

Proteomics of liquid biopsies: Depicting RCC infiltration into the renal vein by MS analysis of urine and plasma

Clizia Chinello^{a*}, Martina Stella^a, Isabella Piga^a, Andrew James Smith^a, Giorgio Bovo^b, Marta Varallo^c, Mariia Ivanova^a, Vanna Denti^a, Marco Grasso^d, Angelica Grasso^d, Marina Del Puppo^a, Apostolos Zaravinos^e and Fulvio Magni^a

^a Department of Medicine and Surgery, University of Milano-Bicocca, Clinical Proteomics and Metabolomics Unit, Vedano al Lambro, Italy

^b Pathology Unit, Vimercate Hospital, Vimercate, Italy

^c Pathology Unit, S.Gerardo Hospital, Monza, Italy

^d Urology Unit, S.Gerardo Hospital, Monza, Italy

^e Department of Life Sciences, School of Sciences, European University Cyprus, 1516, Nicosia, Cyprus.

*Correspondence:

Ph.D. Clizia Chinello

University of Milano-Bicocca,

Department of Medicine and Surgery,

Clinical Proteomics and Metabolomics Unit,

Via Raoul Follereau 3,

20854 Vedano al Lambro, Italy

Tel.: +39-02-64488105

Email:clizia.chinello@unimib.it

1 **Proteomics of liquid biopsies: Depicting RCC infiltration into the**
2 **renal vein by MS analysis of urine and plasma**

3

4 Clizia Chinello^{a*}, Martina Stella^a, Isabella Piga^a, Andrew James Smith^a, Giorgio Bovo^b,
5 Marta Varallo^c, Mariia Ivanova^a, Vanna Denti^a, Marco Grasso^d, Angelica Grasso^d, Marina
6 Del Puppo^a, Apostolos Zaravinos^e and Fulvio Magni^a

7 ^a*Department of Medicine and Surgery, University of Milano-Bicocca, Clinical Proteomics*
8 *and Metabolomics Unit, Vedano al Lambro, Italy*

9 ^b*Pathology Unit, Vimercate Hospital, Vimercate, Italy*

10 ^c*Pathology Unit, S.Gerardo Hospital, Monza, Italy*

11 ^d*Urology Unit, S.Gerardo Hospital, Monza, Italy*

12 ^e*Department of Life Sciences, School of Sciences, European University Cyprus, 1516,*
13 *Nicosia, Cyprus.*

14

15

16 *Correspondence:

17 Ph.D. Clizia Chinello

18 University of Milano-Bicocca,

19 Department of Medicine and Surgery,

20 Clinical Proteomics and Metabolomics Unit,

21 Via Raoul Follereau 3,

22 20854 Vedano al Lambro, Italy

23 Tel.: +39-02-64488105

24 Email:clizia.chinello@unimib.it

25

26 **Abstract**

27 Liquid biopsies, as blood and urine, could offer an invaluable, easily accessible source of
28 biomarkers, and evidences for elucidating the pathological processes. Only few studies
29 integrated the proteomes driven by more than one biofluid. Furthermore, it is not clear which
30 biofluid better mirrors the alterations triggered by disease. Venous infiltrating RCC(Renal
31 Cell Carcinoma) could represent an advantageous model for exploring this aspect. Herein, we
32 investigate how blood and urine ‘proteomically’ reflect the changes occurring during RCC
33 infiltration into renal vein(RV) by label-free nLC-ESI-MS/MS. We found 574 and 58
34 differentially expressed proteins(DEPs) in response to vascular involvement. To the augment
35 of vascular involvement, the abundance of only three proteins in
36 urine(UROM,RALA,CNDP1) and two in plasma(APOA1,K2C1) diminished while increased
37 for twenty-six urinary proteins. 80 proteins were found both in urine and plasma, among
38 which twenty-eight were DEPs. A huge overlap between the two biofluids was highlighted,
39 as expected, being urine the filtrate of blood. However, this consistency decreases when RV-
40 occlusion occurs suggesting alternative protein releases, and a loss of kidney architecture.
41 Moreover, several proteomic and functional signatures were biofluid-specific. In conclusion,
42 the complementarity between the specimens allowed to achieve a deeper level of molecular
43 complexity of the RCC venous infiltration.

44

45

46

47 **1. INTRODUCTION**

48 Tumour-derived proteins carried out by biofluids, as blood and urine, could offer an
49 invaluable and non-invasive source of biomarkers, as well as a font of information regarding
50 the numerous pathological processes related to malignant lesions (primary and metastases)
51 and their evolution.

52 Since blood transports most of the tissue-derived molecules in the organism,
53 connecting all the important organs and collecting the related changes, for decades it has been
54 gained the consensus of the researchers as an optimal biological sample for biomarker
55 discovery. This biofluids is very rich of disease related proteins even if they are technically
56 difficult to be mined, due to the about nine orders of magnitude dynamic range that could
57 hide the more specific alterations generated by the pathological processes [1].

58 A less complex medium, such as urine, is an appreciable alternative for screening
59 disease markers, more easy to be collected in large quantities and frequently. Urinary
60 specimen carries a variety of set of soluble proteins and peptides that are primarily derived
61 from kidney, bladder and prostate as well as filtrated by systemic circulation [2]. Given that
62 urinary protein content is likely to reflect normal kidney physiology as well as systemic
63 physiology. Therefore, alterations of the urinary proteome could be used as an indicator of
64 disease not only for urogenital tract and kidneys but potentially also for other organs [3].

65 In comparison to plasma, urine can be collected over a period of time ensuring an
66 easier monitoring of time-dependent changes of biomarker abundance, and resulted quite
67 stable in terms of peptidome/proteome composition since proteolytic degradation may be
68 complete prior to collection [4]. Moreover, urine, differently from blood, is not under the
69 strict regulation of homeostatic mechanisms [5]. In fact, blood could likely represent a
70 worthy place to find alterations associated to disease, especially for the earliest and most
71 sensitive biomarkers. Indeed some of these changes are contrasted and do not stay in blood
72 enough to be detected in time. Thus blood biomarkers are often uncompensated alterations
73 that persist at a rather later stage of a relatively pathological stable condition (i.e. some long
74 half-life proteins or antibody-based biomarkers) [6]. Thus urine, collecting all wastes from
75 the body, can collect a larger number of variations, both huge and severe. Consequently, their
76 concentration is amplified making more visible biomarkers otherwise not detectable in blood
77 [7].

78 Despite the specific drawbacks, blood and urine indeed could be considered as liquid
79 biopsies easily accessible and able to provide the proteomic landscape of the micro- and
80 macro- changes triggered by a neoplasm. Moreover, the integration of the information driven
81 by both biofluids can not only enrich this molecular scenario but also provide some evidences
82 regarding the handling of tumour-derived proteins. Only a few proteomic studies, mainly
83 investigating secreted biomarkers, have so far focused both on blood and urine [8,9]. In this
84 context, an interesting approach was proposed by Jia L [10], who suggested an integrated
85 strategy to explore kidney function in itemized proteomic language. In this perspective,
86 blood, kidney and urine are investigated in the same context, as a system, instead of isolated
87 specimens. Consequently, the related comparison of the input and output sub-proteome
88 permits to speculate whether a particular protein is blocked, or allowed to be secreted/shed
89 from the kidney. Thus, a similar workflow may outline a picture of the function and state of
90 the organ in physiological conditions, and possibly, also when a modification occurs during a
91 disease/neoplasm progression.

92 Beside the above mentioned studies, biomarker discovery is generally performed
93 using serum/plasma or alternatively urine. However, which biological fluid better reflects the
94 pathological changes caused by the disease within the cells, e.g. of the kidney, is not very
95 clear.

96 One of the most distinctive features of renal cell carcinoma (RCC) is its predilection
97 to extend into the venous system including renal vein, inferior vena cava and right atrium.
98 Indeed, the incidence of involvement of the renal vein (RV) and/or inferior vena cava (IVC)
99 has been reported to range between 4 and 15% [11].
100 Even if the prognostic significance of venous involvement and tumor thrombus level still
101 remain controversial, it has been observed that RCCs with venous tumor thrombus (VTT) are
102 more aggressive and associated with poor prognosis [12,13], and the risk of cancer-specific
103 mortality increases in VTT patients with perinephric fat invasion [14].
104 Moreover, VTT could represent a potential middle ground between the phenotype of primary
105 and metastatic RCC, and it was demonstrated that it has a specific molecular trait different
106 from the locally invading tumour and more representative of its extension [15]. For these
107 peculiar characteristics, venous infiltrating RCC could represent a model for investigating the
108 biological information secreted or shed by cancer cell into biofluids when tumour migrates,
109 adapts and begins to spread into circulating system.

110

111 Therefore, we investigate by nLC-ESI MS/MS approach how blood and urine mirror
112 the alterations of the proteome during RCC invasion into the renal vein (RV): moving from
113 the tumour infiltration into the circulating system across the vessel wall of this vein until its
114 complete obstruction.

115

116 **2. MATERIALS & METHODS**

117 **2.1 Reagents:**

118 Trifluoroacetic acid, ammonium bicarbonate, porcine trypsin, DTT (dithiothreitol),
119 IAA (Iodoacetamide), Urea , Ammonium Bicarbonate (ABC), HPLC grade water,
120 acetonitrile, acetone were purchased from Sigma-Aldrich (Sigma-Aldrich Chemie GmbH,
121 Buchs, Switzerland). HPLC-grade water is used for all solutions for MS analysis. Amicon
122 Ultra Centrifugal Filters Ultracel 4 ml 30,000 MW, and Amicon Ultra-0.5 mL 30 kDa were
123 from Millipore.

124

125 **2.2 Sample collection**

126 Urine and plasma samples were collected from patients affected by Renal Cell
127 Carcinoma (RCC) the day before surgery at San Gerardo Hospital (Monza, Italy). All
128 subjects had signed an informed consent prior to sample donation and analyses were carried
129 out in agreement with the Declaration of Helsinki. Study protocols and procedures were
130 approved by the local ethic committee (Comitato Etico Azienda Ospedaliera San Gerardo,
131 Monza, Italy). Second morning midstream urine was collected in sterile urine tubes (Anicrin
132 s.r.l., Italy). After centrifugation at 3000 rpm for 10 min, samples were kept at -80°C [16].
133 Plasma samples were collected in Vacutainer® K3E containing EDTA (Becton Dickinson
134 Italia S.p.A.), centrifuged at 3700rpm for 10 minutes and then stocked at -80°C.

135

136 **2.3 Trypsin digestion by FASP workflow**

137 The enzymatic digestion protocol was based on Filter Aided Sample Preparation
138 (FASP) technique [17]. Before sample processing for LC-MS analysis, equal volumes of
139 plasma samples were pooled according to three different levels of renal vein infiltration

140 (A=vascular infiltration; B=RV infiltration; C=RV thrombosis). Each pool was derived from
141 three different patients. Plasma samples were pooled using same volume before concentration
142 and digestion. Urine samples, instead, due to the inaccuracy of the determination of protein
143 concentration probably for the presence of interfering compounds, were pooled only after
144 trypsinization obtaining equally represented sample in the pools. 3 ml of each plasma pool
145 and urine sample was concentrated using 30 kDa MWCO centrifugal filter unit more than
146 ten-fold. A buffer exchange with water was applied. Protein concentration was determined
147 using bicinchoninic acid assay (Pierce -Thermo Fisher Scientific).

148 In particular, a volume corresponding to 200 µg of proteins for each sample was used
149 both for urine and plasma specimens and mixed with an equal volume of denaturing buffer
150 (0.1M DTT, 4%SDS in Tris HCl 0.1M pH7.6). The solutions were then incubated at 95°C for
151 5 minutes. After disulphide bond reduction, samples were transferred into the ultrafiltration
152 units (Amicon Ultra-0.5 mL 30 kDa, Millipore), made up to 0.5ml with 8 M urea in 100 mM
153 Tris-HCl, pH 8.5 (UA pH8.5 solution), and centrifuged at 14 000 g for 15 min. FASP
154 digestion was performed as already described [17,18]. Briefly, the centrifugation was
155 repeated after adding UA pH8.5 solution to the filter unit. For the alkylation, 200 µl of a
156 0.05M iodoacetamide IAA (Sigma Aldrich) in UA pH8.5 were added and incubated for 20
157 min at dark . Filter units were centrifuged at 14,000 g for 10 min, and submitted to four
158 washes, two of which adding 100 µl of UA pH7.9 solution each, and the remaining two using
159 100 µl of 50 mM Ammonium Bicarbonate (ABC) for each wash (14,000 g for 15 min).
160 Protein digestion was performed overnight at 37 °C adding 2 µg of trypsin. Filtered tryptic
161 peptides were collected in a new tube, and the filters were washed with 50µl of 50 mM ABC
162 and 0.5M NaCl. Tryptic peptides were quantified by NanoDrop assay (Thermo Scientific,
163 Sunnyvale, CA) after acidification with TFA.

164

165 **2.4 nLC-ESI MS/MS label-free quantification**

166 Digested samples were desalted and concentrated using Ziptip™ µ-C8 pipette tips.
167 About 1 µg of peptide mixtures were injected into UHPLC system (Ultimate™ 3000
168 RSLCnano, Thermo Scientific, Sunnyvale, CA) coupled online with Impact HD™ UHR-
169 QqToF (Bruker Daltonics, Germany). Each sample was analysed at least three times to
170 minimize technical variability. Samples were loaded onto a pre-column (Dionex, Acclaim
171 PepMap 100 C18, cartridge, 300 µm) followed by a 50 cm nano-column (Dionex, ID

172 0.075mm, Acclaim PepMap100, C18). The separation was performed at 40°C and at a flow
173 rate of 300 nL/min using multistep 4 hours gradients of acetonitrile as already reported [19].
174 The column was on-line interfaced to a nanoBoosterCaptiveSpray™ ESI source (Bruker
175 Daltonics). Data-dependent-acquisition mode was applied based on CID fragmentation
176 assisted by N₂ as collision gas. Mass accuracy was improved using a specific lock mass
177 (1221.9906 m/z) and a calibration segment (10 mM sodium formate cluster solution) before
178 the beginning of the gradient for each single run. Acquisition parameters was set as already
179 described [20].

180 Data elaboration were performed through DataAnalysis™ v.4.1 Sp4 (Bruker
181 Daltonics, Germany) and protein identification was achieved using an in-house Mascot
182 search engine (version: 2.4.1), through Mascot Daemon tool. Human swissprot database
183 (accessed Feb 2017, 553,655 sequences; 198,177,566 residues) was used. Searching
184 parameters were set as following: Trypsin as enzyme; carbamidomethyl as fixed
185 modifications; 20 ppm as precursor mass tolerances and 0.05 Da for the product ions.
186 Automatic decoy database search was applied for FDR calculation and a built-in Percolator
187 algorithm for rescores peptide-spectrum matches. Only proteins with at least one unique and
188 significant (p-value < 0.05) peptide were considered identified.

189 Progenesis QI for proteomics (Non-linear Dynamics, Newcastle, England) was used
190 as label-free quantification platform as already reported [18]. Briefly, raw data were imported
191 and the ion intensity maps of all runs (9 for each biofluid) used for the alignment process to
192 compensate for between-run variation in the LC separation technique. For granting the
193 maximal overlay across the data, only alignment scores above 60% were accepted. Peak
194 peaking was performed using the default sensitivity and a peak width of 0.2 min. The survey
195 scan data is used for the quantification of peptide ions without MS/MS data. Data is then
196 normalised to all proteins. Protein identification was achieved using an in-house Mascot
197 search engine as described above. Protein abundance was calculated using the sum of all
198 unique peptide normalised ion abundances for that protein on each run. The peptide
199 abundance was based on the sum of the intensities within the isotope boundaries. Fold
200 changes were calculated selecting only non conflicting peptides (unique) in order to provide a
201 more confidently unambiguous read-out of protein abundance, preventing the overlapping of
202 trends derived from different proteins, that shared the same peptides. Statistical tools were
203 used to evaluate the quantitative differences between groups. To indicate the statistical
204 significance of them in group expression data, Anova test was applied (p-value<0.05). For

205 the power analysis and the estimation of sample size, a threshold of 80% was chosen.
206 Moreover, to afford the multiple testing problem, the FDR adjusted p-values, named q-value,
207 is also provided (q-value<0.05).

208

209 **2.5 Bioinformatics analysis**

210 The PANTHER (protein analysis through evolutionary relationship) Classification
211 System [21] version 12.0 (released 2017-07-10) (<http://pantherdb.org>) was utilized for gene
212 ontology (GO) analysis. In particular, PANTHER Statistical overrepresentation test with GO-
213 Slim Biological Process annotation data was applied; GO terms with $p \leq 0.05$ after Bonferroni
214 correction were deemed significant.

215 Differentially expressed proteins (DEPs) were subjected to Core Expression Analysis
216 and investigated for network interrelation by Ingenuity Pathway Analysis (IPA; Qiagen
217 Bioinformatics). IPA scans the set of input proteins to identify networks by using Ingenuity
218 Knowledge Base for interactions between identified “Focus Genes.” The UniProt/Swiss-Prot
219 Accession was used as identifier in the dataset. In this study, the DEPs between Renal Vein
220 invasion (B) and Vascular endothelium infiltration (A), as well as between Renal Vein
221 Thrombosis (C) and Renal Vein invasion (B), along with hypothetical interacting genes
222 stored in the knowledge base in IPA software, were used to generate a set of networks with a
223 maximum network size of 70 genes/proteins. The ratio values in the datasets were converted
224 to fold change values, where the negative inverse (-1/x) was taken for values between 0 and
225 1. Networks were displayed graphically as genes/gene products (“nodes”) and the biological
226 relationships between the nodes (“edges”). All edges are from canonical information stored in
227 the Ingenuity Pathways Knowledge Base. Networks of these genes were generated based on
228 their connectivity and a score ranked each. This score indicates the likelihood of the focus
229 molecules in a network from Ingenuity's knowledge base being found together due to random
230 chance. It is based on the hypergeometric distribution, calculated with the right-tailed
231 Fisher's Exact Test, and corresponds to the negative log of this p-value. A score of Ratio
232 (Expression Fold Change) = 1.5 and p-value (Anova) = 0.05 were set as cutoffs for
233 identifying networks. Furthermore, we used IPA in order to identify the top deregulated
234 molecules and the top canonical pathways in which they participate. In addition, IPA was
235 used to reveal the top molecular and cellular functions, as well as the top upstream regulators,
236 top diseases and biological functions of the DEPs.

238 **3. RESULTS**

239 **3.1 Experimental design**

240 A cohort of nine patients affected by clear cell RCC with vascular infiltration was
241 studied through a quantitative proteomic approach based on nLC-ESI-MS/MS. All of the
242 patients were subjected to surgical nephrectomy and the diagnosis was confirmed by the
243 histological examination. Patients were classified in agreement with the 2009 TNM (tumor-
244 node-metastasis) system classification [22] and to their clinical characteristics
(Supplemental Table 1).

246 Assessment of vascular infiltration was achieved by CT-scan (Computed Assisted
247 Tomography) following morphological description after surgery. Based on these
248 examinations, the dataset was divided into 3 groups according to the level of RCC extension
249 into renal vein: -(A) patients with the evidence of vascular invasion in renal site (not
250 otherwise distributed); -(B) patients with the evidence of vascular invasion in renal site and
251 renal vein invasion; -(C) patients with evidence of renal vein thrombosis. The experimental
252 design was illustrated in **Figure 1.**

253

254 **3.2 Biofluid proteome variation in response to RCC extension into renal vein**

255

256 **3.2.1 Urinary proteome changes**

257 A label-free proteomic approach was applied to urine sample pools in order to
258 identify and quantify urinary proteins whose abundance is significantly different depending
259 on RCC infiltration level into renal vein.

260 From 1207 identified proteins (**Supplemental Table 2**), 574 proteins were observed
261 as differentially expressed in at least one of the three conditions, using the following filters: at
262 least 2 unique peptides; fold change ≥ 1.5 ; anova test p-value ≤ 0.05 ; power ≥ 0.8 ; q-
263 values ≤ 0.05 (**Supplemental Table 3**). The number of proteins identified in each of the runs
264 was reported in **Supplemental Figure 1**. The proteins were then grouped according to their
265 fold changes calculated comparing the three studied conditions. In particular among all
266 possible combinations, four groups (i-with an ascending concentration trend; ii- with a
267 descending concentration trend; iii-an increase of concentration from condition A to B and
268 then a decrease from B to C; iv-a decrease of concentration from condition A to B and then a

269 increase from B to C) were considered to better provide information of the tumour invasion
270 (**Figure 2A**).

271

272 Among these differentially expressed proteins, only three of them (uromodulin, Ras-
273 related protein Ral-A, Beta-Ala-His dipeptidase), diminished proportionally to RCC
274 infiltration while twenty-six proteins seems to be positively influenced by the increase of
275 renal vein involvement (**Figure 2A**). The remaining 318 proteins showed a positive (268) or
276 negative (51) variations of their fold changes at the beginning of the invasion inside the
277 lumen of the vein.

278

279 **3.2.2 Plasma proteome changes**

280 Plasma samples pools were also investigated by label-free LC-ESI-MS/MS relative
281 quantitation in order to highlight differentially expressed proteins in response to RCC
282 vascular infiltration, similarly to previously described approach on urine.

283 156 different proteins were identified (**Supplemental Table 4**) and, among them, 58
284 DEPs were found filtering based on following criteria: at least 2 unique peptides; fold change
285 ≥ 1.5 ; Anova test p-value ≤ 0.05 ; power ≥ 0.8 ; q-values ≤ 0.05 (**Supplemental Table 5**). The
286 number of proteins identified in each of the runs was reported in **Supplemental Figure 1**.

287 Twenty-eight proteins were present in the fourth group according to their fold changes
288 calculated comparing the three studied conditions as above described for urine (**Figure 2B**).
289 None of them showed an increase of their abundance consistently to the augment of RCC
290 infiltration and only two, including Apolipoprotein A-1 and an isoform belonging to keratin
291 family (type II cytoskeletal 1), appeared to be inversely correlated to renal vein invasion
292 (**Figure 2B**).

293

294 **3.2.3 Comparison between urine and plasma proteome alterations**

295 Urinary protein content derived from patients affected by ccRCC at different vascular
296 infiltration levels was compared with related plasma proteome belonged to the same cohort of
297 patients.

298 Eighty proteins identified with at least 2 unique peptides were found to be shared
299 between urine and plasma datasets, equivalent to about 75% of all plasmatic proteins and
300 about 11% of urinary proteins detected in the sample pools (**Figure 3**). Among the common

301 protein IDs, 28 shown a significant variation of their expression (fold change \geq 1.5),
302 comparing the three conditions (A=vascular infiltration; B=RV infiltration; C=RV
303 thrombosis) (**Figure 3A**). In this subset, two proteins, Complement C1s subcomponent and
304 Immunoglobulin heavy constant alpha 2, were shown to be varied in urine only in the
305 comparison of RCC renal vein obstruction (C) respect than the initial RCC vascular invasion
306 (A) (Ratio \geq 1.5 or \leq 0.67). The remaining 26 differentially modulated during RV infiltration
307 and RV thrombosis phase were listed in **Figure 3B**. A high level of concordance of ranging
308 from 58% to 81% has been observed comparing the expression trend (up- or down-
309 regulation) of proteins between urine and plasma (**Figure 3C**). This coherence is remarkably
310 higher (81%) considering only changes belonging to RV invasion (B/A).

311 From a functional point of view, proteins present in this panel are involved mainly in
312 immune-system process and defense (Lactotransferrin, Haptoglobin, Annexin A1
313 Myeloperoxidase, Leukocyte elastase inhibitor, Plastin-2, Annexin A3, Lysozyme C,
314 Annexin A2, Neutrophil defensin 1, Protein S100-A11, Neutrophil elastase, Immunoglobulin
315 lambda-like polypeptide 1, Complement component C1q receptor, CD166 antigen). Some of
316 them are likely to be associated also to protein binding (Annexin A1, Annexin A3, Annexin
317 A2, Protein S100-A11), and to pentose phosphate pathway (Transaldolase and 6-
318 phosphogluconate dehydrogenase, decarboxylating).

319

320

321 **3.3. Functional and network analysis**

322

323 **3.3.1 Biological processes modulated by RCC vascular invasion in urine and plasma**
324 **protein**

325 A meta-analysis based on functional annotation tools was performed in order to
326 highlight which biological process or pathways altered depending on the free RCC extension
327 into renal vein (RV), are reflected by the two biofluids.

328 For this purpose, only proteins with significant changes in their abundance (574 IDs in urine
329 and 58 IDs in plasma as shown in **Figure 2A**) were included. Moreover, to better isolate the
330 changes related to RV invasion, plasma and urine proteomes were grouped into four datasets
331 for each biofluid, taking into account the three possible levels of infiltration based on the

332 experimental design (**Figure 1**). In particular, the lists of DEPs were divided considering only
333 those proteins that resulted consistently down- or up-regulated in RV invasion (condition B
334 respect than condition A), and in RV thrombosis (condition C respect than condition B), as
335 displayed in **Supplemental Figure 2**.

336 Initially, these eight lists were separately submitted to a statistical overrepresentation
337 test on PANTHER gene analysis tool for pinpointing the most significant biological
338 processes enriched during RV infiltration and thrombosis and for evaluating the degree of
339 coherence between urine and plasma proteome from a functional point of view (**Figure 4**).
340 The biological processes in particular were grouped into macro-categories, according to GO-
341 term classification, with the aim of better detecting typical functional traits characterizing
342 tumour vascular invasion steps, and if these traits were represented similarly in urine and
343 plasma.

344 As displayed in **Figure 4**, most of the bioprocesses varied in urine overlapped with those
345 found altered in plasma. Among these shared categories, no inconsistent pattern was shown.
346
347

348 **3.3.2 Ingenuity Pathway Analysis of the liquid biopsies proteomes.**

349 IPA software was used to deeply explore functions and pathways that resulted
350 differently modulated in the biofluids in response to RCC infiltration into the renal vein
351 (**Supplemental Figure 3-6**). Similarly to the previous analysis by Panther search, a
352 comparison between urine and plasma DEPs was carried on considering both the changes
353 occurred in patients with RV invasion and in those whose RV was obstructed. As shown in
354 **Supplemental Figure 3** and **Supplemental Figure 4**, several functional features were shared
355 between patients showing evidence of RCC infiltration in RV, independently from the
356 presence of RV thrombus. However, it has to be noticed that some GO-terms, including
357 networks and molecular functions, appear to be more specific of the level of vein invasion.
358 On the other hand, there was a remarkable overlap in the comparison of the information
359 gathered in urine versus the one received from plasma (**Supplemental Figure 5** and
360 **Supplemental Figure 6**). This level of overlapping is very high in the case of the
361 physiological system development and functions section, while it tends to disappear for top
362 networks and top upstream regulators. Moreover, in terms of pathways consideration, and
363 disease and biofunctions, the concordance between the two liquid biopsies was slightly
364 higher for RV invasion patients in respect to RV thrombosis subjects. On the other hand, the

365 concordance increased in RV thrombosis samples in the case of molecular and cellular
366 functions.

367

368 **4. DISCUSSION**

369

370 The kidney, urine and plasma proteomes are not isolated compartments, rather, they
371 are closely related and could be considered an interconnected system: kidney filtered plasma
372 proteins and waste products into urine via excretory system, and furthermore renal cells may
373 secrete proteins directly into blood or release them into the urine. Many large scale shotgun
374 analyses have investigated the proteome of these specimens, but only few of them have
375 compared the information gathered from these sources, especially for the biofluids [23]. The
376 human proteome atlas for kidney, urine and plasma described by Farrah *et al.* has been built
377 assembling proteins identified using different sources (e.g. glomerulus, urinary exosomes,
378 urine from healthy subjects, etc.) and different analytical approach. These databases result
379 certainly useful because they provide a reliable storage of proteins of different origin.
380 However, they do not compare the proteome from different specimens belonging to the same
381 subject, do not provide information about the relative changes of these proteins in different
382 conditions and do not clarify which pathways or network are more represented comparing
383 blood and urine. Moreover, the proteome data used for comparing the specimens were often
384 obtained by different databases and by diverse analytical methods.

385 One of the first work concerning the integration of more than one specimen was
386 provided by RF Andersen and co-workers [24]. The authors through a nano-LC-MS/MS
387 quantitative approach based on iTRAQ labelling identified DEPs in urine and plasma during
388 childhood idiopathic nephrotic syndrome (NS) compared with remission. About 149 proteins
389 were found to be present in both the biofluids, although none of these shared proteins were
390 observed as significantly altered following NS remission. Li *et al* investigated urinary and
391 plasmatic proteome by LC-MS/MS to determine the best source for a more sensitive
392 detection of protein markers characterizing the effects of two anticoagulants (heparin or
393 argatroban) in six SD rats before and after treatments [7]. Recently, Welton *et al.* applying a
394 semi-quantitative aptamer-based protein array, identified about 1,000 proteins, of which
395 almost 400 were present at comparable quantities in plasma in respect to urinary vesicles [8].
396 Concerning the study of kidney, data integration between different biofluids was concerned
397 more with the study of physiology of this organ [25] than being finalized to enrich the
398 molecular scenario of a specific disease, as renal cancer.

399 Herein, for the first time we applied a shotgun label-free LC MS/MS approach to
400 compare the proteomes between urine and plasma that derives from the same ccRCC patient
401 cohort with different levels of tumour infiltration into the renal vein, from the vascular
402 invasion without the involvement of the RV to the complete occlusion of this vessel. Plasma
403 and urine were collected selecting three subjects for each the three conditions in a wide
404 cohort of RCC patients using a stringent criterion of inclusion. The appropriateness and the
405 risk of the biological averaging assumption in sample pooling must be seriously take into
406 account during the choice of the proteomic workflow, especially for investigations involving
407 class discovery and class prediction in the context of diagnostics and prognostics analysis
408 [26]. Due to the low number of subjects showing the desired defined tumour characteristics,
409 and to the nature of the study far from being diagnostic- or prognostic-oriented, samples were
410 pooled based on the condition and the specimen of origin. For each sample pool at least three
411 technical replicates were analysed and statistical thresholds were considered as described in
412 methods section. Moreover, plasma was preferred to serum and collected avoiding
413 unnecessary manipulation (e.g. depletion), in order to make the biofluid comparison more
414 reliable and reduce exogenous modifications.

415 Overall, 1207 and 156 proteins were identified in urine and in plasma, respectively; while
416 574 urinary and 58 plasma protein IDs were observed as differentially expressed in at least
417 one of the three conditions. The discrepancy of identification power ~10-fold higher in urine
418 could be ascribe to the higher protein dynamic range in blood compared to urine. A lot of
419 approaches have been applied to overcome this issue, including depletion of high abundant
420 proteins (such as albumin or IgG), often coupled with different strategies of off-line peptide
421 fractionation [27,28]. However, to limit the variability of the results that could be introduced
422 by removing the most abundant proteins, and also to keep the quantitation more reliable, we
423 decided to analyse the two biofluids using the same protocol. The message that urine reflect
424 more information remains valid despite the low identification power for plasma.

425

426 A panel of 26 urinary proteins were found to be directly correlated with the extension
427 of RCC into RV, showing an increase in their abundance levels, parallel with those related to
428 the infiltration level. This trend was not observed in plasma, probably due to the limited
429 protein number. The panel of positive markers in urine is largely composed by proteins
430 implicated in biological process that seem strongly related to the tumour invasion,

431 inflammatory process, and energetic metabolism, as described in results paragraph 3.2.3.
432 Only three proteins (uromodulin, Ras-related protein Ral-A, Beta-Ala-His dipeptidase) in
433 urine and two in plasma (Apolipoprotein A-1 and a keratin type II cytoskeletal 1) are
434 negatively influenced by the increase of infiltrative process. Interestingly, Ral-A GTPase was
435 reported to be associated with advanced kidney cancer, being involved into malignancy
436 invasion processes, through a signal pathway induced by proinflammatory cytokine
437 prostaglandin E2 (PGE2) [29]. Similarly, Apolipoprotein A-1 (APOA1) is shown to be
438 correlated with RCC prognosis in agreement with the findings of a recent investigation that
439 demonstrated in a retrospectively study of 786 patients with RCC that a low APOA1 serum
440 level has been associated to a worse overall survival and to shortened disease-free survival
441 [30]. Moreover, activity of Beta-Ala-His dipeptidase encoded by CNDP1 gene has been
442 observed to be correlated with a potential long term protection of complications linked to
443 reactive metabolites accumulating, e.g. in diabetes and chronic renal failure [31].

444 The comparison of protein content identified in the two biofluids highlighted a huge
445 overlap between plasma and urine, being about three-quarters of all plasmatic proteins
446 included in urine dataset and about half of DEPs found in plasma (**Figure 3A**). This could be
447 expected since urine is mainly the result of blood filtration encompassing the most abundant
448 and less represented proteins. However, this overlap allowed us to gain a new insight from a
449 different perspective about the pathological processes connected to the RCC vascular
450 invasion.

451 The list of biofluid-shared DEPs, included a high percentage of immunoglobulin (about
452 30%), components and factors of complement cascade, modulators of acute-phase response
453 and defense involved in complement activation, innate immune-system, platelet
454 degranulation and scavenging of heme from plasma (**Supplemental Table 6**).
455 Overall, they are consistently regulated if we compared plasma and urine, and some of them
456 appear to be significantly related to the infiltration grade of RCC (**Figure 3B**). However, if
457 we perform this comparison considering the different steps of the extension of RCC, the
458 percentage of variability appear different. In fact, more advanced is the stage of vascular
459 invasion, more discrepancies between blood and urine are present (**Figure 3C**). A possible
460 explanation for this behaviour could be found considering the 3D-development of the tumour
461 mass. In fact, urinary proteome alterations could be influenced by the extent of RCC not only
462 into vascular system but also into the organ itself, which can lose its architecture. If kidney
463 structure and function is damaged, it can be supposed that non-regulated protein deliveries to

464 renal basin can be present. If we consider the urinary albumin level, generally associated with
465 functional status of the glomerular filtration barrier, no statistically significant proteomic
466 variation is detectable comparing the three groups. However, the histological exams
467 displayed a very advanced tumour progression in the three patients showing RV thrombosis.
468 Therefore, an impairment of kidney functionality cannot be excluded in the surrounding
469 tumour area.

470 Alterations associated with disease generally require an entire set of effectors to be
471 completed. By now, modern proteomic approaches despite conspicuous advancements can
472 provide only a partial list of them. A comprehensive study of regulatory networks and
473 pathways could compensate these lacks and effectively increase the understanding of the
474 intricate system of functions that are turned on or off during disease process. This contribute
475 is more evident if we apply an integrated strategy. Therefore, DEPs datasets found in urine
476 and plasma were subjected to a functional classification and outcomes were visualized
477 filtering both RV infiltration steps and biofluid type. Firstly, it can be noticed that the
478 biological processes shared between urine and plasma showed a high grade of concordance in
479 terms of positive or negative regulation (**Supplemental Figure 7**). No inconsistency was
480 revealed between urine and plasma supporting the hypothesis that urine is a good mirror of
481 what is happening in blood. Moreover, it is likely that both urine and blood carry a specific
482 ‘biofluid functional signature’. Several processes were enriched mainly (e.g. glucose and
483 RNA metabolism, catabolic process, adhesion), or exclusively in urine (such as transcription
484 related processes). Others are likely more represented in plasma (e.g. B-cell mediated
485 immunity or blood coagulation).

486 Furthermore, if we integrate the information carried by the two biofluids, also a
487 functional signatures associated to biological processes could be mined. To better visualize
488 them, the enriched biological processes were categorized into macro-groups using
489 PANTHER gene-ontology (**Figure 4**). Results suggest specific trends characterised by
490 processes that are on or off depending on the infiltration phase. Proteins related to catabolic
491 processes, proteolysis and cell to cell adhesion were up-represented both in RV infiltration
492 and RV thrombosis, while energetic metabolic processes including glycolysis and regulation
493 of transcription appeared on during RV invasion and down-represented when RV is occluded.
494 On the other hand, in patients with the evidence of RCC thrombosis, immunity system related
495 proteins, including those involved in the complement activation, and defence mechanisms,

496 endocytosis and cell recognition were found significantly increased and proteins implicated
497 in blood circulation decreased.
498 These data were also confirmed by the functional annotation classification provided by
499 Ingenuity Pathway analysis (**Supplemental Figure 3-6**). Despite a certain overlap between
500 urine and plasma, only the combination of the two datasets permits to highlight specific traits
501 of renal cancer vascular invasion. On one side, the analysis underlined as already suggested
502 that ccRCC is basically a metabolic disorder, since malignant cells handle a number of
503 biological pathways to achieve their aggressive phenotype and spread into circulating system
504 [32]. On the other side, particularly for these infiltrative forms, ccRCC is likely to behave as
505 an immunological disease, involving immune cell trafficking, humoral immune response and
506 positive and negative acute response.

507

508 In conclusion, the comparison between the functional classification of urine and
509 plasma proteome confirms the complementary of the information delivered by these biofluids
510 and shed light to those processes and pathways that are likely to be switched on or off during
511 malignancy spreading into renal vein. On the other hand, it suggests that the loss of kidney
512 architecture during advanced stages of tumour growth could have a detectable repercussion to
513 biofluids proteome. Therefore the integration of information between urine and plasma
514 changes at a proteomic level may provide a more complete landscape of such a dynamic
515 system as growing cancer cells are, also from a functional point of view.

516

517 **ACKNOWLEDGMENTS**

518 This work was supported by grants from the MIUR: FIRB 2007 (RBRN07BMCT_11), FAR
519 2013–2016; and in part by Fondazione Gigi & Pupa Ferrari Onlus.

520

521 **References:**

- 522 [1] M. Rainer, C. Sajdik, G.K. Bonn, Mass spectrometric profiling of low-molecular-weight
523 proteins, *Methods Mol. Biol.* 1023 (2013) 83–95. doi:10.1007/978-1-4614-7209-4-5.
- 524 [2] C. Chinello, V. L’Imperio, M. Stella, A.J. Smith, G. Bovo, A. Grasso, M. Grasso, F.
525 Raimondo, M. Pitto, F. Pagni, F. Magni, The proteomic landscape of renal tumors, *Expert*
526 *Rev. Proteomics.* 13 (2016). doi:10.1080/14789450.2016.1248415.
- 527 [3] A. Di Meo, I. Batruch, A.G. Yousef, M.D. Pasic, E.P. Diamandis, G.M. Yousef, An integrated
528 proteomic and peptidomic assessment of the normal human urinome, *Clin. Chem. Lab. Med.*
529 55 (2017) 237–247. doi:10.1515/cclm-2016-0390.
- 530 [4] H. Husi, R.J.E. Skipworth, A. Cronshaw, K.C.H. Fearon, J.A. Ross, Proteomic identification
531 of potential cancer markers in human urine using subtractive analysis, *Int. J. Oncol.* 48 (2016).
532 doi:10.3892/ijo.2016.3424.
- 533 [5] M. Li, Urine reflection of changes in blood, *Adv. Exp. Med. Biol.* 845 (2015) 13–19.
534 doi:10.1007/978-94-017-9523-4_2.
- 535 [6] Y. Gao, Urine is a better biomarker source than blood especially for kidney diseases, *Adv.*
536 *Exp. Med. Biol.* 845 (2015) 3–12. doi:10.1007/978-94-017-9523-4_1.
- 537 [7] M.L. Li, M.D. Zhao, Y.H. Gao, Changes of proteins induced by anticoagulants can be more
538 sensitively detected in urine than in plasma, *Sci. China Life Sci.* 57 (2014) 649–656.
539 doi:10.1007/s11427-014-4661-y.
- 540 [8] J.L. Welton, P. Brennan, M. Gurney, J.P. Webber, L.K. Spary, D.G. Carton, J.M. Falcón-
541 Pérez, S.P. Walton, M.D. Mason, Z. Tabi, A. Clayton, Proteomics analysis of vesicles isolated
542 from plasma and urine of prostate cancer patients using a multiplex, aptamer-based protein
543 array, *J. Extracell. Vesicles.* 5 (2016). doi:10.3402/jev.v5.31209.
- 544 [9] A. Katafigioti, I. Katafigiotis, S. Sfougaristos, C. Alamanis, K. Stravodimos, I. Anastasiou, E.
545 Roumelioti, M. Duvdevani, C. Constantinides, In the search of novel urine biomarkers for the
546 early diagnosis of prostate cancer. Intracellular or secreted proteins as the target group? Where
547 and how to search for possible biomarkers useful in the everyday clinical practice, *Arch. Ital.*
548 *Di Urol. E Androl.* 88 (2016) 195–200. doi:10.4081/aiua.2016.3.195.
- 549 [10] L. Jia, Comparing plasma and urinary proteomes to understand kidney function, *Adv. Exp.*
550 *Med. Biol.* 845 (2015) 187–193. doi:10.1007/978-94-017-9523-4_18.
- 551 [11] H. Miyake, T. Terakawa, J. Furukawa, M. Muramaki, M. Fujisawa, Prognostic significance of
552 tumor extension into venous system in patients undergoing surgical treatment for renal cell
553 carcinoma with venous tumor thrombus., *Eur. J. Surg. Oncol.* 38 (2012) 630–6.
554 doi:10.1016/j.ejso.2012.03.006.
- 555 [12] Q. Tang, Y. Song, X. Li, M. Meng, Q. Zhang, J. Wang, Z. He, L. Zhou, Prognostic outcomes
556 and risk factors for patients with renal cell carcinoma and venous tumor thrombus after radical
557 nephrectomy and thrombectomy: The prognostic significance of venous tumor thrombus level,
558 *Biomed Res. Int.* 2015 (2015). doi:10.1155/2015/163423.
- 559 [13] M.W. Ball, M.A. Gorin, K.T. Harris, K.M. Curtiss, G.J. Netto, C.P. Pavlovich, P.M.
560 Pierorazio, M.E. Allaf, Extent of renal vein invasion influences prognosis in patients with
561 renal cell carcinoma, *BJU Int.* 118 (2016) 112–117. doi:10.1111/bju.13349.
- 562 [14] V.L. Weiss, M. Braun, S. Perner, A. Harz, R. Vorreuther, G. Kristiansen, S.C. Müller, J.
563 Ellinger, Prognostic significance of venous tumour thrombus consistency in patients with renal
564 cell carcinoma (RCC), *BJU Int.* 113 (2014) 209–217. doi:10.1111/bju.12322.
- 565 [15] A. Laird, F.C. O’Mahony, J. Nanda, A.C.P. Riddick, M. O’Donnell, D.J. Harrison, G.D.

- 566 Stewart, Differential Expression of Prognostic Proteomic Markers in Primary Tumour, Venous
567 Tumour Thrombus and Metastatic Renal Cell Cancer Tissue and Correlation with Patient
568 Outcome, PLoS One. 8 (2013). doi:10.1371/journal.pone.0060483.
- 569 [16] N. Bosso, C. Chinello, S.C.M. Picozzi, E. Gianazza, V. Mainini, C. Galbusera, F. Raimondo,
570 R. Perego, S. Casellato, F. Rocco, S. Ferrero, S. Bosari, P. Mocarelli, M.G. Kienle, F. Magni,
571 Human urine biomarkers of renal cell carcinoma evaluated by ClinProt, Proteomics - Clin.
572 Appl. 2 (2008) 1036–1046. doi:10.1002/prca.200780139.
- 573 [17] J.R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method
574 for proteome analysis, Nat Methods. 6 (2009) 359–362. doi:10.1038/nmeth.1322\rmeth.1322
575 [pii].
- 576 [18] F. Raimondo, S. Corbetta, A. Savoia, C. Chinello, M. Cazzaniga, F. Rocco, S. Bosari, M.
577 Grasso, G. Bovo, F. Magni, M. Pitto, Comparative membrane proteomics: A technical
578 advancement in the search of renal cell carcinoma biomarkers, Mol. Biosyst. 11 (2015).
579 doi:10.1039/c5mb00020c.
- 580 [19] C. Chinello, M. Cazzaniga, G. Sio, A.J. Smith, A. Grasso, B. Rocco, S. Signorini, M. Grasso,
581 S. Bosari, I. Zoppis, G. Mauri, F. Magni, Tumor size, stage and grade alterations of urinary
582 peptidome in RCC, J. Transl. Med. 13 (2015). doi:10.1186/s12967-015-0693-8.
- 583 [20] X. Liu, C. Chinello, L. Musante, M. Cazzaniga, D. Tataruch, G. Calzaferri, A.J. Smith, G. De
584 Sio, F. Magni, H. Zou, H. Holthofer, Intraluminal proteome and peptidome of human urinary
585 extracellular vesicles, Proteomics - Clin. Appl. 9 (2015) 568–573.
586 doi:10.1002/prca.201400085.
- 587 [21] H. Mi, A. Muruganujan, P.D. Thomas, PANTHER in 2013: Modeling the evolution of gene
588 function, and other gene attributes, in the context of phylogenetic trees, Nucleic Acids Res. 41
589 (2013). doi:10.1093/nar/gks1118.
- 590 [22] B. Ljungberg, K. Bensalah, S. Canfield, S. Dabestani, F. Hofmann, M. Hora, M.A. Kuczyk, T.
591 Lam, L. Marconi, A.S. Merseburger, P. Mulders, T. Powles, M. Staehler, A. Volpe, A. Bex,
592 EAU guidelines on renal cell carcinoma: 2014 update, Eur. Urol. 67 (2015) 913–924.
593 doi:10.1016/j.eururo.2015.01.005.
- 594 [23] T. Farrah, E.W. Deutsch, G.S. Omenn, Z. Sun, J.D. Watts, T. Yamamoto, D. Shteynberg,
595 M.M. Harris, R.L. Moritz, State of the human proteome in 2013 as viewed through
596 peptideatlas: Comparing the kidney, urine, and plasma proteomes for the biology- and disease-
597 driven human proteome project, J. Proteome Res. 13 (2014) 60–75. doi:10.1021/pr4010037.
- 598 [24] R.F. Andersen, J. Palmfeldt, B. Jespersen, N. Gregersen, S. Rittig, Plasma and urine proteomic
599 profiles in childhood idiopathic nephrotic syndrome, Proteomics - Clin. Appl. 6 (2012) 382–
600 393. doi:10.1002/prca.201100081.
- 601 [25] L. Jia, L. Zhang, C. Shao, E. Song, W. Sun, M. Li, Y. Gao, An attempt to understand kidney's
602 protein handling function by comparing plasma and urine proteomes, PLoS One. 4 (2009).
603 doi:10.1371/journal.pone.0005146.
- 604 [26] A.L. Oberg, O. Vitek, Statistical design of quantitative mass spectrometry-based proteomic
605 experiments, J. Proteome Res. (2009). doi:10.1021/pr8010099.
- 606 [27] H. Keshishian, M.W. Burgess, H. Specht, L. Wallace, K.R. Clouser, M.A. Gillette, S.A. Carr,
607 Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker
608 discovery by mass spectrometry, Nat. Protoc. (2017). doi:10.1038/nprot.2017.054.
- 609 [28] S. Chutipongtanate, S. Chatthen, J. Svasti, Plasma prefractionation methods for proteomic
610 analysis and perspectives in clinical applications, Proteomics - Clin. Appl. (2017).
611 doi:10.1002/prca.201600135.

- 612 [29] Z. Li, Y. Zhang, W.J. Kim, Y. Daaka, PGE2 promotes renal carcinoma cell invasion through
613 activated RalA, *Oncogene*. 32 (2013) 1408–1415. doi:10.1038/onc.2012.161.
- 614 [30] S. Guo, X. He, Q. Chen, G. Yang, K. Yao, P. Dong, Y. Ye, D. Chen, Z. Zhang, Z. Qin, Z. Liu,
615 Z. Li, Y. Xue, M. Zhang, R. Liu, F. Zhou, H. Han, The Effect of Preoperative Apolipoprotein
616 A-I on the Prognosis of Surgical Renal Cell Carcinoma: A Retrospective Large Sample
617 Study., *Medicine (Baltimore)*. 95 (2016) e3147. doi:10.1097/MD.0000000000003147.
- 618 [31] V. Peters, J. Zschocke, C.P. Schmitt, Carnosinase, diabetes mellitus and the potential relevance
619 of carnosinase deficiency, *J. Inherit. Metab. Dis.* (2017). doi:10.1007/s10545-017-0099-2.
- 620 [32] A. Zaravinos, M. Pieri, N. Mourmouras, N. Anastasiadou, I. Zouvani, D. Delakas, C. Deltas,
621 Altered metabolic pathways in clear cell renal cell carcinoma: A meta-analysis and validation
622 study focused on the deregulated genes and their associated networks., *Oncoscience*. 1 (2014)
623 117. doi:10.18632/oncoscience.13.

624

Figure legends

Figure 1: Experimental design and workflow

Figure 2: Protein expression trend correlated to RCC extension into RV (RV=Renal Vein, A= ccRCC patients that show vascular infiltration, B= ccRCC patients that show tumour infiltration into renal vein; C= ccRCC patients that show renal vein thrombosis) in urine (**panel A**) and in plasma (**panel B**)

Figure 3: **A-** Number of proteins isoforms identified and quantified in urine and plasma samples through nLC-ESI MS/MS. Data were elaborated through Progenesis platform and a Venn Diagram between urine and plasma outcomes is shown both for identified proteins and for differentially regulated proteins (Ratio ≥ 1.5). **B-** DEPs in common between urine and plasma samples. **C-** Proteome expression consistency between urine and plasma samples regarding shared DEPs.

Figure 4: Overview of the biological processes enriched in urine and plasma in RCC patients with RV infiltration or RV thrombosis, using overrepresentation test on PANTHER gene analysis tool ($p < 0.05$). The up- or down-regulation referred to condition B respect to condition A for RV infiltration label (pink), and condition C respect to condition B for RV Thrombosis label (purple). # = number of genes belonging to related PANTHER GO-Slim Biological process

Supplemental Figure legends

Supplemental Figure 1: The charts show the number of proteins identified in each of the runs both for urine and plasma. Each run is shown according to its experimental condition (A= ccRCC patients that show vascular infiltration, B= ccRCC patients that show tumour infiltration into renal vein; C= ccRCC patients that show renal vein thrombosis)

Supplemental Figure 2: Venn diagram of identified and quantified proteins in plasma and urine. The proteins are grouped depending on their up- or down-regulation in patients showing RCC infiltrated into RV respect than others that showed RCC vascular infiltration without RV involvement (RV infiltration), and in subjects with RV thrombosis respect than the ones with RV infiltration (RV Thrombosis).

Supplemental Figure 3: Summary of some of the top functional categories performed by Ingenuity Pathway Analysis on DEPs during the RV invasion (BtoA) and RV thrombosis (CtoB) in plasma. The up- or down-regulation referred to condition B respect to condition A for RV infiltration label (pink), and condition C respect to condition B for RV Thrombosis label (purple). Green labels refer to the matches comparing top functional categories of RV invasion (BtoA) vs RV Thrombosis (CtoB) in plasma.

Supplemental Figure 4: Summary of some of the top functional categories performed by Ingenuity Pathway Analysis on DEPs during the RV invasion (BtoA) and RV thrombosis (CtoB) in urine. The up- or down-regulation referred to condition B respect to condition A for RV infiltration label (pink), and condition C respect to condition B for RV Thrombosis label (purple). Green labels refer to the matches comparing top functional categories of RV invasion (BtoA) vs RV Thrombosis (CtoB) in urine.

Supplemental Figure 5: Summary of some of the top functional categories performed by Ingenuity Pathway Analysis on DEPs during the RV invasion (BtoA) in plasma and urine. Yellow labels refer to the matches comparing top functional categories of urine vs plasma for RV invasion (BtoA).

Supplemental Figure 6: Summary of some of the top functional categories performed by Ingenuity Pathway Analysis on DEPs during the RV invasion (BtoA) in plasma and urine. Yellow labels refer to matches comparing top functional categories of urine vs plasma for RV Thrombosis (CtoB)).

Supplemental Figure 7: Biological processes enriched both in urine and plasma, or only in urine, or only in plasma related to RCC RV infiltration or RCC RV thrombosis.

Overrepresentation test on PANTHER gene analysis tool ($p < 0.05$) was used. The up- or down-regulation referred to condition B respect to condition A for RV infiltration label (pink), and condition C respect to condition B for RV Thrombosis label (purple). # = number of genes belonging to related PANTHER GO-Slim Biological process

Supplemental Table legends

Supplemental Table 1: Clinical characteristics of patient cohort, using 2009 TNM (tumor-node-metastasis) system classification. M=male. F=female.

Supplemental Table 2: list of all urinary proteins identified with at least one unique peptide. ‘Peptide number’ refers to the number of peptide identifying the related protein. ‘Unique peptides’ refer to number of peptides unique to that protein. Confidence score refers to combined protein score for all peptide (calculated using Progenesis QI for proteomics).

Supplemental Table 3: list of urinary proteins that resulted significantly ($p < 0.05$) varied in at least one of the three conditions (A= ccRCC patients that show vascular infiltration, B= ccRCC patients that show tumour infiltration into renal vein; C= ccRCC patients that show renal vein thrombosis).Score, p-value and normalized abundancies were calculated using Progenesis QI for proteomics. Peptide count’ refers to the number of peptide identifying the related protein. ‘Unique peptide’ refers to number of peptides unique to that protein, not belonging to another protein hit.

Supplemental Table 4: list of all plasma proteins identified with at least one unique peptide. ‘Peptide number’ refers to the number of peptide identifying the related protein. ‘Unique peptides’ refer to number of peptides unique to that protein. Confidence score refers to combined protein score for all peptide (calculated using Progenesis QI for proteomics).

Supplemental Table 5: list of plasma proteins that resulted significantly ($p < 0.05$) varied in at least one of the three conditions (A= ccRCC patients that show vascular infiltration, B= ccRCC patients that show tumour infiltration into renal vein; C= ccRCC patients that show renal vein thrombosis).Score, p-value and normalized abundancies were calculated using Progenesis QI for proteomics. Peptide count’ refers to the number of peptide identifying the related protein. ‘Unique peptide’ refers to number of peptides unique to that protein, not belonging to another protein hit.

Supplementary Table 6: PANTHER GO-Slim Biological processes and PANTHER Reactome pathways enrichment of the 26 protein IDs shared between urine and plasma samples. Bonferroni correction for multiple testing was used. # = number of genes included in the related PANTHER GO-process or pathway. +/- refers to positive/negative significance of results.

Figure 1

[Click here to download high resolution image](#)

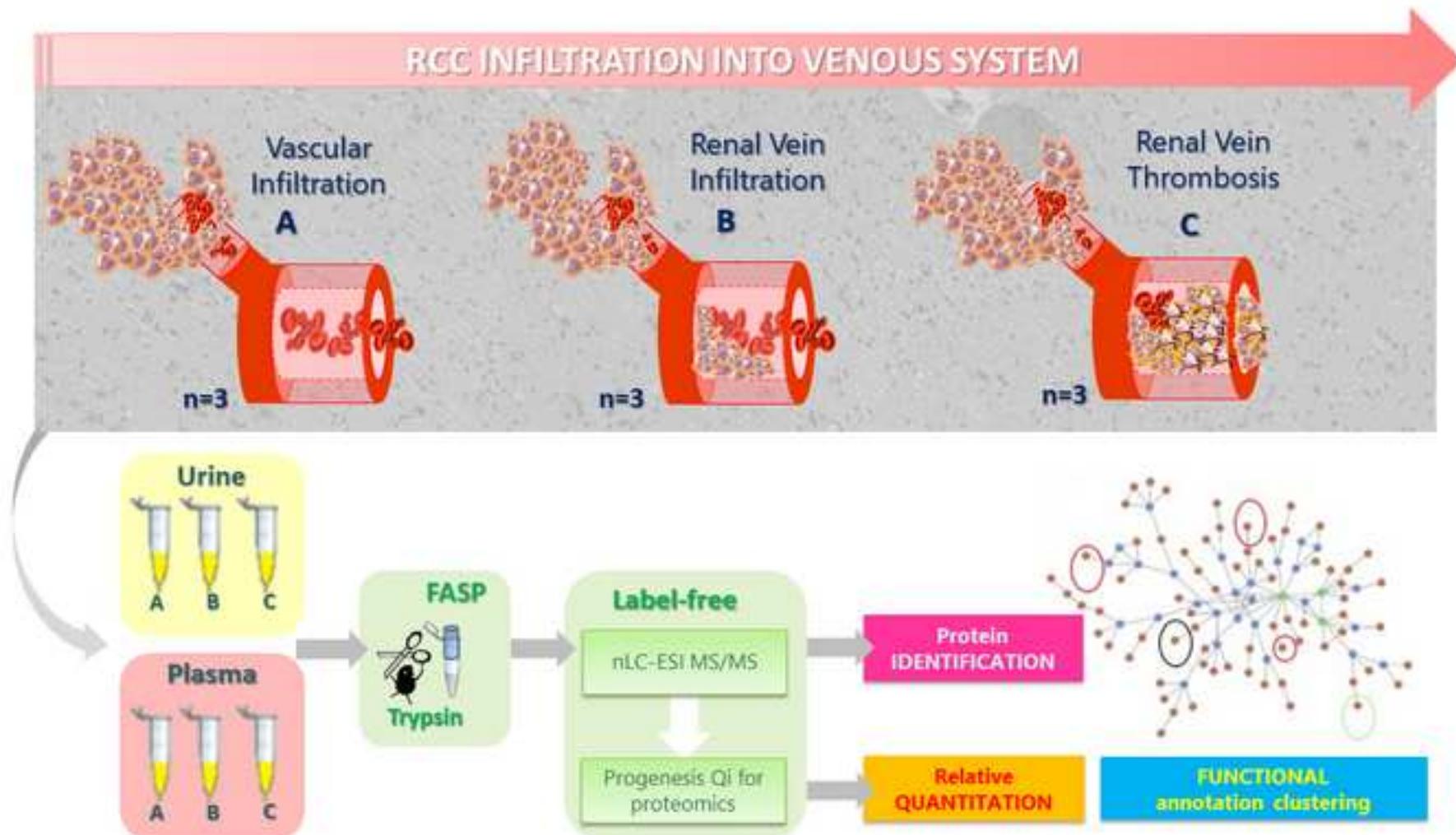


Figure 2

[Click here to download high resolution image](#)

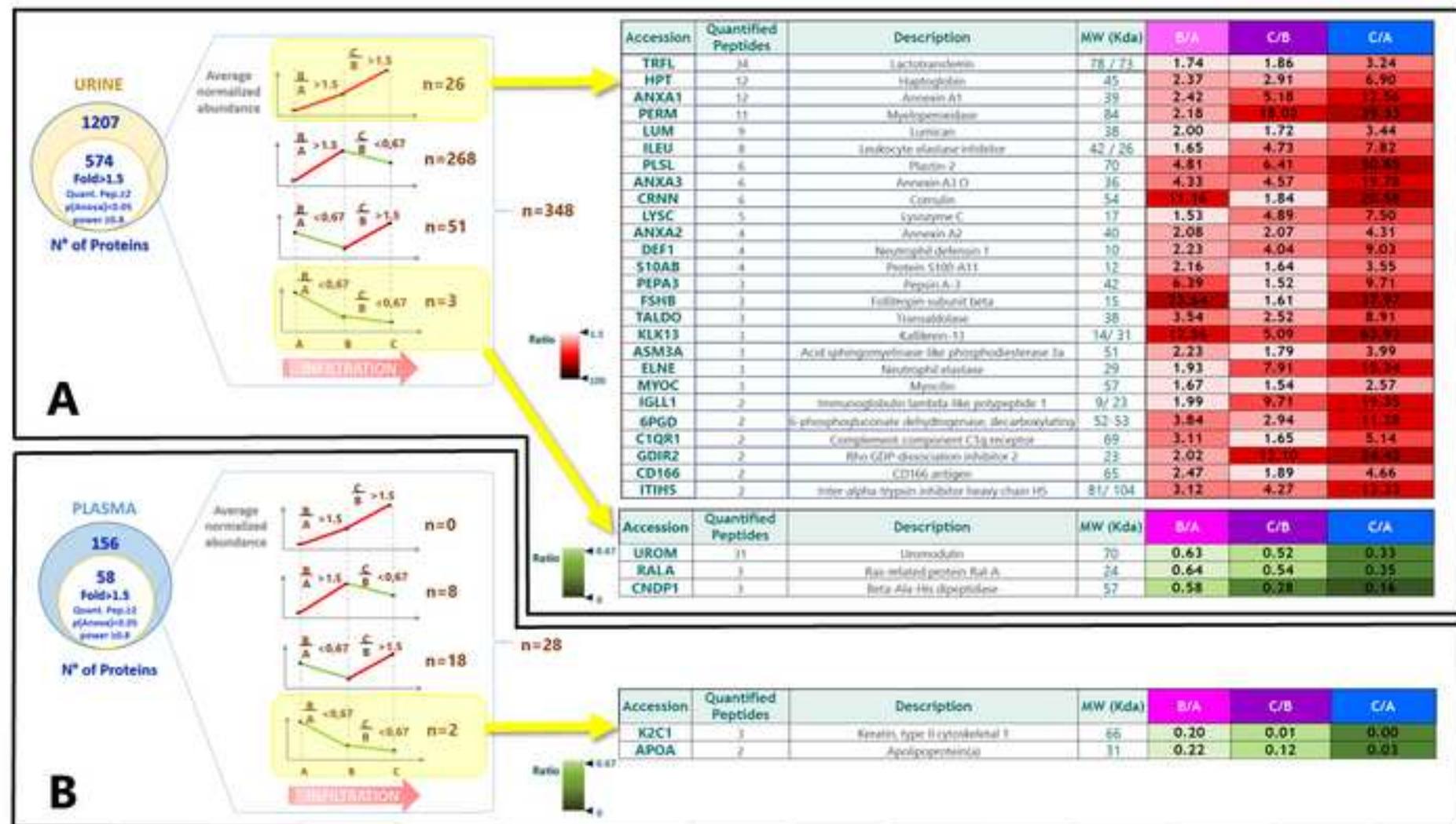
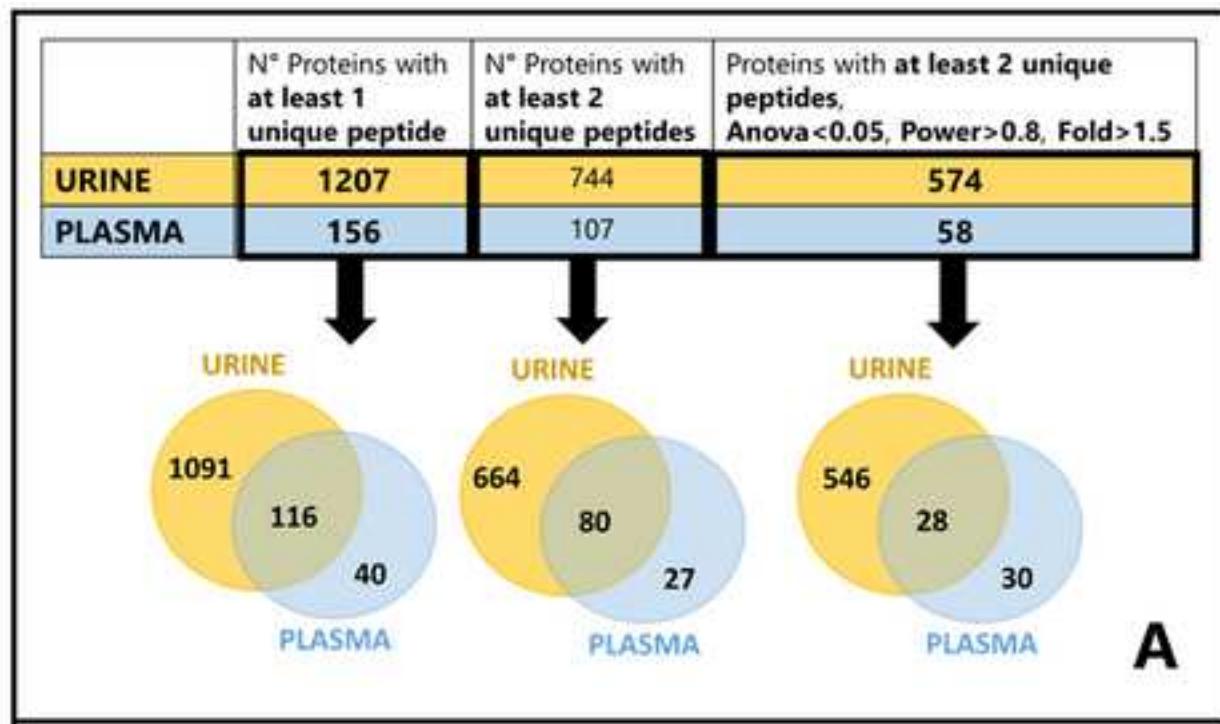


Figure 3

[Click here to download high resolution image](#)



Protein ID	Description	URINE		PLASMA		Ratio
		Ratio (B/A)	Ratio (C/B)	Ratio (B/A)	Ratio (C/B)	
HPT	Haptoglobin	1.31	0.69	0.40	1.46	>2
TTHY	Transthyretin	0.04	0.61	0.75	0.65	1.5<x<2
KALN	Kallistatin SERPINA4	1.67	0.41	2.48	0.53	0.5<x<0.67
TETN	Tetranectin	2.81	0.37	1.63	0.63	<0.5
PGRP2	N-acetylmuramoyl-L-alanine amidase	2.31	0.60	1.77	0.82	0.67<x<1.5
FINC	Fibronectin	1.79	0.67	0.86	0.50	>2
APOE	Apolipoprotein E	1.64	0.47	1.94	1.05	1.5<x<2
ITIH2	Inter-alpha-trypsin inhibitor heavy chain H2	0.99	0.74	1.75	0.84	0.5<x<0.67
IGJ	Immunoglobulin J chain	0.93	1.16	1.69	1.35	<0.5
IGHA1	Ig alpha-1 chain C region	1.50	1.22	0.54	1.84	0.67<x<1.5
C99	Complement component C9	0.67	1.62	0.39	2.26	>2
IGHG4	Ig gamma-4 chain C region	0.62	1.00	1.00	1.74	1.5<x<2
MV120	Ig heavy chain V-III region GAL	0.52	1.40	0.66	1.69	0.5<x<0.67
IGHM	Ig mu chain C region	0.46	1.33	0.83	1.92	<0.5
K2C1	Keratin, type II cytoskeletal 1	0.35	1.98	0.20	0.01	0.67<x<1.5
LV302	Ig lambda chain V-III region LOI	0.33	1.49	0.46	1.34	1.5<x<2
A2GL	Leucine-rich alpha-2-glycoprotein	0.21	1.09	0.29	1.98	0.5<x<0.67
A1AGT	Alpha-1-acid glycoprotein 1	0.21	0.99	0.45	1.76	<0.5
CRP	C-reactive protein	0.18	0.39	0.11	1.62	0.67<x<1.5
IGHG3	Ig gamma-3 chain C region	0.50	1.47	0.36	1.42	1.5<x<2
IGHG2	Ig gamma-2 chain C region	0.39	1.05	0.61	1.30	0.5<x<0.67
AACT	Alpha-1-antichymotrypsin -SERPINA3	0.46	1.33	0.57	1.83	<0.5
APOA	Apolipoprotein(a)	0.55	0.76	0.22	0.12	0.67<x<1.5
CERU	Ceruloplasmin	0.32	1.11	0.60	1.80	1.5<x<2
CFAB	Complement factor B	1.20	1.38	0.68	1.73	0.5<x<0.67
CFI	Complement factor I	1.17	0.62	0.50	1.82	<0.5

RCC infiltration level	Fold change	Proteome expression consistency (%) between urine and plasma
Both RV invasion & RV thrombosis	Fold change B/A and C/B ≥ 1.5	58%
RV invasion	Fold change B/A ≥ 1.5	81%
RV thrombosis	Fold change C/B ≥ 1.5	65%

Figure 4

[Click here to download high resolution image](#)

	PANTHER GO-Slim Biological Process	#	URINE				PLASMA			
			RV infiltration		RV Thrombosis		RV infiltration		RV Thrombosis	
			UP	DOWN	UP	DOWN	UP	DOWN	UP	DOWN
catabolism - proteolysis	protein metabolic process	2062	63						10	
	proteolysis	598	38				32		6	
	catabolic process	789	32							
adhesion	cell adhesion	481	28	10	10	26				
	biological adhesion	481	28	10	10	26				
	cell-cell adhesion	305	22	7		21			5	
metabolism	nucleobase-containing compound metabolic process	3160	25	1		16				
	monosaccharide metabolic process	120	10							
	sulfur compound metabolic process	127	10							
transcription	glycolysis	34	6							
	RNA metabolic process	2051	3			1				
	carbohydrate metabolic process	476				25				
immunity	regulation of transcription from RNA polymerase II promoter	976	3			1				
	transcription from RNA polymerase II promoter	1219	3			1				
	transcription, DNA-dependent	1521	3			1				
endocytosis-phagocytosis	immune system process	1269		14	18			11	11	
	complement activation	131		8	6			9	8	
	immune response	717						9	8	
cell recognition	B cell mediated immunity	214						4	4	
	phagocytosis	116		5	5			4	4	
	endocytosis	418						5	5	
blood related	receptor-mediated endocytosis	233						4	4	
	cell recognition	103		5				4	4	
	blood coagulation	91		6				4	4	
	blood circulation	140								3

N° DOWN ID proteins

N° UP ID proteins