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

Kazuhiko Uchiyama1, Yuji Naito1*, Nobuaki Yagi1, Katsura Mizushima1, Yasuki Higashimura1, Yasuko Hirai1, Tetsuya Okayama1, Naohisa Yoshida1, Kazuhiro Katada1, Kazuhiro Kamada1, Osamu Handa1, Takeshi Ishikawa1, Tomohisa Takagi1, Hideyuki Konishi1, Daisuke Nonaka2, Kyoichi Asada2, Lyang-Ja Lee2, Kenji Tanaka2, Yoshiaki Kuriu3, Masayoshi Nakanishi3, Eigo Otsuji3 and Yoshito Itoh1

1Department of Gastroenterology and Hepatology, Kyoto Prefectural University of Medicine, 465 Kajicho Hirokoji Kawaramachidori, Kamiyogo-ku Kyoto 602-8566, Japan
2Membrane Protein and Ligand Analysis Center, Protosera Inc., Amagasaki 660-0083, Japan
3Division of Digestive Surgery, Department of Surgery, Kyoto Prefectural University of Medicine, 465 Kajicho Hirokoji Kawaramachidori, Kamiyogo-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 Kajicho Hirokoji Kawaramachi Kamiyogo-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


Copyright: © 2015 Uchiyama K, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
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 for peptide identification. Peptides were extracted using a Sep-Pak® (Protosera Inc., Amagasaki, Japan), MALDI matrix, a-cyano-4-hydroxycinnamic acid (CHCA) (Sigma-Aldrich Co., MO, USA), was applied directly onto BLOTCHIP® and peptide 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 CRC patients and healthy 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 Operating 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 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).
To Tokyo, Japan). Next, the solution was diluted with 400 µL of 2% w/v ACN in water containing 0.065% TFA (eluent A) and applied to an AKTA 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 linear gradient of 0-100% of 80% w/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-10 centrifugal concentrator up to 10 µL. The peptide sequences were analyzed using MALDI-TOF/TOF (UltraFlex II 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 analyzed using 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 Cys338 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 (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 peptide 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.

**Table 2: List of identified peptides.**

<table>
<thead>
<tr>
<th>#</th>
<th>Protein</th>
<th>Calculated monoisotopic mass [M+H]+</th>
<th>Amino acid number (N/C-terminus)</th>
<th>Ions score</th>
<th>Swiss-prot accession number</th>
<th>Peptide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alpha-1-antichymotrypsin</td>
<td>4352.34</td>
<td>387-423</td>
<td>102</td>
<td>P01011</td>
<td>VETRTIGPNRFLMIVPDTQNIFFMSKVTNPKQA</td>
</tr>
<tr>
<td>2</td>
<td>Alpha-1-antitrypsin</td>
<td>2390.26</td>
<td>398-418</td>
<td>107</td>
<td>P01009</td>
<td>MEIQNKTPLFMGKVKVPTQGK</td>
</tr>
<tr>
<td>3</td>
<td>Alpha-2-HS-glycoprotein</td>
<td>2739.53</td>
<td>341-367</td>
<td>85</td>
<td>P02765</td>
<td>TVQVPSNGAAAGPVVPPCPGRIRHKFK</td>
</tr>
<tr>
<td>4</td>
<td>Alpha-2-HS-glycoprotein</td>
<td>2858.42</td>
<td>341-367</td>
<td>77</td>
<td>P02765</td>
<td>TVQVPSNGAAAGPVVPPCPGRIRHKFKV</td>
</tr>
<tr>
<td>5</td>
<td>Coagulation factor XIII subunit A</td>
<td>3949.98</td>
<td>2-38</td>
<td>58</td>
<td>P00488</td>
<td>SETSTAFGGRRAVVPNSNAAAEDDLPTVELGQVPR</td>
</tr>
<tr>
<td>6</td>
<td>Fibrinogen alpha chain</td>
<td>1465.66</td>
<td>21-35</td>
<td>109</td>
<td>P02671</td>
<td>DSQEGDFAE8GGRGVS</td>
</tr>
<tr>
<td>7</td>
<td>Fibrinogen alpha chain</td>
<td>1616.66</td>
<td>20-35</td>
<td>110</td>
<td>P02671</td>
<td>ADSQEGDFAE8GGRGVS</td>
</tr>
<tr>
<td>8</td>
<td>Fibrinogen alpha chain</td>
<td>2768.23</td>
<td>576-600</td>
<td>90</td>
<td>P02671</td>
<td>SSSYSKQTSTSSSYNGDSTFESKS</td>
</tr>
<tr>
<td>9</td>
<td>Fibrinogen alpha chain</td>
<td>5078.35</td>
<td>528-574</td>
<td>22</td>
<td>P02671</td>
<td>TFGFPPSMGLFMVSETQSEKSGIFNTKSSSHHPGI-AEPSPRG (oxidation at Met536)</td>
</tr>
<tr>
<td>10</td>
<td>Platelet factor 4</td>
<td>7759.18</td>
<td>32-101</td>
<td>-</td>
<td>P02776</td>
<td>EAEEDGOLQLOCVKTSVQRHRHVSRSNQPHICTAOQ-L-ATJUQGRRKQCLQAPLYKQKIKLILES</td>
</tr>
<tr>
<td>11</td>
<td>Prothrombin</td>
<td>4089.02</td>
<td>328-363</td>
<td>169</td>
<td>P00734</td>
<td>TFGSGAEDGCLRPLFEKKSLEDKTERELSESYIDGR</td>
</tr>
<tr>
<td>12</td>
<td>Vasodilator-stimulated phosphoprotein</td>
<td>3622.78</td>
<td>128-160</td>
<td>61</td>
<td>P05502</td>
<td>SVNPNGPSPVEEQOKROGQPSSEHREVNSAG</td>
</tr>
<tr>
<td>13</td>
<td>Vinculin</td>
<td>4038.05</td>
<td>859-895</td>
<td>35</td>
<td>P18206</td>
<td>APPKPPLEPGEVPPFRPPPEEKEDEEFPEKAGEVIN</td>
</tr>
<tr>
<td>14</td>
<td>Vinculin</td>
<td>4152.99</td>
<td>871-906</td>
<td>34</td>
<td>P18206</td>
<td>PPPRRPPPEEKEDEEFPEKAGEVINQPMMAAROLH (oxidation at Met899 and Met900)</td>
</tr>
</tbody>
</table>

*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).**
The peaks obtained by the statistical analyses were examined by manual inspection to obtain 31 peaks that were sharp and strong.

Tests were calculated using ClinProTools 2.2.

Selected as biomarker candidate peptides for CRC diagnosis. Fold changes indicate the ratio of median values (CRC patients/controls).

4038.05 and m/z 3622.78 was formed by proteolysis at the middle region of VASP.

m/z 4152.99 was oxidized at Met899 and Met900.

Region of vinculin.

2858.42 (AHSG)

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-antitrypsin 0.67, AHSG 0.69, 0.75, and alpha-1-antitrypsin alpha chain), 2390.26 (alpha-1-antitrypsin), 2858.42 (AHSG S-cysteinylated form), 3949.98 (F. XIIIa).

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:

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

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 A/U (0.937), sensitivity (91%),
A) CRC(II+IIa / IIIb+IV)  B) CRC(II+IIa)  C) CRC(IIIb+IV)

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

<table>
<thead>
<tr>
<th>Stage</th>
<th>Number</th>
<th>Diagnostic parameters for CRC detection</th>
<th>Median probability</th>
<th>p-value*1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AUC (95%CI)</td>
<td>sensitivity (%)</td>
<td>specificity (%)</td>
</tr>
<tr>
<td>II+IIa/IIIb+IV</td>
<td>72</td>
<td>0.924 (0.876-0.962)</td>
<td>83</td>
<td>92</td>
</tr>
<tr>
<td>II+IIa</td>
<td>50</td>
<td>0.917 (0.861-0.962)</td>
<td>80</td>
<td>92</td>
</tr>
<tr>
<td>IIIb+IV</td>
<td>22</td>
<td>0.937 (0.878-0.993)</td>
<td>91</td>
<td>92</td>
</tr>
</tbody>
</table>

*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.

Discussion

In our BLOTCHIP®-MS-based peptidomic analysis of CRC patient sera, 159 peaks were detected in the m/z range 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 transference 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 reproducibility 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 (R² > 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 peptides showed the highest AUC (0.803) as a single marker.
PF4 (m/z 7759.18), (2) subunits: the b chain of AHSG released from holo 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 or serum preparation by which biomarker candidate peptides were produced here is related to activation of blood clotting factors, precursor protein. These peptides were not degraded by proteolysis. Another mechanism is related to activation of blood clotting factors, by which F XIIIa 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 13,302 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 of CRC is critical. Importantly, high sensitivity (80%), specificity (92%), and AUC (0.917) were also obtained for CRC stages IIA+IIIb+IIC and IIIa. We would expect a high survival rate by early detection of stages IIA+IIIb+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.


