. The grouping is in correlation to the MSI
874 clustering data. On the y axis two protein groups can be observed distinguishing the two x-
875 axis-cluster. One group is upregulated in Cluster 1 the other group is upregulated in Cluster 2.

876 D) StringDB network analysis of upregulated proteins in MSI-Cluster-2 (top, sample: A-D) and
877 upregulated proteins in MSI-Cluster 1 (bottom, sample: E-F). Cluster 2 shows higher
878 abundance of mitochondrial associated proteins whereas cluster 1 shows increase
879 abundance of cytoskeletal protein binding proteins.

880