Urinary proteomic analysis in the context of cancer biomarkers

Iwona Magdalena Bełczącka

Urinary proteomic analysis in the context of cancer biomarkers

Von der Medizinischen Fakultät
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades einer Doktorin der Theoretischen Medizin
genehmigte Dissertation

vorgelegt von

Iwona Magdalena Bełczącka

aus Świdnik (Polen)

Berichter: Frau apl. Professorin
Dr. rer. nat. Vera Jankowski

Frau Universitätsprofessorin
Dr. phil. nat. Gabriele Pradel

Tag der mündlichen Prüfung: 17.12.2019

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

D82 (Diss. RWTH Aachen University, 2019)

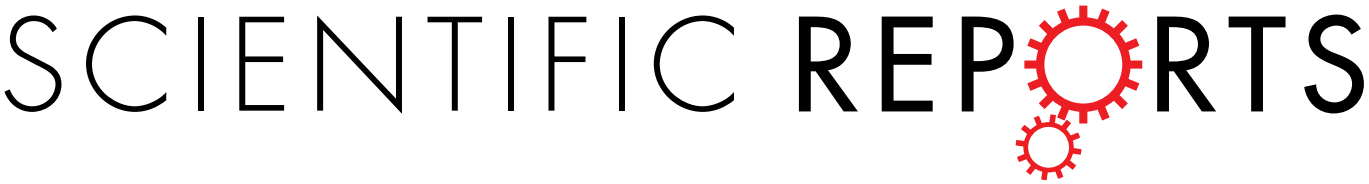
Publications in the dissertation:

1. **Iwona Belczacka**, Agnieszka Latosinska, Justyna Siwy, Jochen Metzger, Axel S. Merseburger, Harald Mischak, Antonia Vlahou, Maria Frantzi, Vera Jankowski. *Urinary CE-MS peptide marker pattern for detection of solid tumors*. Scientific Reports, March 27. 2018. doi: 10.1038/s41598-018-23585-y.
2. **Iwona Belczacka**, Martin Pejchinovski, Magdalena Krochmal, Pedro Magalhães, Maria Frantzi, William Mullen, Antonia Vlahou, Harald Mischak, Vera Jankowski. *Urinary glycopeptide analysis for the investigation of novel biomarkers*. Proteomics Clinical Applications, October 18. 2018. doi: 10.1002/prca.201800111.

Other publications:

1. Pierbruno Ricci, Pedro Magalhaes, Magdalena Krochmal, Martin Pejchinovski, Erica Daina, Maria Rosa Caruso, Laura Goea, **Iwona Belczacka,** Giuseppe Remuzzi, Muriel Umbhauer, Jens Drube, Lars Pape, Harald Mischak, Stéphane Decramer, Franz Schaefer, Joost P. Schanstra, Silvia Cereghini, Petra Zurbig. *Urinary proteome siganture of Renal Cysts and Diabetes syndrome in children*. Scientific Reports, February 18. 2019. Doi: 10.1038/s41598-019-38713-5.
2. David Marx, Jochen Metzger, Jérôme Olagne, **Iwona Belczacka**, Stanislas Faguer, Magali Colombat, Holger Husi, William Mullen, Wilfried Gwinner, Sophie Caillard. *Proteomics in kidney allograft transplantation – Application of molecular pathway analysis for kidney allograft disease phenotypic biomarker selection*. Proteomics Clinical Applications, January 25. 2019. doi.org/10.1002/prca.201800091.
3. Maria Frantzi, Agnieszka Latosinska, **Iwona Belczacka**, Harald Mischak. *Urinary proteomic biomarkers in oncology: ready for implementation?*. Expert Reviews of Proteomics, November 8. 2018. doi: 10.1080/14789450.2018.1547193.
4. **Iwona Belczacka**, Agnieszka Latosinska, Jochen Metzger, David Marx, Antonia Vlahou, Harald Mischak, Maria Frantzi. *Proteomics biomarkers for solid tumor: Current status and future prospects*. Mass Spectrometry Reviews, June 11. 2018. doi: 10.1002/mas.21572.

5. David Marx, Jochen Metzger, Martin Pejchinovski, Ryan Bruce Gil, Maria Frantzi, Agnieszka Latosinska, **Iwona Belczacka**, Silke Sophie Heinzmann, Holger Husi, Jerome Zoidakis, Matthias Klingele, MD, Stefan Herget-Rosenthal. *Proteomics and Metabolomics for AKI Diagnosis*. Seminars in Nephrology, January 2018. doi.org/10.1016/j.semnephrol.2017.09.007.



OPEN

Urinary CE-MS peptide marker pattern for detection of solid tumors

Received: 7 August 2017
Accepted: 9 March 2018
Published online: 27 March 2018

Iwona Belczacka[1,2], Agnieszka Latosinska[1], Justyna Siwy[1], Jochen Metzger[1], Axel S. Merseburger[3], Harald Mischak[1,4], Antonia Vlahou[5], Maria Frantzi[1] & Vera Jankowski[2]

Urinary profiling datasets, previously acquired by capillary electrophoresis coupled to mass-spectrometry were investigated to identify a general urinary marker pattern for detection of solid tumors by targeting common systemic events associated with tumor-related inflammation. A total of 2,055 urinary profiles were analyzed, derived from a) a cancer group of patients (n = 969) with bladder, prostate, and pancreatic cancers, renal cell carcinoma, and cholangiocarcinoma and b) a control group of patients with benign diseases (n = 556), inflammatory diseases (n = 199) and healthy individuals (n = 331). Statistical analysis was conducted in a discovery set of 676 cancer cases and 744 controls. 193 peptides differing at statistically significant levels between cases and controls were selected and combined to a multi-dimensional marker pattern using support vector machine algorithms. Independent validation in a set of 635 patients (293 cancer cases and 342 controls) showed an AUC of 0.82. Inclusion of age as independent variable, significantly increased the AUC value to 0.85. Among the identified peptides were mucins, fibrinogen and collagen fragments. Further studies are planned to assess the pattern value to monitor patients for tumor recurrence. In this proof-of-concept study, a general tumor marker pattern was developed to detect cancer based on shared biomarkers, likely indicative of cancer-related features.

Cancer is one of the leading causes of death worldwide, contributing to 21% of the overall mortality rate[1]. It is also the second leading disease condition affecting people in developing countries[1]. In 2012, 14.1 million newly diagnosed cancer cases and 8.2 million cancer-related deaths were recorded[1]. Moreover, by the year 2030, the global cancer incidence is predicted to be nearly doubled[2]. The major cause of cancer morbidity and mortality is cancer metastasis, accounting for approximately 90% of all cancer related deaths[3]. As the molecular mechanisms that trigger metastasis are still under investigation, it has been shown that cancer metastasis involves a series of pathological changes, starting from local invasion, followed by tumor cell intravasation towards lymphatic and blood vessels and subsequent extravasation to distant organs and tissues leading to secondary tumor formation. In addition, this complex process of human tumor pathogenesis is accompanied by intrinsic inflammatory responses[4].

Despite the remarkable research progress and drug development, the low long-term survival rates of most cancer patients leaves room for improvement in the management of cancer[5]. A cornerstone is, on the one hand, timely detection, when curative treatment is still possible, thus the chances of positive outcome are higher. At the same time, there is also an increasing clinical need for monitoring of cancer relapse and response to treatment. The current diagnostic/monitoring approaches typically involve the combination of imaging techniques and invasive tissue biopsies[6]. For diagnostic screening, the generally preferred specimens are easily accessible body fluids, such as blood serum/plasma or urine due to the relatively non-invasive manner of collection[7,8]. In an effort to increase the chances of detecting cancer earlier, non-invasive screening tests have been introduced into clinical practice, such as: a) carcinoembryonic antigen (CEA)[9], mainly used for monitoring of cholangiocarcinoma, b) cancer antigen 125 (CA-125)[10] primarily used to monitor treatment of ovarian cancer, c) prostate specific antigen (PSA)[11] used for the detection of prostate cancer and d) carbohydrate antigen 19-9 (CA19-9)[12] mainly applied for the detection of pancreatic cancer.

[1]Mosaiques Diagnostics GmbH, Hannover, Germany. [2]University Hospital RWTH Aachen, Institute for Molecular Cardiovascular Research (IMCAR), Aachen, Germany. [3]Department of Urology, University of Lübeck, Lübeck, Germany. [4]University of Glasgow, Institute of Cardiovascular and Medical Sciences, Glasgow, United Kingdom. [5]Biotechnology Division, Biomedical Research Foundation, Academy of Athens (BRFAA), Athens, Greece. Correspondence and requests for materials should be addressed to V.J. (email: vjankowski@ukaachen.de)

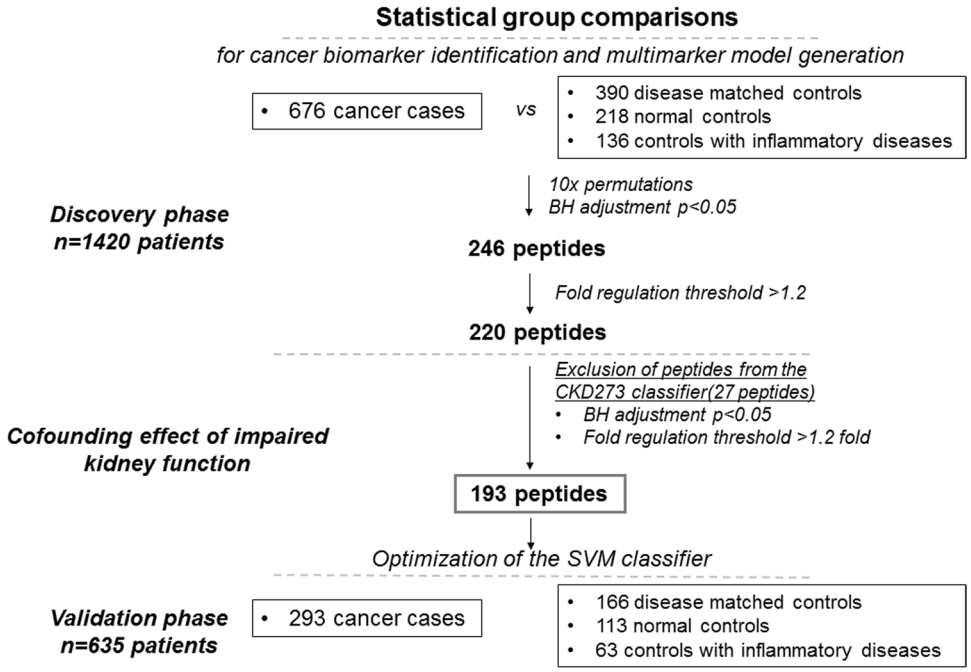


Figure 1. Schematic representation of the study design and workflow of the statistical analyses for the development of the urine-based peptide biomarker pattern. The statistical analysis resulted in the discovery of 193 urinary peptides discriminating between cancer cases and non-malignant controls ($p < 0.05$). ***BH***, *Benjamini-Hochberg adjustment*; ***CKD273***, *chronic kidney disease 273 peptide classifier;* ***p, p****-value.*

Several proteomics studies involving capillary electrophoresis coupled to mass-spectrometry (CE-MS) particularly in the field of cancer[8,13–16] and chronic diseases[17] have demonstrated a good diagnostic potential of urinary peptide biomarkers, when combined to multi-marker patterns. These biomarkers have been identified in the context of a single type of cancer (e.g. bladder, prostate, pancreatic, renal cell carcinoma, cholangiocarcinoma etc.[8,13–16]).

In this proof-of-concept study, the main objective was to investigate pre-existing CE-MS urinary profiling data from different tumors, aiming at the identification of cancer-related markers in urine. The research hypothesis was that cancer progression is associated with common systemic molecular changes, likely reflecting tumor invasion and inflammation. Based on the observation that at least in part, these changes are detectable in urine, we further hypothesized that a combination of such urinary peptides into a multi-marker pattern, may serve as a non-invasive test to monitor cancer progression and metastasis. For this purpose, we evaluated datasets that were previously obtained from five different types of solid tumors including bladder (BCa) and prostate (PCa) cancers, renal cell carcinoma (RCC), cholangiocarcinoma (CC) and pancreatic cancer. To address the expected heterogeneity across different cancer types, urinary peptides were combined into a multi-marker classifier, using a support vector machine (SVM)-based algorithm[13].

Results

Optimization of a urinary peptide multi-marker pattern for detection of solid tumors. The previously acquired CE-MS urinary peptide profiles were compared in a case- control study in a discovery set of 1,420 datasets, aiming at identifying common significant peptides in different types of cancer. The cancer group consisted of 676 cancer cases and the control group of 390 patients with tumor-related benign diseases, 218 healthy individuals and 136 patients with inflammatory diseases (Fig. 1). Statistical comparison in the discovery set between the group of cancer cases (n = 676) and non-cancer controls (n = 744), revealed 246 significantly altered peptides, after adjustment for multiple testing. By applying a fold regulation threshold of 1.2, 220 urinary peptides were retained as potential biomarkers. Of these, 100 were detected with increased and 120 with decreased abundance in the cancer group.

Since impaired kidney function may be a confounder in cancer patients, due to their advanced age, the impact of progressive loss of kidney function was assessed by investigating the potential overlap between the cancer related urinary peptides that were identified in the present study and the previously well-established chronic kidney disease (CKD) specific biomarkers[17]. By comparing the 220 cancer peptides of this study with the 273 CKD specific biomarkers that were previously reported[17], 27 peptides were found to be overlapping and were excluded from the list of the cancer specific biomarkers (Fig. 1). The remaining 193 peptides were selected as potentially cancer specific biomarkers for the establishment of a peptide marker pattern (termed "193-general tumor marker pattern") for detection of solid tumors using SVM (Supplementary Table S1). ROC analysis of the SVM-model on the discovery set, revealed an area under the ROC curve (AUC) of 0.85 [0.83 to 0.87 (95% CI); $p < 0.0001$]. In comparison, the diagnostic performance of the 193 peptides when assessed individually, ranged from 0.54–0.62

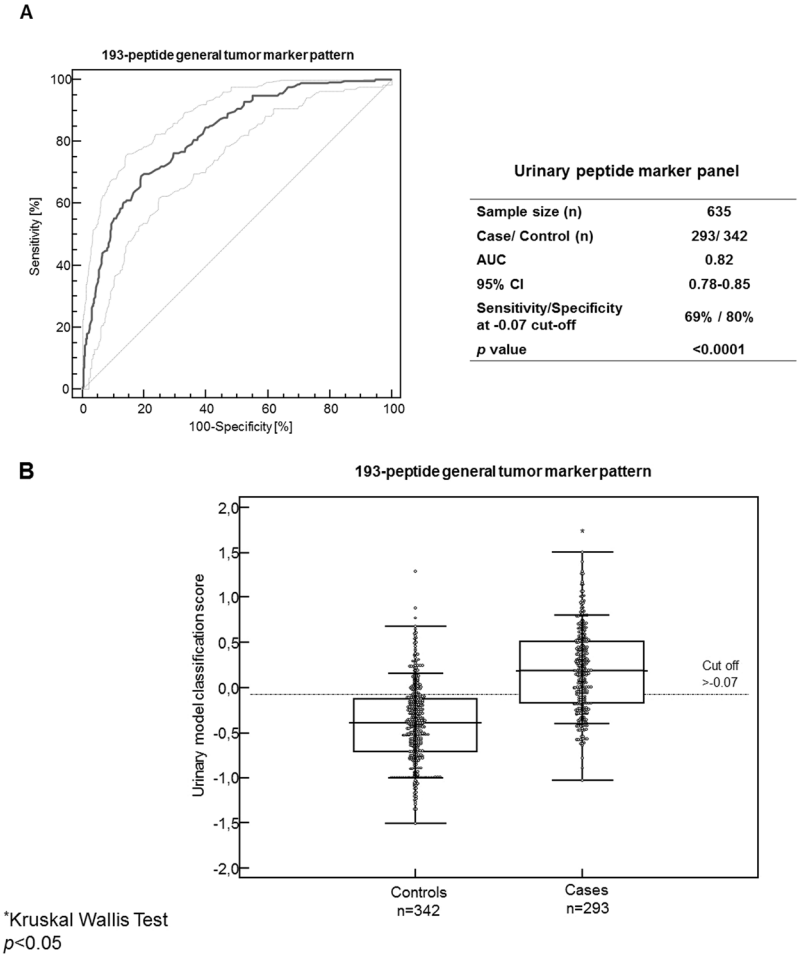
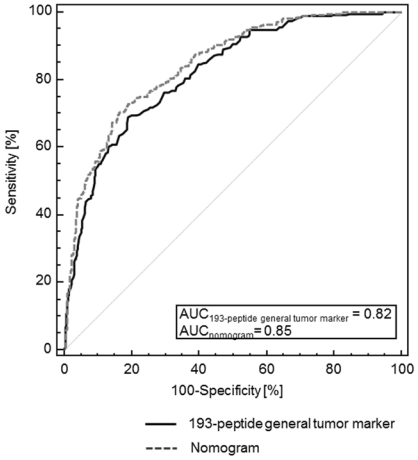


Figure 2. (**A**) ROC curve analysis for the urinary biomarker peptide pattern, consisting of 193 peptides, as performed in the independent validation cohort. The AUC, 95% CI, and ***p*** value are also provided for the classification of cancer patients. (**B**) The classification scores presented in Box-and-Whisker plots are displaying the level of discrimination between cancer cases and non-malignant controls and the distribution of the classification values in each population group. (**A**) Rank-test was performed using Kruskal-Wallis test. The average rank differences were significantly different ($p < 0.05$) between the controls and cancer cases. ***AUC**, area under the curve; **CI**, confidence intervals, **ROC**, receiver operating characteristics; **p, p**-value.*

(AUC values), as presented in Supplementary Table S1. Seventeen of these peptides appear to exhibit certain cancer-type specificity as suggested following a detailed investigation of the distribution of all 193 peptides across the different cancer types (Supplementary Table S2). Since the study was not powered for this analysis, confirmation of this preliminary observation in larger datasets per cancer type is necessary to substantiate any claim of tumor specificity.

Validation of the 193- general tumor marker pattern. The 193-general tumor marker pattern was subsequently validated in an independent cohort of 635 individuals, consisting of 293 cancer cases and 342 controls including: 166 patients with benign diseases (cancer-matched controls), 113 healthy individuals and 63 patients with inflammatory diseases. ROC analysis on this validation set revealed an AUC of 0.82 [0.79 to 0.85 (95% CI); $p < 0.0001$]. At the pre-determined cut-off level of −0.07, the peptide marker pattern's sensitivity and specificity were estimated at 69% and 81%, respectively (Fig. 2A). As presented in the Box-and-Whisker plot representation of Fig. 2B, the classification scores of the peptide marker pattern between the cancer case and non-cancer control groups were significantly different in post-hoc analysis after a positive Kruskal-Wallis test ($p < 0.0001$).

To test if the 193-general tumor marker pattern could be applied to both solid and non-solid cancer types, we performed a second-phase validation in additional sets of cancer patients, available in our database of urinary peptide profiles including: a) acute myeloid leukemia (AML), n = 27 and b) patients with other solid tumor types, n = 6 [neuroendocrine cancer, n = 2 and single patients with breast cancer, small cell lung carcinoma (SCLC), bronchial cancer and hemangioma]. In the case of AML, 20 out of 27 patients were classified positively corresponding to a sensitivity of 74%. In other type of tumors (n = 6), 5 out of 6 patients scored positive resulting in a sensitivity of 83%. Additionally, we validated the 193-general tumor marker pattern in a small group of metastatic



Logistic regression analysis of clinical variables

Variable	*Regression coefficient*	*Standard error*	*Significance level*
193-peptide general tumor marker	2.68	0.25	<0.0001
Age	0.04	0.008	<0.0001

Pairwise comparison of ROC curves

Sample size (n)	635
Case/Controls (n)	293/342
95% Confidence Interval	0.01 - 0.04
Significance level	$p = 0.002$

Figure 3. Logistic regression analysis of ROC curve and age as independent variable. ***ROC****, receiver operating characteristics;* ***AUC****, area under the ROC curve.*

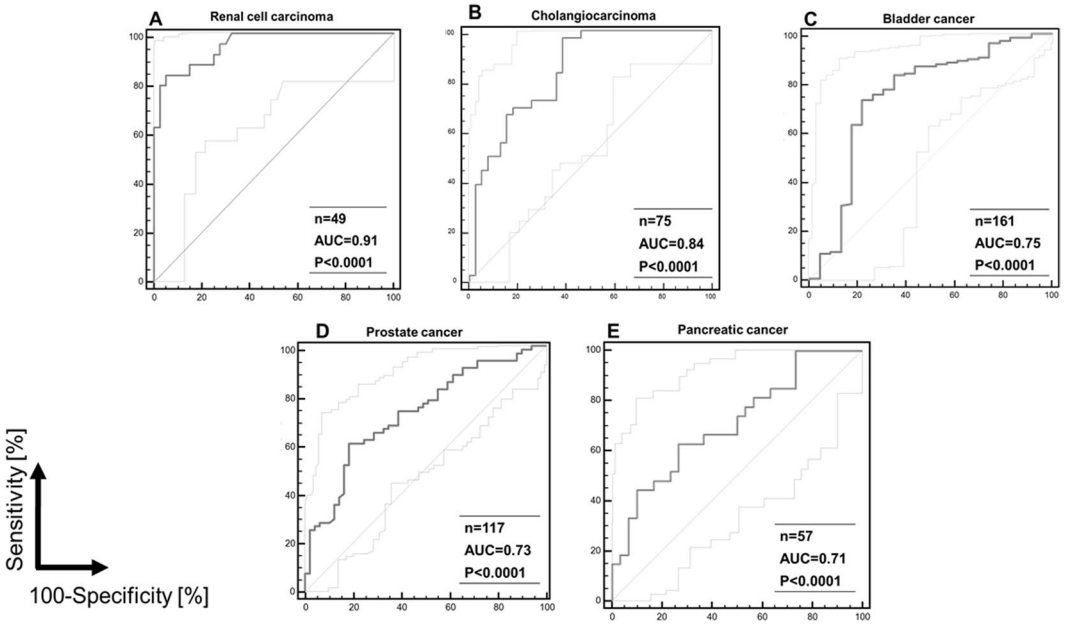


Figure 4. ROC curve analysis for the performance of combined patterns for: (**A**) renal cell carcinoma; (**B**) cholangiocarcinoma; (**C**) bladder cancer; (**D**) prostate cancer; (**E**) pancreatic cancer. ***ROC****, receiver operating characteristics;* ***AUC****, area under the ROC curve.*

cancer cases (metastatic colon cancer, n = 1; metastatic breast cancer, n = 3, metastatic stomach cancer, n = 3). Six out of 7 of these cases scored positive corresponding to a sensitivity of 86%.

Establishment of a diagnostic nomogram. Consecutive logistic regression analysis was performed to investigate the impact of age and gender as independent variables, on the disease occurrence. While gender did not affect significantly the occurrence, combination of the 193- general tumor marker pattern with age ($p < 0.0001$) slightly, but significantly improved the AUC value from 0.82 to 0.85 [0.009–0.040 (95% CI); $p < 0.002$] (Fig. 3). As proven superior, the above nomogram was considered for further analyses. The aforementioned nomogram, generated based on the combination of the 193- general tumor marker pattern with age, was further tested separately in each group of cancer patients from the validation set (specific cancer cases and respective disease-matched controls as listed in–Supplementary Table S3) resulting in AUC values ranging from 0.71 to 0.91. Particularly, in the case of 24 RCC patients and 25 negative for RCC controls, the nomogram resulted in a high AUC value of 0.91 [0.79–0.98 (95% CI); $p < 0.0001$] (Fig. 4A). Similarly, for the 36 CC patients and 39 patients with benign biliary diseases an AUC value of 0.84 was obtained [0.74–0.92 (95% CI); $p < 0.0001$) (Fig. 4B). For the discrimination of the 138 BCa patients from the 23 respective urological controls (patients presenting with haematuria or other benign urological disorders) an AUC value of 0.75 [0.67–0.81 (95% CI); $p < 0.0001$] (Fig. 4C) was received, whereas 68 PCa patients and 49 patients with BPH or prostatitis

	Nomogram*	Previously published model	*Significance level (p-value)*	Combined patterns**	*Significance level (p-value)*	Reference
Bladder cancer	AUC = 0.75	AUC = 0.79	0.62	AUC = 0.84	0.16	Frantzi M. *et al.* 2015[14]
Prostate cancer	AUC = 0.66	AUC = 0.80	0.19	AUC = 0.84	0.30	Theodorescu D. *et al.* 2008[16]
Cholangio-carcinoma	AUC = 0.84	AUC = 0.89	0.50	AUC = 0.93	0.24	Metzger J. *et al.* 2016[8]
Pancreatic cancer	AUC = 0.73	AUC = 0.87	0.12	AUC = 0.87	0.35	Schonemeier B. *et al.* 2016[15]
Renal cell carcinoma	AUC = 0.97	AUC = 1.00	0.42	AUC = 1.00	1.00	Frantzi M. *et al.* 2014[13]

Table 1. Comparative analysis of the nomogram with the previously published specific biomarker panels. AUC, Area under the curve. *Nomograms were generated based on the combination the 193-general tumor marker pattern with age as independent variable. Significance p-value was investigated between the nomograms and the previously published models. **Combined patterns were developed based on the previously published specific peptide marker patterns for each type of cancer and the developed nomogram. Significance p-value was investigated between the combined patterns and the previously published models.

were discriminated with an AUC value of 0.73 [0.64–0.80 (95% CI); $p < 0.0001$] (Fig. 4D). For the 27 patients with pancreatic cancer and 30 patients with chronic pancreatitis the AUC value was 0.71 [0.58–0.82 (95% CI); $p < 0.0001$] (Fig. 4E).

Comparative analysis with the previously published CE-MS marker patterns. A comparison of the scoring data obtained with the use of the present nomogram and the previously published specific classifiers for each individual cancer type[8,13–16] was subsequently performed. Only patients that were common in the independent validation sets of the previously published and the present study were compared. A direct comparison was possible for n = 161 BCa, n = 41 PCa, n = 33 CC, n = 17 RCC and n = 31 pancreatic cancer study subjects. Based on the results of these comparative ROC analyses, the previously described biomarkers for each cancer type exhibited slightly higher accuracy in cancer detection in all cases, even though the difference did not reach statistical significance for the specific sample sets (Table 1, Supplementary Figure S1).

To investigate the added value of combining the introduced nomogram score with that of the specific cancer patterns, logistic regression analysis was performed. When examining all data combined (BCa, PCa, CC, RCC and pancreatic cancer) the combination of the tumor-specific patter with the nomogram resulted in slightly higher accuracy, (yet again not reaching statistical significance, in the available small sample set), comparing to the nomogram alone (Table 1). In detail, the AUC values as received by the combined peptide marker patterns were for BCa: $AUC_{combined\ patterns} = 0.84$, for PCa: $AUC_{combined\ patterns} = 0.84$, for CC: $AUC_{combined\ patterns} = 0.93$, for RCC: $AUC_{combined\ patterns} = 1.00$, for pancreatic cancer: $AUC_{combined\ patterns} = 0.87$ (Table 1).

Comparative analysis of the nomogram with the common laboratory single biomarker tests. A direct comparison of the nomogram with common laboratory single biomarker tests was possible for CA19-9 and CEA levels in the case of pancreatic cancer and PSA levels in the case of PCa. When comparing the nomogram score in pancreatic cancer patients (n = 57) with the CA19-9 and CEA levels, the nomogram presented slightly higher performance than CA19-9 ($AUC_{nomogram} = 0.71$ vs $AUC_{CA19\text{-}9\ level} = 0.70$, $p = 0.88$) and lower performance than CEA ($AUC_{nomogram} = 0.71$ vs $AUC_{CEA\ level} = 0.73$, $p = 0.83$); however in both comparisons the differences did not reach statistical significance. In the case of prostate cancer, (n = 106) the nomogram presented higher overall AUC value (0.74) in PCa detection compared to PSA, with the latter presenting an AUC of 0.63; however this difference was not significant ($p = 0.10$). Logistic regression analysis was used for the combination of the nomogram with the commonly used single biomarker tests. In case of pancreatic cancer, combination of the nomogram with CA19-9 and CEA resulted in slightly increased AUC values ($AUC_{nomogram + CA19\text{-}9\ level} = 0.81$, $p = 0.87$; $AUC_{nomogram + CEA\ level} = 0.83$, $p = 0.07$). In the case of prostate cancer, the combination of the nomogram with PSA presented slightly higher performance ($AUC_{nomogram + PSA\ levels} = 0.75$, $p = 0.10$). In both cases, the analysis did not reach statistical significance.

Assessment of invasion and tumor-related inflammation. Based on the initial hypothesis that urinary peptides may reflect tumor invasion, the nomogram consisting of the 193-general tumor marker pattern and age, was further investigated for its potential to detect invasion. The cancer cases from the validation set were grouped into invasive (n = 54) and non-invasive cases (n = 98). Surprisingly, and as presented by Box-and-Whisker plot (Supplementary Figure S2), the 193-general tumor marker pattern showed similar discrimination abilities for detecting non-invasive and invasive cancer cases (AUC = 0.856 for non-invasive vs AUC = 0.863 for invasive cancer cases).

As presented, the data indicate that the 193-general tumor marker pattern does not appear to reflect local invasion of the tumor but can detect metastasis with increased accuracy. As part of cancer related systemic features, systemic inflammation was also considered. To investigate the specificity of the peptides in terms of inflammation related to cancer, an additional set of patients with autoinflammatory diseases (hypersensitivity vasculitis (HV), n = 58, and systemic lupus erythematosus (SLE), n = 34) was used to further validate the 193-general tumor marker pattern (Table 2). The mean age of patients with autoinflammatory diseases was significantly different from the cancer patients in the validation set, therefore the nomogram could not be applicable. The specificity values were evaluated at the pre-determined cut-off of −0.07, where the sensitivity for cancer diagnosis was estimated to be 69%. At the indicated cut-off, specificity values for different autoinflammatory related diseases

Type of cancer	Group	Discovery set	Mean age (range)	Validation set	Mean age (range)	Sample size (%)
Prostate cancer	Case	N = 138	68 (41–87)	N = 68	68 (45–86)	10%
	Control	N = 130	66 (51–87)	N = 49	66 (51–83)	9%
Pancreatic cancer	Case	N = 61	65 (35–88)	N = 27	63 (33–90)	4%
	Control	N = 62	52 (27–77)	N = 30	52 (25–73)	4%
Renal cell carcinoma	Case	N = 46	62 (41–81)	N = 24	61 (33–79)	3%
	Control	N = 80	58 (23–84)	N = 25	59(23–84)	5%
Cholangio-carcinoma	Case	N = 82	56 (22–81)	N = 36	58 (30–86)	6%
	Control	N = 83	49 (20–83)	N = 39	50 (22–85)	6%
Bladder cancer	Case	N = 349	67 (28–95)	N = 138	69 (27–87)	24%
	Control	N = 23	60 (43–88)	N = 23	65 (33–93)	3%
Normal controls	Control	N = 167	60 (19–86)	N = 89	60 (22–86)	12%
	Individual with smoking history	N = 51	43(19–72)	N = 24	46(20–72)	4%
Inflammation controls	Control	N = 136	46(16–88)	N = 63	47(19–78)	10%
Total	**Case**	**N = 676**		**N = 293**		**47%**
	Control	**N = 744**		N = **342**		**53%**
Additional cancer types						
Acute myeloid leukemia				N = 27		
Other tumor types				N = 13		
Independent specificity analysis						
Inflammatory conditions		N = 92		36 (19–63)		

Table 2. Clinical data of the patients used in the discovery and validation phases of the study. The patient cohorts were adjusted for age to be further sub-grouped in a discovery and a validation set.

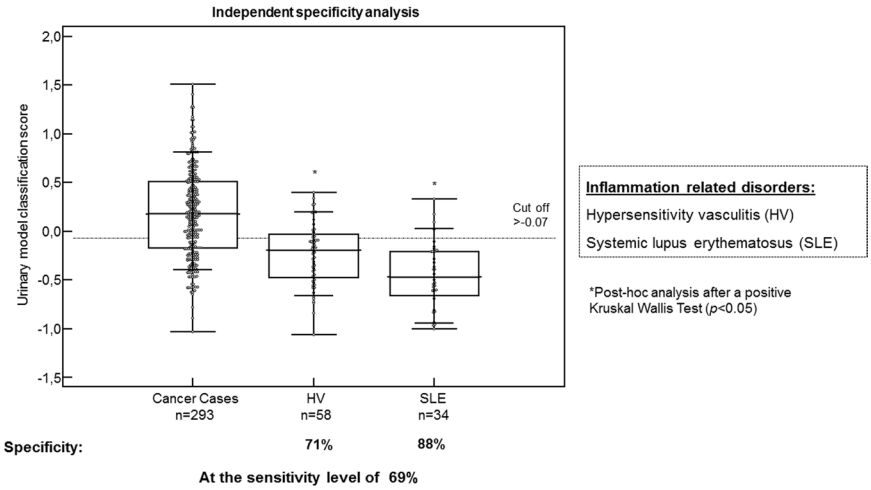


Figure 5. Distribution of classification values for patients with autoinflammatory diseases. The following disease aetiologies were included in the specificity analysis as reference controls: hypersensitivity vasculitis (n = 58) and systemic lupus erythematosus (n = 34). ***HV**, hypersensitivity vasculitis; **SLE**, systemic lupus erythematosus.*

were 71% and 88%, for HV and SLE, respectively. As presented in Fig. 5, the scoring of the patients with inflammatory disease was not significantly different from the cancer or normal controls: the 193- general tumor marker pattern was able to significantly discriminate cancer cases from autoinflammatory diseases, as demonstrated in a post-hoc statistical analysis after a positive Kruskal-Wallis test ($\boldsymbol{p} < 0.05$) (Fig. 5).

Peptide sequence identification. The peptide amino acid sequence was obtained for 133 out of the 193 urinary peptide biomarkers, representing a percentage of 69%. In our study, the vast majority (71%) of the sequenced CE-MS general cancer biomarkers are fragments of various collagen chains, most likely excreted into the urine by proteolytic cleavage of the extracellular matrix. Several fibrinogen α and β chain peptide fragments were also included in the marker pattern, which might be also the remark of inflammatory processes. Besides collagens, several peptide fragments originated from α- and γ-subunits of Na/K-ATPase. Single peptides from stablin-2, 14-3-3 sigma protein, keratin-1, mucin-12 and mucin-16, myeloperoxidase and apolipoprotein A-IV were also among the sequenced peptide markers. Some of the CE-MS ion peaks could not be identified by MS/MS

likely due to incomplete fragmentation and/or unknown post-translational modifications that prevent alignment of the fragment spectrum to the available sequence databases[18].

To assess the "tissue invasion" hypothesis, we further selected only the collagen and fibrinogen peptide fragments and evaluated their diagnostic capabilities. ROC analysis in the independent validation set revealed an AUC value of 0.77 [0.73 to 0.80 (95% CI); $p < 0.0001$]. The AUC values for each particular cancer type were: for bladder cancer: AUC = 0.68; for prostate cancer: AUC = 0.66, for cholangiocarcinoma AUC = 0.82, for renal cell carcinoma AUC = 0.85 and for pancreatic cancer: AUC = 0.54. The SVM based pattern that was developed based on collagen and fibrin peptide fragments alone, exhibited significantly lower performance than the 193- general tumour marker pattern (AUC = 0.82, $p < 0.0001$).

Discussion

In this study, we aimed at the development of a general tumor marker panel based on pre-existing CE-MS urinary profiling data, being available from previously published studies[8,13–16]. The study was based on the hypothesis that tumor growth is associated with general molecular changes, representing systemic inflammation triggered by tissue invasion and metastasis. Therefore, general tumor-associated changes may be valuable in monitoring patients for response to therapy, and/or tumor relapse. As such, the clinical context for application of such panel is not envisioned in cancer initial detection/screening, (where identification of the specific tumor type is obviously a requirement), but rather as an adjunct during disease monitoring, most likely in combination also with tumor-specific markers, as available (such as CA125, PSA, as described below). As a proof-of-concept study, we evaluated previously published CE-MS peptide profiles of more than 2,000 cancer patients and individuals with non-malignant diseases and developed a marker panel of 193 peptides presenting statistical significant changes between the (multi)cancer group and controls. Based on previous results[8,13,14], the combination of several individual peptide markers in an SVM-based multi-marker pattern allows counteracting variability in single biomarkers, hence for higher accuracy, particularly regarding for cancer types characterised by very high intra-tumor heterogeneity.

The sensitivity and specificity levels of the 193 peptide panel revealed satisfactory discriminatory ability in solid, but also non-solid tumors, with sensitivity values to be higher in detecting metastasis at least based on a small set of available metastatic samples. Compared to the previously published cancer type-specific patterns (BCa, PCa, CC, RCC and pancreatic cancer patterns), and with the acknowledgement that the power of this comparison is compromised by the available small sample sizes, the 193- general tumor marker pattern performed with slightly lower AUC values for each individual cancer type. However, combination of the two tests increased accuracy in cancer detection, further suggesting that such a panel may be useful as an adjunct during cancer monitoring, in combination to more cancer type specific tests.

Along these lines, a direct comparison with commonly applied diagnostic tests when available, (CA19-9 and CEA levels for pancreatic cancer and PSA levels for prostate cancer), improved the discrimination of cancer patients, even though the difference was again not significant. These results collectively underscore the need for validation studies further testing the accuracy of the 193-peptide marker pattern in comparison and/or in combination to cancer type specific markers, as an effort to combine knowledge on cancer "systemic" reactions with cancer-type specific events towards improving disease monitoring.

From the 193 tumor-specific peptides, sequence information could be retrieved for 133 peptides. Most of the tumor-specific peptides are fragments of collagen chains, possibly reflecting molecular changes in the extracellular matrix (ECM) organization and the altered activity of ECM-degrading proteases during tumor progression[19]. In the present study, increased urinary levels of fibrinogen-derived peptide fragments were identified in cancer patients. Importantly, one specific peptide, produced by the cleavage of fibrinogen beta chain (FGB) was frequently identified as cancer biomarker in three previously published CE-MS studies, investigating bladder cancer, renal cell carcinoma and pancreatic cancer respectively[13–15]. Fibrinogen is important in tumor stroma-formation, as one of the major components linking the tumor cells to the ECM and once converted to insoluble fibrin by activated thrombin, is responsible for the promotion of platelet coagulation. In addition, fibrinogen is considered as an acute phase reactant protein and is elevated during inflammation processes. In leukemias, high fibrinogen levels are associated with systemic inflammatory response, that is frequently present in AML, and is associated with poor outcome[20].

Additionally, in the presented study, decreased urinary levels of a myeloperoxidase (MPO)- derived peptide were detected in urine of cancer patients, in comparison to the non-cancer controls. MPO is a lysosomal enzyme produced in high amounts by neutrophilic granulocytes in respect to inflammatory responses[21]. Multiple studies link the G463A polymorphism in the MPO promoter region, which decreases the transcription and activity of MPO, with a lower cancer risk for a variety of human cancers e.g. pancreatic cancer[22], bladder cancer[23] and gastric cancer[24]. In addition, peptides from mucin 16 (MUC16) were detected with decreased levels in the urine of cancer patients. Mucins are heavily O-glycosylated proteins released from the cell surface as a result of proteolytic cleavage of their transmembrane segment[25]. MUC16 has been reported to be an important factor in tumor diagnosis since it contains the CA125 cancer antigen which is strongly deregulated in pancreatic cancer. Moreover, MUC16 contributes to the ovarian cancer growth and its metastatic activity[25]. The cancer cells expressing low levels of MUC16 on their surface are the preferential target for the Natural killer (NK) cells, contributing to immunoediting[25].

To further investigate the hypothesis that specific urinary peptides are related to tissue invasion, we selected only the peptides that are expected to be involved in ECM remodelling and tissue rearrangement during cancer invasion, such as collagens and fibrinogen fragments. However, by combining them in specific peptide pattern, the performance was significantly lower (AUC = 0.77, $p < 0.0001$), compared to the 193- general tumor marker pattern. When investigating the discriminatory potential in the differentiation of non-invasive and invasive cancer cases, the results also supported that there is no significant difference in the diagnostic potential between

non-invasive and invasive tumors. This observation was further supported by the fact that the general tumor marker pattern detected AML patients with good sensitivity (74%). Considering the above results and the fact that AML is a non-solid tumor, we concluded that the molecular changes reflected by the urinary peptide profiles are more likely related to general systemic effects during cancer progression rather than local tumor invasion. This is also suggested by the fact that the vast majority of peptides included in the panel do not appear to be cancer-type specific.

As revealed by the specificity analysis in patients that present inflammation but have no cancer, the established nomogram can well differentiate between the patients presenting autoinflammatory diseases and those suffering from cancer. The good discriminatory capability of the established general tumor pattern suggests that the urinary peptides included in marker pattern may be specific to cancer-related inflammatory processes, distinctive from inflammation properties that occur during autoimmune disease. This observation is in line with the literature, suggesting that tumor-associated inflammation seems to be promoted through a distinct intrinsic pathway[26]. More specifically, a unique orchestration of chemokines, secreted factors and immune cells is present at all stages of tumorigenesis, contributing to the tumor initiation and progression. Even more prominently, particular molecules are reported to be involved specifically during metastasis, where inflammation seems to promote and assist the colonization of the tumor cells in the metastatic niche[4,26].

There are some limitations to be noted in the current work which warrant further consideration. Although performed in a large cohort, the presented proof-of-concept study was retrospective.

As such, the clinical and pathological data in some cases were missing, among others the level of the standard biochemical biomarkers (PSA, CA19-9 and CEA) and stage. Moreover, information about the patient's comorbidities and additional information on the administered treatment was not possible to be retrieved. Therefore, the diagnostic ability of the 193-peptide marker pattern needs to be further confirmed prospectively in a proper clinical setting: during cancer monitoring and in parallel to well established cancer-type specific markers.

Based on the results, the 193 tumor-specific peptides combined into a marker pattern presented good accuracy, in comparison to the single peptide biomarkers alone. Addition of age as independent covariate slightly, yet significantly, improved the performance of the multi-marker pattern. However, considering the advanced age of cancer patients (mean of age 67) the developed nomogram may not be applicable to significantly younger individuals (e.g. AML patients with mean value for age of 50).

In addition, some of the peptides (31%) included in the general tumor marker could not be sequenced, mainly due to the incomplete fragmentation or presence of unknown post-translational modifications. Nonetheless, a large number of peptide sequences were retrieved and found to reflect processes involved mainly in cancer related systemic events, further supporting the biological validity of the findings.

Collectively, through presented study we aimed to identify commonly-shared peptide markers between five cancer types that are more likely related to general systemic effects during cancer progression. The envisioned application is a general tumor marker that can detect relapse of tumor is the primary site and/or distant metastasis even after the primary tumor has been removed. Such a marker pattern would have potential not only in monitoring cancer recurrence, but also response to the treatment, e.g. to chemotherapy. However, to support these aims, a prospective study in an appropriate number of patients is needed, where the biomarker pattern is tested in comparison to the current state-of-the-art.

Methods

Patient selection. In this multi-centre study, a case-control comparison was performed based on available CE-MS datasets from previous studies[8,13–16,27]. The study was performed in accordance with the guidelines for biomarker identification and reporting in clinical proteomics[28] and the REMARK Reporting Recommendations[29]. The study was conducted in accordance with the Declaration of Helsinki and ethical approval was obtained by the Ethics Committee from Hannover Medical School in Germany (ID: 3409-2016). Proper informed consent procedures were followed under Institutional Review Board–approved protocols of the participating centers. All experimental protocols for sample collection and processing were performed according to relevant guidelines and the standard protocols for urine collection, as defined by the European Kidney and Urine Proteomics (EuroKUP) and Human Kidney and Urine Proteome Project (HKUPP) networks and described in the respective publications. Available urinary CE-MS peptide profiling datasets from patients presenting five types of solid tumors were selected from an internal database[27]. The distribution of the sample size per cancer type is presented in Supplementary Table S3. The clinical and demographical data per patient are reported in Supplementary Table S4. A total of 2,055 urinary profiles from case and control groups (Table 2) were statistically evaluated for differences in their peptide composition. A total number of 487 urinary peptide profiling datasets from patients with bladder cancer and 58 from disease-matched controls with non-malignant urological pathologies (e.g. haematuria, benign urologic diseases) were included in the analysis, as described in[14]. Peptide datasets from urine samples collected from prostate cancer patients (n = 206) and patients with benign prostate diseases (e.g. benign prostatic hyperplasia or prostatitis; n = 179), described in[16] were also included. Previously published CE-MS profiling datasets[13] from 70 renal cell carcinoma patients and 105 individuals negative for malignancy were also included. Urinary CE-MS datasets from 240 patients with cholestasis were considered[8]. Among the 240 patients, 118 patients were diagnosed with cholangiocarcinoma and 122 patients with benign diseases (e.g. primary sclerosing cholangitis (PSC) or benign biliary disorders (BBD)), were used as disease matched controls[8]. Moreover, urinary datasets from patients with pancreatic cancer (n = 88) or chronic pancreatitis (n = 92)[15] were included in the analysis. Additional urinary CE-MS datasets included in the control group were from non-diseased subjects (n = 256), individuals with smoking history (n = 75) and patients with inflammatory diseases (n = 199) (e.g. arteriosclerosis, IgA nephropathy and systemic lupus erythematosus).

The general tumor marker pattern was tested in additional groups of cancers including: acute myeloid leukemia (AML), n = 27[30] and patients with other solid tumors: neuroendocrine cancer, n = 2; metastatic colon

cancer, n = 1; metastatic stomach cancer, n = 3; metastatic breast cancer, n = 3; breast cancer, n = 1; small cell lung carcinoma, n = 1; bronchial cancer, n = 1; and hemangioma, n = 1[27]. Additionally, an independent analysis was performed to assess the specificity of the general tumor marker pattern in a control group of patients with inflammation-related diseases, like HV (n = 58) and SLE (n = 34), as previously published (Table 2)[31].

Evaluation of the CE-MS acquired datasets. In the context of this study, previously published CE-MS data sets were investigated[8,13–16]. In all of these, the same analytical workflow had been followed following standard operating protocols (SOPs) for routine diagnostics[8,32]. The established SOPs and ISO standards, allow to obtain highly comparable data, which are further integrated in the database, in order to perform comparisons targeting a wide spectrum of clinical applications.

The sample collection was performed in according to the established SOPs described in the respective publications. In brief, 10 ml of sample were collected, splitted into aliquots and immediately frozen at −20 °C. No protease inhibitors were added prior to the analysis. The sample preparation was conducted based on established SOPs, as previously described[8,32]. Briefly, 700 µl urine aliquots were diluted 1:2 in an alkaline buffer containing 2 M urea, 10 mM NH_4OH and 0.02% SDS (pH 10.5). Subsequently, the samples were filtered by Centrisart ultracentrifugation filters (Sartorius, Göttingen, Germany) to remove proteins below 20 kDa, followed by desalting over PD-10 columns (GE Healthcare, Munich, Germany). The peptide extracts were lyophilized and resuspended in high-performance-liquid-chromatography-grade water. CE-MS analysis and data processing was performed according to ISO13485 standards yielding quality controlled urinary data sets[32]. Mass spectral ion peaks representing identical molecules at different charge states were de-convoluted into single masses using MosaiquesVisu software[13,32]. The peak list characterizes each peptide by its molecular mass [kDa], normalized migration time [min] and normalized signal intensity [AU][13]. Normalization of the CE-MS data was based on 29 collagen fragments that are not affected by disease and serve as internal standards[27]. After normalisation, all detected peptides were deposited, matched, and annotated in a Microsoft SQL database and used as input in the presented study. As proteomics datasets frequently contains missing values, due to biological and/ or technical factors, in this study missing values ("0 values") were replaced by the minimum intensity value per dataset. Subsequent transformation of the data (log-transformation) was performed, as previously described[33].

Statistical analysis. A case-control statistical comparison was conducted to detect potentially significant cancer specific-peptide biomarkers (Fig. 1). The datasets were grouped in a cancer (n = 969) and a control set (n = 1,086) and divided into a discovery and validation set, as described below and presented in Fig. 1.

Discovery phase. The peptide profiles were compared for differences in the peptide urinary excretion levels between cancer cases (n = 676) and non-cancer controls (n = 744) by applying the Wilcoxon rank sum test[33]. To increase the validity of the statistical approach, permutation analysis was performed by excluding randomly 30% of the samples, further repeated ten times. Statistical correction of the estimated $\boldsymbol{p}$ values for multivariate testing was performed based on the Benjamini-Hochberg method[34]. The common significant identified peptides ($\boldsymbol{p} < 0.05$ after Benjamini-Hochberg adjustment), obtained in all ten permutation analyses and exhibiting at least a 1.2-fold change between cases and controls were considered for further analysis. The presence of cancer-type specific peptides among the 193 peptides was investigated by performing a similar analysis in each cancer type set separately.

SVM Optimization. The urinary marker pattern was optimized in the training set, using the SVM-based MosaCluster software (version 1.7.0), developed for the classification of samples in a high dimensional parameter space according to SVM-based algorithms. The classifier was optimised based on the shortlisted cancer specific biomarkers with each biomarker representing one dimension in the n-dimensional parameter space[13]. The SVM kernel parameters were defined and applied during classifier validation as following: C = 4.096, gamma = 0.0064 and epsilon = 0.001.

Validation phase. Validation was performed in an independent set of 635 samples, grouped into 293 cancer cases and 342 non-malignant controls. The sensitivity and specificity estimates for the SVM-based peptide marker pattern were calculated based on the number of correctly classified samples. The receiver operating characteristic (ROC) plots and the confidence intervals (95% CI) were based on exact binomial calculations performed in MedCalc 12.7.5.0 (Mariakerke, Belgium). The optimal cut-off level was pre-determined according to the criterion value calculated based on the Youden index J, giving an equal weight to sensitivity and specificity (MedCalc). Statistical comparisons of the classification scores between the cancer cases and control groups of the validation cohort were performed using Kruskal-Wallis rank sum test (MedCalc).

Logistic regression analysis. Logistic regression analysis was used to evaluate clinical impact of independent clinical variables (e.g. age, gender, PSA, CA 19-9 and CEA levels) on the diagnostic outcome of the multi-marker pattern. Diagnostic nomograms were developed, by combining the 193-general tumor marker pattern with age, CEA, CA19-9 and PSA levels (MedCalc). The same method was applied to assess the added value of the 193-general tumor marker pattern over the previously published CE-MS-based classifiers.

Peptide sequencing and matching. The amino acid composition was acquired by MS/MS analysis using either a PACE CE or a Dionex Ultimate 3000 RSLS nanoflow system (Dionex, Camberly UK) coupled to an Orbitrap Velos instrument (Thermo Scientific), as previously described[18]. Protein matching and data analysis was based on Proteome Discoverer 1.2 (activation type: HCD; precursor mass tolerance: 5 ppm; fragment mass tolerance: 0.05 Da). The data were searched against the UniProt human database[35] without enzyme specificity.

Matching of the amino acid sequences with the CE-MS acquired ion peaks was based on mass correlation between CE-MS and liquid chromatography-tandem mass spectrometry analysis (LC-MS/MS). Further validation of the obtained peptide identifications is based on the assessment of the peptide charge at the working pH of 2.2 and the CE-migration time results[32].

Data avability. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Ferlay, J. *et al*. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J Cancer* **136**, E359–E386 (2015).
2. Thun, M. J., DeLancey, J. O., Center, M. M., Jemal, A. & Ward, E. M. The global burden of cancer: priorities for prevention. *Carcinogenesis* **31**, 100–110 (2010).
3. Mehlen, P. & Puisieux, A. Metastasis: a question of life or death. *Nature Reviews Cancer* **6**, 449–458 (2006).
4. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, Inflammation, and Cancer. *Cell* **140**, 883–899 (2010).
5. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2017. *Ca-a Cancer Journal for Clinicians* **67**, 7–30 (2017).
6. Kulasingam, V. & Diamandis, E. P. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature Clinical Practice Oncology* **5**, 588–599 (2008).
7. Hortin, G. L. & Sviridov, D. Diagnostic potential for urinary proteomics. *Pharmacogenomics* **8**, 237–255 (2007).
8. Metzger, J. *et al*. Urine proteomic analysis differentiates cholangiocarcinoma from primary sclerosing cholangitis and other benign biliary disorders. *Gut* **62**, 122–130 (2013).
9. Ince, A. T. *et al*. Roles of serum and biliary CEA, CA19-9, VEGFR3, and TAC in differentiating between malignant and benign biliary obstructions. *Turkish Journal of Gastroenterology* **25**, 162–169 (2014).
10. Kristjansdottir, B., Levan, K., Partheen, K. & Sundfeldt, K. Diagnostic performance of the biomarkers HE4 and CA125 in type I and type II epithelial ovarian cancer. *Gynecologic Oncology* **131**, 52–58 (2013).
11. Singer, E. A. *et al*. Active surveillance for prostate cancer: past, present and future. *Current Opinion in Oncology* **24**, 243–250 (2012).
12. Poruk, K. E. *et al*. The Clinical Utility of CA 19-9 in Pancreatic Adenocarcinoma: Diagnostic and Prognostic Updates. *Current Molecular Medicine* **13**, 340–351 (2013).
13. Frantzi, M. *et al*. Discovery and validation of urinary biomarkers for detection of renal cell carcinoma. *J Proteomics* **98**, 44–58 (2014).
14. Frantzi, M. *et al*. Development and Validation of Urine-based Peptide Biomarker Panels for Detecting Bladder Cancer in a Multi-center Study. *Clin. Cancer Res* **22**, 4077–4086 (2016).
15. Schonemeier, B. *et al*. Urinary Peptide Analysis Differentiates Pancreatic Cancer From Chronic Pancreatitis. *Pancreas* **45**, 1018–1026 (2016).
16. Theodorescu, D. *et al*. Discovery and validation of urinary biomarkers for prostate cancer. *Proteomics Clin. Appl* **2**, 556–570 (2008).
17. Good, D. M. *et al*. Naturally Occurring Human Urinary Peptides for Use in Diagnosis of Chronic Kidney Disease. *Molecular & Cellular Proteomics* **9**, 2424–2437 (2010).
18. Klein, J., Papadopoulos, T., Mischak, H. & Mullen, W. Comparison of CE-MS/MS and LC-MS/MS sequencing demonstrates significant complementarity in natural peptide identification in human urine. *Electrophoresis* **35**, 1060–1064 (2014).
19. Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nature Reviews Molecular Cell Biology* **15**, 786–801 (2014).
20. Berger MD, H. A. *et al*. Increased fibrinogen levels at diagnosis are associated with adverse outcome in patients with acute myeloid leukemia. *Hematol Oncol* (2016).
21. Bergt, C. *et al*. Human neutrophils employ the myeloperoxidase/hydrogen peroxide/chloride system to oxidatively damage apolipoprotein A-I. *European Journal of Biochemistry* **268**, 3523–3531 (2001).
22. Wheatley-Price, P. *et al*. Myeloperoxidase and superoxide dismutase polymorphisms are associated with an increased risk of developing pancreatic adenocarcinoma. *Cancer* **112**, 1037–1042 (2008).
23. Hung, R. J. *et al*. Genetic polymorphisms of MPO, COMT, MnSOD, NQO1, interactions with environmental exposures and bladder cancer risk. *Carcinogenesis* **25**, 973–978 (2004).
24. Zhu, H. *et al*. Myeloperoxidase G-463A polymorphism and the risk of gastric cancer: a case-control study. *Carcinogenesis* **27**, 2491–2496 (2006).
25. Felder, M. *et al*. MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. *Molecular Cancer* **13** (2014).
26. Mantovani, A., Allavena, P., Sica, A. & Balkwill, F. Cancer-related inflammation. *Nature* **454**, 436–444 (2008).
27. Siwy, J., Mullen, W., Golovko, I., Franke, J. & Zurbig, P. Human urinary peptide database for multiple disease biomarker discovery. *Proteomics Clin. Appl* **5**, 367–374 (2011).
28. Mischak, H. *et al*. Recommendations for Biomarker Identification and Qualification in Clinical Proteomics. *Science Translational Medicine* **2** (2010).
29. McShane, L. M. *et al*. REporting recommendations for tumor MARKer prognostic studies (REMARK). *Nature Clinical Practice Urology* **2**, A416–A422 (2005).
30. Weissinger, E. M. *et al*. Proteomic peptide profiling for preemptive diagnosis of acute graft-versus-host disease after allogeneic stem cell transplantation. *Leukemia* **28**, 842–852 (2014).
31. Ru Wei, B. G. *et al*. Alterations in urinary collagen peptides in lupus nephritis subjects correlate with renal dysfunction and renal histopathology. *Nephrol Dial Transplant* 1–10 (2016).
32. Zurbig, P. *et al*. Biomarker discovery by CE-MS enables sequence analysis via MS/MS with platform-independent separation. *Electrophoresis* **27**, 2111–2125 (2006).
33. Dakna, M. *et al*. Addressing the challenge of defining valid proteomic biomarkers and classifiers. *BMC Bioinformatics* **11**, 594 (2010).
34. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289–300 (1995).
35. UniProt, C. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **45** (2017).

Acknowledgements

This work was supported by the Chromatin3D grant n°622934 (Chromatin3D-MSCA-ITN) project funded by the European Commission.

Author Contributions

I.B., M.F. and A.L. wrote the main manuscript text. I.B., M.F., A.L., J.S. and J.M. analyzed the peptidomics data and performed the statistical analysis. V.J., A.V., J.S., J.M., A.S.M. and H.M. contributed to the writing of the manuscript. I.B. prepared all figures, tables and supplementary materials. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-23585-y.

Competing Interests: HM is the founder and co-owner of Mosaiques Diagnostics. IB, MF, AL, JS and JM are employed by Mosaiques Diagnostics GmbH.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018

SCIENTIFIC REPORTS
nature research

OPEN

Author Correction: Urinary CE-MS peptide marker pattern for detection of solid tumors

Published online: 03 September 2019

Iwona Belczacka[1,2], Agnieszka Latosinska[1], Justyna Siwy[1], Jochen Metzger[1], Axel S. Merseburger[3], Harald Mischak[1,4], Antonia Vlahou[5], Maria Frantzi[1] & Vera Jankowski[2]

Correction to: *Scientific Reports* https://doi.org/10.1038/s41598-018-23585-y, published online 27 March 2018

The Author Contributions section in this Article is incorrect.

"I.B., M.F. and A.L. wrote the main manuscript text. I.B., M.F., A.L., J.S. and J.M. analyzed the peptidomics data and performed the statistical analysis. V.J., A.V., J.S., J.M., A.S.M. and H.M. contributed to the writing of the manuscript. I.B. prepared all figures, tables and supplementary materials. All authors reviewed the manuscript."

should read:

"I.B. wrote the manuscript, performed data collection and evaluation, performed experiments, interpretation of peptidomics data and statistical evaluation. I.B., A.L., J.S., H.M., M.F. and V.J. participated in study supervision. I.B., H.M., A.V., M.F. and V.J. served as scientific advisors in consulting the study design. I.B., J.S. and J.M. contributed to examination of study participants. I.B., A.L., J.S., J.M. and M.F. contributed to data evaluation, interpretation of peptidomics data and statistical evaluation. I.B., J.M., A.S.M., H.M. and M.F. participated in delivery of the materials. I.B., A.L., J.S. and M.F. participated in performing experiments (sequencing of peptides). I.B., A.L. and M.F. contributed to performing experiments (specificity analysis and validation in additional set of cancer patients). I.B. prepared all figures, tables and supplementary materials. All authors corrected the manuscript and gave the final approval of the version to be published."

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

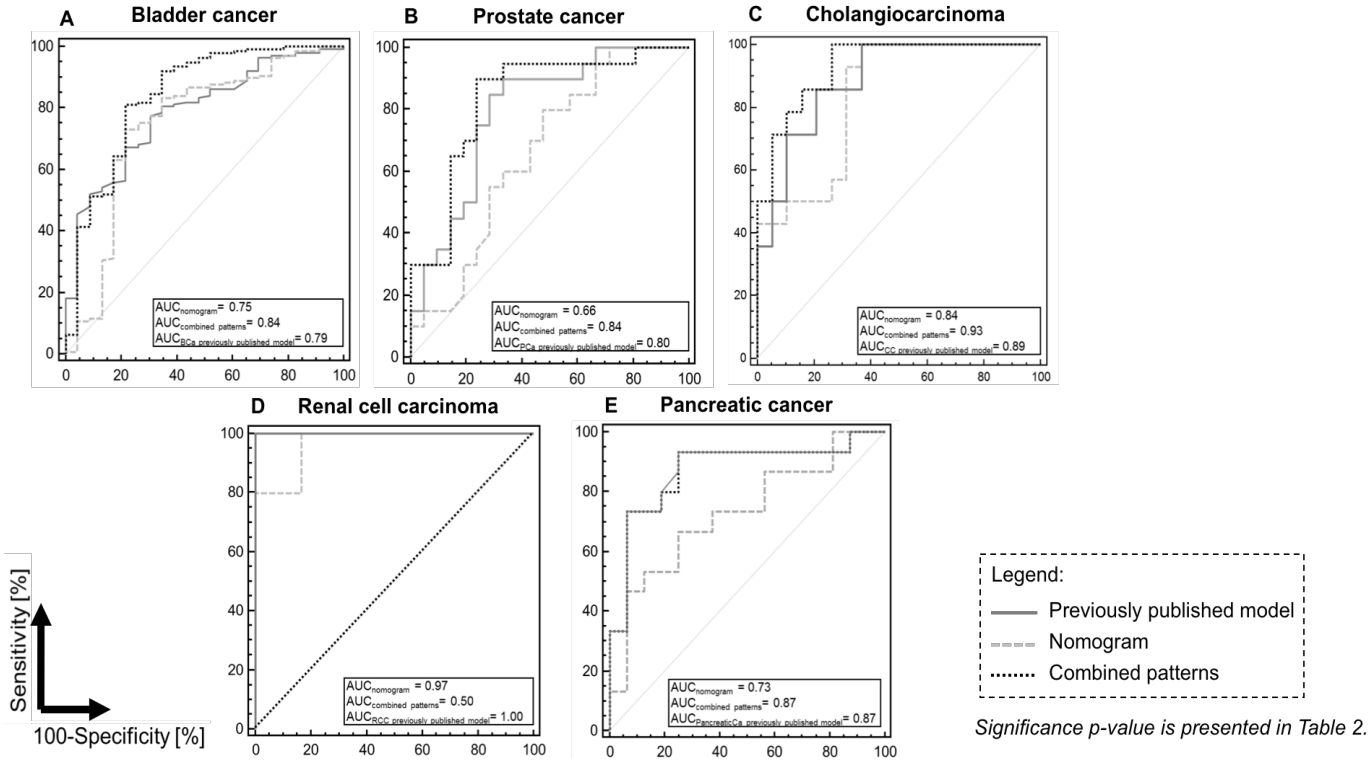
© The Author(s) 2019

[1]Mosaiques Diagnostics GmbH, Hannover, Germany. [2]University Hospital RWTH Aachen, Institute for Molecular Cardiovascular Research (IMCAR), Aachen, Germany. [3]Department of Urology, University of Lübeck, Lübeck, Germany. [4]University of Glasgow, Institute of Cardiovascular and Medical Sciences, Glasgow, United Kingdom. [5]Biotechnology Division, Biomedical Research Foundation, Academy of Athens (BRFAA), Athens, Greece. Correspondence and requests for materials should be addressed to V.J. (email: vjankowski@ukaachen.de)

Supplementary data

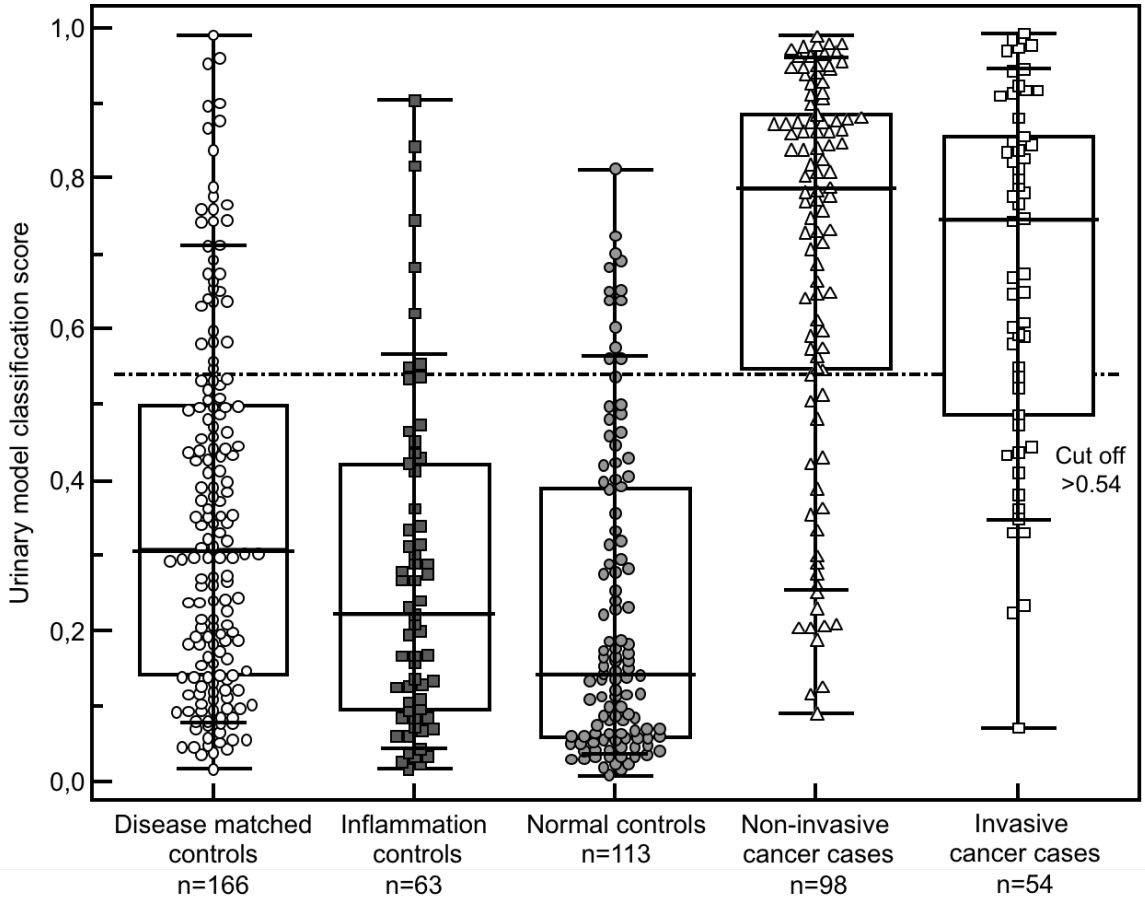
Supplementary tables can be found online and on the CD.

Supplementary Figure S1. Comparison of the classification scores with the previously published peptide marker patterns for the detection of **A.** bladder cancer, **B.** prostate cancer, **C.** cholangiocarcinoma, **D.** renal cell carcinoma, **E.** pancreatic cancer. ***AUC**, Area under the curve.*



Significance p-value is presented in Table 2.

Supplementary Figure S2. Representation of the nomogram classification scores between non-invasive and invasive cancer cases. The optimal cut-off (>0.54) was estimated based on the nomogram scores according to the Youden index.



Proteomics
Clinical Applications

www.clinical.proteomics-journal.com

Urinary Glycopeptide Analysis for the Investigation of Novel Biomarkers

*Iwona Belczacka, Martin Pejchinovski, Magdalena Krochmal, Pedro Magalhães, Maria Frantzi, William Mullen, Antonia Vlahou, Harald Mischak, and Vera Jankowski**

Purpose: Urine is a rich source of potential biomarkers, including glycoproteins. Glycoproteomic analysis remains difficult due to the high heterogeneity of glycans. Nevertheless, recent advances in glycoproteomics software solutions facilitate glycopeptide identification and characterization. The aim is to investigate intact glycopeptides in the urinary peptide profiles of normal subjects using a novel PTM-centric software—Byonic.

Experimental design: The urinary peptide profiles of 238 normal subjects, previously analyzed using CE–MS and CE–MS/MS and/or LC–MS/MS, are subjected to glycopeptide analysis. Additionally, glycopeptide distribution is assessed in a set of 969 patients with five different cancer types: bladder, prostate and pancreatic cancer, cholangiocarcinoma, and renal cell carcinoma.

Results: A total of 37 intact O-glycopeptides and 23 intact N-glycopeptides are identified in the urinary profiles of 238 normal subjects. Among the most commonly identified O-glycoproteins are Apolipoprotein C-III and insulin-like growth factor II, while titin among the N-glycoproteins. Further statistical analysis reveals that three O-glycopeptides and five N-glycopeptides differed significantly in their abundance among the different cancer types, comparing to normal subjects.

Conclusions and clinical relevance: Through the established glycoproteomics workflow, intact O- and N-glycopeptides in human urine are identified and characterized, providing novel insights for further exploration of the glycoproteome with respect to specific diseases.

1. Introduction

Mass spectrometry (MS) in combination with the latest proteomic software solutions, allows in-depth analysis of proteins and peptides in a range of biological specimens.[1] Indeed, MS/MS analysis generates thousands of fragmentation spectra, requiring well-developed computational tools, to assess them.[2] Matching of MS/MS spectra to the correct peptide candidate using search engines is a key step of the analysis.[3] The process becomes more challenging when intact endogenous peptides are examined, as compared to tryptic peptides, given that optimal search parameters are not restricted by the enzyme used.[4] In addition, the presence of posttranslational modifications (PTMs) is another major obstacle for peptide characterization and identification.[5]

Assessment of PTMs is critical when studying protein function/activity, folding, and molecular interactions in relation to disease.[6,7] Glycosylation, as the most frequent PTM, is predicted to affect approximately 50% of the human proteome.[8] This enzymatic process, catalyzed via glycosyltransferases and glycosidases,[8] produces different types of saccharides, starting from simple monosaccharides to complex polysaccharides linked via amide or glycosidic bonds.[9]

The presence of glycans determines, among others, correct protein folding within the endoplasmic reticulum and trafficking of newly synthesized proteins.[10] Additionally, glycans protect proteins from proteolytic degradation,[11] and mediate cell–cell and cell–matrix interactions, hence demonstrate a significant impact on multiple biological processes including pathological conditions, such as cancer invasion,[12] diabetes mellitus, cardiovascular disorders, rheumatoid arthritis, and neurological diseases.[13]

In general, there are three distinct types of glycosylation: a) N-glycans (attached to asparagine side chain in an NXT or NXS amino acid consensus sequence with X ≠ proline); b) O-glycans (attached to serine or threonine side chains); and c) glycosaminoglycans (bound to a serine side chain).[14] N-linked and O-linked glycosylation is considered the most common and structurally

I. Belczacka, Dr. M. Pejchinovski, Dr. M. Krochmal, Dr. P. Magalhães
Dr. M. Frantzi, Prof. Dr. H. Mischak
Mosaiques Diagnostics GmbH
30659 Hannover, Germany

I. Belczacka, Dr. V. Jankowski
University Hospital RWTH Aachen
Institute for Molecular Cardiovascular Research (IMCAR)
52074 Aachen, Germany
E-mail: vjankowski@ukaachen.de

Dr. W. Mullen
Institute of Cardiovascular and Medical Sciences
University of Glasgow
G128QQ Glasgow, UK

Dr. A. Vlahou
Biotechnology Division
Biomedical Research Foundation Academy of Athens (BRFAA)
11527 Athens, Greece

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/prca.201800111

DOI: 10.1002/prca.201800111

 © 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

diverse PTMs found on secreted proteins and on the extracellular parts of membrane-bound proteins.[15] Based on their specific sequence consensus NXS/T/C motif, N-glycans can be released using a universal enzyme, that is, PNGase F (peptide N-glycosidase F).[16] Characterization of O-glycosylated peptides is more challenging, owing to the lack of a common amino acid consensus motif.[17] Multiple modification sites are possible and there is no universal enzyme to remove the O-glycans. O-glycans can be released chemically, either in solution or from samples immobilized, followed by the reductive β-elimination.[17]

Glycosylation has been repeatedly implicated in many cancer types in multiple studies.[18,19] Understanding the O- and N-linked glycan structure and attachment may be highly relevant in oncology.[20] Characterization of the site-specific N- or O- linked glycosylation events and identification of the modified proteins are being pursued by the proteomic community.[21,22] Urine peptides have been extensively studied and connected to pathophysiological processes in many diseases.[13,21] They were demonstrated to enable monitoring of disease activity and drug response.[23,24] Nevertheless, a significant part of the urine peptidome remains unknown (unsequenced), partially attributed to the presence of PTMs.[25,26] Identification of such disease-related modified structures could be highly informative on disease pathophysiology.

Although, there is progress in the identification of PTMs due to the technology development, it still remains a challenging task, relying heavily on the employed analytical tools for data evaluation.[27–29] Traditionally, proteomic data analysis relies on database searches using standard parameters that do not consider the potential presence of glycopeptide fragment ions, which are thereby not typically assessed. This is reflected by the scarce number of reports on glycoproteomic characterization, also in urine.[30]

Byonic, a PTM-centric search engine targeting the identification of intact N- and O-linked glycopeptides from MS/MS spectral input, has been successfully applied in glycoproteomics.[31,32] Byonic identifies glycopeptides by searching separate predefined or user-defined protein and glycan databases. Consequently, peptide sequence as well as structural information on the glycan moieties and attachment site are assigned.[32]

In the present study, we used Byonic software to identify intact O- and N-glycopeptides in human urine analyzed by CE–MS. CE and LC coupled to MS/MS were used both to identify glycopeptide sequences. Furthermore, we estimated the abundance levels of these glycopeptides in urine from five different cancer types, that is, bladder cancer (BCa), prostate cancer (PCa), pancreatic cancer, cholangiocarcinoma (CC), and renal cell carcinoma (RCC).

Clinical Relevance

Glycosylation is the most abundant and diverse posttranslational modification of proteins, affecting approximately half of human proteome. In recent years, it has been demonstrated that glycoproteins are related to many diseases. The emerging technologies in glycoproteomics provide new insights into the study of glycoproteins, focusing on the characterization of glycan structures and their attachment sites. In the present study, we established a glycoproteomics workflow for the identification of intact glycopeptides. The methodology and the results presented in this study appear to be of value in the exploration of the urinary glycoproteome and the search of specific glycosylation signatures in health and disease.

2. Experimental Section

2.1. Sample Selection

MS datasets representing urinary peptide profiles of 238 normal subjects were retrieved from the Human Urinary Protein Database[33] and were used as a source for glycopeptides search. Furthermore, glycopeptide distribution was assessed in previously published datasets representing urinary profiles of patients with BCa ($n = 487$), PCa ($n = 206$), CC ($n = 118$), pancreatic cancer ($n = 88$), and RCC ($n = 70$).[34] In addition, we included a second group of normal subjects at more advanced age ($n = 226$, mean of age 60) to ascertain that the differences in glycopeptide distributions are not related to age. Demographic characteristics of the subjects are presented in **Table 1** and Table S1, Supporting Information. The study was conducted fulfilling all the laws on the protection of individuals being involved in medical research and in accordance with the principles of the Declaration of Helsinki. Ethical approval for the study was obtained by the Ethics Committee from Hannover Medical School in Germany (ID: 3409-2016).

2.2. Sample Preparation

All CE-MS datasets were acquired following the same analytical workflow, employing relevant guidelines and standard operating procedures (SOPs)[35] for urine collection, as defined by EuroKUP and described in respective publications.[36] The samples were spontaneously voided urine, collected according to the established SOPs as described previously.[35,36] Urine aliquots of 700 μL were thawed and mixed with 700 μL of 2 M urea, 10 mM NH_4OH containing 0.02% SDS. Subsequently, samples were ultrafiltered using a Centristat 20 kDa cutoff centrifugal filter device (Satorius, Göttingen, Germany) to eliminate high molecular weight proteins. The obtained filtrate was desalted using a PD-10 column (GE Healthcare Bio Sciences, Uppsala, Sweden) to remove urea, electrolytes, and salts. The samples were lyophilized and stored at 4 °C before usage. Shortly before CE-MS analysis, the samples were resuspended in HPLC-grade H_2O.

2.3. CE-MS Analysis and Data Processing

CE-MS analysis was performed as previously described.[37] A P/ACE MDQ capillary electrophoresis system (Beckman Coulter, Fullerton, CA) was coupled with a Micro-TOF MS (Bruker Daltonic, Bremen, Germany). A solution of 20% acetonitrile (Sigma–Aldrich, Taufkirchen, Germany) in HPLC-grade water (Roth, Karlsruhe, Germany) supplemented with 0.94% formic

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Characteristics of the patients used in the study.

Group	Number	Mean age (years; range)	Gender	
			Female/male	No gender data available
Normal subjects				
Normal subjects (*used for glycopeptide identification*)	$N = 238$	30 (14–75)	115/106	17
Older normal subjects (*used for additional analysis*)	$N = 226$	60 (38–86)	95/131	–
Cancer datasets used to assess the glycopeptide distribution				
Bladder cancer	$N = 487$	68 (27–95)	93/393	1
Prostate cancer	$N = 206$	68 (41–87)	–/206	–
Pancreatic cancer	$N = 88$	68 (33–95)	34/44	10
Renal cell carcinoma	$N = 70$	61 (33–81)	25/45	–
Cholangiocarcinoma	$N = 118$	61 (22–86)	43/75	1

acid (Sigma–Aldrich) was used as running buffer. For CE-MS analysis, the electrospray ionization interface from Agilent Technologies (Palo Alto, CA) was set to a potential of –4.0 to –4.5 kV. Spectra were recorded over m/z range of 350–3000 and accumulated for 3 s each. Monoisotopic mass signals were determined for $z \leq 6$ with resolution in the range of 10 000 and a limit of detection of $\approx$ 1 fmol, as reported in.[35]

Ion peaks representing identical molecules at different charge states were deconvoluted into single masses using the MosaVisu software.[38] The result is a peak list, characterizing each protein and peptide by its molecular mass (Da) and normalized CE migration time (min), as described in detail in Kaiser et al. [39]. Migration time and signal intensity (amplitude) of the urinary peptides were normalized by using reference signals from 29 highly abundant peptides as "internal standard," using locally linear regression.[40] These 29 internal standards are collagen peptides that have been and are routinely used for normalization in all CE–MS-based urine proteomics studies. These peptides are not significantly associated with any diseases investigated to date.[40] All detected peptides were deposited, matched, and annotated in a Microsoft SQL database allowing further statistical analysis.[40,41] The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD011214.

2.4. MS/MS Analysis

MS/MS analysis (coupled either to CE or LC to maximize coverage) was used for peptide sequencing. CE–MS/MS analysis was performed as described before.[37] Briefly, the CE was connected to an Orbitrap Velos FTMS (Thermo Finnigan, Bremen, Germany). The sample was ionized using a modified Proxeon nano spray source fitted with an Agilent ESI sprayer operating in positive ion mode. Ionization voltage was 3.4 kV and the capillary temperature was 275 °C. The mass spectrometer was operated in MS/MS mode scanning from 350 to 1500 amu. The threshold was set to 5000 counts to trigger sequencing. The top five multiply charged ions were selected from each scan for MS/MS analysis using higher-energy collision-induced dissociation (HCD) at 40% collision energy.[40] The resolution of ions in MS1 was 60 000 and 7500 for MS2.

For LC–MS/MS analysis, aliquots of 5 μL sample were separated on a nano-flow system (Dionex Ultimate 3000 RSLC, Dionex, Camberly, UK) as previously described.[37] Samples were loaded onto a Dionex nano trap column (C18, 0.1 × 20 mm, 5μm) at a flow rate of 5 μL min^{-1} in 98% H_2O, 0.1% formic acid, and 2% acetonitrile. The trap and nano-flow column were maintained at 35 °C. Separation was performed on an Acclaim PepMap nano column (C18, 75 μm × 50 cm, 2 μm 100 Å) at a flowrate of 0.3 μL min^{-1}. Samples were eluted with a gradient of solvent A (97.9% H_2O, 0.1% formic acid, and 2% acetonitrile) versus solvent B (80% acetonitrile, 19.9% H_2O, and 0.1% formic acid). Separation was started using 1% solvent B (5 min) followed by gradual increase to 20% (after 90 min) and finally to 40% (after 120 min). The eluant was ionized using a Proxeon nano spray ESI source operating in a positive ion mode connected to an Orbitrap Velos FTMS. Ionization voltage was 2.6 kV and the capillary temperature was 250 °C. The mass spectrometer was operated in HCD MS/MS mode. Using HCD, the top 20 multiply charged ions were selected from each scan for MS/MS analysis, rejecting singly charged ions.[37] In addition, the Orbitrap Velos used to identify the glycopeptides allows automatic fragmentation of a glycopeptide by both HCD and then electron-transfer dissociation (ETD) to provide two fragmentation spectra that can be combined in Byonic for a higher certainty of the correct glycopeptide identification, as described by Yin et al.[42] The detection limit for the CE– or LC–MS/MS analysis using the Orbitrap Velos mass spectrometer, with 60 000 resolution for MS1 and with 7500 resolution for MS2, was in the range of 0.05–0.2 fmol.

2.5. Glycopeptide Identification Using the Byonic Software

MS/MS data were further processed in Byonic software[31] (Protein Metrics INC, USA). The data were searched against the entire human proteome (FASTA file, 20.303 reviewed entries in UniProtKB, released January 2018), against the mammalian common O-glycome database (six most common O-glycans–default database in Byonic) and the common N-glycome database (309 common mammalian N-glycans–default database in

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

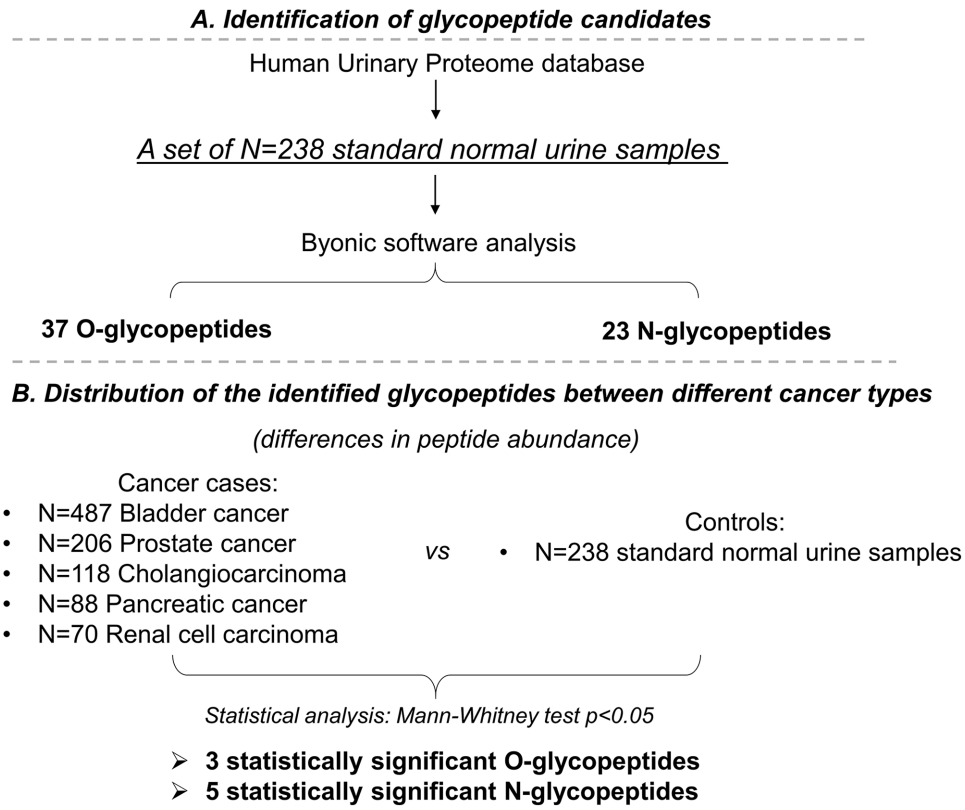


Figure 1. Schematic representation of the study design for the identification of the urinary glycopeptides.

Byonic). The mass tolerance was up to 10 ppm for the precursor ions and 0.05 Da for the fragment ions. Oxidation of methionine and proline as variable modification was selected and O- or N-linked glycan residues were used as variable modifications. The mass range for the search was set between 350 and 5000 Da. Glycopeptide sequences reported by Byonic were only accepted if they migrated at the predicted position, as described in detail by Zurbig et al.[43]

2.6. Statistical Analysis

Statistical comparison between urinary profiles of different groups was performed using the Mann–Whitney *U*-test. The reported *p*-values were adjusted by multiple testing with the method described by Benjamini and Hochberg[44] and $p < 0.05$ was considered as statistically significant. Box and Whisker plots were generated using the R statistical software version 3.3.3 (packages ggplot2, plotly).

3. Results

3.1. Identification of O-Glycopeptides

The O-glycopeptide identification and glycosylation site determination followed the study design outlined in **Figure 1**. A total of 37 glycopeptides were detected with different frequencies in the 238 urinary peptide profiles of normal subjects. These 37 glycopeptides corresponded to 17 glycoproteins. The glycopeptide mass, sequence information, glycan composition, and glycosylation sites are presented in **Table 2**. The most frequently identified glycopeptides were derived from Apolipoprotein C-III (APOC3; $n = 10$) and insulin-like growth factor II (IGF2; $n = 7$). Five glycoproteins: vitamin K-dependent protein C (PROC), alpha-2-HS-glycoprotein (FETUA), fractalkine (CX3CL1), peptidase inhibitor 16 (PI16), and ProSAAS (PCSK1) were represented by two peptides. An example of the annotated HCD/ETD-MS/MS spectra of two O-glycopeptides, as derived from APOC3 and IGF2, are presented in **Figure 2**. Interestingly, a limited number of glycopeptides, that is, *WDLDPEVRPTSA* and *DLDPEVRPTSA* (derived from APOC3) and *DVSTPPTVLPDNFPRYPVGKF* (derived from IGF2), were identified as different glycoforms exhibiting identical peptide sequence, however bearing different carbohydrate structures ($HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_2$; presented in Table 2).

3.2. Identification of N-Glycopeptides

The N-glycopeptide identification workflow is also schematically depicted in Figure 1. In total, 23 N-glycopeptides were detected with different frequencies in the 238 urinary peptide profiles of normal subjects. These 23 glycopeptides corresponded to 20 glycoproteins. The N-glycopeptide mass, sequence information, glycan composition, and glycosylation sites are presented in

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. List of the 37 O-glycopeptides and 23 N-glycopeptides identified using Byonic software. Detailed information of the glycan composition and amino acid position is specified. Protein names marked with asterisk (*) were not unique candidates for the sequence.

Mass (Dalton)	CE-time (min)	Amino acid sequence	Protein symbol	Protein name	Glycan position/ mass	Glycan composition	Start Amino Acid	Stop Amino Acid	Consistent with Halim et al.	Reported attachment site[a]
O-glycopeptides										
1451.63	36.01	EAGSPPVVE	ANKRD17	Ankyrin repeat domain-containing protein 17	S4 (568.21)	$HexNAc_2Hex_1$	2078	2086	–	–
1669.65	46.08	FSNGADL	A1AT	Alpha-1-antitrypsin	S2 (947.32)	$HexNAc_1Hex_1NeuAc_2$	336	342	√	–
1574.71	40.59	VVQPSV	FETUA	Alpha-2-HS-glycoprotein	S5 (947.32)	$HexNAc_1Hex_1NeuAc_2$	342	347	√	S346
1866.84	39.96	TQPVTSQPQPE	FETUA	Alpha-2-HS-glycoprotein	T5 (656.23)	$HexNAc_1Hex_1NeuAc_1$	252	262		T256
1783.79	31.90	DLDPEVRPTS	APOC3	Apolipoprotein C-III	T9 (656.23)	$HexNAc_1Hex_1NeuAc_1$	86	95	√	T94
1854.84	27.04	DLDPEVRPTSA	APOC3	Apolipoprotein C-III	T9 (656.23)	$HexNAc_1Hex_1NeuAc_1$	86	96		
2145.89	32.55	DLDPEVRPTSA	APOC3	Apolipoprotein C-III	T9 (947.32)	$HexNAc_1Hex_1NeuAc_2$	86	96		
1626.70	29.75	DPEVRPTSA	APOC3	Apolipoprotein C-III	T7 (656.23)	$HexNAc_1Hex_1NeuAc_1$	88	96		
1705.74	36.77	EVRPTSA	APOC3	Apolipoprotein C-III	T5 (947.32)	$HexNAc_1Hex_1NeuAc_2$	90	96		
2187.97	32.82	FWDLDPEVRPTSA	APOC3	Apolipoprotein C-III	T11 (656.23)	$HexNAc_1Hex_1NeuAc_1$	84	96		
1802.79	30.90	PEVRPTSA	APOC3	Apolipoprotein C-III	T6 (947.32)	$HexNAc_1Hex_1NeuAc_2$	89	96		
1576.68	29.47	VRPTSA	APOC3	Apolipoprotein C-III	S5 (947.32)	$HexNAc_1Hex_1NeuAc_2$	91	96		
2040.90	32.97	WDLDPEVRPTSA	APOC3	Apolipoprotein C-III	T10 (656.23)	$HexNAc_1Hex_1NeuAc_1$	85	96		
2331.90	33.42	WDLDPEVRPTSA	APOC3	Apolipoprotein C-III	T10 (947.32)	$HexNAc_1Hex_1NeuAc_2$	85	96		
2046.89	39.84	PGQTGSPGPTGEPGPSG	COL5A1	Collagen alpha-1(V) chain	S6 (568.21)	$HexNAc_2Hex_1$	1371	1387	–	–
3314.48	20.21	DEAGSEADHEGTHSTKRGHAKSRPV	FGA	Fibrinogen alpha chain	T12 (656.22)	$HexNAc_1Hex_1NeuAc_1$	305	329	–	–
1705.70	45.88	TTTAPTPA	HUWE1	E3 ubiquitin-protein ligase	T2 (947.32)	$HexNAc_1Hex_1NeuAc_2$	1090	1097	–	–
1515.66	39.16	SVVPVSSEG	CX3CL1	Fractalkine	S7 (656.23)	$HexNAc_1Hex_1NeuAc_1$	287	295	√	–
1728.76	36.62	ESVVLEPEAT	CX3CL1	Fractalkine	T10 (656.23)	$HexNAc_1Hex_1NeuAc_1$	123	132		
2109.94	25.49	KPEYLEELPTMMQH	GSTM4	Glutathione S-transferase Mu 4	T10 (365.13)	$HexNAc_1Hex_1$	124	137	–	–
2380.07	35.57	EDLATQLNLKTST	CUX1	Homeobox protein cut-like 1	T5 (947.32)	$HexNAc_1Hex_1NeuAc_2$	1275	1287	–	–
O-glycopeptides										
2861.25	34.84	DVSTPPTVLPDNFPRYP	IGF2	Insulin-like growth factor II	T7 (947.32)	$HexNAc_1Hex_1NeuAc_2$	93	109	√	T99
3017.41	34.62	DVSTPPTVLPDNFPRYPVG	IGF2	Insulin-like growth factor II	T7 (947.32)	$HexNAc_1Hex_1NeuAc_2$	93	111		T99
2854.31	28.77	DVSTPPTVLPDNFPRYPVGK	IGF2	Insulin-like growth factor II	T7 (656.23)	$HexNAc_1Hex_1NeuAc_1$	93	112		T99
3001.30	26.91	DVSTPPTVLPDNFPRYPVGKF	IGF2	Insulin-like growth factor II	T7 (656.23)	$HexNAc_1Hex_1NeuAc_1$	93	113		T99
3292.51	29.55	DVSTPPTVLPDNFPRYPVGKF	IGF2	Insulin-like growth factor II	T7 (947.32)	$HexNAc_1Hex_1NeuAc_2$	93	113		T99
2758.24	41.85	LIALPTQDPAHGGAPPEMASN	IGF2	Insulin-like growth factor II	T6 (656.23)	$HexNAc_1Hex_1NeuAc_1$	158	178		T163
2269.00	34.06	TPPTVLPDNFPRYP	IGF2	Insulin-like growth factor II	T4 (656.23)	$HexNAc_1Hex_1NeuAc_1$	96	109		T99
2039.92	35.24	TTPPNKELPP	MAP4	Microtubule-associated protein 4	T2 (947.32)	$HexNAc_1Hex_1NeuAc_2$	686	695	–	–

(*Continued*)

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. Continued.

Mass (Dalton)	CE-time (min)	Amino acid sequence	Protein symbol	Protein name	Glycan position/ mass	Glycan composition	Start Amino Acid	Stop Amino Acid	Consistent with Halim et al.	Reported attachment site[a]
1227.56	22.39	KTKVP	PI16	Peptidase inhibitor 16*	T2 (656.23)	$HexNAc_1Hex_1NeuAc_1$	325	330	√	–
1605.73	24.02	KTKVPS	PI16	Peptidase inhibitor 16*	T2 (947.32)	$HexNAc_1Hex_1NeuAc_2$	325	331		
1345.55	28.90	SRLLDS	IC1	Plasma protease C1 inhibitor*	S1 (656.23)	$HexNAc_1Hex_1NeuAc_1$	275	280	√	–
1950.88	23.64	AETGAPRRF	PCSK1	ProSAAS	T3 (947.32)	$HexNAc_1Hex_1NeuAc_2$	51	59	√	T53
1989.95	25.63	LETPAPQVPARR	PCSK1	ProSAAS	T3 (656.23)	$HexNAc_1Hex_1NeuAc_1$	245	256		T247
1772.80	26.21	KQLSLTH	HEG1	Protein HEG homolog 1*	S4 (947.32)	$HexNAc_1Hex_1NeuAc_2$	879	885	√	–
2647.21	27.68	TPAPLDSVFSSSERAH	PROC	Vitamin K-dependent protein C	S10 (947.32)	$HexNAc_1Hex_1NeuAc_2$	19	34	√	–
3143.40	24.44	TPAPLDSVFSSSERAHQVLR	PROC	Vitamin K-dependent protein C	S12 (947.32)	$HexNAc_1Hex_1NeuAc_2$	19	38		
N-glycopeptides										
1422.58	21.78	ANTSLMG	ANKRD17	Ankyrin repeat domain-containing protein 17	N2 (730)	$HexNAc_2Hex_2$	1828	1834	–	–
1452.64	24.84	RYFFNVSD	ITGA2	Integrin alpha-2	N5 (406)	$HexNAc_2$	339	346	–	N343
1623.68	22.59	GFRNSSYPE	CDK17	Cyclin-dependent kinase 17	N4 (568)	$HexNAc_2Hex_1$	502	510	–	–
1688.77	23.84	TAPPNFSHTGPTFP	WNK1	Serine/threonine-protein kinase WNK1	N5 (203), P[16]	$HexNAc_1$	1321	1334	–	–
1848.81	37.29	PPPRDNSSLH	MAST4	Microtubule-associated serine/ threonine-protein kinase 4	N6 (730)	$HexNAc_2Hex_2$	2053	2062	–	–
1960.89	32.41	PGPPLNVTI	TTN	Titin	N6 (1054)	$HexNAc_2Hex_4$	14 118	14 126	–	–
2314.02	34.75	VPDAPGIPEPSNIT	TTN	Titin	N13 (892), P[16]	$HexNAc_2Hex_3$	29 078	29 091		–
2479.06	47.73	NKTPIR*	TTN	Titin	N1 (1752)	$HexNAc_4Hex_4NeuAc_1$	21 153	21 158		–
2391.92	33.92	QNGADPNI	ANKRD36B	Ankyrin repeat domain-containing protein 36B	N7 (1565)	$HexNAc_3Hex_4NeuGc_1$	107	114	–	–
2443.06	34.94	PGDLINPLNL	ITGA3	Integrin alpha-3	N6 (1378)	$HexNAc_2Hex_6$	849	858	–	–
2567.11	34.83	ATPLYINI	FAT1	Protocadherin Fat 1	N7 (1664)	$HexNAc_5Hex_4$	654	661	–	N660
3034.34	35.21	NITVTSLVHL	KCNA4	Potassium voltage-gated channel subfamily H member 4	N1 (1939)	$HexNAc_5Hex_3Fuc_1NeuAc_1$	326	335	–	–
3046.32	20.80	NITPLLL*	H2AFB2	Histone H2A-Bbd type 1	N1 (2264)	$HexNAc_5Hex_5Fuc_3$	82	88	–	–
3182.47	36.01	NITILVVVLV	ADAM15	Disintegrin and metalloproteinase domain-containing protein 7	N1 (2101)	$HexNAc_5Hex_4Fuc_1NeuAc_1$	668	677	–	–
3201.54	36.01	KYIDQEELNKTKPIWTRNPD	HSP90AA2P	Heat shock protein HSP 90-alpha A2	N9 (714)	$HexNAc_2Hex_1Fuc_1$	283	302	–	–
3223.41	38.84	NKTLIYIL	UBAC2	Ubiquitin-associated domain-containing protein 2	N1 (2247)	$HexNAc_5Hex_4Fuc_2NeuAc_1$	161	168	–	–
3280.59	31.98	TSPPGVNFTARSKPFAVLPVT	PKHD1	Fibrocystin	N7 (1095)	$HexNAc_3Hex_3$	3832	3852	–	N3838

(*Continued*)

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. Continued.

Mass (Dalton)	CE-time (min)	Amino acid sequence	Protein symbol	Protein name	Glycan position/ mass	Glycan composition	Start Amino Acid	Stop Amino Acid	Consistent with Halim et al.	Reported attachment site[a]
N-glycopeptides										
3437.56	31.94	PGPLLVPFTLNFTITN	MUC16	Mucin-16	N9 (1695)	$HexNAc_3Hex_4Fuc_1NeuAc_1$	12 384	12 398	–	–
3553.64	31.90	DGIPQGVLERNNVITLNI	GLB1	Beta-galactosidase	N11 (1590)	$HexNAc_4Hex_3NeuAc_1$	448	465	–	–
3583.66	26.32	TAVTVNQSLLSPLVLEVDP	KRT8	Keratin, type II cytoskeletal 8	N6 (1590)	$HexNAc_4Hex_3NeuAc_1$	64	82	–	–
4020.74	33.65	STPARATVTINVTDVNDNPPNID	PCDH9	Protocadherin-9	N11 (1582), P[16]	$HexNAc_3Hex_6$	337	359	–	N347
4075.65	40.08	NLTVVVR*	PCDHAC2	Protocadherin alpha-C2	N1 (3276)	$HexNAc_6Hex_{10}Fuc_1NeuAc_1$	657	663	–	N657
4916.10	24.11	NLSLVILVPQNL	SERPING1	Plasma protease C1 inhibitor	N1 (3594)	$HexNAc_3Hex_{10}Fuc_1$	352	363	√	N352

[a]Glycan attachment site reported in UniProtKB/Swiss–Prot database.

Table 2. Among the identified N-glycopeptides, three peptides were derived from titin (TTN). Three identified glycopeptides originated from the Cadherin's protein superfamily: protocadherin alpha-C2 (PCDHAC2), protocadherin Fat1 (FAT1), and protocadherin 9 (PCDH9). In addition, two glycopeptides were derived from Ankyrin repeat domain-containing protein 17 (ANKRD17) and Ankyrin repeat domain-containing protein 36B (ANKRD36B), as well as from integrin alpha-2 (ITGA2) and integrin alpha-3 (ITGA3) glycoproteins. The remaining 13 N-glycopeptides originated from different proteins, each.

3.3. Distribution of the Glycopeptides in Different Cancer Types

To assess potential association of the 37 O-glycopeptides and 23 N-glycopeptides with malignancies, we investigated their distribution across the urinary peptide profiles of patients with five different cancer types, in comparison to controls (Figure 1). We screened urinary peptidome profiles of patients with BCa ($n = 487$), PCa ($n = 206$), pancreatic cancer ($n = 88$), CC ($n = 118$), and RCC ($n = 70$).[34] The majority of these glycopeptides were detected in the cancer datasets as summarized in Table S2, Supporting Information. Subsequently, Mann–Whitney test was used to detect significant differences between two compared groups (each individual cancer type and normal controls). In case of O-glycopeptides, this analysis resulted in the identification of seven glycopeptides significantly different between normal subjects and cancer groups (BH adjustment, $p < 0.05$): O-glycopeptides derived from ANKRD17 (*EAGSPPVVE*), CX3CL1 (*ESVVLEPEAT*), fibrinogen alpha chain (FGA; *DEAGSEADHEGTHSTKRGHAKSRP*), PROC (*TPAPLDSVFSSSERAH*), alpha-1-antitrypsin (A1AT; *FSNGADL*), PI16 (*KTKVPS*), and E3 ubiquitin-protein ligase (HUWE1; *TTTAPTPA*). Three O-glycopeptides, originating from A1AT, PI16, and PROC, were observed at less than 30% frequency in all groups, therefore they were excluded from further analysis. The FGA glycopeptide was the only O-glycopeptide identified at increased urinary levels in all five types of cancer comparing to normal subjects ($p < 0.05$; **Figure 3**A). N-glycopeptides derived from ANKRD17 (*ANTSLMG*), TTN (*NKTPIR*), FAT1 (*ATPLYINI*), ubiquitin-associated domain-containing protein 2 (UBAC2; *NKTLIYIL*), and keratin, type II cytoskeletal 8 (KRT8; *TAVTVNQSLLSPLVLEVDP*) were significantly different between normal subjects and cancer groups, all observed at frequencies above 30% in all groups (Table S3, Supporting Information). Moreover, the FAT1-derived glycopeptide was identified at increased urinary levels in all five types of cancer comparing to normal subjects ($p < 0.05$) (Figure 3B).

3.4. Correlation of the Glycopeptides with Age

Considering that the group of normal subjects was much younger (mean of age 30) than the cancer groups (means of age ranging between 61 to 68 years), we investigated further if the differences in O-glycopeptide and N-glycopeptide abundance may be related to age. Investigation of previous data on the association of urinary peptides with age[45] in over 10 000 subjects

 © 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

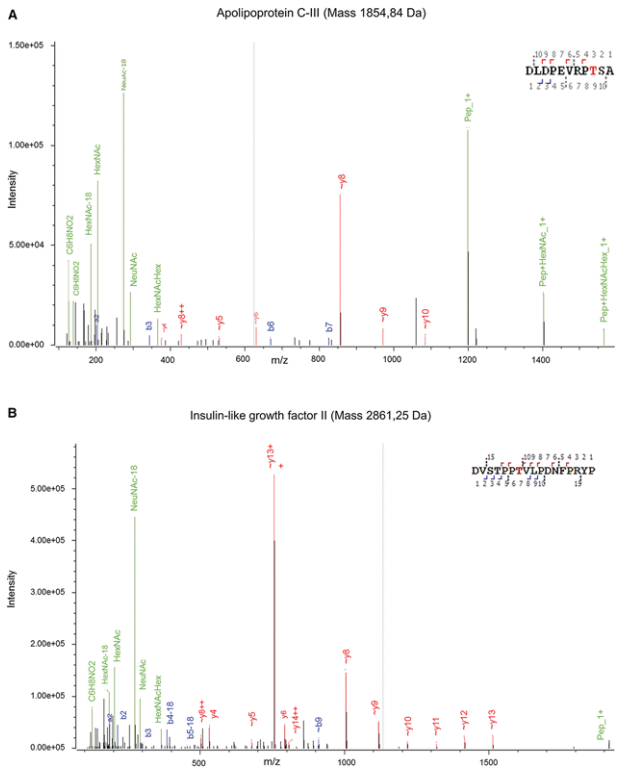


Figure 2. Representation of O-glycopeptide fragmentation spectra, provided by Byonic, identified with HCD/ETD-MS/MS analysis: (A) APOC3 and (B) IGF2.

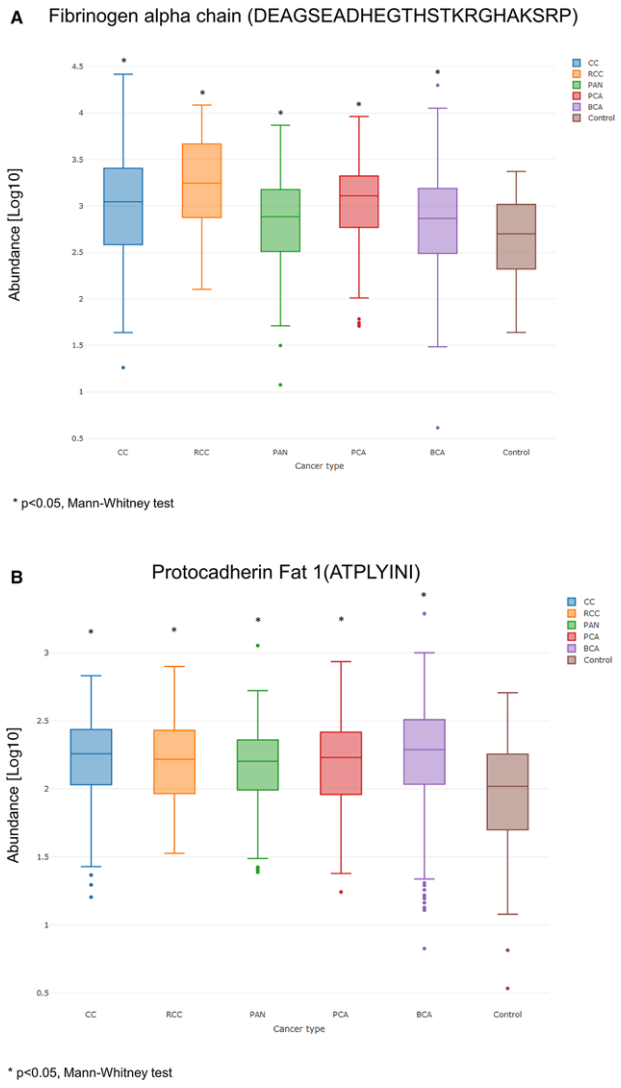


Figure 3. Distribution of the abundance (Log10 transformed) of (A) FGA O-glycopeptide (*DEAGSEADHEGTHSTKRGHAKSRP*) and (B) FAT1 N-glycopeptide (*ATPLYINI*), identified as increased in its urinary levels in all five different cancer types comparing to normal controls ($p < 0.05$). CC, cholangiocarcinoma; RCC, renal cell carcinoma; PAN, pancreatic cancer; PCA, prostate cancer; BCA, bladder cancer.

revealed no significant association of any of the O-glycopeptides and N-glycopeptides with age. We further extracted data from an additional group of older, apparently healthy subjects ($n = 226$, mean of age 60, similar to the age range of the cancer groups). Subsequently, we performed statistical analysis to detect differences between two compared groups (older normal subjects and each individual cancer type). In result, the aforementioned four O-glycopeptides (derived from ANKRD17, CX3CL1, HUWE1, and FGA) and five N-glycopeptides (obtained from ANKRD17, TTN, FAT1, UBAC2, and KRT8) remained significantly different between older normal subjects and cancer groups (BH adjustment, $p < 0.05$). Following this, we further investigated the differences between the urinary profiles of younger (mean of age 30) and older normal subjects (mean of age 60). As a result, we observed that the O-glycopeptide derived from HUWE1 was significantly different between two groups ($p < 0.05$). The mean of HUWE1 glycopeptide abundance was higher in the group of older than in younger normal controls (Log10 transformed: means of 2.87 and 2.44, respectively; fold change 0.84, p-value $= 5.85 \times 10^{-8}$). Based on this, we decided to dismiss the association of HUWE1 glycopeptide with cancer, which resulted in a final list of three O-glycopeptides derived from FGA, ANR17, and CX3CL1, as presented in Table S3, Supporting Information.

4. Discussion

Recent advancements in glycoproteomics for the characterization of glycoproteins open research avenues toward better understanding of their (patho)physiological role. Increased attention has been given to glycoprotein/glycopeptide composition, determining the glycosylation form, glycan structure, and attachment sites.[32,46]

Glycan and glycoprotein expression patterns are considered as an attractive option for the development of novel cancer biomarkers. This is supported by the fact that several glycans and glycoproteins have been approved by the U.S. Food and Drug Administration (FDA) as cancer biomarkers, such as the CA19-9,[47] or the CA125,[48] PSA,[49] CEA,[50] and AFP.

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Despite the improvement in MS technologies and software solutions for protein sequencing, identification of protein/peptide modifications remains challenging, with the main difficulties related to the complexity and high glycan heterogeneity.[51] Both O-glycosylation and N-glycosylation have been implicated in various biological processes, including inflammatory responses, angiogenesis, autoimmunity, and cancer.[17] The purpose of our study was to investigate the intact urinary glycopeptidome, also inspired by the hypothesis that intact glycopeptides may provide valuable information regarding disease-related changes.[21] The most frequently identified O-glycopeptides in our study were APOC3 and IGF2, followed by glycopeptides originating from PROC, ProSAAS, FETUA, CX3CL1, PI16, FGA, and A1AT. Among the identified N-glycopeptides were TTN, ITGA2, ITGA3, KRT8, and FAT1.

These findings are consistent with the study conducted by Halim et al.,[21] where O- and N-glycopeptides (tryptic digests) were detected in urine samples from one healthy control collected in five different time points, using hydrazide chemistry enrichment prior to LC–MS/MS. In our study, we investigated O- and N-glycopeptides in a larger cohort ($n = 238$) using an alternative glycoproteomics workflow for the analysis of native peptides (without enrichment strategy and trypsin digestion but involving specialized analysis of high resolution LC–MS/MS data). The majority of the identified O-glycopeptides were derived from the same proteins as reported by Halim et al. (summarized in Table 2). Moreover, we were able to identify the exact glycan attachment sites in these O-glycopeptides. For example, the glycan attachment sites for IGF2 and APOC3 were at the position of Thr99 and Thr94, respectively, as also reported in previous studies.[21,52] In case of FEUTA, the glycan positions were at Ser346 and Thr256, as also confirmed by Halim et al.[21] In case of N-glycopeptides, only one N-glycopeptide derived from SERPING1 was consistent with the findings presented by Halim et al.[21] However, the SERPING1 glycan attachment site at position of N352 was confirmed by several studies.[53–55]

To date, only a limited number of studies investigated urinary glycopeptides[56,57] and none of the existing reports described the intact glycopeptide content. Enrichment strategies based on lectin affinity chromatography compromises the unbiased assessment of the glycopeptides, as a result of their glyco-specificity. In our workflow, glycopeptide enrichment was not performed, which affects sensitivity in peptide detection, but allows a more unbiased assessment.

Some glycopeptides were identified with the same peptide sequence and glycan attachment site but bearing different carbohydrate structures. For instance, two APOC3 glycopeptides, *WDLDPEVRPTSA* and *DLDPEVRPTSA*, and one glycopeptide obtained from IGF2—*DVSTPPTVLPDNFPRYPVGKF*, were identified with two different carbohydrate structures ($HexNAc_1Hex_1NeuAc_1$ and $HexNAc_1Hex_1NeuAc_2$), which resulted in different molecular masses. These data demonstrate that the exact attachment site in the peptide backbone can be occupied by different glycans, increasing the site-specific glycosylation microheterogeneity.[16]

Interestingly, several O-glycopeptides derived from IGF2, CX3CL1, and FGA have been already reported in the literature in the context of glycosylation changes occurring during tumorigenesis. In cancer, processing of IGF2 is incomplete and high molecular weight isoforms of IGF2 are secreted resulting from aberrant posttranslational processing of the precursor of IGF2 (pro-IGF2).[58–60] However, possibly due to instability of aberrantly glycosylated IGF2, we could not detect a significant change in our dataset. Of note, we could not find any unglycosylated IGF2 peptide, suggesting glycosylation as mandatory prerequisite.

We observed increased urinary levels of an O-glycopeptide derived from CX3CL1 (*ESVVLEPEAT*) in RCC. CX3CL1 has been involved in various diseases, such as cancer,[61] atherosclerosis, and rheumatoid arthritis.[62] CX3CL1 functions as a chemokine, attracting and activating cytotoxic lymphocytes, which may lead to vascular and tissue damage.[63] CX3CL1 contains a chemokine domain, a mucin-like stalk, a single transmembrane region, and a short intracellular C terminus. In particular, glycosylation of its mucin stalk improves its adhesion potency by ensuring the accessibility of the chemokine domain. Both glycopeptides derived from CX3CL1 described in this study are from the mucin stalk region.

A specific FGA O-glycopeptide (*DEAGSEADHEGTHSTKRGHAKSRPV*) was identified at increased levels in all examined cancer types. Fibrinogen, a major plasma glycoprotein, can link tumor cells to extracellular matrix and is important in tumor stroma formation.[64] The FGA O-glycosylation is not well described, however, it has been detected in human using lectin binding.[65] Additionally, the O-glycopeptide as a C-terminal peptide of the fibrinogen α-chain, carrying a disialylated T-antigen, was described as a novel biomarker for the identification of urinary tract infection.[66]

In case of N-glycopeptides, FAT1 (*ATPLYINI*) was identified at increased levels in all examined cancer types. Protocadherin Fat1 is a transmembrane protein belonging to Cadherin superfamily. The FAT1 N-glycosylated peptide has been reported to be involved in the establishment of a proper epithelial cell polarity that requires glycosylation and association of the protein with detergent-resistant membranes.[67]

An N-glycopeptide from KRT8 (*TAVTVNQSLLSPLVLEVDP*) was identified upregulated in urine from prostate cancer, bladder cancer, and cholangiocarcinoma. Keratin type II cytoskeletal 8 is one the major intermediate filament proteins that together with microfilaments and microtubules forms the cytoskeleton in mammalian cells.[68] Aberrant O-glycosylation of KRT8 has been already reported in the literature, resulting in increased solubility and decreased stability by inducing proteasomal degradation.[69] However, to our knowledge, the N-glycosylation of KRT8 has not been reported yet in the context of cancer.

To investigate if the observed glycosylation may be specifically associated with malignant transformation, we aimed at comparison of modified and unmodified peptide. Unfortunately, based on the peptide sequence, we were able to identify only one peptide, derived from FGA (*DEAGSEADHEGTHSTKRGHAKSRP*), also being present in an unglycosylated form. Like the glycosylated FGA derived peptide, the unmodified peptide was significantly upregulated ($p < 0.05$, BH adjustment) in all five cancer types compared to controls.

The present study was only possible based on high mass accuracy of the data obtained. As outlined previously,[70] low mass accuracy in both MS and MS/MS results in the assignment of

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

incorrect sequences. This is even more pronounced when accounting for protein modifications, which further increase the degree of freedom in the sequence assignment, hence the error rate. A further measure to increase the validity of the findings was the application of both HCD and ETD. The use of these two different fragmentation methods, resulting in different fragmentation patterns and the strict requirement for identical sequence assignment, highly increases the validity of the results, but also reduces the number of peptides identified. However, an extensive list of questionable sequence assignments appears of little, if any value.

There are several limitations in the present study that warrant further consideration. For some of the identified glycopeptide sequences, multiple protein assignments were possible. For those sequences, we selected the protein candidates that where already presented as urinary glycoproteins in the study by Halim et al.[21] Nevertheless, for these cases, additional information is needed to confidently assign the glycopeptide candidates to glycoproteins. Furthermore, the study represents a retrospective study based on data obtained in previous work; therefore, prospective collection of samples would be required to properly assess a potential value of any of the identified glycopeptides as biomarkers.

Collectively, in this study, several O- and N-glycan modifications were detected in a number of intact urinary glycopeptides, including Apolipoprotein C-III, IGF 2, fractalkine, fibrinogen alpha, plasma protease C1 inhibitor, protocadherin Fat1, and TTN, in good agreement with the previously reported findings,[21,55] confirming the validity of the approach. To the best of our knowledge, this study is the first report on intact urinary glycopeptides. The present workflow allows the analysis of O- and N-glycosylated intact urinary peptides without applying glycan separation and/or enrichment strategy, which maintains the information regarding glycosylation attachment site that can be used to correlate the glycosylation variants with unglycosylated forms and disease.

Mining of the idenitifed glycopeptides in the urinary peptide profiles of patients with different cancer types suggested significant differences of the glycopeptide abundance in specific cancers. Importantly, as revealed by the additional analysis performed in a second age-matched control group, we were able to confirm that significant changes in glycopeptide abundances are related to cancer, not age. Further investigation of these findings aiming at understanding the biological mechanism and relevance of this modifications in disease and their value as potential biomarkers are with no doubt required and will be pursued.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

We thank Dr. Jenny Ho from Thermo Fisher Scientific, Hemel Hempstead, United Kingdom for the assistance with the LC-MS/MS analysis.

This work was supported by the Chromatin3D grant no. 642934 (Chromatin3D-MSCA-ITN) project funded by the European Commission. The first sentence of the Introduction was corrected on Octorber 31, 2018.

Conflict of Interest

H.M. is the founder and co-owner of Mosaiques Diagnostics. I.B., P.M., M.K., M.F., and M.P. are employed by Mosaiques Diagnostics.

Keywords

glycopeptides, glycosylation, protein biomarkers, posttranslational modifications, urine

Received: June 25, 2018
Revised: August 16, 2018
Published online:

[1] W. Kolch, H. Mischak, A. R. Pitt, *Clin. Sci.* **2005**, *108*, 369.
[2] A. Michalski, E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Mueller, R. Viner, J. Schwartz, P. Remes, M. Belford, J.-J. Dunyach, J. Cox, S. Horning, M. Mann, A. Makarov, *Mol. Cell. Proteomics* **2012**, *11*.
[3] X. Yang, L. Hu, M. Ye, H. Zou, *Anal. Chim. Acta* **2014**, *829*, 40.
[4] M. Pejchinovski, D. Hrnjez, A. Ramirez-Torres, V. Bitsika, G. Mermelekas, A. Vlahou, P. Zuerbig, H. Mischak, J. Metzger, T. Koeck, *Proteomics Clin. Appl.* **2015**, *9*, 453.
[5] Y. Shen, N. Tolic, S. O. Purvine, R. D. Smith, *J. Proteome Res.* **2012**, *11*, 668.
[6] P. Chi, C. D. Allis, G. G. Wang, *Nat. Rev. Cancer* **2010**, *10*, 457.
[7] K. E. Krueger, S. Srivastava, *Mol. Cell. Proteomics* **2006**, *5*, 1799.
[8] E. Song, Y. Mechref, *Biomark. Med.* **2015**, *9*, 835.
[9] R. G. Spiro, *Glycobiology* **2002**, *12*, 43R.
[10] K. D.W., *Nat. Rev. Cancer* **2009**, *9*, 874.
[11] D. Russell, N. J. Oldham, B. G. Davis, *Carbohydr. Res.* **2009**, *344*, 1508.
[12] R. D. Bektas M, *J. Biomed. Res.* **2011**, *25*, 227.
[13] R. E. Schweppe, C. E. Haydon, T. S. Lewis, K. A. Resing, N. G. Ahn, *Acc. Chem. Res.* **2003**, *36*, 453.
[14] Y. Mechref, Y. L. Hu, A. Garcia, A. Hussein, *Electrophoresis* **2012**, *33*, 1755.
[15] A. Berk, H. Lodish, S.L. Zipursky, Section 17.7, W.H. Freeman, 2000, Section 17.7 Molecukar Cell Biology, 4th edition New York.
[16] S. T. Eshghi, W. M. Yang, Y. W. Hu, P. Shah, S. S. Sun, X. D. Li, H. Zhang, *Sci. Rep.* **2016**, *6*, 37189.
[17] P. H. Jensen, D. Kolarich, N. H. Packer, *FEBS J.* **2010**, *277*, 81.
[18] A. Cazet, S. Julien, M. Bobowski, J. Burchell, P. Delannoy, *Breast Cancer Res.* **2010**, *12*, 203.
[19] T. Z. Ju, Y. C. Wang, R. P. Aryal, S. D. Lehoux, X. K. Ding, M. R. Kudelka, C. Cutler, J. W. Zeng, J. M. Wang, X. D. Sun, J. Heimburg-Molinaro, D. F. Smith, R. D. Cummings, *Proteomics Clin. Appl.* **2013**, *7*, 618.
[20] S. Mulagapati, V. Koppolu, T. S. Raju, *Biochemistry* **2017**, *56*, 1218.
[21] A. Halim, J. Nilsson, U. Ruetschi, C. Hesse, G. Larson, *Mol. Cell. Proteomics* **2012**, *11*.
[22] M. Hoffmann, K. Marx, U. Reichl, M. Wuhrer, E. Rapp, *Mol. Cell. Proteomics* **2016**, *15*, 624.
[23] M. Frantzi, J. Metzger, R. E. Banks, H. Husi, J. Klein, M. Dakna, W. Mullen, J. J. Cartledge, J. P. Schanstra, K. Brand, M. A. Kuczyk, H. Mischak, A. Vlahou, D. Theodorescu, A. S. Merseburger, *J. Proteom.* **2014**, *98*, 44.
[24] J. Metzger, A. A. Negm, R. R. Plentz, T. J. Weismueller, J. Wedemeyer, T. H. Karlsen, M. Dakna, W. Mullen, H. Mischak, M. P. Manns, T. O. Lankisch, *Gut* **2013**, *62*, 122.
[25] H. Mischak, J. J. Coon, J. Novak, E. M. Weissinger, J. P. Schanstra, A. F. Dominiczak, *Mass Spectrom. Rev.* **2009**, *28*, 703.
[26] C. Pontillo, S. Filip, D. M. Borras, W. Mullen, A. Vlahou, H. Mischak, *Proteom. Clin. Appl.* **2015**, *9*, 322.

 © 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[27] C. H. Ahrens, E. Brunner, E. Qeli, K. Basler, R. Aebersold, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 789.
[28] M. M. Cox *J. Cell* **2007**, *130*, 395.
[29] A. I. Nesvizhskii, O. Vitek, R. Aebersold, *Nat. Methods* **2007**, *4*, 787.
[30] N. Yang, S. Feng, K. Shedden, X. L. Xie, Y. S. Liu, C. J. Rosser, D. M. Lubman, S. Goodison, *Clin. Cancer Res.* **2011**, *17*, 3349.
[31] M. Bern, Y. J. Kill, C. Becker, *Curr. Protoc. Bioinformatics* **2012**, *40*, 13.20.1.
[32] L. Y. Lee, E. S. X. Moh, B. L. Parker, M. Bern, N. H. Packer, M. Thaysen-Andersen, *J. Proteome Res.* **2016**, *15*, 3904.
[33] J. Siwy, W. Mullen, I. Golovko, J. Franke, P. Zurbig, *Proteom. Clin. Appl.* **2011**, *5*, 367.
[34] I. Belczacka, A. Latosinska, J. Siwy, J. Metzger, A. S. Merseburger, H. Mischak, A. Vlahou, M. Frantzi, V. Jankowski, *Sci. Rep.* **2018**, *8*, 5227.
[35] H. Mischak, A. Vlahou, J. P. A. Ioannidis, *Clin. Biochem.* **2013**, *46*, 432.
[36] H. Mischak, W. Kolch, M. Aivaliotis, D. Bouyssie, M. Court, H. Dihazi, G. H. Dihazi, J. Franke, J. Garin, A. G. de Peredo, A. Iphofer, L. Jansch, C. Lacroix, M. Makridakis, C. Masselon, J. Metzger, B. Monsarrat, M. Mrug, M. Norling, J. Novak, A. Pich, A. Pitt, E. Bongcam-Rudloff, J. Siwy, H. Suzuki, V. Thongboonkerd, L. S. Wang, J. Zoidakis, P. Zurbig, J. P. Schanstra, A. Vlahou, *Proteom. Clin. Appl.* **2010**, *4*, 464.
[37] J. Klein, T. Papadopoulos, H. Mischak, W. Mullen, *Electrophoresis* **2014**, *35*, 1060.
[38] T. Kaiser, A. Hermann, J. T. Kielstein, S. Wittke, S. Bartel, R. Krebs, F. Hausadel, M. Hillmann, I. Golovko, P. Koester, H. Haller, E. M. Weissinger, D. Fliser, H. Mischak, *J. Chromatogr. A* **2003**, *1013*, 157.
[39] T. Kaiser, S. Wittke, I. Just, R. Krebs, S. Bartel, D. Fliser, H. Mischak, E. M. Weissinger, *Electrophoresis* **2004**, *25*, 2044.
[40] J. Jantos-Siwy, E. Schiffer, K. Brand, G. Schumann, K. Rossing, C. Delles, H. Mischak, J. Metzger, *J. Proteome Res.* **2009**, *8*, 268.
[41] D. Theodorescu, H. Mischak, *World J. Urol.* **2007**, *25*, 435.
[42] X. K. Yin, M. Bern, Q. R. Xing, J. Ho, R. Viner, M. Mayr, *Mol. Cell. Proteom.* **2013**, *12*, 956.
[43] P. Zurbig, M. B. Renfrow, E. Schiffer, J. Novak, M. Walden, S. Wittke, I. Just, M. Pelzing, C. Neususs, D. Theodorescu, K. E. Root, M. M. Ross, H. Mischak, *Electrophoresis* **2006**, *27*, 2111.
[44] Y. Benjamini, Y. Hochberg, *J. Roy. Statist. Soc. Ser. B (Methodological)* **1995**, *57*, 289.
[45] E. Nkuipou-Kenfack, A. Bhat, J. Klein, V. Jankowski, W. Mullen, A. Vlahou, M. Dakna, T. Koeck, J. P. Schanstra, P. Zurbig, K. L. Rudolph, B. Schumacher, A. Pich, H. Mischak, *Oncotarget* **2015**, *6*, 34106.
[46] R. C. Bollineni, C. J. Koehler, R. E. Gislefoss, J. H. Anonsen, B. Thiede, *Sci. Rep.* **2018**, *8*, 2117.
[47] A. C. Berger, M. Garcia, J. P. Hoffman, W. F. Regine, R. A. Abrams, H. Safran, A. Konski, A. B. Benson, J. MacDonald, C. G. Willett, *J. Clin. Oncol.* **2008**, *26*, 5918.
[48] B. Kristjansdottir, K. Levan, K. Partheen, K. Sundfeldt, *Gynecol. Oncol.* **2013**, *131*, 52.
[49] M. P. Wirth, M. Froehner, *Eur. Urol.* **2014**, *65*, 1218.
[50] M. J. Duffy, R. Lamerz, C. Haglund, A. Nicolini, M. Kalousova, L. Holubec, C. Sturgeon, *Int. J. Cancer* **2014**, *134*, 2513.
[51] L. Cao, Y. Qu, Z. R. Zhang, Z. Wang, I. Prytkova, S. Wu, *Expert Rev. Proteomics* **2016**, *13*, 513.
[52] R. U. Nilsson J., Halim A., Hesse C., Carlsohn E., Brinkmalm G., Larson G., *Nat. Methods* **2009**, *6*, 809.
[53] J. Bunkenborg, B. J. Pilch, A. V. Podtelejnikov, J. R. Wisniewski, *Proteomics* **2004**, *4*, 454.
[54] W. Jia, Z. Lu, Y. Fu, H. P. Wang, L. H. Wang, H. Chi, Z. F. Yuan, Z. B. Zheng, L. N. Song, H. H. Han, Y. M. Liang, J. L. Wang, Y. Cai, Y. K. Zhang, Y. L. Deng, W. T. Ying, S. M. He, X. H. Qian, *Mol. Cell. Proteomics* **2009**, *8*, 913.
[55] J. Nilsson, U. Ruetschi, A. Halim, C. Hesse, E. Carlsohn, G. Brinkmalm, G. Larson, *Nat. Methods* **2009**, *6*, 809.
[56] D. E. Sleat, H. Y. Zheng, P. Lobel, *Biochim. Biophys. Acta, Proteins Proteomics* **2007**, *1774*, 368.
[57] L. J. Wang, F. X. Li, W. Sun, S. Z. Wu, X. R. Wang, L. Zhang, D. X. Zheng, J. Wang, Y. H. Gao, *Mol. Cell. Proteomics* **2006**, *5*, 560.
[58] S. J. Duguay, Y. Jin, J. Stein, A. N. Duguay, P. Gardner, D. F. Steiner, *J. Biol. Chem.* **1998**, *273*, 18443.
[59] I. Fukuda, N. Hizuka, Y. Ishikawa, K. Yasumoto, Y. Murakami, A. Sata, J. Morita, M. Kurimoto, Y. Okubo, K. Takano, *Growth Horm. IGF Res.* **2006**, *16*, 211.
[60] S. A. Greenall, J. D. Bentley, L. A. Pearce, J. A. Scoble, L. G. Sparrow, N. A. Bartone, X. W. Xiao, R. C. Baxter, L. J. Cosgrove, T. E. Adams, *J. Biol. Chem.* **2013**, *288*, 59.
[61] J. Guo, T. Chen, B. Wang, M. Zhang, H. An, Z. Guo, Y. Yu, Z. Qin, X. Cao, *Immunol. Lett.* **2003**, *89*, 1.
[62] J. H. Ruth, M. V. Volin, G. K. Haines III, D. C. Woodruff, K. J. Katschke Jr., J. M. Woods, C. C. Park, J. C. M. Morel, A. E. Koch, *Arthritis Rheum.* **2001**, *44*, 1568.
[63] J. F. Bazan, K. B. Bacon, G. Hardiman, W. Wang, K. Soo, D. Rossi, D. R. Greaves, A. Zlotnik, T. J. Schall, *Nature* **1997**, *385*, 640.
[64] C. Gerner, W. Steinkellner, K. Holzmann, A. Gsur, R. Grimm, C. Ensinger, P. Obrist, G. Sauermann, *Thromb. Haemost.* **2001**, *85*, 494.
[65] C. L. Lhote, S. Berger, S. Bourgerie, Y. Duvaliflah, R. Julien, Y. Karamanos, *Infect. Immun.* **1995**, *63*, 1927.
[66] G. Zauner, M. Hoffmann, E. Rapp, C. A. M. Koeleman, I. Dragan, A. M. Deelder, M. Wuhrer, P. J. Hensbergen, *J. Proteome Res.* **2012**, *11*, 5804.
[67] X. Y. Li, J. Jiang, X. Y. Zhao, Y. Zhao, Q. C. Cao, Q. Zhao, H. H. Han, J. F. Wang, Z. X. Yu, B. Peng, W. T. Ying, X. H. Qian, *Oncotarget* **2016**, *7*, 22031.
[68] N. O. Ku, X. J. Zhou, D. M. Toivola, M. B. Omary, *Am. J. Physiol. Gastrointest. Liver Physiol.* **1999**, *277*, G1108.
[69] N. O. Ku, D. M. Toivola, P. Strnad, M. B. Omary, *Nat. Cell Biol.* **2010**, *12*, 876.
[70] M. Pejchinovski, J. Klein, A. Ramirez-Torres, V. Bitsika, G. Mermelekas, A. Vlahou, W. Mullen, H. Mischak, V. Jankowski, *Proteom. Clin. Appl.* **2015**, *9*, 531.

© 2018 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Credits: Belczacka Iwona et al.: "Urinary glycopeptide analysis for the investigation of novel biomarkers". *Proteomics Clinical Applications*. 2019. Volume 13. Issue 3. Page: 1800111. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

Supplementary data

Supplementary tables can be found online and on the CD.

Acknowledgements

I would like to thank all the people that contributed to this work. In particular, I would like to express my appreciation to:

Priv. Doz. Dr. Vera Jankowski for her supervision and scientific guidance.

Prof. Harald Mischak for giving me the opportunity to develop my project as well as his guidance and advices in any aspects of my PhD.

Dr. Antonia Vlahou for valuable advices and assistance in preparation of the research manuscripts.

Dr. Maria Frantzi and Dr. Agnieszka Latosinska for all help, assistance and guidance in conducted research.

Dr. Jochen Metzger and Dr. Justyna Siwy to guide me to critically evaluate my research and for many interesting discussions on biomarkers and proteomics.

Dr. Martin Pejchinovski and Dr. William Mullen for helping me in the evaluation of CE-MS and LC-MS/MS data.

All the people working at Mosaiques Diagnostics to be present in any aspects of the project.

All the partners and PhD students of the Chromatin3D consortium for many interesting discussions and exciting meetings and collaborations.

Very special thanks to my family and friends, for their love and never-ending support.

Affidavit according to § 5 (1) for Data Retention

I hereby declare that the original data forming the basis of this doctoral thesis are stored by me personally, ***Iwona Magdalena Bełczącka****,* ***Bergiusstraße 17****,* ***30655 Hannover***, ***Germany****.*

Affidavit according to § 5 (1) and (2), and § 11 (3) 12 of the doctoral studies regulations

I hereby declare on oath, **Iwona Magdalena Bełczącka**, that I have contributed a significant part, and thus majority, of the publication:

Publication 1:
Iwona Belczacka, Agnieszka Latosinska, Justyna Siwy, Jochen Metzger, Axel S. Merseburger, Harald Mischak, Antonia Vlahou, Maria Frantzi, Vera Jankowski. *Urinary CE-MS peptide marker pattern for detection of solid tumors*; Scientific Reports, March 27. 2018. doi: 10.1038/s41598-018-23585-y

The contributions to the publication were as follows:

Names--→	Iwona Belczacka	Agnieszka Latosinska	Justyna Siwy	Jochen Metzger	Axel S. Merseburger	Harald Mischak	Antonia Vlahou	Maria Frantzi	Vera Jankowski	Sum (%)
Study supervision	40	10	10			10		10	20	100
Study design/conception	60					10	10	10	10	100
Examination of the study participants	55		35	10						100
Data evaluation	60	10	10	10				10		100
Performing experiments: Sequencing of peptides	50	10	30					10		100
Statistical evaluation	65	10	10	5				10		100
Delivery of materials	50			20	10	10		10		100
Interpretation of peptidomics data	65	10	10	5				10		100
Performing experiments: Specificity analysis, Validation in additional set of cancer patients	80	10						10		100
Writing of draft of the manuscript	100									100
Correction of the manuscript	25	10	5	5	5	10	10	10	20	100

The position as the first author obviously arises from this significant contribution.

_____
Signature of the doctoral candidate

As supervisor and corresponding author I confirm the statements of Iwona Magdalena Bełczącka.

_____
Signature of the doctoral supervisor

As co-author I endorse the statement of PD. Dr. Vera Jankowski

_____
Names and signatures of all co-authors

Agnieszka Latosinska | Justyna Siwy | Jochen Metzger

Axel S. Merseburger | Harald Mischak | Antonia Vlahou

Maria Frantzi

Affidavit according to § 5 (1) and (2), and § 11 (3) 12 of the doctoral studies regulations

I hereby declare on oath, **Iwona Magdalena Bełczącka**, that I have contributed a significant part, and thus majority, of the publication:

Publication 2:
Iwona Belczacka, Martin Pejchinovski, Magdalena Krochmal, Pedro Magalhaes, Maria Frantzi, William Mullen, Antonia Vlahou, Harald Mischak, Vera Jankowski**:** *Urinary glycopeptide analysis for the investigation of novel biomarkers*, Proteomics Clinical Applications, October 18. 2018. doi: 10.1002/prca.201800111

The contributions to the publication were as follows:

Names ---→	Iwona Belczacka	Martin Pejchinovski	Magdalena Krochmal	Pedro Magalhaes	Maria Frantzi	William Mullen	Antonia Vlahou	Harald Mischak	Vera Jankowski	Sum (%)
Study supervision	50					10	10	10	20	100
Study design/conception	60					20		10	10	100
Examination of the study participants	55	35				10				100
Data evaluation	80	10	10							100
Performing experiments: Sequencing of peptides	50	20		10		20				100
Statistical evaluation	50		50							100
Delivery of materials	50	20				30				100
Interpretation of glycoproteomics data	60	10	20		5	5				100
Performing experiments: Distribution analysis of identified glycopeptides	60	10	10			20				100
Writing of draft of the manuscript	100									100
Correction of the manuscript	25	10	10	5	5	5	10	15	15	100

The position as the first author obviously arises from this significant contribution.

_____
Signature of the doctoral candidate

As supervisor and corresponding author I confirm the statements of Iwona Magdalena Bełczącka.

_____
Signature of the doctoral supervisor

As co-author I endorse the statement of PD. Dr. Vera Jankowski

_____
Names and signatures of all co-authors

Martin Pejchinovski Magdalena Krochmal Pedro Magalhaes

Maria Frantzi William Mullen Antonia Vlahou

Harald Mischak

ResearchScientist

Iwona Magdalena Bełczącka

Address: Bergiusstraße 17, 30655 Hannover, Germany
Born: 29/05/1990, Świdnik, Poland
Email: i.belczacka@gmail.com

WORK EXPERIENCE

X 2015- X 2018 | Researcher

MOSAIQUES DIAGNOSTICS GMBH, HANNOVER, GERMANY
MARIA SKLODOWSKA-CURIE EUROPEAN TRAINING NETWORK "CHROMATIN 3D"

II-V 2015 | Junior Scientist

MEDICAL UNIVERSITY OF VIENNA, VIENNA, AUSTRIA
INSTITUTE OF VASCULAR BIOLOGY AND THROMBOSIS RESEARCH

VIII-X 2014 | Erasmus Program Internship

MAX DELBRUCK CENTER FOR MOLECULAR MEDICINE, BERLIN, GERMANY
PROF. NORBERT HÜBNER LAB: GENETICS AND GENOMICS OF CARDIOVASULAR DISEASES

EDUCATION

X 2015- 2019 | PhD

RWTH AACHEN UNIVERSITY MEDICAL FACULTY, AACHEN, GERMANY
INSTITUTE FOR MOLECULAR CARDIOVASCULAR RESEARCH
PhD Thesis: "Urinary proteomic analysis in the context of cancer biomarkers". Thesis advisor: Priv. Doz. Vera Jankowski

2012-2014 | Master of Science in Biotechnology

MARIA SKLODOWSKA-CURIE UNIVERSITY IN LUBLIN, POLAND
DEPARTMENT OF GENETICS AND MICROBIOLOGY
Thesis: "Association of the CYP17 gene polymorphism in the cytochrome P450 and HTR2A gene polymorphism with the risk of urinary incontinence". Thesis advisor: Prof: Anna Skorupska.

2009-2012 | Bachelor of Science in Biotechnology

MARIA SKLODOWSKA-CURIE UNIVERSITY IN LUBLIN, POLAND
DEPARTMENT OF GENETICS AND MICROBIOLOGY
Thesis: "Quorum sensing as a communication between bacteria". Thesis advisor: Prof: Anna Skorupska.

SKILLS/ COMPETENCES

SOFT SKILLS

- Project management experience: good attitude in estimating workload and resources needed and in scheduling timelines according to tasks.
- Ability to effectively communicate with specialist from different fields. International experience developed in: English.
- Interacted with many researchers from diverse fields for co-authored research work. Delivered presentations of research works at several international meetings and conferences, at group and project meetings, and in front of the company's scientific advisory board.

SOFTWARE ENGINEERING

- Operating systems: XOS, Microsoft Windows, Linux/Unix
- Programming languages: R/Bioconductor, Python
- Databases: MySQL
- Visualization tools: Tableau Software
- Statistical tools: SPPS and MedCalc

BIOINFORMATICS

- Applications: Extensive working knowledge in analysis of large mass spectrometry/proteomics data using R/Bioconductor and Linux shell scripting. Data analysis and visualization using R packages.
- Statistics: Good understanding of biostatistical methodology and tools. Main statistical methodologies: ROC curve analysis, survival analyses, Cox regression, anova test, logistic regression, linear regression, hypothesis testing.
- Processing and analysis of large raw proteomics data using Proteome Discoverer.
- Assessment of data reliability: comparison of data with public repositories
- Data correlation with different -omics datasets (transcriptomics)
- Functional annotation: UniProt Database, NextProt, Gene Ontology
- Pathway and Interactome analysis/visualization: Ingenuity Pathway Analysis, Cytoscape (plugin ClueGO i.e. Reactome, Kegg pathways etc.), String, Impala
- Literature mining in PubMed, Web of Science

LANGUAGES

- Polish (native), English (fluent), German (basic)

PUBLICATIONS

ORIGINAL RESEARCH ARTICLES

- Ricci P, Magalhaes P, Krochmal M, Pejchinovski M, Daina E, Caruso MR, Goea L, **Belczacka I**, Remuzzi G, Umbhauer M, Drube J, Pape L, Mischak H, Decramer S, Schaefer F, Schanstra JP, Cereghini S, Zürbig P. "Urinary proteome signature of Renal Cysts and Diabetes syndrome in children". Scientific Reports, 2019 February 18
- **Belczacka I,** Latosinska A, Siwy J, Metzger J, Merseburger A.S, Mischak H, Antonia Vlahou A, Frantzi M, Jankowski V. "Urinary CE-MS peptide marker pattern for detection of solid tumors". Scientific Reports, 2018 March 27
- **Belczacka I**, Pejchinovski P, Krochmal M, Magalhaes P, Frantzi M, Mullen W, Vlahou A, Mischak H, Jankowski V. "Urinary glycopeptide analysis for the investigation of novel biomarkers". Proteomics Clinical Applications, 2018 October 08

REVIEW ARTICLES

- Marx D, Metzger J, Olagne J, **Belczacka I**, Faguer S, Colombat M, Husi H, Mullen W, Gwinner W, Caillard S. "Proteomics in kidney allograft transplantation – Application of molecular pathway analysis for kidney allograft disease phenotypic biomarker selection". Proteomics Clinical Applications, 2019 January 25
- Frantzi M, Latosinska A, **Belczacka I**, Mischak H. "Urinary proteomic biomarkers in oncology: ready for implementation?", Expert Review of Proteomics, 2018 November
- **Belczacka I**, Latosinska A, Franzi M, Mischak H. „Urinary CE-MS peptide marker pattern for detection of solid tumors – Beyond the abstract", UroToday online Commentary, 2018 May 10
- **Belczacka I,** Latosinska A, Metzger J, Marx D, Vlahou A, Mischak H, Frantzi M. „Proteomics biomarkers for solid tumors: Current status and future prospects.", Mass Spectrometry Reviews, 2018 June 11

- Marx D, Metzger J, Pejchinovski M, Gil BR, Frantzi M, Latosinska A, **Belczacka I,** Heinzmann S, Husi H, Zoidakis J, Klingele M, Herget-Rosenthal S. "Proteomics and Metabolomics for AKI diagnosis". Seminars of Nephrology, 2018 Jan;38(1):63-87

CONFERENCES AND SEMINARS

- Chromatin3D Project Conference „Genome and epigenome integrity" in Edinburgh, Stockholm (23 September 2018).
- LKI- Oncoforum 15th Edition in Leuven (08 June 2018). Short Talk Presentation: "Urinary CE-MS

peptide marker pattern for detection of solid tumors"

- Epigenetic Mechanisms in Health and Disease. Barcelona Conference on Epigenetics and Cancer (25-27 October 2017), Barcelona, Spain. Poster Presentation: "Urinary CE-MS peptide marker pattern for detection of solid tumors"
- Proteomics Bioinformatics Course, EMBL Institute, Cambridge, UK (04-08 December 2016). Poster Presentation: "Urinary CE-MS peptide marker pattern for detection of solid tumors"
- Conference: FEBS Chromatin Proteomics Workshop 2016, Heraklion, Crete (03-08 October 2016). Poster Presentation: "Urinary CE-MS peptide marker pattern for detection of solid tumors"

AWARDS, RECOGNITIONS AND FELLOWSHIPS

- Travel award for the Proteomics Bioinformatics Course (EMBL-EBI, December 2016, Hinxton, Cambridge, UK)
- Marie Sklodowska Curie Actions (MSCA) Innovative Training Network (ITN) Horizon2020 Framework Programme, Chromatin 3D Programme (October 2015 – October 2018)
- Erasmus Plus Funding programme (August-October 2014) for Internship in Max Delbruck Center for Molecular Medicine, Berlin, Germany

Iwona Magdalena Bełczącka