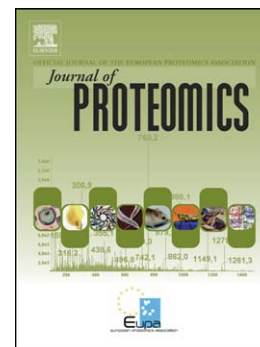# Accepted Manuscript

Alterations of the serum peptidome in renal cell carcinoma discriminating benign and malignant kidney tumors

Erica Gianazza, Clizia Chinello, Veronica Mainini, Marta Cazzaniga, Valeria Squeo, Giancarlo Albo, Stefano Signorini, Salvatore S. Di Pierro, Stefano Ferrero, Simone Nicolardi, Yuri E.M. van der Burgt, André M. Deelder, Fulvio Magni

Please cite this article as: Gianazza Erica, Chinello Clizia, Mainini Veronica, Cazzaniga Marta, Squeo Valeria, Albo Giancarlo, Signorini Stefano, Di Pierro Salvatore S., Ferrero Stefano, Nicolardi Simone, van der Burgt Yuri E.M., Deelder André M., Magni Fulvio, Alterations of the serum peptidome in renal cell carcinoma discriminating benign and malignant kidney tumors, *Journal of Proteomics* (2012), doi: 10.1016/j.jprot.2012.07.032

Alterations of the serum peptidome in renal cell carcinoma discriminating benign and malignant kidney tumors

[a*]Erica Gianazza, [a*]Clizia Chinello, [a]Veronica Mainini, [a]Marta Cazzaniga, [a]Valeria Squeo, [b]Giancarlo Albo, [c]Stefano Signorini, [b]Salvatore S. Di Pierro, [d]Stefano Ferrero, [e]Simone Nicolardi, [e]Yuri E.M. van der Burgt, [e]André M. Deelder and [a]Fulvio Magni.

[a] Department of Experimental Medicine, University of Milano-Bicocca, Monza, Italy

[b] Department of Specialistic Surgical Sciences, Urology unit, Ospedale Maggiore Policlinico Foundation, Milano, Italy

[c] Department of Laboratory Medicine, Hospital of Desio, Desio, Italy

[d] Department of Medicine, Surgery and Dental Sciences, Pathology Unit, IRCCS-Policlinico Foundation, Mangiagalli and Regina Elena, University of Milan, Milano, Italy

[e] Department of Parasitology, Biomolecular Mass Spectrometry Unit, Leiden University Medical Center, Leiden, The Netherlands

[*] Equally contributing authors

Corresponding author: Fulvio Magni

Department of Experimental Medicine (DIMS)

University of Milano-Bicocca

Via Cadore, 48

20900 Monza, Italy

Phone: +39-02-64488213

Fax: +39-02-64488252

E-mail: Fulvio.Magni@unimib.it

Abstract

Renal Cell Carcinoma (RCC) is typically asymptomatic and surgery usually increases patient's life only for early stage tumors. However, some cystic and solid renal lesions cannot be confidently differentiated from clear-cell-RCC. Therefore possible markers for early detection and to distinguish malignant kidney tumors are needed. To this aim, we applied MALDI-TOF and LC-MS/MS analysis to RPC18 MB purified serum of ccRCC, non-ccRCC patients and controls. A cluster of five signals differentiate malignant tumors from benign renal masses and healthy subjects. Moreover, a combination of six ions showed the highest specificity and sensitivity to distinguish ccRCC from controls. Healthy subjects were also differentiated from non-ccRCC by three features. Peptide ratios obtained by MALDI-TOF were compared with those from label-free LC-ESI and no statistical difference was found ($p>0.05$). ESI-results were linked with MALDI profiles by both TOF/TOF sequencing and MALDI FT-ICR accurate mass measurements. About 200 unique endogenous peptides, originating from 32 proteins, were identified. Among them, SDPR and ZYX were found down-expressed, while SRGN and TMSL3 were up-expressed. In conclusion, our results suggest the possibility to discriminate malignant kidney tumors based on a cluster of serum peptides. Moreover, label-free may represent a valid method to verify results obtained by MALDI-TOF.

1.Introduction

Renal cell carcinoma (RCC) is reported as the most common form of kidney cancer, accounting for about 3-6 % of all human cancers and for more than 100,000 deaths/year [1]. The majority of RCC cases present histologically as clear-cell (ccRCC), while the rest of the subtypes are the papillary, chromophobe, oncocytoma, collecting duct, medullary, or unclassified cell type [2].

RCC is typically asymptomatic and commonly diagnosed incidentally either by computed-tomography (CT) scans or when patients reach the hospital, with symptoms that most generally include blood in the urine and/or abdominal pain [3, 4]. Even in well developed countries the detection of renal cell carcinoma in its early stage is increased as a consequence of the wider use of imaging procedures. However, the nature of some cystic and solid renal lesions cannot be confidently diagnosed by imaging techniques and differentiated from RCC.

The current gold standard for the RCC treatment is the surgical tumor mass removal, either partial or radical nephrectomy. Surgery usually increases the duration of patient's life only for early stage tumors. However prognosis of patients with advanced stage or metastatic disease is poor due to chemo- and radio-therapy resistance of the RCC [5].

Several studies showed that the survival rate for RCC patients is inversely proportional to disease stage [6]. Moreover, the recent implementation in the follow-up protocols for RCC patient treatments with different compounds resulted in the need for evaluating risk stratification and correcting therapy planning. Therefore, it appears important to find biomarkers for early detection, for differential diagnosis of RCC from benign renal lesions, for prognosis, and follow-up. These biomarkers would significantly improve patient diagnosis and outcome. However, there are currently no biomarkers yet, approved or with a wide consensus from the scientific community, for a reliable screening for RCC [7-9]. Most of the studies aimed to

identify RCC-specific biomarkers have investigated at gene and protein levels either in cells, primary culture [10] or stable lines [11], or in tissues [12-14]. These candidate markers identified by these strategies may not be present or their profile in biological fluids (*e.g.,* urine, serum and plasma) may not be altered, as would be necessary for screenings based on biological material collected with noninvasive or less invasive methods. In fact, serum and plasma are easily collected and available for biomarker discovery. Several investigations described the diagnostic and prognostic potentiality of many proteins singularly considered. Vascular Endothelial Growth Factor [15], Serum Amyloid A [16] and Osteopontin [16, 17] are some of the proteins evaluated in serum and plasma. Recently, there has been a growing interest in applying proteomics to research on clinical diagnostics and predictive medicine. Main proteomic aim is the identification of proteins that have distinctive quantitative/qualitative differences between normal and disease samples. There are different strategies for the search of protein biomarkers in biological fluids. Based on the proteomics feature to identify and detect several alterations simultaneously, cluster of possible markers can be built thus improving the diagnostic / prognostic capability compared to a single biomarker. Moreover, serum is an extremely complex biofluid containing a plethora of high-abundant proteins that hamper detection and differential quantification of potential biomarkers, which for early stage malignancies in particular, are likely to be present at relatively low abundance. Therefore several techniques are used to simplify or to enrich in low abundant proteins/peptides. Immunodepletion based on antibodies removing from 6 to 14 of the most abundant proteins [18], beads or MALDI target with an activated surface are now widely used. Beads with a library of peptides [19] , mesoporous silica particles for the selective enrichment of low molecular weight proteins [20, 21]or beads with a magnetic core are used to pre-purify the biological fluids before mass spectrometry analysis, thus removing/diminishing the abundant proteins. This last approach has been successfully used to investigate in pilot studies, with a small number of ccRCC patients and control subjects, serum and urine as a source of possible biomarkers not

only for RCC [22, 23], but also for different type of tumors [24]. A technique based on activated target surface, well known as SELDI-MALDI-TOF, has also been successfully used to provide preliminary evidence that a cluster of peptides can be used to build a decision tree flow able to differentiate not only ccRCC from normal subjects [25], but also from benign renal tumors in combination with CT [26]. Additional evidence of the usefulness of this protein profiling approach was shown in a pilot study aimed to detect candidate serum markers in chronic renal disease [27]. More interesting is the possibility offered by proteomics strategies to detect protein alterations with prognostic capability in serum of patients. Urinary expression of Cathepsin D, initially observed to be altered in human cell lines by 2DE profiling, was reported as a possible independent prognostic marker [28]. SELDI strategy was used by the same group to investigate serum of RCC patients [29]. They detected several peptides with an independent prognostic value. Among them, Serum Amyloid A (SAA) protein was evaluated in serum. Results showed that pre-operative level of SAA has an independent statistical significance for prognosis. Most, if not all, of these evidences about the usefulness of protein profiling techniques to investigate biological fluids to biomarker discovery were provided in pilot studies thus limiting their strength. Therefore we apply the proteomic approach based on magnetic beads pre-purification followed by MALDI-TOF analysis and by nanoLC-MS/MS to serum of a large cohort of ccRCC, non-ccRCC patients and control subjects.

2. Materials and Methods

2.1 Chemicals and standards

Profiling Kit Dynabeads® RPC18 were purchased from Invitrogen (Carlsbad, CA, USA); α-cyano-4-hydroxycinnamic acid, Protein Calibration Standard I (ProtMix I) and Peptide Calibration Standard II (PepMix II) were supplied by Bruker Daltonics GmbH (Bremen, Germany); Water, Acetone, Acetonitrile, Ethanol (HPLC-grade) and Trifluoroacetic acid were purchased by Sigma-Aldrich Chemie GmbH (Buchs, Switzerland).

2.2 Blood samples collection and handling procedure

Blood samples were collected from patients the day before surgery and from healthy volunteers at "Ospedale Maggiore Policlinico" Foundation (Milan, Italy) and Desio Hospital (Desio, Italy). All subjects signed an informed consent prior to sample donation. Study protocols and procedures were approved by the local ethic committee and analysis was carried out in agreement with the Declaration of Helsinki. Serum samples were prepared as follows: blood was collected in a 10 mL BD vacutainer tube and let clot at room temperature (RT) for 30 minutes in upright position in the rack. About 2 mL of serum were obtained after centrifugation at 2000 rpm for 10 minutes at 4 °C, transferred to small aliquots and stored at -80 °C until further analysis. A tube per sample was thawed only once for the automated peptide isolation procedure.

2.3 Peptidome separation with magnetic beads

Reversed-phase C18-functionalized magnetic beads (RPC18 MB) were used to capture peptides/proteins in serum samples. The same profiling kit was employed to purify all samples and the good functionality of this RPC18 beads suspension was ensured before use by an optical microscopic investigation.

The extraction procedures were automatically achieved using an 8-channel ClinProt pipetting robot (Bruker Daltonics, Germany) mainly based on the manufacturer's instructions. Briefly, 10 μL of RPC18 Dynabeads were mixed with 5 μL of serum sample after four washes with 40 μL of a 0.1 % TFA solution. After an incubation of 1 min at RT into a magnetic separator, the beads with the peptides bound to their surface were then washed three times with 25 μL of a 0.1 % TFA solution. Finally, the peptides were released with 15 μL elution solution (50 % acetonitrile in water). A small volume of eluates was used for MALDI-spotting procedure, while the residue was frozen and stored at -80 °C for the following nano-LC-ESI MS/MS analysis. Two samples were also prepared by mixing 50 μL of serum from 80 patients (ccRCC n=62 plus non-ccRCC n=18) and 80 control subjects and manually purified with RPC18 MB for the label-free quantitation (LF).

2.4 MALDI-TOF peptide profiling

MALDI spotting of RPC18 eluates was performed using the 8-channel pipetting robot. In detail, 2 μL of the eluted peptide fraction were mixed with 10 μL CHCA matrix solution (0.3 g/L in ethanol/acetone 2:1) in a 96-well plate. Then, 0.8 μL of this mixture were spotted in quadruplicate directly onto a MALDI AnchorChip™ 600/384 target plate (Bruker Daltonics, Germany). The target plates were air-dried and immediately transferred into the mass spectrometer MALDI-TOF/TOF. Purified samples were analyzed using an UltrafleXtreme[TM] MALDI-TOF/TOF MS instrument (Bruker Daltonics, Germany) and mass spectra were automatically acquired in positive linear mode. Peptides were detected in a mass range of *m/z* 1 to 12 kDa and the external calibration was achieved using a commercially available mixture of standard peptides/proteins, ProtMix I. Analyses were performed using the AutoXecute tool (version 3.3.108.0) of FlexControl software version 3.3 (Bruker Daltonics, Germany) with following acquisition settings: Ion Source 1, 25.13 kV; Ion Source 2, 23.77 kV; lens 6.46 kV; pulsed ion extraction 300 ns; high gating factor with matrix suppression up to 900 Da. Ionization was performed with a

Smartbeam™-II solid-state laser operating at a frequency of 1000 Hz. For each MALDI spot, spectra were recorded from six different spot positions (200 shots per position) and summed up (1200 satisfactory shots). Data acquisition was carried out setting the laser power at 60 % of the maximum laser energy. The accurate mass of the peptides was determined in positive reflector mode. Mass spectra were externally calibrated using standard peptides of PepMix II. Instrument parameters were: Ion Source 1, 25.13 kV; Ion Source 2, 22.43 kV; lens 7.45 kV; deflection mode with matrix suppression up to 400 Da.

2.5 Peptide identification by MALDI-TOF/TOF

For peptide identification, the LIFT-TOF/TOF spectra were recorded on the UltrafleXtreme™ MALDI-TOF/TOF MS instrument. No additional collision gas was applied to acquire the MS/MS spectra. The fragment masses were analyzed after their ion reflector detection. Analyses were performed using the following acquisition settings: Ion Source 1, 7.5 kV; Ion Source 2, 6.7 kV; lens 3.6 kV; reflector, 29.5 kV; reflector 2, 13.95 kV; Lift 1, 19 kV; Lift 2, 3.15 kV; pulsed ion extraction 80 ns. Raw MS/MS data were processed via FlexAnalysis™ 3.3 software (Bruker Daltonics, Germany). Database searching was performed by in house Mascot search engine (Version: 2.3.02) using following parameters: taxonomy was restricted to human Swissprot (accessed Feb 2012- 20,317 sequences) database; no enzyme and fixed and variable modifications were set; MS and MS/MS tolerances were generally set at 1 Da. Only identifications with a score higher than Mascot identity thresholds were accepted.

2.6 MALDI FT-ICR peptide precision profiling

MALDI-Fourier transform ion cyclotron resonance (FT-ICR) experiments were performed on a Bruker 15 tesla solariX™ FT-ICR mass spectrometer equipped with a novel CombiSource (Bruker Daltonics, Germany), as reported previously [30]. The MALDI-FT-ICR system was controlled by Compass solariXcontrol software and

equipped with a Bruker Smartbeam-II™ Laser System that operated at a frequency of 500 Hz. The "medium" predefined shot pattern was used for the irradiation. Each mass spectrum was obtained from a single scan of 600 laser shots using 512 K data points. Typically, the target plate offset was 100 V with the deflector plate set at 180 V. The ion funnels operated at 100 V and 6.0 V, respectively, with the skimmers at 15 V and 5 V. The trapping potentials were set at 0.60 V and 0.55 V, the analyzer entrance was maintained at -7 V, and side kick technology was used to further optimize peak shape and signal intensity. The required excitation power was 28 % with a pulse time of 20.0 µs. The MALDI FT-ICR spectra were internally calibrated using the same six peptides as reported previously [30]. Peaks were detected using the FTMS algorithm with a signal-to-noise threshold of 5 and using the centroid for peak position with a percentage height of 85. Mass measurement errors (MME's) were calculated from multiple FT-ICR-profiles of multiple replicates obtained from a serum pool.

2.7 Profile processing

Visualization of mass spectra and initial data processing were performed with FlexAnalysis^TM 3.3 software. Pre-processing was necessary to reduce experimental variance within the data set before subsequent statistical analysis and the workflow consisted of baseline correction and realignment of raw spectra using the *Batch Process* tool. Baseline subtraction of all spectra was applied using the "TopHat" algorithm. Then a subset of peaks common to most if not all sample spectra was identified in order to realign them. For linear profile data, 6 frequent peaks (median values assigned to these peaks: *m/z* 1206.9, *m/z* 1466.1, *m/z* 2931.7, *m/z* 3261.8, *m/z* 4207.9 and *m/z* 5902.8) were chosen based on a manual inspection of few spectra to create a calibration mass control list with a peak tolerance of 2000 ppm. An internal calibration was also performed on reflector data using 11 common peaks and a tolerance in peak assignment of 200 ppm except for the *m/z* 1350.7, 1465.7, 1536.7 that was 100 ppm. A quadratic calibration algorithm was used for the alignment.

Peaks were detected using the SNAP peak detection algorithm with signal-to-noise (S/N) threshold of 1.

2.8 Data Analysis

Multiple spectra comparison and peptide pattern recognition were achieved using ClinProTools[TM] 2.2 software (Bruker Daltonics, Germany). ClinProTools[TM] was used for all MS data interpretation procedures, including comparison of investigated classes spectra profiles and identification of disease marker candidates. Normalization of spectra to their total ion current followed by an internal signal alignment using prominent *m/z* peak values and by peak picking procedure were performed before the peak-area statistic calculation. Spectra pre-treatment parameters were tested and optimized for the further data elaboration. Data selection was achieved based on these filters: resolution value of 800, "Top Hat" baseline correction with a minimum baseline width of 10 %, peaks smoothing by applying the Savitzky-Golay algorithm and null or not recalibratable spectra exclusion. These filters aimed on selecting only "good" spectra and excluding those of lower quality. The mean spectrum obtained from each subject's data set was used for the next statistical analysis by enabling "Support Spectra Grouping" option. Peak detection was carried out with a signal-to-noise ratio (S/N) of 5 and peak areas were calculated using end-point level integration type on total average spectrum. ClinProTools[TM] automatically provided a list of peaks sorted according to the statistical relevance to differentiate between classes with their corresponding p-value (<0.05). Statistical value of differentially expressed peaks was determined by *Student's t-test*. Selection of signal clusters able to differentiate the three studied populations was performed using Support Vector Machine (SVM) and Genetic Algorithm (GA). Specificity and sensitivity values were calculated for each selected model and the diagnostic performance of the signals was evaluated within the whole subjects by a cross-validation process using *k-fold* test ($k = 10$). Data were divided randomly into 10 equally folds; subsequently 10 iterations of training and validation were performed

such that within each iteration k-1 folds were used for learning while the remaining omitted part of data was classified against the calculated model. The obtained classification results were averaged and returned as the prediction capability [31]. The receiver operating characteristic curve analysis and area under the curve (AUC) in order to determine the diagnostic efficacy of each single marker were evaluated as previously described [23].

2.9 Peptide identification and profiling by nLC-ESI MS/MS

Serum endogenous peptides in the enriched fraction obtained by RPC18 MB purification either of patients and controls were identified and also profiled by nano liquid chromatography (nLC)-ESI MS/MS. For peptide identification, eluates collected from both automatically and manually serum RPC18 MB pre-fractionation, were pooled and analyzed. Collected eluates were first concentrated with Heto Speed-Vac apparatus (Heto Lab. Equipment A/S, Copenhagen, Denmark), then desalted and residual beads removed to avoid capillary clogging using Ziptip^{TM} μ-C18 Pipette Tips following the standard protocol provided by Millipore. Desalted fractions were injected with a draw speed of 2 ul/min into EasynLC^{TM} system (Proxeon Biosystems, Odense, Denmark) coupled with a MaXis hybrid UHR-QToF system (Bruker Daltonics, Germany).

Two pools of sera from 80 patients and 80 healthy subjects were used to evaluate patients/controls peptide ratios by LC-ESI. The two samples, purified by RPC18 MB as described above, were diluted before the injection of a fold to avoid signal saturation. To increase the number of peptide identifications, independently of the quantitative evaluation, more concentrated samples were analyzed [32]. LC-MS analysis was performed as described in Mainini *et al*. [33]. Raw MS/MS data were lock-mass corrected (at *m/z* 1221.9906), deconvoluted and converted to XML peaklists via Compass DataAnalysis^{TM} v.4.0 Sp4 (Bruker Daltonics, Germany). Peakfinder (sumpeaks) was set to exclude any ion with <1 S/N and <20 counts intensity. XML data were interrogated using in house Mascot search engine (Version:

2.3.02). Database searching was restricted to human Swissprot (accessed Feb 2012-20,317 sequences) database. No enzyme and fixed modifications were set in search parameters. Any variable modifications setting as Mascot search parameters were specified in the results (selected variable modifications: Acetyl (N-term); Acetyl (K), Amidated (C-term), Oxidation (M); Phospho (Y) / Phospho (ST), Sulfo (S) / Sulfo (Y) / Sulfo (T)). Mass tolerances for all identifications were generally set at 5-10 ppm MS and 0.1-0.5 Da MS/MS. Mascot thresholds score for homology and identity and decoy database were used as peptide level filters of peptide significance.

2.10 Label-free peptide / protein abundance determination

Peptide abundance in the eluted fractions from RPC18 magnetic beads purification of a serum pool of either 80 patients or 80 control individuals was evaluated by IDEAL-Q software (vers 1.6.0.2). The raw data of the three analyses for each sera pool were converted to mzXML using CompassXport vers 3.0.4 (Bruker Daltonics, Germany). Peptides identity was obtained using Mascot and the output of the DataAnalysis^{TM} elaboration as Mascot generic format (mgf) file. In order to allow spectra data alignment (mzXML) with the search results the compound number (Cmpd) in the mgf file was replaced with the scan number. Mascot search results obtained using the above described parameters were exported as XML file format. The mzXML and their corresponding XML files were used to determine the serum patients/controls peptide ratios. The IDEAL-Q data input was set to centroid MS1 data mode. Default parameters for Q-TOF instruments were used for data peak alignments. Normalization was performed using the median of peptide abundance and quantile of the peptide ratio level [34]. Protein abundances were determined using the same parameter settings and a peptide score threshold of 37 (score for identity or for extensive homology) as calculated by Mascot algorithm.

Then peptide ratios obtained with ESI-MS/MS were compared with those calculated from MALDI-TOF spectra. To this end, the mass spectra obtained by MALDI-TOF in linear mode (MALDI-LN) for each patient and control whose serum was pooled

were used to calculate the patients/controls *m/z*-ratio. Alignment of the MALDI-LN *m/z*-ratios with those from LC-ESI analysis was carried out based on the mass spectra acquired by MALDI-TOF in reflector mode, by ensuring the identity for some of them by MALDI-TOF/TOF and by an additional validation from accurate mass measurements in MALDI FT-ICR profiles. Moreover *paired samples t-test* was used to test that the average of the differences between the MALDI peptide ratios and the corresponding normalized ESI ratios, quantified with a label-free approach, are not statistically different ($p > 0.05$).

3. Results

3.1 Clinical data and study design

Three cohorts of 92 control subjects (ctrl) (68 men, 24 women), 85 clear cell RCC (54 men, 31 women) and 29 other different histological subtypes (non-ccRCC) (15 men, 14 women) patients were studied. Mean age for controls was 46 with a range of 30 - 64 years, while for patients 64 with a range of 33 - 87 years. Patients' clinical characteristics, classified according to the 2009 TNM (tumor-node-metastasis) system classification, were summarized in Table 1 [35]. Histological analysis was performed on patients and comprised Fuhrman grading system, histological subtype, sarcomatoid and cystic differentiation, tumor necrosis, microvascular and collecting urinary infiltration. Patients are divided in groups according to their pathologies: ccRCC (n=85) and non-ccRCC (n=29). Tumor patients underwent surgical excision of the renal lesion.

3.2 Protein profiles of the three studied populations and cluster analysis

A large number of ccRCC patients and control subjects including also non-ccRCC tumor patients were evaluated by serum proteome pre-fractionation with reversed-phase C18 magnetic beads before MALDI-TOF mass spectrometry analysis. Serum MALDI-TOF spectra were acquired, aligned and statistically evaluated aiming for discriminating cluster of signals discovery. The average serum peptide profiles of controls, ccRCC and non-ccRCC patients after RPC18 samples pre-purification are reported in Supplemental Figure 1. The MALDI-TOF analysis provided high quality profiles, in which more than 100 signals common to the three studied populations have been detected in the mass range 1 - 12 kDa and about 40 were observed with a statistically different area ($p<0.05$ by one-way ANOVA analysis) (Supplemental Table 1). Because of the high complexity of spectra profiles statistical analysis and discriminant pattern recognition were performed using two of the algorithms supplied

by ClinProTools[TM] software, Support Vector Machine and Genetic Algorithm, for the possible accurate classification of the three studied groups.

The entire data set was initially split into malignant tumors (n=102) and benign renal masses (n=12) plus healthy subjects (n=92) and a discriminant cluster of five peptides (linear m/z 1264, 1419, 2423, 2789 and 4962) was built with 90 % and 94 % of specificity and sensitivity (in training phase), respectively (Table 2A). Based on the receiver operating characteristic (ROC) curves, all five ions displayed good diagnostic power according to the criteria suggested by Swets [36] with the value of area under the curve (AUC) that is included in a range of 0.84 − 0.75. The combination of these peaks showed an improvement in the performance compared to each single ion. Among these peaks, only the ion at m/z 4962 was under-represented ($p<0.05$), while the others were over-represented ($p<0.05$) in malignant kidney tumor patients compared to controls and subjects with benign renal masses. The *k-fold* cross-validation test (k=10) was applied to evaluate the prediction capability of this model and the obtained classification results showed 84 % and 89 % of specificity and sensitivity, respectively.

In addition, using an interactive analysis three different case/control classification tasks were also performed: (i) controls *versus* ccRCC; (ii) controls *versus* non-ccRCC; (iii) ccRCC *versus* non-ccRCC patients. Classification performances for each different case/control comparison are reported in Table 2B-2C. A cluster of two peptides at m/z 1466 (AUC = 0.89) and at m/z 4052 (AUC = 0.85), determined by MALDI-TOF in linear mode, was selected by a statistical analysis of spectra and both peaks showed a very high accuracy on discriminating ctrl *versus* ccRCC. Their simultaneous combination resulted in a specificity of 92 % and a sensitivity of 92 % in the training phase, while in the validation test the specificity and sensitivity were 88 % and 85 %, respectively. Aiming to improve the discriminating efficacy of this predictive model based on only two ions new additional ions were combined. A pattern of four (linear m/z 1466, 2378, 2789 and 4052) and of six signals (linear m/z 1466, 2378, 2789, 4052, 4089 and 4151) allowed to improve the classification power

(Table 2B). The area under the curve (AUC) of the peaks used in these models, evaluated by the ROC curve analysis, was comprised in the range 0.89 – 0.70, which corresponds to a moderately accurate test [36]. However, the combination of the six ions of the latest cluster showed higher specificity (96 %) and sensitivity (94 %) compared to the values obtained with the other smaller ones. The cross-validation analysis confirmed the good performance obtained in the training phase (93.5 % and 91 % of specificity and sensitivity, respectively). Two of these ions at *m/z* 1466 and 2789 were observed to be in higher amount in ccRCC patients, while the others showed a higher concentration in control subjects. In order to avoid overfitting the number of ions included in the discriminant cluster was not increased even more; however any improvement was observed even if new signals were added in the pattern.

A high-performing pattern of three ions (linear *m/z* 1419, 2603 and 4264) was also noticed to properly distinguish healthy subjects from non-ccRCC patients (ctrl *versus* non-ccRCC) with very good diagnostic capability (96 % and 90 %, in the training, and 95 % and 90 %, in the validation phases, of specificity and sensitivity, respectively) (Table 2C). The ion at *m/z* 1419 (AUC = 0.85) was over-represented (P<0.05), while the others were under-represented in non-ccRCC patients compared to healthy subjects.

The same approach for signal selection based on two algorithms with inclusion of six or even more ions in the cluster failed to discriminate ccRCC from non-ccRCC patients.

3.3 Peptide identifications

Analysis by nLC-ESI MS/MS of the enriched fractions obtained with RPC18 MB pre-fractionation of serum of controls and patients affected by kidney tumor allowed us to build a library of identified endogenous peptides (Supplemental Table 2). In particular, it included the identification of about 200 unique peptides, with a score above peptide identity threshold with a false discovery rate <1.5 %, belonging to 32

different protein and to 262 peptides using as cut-off the score for homology threshold with a false discovery rate <5 %, (Table 3). About 60 % of the 201 peptide isoforms identified with score above identity threshold was detected in both tumor and control samples, while 30 peptides were identified only in patients and 53 only in healthy subjects. More than 25 % of all identified peptides present in this library was matched setting in Mascot search parameter possible post-translational modifications. Sulphation, phosphorylation, N-terminal acetylation, and C-terminal amidation were the most frequent hypothetical modifications detected, as already reported for endogenous peptides [37].

Forty-three of the all 101 signals present in MALDI-linear mode (LM) spectra were most likely recognized in the identification library based on their matching Mr evaluated by LC-ESI with those determined by MALDI-TOF in linear and reflector mode (RM) and also by MALDI FT-ICR (Supplemental Table 3). The average differences between MALDI and ESI molecular weights for quantified peptides above identity were 12.0 ± 26.1 ppm for RM and 101 ± 192 ppm for LM. The average mass measurement error from FT-ICR precision profiles compared with exact masses of peptides identified in ESI-LC-MS data was determined at 0.11 ± 0.88 ppm.

3.4 Validation of ratios

With the aim of validating the peptide profile results obtained by MALDI-TOF analysis, the peptide ratios obtained by MALDI-TOF based strategy were compared with those resulted from label-free (LF) relative quantification approach performed on the same subjects.

To this purpose, two representative sample groups of patients and controls were chosen for LF analysis from whole cases of subjects evaluated by MALDI based technique. Therefore two serum pools were prepared: one with 80 controls and one with 62 ccRCC and 18 non-ccRCC patients.

The mass spectra obtained by MALDI-TOF in linear mode for each patient and control, whose serum was exploited for the two pools, were used to calculate the patients/controls *m/z*-ratio. Alignment of the MALDI-LN *m/z*-ratios with those from LC-ESI analysis was made based on the mass spectra acquired by MALDI-TOF in reflector mode (RM). The average differences between Mr determined by MALDI-RM and by ESI for the quantified peptides with a score above identity threshold was 8.4 ± 13 ppm. Additional validation of mass alignment was supported by accurate mass measurements in MALDI FT-ICR profiles. The average MME of FT-ICR precision profiles compared with ESI calculated mass for quantified peptides identified with a score above identity threshold was determined at 0.14 ± 0.92 ppm. Three of these ions, which are matched by *m/z* value in MALDI-LM and MALDI-RM and in LC-ESI MS/MS, were also identified by MALDI-TOF/TOF analysis. All of these peptides were identified based on their MS/MS spectra as fragments of Fibrinogen alpha-chain (SwissProt accession number P02671) with a Mascot score exceeding the significant threshold of identity ($p < 0.05$). In particular, signal at *m/z* 1263.63 determined by MALDI-TOF in reflector mode (Mr calc 1262.589) was identified (score=93) as peptide with the sequence GEGDFLAEGGGVR. The other two ions at *m/z* 1350.66 (Mr calc 1349.621) and *m/z* 1465.69 (Mr calc 1464.648) were identified as peptides with a sequence SGEGDFLAEGGGVR (score=112) and DSGEGDFLAEGGGVR (score=130) , respectively.

In particular, 26 signals detected in MALDI-RM plus 6 in MALDI-LM were most likely identified by LF approach with a Mascot score above identity threshold and two observed in MALDI-RM with a score above homology threshold (Table 4).

In only two cases it was not possible to effectively distinguish two peptides, identified in LC-ESI MS/MS and with a Mr difference of about 1 Da, in the MALDI-TOF mass spectrum obtained either in RM and LM, even with the additional information deriving from FT-ICR analysis. As shown in Table 4, the ions at Mr 2601.30 and 2602.17 identified in the LC- MS/MS analysis at different retention time (r.t.) could correspond to the signal at *m/z* 2603 and at 2602.28 observed in the

MALDI-TOF spectrum acquired in linear mode and reflector mode, respectively. The first signal was identified as a fragment of F13A (r.t. 82 min) while the second as a fragment of SRGN (r.t. 77 min). Both these peptides were observed in MALDI FT-ICR profiles, namely at *m/z* 2602.311 and at *m/z* 2603.168. Furthermore, we observed with LF approach two peptides, one at Mr 3155.62 (r.t. 74 min) identified as fragment of ITIH4 and the other at Mr 3156.61 (r.t. 93 min) as fragment of TTHY, that could match to the signal at *m/z* 3156.64 observed in MALDI-RM spectrum. However, in MALDI FT-ICR profiles only the ion at *m/z* 3156.630 could be linked to the LC-MS data. The ion at *m/z* 4786 detected in MALDI-LM could not be observed in corresponding reflector spectra. MALDI FT-ICR analysis links this peptide to the one identified as FIBA fragment corresponding to an Mr of 4784.03 Da (Table 4). Peptide ratios obtained from LC ESI-MS/MS were compared with those calculated from MALDI-TOF spectra. No statistical difference resulted by *t-student test for paired data* between the MALDI-TOF and LC-ESI MS/MS m/z ratios (p=0.4547). Moreover also splitting the dataset in the two groups of under-represented (MALDI-LM ratio<1) and over-represented peaks (MALDI-LM ratio>1) no statistical difference was noticed (p=0.3044 and p =0.5378, respectively).

4. Discussion

4.1 Clinical perspectives

Proteins and peptides that are differentially represented as a consequence of a disease are very useful in clinical research as they can be used as biomarkers for screening, diagnostic and prognostic purposes [38]. Several high-throughput MS-based protein profiling techniques have been developed in order to discover potential biomarkers in body fluids able to characterize and predict multifactorial diseases. They are based on the assumption that biological fluids proteome would reflect the body status as they carry proteins secreted from tissues. Therefore, the direct investigation in peripheral fluids for proteins related to an altered physiological state represents a valid approach because samples can be easily obtained through non- or very low-invasive methods. Nowadays, protein profiles of biological fluids, which permit to select a distinctive pattern of signals able to characterize their pathological state, can be obtained from patients [39]. Highly sensitive profiling studies require a combination of MS and separation technologies [40]. RCC-specific biomarkers have been searched in primary culture, stable lines, and in solid biological material (primary cell cultures, stable cell lines and tissues) with classical proteomics approaches [10-13]. Sample fractionation can be used to increase the number of proteins detectable with MS (*e.g.,* enriches for the low abundant proteins) such as by bi-dimensional liquid chromatography (MudPIT) or to reduce the sample complexity by activated surfaces. Solid-phase extraction technique based on an off-line fractionation of the proteome using magnetic beads or solid MALDI target, with activated surface, before MS analysis, such as ClinProt and SELDI, permits a high-throughput search in peripheral fluids for proteins/peptides altered by the disease [41-43]. In particular the one based on activated target surface (SELDI-MALDI-TOF) was used to provide preliminary evidence about a serum cluster of peptides able to differentiate ccRCC from normal subjects [25] and from benign renal tumors in association with CT [26]. SELDI strategy was also used to investigate serum of RCC patients with the aim of searching possible prognostic biomarker [29]. We have also previously identified in a very

small pilot study a cluster of peptides able to discriminate RCC patients from healthy subjects thus encouraging the prosecution of this investigation with a wider number of cases based on magnetic bead pre-fractionation before MALDI-TOF analysis both on serum and urine [22, 23]. Here we present a proteomic study based on serum peptidome profiling with ClinProt-MALDI-TOF in Renal Cell Carcinoma. Our aim was to evidence markers which might be useful for RCC early detection, as well as differential diagnosis of RCC from benign renal lesions. Therefore we analyzed a large number of ccRCC patients and control subjects including also non-ccRCC tumor patients. In this study we focused more our analysis on the serum small peptides (Mr <5000Da) obtained by pre-fractionation with reversed-phase (RPC18) magnetic beads. The average mass spectra of the three studied population (Supplemental Figure 1) revealed a number of potentially interesting peaks with statistically different intensities ($p < 0.05$) (Supplemental Table 1). RCC is a complex and multifactorial cancer, as most of the human diseases, therefore combination of several markers should result in a more performing cluster for screening or diagnosis. Primary clinical question and probably also the most relevant about this disease is not about the nature of the solid kidney mass (*e.g.*, ccRCC or papillary) but on its malignancy. Therefore we first focused our statistical elaboration on this aim. High discriminating capability was observed for a cluster of five signals able to differentiate malignant tumors from benign renal masses plus healthy subjects as suggested by the 90 % and 94 % of specificity and sensitivity, respectively, (Table 2A) obtained in the training phase which was also confirmed in the validation test (83.7 % and 89.2 % of specificity and sensitivity, respectively). Then we verified also the possibility to investigate the nature of the kidney lesions making different group comparisons. Several clusters, with an increased complexity from 2 up to 6 ions, were tested. Performances of the various clusters for control *versus* ccRCC discrimination, are quite comparable (Table 2B). The best discrimination was showed when six peptides (linear *m/z* 1466, 2378, 2789, 4052, 4089 and 4151) were included in the cluster. We limited the inclusion in the cluster up to six signals to avoid

overfitting. However no diagnostic improvement was observed increasing the number of signals to more than six ions. The non-ccRCC could be also well differentiated from controls with a simple cluster of three signals (Table 2C). On the contrary the non-ccRCC group of patients could not be discriminated from ccRCC patients probably because we investigated a small number of samples and a very heterogeneous group of patients.

4.2 Ratio Validation

Optimally, an independent validation method is performed to verify results obtained by analytical methodologies such as protein profiling aiming at the discovery of potential biomarkers. This validation is generally achieved by western-blot based on immunodetection with available antibodies. However , such validation may not be possible *e.g*., in the case in which an antibody is not available or for small peptides that may not contain antigenic amino acid sequence and/or antibody epitopes. Moreover we should also consider that, as well known in the common experience of researchers working in this field, antibody binding not always works properly and the cross-reactivity may lead to wrong results. Recently, a combined method was suggested for the analysis of trypsin digested proteins extracted from human urine using high-throughput MALDI-TOF analyzing each individual sample to maintain the biological variability [44]. In this approach LC-ESI- MS/MS is used in combination with stable isotope labeling for analyzing the pooled urine from normal and benign prostatic hyperplasia patients, which provided protein identity and quantification ratio [44]. In our work we have proposed to apply a relative quantification label-free (LF) approach based on an alternative ionization method to validate the peptide profile results obtained by RPC18 MB pre-purification combined with MALDI-TOF analysis. In particular we compared the peptide ratios obtained by this MALDI strategy with those resulted from the LF approach performed on serum pools from the same control subjects and patients. LF proteomic methodology was

herein applied for the quantitative evaluation at peptidomic level and not targeted to protein level as usually reported for complex tryptic mixtures [45].

The comparison of the ratios of peptide signals quantified both in MALDI and ESI was feasible through an accurate mass alignment. Many of the identifications performed by LC-ESI-MS/MS were indeed well matched with signals present in the MALDI-LM profiles of the analyzed serum samples (Supplemental Table 3). Alignment of the MALDI-LM *m/z*-ratios with those from LC-ESI analysis was done through the mass spectra acquired by MALDI-TOF in reflector mode (RM) and further confirmed by MALDI FT-ICR profiles and MALDI MS/MS identifications. MMEs were found to be at a low-ppm level as reported in the results. Only in two cases we identified by LC-ESI-MS/MS two peptides that could be matched to a single ion present in the MALDI-LM spectra (*m/z* 2603 and 3157). However the ratios for both the two possible peptides identified by ESI-MS/MS, that can be matched to the ions observed in MALDI-TOF, were similar. The ratio of the signal at *m/z* 2603 was determined as down-represented either for the two peptides identified by ESI and for the signal observed by MALDI. For the other signal at *m/z* 3157, found to be probably present in a higher amount in controls based on MALDI results, no alteration was observed for the peptide at Mr 3156.61 and a higher concentration in healthy subjects for the one at Mr 3155.62 by ESI.

Twenty-eight of the 101 ions present in the MALDI-LM spectra were quantified also with LF approach (Table 4). Statistical analysis did not show any difference (p>0.05) between peptide ratios determined by MALDI and ESI thus reinforcing their validity.

4.3 Peptide identification

The successful discovery of a peptide profile correlated to an altered state through technologies based on the biological fluids proteome magnetic beads (MB) and ProteinChip pre-fractionation combined with MS analysis has been reported in various human diseases [38] including renal cell carcinoma in urine [22] and serum [23]. These selected clusters of peaks could represent promising diagnostic tools even

independently from their structures identity [22]. Nevertheless the identification of the endogenous peptides present in the pre-fractionated eluates especially for the altered signals is important. However a drawback for further validation of potential biomarkers with other approaches is due to the lack of their identity. Moreover specific assay development, including immunoassays that might potentially overcome some of the analytical limitations of MALDI MS profiling, could be later developed and could increase biological insight exploring the function and the regulation of bioactive molecules and degradome products in biological fluids based on their identity [37]. The molecular weights of these distinguishing peaks are usually under 30 kDa [46] and refer more to peptidome, providing a broader view of the entire peptide pool in a biological sample, than proteome [46, 47]. Modern peptidomics approaches are widely performed by LC-MS/MS due to their high detection and identification capability for low molecular weight proteins and peptides from biological samples. Hundreds of peptides and proteins have been identified either in serum or plasma by LC-ESI MS/MS [48, 49]. Peptide identification by MALDI is more challenging, usually less sensitive and informative, mainly because of the single charging of the precursor ion compared to multiply charged species obtained in ESI-LC MS/MS. MALDI MS profiles have usually a low relation with identification [50, 51]. In fact, only few ions present on MALDI MS profile spectrum have been identified so far. The most comprehensive list of peptides identified in human serum was reported by Villanueva *et al* [24]. Also *Tiss el al*. could identify more than 90 peaks in MALDI MS profiles of human serum with little overlap with already published data sets and originating from 13 proteins [43]. However MALDI-TOF based approach is faster and allows to reduce time for sample preparation (*e.g*., these steps can be even done automatically by a robot) and for the analyses.

We applied nLC-ESI MS/MS based method to build a library of endogenous peptides identified in our enriched fraction obtained with RPC18 MB pre-fractionation of serum of controls and patients. The strength of this approach for peptide sequencing is mainly due to its features, such as accuracy and sensitivity. However, the study of

endogenous peptides using tandem MS is somehow challenging and often requires non CID fragmentation techniques [52], since these endogenous peptides can be relatively large (compared to tryptic peptides) and possibly contain multiple basic residues and labile PTMs [53]. The increase of CID fragmentation energies and the optimization of the MS/MS accumulation times together with the appropriate setting of acquisition parameters allowed us to reach a better fragmentation of endogenous peptides. These settings resulted in a larger number of diagnostic sequence ions and a more confident identification. Exploring such a simplified biological matrix at peptidome level more than 200 different peptides were identified above identity significance threshold belonging to 32 different proteins and about 260 peptides above Mascot homology threshold (Supplemental Table 2).

The matching between nLC-ESI MS/MS identifications and MALDI signals was enforced by the low ppm MMEs of the mass alignment as previously described (Supplemental Table 3). Some of these identified peptides belonged to the selected clusters signals used in diagnostic and screening models (Supplemental Table 4). These peptides mainly derived from proteins related to blood coagulation or complement system activation, in good agreement with those observed by Villanueva *et al*. [24]. These authors indeed have already shown that some biomarkers are not directly expressed by the diseased tissues, but are breakdown products from exopeptidase activities superimposed on the proteolytic events of the *ex vivo* coagulation and complement cascade [24, 54]. They studied in particular breast, bladder and prostate tumors because they involve the transformation of altered cells types that produce specific proteins. Since the same hypothesis can be applied to every type of disease, Villanueva has suggested that specific serum exoproteases may originate specific peptide signatures for different diseases. The exoprotease activities contribute to generation of not only cancer-specific but also "cancer type"-specific serum peptides, implying that different tumor types secrete distinct proteases that through their catalytic activity may generate unique serum peptide profiles [24]. Therefore blood proteins seem to be an endogenous substrate pool for tumor-derived

proteases, suggesting a direct correlation between disease peptide profiles and differential protease activities [55].

It is noteworthy that if we consider the whole identification library, we could observe that twenty-one of the 43 ESI MS/MS identified peptides correlated to signals in MALDI profiles have been reported previously by Villanueva *et al.* [24], although different stationary phases (RP C18 and C8) for coated magnetic beads were used (Table 5 lists and compares these peptides patterns).

The overlapping among several peptides involved in different tumor types highlights even more the importance of the protease activity in blood acting on substrate proteins also depending on their abundance. Indeed high abundance proteins are often only the target of the exoproteases and are not themselves really overexpressed. The concentration of these peptides thus in these cases is not a measure of the abundance of the parent proteins. If we used a very conservative and robust criteria to evaluate down - (ratio fold change ≤0.5) or up-concentration(ratio fold change ≥1.5) as already suggested [56, 57], it could be noticed that the pattern expression trend between kidney tumor and the other three pathologies shows a low yield of coherence of 38 % for prostate cancer, 19 % for bladder tumor and 24 % for breast cancer (Table 5). Therefore each pattern may represent an exclusive "cancer type"-specific degradomic signature.

Seven among twelve signals considered in the discriminant patterns were identified and resulted as peptides mainly belonged to blood coagulation and fibrinolysis (Supplemental Table 4). Three ions at *m/z* 1264, 1466 and 2423 observed in MALDI-LM spectra, identified as fragments of Fibrinogen alpha-chain (FIBA), are present in higher concentration in RCC patients. The first two peptides were fragments of the Fibrinopeptide A (FPA) (Mr 1535.69) which derives from Fibrinogen degradation to fibrin releasing Fibrinopeptides A and B. Instead, the ion at m/z 2423 matched to the sequence between Asp101 and Asn122 of the Fibrinogen alpha-chain (FIBA). Actually, an additional ion with Mr 2422.07 was observed by LC-ESI MS/MS, that could correspond to the same signal at m/z 2423 detected in the MALDI-TOF

spectrum acquired in linear mode and at m/z 2423.09 observed in MALDI-RM. Also this ion was identified as another fragment of the Fibrinogen alpha-chain (FIBA) with the sequence between Phe583 and Met603. Fragments of Fibrinogen A alpha-chain have already been reported to be altered in several pathologies [24, 42, 58]. In contrast with our results, Villanueva *et al*. [24] described the lower concentration of these two peptides (decreased levels from 20 to 80 %) in the serum of patients with prostate, bladder and breast cancer. Their enhanced levels in RCC serum may reflect the increased expression and activation of cancer-type specific proteases. The other four identified peptides (Mr 2377.20, 2601.30, 4088.02 and 4150.17) were under-represented in tumor patients and derive from Complement C4-A or B, Coagulation Factor XIII A Chain (F13A), Prothrombin (THRB) and Plasma Protease C1 Inhibitor (IC1). It is to be noticed that we observed two ions at Mr 2601.30 and 2602.16 in the LC-ESI-MS/MS analysis at different RT, that could correspond to the signal at *m/z* 2603 detected in the MALDI-TOF spectrum acquired in linear mode as previously described. Both peptides are down-represented in patients. The peptide at Mr 2601.30 was a fragment of the propeptide form of F13A, the active peptide released by Thrombin, while the second was a fragment of Serglycin (SRGN), a proteoglycan with serine-glycine dipeptide repeats and characterized by various glycosaminoglycan side chains. Serglycin core protein has 158 amino acid residues that form three functional domains and the peptide we identified in our study is a part of the glycosylation domain [59]. Ueda *et al.* have already reported that the serum level of these two peptides are altered in lung adenocarcinoma patients, as well as other peptides included in our discriminatory clusters, but no indication was provided about their different concentration [60]. The ion at *m/z* 4089 detected in MALDI-LM was identified as Thrombin light chain, cleaved by the factor Xa. Among the others 5 signals, not involved in blood coagulation or complement system, only for one, at *m/z* 4052 in MALDI-LM, a possible identity was obtained as fragment of AKA11 (score of 31 above homology threshold>26). Other two possible amino acid sequence were returned by database search with lower score and possible variable modifications

(ZC12C with a score of 26 and UBP2 with a score of 25). The A-Kinase Anchor Proteins (AKAPs) are a group of cytoplasmic proteins able to bind the regulatory subunit of Protein Kinase A (PKA) and AKA11 is expressed in several organs among which kidney.

4.4 Protein expression

We have also investigated for possible rules or trends to understand or predict how a specific protein can generate a different cluster of fragments for each considered cancer type based on Villanueva *et al.* [24] and including our results. Apparently no evident/clear processes for multivariate predictive tool can be deduced from these data. Thus the divergent relative abundance ratios of peptides deriving from the same protein (*i.e*., from FIBA and other high abundant coagulation factors), observed either by Villanueva *et al.* and by us, with different quantitative approaches most likely can be explained by "cancer type"–specific exopeptidase activities as previously described.

However, for same proteins, label-free ratio values of their peptides, identified with a score above the identity or extensive homology threshold (>37), are all related, within the single protein, to an up- or a down- peptide concentration (Table 6). For these proteins, not described by Villanueva *et al* [24] it could be reasonable assuming that this high or low peptide abundance may reflect serum protein concentration and in some cases being lastly correlated to their biological meaning.

Among them only one protein showed a high overexpression: the Prothrombin (THRB), or Coagulation Factor II. Almost all nine peptides identified and quantified as Prothrombin fragments with LF approach were found with a very high peptide ratios leading to a remarkable higher concentration value in patients coherently with the related protein ratio (Table 6). This protein consists of three regions: fragment 1 (Ala_{44}-Arg_{198}), fragment 2 (Ser_{199}-Arg_{327}), prothrombin light chain (Thr_{328}-Arg_{363}) and prothrombin heavy chain (Ile_{364}-Glu_{622}). Prothrombin is the inactive precursor of Thrombin, it is produced in the liver and after its modification to Glu residues by

gamma-Glutamylcarboxylase in the presence of vitamin K, O_2 and CO_2 it is released into the bloodstream. It is noteworthy that all these nine identified fragments cover only THRB protein sequence from Arg_{314} to Arg_{363} and no other fragments were observed (Supplemental Table 2). This is in agreement with the cleavable sites of the activation process of Prothrombin. It is known that on the surface of a cellular phospholipid membrane that binds the N-term of Prothrombin, Factor Xa removes the activation peptide and cleaves the remaining part into light and heavy chains at Arg_{327} and at Arg_{363}. It is not clear whether other smaller peptides, with additional cleavage after Arg_{314}, could be released in natural blood clotting. Interestingly, we found fragments over-represented in serum of patients affected by kidney tumors some of which nearly completely cover the light chain. Moreover other four fragments, of those present in higher concentration in patients, cover all sequence of THRB fragment 2, from Arg_{314} to Arg_{327}. It has also to be noticed that increased serum level of the THRB fragment $T_{315}ATSEYQTFFNPR_{327}$ (signal at m/z 1561.77; Table 4) has been already described in stage IV of breast cancer [61]. Moreover, in a very recent paper it has been reported an increased concentration in human urine of patients with bladder cancer of the intact form of this protein, which was included in a diagnostic protein panel due to its great efficacy to discriminate the two populations [62].

Serum Deprivation-Response Protein (SDPR), a citosolic protein present also in caveolae, was found to be down-expressed (Table 6). Interestingly, recent data suggest the potential use of SDPR as a biomarker that lies downstream of Fhl1 (four and a half LIM domains 1) [63]. In this study, the authors showed that SDPR and FHL1 gene expression are significantly reduced in breast (p <0.001 and p<0.02), kidney (p<0.01), and prostate (p<0.05) tumors. The suppression of Fhl1 and SDPR expression is driven by Src independently of Mitogen-Activated Protein Kinase (MAPK) [63]. Indeed their outcome establish that Fhl1 induces the expression of SDPR in Src-transformed cells and as with Fhl1, Src utilizes Cas (Crk-associated substrate) to suppress SDPR expression. These data are consistent with other

findings. First, this gene was associated with the ability of normal cells to control the growth of adjacent tumor cells [64]. SDPR was also found to be widely expressed in human tissues, but decreased in many cancer cells, including those of the prostate, kidney, and luminal breast epithelium [65]. Additionally, the SDPR gene is located on chromosome region 2q33 [66], which undergoes a high incidence of loss of heterozygosity in some lung cancers [67]. Finally, SDPR relative protein called SRBC (SDR-related gene product that binds to c-kinase) is also induced by serum deprivation and associated with Protein Kinase C [68]. SRBC expression is suppressed in lung and breast cancers [69], apparently by gene deletion [70] or methylation [69].

Also for the Zyxin a down-expression can be estimated based on the concentration of its six identified fragments (Table 6), in good agreement to actual knowledge about its expression in tumors. Zyxin is a dual-function LIM domain protein that regulates actin dynamics in response to mechanical stress and shuttles between focal adhesions and the cell nucleus. Several studies have already suggested Zyxin as a tumor suppressor [71, 72]. Hervy *et al.* showed also a role for Zyxin in promoting apoptosis through the binding to Cell Cycle and Apoptosis Regulator Protein-1 (CARP-1) [73]. In a more recent work, by RT-PCR, Western blot, and EGFP report system it was confirmed that Zyxin is a direct target gene of miR-16, a microRNA especially upregulated in laryngeal carcinoma cells when compared with normal tissues [74]. MiR-16 could have deep influence in the adhesion and migration by targeting Zyxin. These outcomes are corresponding with the consequence that genetic ablation of Zyxin causes Mena/VASP mislocalization, increased motility, and deficits in Actin remodeling [75] and may agree with our findings. Furthermore Zyxin participates in actin dynamics by binding VASP, an interaction that occurs via proline-rich N-terminal ActA repeats. Since Zyxin binding to several partners, via the LIM domains, requires phosphorylation, it seems likely that Zyxin phosphorylation might alter the head-tail interaction and, thus, Zyxin activity. That Zyxin point mutants at a known phosphorylation site, serine 142, alter the ability of a Zyxin fragment to directly bind

a separate Zyxin LIM domain fragment protein [76]. These data suggest that Zyxin phosphorylation at serine 142 results in the release of the head-tail interaction, changing Zyxin activity at cell-cell contacts. Interestingly, we identified only in control serum samples two Zyxin endogenous fragments contiguous in the sequence which are proteolytic cleaved at this phosphorylation site Ser_{142}-Ser_{143}: one fragment at Mr 3426.70 identified above homology threshold with a score of 28 corresponding to the sequence E_{11}EIFPSPPPPPEEEGGPEAPIPPPPQPREKVS$_{142}$, and a second one at Mr 3008.50 identified related to the sequence S_{143}IDLEIDSLSSLLDDMTKNDPFKARVS$_{169}$. Recently, Sansing $et\ al.$ investigated the integrin effectors in oral carcinoma [77] using a mass spectrometry based proteomics approach. These authors could determine that a reduction in Zyxin expression lead to a modulation of cell spreading, a reduced cancer cell line resistance to the chemotherapeutic drug cisplatin, and a decreased proliferation. Among the integrin effectors studied in this report [77], the targeted reduction of Talin had the most potent impact on adhesion-dependent processes including invasion, spreading and chemoresistance. In this respect we could identify fragment F_{409}GLEGDEESTMLEDSVSPKKSTVLQ$_{433}$ (Mr 2725.30) as the of Talin-1 (TLN1) that was detected only in control subject samples.

Two of the identified proteins reported in Table 6 as down-expressed show a ratio trend in contrast with literature data.

Serglycin (SRGN) is a low mass weight secretory granule proteoglycan protein with serine-glycine dipeptide repeats and modified with various glycosaminoglycanside chains. SRGN is known as promoting factor of cancer cell metastasis [59] due to this glycosylation. Serglycins were found up-expressed in many cancerous tissues, as drug resistant tumor cell of hematopoietic origin [78] and nasopharingeal carcinoma [79].

The other protein is the Thymosin β-4-Like Protein 3 (TMSL3) which is heavily involved in the organization of the cytoskeleton. Our result is different from that reported in a very recent study indicating that the expression of TMSL3 promotes

tumorigenicity and metastatic potential and induces drug resistance of malignant cells [80].

However, since our working is at peptide level, these differences could be ascribed to the presence of peptides shared among multiple protein isoforms having different abundance [81, 82] or by the presence of post-translational modification of the protein with a different concentration of the modified peptide.

In conclusion our results suggest the possibility to discriminate malignant kidney tumor from benign one and from controls based on a cluster of serum peptides. Moreover, a label-free LC-MS approach may represent a valid method to verify results obtained by MALDI-TOF profiling strategy applied to biomarker discovery.

Acknowledgments

References

[1] Gupta K, Miller JD, Li JZ, Russell MW, Charbonneau C. Epidemiologic and socioeconomic burden of metastatic renal cell carcinoma (mRCC): a literature review. Cancer Treat Rev. 2008;34:193-205.

[2] Rini BI, Campbell SC, Escudier B. Renal cell carcinoma. Lancet. 2009;373:1119-32.

[3] Banks RE, Craven RA, Harnden P, Madaan S, Joyce A, Selby PJ. Key clinical issues in renal cancer: a challenge for proteomics. World J Urol. 2007;25:537-56.

[4] Rathmell WK, Godley PA. Recent updates in renal cell carcinoma. Curr Opin Oncol. 2010;22:250-6.

[5] Mulders P, Figlin R, deKernion JB, Wiltrout R, Linehan M, Parkinson D, et al. Renal cell carcinoma: recent progress and future directions. Cancer Res. 1997;57:5189-95.

[6] Pantuck AJ, Zisman A, Belldegrun AS. The changing natural history of renal cell carcinoma. J Urol. 2001;166:1611-23.

[7] Ljungberg B. Prognostic markers in renal cell carcinoma. Curr Opin Urol. 2007;17:303-8.

[8] Seliger B, Dressler SP, Lichtenfels R, Kellner R. Candidate biomarkers in renal cell carcinoma. Proteomics. 2007;7:4601-12.

[9] Seliger B, Menig M, Lichtenfels R, Atkins D, Bukur J, Halder TM, et al. Identification of markers for the selection of patients undergoing renal cell carcinoma-specific immunotherapy. Proteomics. 2003;3:979-90.

[10] Perego RA, Bianchi C, Corizzato M, Eroini B, Torsello B, Valsecchi C, et al. Primary cell cultures arising from normal kidney and renal cell carcinoma retain the proteomic profile of corresponding tissues. J Proteome Res. 2005;4:1503-10.

[11] Aggelis V, Craven RA, Peng J, Harnden P, Cairns DA, Maher ER, et al. Proteomic identification of differentially expressed plasma membrane proteins in renal cell carcinoma by stable isotope labelling of a von Hippel-Lindau transfectant cell line model. Proteomics. 2009;9:2118-30.

[12] Raimondo F, Salemi C, Chinello C, Fumagalli D, Morosi L, Rocco F, et al. Proteomic analysis in clear cell renal cell carcinoma: identification of differentially expressed protein by 2-D DIGE. Mol Biosyst. 2012;8:1040-51.

[13] Seliger B, Dressler SP, Wang E, Kellner R, Recktenwald CV, Lottspeich F, et al. Combined analysis of transcriptome and proteome data as a tool for the identification of candidate biomarkers in renal cell carcinoma. Proteomics. 2009;9:1567-81.

[14] Morris MR, Ricketts CJ, Gentle D, McRonald F, Carli N, Khalili H, et al. Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma. Oncogene. 2011;30:1390-401.

[15] Jacobsen J, Rasmuson T, Grankvist K, Ljungberg B. Vascular endothelial growth factor as prognostic factor in renal cell carcinoma. J Urol. 2000;163:343-7.

[16] Ramankulov A, Lein M, Johannsen M, Schrader M, Miller K, Loening SA, et al. Serum amyloid A as indicator of distant metastases but not as early tumor marker in patients with renal cell carcinoma. Cancer Lett. 2008;269:85-92.

[17] Ramankulov A, Lein M, Kristiansen G, Meyer HA, Loening SA, Jung K. Elevated plasma osteopontin as marker for distant metastases and poor survival in patients with renal cell carcinoma. J Cancer Res Clin Oncol. 2007;133:643-52.

[18] Echan LA, Tang HY, Ali-Khan N, Lee K, Speicher DW. Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. Proteomics. 2005;5:3292-303.

[19] Santucci L, Candiano G, Bruschi M, D'Ambrosio C, Petretto A, Scaloni A, et al. Combinatorial peptide ligand libraries for the analysis of low-expression proteins: Validation for normal urine and definition of a first protein MAP. Proteomics. 2012;[Epub ahead of print].

[20] Terracciano R, Pasqua L, Casadonte F, Frasca S, Preiano M, Falcone D, et al. Derivatized mesoporous silica beads for MALDI-TOF MS profiling of human plasma and urine. Bioconjug Chem. 2009;20:913-23.

[21] Fan J, Gallagher JW, Wu HJ, Landry MG, Sakamoto J, Ferrari M, et al. Low molecular weight protein enrichment on mesoporous silica thin films for biomarker discovery. J Vis Exp. 2012.

[22] Bosso N, Chinello C, Picozzi SC, Gianazza E, Mainini V, Galbusera C, et al. Human urine biomarkers of renal cell carcinoma evaluated by ClinProt. Proteomics Clin Appl. 2008;2:1036-46.

[23] Chinello C, Gianazza E, Zoppis I, Mainini V, Galbusera C, Picozzi S, et al. Serum biomarkers of renal cell carcinoma assessed using a protein profiling approach based on ClinProt technique. Urology. 2010;75:842-7.

[24] Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest. 2006;116:271-84.

[25] Xu G, Xiang CQ, Lu Y, Kang XN, Liao P, Ding Q, et al. Application of SELDI-TOF-MS to identify serum biomarkers for renal cell carcinoma. Cancer Lett. 2009;282:205-13.

[26] Xu G, Xiang CQ, Lu Y, Wang WJ, Kang XN, Liao P, et al. SELDI-TOF-MS-based serum proteomic screening in combination with CT scan distinguishes renal cell carcinoma from benign renal tumors and healthy persons. Technol Cancer Res Treat. 2009;8:225-30.

[27] Thompson D, Develter W, Cairns DA, Barrett JH, Perkins DA, Stanley AJ, et al. A pilot study to investigate the potential of mass spectrometry profiling in the discovery of novel serum markers in chronic renal disease. Proteomics Clin Appl. 2011;5:523-31.

[28] Vasudev NS, Sim S, Cairns DA, Ferguson RE, Craven RA, Stanley A, et al. Pre-operative urinary cathepsin D is associated with survival in patients with renal cell carcinoma. Br J Cancer. 2009;101:1175-82.

[29] Wood SL, Rogers M, Cairns DA, Paul A, Thompson D, Vasudev NS, et al. Association of serum amyloid A protein and peptide fragments with prognosis in renal cancer. Br J Cancer. 2010;103:101-11.

[30] Nicolardi S, Palmblad M, Hensbergen PJ, Tollenaar RA, Deelder AM, van der Burgt YE. Precision profiling and identification of human serum peptides using Fourier transform ion cyclotron resonance mass spectrometry. Rapid Commun Mass Spectrom. 2012;25:3457-63.

[31] Refaeilzadeh P, Tang L, Liu H. Cross-Validation, Encyclopedia of Database Systems. Springer. 2009:532-8.

[32] Nagaraj N, Mann M. Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. J Proteome Res. 2011;10:637-45.

[33] Mainini V, Gianazza E, Chinello C, Bilo G, Revera M, Giuliano A, et al. Modulation of urinary peptidome in humans exposed to high altitude hypoxia. Mol Biosyst. 2011;8:959-66.

[34] Tsou CC, Tsai CF, Tsui YH, Sudhir PR, Wang YT, Chen YJ, et al. IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation. Mol Cell Proteomics. 2010;9:131-44.

[35] Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010;17:1471-4.

[36] Swets JA. Measuring the accuracy of diagnostic systems. Science. 1988;240:1285-93.

[37] Tinoco AD, Saghatelian A. Investigating endogenous peptides and peptidases using peptidomics. Biochemistry. 2011;50:7447-61.

[38] Magni F, Van Der Burgt YE, Chinello C, Mainini V, Gianazza E, Squeo V, et al. Biomarkers discovery by peptide and protein profiling in biological fluids based on functionalized magnetic beads purification and mass spectrometry. Blood Transfus. 2010;8 Suppl 3:s92-7.

[39] Latterich M, Abramovitz M, Leyland-Jones B. Proteomics: new technologies and clinical applications. Eur J Cancer. 2008;44:2737-41.

[40] Choudhary J, Grant SG. Proteomics in postgenomic neuroscience: the end of the beginning. Nat Neurosci. 2004;7:440-5.

[41] Schwamborn K, Krieg RC, Grosse J, Reulen N, Weiskirchen R, Knuechel R, et al. Serum proteomic profiling in patients with bladder cancer. Eur Urol. 2009;56:989-96.

[42] Gianazza E, Mainini V, Castoldi G, Chinello C, Zerbini G, Bianchi C, et al. Different expression of fibrinopeptide A and related fragments in serum of type 1 diabetic patients with nephropathy. J Proteomics. 2010;73:593-601.

[43] Tiss A, Smith C, Menon U, Jacobs I, Timms JF, Cramer R. A well-characterised peak identification list of MALDI MS profile peaks for human blood serum. Proteomics. 2010;10:3388-92.

[44] Cheng HL, Huang HJ, Ou BY, Chow NH, Chen YW, Tzai TS, et al. Urinary CD14 as a potential biomarker for benign prostatic hyperplasia - discovery by combining MALDI-TOF-based biostatistics and ESI-MS/MS-based stable-isotope labeling. Proteomics Clin Appl. 2011;5:121-32.

[45] Coombs KM. Quantitative proteomics of complex mixtures. Expert Rev Proteomics. 2011;8:659-77.

[46] Zheng ZG, Yu HQ, Ling ZQ, Mou HZ, Deng HT, Mao WM. Comprehensive profiling of the low molecular weight proteins and peptides in weak cation exchange beads human serum retentate. Protein Pept Lett. 2011;18:498-506.

[47] Svensson M, Skold K, Svenningsson P, Andren PE. Peptidomics-based discovery of novel neuropeptides. J Proteome Res. 2003;2:213-9.

[48] Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics. 2005;5:3226-45.

[49] Schenk S, Schoenhals GJ, de Souza G, Mann M. A high confidence, manually validated human blood plasma protein reference set. BMC Med Genomics. 2008;1:41.

[50] Harper RG, Workman SR, Schuetzner S, Timperman AT, Sutton JN. Low-molecular-weight human serum proteome using ultrafiltration, isoelectric focusing, and mass spectrometry. Electrophoresis. 2004;25:1299-306.

[51] Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD. Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics. 2003;2:1096-103.

[52] Altelaar AF, Mohammed S, Brans MA, Adan RA, Heck AJ. Improved identification of endogenous peptides from murine nervous tissue by multiplexed peptide extraction methods and multiplexed mass spectrometric analysis. J Proteome Res. 2009;8:870-6.

[53] Menschaert G, Vandekerckhove TT, Baggerman G, Schoofs L, Luyten W, Van Criekinge W. Peptidomics coming of age: a review of contributions from a bioinformatics angle. J Proteome Res. 2010;9:2051-61.

[54] Villanueva J, Nazarian A, Lawlor K, Yi SS, Robbins RJ, Tempst P. A sequence-specific exopeptidase activity test (SSEAT) for "functional" biomarker discovery. Mol Cell Proteomics. 2008;7:509-18.

[55] Villanueva J, Martorella AJ, Lawlor K, Philip J, Fleisher M, Robbins RJ, et al. Serum peptidome patterns that distinguish metastatic thyroid carcinoma from cancer-free controls are unbiased by gender and age. Mol Cell Proteomics. 2006;5:1840-52.

[56] Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics. 2005;4:1487-502.

[57] Wang H, Alvarez S, Hicks LM. Comprehensive comparison of iTRAQ and label-free LC-based quantitative proteomics approaches using two Chlamydomonas reinhardtii strains of interest for biofuels engineering. J Proteome Res. 2012;11:487-501.

[58] Nomura F, Tomonaga T, Sogawa K, Ohashi T, Nezu M, Sunaga M, et al. Identification of novel and downregulated biomarkers for alcoholism by surface enhanced laser desorption/ionization-mass spectrometry. Proteomics. 2004;4:1187-94.

[59] Li XJ, Qian CN. Serglycin in human cancers. Chin J Cancer. 2011;30:585-9.

[60] Ueda K, Saichi N, Takami S, Kang D, Toyama A, Daigo Y, et al. A comprehensive peptidome profiling technology for the identification of early detection biomarkers for lung adenocarcinoma. PLoS One. 2011;6:1-12.

[61] Schaub NP, Jones KJ, Nyalwidhe JO, Cazares LH, Karbassi ID, Semmes OJ, et al. Serum proteomic biomarker discovery reflective of stage and obesity in breast cancer patients. J Am Coll Surg. 2009;208:970-8.

[62] Chen YT, Chen HW, Domanski D, Smith DS, Liang KH, Wu CC, et al. Multiplexed quantification of 63 proteins in human urine by multiple reaction monitoring-based mass spectrometry for discovery of potential bladder cancer biomarkers. J Proteomics. 2012:[Epub ahead of print].

[63] Alexander DB, Ichikawa H, Bechberger JF, Valiunas V, Ohki M, Naus CC, et al. Normal cells control the growth of neighboring transformed cells independent of gap junctional communication and SRC activity. Cancer Res. 2004;64:1347-58.

[64] Li X, Jia Z, Shen Y, Ichikawa H, Jarvik J, Nagele RG, et al. Coordinate suppression of Sdpr and Fhl1 expression in tumors of the breast, kidney, and prostate. Cancer Sci. 2008;99:1326-33.

[65] Allinen M, Beroukhim R, Cai L, Brennan C, Lahti-Domenici J, Huang H, et al. Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell. 2004;6:17-32.

[66] Gustincich S, Schneider C. Serum deprivation response gene is induced by serum starvation but not by contact inhibition. Cell Growth Differ. 1993;4:753-60.

[67] Otsuka T, Kohno T, Mori M, Noguchi M, Hirohashi S, Yokota J. Deletion mapping of chromosome 2 in human lung carcinoma. Genes Chromosomes Cancer. 1996;16:113-9.

[68] Izumi Y, Hirai S, Tamai Y, Fujise-Matsuoka A, Nishimura Y, Ohno S. A protein kinase Cdelta-binding protein SRBC whose expression is induced by serum starvation. J Biol Chem. 1997;272:7381-9.

[69] Xu XL, Wu LC, Du F, Davis A, Peyton M, Tomizawa Y, et al. Inactivation of human SRBC, located within the 11p15.5-p15.4 tumor suppressor region, in breast and lung cancers. Cancer Res. 2001;61:7943-9.

[70] Kohno T, Otsuka T, Inazawa J, Abe T, Yokota J. Breakpoint junction of interstitial homozygous deletion at chromosome 2q33 in a small cell lung carcinoma. DNA Res. 1996;3:421-4.

[71] Sanchez-Carbayo M, Socci ND, Charytonowicz E, Lu M, Prystowsky M, Childs G, et al. Molecular profiling of bladder cancer using cDNA microarrays: defining histogenesis and biological phenotypes. Cancer Res. 2002;62:6973-80.

[72] Amsellem V, Kryszke MH, Hervy M, Subra F, Athman R, Leh H, et al. The actin cytoskeleton-associated protein zyxin acts as a tumor suppressor in Ewing tumor cells. Exp Cell Res. 2005;304:443-56.

[73] Hervy M, Hoffman LM, Jensen CC, Smith M, Beckerle MC. The LIM Protein Zyxin Binds CARP-1 and Promotes Apoptosis. Genes Cancer. 2010;1:506-15.

[74] Wu H, Liu T, Wang R, Tian S, Liu M, Li X, et al. MicroRNA-16 targets zyxin and promotes cell motility in human laryngeal carcinoma cell line HEp-2. IUBMB Life. 2011;63:101-8.

[75] Hoffman LM, Jensen CC, Kloeker S, Wang CL, Yoshigi M, Beckerle MC. Genetic ablation of zyxin causes Mena/VASP mislocalization, increased motility, and deficits in actin remodeling. J Cell Biol. 2006;172:771-82.

[76] Call GS, Chung JY, Davis JA, Price BD, Primavera TS, Thomson NC, et al. Zyxin phosphorylation at serine 142 modulates the zyxin head-tail interaction to alter cell-cell adhesion. Biochem Biophys Res Commun. 2011;404:780-4.

[77] Sansing HA, Sarkeshik A, Yates JR, Patel V, Gutkind JS, Yamada KM, et al. Integrin alphabeta1, alphavbeta, alpha6beta effectors p130Cas, Src and talin regulate carcinoma invasion and chemoresistance. Biochem Biophys Res Commun. 2011;406:171-6.

[78] Beyer-Sehlmeyer G, Hiddemann W, Wormann B, Bertram J. Suppressive subtractive hybridisation reveals differential expression of serglycin, sorcin, bone marrow proteoglycan and prostate-tumour-inducing gene I (PTI-1) in drug-resistant and sensitive tumour cell lines of haematopoetic origin. Eur J Cancer. 1999;35:1735-42.

[79] Li XJ, Ong CK, Cao Y, Xiang YQ, Shao JY, Ooi A, et al. Serglycin is a theranostic target in nasopharyngeal carcinoma that promotes metastasis. Cancer Res. 2011;71:3162-72.

[80] Wang ZH, Ding KF, Yu JK, Zhai XH, Ruan SQ, Wang SW, et al. Proteomic analysis of primary colon cancer-associated fibroblasts using the SELDI-ProteinChip platform. J Zhejiang Univ Sci B. 2012;13:159-67.

[81] Nanni P, Mezzanotte L, Roda G, Caponi A, Levander F, James P, et al. Differential proteomic analysis of HT29 Cl.16E and intestinal epithelial cells by LC ESI/QTOF mass spectrometry. J Proteomics. 2009;72:865-73.

[82] Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data: the protein inference problem. Mol Cell Proteomics. 2005;4:1419-40.

Table 1: Patients' clinical characteristics, classified according to the 2009 TNM (tumor-node-metastasis) system classification.

	N° of PATIENTS
ALL	114
Mean ± SD age at diagnosis	64 ± 12.04
Median age at diagnosis (range)	33-87
GENDER	
Males	69
Females	45
STAGING	
Primary Tumor (T)	
pT1	74
pT2	23
pT3	1
pT4	0
unknown	4
Regional lymph nodes (N)	
NX	68
N0	28
N1	1
unknown	5
GRADE	
G1	4
G2	68
G3	21
G4	1
unknown	8
HISTOLOGY	
Clear cell RCC	85
Papillary RCC	9
Clear cell + Papillary RCC	3
Chromophobe	2
Oncocytoma	6
Angiomyolipoma	5
Other subtypes	4
TUMOR TYPE	
Malignant	102
Benign	12

Table 2: Classification performances for each different case/control comparison. A) malignant versus benign kidney lesion plus healthy subjects. B) control subjects versus ccRCC and C) control subjects versus non-ccRCC.

A

m/z	MALIGNANT vs (CTRL + BENIGN)			MALIGNANT vs (CTRL + BENIGN)			
1264		Specificity	Sensitivity		true false	true true	classe precision
1419	TRAINING PHASE	90.4%	94.1%	Pred false	85	5	94.4%
2423	TEST PHASE	83.7%	89.2%	Pred true	9	87	90.6%
2789				spec, sens	90.4%	94.1%	
4962							

B

m/z	CTRL vs ccRCC			CTRL vs ccRCC			
1466		Specificity	Sensitivity		true false	true true	classe precision
4052	TRAINING PHASE	92.4%	91.8%	Pred false	77	6	92.8%
	TEST PHASE	88.1%	84.7%	Pred true	6	71	92.2%
				spec, sens	92.4%	91.8%	

m/z	CTRL vs ccRCC			CTRL vs ccRCC			
1466		Specificity	Sensitivity		true false	true true	classe precision
2378	TRAINING PHASE	96.7%	90.6%	Pred false	80	7	92.00%
2789	TEST PHASE	89.1%	89.4%	Pred true	3	70	95.9%
4052				spec, sens	96.7%	90.6%	

m/z	CTRL vs ccRCC			CTRL vs ccRCC			
1466		Specificity	Sensitivity		true false	true true	classe precision
2378	TRAINING PHASE	95.7%	94.1%	Pred false	79	5	94.00%
2789	TEST PHASE	93.5%	90.6%	Pred true	4	72	94.8%
4052				spec, sens	95.7%	94.1%	
4089							
4151							

C

m/z	CTRL vs NON-ccRCC			CTRL vs NON-ccRCC			
1419		Specificity	Sensitivity		true false	true true	classe precision
2603	TRAINING PHASE	95.7%	89.7%	Pred false	79	3	96.3%
4264	TEST PHASE	94.6%	89.7%	Pred true	4	23	85.2%
				spec, sens	95.7%	89.7%	

Table 3: Overview of identified peptide by nLC-ESI MS/MS in the RPC18 MB pre-fractionated serum of patients and control subjects. Peptide identification was based either on Mascot score above identity (red) or above homology thresholds score (blue). PTM=post-translational modifications (see materials and methods section).

	QUALITY of PEPTIDE IDENTIFICATION in SAMPLES		N° of PEPTIDE IDENTIFICATIONS					
ALL pre-fractionated serum samples	Kidney tumor	Controls	With PTM	Without PTM	TOTAL	With PTM	Without PTM	TOTAL
IDENTITY	IDENTITY	IDENTITY	7	79	86	17	184	201
	IDENTITY	HOMOLOGY	1	13	14			
	IDENTITY	not identified	1	29	30			
	HOMOLOGY	IDENTITY	0	18	18			
	not identified	IDENTITY	8	45	53			
HOMOLOGY	HOMOLOGY	HOMOLOGY	6	19	25	111	151	262
	HOMOLOGY	not identified	76	63	139			
	not identified	HOMOLOGY	29	69	98			
Total (IDENTITY+ HOMOLOGY)						128	335	463

Table 4: Comparison of the peptide ratios obtained analyzing by MALDI and ESI two pools of sera samples. One pool was made with serum from controls (n=80) and one from patients affected by kidney tumor (62 ccRCC+18 non-ccRCC). Peptide identification was based either on Mascot score above identity (red) or above homology thresholds score (blue). Related MALDI FT-ICR m/z values were provided.

MALDI-TOF			MALDI FT-ICR	nLC-ESI MS/MS					
Reflectron mode m/z	Linear mode m/z	Linear mode Ratio (tumor/CTRL)	FT-ICR m/z	Calc .m/z of [M+H]+	Calc. mass (Da) in serum samples pools	Charge state	Sequence	Accesion (UNIPROT)	label-free peptide ratio (tumor/CTRL)
1077.57	1077.72	0.75	1077.533	1077.532	1076.525	2+	GDFLAEGGGVR	FIBA_HUMAN	0.41
1206.61	1206.94	2.54	1206.575	1206.575	1205.568	2+	EGDFLAEGGGVR	FIBA_HUMAN	2.79
1263.63	1263.93	5.76	1263.596	1263.596	1262.589	2+	GEGDFLAEGGGVR	FIBA_HUMAN	2.82
1350.66	1351.06	3.63	1350.628	1350.628	1349.621	2+/3+	SGEGDFLAEGGGVR	FIBA_HUMAN	3.02
1449.77	1450.13	1.02	1449.760	1449.760	1448.752	2+/3+	THRIHWESASLL	CO3_HUMAN	0.78
1465.69	1466.13	4.73	1465.655	1465.655	1464.648	2+/3+	DSGEGDFLAEGGGVR	FIBA_HUMAN	1.01
1536.72	1537.15	1.92	1536.692	1536.692	1535.685	2+/3+	ADSGEGDFLAEGGGVR	FIBA_HUMAN	2.98
1561.77	1562.25	1.85	1561.729	1561.728	1560.721	2+/3+	TATSEYQTFFNPR	THRB_HUMAN	4.79
1739.95	1740.44	1.06	1739.929	1739.930	1738.923	3+	NGFKSHALQLNNRQI	CO4A_HUMAN / CO4B_HUMAN	1.28
1865.02	1865.38	0.98	1865.002	1865.003	1863.995	3+/4+	SSKITHRIHWESASLL	CO3_HUMAN	1.17
1896.04	1896.64	1.15	1896.031	1896.031	1895.020	3+/4+	NGFKSHALQLNNRQIR	CO4A_HUMAN /	1.39

								CO4B_HUMAN	
2021.12	2021.65	1.13	2021.104	2021.104	2020.097	3+/4+	SSKITHRIHWESASLLR	CO3_HUMAN	1.38
2271.11	2271.80	0.40	2271.130	2271.131	2270.123	3+/4+	SRQLGLPGPPDVPDHAAYHPF	ITIH4_HUMAN	0.44
2378.20	2378.33	0.55	2378.209	2378.210	2377.200	3+/4+	DDPDAPLQPVTPLQLFEGRRN	CO4A_HUMAN / CO4B_HUMAN	0.99
2422.11	2422.87	2.94	2422.129	2422.130	2421.123	3+	DSHSLTTNIMEILRGDFSSANN	FIBA_HUMAN	12.25
2423.09	2422.87	2.94	2423.081	2423.082	2422.070	3+/4+	FTSSTSYNRGDSTFESKSYKM	FIBA_HUMAN	2.74
2553.08	2553.51	1.43	2553.101	2553.101	2552.094	3+	SSSYSKQFTSSTSYNRGDSTFES	FIBA_HUMAN	1.48
2602.28	2603.29	0.31	2602.311	2602.311	2601.303	3+/4+	AVPPNNSNAAEDDLPTVELQGVVPR	F13A_HUMAN	0.36
		0.31	2603.168	2603.168	2602.161	3+/4+	SLDRNLPSDSQDLGQHGLEEDFM	SRGN_HUMAN	0.44
2753.40	2753.79	0.65	2753.436	2753.437	2752.430	5+	DAHKSEVAHRFKDLGEENFKALVL	ALBU_HUMAN	0.31
2768.22	2768.58	1.20	2768.228	2768.228	2767.221	3+/4+	SSSYSKQFTSSTSYNRGDSTFESKS	FIBA_HUMAN	0.95
2861.34	2861.87	1.22	2861.334	2861.334	2860.330	5+	MADEAGSEADHEGTHSTKRGHAKSRPV	FIBA_HUMAN	4.07
2931.29	2931.71	1.36	2931.292	2931.292	2930.284	3+/4+	SSSYSKQFTSSTSYNRGDSTFESKSY	FIBA_HUMAN	0.94
	2989.85	1.62	2989.428	2989.429	2988.420	5+	KMADEAGSEADHEGTHSTKRGHAKSRPV	FIBA_HUMAN	6.53
3156.64	3157.34	0.47	3156.630	3156.627	3155.619	4+/5+/6+	NVHSGSTFFKYYLQGAKIPKPEASFSPR	ITIH4_HUMAN	0.23
		0.47	n.d.	3157.616	3156.609	4+	DSGPRRYTIAALLSPYSYSTTAVVTNPKE	TTHY_HUMAN	0.98
3190.45	3190.87	1.49	3190.428	3190.427	3189.420	4+	SSSYSKQFTSSTSYNRGDSTFESKSYKM	FIBA_HUMAN	0.74
3239.57	3240.02	0.86	3239.526	3239.524	3238.517	5+/6+/7+	SYKMADEAGSEADHEGTHSTKRGHAKSRPV	FIBA_HUMAN	1.90
3261.50	3261.85	1.50	3261.465	3261.464	3260.457	4+	SSSYSKQFTSSTSYNRGDSTFESKSYKMA	FIBA_HUMAN	0.73
	4089.11	0.72	4089.021	4089.023	4088.020	6+	TFGSGEADCGLRPLFEKKSLEDKTERELLESYIDGR	THRB_HUMAN	1.41
	4151.45	0.60	4151.162	4151.173	4150.166	4+/5+	TLLVFEVQQPFLFVLWDQQHKFPVFMGRVYDPRA	IC1_HUMAN	0.53
	4786.26	0.60	4785.045	4785.040	4784.030	6+	SSSYSKQFTSSTSYNRGDSTFESKSYKMADEAGSEADHEGTHST	FIBA_HUMAN	0.89
		0.60	n.d.	4787.574	4786.570	6+	LEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK	A1AT_HUMAN	0.47

	506 2.99	1.03	5062 .378	5062. 359	5061 .352	6+/7 +	TFPGFFSPMLGEFVSETESRGSESGIFT NTKESSSHHPGIAEFPSRG	FIBA_H UMAN	0.62
	533 4.92	1.22	5334 .359	5334. 354	5333 .346	6+/8 +	SSSYSKQFTSSTSYNRGDSTFESKSYK MADEAGSEADHEGTHSTKRGHA	FIBA_H UMAN	0.96

Table 5: Serum peptide ratios determined for kidney tumor patients, compared to those reported by Villanueva et al. for advanced prostate, bladder and breast cancer [20] versus healthy subjects. Peptide identification was based either on Mascot score above identity (red) or above homology thresholds score (blue).

MALDI-TOF			nLC-ESI MS/MS	PEPTIDE IDENTIFICATION		Villanueva at al [11]Ratio of medians (patients/controls)		
Reflectron mode monoisotopic m/z	Linear mode monoisotopic m/z	Linear mode Ratio (kidney tumor/CTRL)	Calc. mass (Da) in all pre-fractionated serum samples by nLC-ESI MS/MS	Sequence	Accesion (UNIPROT)	Prostate cancer	Bladder cancer	Breast cancer
1077.565	1077.720	0.750	1076.525	GDFLAEGGGVR	FIBA_HUMAN	0.54	0.50	0.95
1206.609	1206.940	2.540	1205.568	EGDFLAEGGGVR	FIBA_HUMAN	0.50	0.44	0.69
1263.632	1263.930	5.760	1262.589	GEGDFLAEGGGVR	FIBA_HUMAN	0.20	0.24	0.23
1350.662	1351.060	3.634	1349.621	SGEGDFLAEGGGVR	FIBA_HUMAN	0.47	0.46	0.35
1449.767	1450.130	1.019	1448.752	THRIHWESASLL	CO3_HUMAN	1885	2646	437
1465.690	1466.130	4.734	1464.648	DSGEGDFLAEGGGVR	FIBA_HUMAN	0.66	0.55	0.80
1536.717	1537.150	1.924	1535.685	ADSGEGDFLAEGGGVR	FIBA_HUMAN	0.58	0.00	0.54
1739.948	1740.440	1.063	1738.923	NGFKSHALQLNNRQI	CO4A_HUMAN / CO4B_HUMAN	0.75	0.66	2.76
1865.019	1865.380	0.976	1863.995	SSKITHRIHWESASLL	CO3_HUMAN	2.18	3.33	0.30
1896.037	1896.640	1.147	1895.020	NGFKSHALQLNNRQIR	CO4A_HUMAN / CO4B_HUMAN	1.27	3.33	2.95
2021.117	2021.650	1.130	2020.097	SSKITHRIHWESASLLR	CO3_HUMAN	1.05	0.86	0.93
2271.108	2271.800	0.396	2270.123	SRQLGLPGPPDVPDHAAYHPF	ITIH4_HUMAN	4.58	2.64	1.46

2553.083	2553.510	**1.430**	2552.094	SSSYSKQFTSSTSYNRGDSTFES	FIBA_HUMAN	**1.35**	**1.31**	**1.50**
2602.284	2603.290	**0.314**	2601.303	AVPPNNSNAAEDDLPTVELQGVVPR	F13A_HUMAN	**2.84**	**2.31**	**4.73**
2659.242	2659.730	**0.658**	2658.250	DEAGSEADHEGTHSTKRGHAKSRPV	FIBA_HUMAN	**1.37**	**1.18**	**2.45**
2768.223	2768.580	**1.195**	2767.221	SSSYSKQFTSSTSYNRGDSTFESKS	FIBA_HUMAN	**0.77**	**0.03**	**0.98**
2931.291	2931.710	**1.357**	2930.284	SSSYSKQFTSSTSYNRGDSTFESKSY	FIBA_HUMAN	**1.24**	**0.13**	**0.97**
3156.637	3157.340	**0.470**	3155.619	NVHSGSTFFKYYLQGAKIPKPEASFSPR	ITIH4_HUMAN	**3.43**	**10.68**	**1.33**
3190.448	3190.870	**1.485**	3189.420	SSSYSKQFTSSTSYNRGDSTFESKSYKM	FIBA_HUMAN	**0.52**	**0.00**	**0.94**
3239.571	3240.020	**0.862**	3238.517	SYKMADEAGSEADHEGTHSTKRGHAKSRPV	FIBA_HUMAN	**1.00**	**1.00**	**1.00**
3261.497	3261.850	**1.501**	3260.457	SSSYSKQFTSSTSYNRGDSTFESKSYKMA	FIBA_HUMAN	**0.33**	**0.08**	**0.74**

Table 6: Differential expression ratio of proteins, not cited by Villanueva *et al*. [20], with a coherent expression trend of their deriving endogenous peptides determined by label-free approach. Peptide identification was assessed based on Mascot score above 37 (threshold for identity or extensive homology). The ratio was calculated between the two serum pools of 80 controls and 62 ccRCC plus18 non-ccRCC patients.
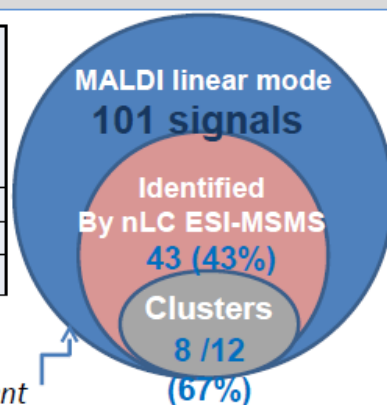
Accession (UNIPROT)	Description	Protein ratio (patients/controls)
SDPR_HUMAN	Serum deprivation-response protein OS=Homo sapiens GN=SDPR PE=1 SV=3	0.56
THRB_HUMAN	Prothrombin OS=Homo sapiens GN=F2 PE=1 SV=2	9.39
ZYX_HUMAN	Zyxin OS=Homo sapiens GN=ZYX PE=1 SV=1	0.31
SRGN_HUMAN	Serglycin OS=Homo sapiens GN=SRGN PE=1 SV=3	0.52
TMSL3_HUMAN	Thymosin beta-4-like protein 3 OS=Homo sapiens GN=TMSL3 PE=2 SV=1	0.47

Graphical abstract

Highlights:

1- Markers for early ccRCC detection and for distinguish malignant kidney tumors.

2- Serum was purified by RPC18 magnetic beads followed by MALDI-TOF and LC-MS/MS.

3- Five signals differentiate malignant tumors from benign renal masses and controls.

4- No statistical difference between peptide ratios by MALDI and label-free strategy.