Journal of Proteomics & Bioinformatics
ISSN: 0974-276X

Uchiyama et al., J Proteomics Bioinform 2015, S5
DOI: 10.4172/jpb.S5-005

Research Article Open Access

Peptidomic Analysis via One-Step Direct Transfer Technology for Colorectal Cancer Biomarker Discovery

Kazuhiko Uchiyama[1], Yuji Naito[1]*, Nobuaki Yagi[1], Katsura Mizushima[1], Yasuki Higashimura[1], Yasuko Hirai[1], Tetsuya Okayama[1], Naohisa Yoshida[1], Kazuhiro Katada[1], Kazuhiro Kamada[1], Osamu Handa[1], Takeshi Ishikawa[1], Tomohisa Takagi[1], Hideyuki Konishi[1], Daisuke Nonaka[2], Kyoichi Asada[2], Lyang-Ja Lee[2], Kenji Tanaka[2], Yoshiaki Kuriu[3], Masayoshi Nakanishi[3], Eigo Otsuji[3] and Yoshito Itoh[1]

[1]Department of Gastroenterology and Hepatology, Kyoto Prefectural University of Medicine, 465 Kajiicho Hirokoji Kawaramachidori, Kamigyo-ku Kyoto 602-8566, Japan

[2]Membrane Protein and Ligand Analysis Center, Protosera Inc., Amagasaki 660-0083, Japan

[3]Division of Digestive Surgery, Department of Surgery, Kyoto Prefectural University of Medicine, 465 Kajiicho Hirokoji Kawaramachidori, Kamigyo-ku Kyoto 602-8566, Japan

Abstract

Colorectal cancer (CRC) is a major cause of cancer-related mortality worldwide. Early CRC diagnosis is critical, since patients diagnosed at an early stage have an increased five-year survival rate after surgical resection. Serum biomarkers for CRC detection have been described and peptidomic analysis is a promising approach because of its diagnostic potential. A total of 72 CRC patients and 63 healthy controls were investigated. We used a comprehensive peptide analysis technique, BLOTCHIP®-MS analysis, a combination of electrophoresis and mass spectrometry, for high sensitivity detection of trace amounts of serum peptides. The prediction model comprised five peptides: *m/z* 1616.66 (fibrinogen alpha chain), *m/z* 2390.26 (alpha-1-antitrypsin), *m/z* 2858.42 (AHSG *S*-cysteinylated form), *m/z* 3622.78 (VASP), and *m/z* 3949.98 (F. XIIIa). The three CRC groups, stages II to IV, II + IIIa, and IIIb + IV, were discriminated from controls. High diagnostic performance was suggested by AUC (0.924), sensitivity (83%), specificity (92%), and median probability ratio (6.80) to CRC stage II to IV. We describe a prediction model for CRC diagnosis using candidate biomarker peptides discovered by a one-step direct transfer technology (BLOTCHIP®-MS analysis). The high throughput technology has high reproducibility and is applicable for peptide quantification and differential analysis for biomarker discovery.

Keywords: Colorectal cancer; Biomarker; Peptidome; Peptidomic analysis; Diagnosis; Mass spectrometry; BLOTCHIP®-MS analysis; High throughput discovery technology; Prediction model; Logistic regression model

Introduction

Colorectal cancer (CRC) is one of the most common causes of cancer-related mortality. In 2013, 1.2 million people worldwide were diagnosed with CRC and more than half will likely die [1]. Since patients diagnosed at stage I have a five-year survival rate of 90–95% after surgical resection, early CRC diagnosis is particularly important [2]. In contrast, patients at advanced stages have a five-year survival rate of only 5-10%. Though colonoscopy is a more accurate and reliable approach for diagnosing CRC, it is invasive and expensive. The fecal occult blood test (FOBT) and fecal immune test have been clinically accepted as noninvasive diagnostic tests for CRC [3]. However, those tests have been reported to provide low sensitivity for CRC detection, especially at an early stage [4]. Therefore, the development of new screening methods that are highly sensitive, specific, and noninvasive are critical for early diagnosis and timely treatment of CRC.

Recently, a number of new CRC detection methods have been developed based on stool samples, including stool DNA and microRNA (miRNA) testing, which have demonstrated a potential for noninvasive CRC detection. Hypermethylation of an increasing number of genes has been associated with colorectal tumorigenesis [5,6], and detection of the methylation of CpG islands in human DNA isolated from stool samples has been reported as a potential approach for noninvasive screening and early CRC diagnosis [7-10]. Fecal miRNA measurements showed higher expression of miR-21 and miR-106 in patients with CRC compared with individuals free of colorectal neoplasia; therefore, they could also be used as diagnostic biomarkers [11].

Use of serum-based markers including proteins, peptides, and metabolites, would be highly attractive for CRC screening, since samples could be obtained noninvasively and could be integrated in a routine health evaluation without the need for additional stool sampling. Using capillary electrophoresis-mass spectrometry or gas-chromatography/mass-spectrometry, comprehensive and quantitative analyses of charged metabolites in tumor and normal tissues from CRC and gastric cancer patients were previously performed [12,13]. However, intraday variances in the levels of some amino acids were observed, which affected the results of these metabolomics studies. Serum proteome analysis has also been reported as a method for diagnosis and therapeutic monitoring of CRC. Hundreds of small- to medium-sized peptides can now be detected using microliters of serum because of advances in Mass Spectrometry (MS), such as matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) [14-17]. Recently, several peptide-profiling studies using these techniques have been reported to diagnose CRC [15,18-24] and showed reliable, reproducible results. Most peptidomic analyses

***Corresponding author:** Yuji Naito, Department of Gastroenterology and Hepatology, Kyoto Prefectural University of Medicine, 465 Kajiicho Hirokoji Kawaramachi Kamigyo-ku, Kyoto, Japan 602-8566, Tel: +81-75-251-5518; Fax: +81-75-251-0710; E-mail: ynaito@koto.kpu-m.ac.jp

Received February 14, 2015; **Accepted** March 13, 2015; **Published** March 17, 2015

Citation: Uchiyama K, Naito Y, Yagi N, Mizushima K, Higashimura Y, et al. (2015) Peptidomic Analysis via One-Step Direct Transfer Technology for Colorectal Cancer Biomarker Discovery. J Proteomics Bioinform S5: 005. doi:10.4172/jpb.S5-005

required the removal of large amounts of plasma proteins with variable unstandardized methods before MS analysis, which resulted in exclusion of endogenous peptides that could not be quantified. Thus, many valuable biomarker candidate peptides have been overlooked [25,26].

To compensate for the drawbacks of conventional proteomic/peptidomic technologies, we successfully developed a one-step direct transfer technology using a new target plate "BLOTCHIP®" for MALDI-TOF/MS analysis. No pretreatment is required for BLOTCHIP®-MS analysis of blood samples because peptides are effectively dissociated from major blood proteins by the detergent component and move far from the plasma/serum proteins during one-dimensional (1-D) polyacrylamide gel electrophoresis process [27].

Accordingly, this new technology enables the detection of peptides that would otherwise be adsorbed to abundant proteins and would thus escape detection [27]. Additionally, the biggest advantage of BLOTCHIP®-MS analysis is its high throughput capacity for discovery of peptide biomarkers. The aim of the present study was to identify and validate serum biomarker peptides and to demonstrate the potential usefulness of these candidates for CRC diagnosis.

Materials and Methods

Ethics

The study was approved by the ethics committee of the Kyoto Prefectural University of Medicine carried out in accordance with the World Medical Association's Declaration of Helsinki [28] and performed between August 2012 and August 2013. Asian patients were attending the gastroenterology outpatient clinic at the Hospital of the Kyoto Prefectural University of Medicine. Patients investigated in the present study provided written informed consent.

Peripheral blood sample collection

Serum from CRC patients (n=72) and healthy controls (n=63) was used in the study; subject details are stated in Table 1. All subjects were examined by large intestine endoscopy and diagnosed as either CRC or healthy control. CRC patients were classified as stages II to IV based on Union for International Cancer Control (UICC) stage. Blood samples were drawn from superficial veins using a 21-gauge needle and collected in 10 mL vacuum blood collecting tubes (Venoject II®, Terumo Co., Tokyo, Japan). Samples were allowed to stand at room temperature (RT) for 1 h and centrifuged at 3,000 rpm for 10 min at RT to obtain sera. The supernatant was stored in aliquots at -80°C until further use. Samples were not subjected to more than two freeze/thaw cycles.

		Control	CRC stage					
			IIa	IIb	IIc	IIIa	IIIb	IV
Number		63	25	3	1	21	9	13
Gender	Male	34	19	2	0	13	6	6
	Female	29	6	1	1	8	3	7
Age	Mean	62.9	67.9	69.0	58.0	63.0	60.0	68.2
	S.D.	13.7	10.6	2.0	-	18.1	13.7	8.5
Cancer location								
	Ascending colon	-	2	1	0	2	1	2
	Transverse colon	-	2	0	0	2	0	2
	Descending colon	-	4	0	0	0	1	2
	Sigmoid colon	-	11	1	1	7	0	2
	Cecum	-	1	0	0	0	0	0
	Rectum	-	5	1	0	10	7	5

Patients with a previous history of CRC or a previous or current diagnosis of cancer other than CRC were excluded.

Table 1: Information about CRC patients and controls.

BLOTCHIP®-MS analysis

Serum peptidomic analysis was conducted by newly-established one-step direct transfer technology "BLOTCHIP®-MS analysis", a rapid quantitative technology for peptidomic analysis [27]. Serum samples were subjected to sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis to separate peptides far from proteins. Next, peptides in the gel were electroblotted onto BLOTCHIP® (Protosera Inc., Amagasaki, Japan). MALDI matrix, α-cyano-4-hydroxycinnamic acid (CHCA) (Sigma-Aldrich Co., MO, USA), was applied directly onto BLOTCHIP® and peptidome profiles were obtained in a linear mode of UltraFlexII TOF/TOF (Bruker Daltonics Inc. MA, USA), as previously described in detail [29]. All sample measurements were repeated four times.

All statistical analyses of MS spectral data were conducted using ClinProTools version 2.2 (Bruker Daltonics). MS spectra obtained were baseline-subtracted, normalized, recalibrated, and peak-picked within the software. Peak heights, which show significant statistical differences between two groups (CRC patients versus controls), were analyzed using the Wilcoxon test, a nonparametric test for 2-group comparisons. A probability of p<0.05 was considered statistically significant. All peaks obtained by the statistical data analysis were examined by manual inspection. Peaks with intensity of lower than 20,000 were rejected. Generally, the height of baseline was below 1,000 in BLOTCHIP®-MS analysis. Therefore, the signal to noise ratio threshold was 20.

Statistical analysis of peptides

Diagnostic performance of the peptides was evaluated using R statistical environment software [30]. Receiver Operation Characteristic (ROC) analysis was performed with package 'Epi' [31] within R software. Areas under the curve (AUC) values were calculated from an ROC curve as an indicator of the diagnostic value. The optimal cutoff thresholds for diagnosis were determined according to Youden's index [32]. AUC confidence intervals (CI) were estimated using non-parametric bootstrapping. Resampling and calculation were repeated 10,000 times, and the 2.5 and 97.5 percentiles of the accumulated diagnostic values were used as a 95% CI of AUC. Multiple logistic regression analysis of peptides was conducted for detection of the best combination of peptides discriminating the two groups using R package 'Aod'. To assess validity of the final logistic model, k-fold cross validation was conducted. In k-fold cross-validation, the dataset (72 cancer patients and 63 controls) was divided into k subsets with equal size. In this study, k=8 was chosen. Logistic model using multiple peptides was trained for 8 times with each time leaving out one of the subsets from training. The omitted subset was applied for calculation of the diagnostic values, e.g. AUC. Eight AUC values were averaged and were compared with the diagnostic performance of the final model .

Identification of peptides

Several sera containing high amount of each peptide were mixed for peptide identification. Peptides were extracted using a Sep-Pak C18 solid-phase extraction cartridge (Waters Corporation, Milford, MA, USA) with 80% v/v acetonitrile (ACN) in water containing 0.1% trifluoroacetic acid (TFA) and the eluent was concentrated up to 100 µL using a CC-105 centrifugal concentrator (TOMY SEIKO Co, Ltd.,

Tokyo, Japan). Next, the solution was diluted with 400 μL of 2% *v/v* ACN in water containing 0.065% TFA (eluent A) and applied to an ÄKTA purifier (GE Healthcare UK Ltd, Buckinghamshire, England) equipped with a C18 silica-based column (COSMOSIL 5C18-AR-II) (Nakalai Tesque, Inc., Kyoto, Japan). The eluate was fractionated into 20 fractions (1 mL each) by a liner gradient of 0-100% of 80% *v/v* ACN in water containing 0.05% TFA against eluent A at a flow rate of 1.0 mL/min. Each fraction was concentrated using a CC-105 centrifugal concentrator up to 10 μL. The peptide sequences were analyzed using MALDI-TOF/TOF (UltraFlex II TOF/TOF) and LC-MS/MS (Q-Exactive; Thermo Fisher Scientific Inc., Waltham, MA, USA). For peptide with molecular weight up to approximately 3500, we used MALDI-TOF-MS to conduct MS/MS analysis. LC-MS/MS was applied for the peptide larger than 3500 Da. In search of MALDI-TOF-MS data, MASCOT software version 2.1 was used for a "MS/MS ions search". Parent peptide and MS/MS ions tolerance parameters were set at ± 100 ppm and ± 0.7 Da, respectively. For LC-MS data, we used MASCOT program version 2.4. Parent peptide and MS/MS tolerance parameters were set at ± 10 Da and ± 0.03Da. Since relatively large peptides were analyzed with LC-MS/MS, the value of the parent peptide tolerance (± 10 Da) was set to allow unanticipated modifications in the sequence. Swiss-Prot sequence database, of which taxonomy was limited to "human", was selected for the searches. "Oxidation", "phosphorylation", "*N*-acetylation", and "*S*-cysteinylation" were selected as variable modifications.

Results

Serum peptidomic analysis of CRC patients and controls

Serum peptidome profiles were obtained from 72 CRC patients (stages II to IV) and 63 healthy control subjects using BLOTCHIP®-MS analysis. All measurements were carried out four times and some samples were repeatedly analyzed; thus, over 600 analyses were done in total within two weeks (80 analyses per day). Data from each peptidome profile was stored in a database, and after completion of the MS analysis, batch processing was performed to identify peptides with significantly different concentrations between CRC patients and control groups. In total, 159 peaks were identified in the *m/z* of 1,500-20,000. The shapes of all peaks obtained by the statistical analyses were examined by manual inspection to reject randomly appearing noise peaks during MALDI measurements as well as weak or faint peaks. Lastly, 31 sharp peaks were obtained.

Identified biomarker candidate peptides

MALDI-TOF/TOF and LC-MS/MS peptide sequencing analyses were performed using partially purified serum peptides after reverse-phase chromatography. Fourteen peptides out of the 31 peaks were finally identified as shown in Table 2. Further examination revealed that these peaks corresponded to a chemokine and eight protein degradation products: alpha-1-antichymotrypsin, alpha-1-antitrypsin, alpha-2-HS-glycoprotein (AHSG), coagulation factor XIII subunit a (F. XIIIa), fibrinogen alpha chain, vasodilator-stimulated phosphoprotein (VASP), prothrombin, and vinculin (listed in alphabetical order), as shown in Table 2. Those peptides, *m/z* 4352.34 and *m/z* 2390.26, were C-terminal peptides of alpha-1-antichymotrypsin and alpha-1-antitrypsin, respectively. The *m/z* 2739.53 peptide was AHSG b chain and *m/z* 2858.42 was uniquely S-cysteinylated at the position of Cys358 on *m/z* 2739.53 (AHSG b chain). The *m/z* 3949.98 peptide was an activation peptide released from F. XIIIa resulting from thrombin proteolysis. Four peptide fragments, *m/z* 1465.66, 1616.66, 2768.23, and 5078.35, were proteolyzed at different positions in the middle region of the fibrinogen alpha chain. Additionally, *m/z* 1616.66 was modified at

#	Protein	Calculated monoisotopic mass [M+H]$^+$	Amino acid number (N-/C-terminus)	Ions score	Swiss-prot accession number	Peptide sequence
1	Alpha-1-antichymotrypsin	4352.34	387-423	102	P01011	VETRTIVRFNRPFLMIIVPTDTQNIFFMSKVTNPKQA
2	Alpha-1-antitrypsin	2390.26	398-418	107	P01009	MIEQNTKSPLFMGKVVNPTQK
3	Alpha-2-HS-glycoprotein	2739.53	341-367	85	P02765	TVVQPSVGAAAGPVVPPCPGRIRHFKV
4	Alpha-2-HS-glycoprotein	2858.42	341-367	77	P02765	TVVQPSVGAAAGPVVPPCPGRIRHFKV (*S*-cysteinylation at Cys358)
5	Coagulation factor XIII subunit A	3949.98	2-38	58	P00488	SETSRTAFGGRRAVPPNNSNAAEDDLPTVELQGVVPR (*N*-acetylation at Ser2)
6	Fibrinogen alpha chain	1465.66	21-35	109	P02671	DSGEGDFLAEGGGVR
7	Fibrinogen alpha chain	1616.66	20-35	110	P02671	ADSGEGDFLAEGGGVR (phosphorylation at Ser22)
8	Fibrinogen alpha chain	2768.23	576-600	90	P02671	SSSYSKQFTSSTSYNRGDSTFESKS
9	Fibrinogen alpha chain	5078.35	528-574	22	P02671	TFPGFFSPMLGEFVSETESRGSESGIFTNTKESSSHHPGI-AEFPSRG (oxidation at Met536)
10	Platelet factor 4	7759.18*1	32-101	-	P02776	EAEEDGDLQCLCVKTTSQVRPRHITSLEVIKAGPHCPTAQLI-ATLKNGRKICLDLQAPLYKKIIKKLLES*2
11	Prothrombin	4089.02	328-363	169	P00734	TFGSGEADCGLRPLFEKKSLEDKTERELLESYIDGR
12	Vasodilator-stimulated phosphoprotein	3622.78	128-160	61	P50552	SVPNGPSPEEVEQQKRQQPGPSEHIERRVSNAG
13	Vinculin	4038.05	859-895	35	P18206	APPKPPLPEGEVPPPRPPPPEEKDEEFPEQKAGEVIN
14	Vinculin	4152.99	871-906	34	P18206	PPPRPPPPEEKDEEFPEQKAGEVINQPMMMAARQLH (oxidation at Met899 and Met900)

*1 Detected as a doubly charged ion on BLOTCHIP®-MS peptide profiling.
*2 Although only partial MS/MS spectra was obtained, mass of the whole chain was identical to the calculated value.

Characteristics of the 14 identified peptides are shown. MALDI-TOF/TOF and LC-MS/MS peptide sequencing analyses were performed using partially purified serum peptides after reverse-phase chromatography. Some of the identified peptides were modified (see details in manuscript).

Table 2: List of identified peptides.

#	Protein	Calculated monoisotopic mass [M+H]+	AUC	Sensitivity (%)	Specificity (%)	Cutoff	Fold change (CRC/Control)	ClinProTools *p*-value*1
1	Alpha-1-antichymotrypsin	4352.34	0.690	60	71	10,723	0.67	< 0.000001
2	Alpha-1-antitrypsin	2390.26	0.685	74	56	4,532	0.83	0.000030
3	Alpha-2-HS-glycoprotein	2739.53	0.742	43	94	54,162	0.75	< 0.000001
4	Alpha-2-HS-glycoprotein	2858.42	0.754	75	70	58,384	0.69	< 0.000001
5	Coagulation factor XIII subunit A	3949.98	0.803	76	73	66,666	0.62	< 0.000001
6	Fibrinogen alpha chain	1465.66	0.583	35	83	182,168	1.08	0.000517
7	Fibrinogen alpha chain	1616.66	0.571	51	67	52,321	1.22	0.006910
8	Fibrinogen alpha chain	2768.23	0.589	57	62	28,522	0.81	0.004350
9	Fibrinogen alpha chain	5078.35	0.647	71	54	20,407	0.84	0.000004
10	Platelet factor 4	7759.18	0.567	81	37	24,139	0.97	0.005640
11	Prothrombin	4089.02	0.549	43	75	7,796	0.98	0.012600
12	Vasodilator-stimulated phosphoprotein	3622.78	0.712	56	81	3,460	1.68	< 0.000001
13	Vinculin	4038.05	0.662	50	78	2,112	2.97	0.024400
14	Vinculin	4152.99	0.636	67	57	9,657	1.80	0.000001

*1 Wilcoxon rank-sum test. *p*-value< 0.05 was considered statistically significant.

BLOTCHIP®-MS analysis of subject serum detected 159 statistically significant peaks, which were differentially expressed in CRC patients and controls. Shapes of all the peaks obtained by the statistical analyses were examined by manual inspection to obtain 31 peaks that were sharp and strong. From these, 14 peptides were finally selected as biomarker candidate peptides for CRC diagnosis. Fold changes indicate the ratio of median values (CRC patients/controls). *P*-values of the Wilcoxon rank-sum tests were calculated using ClinProTools 2.2.

Table 3: Diagnostic performance of the peptides.

Ser22 by phosphorylation and *m/z* 5078.35 was oxidized at Met536. The full length of platelet factor 4 (PF4) (*m/z* 7759.18) was detected as a doubly-charged ion (*m/z* 3879.59) by BLOTCHIP®-MS analysis. The *m/z* 4089.02 peptide was a light chain of thrombin activated from prothrombin, resulting from proteolysis by coagulation factor X. *M/z* 3622.78 was formed by proteolysis at the middle region of VASP. *M/z* 4038.05 and *m/z* 4152.99 were formed by proteolysis at the C-terminal region of vinculin. *M/z* 4152.99 was oxidized at Met899 and Met900.

Variation of biomarker candidate peptides in CRC patient serum

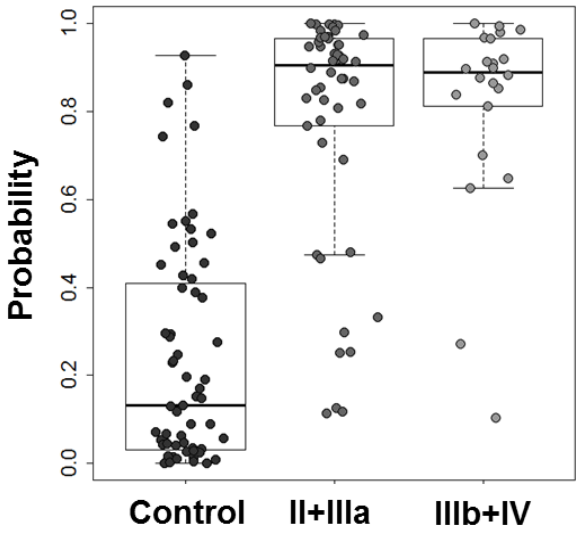
The ratio of concentrations of peptides in CRC serum to control serum varied from 2.97 to 0.62, as represented by 'fold change' in Table 3. Based on the results, it appeared that peptides derived from vinculin (2.97 and 1.80) and VASP (1.68) had a tendency to increase in CRC; conversely, F. XIIIa (0.62) and protein inhibitors (alpha-1-antichymotrypsin 0.67, AHSG 0.69, 0.75, and alpha-1-antitrypsin 0.83) decreased in CRC. Since no major changes were observed in CRC samples, biomarker discovery techniques require a high degree of preciseness and reproducibility.

Multiple logistic regression models

The *m/z* 3949.98 peptide (F. XIIIa) that displayed the highest AUC only reached a maximum of 0.803 as a single marker (Table 3). The model construction was performed with two to five peptides randomly selected from the 14 candidates. After examination of 15,302 combinations of the peptide sets, an optimized model (equation 1 shown below) with the highest AUC of 0.924 was obtained:

$$\text{Probability}=1/(1+\exp[-\{3.836581e\text{-}05\times[m/z\ 1616.66]-5.079913e\text{-}04\times[m/z\ 2390.26]-3.288524e\text{-}05\times[m/z\ 2858.42]+5.243605e\text{-}04\times[m/z\ 3622.78]-5.477792e\text{-}05\times[m/z\ 3949.98]+4.379456\}]) \quad (1)$$

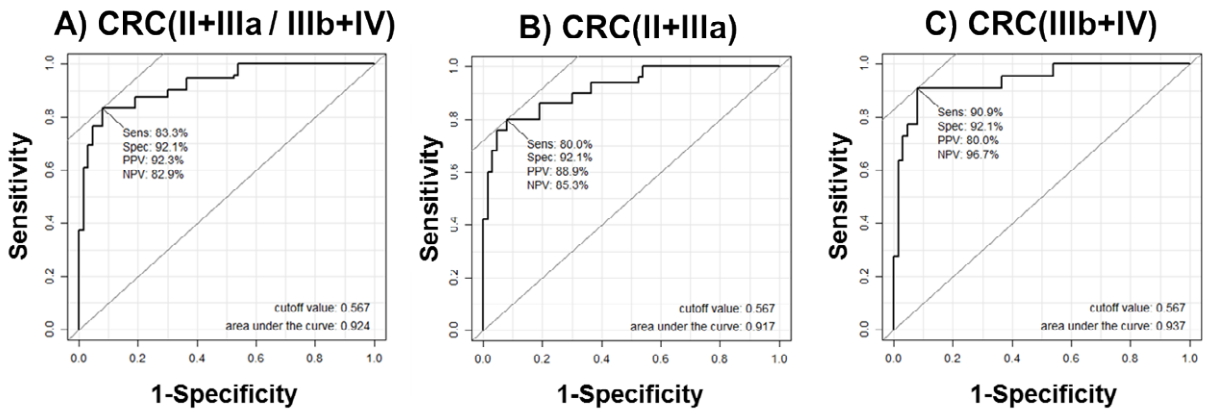
This model was composed of five peptides: *m/z* 1616.66 (fibrinogen alpha chain), *m/z* 2390.26 (alpha-1-antitrypsin), *m/z* 2858.42 (AHSG S-cysteinylated form), *m/z* 3622.78 (VASP), and *m/z* 3949.98 (F. XIIIa).



The probability of CRC stages II + IIIa and IIIb + IV were calculated using the logistic model with five peptides.

Figure 1: Scatterplot and box plot of the probability obtained with the logistic model.

A scatterplot of probability calculated using the model (equation 1) is shown in Figure 1. It is clear that the three CRC groups, stages II to IV (*n*=72), IIa+IIb+IIc + IIIa (*n*=50), and IIIb + IV (*n*=22), were sufficiently discriminated from the control group (*n*=63). Table 4 shows AUC (0.924), sensitivity (83%), specificity (92%), and median probability ratio (6.80) for diagnosis of CRC stage II to IV; AUC (0.917), sensitivity (80%), specificity (92%), median probability ratio (6.86) for diagnosis of stages IIa+IIb+IIc + IIIa; and AUC (0.937), sensitivity (91%),



ROC curves of CRC stages II to IV (A), II + IIIa (B), and IIIb + IV (C) against controls are depicted.

Figure 2: ROC curves for CRC diagnosis using the logistic model with five peptides.

Stage	Number		Diagnostic parameters for CRC detection					Median probability			*p*-value*1
	CRC	Control	AUC (95%CI)		sensitivity (%)	specificity (%)	cutoff	CRC	Control	CRC/ Control	
II+IIIa/IIIb+IV	72	63	0.924	(0.876-0.962)	83	92	0.567	0.898	0.132	6.80	2.40×10^{-17}
II+IIIa	50	63	0.917	(0.861-0.962)	80	92	0.567	0.906	0.132	6.86	2.92×10^{-14}
IIIb+IV	22	63	0.937	(0.878-0.993)	91	92	0.567	0.890	0.132	6.74	1.20×10^{-9}

*1 Mann-Whitney's U-test. p-value<0.05 was considered statistically significant.

The multiple logistic regression model was constructed with 2–5 peptides randomly selected from 14 candidate peptides. After examination of 15,302 combinations of the peptide sets, an optimized model (equation 1 in Results) was obtained with the highest AUC value (0.924), suggesting it was sufficient to discriminate between CRC patient and control groups. Using the same model, the diagnostic parameters for the subsets of CRC groups were also calculated.

Table 4: Diagnostic parameters for CRC diagnosis using the logistic model with five peptides.

specificity (92%), median probability ratio (6.74) for diagnosis of stages IIIb + IV. These values demonstrate the high diagnostic performance of the model. The Hosmer-Lemeshow test indicated an excellent fit of this model to the data (chi-square=8.83 with 8 DF, *p*-value=0.36). Since the final logistic model was constructed using the whole data (72 cancer patients and 63 controls) and its diagnostic performance was evaluated using the same dataset, over-fitting of the model to the dataset might occur. For the purpose of the model validation, we conducted k-fold cross validation. The repeated trainings and tests in 8-fold cross validation resulted in high diagnostic performances with AUC of 0.91 ± 0.07 (average and standard deviation of 8 sample sets), which was comparable to the AUC of the model constructed and self-validated using the whole samples (72 patients and 63 controls) (Table 4). We confirmed that repeating the k-fold cross-validation procedure several times gave essentially the same results. Thus we concluded the effect of over-fitting in the final prediction model was not significant.

Discussion

In our BLOTCHIP®-MS-based peptidomic analysis of CRC patient sera, 159 peaks were detected in the *m/z* of 1,500-20,000. After rejecting peaks from the statistical analysis associated with noise, weakness or faintness, 31 sharp peaks were obtained. Among them, 14 peptides out of the 31 peaks were identified by MALDI-TOF/TOF and LC-MS/MS peptide sequencing analyses. The concentration ratio of these peptides in CRC patients compared with healthy control subjects varied from 2.97 to 0.62. The concentration of coagulation factor XIII subunit a peptide showed the highest AUC (0.803) as a single marker. Finally, the scatterplot of probability calculated by the logistic regression model using five biomarker candidate peptides for the CRC diagnosis displayed high AUC (0.924), sensitivity (83%), specificity (92%), and median probability ratio (6.80).

This was the first study to discover 14 biomarker candidate peptides for diagnosis of CRC using a new technology (BLOTCHIP®-MS analysis). It was composed of two processes: 1-D electrophoresis, during which the peptides were separated far from proteins in the gel under SDS conditions without prior depletion of plasma proteins and MALDI-MS after one-step direct transferring of the peptides (by electro blotting) onto BLOTCHIP®. As a result of these technical improvements, this new technology displayed high throughput characteristics in discovery of novel peptide biomarkers: the analysis capacity reached 80 samples/day [27] and had a high quality of reproducibility. Regarding variance in this new technology, the mean coefficient of variation (CV) was 18% [27]. With regard to the accuracy of the BLOTCHIP®-MS analysis, highly linear standard curves ($R^2 > 0.995$) of both *m/z* 2126.96 (kininogen-1458-477, a candidate biomarker peptide) and *m/z* 5809 (insulin) were obtained at concentrations ranging from 12 fmol to 1.0 pmol per analysis [29]. These experimental data indicated that this new analysis yields satisfactory quantitative results. On the basis of reproducibility, this technology is considered applicable for peptide quantification and differential analysis for biomarker discovery [29,33].

Serum proteomics profiling has been reported to predict the response of treatment and diagnosis of CRC [20] and several candidates have been identified. Biomarker peptides discovered in the present study were classified as follows: (1) full length of chemokine

PF4 (*m/z* 7759.18), (2) subunits: the b chain of AHSG released from hollo protein (*m/z* 2739.53 and 2858.42 S-cysteinylated form) and the light chain of thrombin, (3) activation peptide (*m/z* 3949.98) of F. XIIIa released under physiological and/or pathological blood coagulation, and (4) other peptides derived from functional proteins via proteolysis (owing to unknown disease-specific proteases). We assumed that these peptides might be produced by at least three distinct mechanisms. One of the mechanisms is that peptides were produced in response to pathophysiological state *in vivo*. For example, full-length of PF4 detected in this study was actually existed in the bloodstream. Two alpha-2-HS-glycoprotein peptides, one of which had no modification and the other was *S*-cysteinylated, were full-length B chain of its precursor protein. These peptides were not degraded by proteolysis. Another mechanism is related to activation of blood clotting factors, by which F.XIII subunit a and the light chain of thrombin were released during cancer growing *in vivo* or serum preparation *in vitro*, however, the details are unknown at present. Lastly the interest mechanism by which biomarker candidate peptides were produced is related to proteolysis of membrane protein and high-abundance (plasma) proteins during *in vivo* or serum preparation *in vitro*. Although the biomarker candidate peptides detected here are presumed to be associated with CRC pathophysiology, we must consider further research on the mechanisms relating to their formations and functions for the disease.

The following references discuss the various candidate peptides in more detail: alpha-1-antichymotrypsin [34], alpha-1-antitrypsin [35-38], alpha-2-HS-glycoprotein [20,39-42], F. XIIIa [43], fibrinogen alpha chain [21], PF4 [21], prothrombin [43], vasodilator-stimulated phosphoprotein [44-46], and vinculin [47-50]. However, altered expression of these factors has not been associated with CRC and their known functions are not consistent with CRC pathology.

In the present study, 14 peptides were selected as candidate CRC biomarkers. However, these peptides displayed individual AUC values of 0.571-0.803 and relatively low sensitivity or specificity (Table 3), indicating that single peptide biomarkers are not practical for CRC screening or diagnosis. Thus, the use of multiple biomarkers could be better for biomarker candidates with high sensitivity and specificity. Therefore, we examined a logistic regression model, one of the machine learning methods, to perform highly accurate CRC detection. The upper number of peptides used in the model was limited to five in consideration of future clinical application. After examination of 15,302 combinations of the peptide sets, the optimized model shown in equation 1 was obtained using five peptides: fibrinogen alpha chain, alpha-1-antitrypsin, AHSG S-cysteinylated form, VASP, and F. XIIIa. This screening procedure determined using a multiple logistic regression model showed high sensitivity (83%), specificity (92%), and AUC (0.924) for all advanced CRC cases. As over 90% of patients with stage II CRC are expected to completely recover [49], early diagnosis is critical. Importantly, high sensitivity (80%), specificity (92%), and AUC (0.917) were also obtained for CRC stages IIa+IIb+IIc and IIIa. We would expect a high survival rate by early detection of stages IIa+IIb+IIc and IIIa of CRC. After all the prediction model could contribute to long-term survival and recovery in patients with advanced CRC.

Researchers are focusing their efforts on the discovery of new biomarkers in human DNA, RNA, proteins, peptides and metabolites, in order to advance the field of evidence-based health care and permit early detection and diagnosis of various diseases. Although the potential of peptidomics has been recognized [52,53], results have fallen short of our expectations, owing to technical difficulties. Here, we demonstrated a screening procedure (using a training set in a logistic regression model) for CRC diagnosis using biomarker candidate peptides discovered by an improved peptidomic analysis technology (BLOTCHIP®-MS analysis). The next step is to validate the present prediction model using a validation set to develop a noninvasive and precise diagnostic test for CRC.

Acknowledgement

This work was supported by a Grant-in-Aid for Scientific Research (B) to T.Y. (No. 21390184); a Grant-in-Aid for Challenging Exploratory Research to Yuji Naito (No. 08101559); a Grant-in-Aid for Scientific Research (C) to Kazuhiko Uchiyama (No. 24590665) from the Japan Society for the Promotion of Science; the City Area Program to T.Y. and Yuji Naito from the Ministry of Education, Culture, Sports, Science and Technology, Japan; the Adaptable and Seamless Technology Transfer Program through target-driven R&D to Yuji Naito from the Japan Science and Technology Agency. Part of the data used in this research was originally obtained through the project "The development of peptide biomarker for detection and precise diagnosis in early stage of colorectal cancer" (led by Prof. Nobuaki Yagi) in the Project for Development of Innovative Research on Cancer Therapeutics (P-DIRECT)/Ministry of Education, Culture, Sports, Science and Technology of Japan and was provided through the website of the National Bioscience Database Center (NBDC)/the Japan Science and Technology Agency (JST).

Conflicts of Interest

Y. Itoh and N. Yagi have an affiliation with a donation-funded department from AstraZeneca Co., Ltd., Eisai Co., Ltd., Otsuka Pharmaceutical Co., Ltd., MSD K.K., Dainippon Sumitomo Pharma Co., Ltd., Chugai Pharmaceutical Co., Ltd., FUJIFILM Medical Co., Ltd. and Merck Serono Co., Ltd. Y. Naito received scholarship funds from Otsuka Pharmaceutical Co., Ltd. and Takeda Pharmaceutical Co., Ltd. Y. Itoh received scholarship funds from MSD K.K. and Bristol-Myers K.K. The other authors have no conflicts of interest to declare.

Author Contributions

1. Conception and design of the study: Y Naito, K Tanaka, N Yagi.

2. Analysis and interpretation of data: K Uchiyama, D Nonaka, K Asada, L Lee, K Tanaka.

3. Collection and assembly of data: K Uchiyama, Y Naito, N Yagi, K Mizushima, Y Higashimura, Y Hirai, T Okayama, N Yoshida, K Katada, K Kamada, O Handa, T Ishikawa, T Takagi, H Konishi, Y Kuriu, M Nakanishi, E Otsuji.

4. Drafting of the article: Y Naito, K Uchiyama, K Tanaka.

5. Final approval of the article: Y Naito, Y Itoh.

References

1. Siegel R, Naishadham D, Jemal A (2013) Cancer statistics, 2013. CA: A Cancer journal for clinicians. 63: 11-30.

2. Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, et al. (2003) The case for early detection. Nat Rev Cancer 3: 243-252.

3. Kuipers EJ, Rösch T, Bretthauer M (2013) Colorectal cancer screening--optimizing current strategies and new directions. Nat Rev Clin Oncol 10: 130-142.

4. Morikawa T, Kato J, Yamaji Y, Wada R, Mitsushima T, et al. (2005) A comparison of the immunochemical fecal occult blood test and total colonoscopy in the asymptomatic population. Gastroenterology 129: 422-428.

5. Matsubara N (2012) Epigenetic regulation and colorectal cancer. Dis Colon Rectum 55: 96-104.

6. Lao VV, Grady WM (2011) Epigenetics and colorectal cancer. Nat Rev Gastroenterol Hepatol 8: 686-700.

7. Ahlquist DA (2010) Molecular detection of colorectal neoplasia. Gastroenterology 138: 2127-2139.

8. Azuara D, Rodriguez-Moranta F, de Oca J, Soriano-Izquierdo A, Mora J, et al. (2010) Novel methylation panel for the early detection of colorectal tumors in stool DNA. Clin Colorectal Cancer 9: 168-176.

9. Osborn NK, Ahlquist DA (2005) Stool screening for colorectal cancer: molecular approaches. Gastroenterology 128: 192-206.

10. Ahlquist DA, Zou H, Domanico M, Mahoney DW, Yab TC, et al. (2012) Next-generation stool DNA test accurately detects colorectal cancer and large adenomas. Gastroenterology 142: 248-256.

11. Link A, Balaguer F, Shen Y, Nagasaka T, Lozano JJ, et al. (2010) Fecal MicroRNAs as novel biomarkers for colon cancer screening. Cancer Epidemiol Biomarkers Prev 19: 1766-1774.

12. Hirayama A, Kami K, Sugimoto M, Sugawara M, Toki N, et al. (2009) Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. Cancer Res 69: 4918-4925.

13. Nishiumi S, Kobayashi T, Ikeda A, Yoshie T, Kibi M, et al. (2012) A novel serum metabolomics-based diagnostic approach for colorectal cancer. PLoS One 7: e40459.

14. Huang YJ, Xuan C, Zhang BB, Liao M, Deng KF, et al. (2009) SELDI-TOF MS profiling of serum for detection of nasopharyngeal carcinoma. J Exp Clin Cancer Res 28: 85.

15. Fan NJ, Gao CF, Wang XL, Zhao G, Liu QY, et al. (2012) Serum peptidome patterns of colorectal cancer based on magnetic bead separation and MALDI-TOF mass spectrometry analysis. J Biomed Biotechnol 2012: 985020.

16. Fan NJ, Gao CF, Zhao G, Wang XL, Liu QY (2012) Serum peptidome patterns of breast cancer based on magnetic bead separation and mass spectrometry analysis. Diagn Path 7: 45.

17. Fan NJ, Gao CF, Zhao G, Wang XL, Qiao L (2012) Serum peptidome patterns for early screening of esophageal squamous cell carcinoma. Biotechnol Appl Biochem 59: 276-282.

18. Gemoll T, Roblick UJ, Auer G, Jörnvall H, Habermann JK (2010) SELDI-TOF serum proteomics and colorectal cancer: a current overview. Arch Physiol Biochem 116: 188-196.

19. Helgason HH, Engwegen JY, Zapatka M, Vincent A, Cats A, et al. (2010) Identification of serum proteins as prognostic and predictive markers of colorectal cancer using surface enhanced laser desorption ionization-time of flight mass spectrometry. Oncol Rep 24: 57-64.

20. Fan NJ, Kang R, Ge XY, Li M, Liu Y, et al. (2014) Identification alpha-2-HS-glycoprotein precursor and tubulin beta chain as serology diagnosis biomarker of colorectal cancer. Diagn Pathol 9: 53.

21. Karczmarski J, Rubel T, Mikula M, Wolski J, Rutkowski A, et al. (2013) Pre-analytical-related variability influencing serum peptide profiles demonstrated in a mass spectrometry-based search for colorectal and prostate cancer biomarkers. Acta Biochim Pol 60: 417-425.

22. Liu C, Pan C, Shen J, Wang H, Yong L (2011) MALDI-TOF MS combined with magnetic beads for detecting serum protein biomarkers and establishment of boosting decision tree model for diagnosis of colorectal cancer. Int J Med Sci 8: 39-47.

23. Deng BG, Yao JH, Liu QY, Feng XJ, Liu D, et al. (2013) Comparative serum proteomic analysis of serum diagnosis proteins of colorectal cancer based on magnetic bead separation and maldi-tof mass spectrometry. Asian Pac J Cancer Prev 14: 6069-6075.

24. Bakun M, Karczmarski J, Poznanski J, Rubel T, Rozga M, et al. (2009) An integrated LC-ESI-MS platform for quantitation of serum peptide ladders. Application for colon carcinoma study. Proteomics Clin Appl 3: 932-946.

25. Lowenthal MS, Mehta AI, Frogale K, Bandle RW, Araujo RP, et al. (2005) Analysis of albumin-associated peptides and proteins from ovarian cancer patients. Clin Chem 51: 1933-1945.

26. Granger J, Siddiqui J, Copeland S, Remick D (2005) Albumin depletion of human plasma also removes low abundance proteins including the cytokines. Proteomics 5: 4713-4718.

27. Tanaka K, Tsugawa N, Kim YO, Sanuki N, Takeda U, et al. (2009) A new rapid and comprehensive peptidome analysis by one-step direct transfer technology for 1-D electrophoresis/MALDI mass spectrometry. Biochem Biophys Res Commun 379: 110-114.

28. Association WM (2008) World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. Seoul: From the 59th World Medical Association Assembly.

29. Araki Y, Nonaka D, Tajima A, Maruyama M, Nitto T, et al. (2011) Quantitative peptidomic analysis by a newly developed one-step direct transfer technology without depletion of major blood proteins: its potential utility for monitoring of pathophysiological status in pregnancy-induced hypertension. Proteomics 11: 2727-2737.

30. Dean CB, Nielsen JD (2007) Generalized linear mixed models: a review and some extensions. Lifetime Data Anal 13: 497-512.

31. Carstensen BMP, Laara E, Hills M (2013) Epi: A Package for Statistical Analysis in Epidemiology. R package version 1.1.49. 2013.

32. Youden WJ (1950) Index for rating diagnostic tests. Cancer 3: 32-35.

33. Hashiguchi T, Tanaka K, Lee LJ, Sasaki K, Natsugoe S, et al. (2009) Diagnostic value of serum peptidome analyses for protease activated pathological conditions beyond cancer diagnosis. Med Hypotheses 73: 760-763.

34. Choi JW, Liu H, Shin DH, Yu GI, Hwang JS, et al. (2013) Proteomic and cytokine plasma biomarkers for predicting progression from colorectal adenoma to carcinoma in human patients. Proteomics 13: 2361-2374.

35. Xie LQ, Zhao C, Cai SJ, Xu Y, Huang LY, et al. (2010) Novel proteomic strategy reveal combined alpha1 antitrypsin and cathepsin D as biomarkers for colorectal cancer early screening. J Proteome Res 9: 4701-4709.

36. Shamamian P, Schwartz JD, Pocock BJ, Monea S, Whiting D, et al. (2001) Activation of progelatinase A (MMP-2) by neutrophil elastase, cathepsin G, and proteinase-3: a role for inflammatory cells in tumor invasion and angiogenesis. J Cell Physiol 189: 197-206.

37. Kataoka H, Uchino H, Iwamura T, Seiki M, Nabeshima K, et al. (1999) Enhanced tumor growth and invasiveness *in vivo* by a carboxyl-terminal fragment of alpha1-proteinase inhibitor generated by matrix metalloproteinases: a possible modulatory role in natural killer cytotoxicity. Am J Pathol 154: 457-468.

38. Perez-Holanda S, Blanco I, Menendez M, Rodrigo L (2014) Serum concentration of alpha-1 antitrypsin is significantly higher in colorectal cancer patients than in healthy controls. BMC cancer 14: 355.

39. Nangami GN, Watson K, Parker-Johnson K, Okereke KO, Sakwe A, et al. (2013) Fetuin-A (α2HS-glycoprotein) is a serum chemo-attractant that also promotes invasion of tumor cells through Matrigel. Biochem Biophys Res Commun 438: 660-665.

40. Watson K, Koumangoye R, Thompson P, Sakwe AM, Patel T, et al. (2012) Fetuin-A triggers the secretion of a novel set of exosomes in detached tumor cells that mediate their adhesion and spreading. FEBS Lett 586: 3458-3463.

41. Maxwell F, McGlynn LM, Muir HC, Talwar D, Benzeval M, et al. (2011) Telomere attrition and decreased fetuin-A levels indicate accelerated biological aging and are implicated in the pathogenesis of colorectal cancer. Clin Cancer Res 17: 5573-5581.

42. Quaresima B, Crugliano T, Gaspari M, Faniello MC, Cosimo P, et al. (2008) A proteomics approach to identify changes in protein profiles in serum of Familial Adenomatous Polyposis patients. Cancer Lett 272: 40-52.

43. Kawai K, Watanabe T (2014) Colorectal cancer and hypercoagulability. Surg Today 44: 797-803.

44. Ali M, Rogers LK, Pitari GM3 (2015) Serine phosphorylation of vasodilator-stimulated phosphoprotein (VASP) regulates colon cancer cell survival and apoptosis. Life Sci 123: 1-8.

45. Zuzga DS, Pelta-Heller J, Li P, Bombonati A, Waldman SA, et al. (2012) Phosphorylation of vasodilator-stimulated phosphoprotein Ser239 suppresses filopodia and invadopodia in colon cancer. Int J Cancer 130: 2539-2548.

46. Tu K, Li J, Verma VK, Liu C, Billadeau DD, et al. (2015) Vasodilator-stimulated phosphoprotein promotes activation of hepatic stellate cells by regulating Rab11-dependent plasma membrane targeting of transforming growth factor beta receptors. Hepatology 61: 361-374.

47. Takahashi R, Nagayama S, Furu M, Kajita Y, Jin Y, et al. (2014) AFAP1L, a novel associating partner with vinculin, modulates cellular morphology and motility, and promotes the progression of colorectal cancers. Cancer Med 3: 759-774.

48. Goldmann WH, Auernheimer V, Thievessen I, Fabry B (2013) Vinculin, cell mechanics and tumour cell invasion. Cell Biol Int 37: 397-405.

49. Li T, Guo H, Song Y, Zhao X, Shi Y, et al. (2014) Loss of vinculin and membrane-bound β-catenin promotes metastasis and predicts poor prognosis in colorectal cancer. Mol Cancer 13: 263.

50. Wang Y, Kuramitsu Y, Ueno T, Suzuki N, Yoshino S, et al. (2012) Proteomic differential display identifies upregulated vinculin as a possible biomarker of pancreatic cancer. Oncol Rep 28: 1845-1850.

51. Yamashita K, Watanabe M (2009) Clinical significance of tumor markers and an emerging perspective on colorectal cancer. Cancer Sci 100: 195-199.

52. Check E (2004) Proteomics and cancer: running before we can walk? Nature 429: 496-497.

53. Petricoin EF, Belluco C, Araujo RP, Liotta LA (2006) The blood peptidome: a higher dimension of information content for cancer biomarker discovery. Nat Rev Cancer 6: 961-967.

This article was originally published in a special issue, **Proteomics & Biomarker** handled by Editors. Dr. Lifeng Peng, Victoria University of Wellington, New Zealand; Dr. Sreenivasa Rao Ramisetty, Idexx Laboratories, USA