

Validation of SELDI-TOF MS serum protein profiles for renal cell carcinoma in new populations

Judith YMN Engwegen¹, Niven Mehra², John BAG Haanen³, Johannes MG Bonfrer⁴, Jan HM Schellens^{3,5}, Emile E Voest² and Jos H Beijnen^{1,5}

Currently, no suitable biomarker for the early detection or follow-up of renal cell carcinoma (RCC) is available. We aimed to validate previously reported potential serum biomarkers for RCC obtained with Surface Enhanced Laser Desorption Ionisation-Time of Flight Mass Spectrometry (SELDI-TOF MS) in our laboratory using distinct patient populations. Two sets of sera from RCC patients and healthy controls (HC) were gathered from different institutes and analysed according to published procedures. The first set (40 RCC, 32 HC) consisted of mainly presurgery samples from patients with disease stages I–IV. The second set (26 RCC, 27 HC) were mostly sera from patients with stage-IV disease, drawn after nephrectomy. Only the increased expression of the previously found serum amyloid- α (SAA) peak cluster could be validated in a similar RCC patient subset in both our populations in two independent analyses. It was seen both in early- and late-stage disease and in pre- and postsurgery samples. These results were also confirmed by ELISA. Other previously identified biomarker candidates (mass-to-charge ratio's (m/z) 3900, 4107, 4153, 5352 and 5987) proved difficult to reproduce upon duplicate analysis. Modification of the analytical protocol for these markers resulted in their detection, but we did not achieve satisfactory classification of patients and controls with these alleged biomarkers in any of our two sample sets. Instead, two new peaks (m/z 4289 and 8151) were identified with better performance (sensitivity and specificity ~ 65 – 90%) for separating patients from controls in the first sample set. Concluding, only the SAA peak cluster was validated as a robust RCC biomarker candidate, which is present in a specific subset of these patients, regardless of disease stage or nephrectomy status. In addition, two new peaks were seen which might prove useful as biomarkers, provided these are validated in new populations.

Laboratory Investigation (2007) 87, 161–172. doi:10.1038/labinvest.3700503

KEYWORDS: biomarker; renal cell carcinoma; reproducibility; SELDI-TOF MS; validation

Renal cell carcinoma (RCC) is difficult to diagnose at early stages due to a lack of clear clinical symptoms. When symptoms do occur, about 30% of patients already have metastatic disease. In addition, a similar part of patients with resection of localised disease will have a recurrence.¹ Therefore, a need remains for reliable markers for diagnosis and follow-up of RCC, preferably in easy-accessible body fluids. Surface Enhanced Laser Desorption Ionisation-Time of Flight Mass Spectrometry (SELDI-TOF MS)^{2,3} is being increasingly used to search for new and better tumour markers in (serum) protein profiles, for example, for ovarian,⁴ breast,⁵ prostate⁶ and colorectal cancer.⁷ Its appeal lies in the ease with which a multitude of samples can be

analysed with a minimum of sample preparation in a single SELDI-TOF MS analysis. Indeed, SELDI-TOF MS has also been performed to identify biomarker proteins for early detection of RCC. Two studies describe protein profiling of serum with SELDI-TOF MS.^{8,9} Tolson *et al*⁸ reported a peak cluster at a mass-to-charge ratio (m/z) of approximately 11 000 from serum amyloid α -1 and variants, which was 100% specific for RCC compared to HC, although not very sensitive (detection in eight of 25 patients). Five m/z values of 3900, 4107, 4153, 5352 and 5987 were reported by Won *et al*⁹ in a classification tree separating RCC from healthy and non-RCC urologic disease, with sensitivities and specificities ranging from 80 to 100%.

¹Department of Pharmacy and Pharmacology, The Netherlands Cancer Institute/Slotervaart Hospital, Amsterdam, The Netherlands; ²Department of Medical Oncology, University Medical Centre Utrecht, Utrecht, The Netherlands; ³Department of Medical Oncology, The Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands; ⁴Department of Clinical Chemistry, The Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands and ⁵Division of Drug Toxicology, Department of Biomedical Analysis, Faculty of Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands
Correspondence: JYMN Engwegen, PharmD, Department of Pharmacy and Pharmacology, Slotervaart Hospital, Louwesweg 6, 1066 EC, Amsterdam, The Netherlands. E-mail: apjen@slz.nl

Received 20 June 2006; revised and accepted 10 October 2006

Although these results are promising, concerns have risen regarding reproducibility of protein profiles and the nature of detected biomarkers.^{10–17} For the same cancer types, different biomarkers have been found by different research groups.^{13,18} Yet, reproducible protein profiles can hardly be presumed when using different assay procedures or sample handling, or when samples from a patient population with other characteristics are used.¹⁹ To ensure that supposed biomarker proteins are not due to chance or bias and are robust enough for detection in different laboratories, validation with an independent sample set originating from a different institute, but handled using standard analytical procedures, is imperative. Yet, only few groups have performed such validation. Attempts to validate biomarker proteins for ovarian and breast cancer with samples from other institutes have been reported, although not all with similar success.^{4,20,21} For prostate cancer, a biomarker validation project is ongoing,^{22,23} also standardising analyses among different laboratories. However, the RCC serum protein profiles in the two above-mentioned studies have not been validated in other laboratories or with new populations, leaving the question of their validity and robustness. Thus, instead of starting a new search for RCC biomarkers using SELDI-TOF MS, we attempted to move forward and validate the ones postulated by previous studies. We applied the previously developed analytical protocols^{8,9} to two distinct sample sets of patients and controls from two different institutes and assessed robustness and validity of reported RCC serum protein profiles.

MATERIALS AND METHODS

Chemicals

All used chemicals were obtained from Sigma, St Louis, MO, USA, unless stated otherwise.

Patient Samples

Our first set of samples (set 1) consisted of sera from 40 patients with renal cell carcinoma and 32 healthy controls (HC) obtained from the Netherlands Cancer Institute, Amsterdam, The Netherlands. The second set (set 2) consisted of sera from 26 patients with renal cell carcinoma and 27 HC obtained at the University Medical Centre in Utrecht. All serum samples originated from a serum bank, where they had been collected according to institutional protocols. Sample collection was performed with individuals' informed consent after approval by the institutional review boards.

Protein Profiling

The two serum sample sets were analysed at separate time points in our laboratory. Analyses for both sample sets were repeated after storage at -20°C for a year. A scheme of the experimental set-up is shown in Figure 1. For quality control, we analysed two separate pools of HC sera from set 1 and set 2, respectively, on seven chips in each analysis. Assay procedures were identical to those reported,^{8,9} using CM10 chips

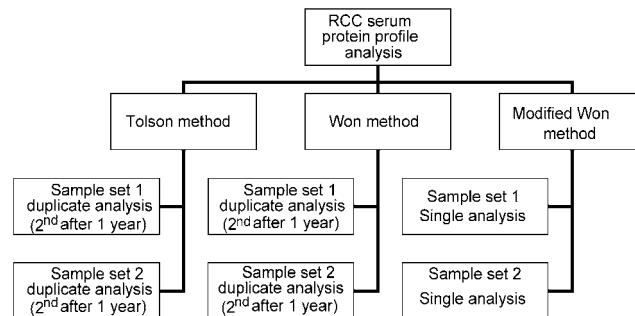


Figure 1 Experimental set-up. Two separate samples sets were analysed according to two previously published methods. In addition, a modified version of one of the methods was used to yield better peaks (see Results section).

with a weak cation exchange chromatography and as matrix a 50% solution of sinapinic acid (SPA) for Tolson's analysis and a saturated solution of α -cyano-4-hydroxycinnamic acid (CHCA) for Won's analysis. Only the albumin depletion for Tolson's method was performed slightly different, using Cibacron Blue spin columns from a different manufacturer (Ciphergen Biosystems Inc., Fremont CA, USA). Briefly, serum was denatured and depleted from albumin according to manufacturer's instructions. Equivalent quantities of denatured serum to Tolson's were loaded onto these columns, taking into account the different binding capacity for albumin. Protein chips were analysed using the PBS-IIc ProteinChip Reader (Ciphergen Biosystems). Data acquisition parameters in our experiments were optimised for detection of the reported biomarkers. Using Tolson's method, spectra were collected to 80 kDa with focus mass kept at 8000 Da according to reported procedures. M/z values for the detected proteins were calibrated externally with a standard peptide mixture (Ciphergen Biosystems) containing [Arg8] vasopressine (1084.3 Da), somatostatine (1637.9 Da), dynorphine (2147.5 Da), ACTH (2933.5 Da), insulin β -chain (bovine) (3495.9 Da), insulin (human recombinant) (5807.7 Da) and hirudin (7033.6 Da).

Statistics and Bioinformatics

Data were analysed with the ProteinChip Software package, version 3.1 (Ciphergen Biosystems). Per sample set, all acquired spectra were compiled and analysed as a whole. Spectra were baseline subtracted and normalised to the total ion current from 1500 Da to the spectrum's end. Spectra with normalisation factors above 2.00 or lower than 0.50 were excluded from clustering. Biomarker Wizard (BMW) software (Ciphergen Biosystems) was used for peak clustering. Peaks were autodetected when occurring in at least 30% of spectra and with first and second pass signal-to-noise (S/N) of 5 and 2, respectively, in a 0.3% cluster mass window. In case of no autodetection of peaks of interest, peak detection was performed manually and the BMW applied including these user-detected peaks. Mean peak intensities between

groups were compared by nonparametric statistical tests with $P < 0.01$ defined as statistically significant. The mean coefficient of variation (CV) in each sample set was calculated with replicate peak intensities from the quality-control samples. Biomarker Patters Software (BPS; Ciphergen Biosystems) was used to generate classification trees for patients and controls.

Protein Identification

The identity of proteins of interest was elucidated using immunocapture with appropriate antibodies. For confirmation of the 11-kDa peak identities, the same mouse monoclonal antibody to serum amyloid- α (SAA) was used as reported (Abcam Ltd., Cambridge, UK). In short, antibody was coupled to protein A ceramic HyperD beads (Pall/Biosepra, Saint-Germain-en-Laye, France) and washed thrice with phosphate-buffered saline (PBS). After a 30-min incubation of serum diluted in PBS, the unbound fraction was collected. Beads were washed five times with PBS and once with deionised water. Finally, bound proteins were eluted from the beads with 0.1 M acetic acid. Unbound fractions as well as eluates were profiled both on gold chips or NP20 chips and on the original CM10 chip surface.

Measurement of Serum SAA Levels

Results of protein profiling for SAA were validated by ELISA using a commercially available kit (Tridelta Development

Ltd, Maynooth, Ireland). All sera were analysed in duplicate according to manufacturer's instructions.

RESULTS

Patient Samples

Patient and sample characteristics are summarised in Table 1. Two distinct patient populations were gathered: In our first sample set, most samples had been drawn before surgery and patients had mostly early-stage disease, as was the case for the samples in the published studies. In contrast, many samples from set 2 were from patients with metastatic stage-IV disease who had undergone nephrectomy. This enabled us to assess the abundance of the previous markers across several disease stages and the influence of primary tumour resection on their presence. No group of patients with other urological diseases was included in our analysis. Sample collection procedures differed between the institutes regarding clotting time and storage temperature (Table 1).

Protein Profiling: Tolson Method

In the first analysis, we obtained optimal spectra using laser intensity 142 and detector sensitivity 7, collecting an average of 165 laser shots per spot. For the second analysis a year later, acquisition parameters had to be re-optimised to laser intensity 155 and detector sensitivity 8, whereas other parameters were kept constant. On both occasions, protein patterns were highly reproducible within and between the

Table 1 Patient and sample characteristics of sample sets 1 and 2 compared to those of Won and Tolson

	Sex			Patient age (years)		RCC stage	Surgical status	Sample handling	Sample age (years)	
	RCC	HC	Non-RCC disease	RCC	HC				RCC	HC
Set 1	25 M	20 M	—	57.1 (11.0)	56.1 (8.79)	16 IV 4 III	33 presurgery 5 postsurgery	30 min coagulation at RT Centrifugation at 1900 g	2.94 ^a (0.56)	1.90 (0.034)
	15 F	12 F				16 I/II 4 Unknown	2 unknown	Storage at -30°C		
Set 2	19 M	7 M	—	63.7 ^b (10.2)	42.9 (11.1)	24 IV	8 presurgery	2-6 h coagulation at RT+ overnight at 4°C	1.05 (0.55)	1.28 (0.21)
	7 F	20 F				2 II/III	18 postsurgery	Centrifugation at 1500 g Storage at -80°C		
Tolson	15 M	26 matched	—	59.3 (14.6)	Matched	1 III 6 IV 7 II 11 I	25 presurgery	Serum supernatant collected Overnight coagulation at 4°C Storage at -70°C	NR	NR
10 F										
Won	15 M+F	6 M+F	15 M+F	NR	NR	NR	15 presurgery	NR	All collected in same period	

Means (s.d.) for patient and sample age.

Abbreviations: F: female, HC: healthy control, M: male, NR: not reported, RCC: renal cell carcinoma, RT: room temperature.

^a $P < 0.000$ independent samples *t*-test.

^b $P < 0.000$ nonparametric Mann-Whitney *U*-test.

Table 2 Comparison of peak detection in Tolson's samples and in our sets 1 and 2

	Peak detection in number of patients/controls					Peak identity
	Sample set 1 (max 40/32)		Sample set 2 (max 26/27)		Sample set Tolson (max 25/26)	
	1st analysis	2nd analysis	1st analysis	2nd analysis		
Excluded from analysis	0/1	4/1	2/2	1/0	NA	
9.2 kDa	4/2 ^a	14/13	1/2 ^a	14/13	15/12	Haptoglobin 1- α
10.85 kDa ^b	1/0 ^c	3/0	1/0 ^c	1/0	16/6	Unknown
11.4 kDa ^b	10/0	11/0	4/0	5/0	7/0	des-RS SAA-1
11.5 kDa ^b	18/0	19/0	5/0	7/0	8/0	des-R SAA-1
11.68 kDa ^b	18/1	20/1	6/0	7/0	8/0	SAA-1
13.7 kDa	30/29	28/29	15/25	17/26	25/26	Transthyretin
15.1 kDa	1/0	1/0 ^d	1/5	0/0 ^d	25/26	α -Globin
15.8 kDa	19/14	13/11	11/16	14/11	25/26	β -Globin
28.0 kDa	39/31	36/31	24/25	25/27	25/26	Apolipoprotein A-I

Peaks were considered present when having an $S/N \geq 2$.

Abbreviations: SAA-1: serum amyloid α -1; R: arginine; S: serine.

^aWhen using a cutoff of $S/N > 1$, detected in 14 RCC patients and nine controls from sample set 1 and 18 patients and 17 controls from sample set 2.

^bTolson biomarker candidate.

^cWhen using a cutoff of $S/N > 1$, detected in four RCC patients in set 1 and four RCC patients in set 2.

^dWhen using a cutoff of $S/N > 1$, detected in one RCC patient and two HC in set 1 and two RCC patients and one HC in set 2.

sample sets and the two analyses. Mean peak CV's for the quality-control sera from each sample set were 20 and 30% for set 1 and 2, respectively, based on the 11 most prominent peaks common to both sample sets. Few samples had to be excluded due to aberrant normalisation factors (Table 2). Comparison of spectra from the duplicate analysis with the one from Tolson shows the presence of the same peaks (Figure 2). Results of the duplicate analysis were highly congruent, except for the 9.2-kDa peak, which was seen more frequently in the second analysis. The 11-kDa peaks were present in both our sample sets in a roughly similar patient subset as Tolson's (Table 2). In two healthy controls from set 1 and one from set 2 this peak was also visible, although with $S/N < 2$ (Figure 3). None of the excluded spectra showed the 11-kDa peaks. The peak at 10.85 kDa, also indicated as a potential biomarker by Tolson, was found in very few of our patients (Table 2). The relative abundances of Tolson's biomarkers in our patients in relation to their surgery status and disease stage are summarised in Table 3. Also for some other (non-biomarker) peaks, the frequencies we found were lower than reported, such as for the 9.2-kDa peak from haptoglobin in the first analysis and for the peaks at 15.8 and 15.1 kDa (β - and α -globin).

When applying the BMW application to these spectra, the 11-kDa peaks came up as most significantly different in both analyses and both sample sets ($P < 0.000$). However, as these peaks only occur in a subset of RCC patients, their sensitivity

for RCC detection is limited. Therefore, we also evaluated other discriminating peaks that were present in all patients and/or controls. Peaks at 8.7, 14.1, 17.3 and 28 kDa were consistently decreased in patients in both analyses and both sample sets ($P < 0.000$, except for 14.1 kDa in set 2: $P < 0.03$). Classification trees were constructed with BPS, but none of the above-mentioned discriminatory peaks, nor any other peaks yielded trees with suitable sensitivity and specificity for both sample sets (data not shown).

Protein Profiling: Won Method

Figure 3 shows spectra from the optimisation of laser intensity and detector sensitivity for the Won peaks at 3.90, 4.10, 4.15, 5.35 and 5.99 kDa. In the first analysis, we obtained optimal spectra using laser intensity 135 and detector sensitivity 7, collecting an average of 165 laser shots per spot. Again, for the second analysis, re-optimisation of acquisition parameters was needed, resulting in laser intensity 128 and detector sensitivity 7 keeping other parameters constant. On both occasions, protein patterns were highly reproducible across the sample sets and the two analyses. However, many spectra were 'empty' and had to be excluded from further data analysis (Table 4). We noted that this was related to the presence of a glassy layer on the chip spot for these samples. This layer is probably due to lack of a washing step with water in the procedure, leaving salt remnants on the chip that cause ion suppression. Owing to this, only two quality-control

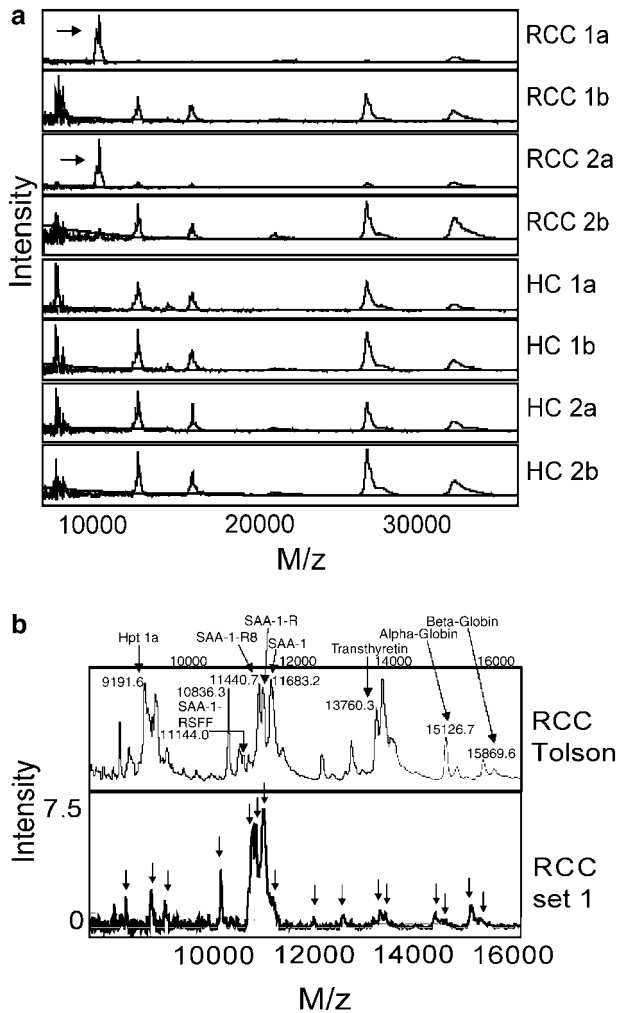


Figure 2 Spectra parts generated according to Tolson's procedure. (a) Samples from set 1 and set 2 are shown. The 11-kDa peaks are visible in a subset of the RCC sera (arrow), but not in those from HC. (b) Comparison of one of our spectra with the one reported by Tolson *et al*. Nearly all peaks present in the Tolson spectrum were visible in our spectrum (arrows), although some with lower peak intensities, which is possibly due to the fact that the published spectrum was acquired with extra high laser intensity for peaks >10 kDa. Tolson spectrum reprinted by permission from Macmillan Publishers Ltd: [Lab Invest],⁸ copyright 2004.

samples from set 1 and three from set 2 were assessable. Mean peak CVs of the seven most prominent peaks in these spectra common to both sample sets were 60 and 23% for set 1 and 2, respectively. Our peak intensities for the peaks of interest were markedly lower than the intensity values reported by Won, except for the peak at 4.10 kDa (Table 5, original procedure and Figure 4). Varying acquisition parameters did not result in higher intensities for these peaks. As a result, only the 4.10-kDa peak met the criterion of an $S/N > 5$ for clustering and the others had to be manually detected. Although the mean peak intensities from the two analyses were comparable, there were some differences between the first and second analysis regarding the group with the highest

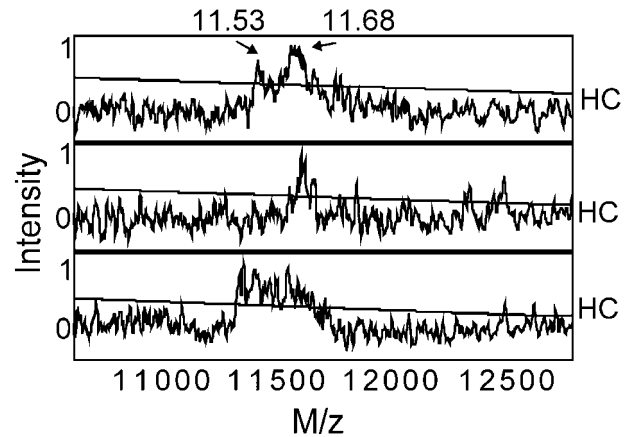


Figure 3 Detection of the 11-kDa cluster in healthy controls (HC) from sample sets 1 and 2. The straight line represents the noise level.

mean peak intensity and the observed P -values. None of the Won peaks was significantly different at the 0.01 level in our analyses (Table 5, original procedure). Some of the peaks did show a (nonsignificant) expression difference between patients and controls, but this was only observed in one of the analyses (m/z 4097, 5348), or was discordant with the expression difference reported by Won (m/z 5350, 5991). The lack of an expression difference in patients and controls for m/z 4153 was in agreement with Won's result, however (Figure 5).

In an attempt to obtain better intensities for the peaks of interest, we also used a slightly modified assay for these samples (single analysis). Instead of 2% 2-mercaptoethanol in the denaturation solution, 2% dithiotreitol (DTT), a more effective reductant, was used and the protease inhibitor phenylmethanesulphonyl fluoride (PMSF) was left out. Instead of saturated CHCA, a 50% SPA solution in 50% acetonitrile (ACN) + 0.5% sinapinic acid (TFA) was used as energy-absorbing matrix. We have applied this matrix solution successfully for masses below 15 kDa in other experiments, among which the Tolson procedure. SELDI-TOF MS acquisition parameters were optimised for these conditions to laser intensity 142 and detector sensitivity 7. From Figure 6 it can be seen that with the modified method, the same samples yielded higher peak intensities for m/z 3901, 4107, 4153, 5352 and 5987, with reproducible spectra among sample sets 1 and 2. In addition, many more peaks were present in the mass spectrum than with the original procedure, and few spectra did not meet normalisation criteria (Table 4). Two significant differences in agreement with Won's result could now be seen at m/z 5996 and 3894 in set 1. However, the found increase in RCC patients of a peak at 4161 Da in set 2 ($P = 0.004$, Table 5, modified procedure) was not according to Won's results.

To assess whether the many patients who had undergone nephrectomy in sample set 2 could be the cause of the lack of significant expression differences, only data from the combined presurgery sera in set 1 and 2 were analysed (41 RCC

Table 3 Abundances of Tolson peaks (peak height ‘averages’ of duplicate analysis and concentrations measured by ELISA) and patient characteristics

	Patient ID	Abundance of peaks of interest			Surgical status	Disease stage
		10.85 kDa	11-kDa cluster	SAA concentration (µg/ml)		
Set 1	RCC 364	—	++	50.9	Presurgery	II
	RCC 367	+/-	+++	67.0	Presurgery	IV
	RCC 368	—	+	3.97	Presurgery	I
	RCC 369	+/-	+++	2.35	Presurgery	IV
	RCC372	—	++	74.5	Presurgery	IV
	RCC 373	—	+++	75.6	Presurgery	IV
	RCC 374	++	+++	72.8	Presurgery	IV
	RCC 375	—	++	50.5	Presurgery	IV
	RCC 376	+/-	++	49.2	Presurgery	IV
	RCC 377	—	+++	36.7	Presurgery	IV
	RCC 378	+/-	+++	71.8	Presurgery	I
	RCC 379	+/-	+++	67.5	Presurgery	I
	RCC 381	—	+++	89.0	Presurgery	II
	RCC 387	+/-	++	43.7	Presurgery	I
	RCC 388	—	+	15.2	Presurgery	I
	RCC 390	—	+	33.1	Unknown	IV
	RCC 392	—	+++	61.0	Presurgery	IV
	RCC 393	—	++	9.59	Presurgery	IV
	RCC 397	—	+	36.3	Presurgery	III
	RCC 399	++	—	11.4	Postsurgery	Unknown
RCC 403	+	—	7.50	Presurgery	I	
Set 2	RCC 94	+/-	+++	3.24	Presurgery	IV
	RCC 96	—	+++	61.1	Presurgery	IV
	RCC 99	—	+	44.9	Presurgery	IV
	RCC 100	++	++	53.2	Postsurgery	IV
	RCC 109	—	++	50.9	Presurgery	IV
	RCC 114	—	+	7.61	Postsurgery	IV
	RCC 116	—	+++	94.0	Presurgery	IV

and 59 HC). After normalisation, 49 (22 RCC, 27 HC) and 39 (15 RCC, 24 HC) spectra from the two analyses with the original procedure and 85 (35 RCC, 50 HC) spectra from the modified procedure were assessable. None of the peaks of interest were discriminative in the presurgery samples with the original procedure. Only in the second analysis we saw a slight difference at m/z 4159 ($P=0.09$, intensity HC 3.89 vs RCC 4.33), which was not in agreement with Won's result, however. In data from the modified assay, the peak at 4097 Da was seen slightly, but significantly decreased in patients compared to controls (intensity 44.8 vs 52.6, $P=0.003$). Although an expression difference at m/z 5996

and 3894 was still present, significance was lost in this analysis ($P=0.11$ and $P=0.12$).

Lastly, we looked for other discriminative peaks than those reported by Won. Few significant differences between the RCC and control groups were found using the original procedure. A significant peak at 8597 Da was observed, which was decreased in RCC patients ($P<0.000$), both with the original and modified procedure using either the combined samples from set 1 and 2, or only the presurgery ones. Considering the presurgery samples only, in both analyses following the original procedure, consistent expression differences were seen for m/z 2826 ($P=0.01$ and $P=0.09$,

respectively) and 4291 ($P=0.04$ and $P=0.01$). However, with the modified procedure many more peaks differed significantly, notably m/z 4289, 3960, 8151, 6198, 3163, 8707,

4303 and 6456 ($P<0.000$ in the combined sample sets and in the presurgery samples).

Table 4 Spectra from the Won method included for data analysis

	Assessable RCC/total RCC		Assessable HC/total HC	
	Set 1	Set 2	Set 1	Set 2
Original procedure, 1st analysis	20/40	18/26	12/32	17/27
Original procedure, 2nd analysis	13/40	11/26	14/32	13/27
Modified procedure	33/40	25/26	26/32	24/27

Generating classification trees, we found some with reasonably good classification of RCC and healthy persons. The (nonsignificant) Won peaks at 4097 and 5348 Da in the second original analysis did not come up as tree classifiers, nor did the significant m/z 's (5996 and 3894) in the data from the modified method (data not shown). Yet, from these data a tree could be built using m/z 4289 as single classifier with sensitivity 73.3% and specificity 66.0% upon 10-fold cross-validation. Also, m/z 8151 showed promising performance as a single classifier (cross-validation sensitivity 90.6%, specificity 81.5%). However, these peaks were good classifiers only for sample set 1. We also investigated whether the significantly different 4097-Da peak in the presurgery serum

Table 5 Observed peak expression differences detected in our sample sets with Won's method (original procedure) and a slightly modified procedure

Reported by Won			Observed set 1 ^a				Observed set 2 ^a			
m/z (Da)	HC peak intensities	RCC peak intensities	m/z (Da)	P ^b	HC peak intensities	RCC Peak intensities	m/z (Da)	P ^b	HC peak intensities	RCC peak intensities
			<i>Original procedure (duplicate)</i>				<i>Original procedure (duplicate)</i>			
3901	22.9 (5.38)	14.4 (5.67)	3901	0.186	1.24 (0.570)	1.08 (0.67)	3904	0.792	2.13 (1.60)	1.84 (1.01)
			3897	0.409	2.35 (2.35)	1.72 (2.24)	3897	0.75	1.46 (1.47)	1.26 (0.650)
4107	29.4 (7.09)	6.48 (7.40)	4100	0.119	3.56 (2.34)	2.33 (1.97)	4104	0.322	7.74 (6.39)	5.54 (4.95)
			4101	0.409	13.7 (9.24)	17.4 (10.1)	4097 ³	0.052	13.7 (7.38)	8.32 (4.33)
4153	9.19 (6.75)	8.44 (4.50)	4155 ^c	0.907	2.96 (3.09)	2.33 (1.58)	4159 ^c	0.146	1.21 (1.11)	1.66 (1.17)
			4157 ^c	0.072	4.29 (6.30)	4.22 (2.44)	4163 ^c	0.339	1.96 (1.16)	1.68 (1.84)
5352	16.5 (3.65)	12.2 (4.37)	5357	0.559	0.962 (0.567)	0.892 (0.665)	5353	0.291	1.64 (1.15)	2.43 (1.97)
			5350	0.022	1.76 (1.46)	3.08 (1.45)	5348 ^c	0.099	3.11 (2.34)	1.68 (1.36)
5987	7.70 (4.14)	4.82 (4.93)	5983	0.111	0.919 (0.480)	0.671 (0.488)	5991	0.056	0.517 (0.489)	1.05 (0.951)
			5996	0.593	1.17 (0.661)	1.15 (0.939)	5991	0.75	1.53 (0.990)	1.38 (0.944)
<i>Classification tree difference</i>			<i>Modified procedure</i>				<i>Modified procedure</i>			
3901	↑ in RCC vs non-RCC		3894 ^c	5.19×10^{-05}	9.01 (4.10)	4.81 (2.13)	3883	0.383	5.33 (2.57)	4.89 (1.70)
4107	↑ in HC vs RCC		4098	0.171	39.2 (11.5)	35.4 (13.8)	4098	0.226	59.0 (9.9)	56.4 (11.8)
4153	↑ in non-RCC vs RCC		4160 ^c	0.543	9.99 (8.77)	8.10 (5.90)	4161	0.004	8.04 (3.70)	11.3 (3.6)
5352	↑ in HC vs non-RCC		5364	0.346	0.862 (0.635)	0.666 (0.658)	5361	0.803	1.95 (2.50)	1.39 (1.03)
5987	↑ in non-RCC vs RCC		5996 ^c	3.49×10^{-05}	1.60 (0.54)	0.961 (0.501)	5995	0.476	1.87 (1.10)	1.65 (0.81)

Mean (s.d.) peak intensities for the RCC and HC groups are given.

^aValues for the first and second analysis, respectively.

^bNonparametric Mann-Whitney U -test.

^cExpression concordant with Won's result.

samples using the modified procedure was a useful classifier. Ten-fold cross-validation with this *m/z* as single classifier yielded limited sensitivity and specificity of 62 and 59%. The other significant *m/z*'s we found by clustering were no classifiers or did not yield trees with reasonable sensitivity and specificity (data not shown).

Protein Identification

The identity of the 11-kDa cluster was confirmed by using a serum amyloid- α mouse monoclonal antibody (Abcam Ltd, Cambridge, UK). A similar peak cluster as the one in the

profiling experiment was seen in the bound fraction of this antibody, both on NP20 and on CM10 chips (Figure 7). The identities of the discriminating peaks at 14.1 and 17.3 kDa in the Tolson analysis and that of 8597 Da in the Won analysis could be deduced from the protein profiling results and were confirmed by similar immunocapture experiments. Both these discriminating Tolson peaks were lost with the Won analysis. Although highly similar, these methods differ in the addition of a reducing agent to the denaturation buffer in the latter procedure. A peak cluster around 14 kDa has been described, identified as transthyretin and several isoforms, among which a glutathionylated form at 14.1 kDa.²⁴ Similarly, from the Swiss-Prot database (<http://www.expasy.org/>) it is known that a 17-kDa homodimer of apolipoprotein A-II is formed by disulphide linkage of two monomers. With the Won procedure, reduction would result in loss of the glutathion group for the 14.1-kDa peak to form a mass of

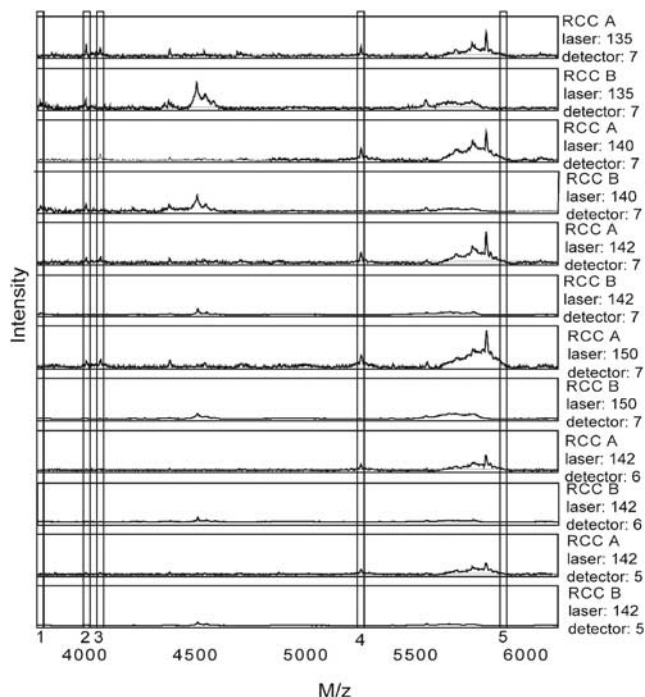


Figure 4 Optimisation of SELDI-TOF MS acquisition parameters for detection of the Won peaks (1–5). Two RCC samples were used for optimisation (A and B). Both laser intensity and detector sensitivity were optimised separately.

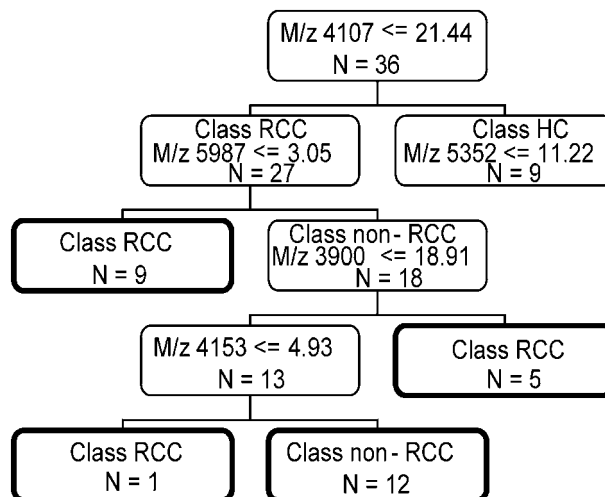


Figure 5 Classification tree reported by Won *et al*. The primary node includes the peak at 4107Da, which roughly separates between patients (left branch) and controls (right branch). Reprinted with permission from Wiley-VCH Verlag GmbH & Co KG.

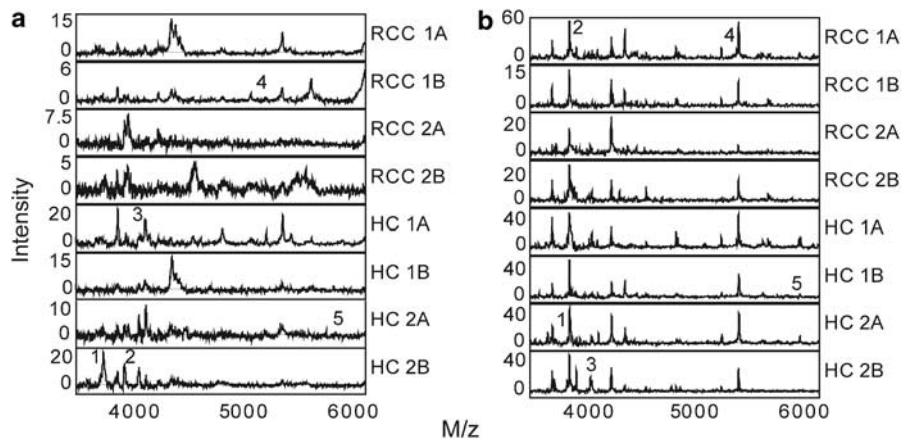


Figure 6 Spectra parts generated with Won's original procedure (a) and with a slightly modified one (b). The same samples in set 1 and set 2 are shown for either assay. Peak intensities for the peaks of interest at 3901, 4107, 4153, 5352 and 5987 Da (peaks 1–5) are higher in the modified procedure.

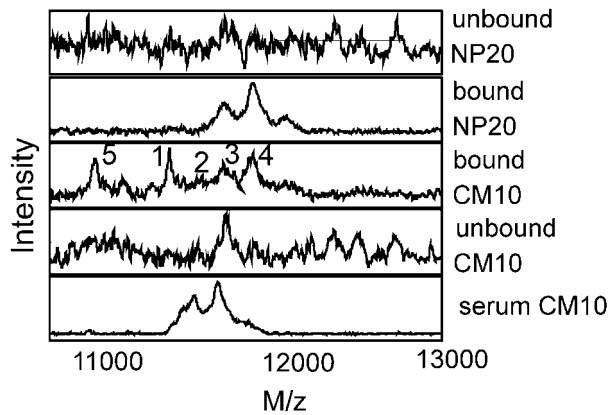


Figure 7 Identification of the serum amyloid- α (SAA) peak cluster by immunocapture. The spectra parts successively show protein profiling on NP20 chip of the serum fraction that has not bound to the SAA-1 antibody, the fraction that has bound to this antibody on NP20 chip and on CM10 chip, the unbound serum fraction on CM10 chip and the original protein profile of whole serum on CM10 chip. Peaks 1–3 represent des-R5 SAA-1 (11.4 kDa), des-R SAA-1 (11.5 kDa) and SAA-1 (11.68 kDa), respectively. The peak at 11.9 kDa (4) may be another post-translationally modified form of SAA-1. The fifth peak at 10.8 kDa may be the same as found by Tolson, although it was not reported to be a form of SAA-1.

13.7 kDa and the formation of two 8.5-kDa monomers from the 17-kDa dimeric protein, which was indeed the case in our results. Therefore, we performed immunocapture with a transthyretin rabbit polyclonal antibody (Abcam Ltd, Cambridge, UK) and an apolipoprotein A-II goat polyclonal antibody (Chemicon Europe Ltd, Hampshire, UK). Thus, the identity of glutathionylated transthyretin was confirmed, as was that of the apolipoprotein A-II fragment missing a terminal glutamine residue in one chain of the dimer (Tolson method, 17.3 kDa) and in the monomer (Won method, 8597 Da).

Measurement of Serum SAA Levels

We observed good agreement between the results of SELDI-TOF MS protein profiling and ELISA regarding the abundance of SAA (Figure 8). It is also clear from this figure that protein profiling is less sensitive than ELISA for the detection of SAA. Similar to Tolson's result, SAA levels below 20 $\mu\text{g/ml}$ were generally not detectable with SELDI-TOF MS in our analyses. However, in several RCC samples, the 11-kDa peaks were detectable despite a measured SAA concentration below 10 $\mu\text{g/ml}$ (Table 3). Most HC had SAA levels below 10 $\mu\text{g/ml}$, except for 6 samples with levels between 10 and 25 $\mu\text{g/ml}$.

DISCUSSION

Serum protein profiling of RCC was performed to assess the validity and robustness of previously reported biomarker proteins for this type of cancer. We were able to detect the discriminating cluster of serum amyloid- α peaks described by Tolson *et al* in about the same subset of RCC patients in both

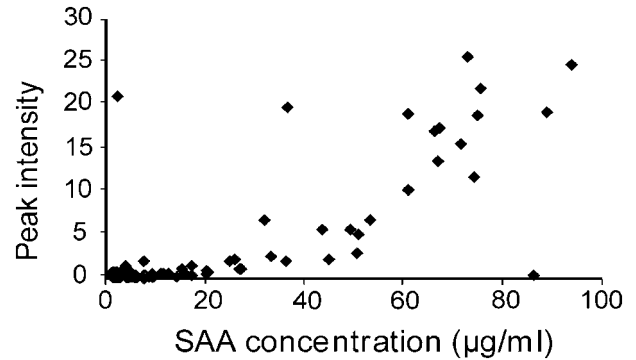


Figure 8 Correlation of serum levels of SAA as measured by ELISA with peak intensities of the 11.68-kDa peak from SELDI-TOF MS protein profiling.

sample sets. Furthermore, the protein profiling result was validated with an ELISA for SAA. The observed expression difference between patients and controls was detected despite differences in sample handling and patient characteristics compared to Tolson's samples. Moreover, it was stably present a year after the first analysis. This indicates SAA is a robust biomarker candidate for RCC, which is also supported by the fact that we found it not only with the assay from Tolson, but also with both the original and modified assay from Won, implying that albumin depletion does not influence the presence of this peak cluster. However, for some other peaks, we found markedly lower frequencies in both analyses, such as for the postulated biomarker of 10.85 kDa. As we did detect it in some of the patients, it is unlikely that the lower frequency is due to the analytical procedure, but more probable, these differences reflect real biological variation in patients. The co-presence of this peak with the SAA peaks as reported by Tolson could not be confirmed in our populations. Furthermore, its postulated prognostic value suggested by its presence in four out of six patients with metastatic disease in Tolson's population does not seem to hold, as most of our metastatic patients did not have this peak.

As we found a similar number of patients with the SAA peak cluster as Tolson, that is, SAA levels $> \sim 20 \mu\text{g/ml}$, this possibly represents a specific sub-population of RCC patients. As can be concluded from Table 3, disease stage or surgical status do not account for this distinction, nor seems the abundance of this peak correlated with disease stage. A study by Le *et al*²⁵ described a high sensitivity and specificity for the presence of SAA and isoforms in prostate cancer patients suffering from bone metastases, but this does not explain the abundance of these peaks in some of our patients with early-stage disease. The cluster of SAA peaks has also been mentioned in several other reports. In a study in ovarian cancer it was found in about half of the patients,²⁶ and a study in nasopharyngeal cancer²⁷ showed an increase of SAA in $\sim 70\%$ of these patients and in $\sim 40\%$ of lung cancer patients. These peaks are thus not specific for RCC, but

might reflect a process that occurs only in a subset of these cancer patients. Despite that we found some other expression differences in both sample sets, these were not large enough for good classification of patients and controls. Furthermore, both a decrease of a 17.3-kDa peak (possibly des-glutamine apolipoprotein A-II) and glutathionylated transthyretin have been described in the sera from patients with other cancers before^{24,28–30} and thus they seem not specific for RCC either. Yet, these proteins might have a role as markers in disease or therapy monitoring.

We experienced difficulties in the analysis with the Won procedure. Many spectra were not assessable, possibly because of ion suppression caused by salt remnants on the chip. Moreover, peak intensities for the peaks of interest were generally much lower than those reported, despite optimisation of MS acquisition parameters, and Won's discriminating peaks were not readily detectable. Thus, this analytical protocol seems less robust than, for example, Tolson's. Furthermore, only nonsignificant expression differences were observed with this procedure, but only in sample set 2, which was the least comparable to Won's. The only similarity with Won's data for this procedure was a lack of an expression difference for m/z 4153. A slightly modified method, including a better reductant and omission of protease inhibitor in the denaturation buffer, as well as a different energy-absorbing matrix generally improved peak detection. Moreover, much fewer spectra suffered from ion suppression. Although also in this modified procedure a final washing step with water was lacking, the use of a different matrix might have precluded ion suppression effects. Using this method, the observed peak differences proved reflective of the patient characteristics: only in sample set 1, most comparable to Won's, we found two of the previously described expression differences. Yet, in the combined presurgery sera from set 1 and set 2, these expression differences became statistically insignificant. Possibly, other patient characteristics than nephrectomy status are responsible for this discrepancy, for example disease stage. Indeed, all presurgery sera in set 2 were from patients with metastatic disease, whereas in set 1 these included 20 patients with RCC stages I–III. In addition, with this modified method, only for the combined presurgery sera we found a peak at 4097 Da higher in controls compared to patients, as reported by Won. However, the difference in mean peak intensities between patients and controls was only slight, a marked difference compared to Won's result. Therefore, despite the observed expression difference, this peak was not identified by the BPS as a main classifier for the classification tree, nor was it able to split patients and controls correctly when manually selected as the main classifier. Thus, we could not reproduce the classification performance of this peak that was suggested by Won's classification tree.

For a peak at 4161 Da, we found an expression difference in sample set 2 with the modified procedure, but not in set 1, nor in the combined presurgery sera. This protein (increased

in non-RCC disease vs RCC according to Won's classification tree) might be a specific inflammatory protein that is neither deregulated in RCC nor in health, which is concordant with the similar mean peak intensities for patients and controls mentioned by Won (8.442 ± 4.496 and 9.192 ± 6.746 , respectively⁹). The increase of this peak in RCC that we saw in sample set 2 might be due to differences in patient characteristics or sample handling compared to Won's samples. Interestingly, a more recent report on serum protein profiling for RCC also describes the increase of a peak at m/z 4151 in RCC patients compared to controls.³¹ This peak was observed under similar assay conditions and thus may well be the same protein as Won's. Its performance to detect RCC in stage I was reported lower than for later stages.³¹ This could explain why we only saw a significant difference in sample set 2 (mostly stage-IV disease) and not in set 1 or the presurgery sera (mainly early-stage samples), and, if Won and co-workers also evaluated earlier disease stages, why they did not find this difference.

Altogether, some of Won's postulated biomarkers seem prone to lack of reproducibility (at least with the original method) and lack of validity. Although differences with our results might be caused by different sample handling conditions or other disease stages, this at least stresses the necessity of careful control of these parameters if the Won proteins were to be used as biomarkers. Furthermore, the few assessable spectra with the original procedure may have hampered our power to detect the expression differences described previously. However, the 4107-Da peak splitting patients and controls in the first node of Won's classification tree is only based on a total of six controls and 15 patients, which is even less than our assessable individuals. With the modified method, we could evaluate many more spectra, but the expression difference at m/z 4107 that we found for the presurgery samples was not large enough for correct classification of patients and controls. Thus, the classification tree reported by Won might well be the result of overfitting of data. For five peaks to fit a population of 36 samples in three groups is a rather high number, as peaks in the lower branches of a classification tree are discriminative for successively lower numbers of samples, thus more likely influenced by outliers or artefacts in the data. Possibly, the other discriminative m/z 's at 4289 and 8151 that we found with this procedure have more potential as biomarkers for RCC. A decrease of a peak at 4289 Da has been mentioned for several other cancers,^{32,33} one group attributing this mass to a fragment of inter- α trypsin inhibitor heavy chain 4.³³ Also, peaks with an m/z similar to 8151 have been reported as biomarker candidates.^{6,34} However, the consistent decrease of these peaks in RCC remains to be established in new sets of presurgery serum samples including patients with early-stage disease. In addition, their identity and their specificity for RCC or cancer in general must be addressed.

Considering our goal to validate previously reported biomarkers for RCC, lack of robustness was the first hurdle to

take, especially for the Won analysis. It is unlikely that chip characteristics, (manual) sample processing or instrument integrity were the cause of this, as the spectra that were assessable with both the original and modified analytical protocol from Won, and those from the Tolson protocol did show reproducible profiles across the sample sets and the duplicate analyses. More probably, critical steps in the analytical protocol were the cause, such as the use of a good reductant (DTT) and suitable energy-absorbing matrix (SPA). Furthermore, duration of all steps might prove critical: as DTT reduces full-length haptoglobin, resulting in the release of the 9.2-kDa 1- α chain, the lack of detection of this peak in the first Tolson analysis may have been caused by incomplete reduction of the full-length protein. Because of this potential sensitivity of the technique to such subtle changes, the structural characterisation of any postulated biomarkers and development of more quantitative assays are needed for routine clinical analyses. The second hurdle in validation is related to the law of large numbers. Even when peaks are reproducibly detected, failure to validate expression differences between patients and controls may still occur as a result of limited sample size: too few samples could result in insufficient statistical power to identify a significant difference for the alleged biomarkers. Alternatively, statistically significant differences found previously could rather be the result of chance or bias, in which case former data were overfitted. These problems seem likely to have occurred with Won's result, which was based on very few samples. However, we cannot rule out that this was also the case for our data, as our sample size was still limited.

It might be argued that the only expression difference we could validate was that of the acute phase protein SAA, and that this is due to the extreme increase of this protein during acute phase reactions, which makes detection with SELDI-TOF MS more feasible. However, the change in protein expression detectable with SELDI-TOF MS does not so much depend on the size of this change, as on the concentration range in which it occurs. As for all equilibrium reactions, the binding of proteins such as SAA to the chromatographic chip surface describes an S-curve. If an expression difference occurs in the linear range of this S-curve, a rather small increase can result in binding of much more protein and thus a large change in SELDI-TOF MS peak intensity. Alternatively, a small change in peak intensity might represent a large expression difference, when this happens in the plateau regions of this S-curve. Unfortunately, in complex matrices such as serum, one can never be sure when to be in the linear or plateau range, due to competition for binding sites from other proteins (which causes a shift of the S-curve to the right). Furthermore, the more abundant proteins, often acute phase proteins, compete and bind preferably to the chip, causing limited sensitivity of SELDI-TOF MS for less abundant proteins compared to antibody assays such as ELISA. Thus, the ability to detect expression differences with SELDI-TOF MS depends more on the abundance of the protein

under investigation than on the extent of the expression difference.

Concluding, we show that, depending on the analytical protocol, some of the alleged RCC biomarkers were difficult to reproduce. From the detectable candidate biomarkers, only the increased expression in RCC of SAA and variants could be unambiguously validated in our populations, with potential specificity for a certain subset of cancer patients. However, two newly identified biomarker candidates merit further evaluation in new sample sets.

1. Cohen HT, McGovern FJ. Renal-cell carcinoma. *N Engl J Med* 2005;353:2477–2490.
2. Fung ET, Thulasiraman V, Weinberger SR, *et al*. Protein biochips for differential profiling. *Curr Opin Biotechnol* 2001;12:65–69.
3. Issaq HJ, Veenstra TD, Conrads TP, *et al*. The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification. *Biochem Biophys Res Commun* 2002;292:587–592.
4. Zhang Z, Bast Jr RC, Yu Y, *et al*. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res* 2004;64:5882–5890.
5. Li J, Zhang Z, Rosenzweig J, *et al*. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem* 2002;48:1296–1304.
6. Adam BL, Qu Y, Davis JW, *et al*. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* 2002;62:3609–3614.
7. Chen YD, Zheng S, Yu JK, *et al*. Artificial neural networks analysis of surface-enhanced laser desorption/ionization mass spectra of serum protein pattern distinguishes colorectal cancer from healthy population. *Clin Cancer Res* 2004;10:8380–8385.
8. Tolson J, Bogumil R, Brunst E, *et al*. Serum protein profiling by SELDI mass spectrometry: detection of multiple variants of serum amyloid alpha in renal cancer patients. *Lab Invest* 2004;84:845–856.
9. Won Y, Song HJ, Kang TW, *et al*. Pattern analysis of serum proteome distinguishes renal cell carcinoma from other urologic diseases and healthy persons. *Proteomics* 2003;3:2310–2316.
10. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* 2004;20:777–785.
11. Diamandis EP. Re: diagnostic potential of serum proteomic patterns in prostate cancer. *J Urol* 2004;171:1244–1245.
12. Diamandis EP, van der Merwe D-E. Plasma protein profiling by mass spectrometry for cancer diagnosis: opportunities and limitations. *Clin Cancer Res* 2005;11:963–965.
13. Diamandis EP. Point: proteomic patterns in biological fluids: do they represent the future of cancer diagnostics? *Clin Chem* 2003;49:1272–1275.
14. Petricoin EF, Liotta LA. Counterpoint: the vision for a new diagnostic paradigm. *Clin Chem* 2003;49:1276–1278.
15. Baggerly KA, Morris JS, Edmonson SR, *et al*. Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J Natl Cancer Inst* 2005;97:307–309.
16. Liotta LA, Lowenthal M, Mehta A, *et al*. Importance of communication between producers and consumers of publicly available experimental data. *J Natl Cancer Inst* 2005;97:310–314.
17. Sorace JM, Zhan M. A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinformatics* 2003;4:24. Available from URL: <http://www.biomedcentral.com/1471-2105/4/24>.
18. Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol Cell Proteomics* 2004;3:367–378.
19. Grizzle WE, Meleth S, Petricoin EF, *et al*. Clarification in the point/counterpoint discussion related to surface-enhanced laser desorption/ionization time-of-flight mass spectrometric identification of patients with adenocarcinomas of the prostate. *Clin Chem* 2004;50:1475–1477.

20. Li J, Orlandi R, White CN, *et al*. Independent validation of candidate breast cancer serum biomarkers identified by mass spectrometry. *Clin Chem* 2005;51:2229–2235.
21. Mathelin C, Cromer A, Wendling C, *et al*. Serum biomarkers for detection of breast cancers: a prospective study. *Breast Cancer Res Treat* 2006;96:83–90.
22. Grizzle WE, Adam BL, Bigbee WL, *et al*. Serum protein expression profiling for cancer detection: validation of a SELDI-based approach for prostate cancer. *Dis Markers* 2003;19:185–195.
23. Grizzle WE, Semmes OJ, Basler J, *et al*. The early detection research network surface-enhanced laser desorption and ionization prostate cancer detection study: a study in biomarker validation in genitourinary oncology. *Urol Oncol* 2004;22:337–343.
24. Fung ET, Yip TT, Lomas L, *et al*. Classification of cancer types by measuring variants of host response proteins using SELDI serum assays. *Int J Cancer* 2005;115:783–789.
25. Le L, Chi K, Tyldesley S, *et al*. Identification of serum amyloid A as a biomarker to distinguish prostate cancer patients with bone lesions. *Clin Chem* 2005;51:695–707.
26. Moshkovskii SA, Serebryakova MV, Kuteykin-Teplyakov KB, *et al*. Ovarian cancer marker of 11.7 kDa detected by proteomics is a serum amyloid A1. *Proteomics* 2005;5:3790–3797.
27. Cho WC, Yip TT, Yip C, *et al*. Identification of serum amyloid a protein as a potentially useful biomarker to monitor relapse of nasopharyngeal cancer by serum proteomic profiling. *Clin Cancer Res* 2004;10:43–52.
28. Langlois RG, Trebes JE, Dalmaso EA, *et al*. Serum protein profile alterations in hemodialysis patients. *Am J Nephrol* 2004;24:268–274.
29. Honda K, Hayashida Y, Umaki T, *et al*. Possible detection of pancreatic cancer by plasma protein profiling. *Cancer Res* 2005;65:10613–10622.
30. Yu JK, Chen YD, Zheng S. An integrated approach to the detection of colorectal cancer utilizing proteomics and bioinformatics. *World J Gastroenterol* 2004;10:3127–3131.
31. Hara T, Honda K, Ono M, *et al*. Identification of 2 serum biomarkers of renal cell carcinoma by surface enhanced laser desorption/ionization mass spectrometry. *J Urol* 2005;174:1213–1217.
32. Banez LL, Prasanna P, Sun L, *et al*. Diagnostic potential of serum proteomic patterns in prostate cancer. *J Urol* 2003;170:442–446.
33. Koomen JM, Shih LN, Coombes KR, *et al*. Plasma protein profiling for diagnosis of pancreatic cancer reveals the presence of host response proteins. *Clin Cancer Res* 2005;11:1110–1118.
34. Gourin CG, Xia ZS, Han Y, *et al*. Serum protein profile analysis in patients with head and neck squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg* 2006;132:390–397.