# Proteomics for early detection of colorectal cancer: recent updates

Abdo Alnabulsi, Graeme I Murray

Pathology, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK.

Address correspondence to: Professor Graeme I Murray

Email g.i.murray@abdn.ac.uk

Phone: +44(0)1224 553794

Fax: +44(0)1224 663002

Number of figure: 2

Number of tables: 0

Words count (excluding, abstract, legends, tables, key issues and references): 5192

Conflict of interest.

Abdo Alnabulsi is a PhD student supported by Vertebrate Antibodies. Graeme Murray is a scientific advisor to Vertebrate Antibodies.

**Abstract**

**Introduction:** Colorectal cancer (CRC) is a common type of cancer with a relatively poor survival rate. The survival rate of patients could be improved if CRC is detected early. Biomarkers associated with early stages of tumour development might provide useful tools for the early diagnosis of colorectal cancer.

**Area covered:** Online searches using PubMed and Google Scholar were performed using keywords and with a focus on recent proteomic studies. The aim of this review is to highlight the need for biomarkers to improve the detection rate of early CRC and provide an overview of proteomic technologies used for biomarker discovery and validation. This review will also discuss recent proteomic studies which focus on identifying biomarkers associated with the early stages of CRC development.

**Expert opinion:** A large number of CRC biomarkers are increasingly being identified by proteomics using diverse approaches. However, the clinical relevance and introduction of these markers into clinical practice cannot be determined without a robust validation process. The size of validation cohorts remains a major limitation in many biomarker studies.

Keywords: biomarker, colorectal cancer, diagnosis, early detection, proteomics, screening

## 1. Introduction

CRC is a common type of malignancy which is the second leading cause of cancer related death in developed countries [1, 2]. The survival rate of CRC patients varies significantly based on the stage of the disease at the time of presentation. The 5-year survival rate of CRC can be as high as 90% for patients with localised disease, declining to around 70% for patients with regional metastasis and 15-20% for patients with distant metastasis [3]. Therefore, improved survival rates could be achieved by detecting CRC early when treatment is more effective. The detection of colorectal adenomas before the development of invasive malignancy may also significantly reduce the risk of CRC and related deaths [4, 5, 6]. However, there is considerable molecular heterogeneity in the development and progression of CRC as multiple molecular pathways are involved [7].

The early stages of CRC development are not often associated with specific symptoms, with some experiencing no symptoms at all [8]. Common symptoms associated with CRC include rectal bleeding, abdominal pain, weight loss and changes in bowel habit [9]. However, only a small minority of patients with these symptoms are diagnosed with CRC [10]. Therefore, population-based screening programs may help in reducing the risk and mortality rates of CRC in part by detecting and removing adenomas [11, 12]. Screening programmes generally rely on risk factors, usually age, to determine which individuals to screen [13]. However, the influence of screening programs for CRC on survival is still being debated [14, 15].

The main methods used in CRC screening programmes are faecal tests (e.g. guaiac-based faecal occult blood test (FOBT), immunochemical faecal occult blood tests (FIT) and faecal multi-DNA tests) and colorectal endoscopy (e.g. colonoscopy and/or flexible sigmoidoscopy). Currently, lower gastrointestinal endoscopy is the optimal method of

detection and removal of colorectal adenomas. According to a recent study, the risk of CRC can be reduced by 30% using a sigmoidoscopy based screening trial even though only the rectum and sigmoid colon are visualised by sigmoidoscopy [16]. However, colorectal endoscopy is invasive, relatively risky (e.g. colon perforation and anaesthetic complications) and expensive [17, 18]. Furthermore, a significant number of adenomas may be missed due to factors related to endoscopic procedure (observation technique of endoscopist, bowel preparation and colonoscopic insertion time) and adenoma (size, number, shape and anatomical location) [19, 20]. Another challenge is which adenomas should be removed/monitored since only a small proportion of adenomas progress to malignancy [21]. Currently, the risk of malignant transformation is mainly determined by histopathological assessment of polyp size, degree of epithelial cell dysplasia and "villousness" [22].

Faecal based tests are cheaper, less invasive and possibly more convenient than colorectal endoscopy. However, the low specificity of the FOBT, the high number of false positives and associated follow-up colonoscopies have raised doubts over its clinical utility as a screening method [23]. The FIT addresses the main analytical problems associated with the FOBT since there is no need for repeated sampling, there are no dietary restrictions and it has a superior sensitivity [24]. Nevertheless, similar to FOBT, the performance of FIT is compromised by the presence of non-bleeding neoplasms and bleeding non-neoplastic conditions [25, 26]. Another clinically approved method for detecting CRC is multi-targeted DNA testing which detects altered DNA markers in cells shed into the stool. Although this test has shown better sensitivity for detecting early CRC and adenomas compared with FIT, the specificity of DNA-based tests was inferior to that of FIT [27]. Therefore, non-invasive detection tools which identify high-risk adenomas and early carcinoma are still needed.

## 2. Proteomic biomarkers

Proteomics describes a wide range of technologies used for large-scale identification, measurement, characterisation and analysis of proteins. Proteomics can be classified into many branches based on the overall objective and technology of proteomic applications (Figure 1). The majority of biomarker studies use quantitative mass spectrometry-based technologies for the identification and profiling of disease-associated or disease-specific protein markers.

The detection and quantification of low-abundant proteins can be challenging in serum samples because of highly abundant and complex mixture of major proteins such as albumin and immunoglobulins [28]. However, the sensitivity of proteomics has significantly improved due to better sample preparation, advances in current technologies and the introduction of new ultrasensitive technologies such as single cell-quantum dot platform [29-32].

A biomarker refers to any measurable molecule that reflects normal or abnormal biological conditions [30]. Different types of molecules can be classified as biomarkers which can be evaluated in specific types of sample using different technologies (Figure 2). Biomarkers can be utilised in screening, diagnosis, prognosis, predicting therapy and monitoring the progression of CRC [33]. While mass spectrometry-based proteomics is mainly used for the discovery of a large number of protein targets, antibody-based techniques are generally essential for the validation of any potential biomarker targets [34-36].

## 3. Recent proteomic studies for the early detection of colorectal neoplasia

Recent proteomic studies were evaluated in terms of assessed protein targets, proteomic methods, validation process, size and quality of sample cohorts, limitations and

potential clinical impact. Based on this evaluation, individual studies were selected for discussion to highlight key findings and potential limitations. The studies, their biomarker targets, proteomic technologies used, patient cohorts and commentary on study selection have been detailed in supplementary information Methods S1 and Table S1. Blood-based samples (serum and plasma), tissue samples, urine and faecal samples and colorectal tumour models (animal models and organoid culture) will be reviewed.

### 3.1. Blood-based biomarkers

Blood is potentially the ideal sample type for early detection markers since samples can be obtained in a straightforward manner at minimal cost, minimal risk and most importantly in a less-invasive manner compared to existing detection methods for example colonoscopy [18]. Moreover, standardised protocols for collecting and processing blood samples can easily be implemented. However, the detection of low abundance proteins remains a challenge.

A potentially useful screening tool for early diagnosis of CRC is the identification of serum-based autoantibodies [37]. Tumour-associated autoantibodies are produced by the immune system as a reaction to the presence of abnormal molecules linked to the presence of a tumour, known as tumour-associated antigens (TAAs). The identification of these molecules in serum samples is mainly achieved through proteomic-based technologies such as ELISA and protein microarrays [38]. For instance, eight TAAs, which were identified previously by protein microarray-based methods, were selected to test their combined ability to detect CRC by a multiplex beads assay using a well-characterised sample cohort containing 307 samples; 135 CRC (stage I=35, stage II=25, stage III=46 and stage IV=29), 65 other cancer types, 14 inflammatory bowel disease and 93 healthy controls [39]. Out of

the eight TAAs, a panel of six TAAs (general transcription factor IIB, EGF-like repeats and discoidin I-like domains 3, HCK proto-oncogene, pim-1 proto-oncogene, serine/threonine kinase 4 and tumour protein P53) diagnosed CRC with 66% sensitivity at 90.0% fixed specificity [39]. Using a similar approach, a panel of tumour-associated autoantibodies (anti-TP53, anti-IMPDH2, anti-MDM2 and anti-MAGEA4) detected early CRC with a sensitivity of 26% (95% CI, 13–45%) and advanced adenomas with a sensitivity of 20% (95% CI, 13–29%) at a specificity of 90% [40]. The discovery cohort comprised of sera samples of 124 healthy controls and 352 CRC (stage I=96, stage II=102, stage III=105 and stage IV=49) and the validation cohort included 100 healthy controls, 29 non-advanced adenomas, 99 advanced adenomas and 45 CRC (stage I=18, stage II=5, stage III=19 and stage IV=3) [40]. Although sensitivity of only 20% is a major limitation [40], both studies present a potentially useful approach whereby multiple TAAs or autoantibodies can be assessed simultaneously to detect early colorectal neoplasms [39, 40]. However, the main limitation in both studies was the size of patient cohorts used to validate the results. There is still a need for additional validation using large and well-characterised cohorts. Furthermore, the clinical utility of multiplex bead assays needs to be verified in external laboratories and needs to be compared to established screening tools before it can be considered for use in clinical practice.

Selected/multiple reaction monitoring-mass spectrometry (S/MRM-MS) is increasingly used as a technology for validating preliminary proteomic discoveries. For example, targeted multiplex MRM-MS assay was used to test a number of protein targets associated with early CRC [41]. The biomarker targets were identified by literature mining of publically available research data. The MRM assay was optimised to enable the analysis of 187 protein targets using liquid chromatography mass spectrometry (LC-MS) [41]. The discovery cohort included 69 healthy controls and 69 CRC cases (stage I=13, stage II=35, stage III=15 and stage IV=6), while the validation cohort included 68 controls and 68 CRC

cases (stage I=16, stage II=35, stage III=14 and stage IV=3). Stage I and II cases were detected with 91% overall accuracy using a protein panel that included 13 targets; alpha-1-acid glycoprotein 1, alpha-1 antitrypsin, amylase alpha 2b, clusterin, complement c9 , enoyl-CoA hydratase 1, ferritin light chain, gelsolin, osteopontin, selenium binding protein 1 , seprase, spondin 2 and tissue inhibitor of metalloproteinases 1 [41].

The suitability of MRM/SRM targeted proteomics as a discovery and validation platform was confirmed by another study [42]. Different protein signature associated with early CRC (caeruloplasmin, serum paraoxonase/arylesterase 1, serpin peptidase inhibitor clade A, leucine-rich alpha-2-glycoprotein and tissue inhibitor of metalloproteinases 1) was identified in plasma samples using LC-MS and validated by SRM-MS [42]. To identify an optimal protein signature this study followed a detailed analytical approach; initial discovery by LC-MS, screening discovery by SRM-targeted MS, training and validation steps using SRM-MS and algorithmic analysis (the patient cohort used in each step is detailed in Table S1). Both studies have shown that SRM/MRM can be used for testing multiple protein targets and may potential be a useful technology in the clinical practice [41, 42]. However, the detection accuracy of SRM assay using a protein biomarker signature was 72% [42] which was not superior to established CRC screening tests such as the FIT (around 80% detection accuracy) [43]. Furthermore, the clinical utility of MRM-based assay is still hindered by lack of standardisation, complex and laborious sample preparation, high cost, low sensitivity and peptide specificity [44]. Therefore, there is a need for further optimisation and validation of the findings using larger cohorts of participants.

Evaluating multi-protein combinations is a strategy that is being increasingly used in many biomarkers studies. Two potentially useful marker panels for the detection of CRC and advanced adenomas were identified and validated by ELISA using well-characterised patient cohort which included plasma samples of 150 CRC (stage I=34, stage II=51, stage III=34 and

stage IV=31), 151 advanced adenomas and 301 healthy controls [45]. The patient cohort was divided into equal discovery and validation cohorts. Advanced adenomas were defined as "1 or more of adenoma size ≥1 cm, sessile serrated polyp ≥1 cm, adenoma with ≥25% villous histologic features and adenoma with high-grade dysplasia". This study evaluated 28 proteins which were identified as potential markers for early CRC in previous research using MRM-targeted MS as discussed above [41]. The optimum performance (diagnostic performance of around 82%) in detecting CRC was observed using a protein panel which included carcinoembryonic antigen, seprase, serpin A3, macrophage migration inhibitory factor, complement component 3, complement component 9, p-selectin glycoprotein ligand 1 and cathepsin D [45]. Advanced adenomas were detected (diagnostic performance of around 65%) using a panel of four proteins consisting of cathepsin D, clusterin, growth differentiation factor 15 and serum amyloid A1 [45]. Nevertheless, the validation cohort was not independent (i.e. internal validation) and included only a small number of samples. Furthermore, there was no rationale for the inclusion of advanced stage CRC as the main focus was the detection of early colorectal neoplasms. Therefore, these findings require to be validated on larger independent patient cohorts comprised of early CRC cases.

Inconsistencies in the findings of different proteomic studies are still being observed as highlighted by the following studies. Using isobaric tag for relative and absolute quantitation-mass spectrometry (iTRAQ-MS), three serpin family proteins (serpin A1, serpin A3 and serpin C1) were identified as being differentially expressed in serum samples of CRC (stage I=2, stage II=2, stage III=4 and stage IV=7) and adenomas (n=15) compared to healthy controls (n=15) [46]. The results were confirmed by ELISA using serum samples of 21 healthy controls and 19 CRC patients (stage I=2, stage II=5, stage III=5 and stage IV=7). An increase in the serum levels of serpin A1 and serpin A3 were observed in CRC patients compared to healthy controls, whereas the level of serpin C1 was lower in CRC patients [46].

The diagnostic accuracy of these markers was 97% for serpin A1, 82% for serpin A3 and 97% for serpin C1. However, these findings are inconsistent with a previous study which measured serpin A3 by ELISA using plasma samples from 311 CRC patients (Dukes A=53, Dukes B=128, Dukes C=107 and Dukes D=23) and 359 healthy controls [47]. This study did not observe a significant change in the plasma level of serpin A3 in CRC patients compared to healthy controls. Further analysis by immunohistochemistry on paired normal colon and CRC tissue samples (Dukes A=17, Dukes B=45, Dukes C=33 and Dukes=9) showed a decrease in the expression of serpin A3 in the early stages of CRC while it increased in the higher stages [47]. Similarly, the level of serpin A1 was associated with advanced stages of CRC, when analysed by immunohistochemistry using 522 CRC samples (lymph node stage: N0=278 and N1-2=244) [48]. Therefore, serpin A1 and serpin A3 proteins might not be suitable markers for the diagnosis of early CRC. Furthermore, larger patient cohorts are needed for validating these preliminary proteomic findings.

More robust findings may be achieved by using a combination of technologies for biomarkers discovery and validation. Biomarker targets associated with CRC were also identified using a combination of proteomic (LC-MS) and metabolomic technologies (ultra-high performance liquid chromatography (UHPLC-MS) and gas chromatography (GS-MS)) [49]. Pyruvate kinase isoenzyme type M2 (M2-PK), gamma enolase, serotonin and 14-3-3 family members were all identified as potential markers for CRC detection. The discovery cohort included plasma samples of 16 CRC patients (stage III=8 and stage IV=8) and 10 healthy controls [49]. The results were confirmed by ELISA analysis using plasma samples from 40 CRC patients (10 for each stage), adenomas (n=20) and healthy controls (n=20). Moreover, immunohistochemical analysis of 14-3-3 epsilon (24 CRC tissue cores with corresponding normal) showed there was an increased expression of this protein in malignant tissue compared to normal colonic tissue. Although this study presented interesting findings,

further validation is required because the sample size was small (n=40) especially for stage I and stage II CRC cases. Additional investigation is especially needed for serotonin and 14-3-3 proteins and gamma enolase, whereas the potential of M2-PK as a marker for detecting early CRC has been extensively studied [50, 51].

An interesting biomarker candidate for the early detection of CRC is microtubule associated protein RP/EB family member 1 (MAPRE1) which has been identified in several studies [52-55]. A combination of LC-MS, antibody array (plasma samples: 60 healthy controls, 60 adenomas and 60 CRC) and immunohistochemistry assessment of fixed tissues (20 healthy controls, 10 adenomas and 66 CRC) was used to determine the association between MAPRE1 and early CRC (Table S1) [52]. The expression of MAPRE1 was found to be higher in both adenoma and CRC when compared to healthy controls. Furthermore, a combination of MAPRE1 with carcinoembryonic antigen and adenylate kinase 1, tested by antibody array, revealed promising results in diagnosing adenoma and early CRC [52]. The increased levels of MAPRE1 (in tissues and plasma) and the relationship between this marker and early CRC have been previously reported in several studies [53-55]. Nevertheless, additional investigation of the role of MAPRE1 in the early stages of CRC development is required.

### 3.2. Tissue-based biomarkers

Tissue samples can be a useful platform for discovery and initial validation of novel biomarkers because large cohorts of well characterised tissue samples are readily available [56]. Moreover, formalin fixed tissue samples have become increasingly more suitable for proteomic analysis because of advances in proteomics especially improvements in the extraction of proteins from formalin fixed wax embedded tissue samples [57].

Protein markers associated with early CRC were identified by LC-MS based proteomics using fixed tissue samples consisting of 36 CRC (pT1N0=16 and pT2N0=20), 20 normal colon samples and 20 diverticulitis inflammatory controls [58]. The validation by immunohistochemistry was performed using 20 healthy controls, 20 diverticulitis controls, 20 low grade adenomas, 20 high grade adenomas and 100 CRC (pT1N0=20, pT2N0=20, pT3=20 and pT4=20). Half of the pT3 and pT4 samples had lymph node metastasis and four samples had metastasis to other organs. The results showed that there was a significant increase in expression of kininogen-1, transport protein Sec24C and olfactomedin-4 in the early stages of CRC compared to that of normal and inflammatory tissues [58]. A similar trend towards increased expression (mainly weak and moderate immunostaining) was also observed in high-grade adenomas compared to lower grade adenomas and normal tissues. This study has therefore identified three markers associated with early CRC and has shown that fixed tissue sample can be a valuable source for proteomic discovery studies. The increased expression of transport protein Sec24C in early CRC is a novel finding that necessitates further investigation. The other two markers, olfactomedin-4 and kininogen-1, have been previously implicated in early CRC [59, 60]. The increased expression of olfactomin-4 in early CRC was detected in a previous study using proteomic-based analysis (iTRAQ labelling and matrix assisted laser desorption ionization time-of-flight (MALDI-TOF/TOF- MS)) of tissue samples [59]. The proteomic findings were validated by immunohistochemistry on 30 adenomas and 84 CRC (stage I=26, stage II=14, stage III=25 and stage IV=19). Hence, olfactomedin-4 might be a potential candidate for early detection of CRC especially since it is also secreted [61]. Similarly, serum levels of kininogen-1 were analysed using MALDI-TOF/TOF-MS and validated by ELISA and immunohistochemistry [60]. The results indicated kininogen-1 might be a useful marker for the early detection of CRC with a diagnostic accuracy of around 66%-70% [60]. This is consistent with previous

research, which indicated that the level of kininogen-1 was higher in advanced adenomas and carcinomas compared to healthy samples [62]. Although promising, further investigation and validation of the role of olfactomedin-4 and kininogen-1 in early carcinoma are still needed since little is known about their roles in CRC.

In addition to fixed tissue, fresh-frozen tissue samples are often used in proteomics. A significant change in the expression of five proteins (S100 calcium-binding protein A9 (S100A9), annexin A3, nicotinamide phosphoribosyltransferase, carboxylesterase 2 and calcium activated chloride channel A1) was detected in CRC tissues when compared to normal colonic tissues [63]. The biomarker targets were identified by performing iTRAQ-MS on 24 fresh-frozen tumour tissues with corresponding normal tissues and by gene microarray analysis of 52 pairs of normal and tumour tissues (stage I=4, stage II=17, stage III=27 and stage IV=4). The results were validated by immunohistochemistry using 18 pairs of fixed normal and tumour tissues (stage I=2, stage II=6, stage III=9 and stage IV=1) and by ELISA using serum samples from 76 healthy controls and 100 CRC cases (stage I=12, stage II=38, stage III=25 and stage IV=25). The serum levels of S100A9 and annexin A3 were significantly higher in CRC patients compared to healthy controls. This is consistent with a recent paper which reported that S100A9 was upregulated in CRC tissues [64]. Furthermore, S100A9 showed a promising performance in differentiating CRC patients from healthy controls (75% sensitivity) by ELISA using 60 serum samples (40 CRC cases and 20 controls) [65]. There is a limited literature on the role of annexin A3 in early stages of CRC development, although there have been many previous reports on the potential role of other annexins (e.g. annexin A2, annexin A4, annexin A5) in tumour development, drug resistance, therapy and prognosis [66-68]. Nonetheless, a likely limitation of the study by Yu, Li and co-workers [63] was the size of validation cohort (76 controls and 100 CRCs). A further limitation in the cohort used for ELISA validation was the significant difference in age

between healthy controls (median age=50 years old) and CRC patients (median age=61 years old). Therefore, although this study presented an effective approach utilising both proteomics and genomics for the identification of protein biomarkers, there is still a requirement to validate the results using much larger cohorts. Additionally, functional assessment of these proteins in the pathogenesis of CRC is also required.

Biomarkers can also be identified by iTRAQ-LC-MS analysis of cancer-associated fibroblasts obtained from tumour tissues and corresponding normal tissues (n=12) [69]. The results were validated by IHC on 121 colon cancer tissues (stage I=31, stage II=53, stage III=9 and stage IV=28), quantitative PCR on 70 colon cancer samples (stage I=8, stage II=26, stage III=22 and IV=14) and using external gene expression datasets (GSE17538, 232 colon cancers (information about tumour stage not stated); GSE33113, 90 stage II colon cancers and GSE12945, 21 stage III colon cancers) [69]. Lysyl oxidase-like 2 (LOXL2) was identified as a promising biomarker for risk classification in early stage CRC patients [69]. LOXL2 was also associated with survival and recurrence, and demonstrated predictive value for adjuvant therapy in stage II colon cancer. Although the main focus was on identifying prognostic markers, the study presented a valuable approach for proteomic analysis of fibroblasts from the stromal compartment of tumours [69].

### 3.3. Faecal and urine-based biomarkers

Urine and faeces are potentially useful samples for early detection markers since they can be obtained in a straightforward and non-invasive manner. Nonetheless, the availability of large and well-characterised patient cohorts may be lacking compared to tissue-based samples.

Faecal M2-PK is one of the most promising marker for early detection of CRC. According to a meta-analysis of eight clinical studies including 2,654 participants, the M2-PK test demonstrated a pooled sensitivity of 79% and specificity of 80% [70]. However, the main limitations of studies included in the meta-analysis were a significant number of false positives in some studies, lack of standardisation in cut-off values, selection bias of participants and heterogeneity of patient/participant cohorts. Moreover, the sensitivity of the M2-PK test for adenomas is still debatable [71]. Therefore, to accurately assess the potential of M2-PK as a marker for early CRC, the diagnostic performance of M2-PK need to be evaluated using a large screening population.

Clinically useful markers can also be identified in urine samples using mass spectrometry technology. A recent study showed there was a relationship between high-risk adenomas and the levels of prostaglandin metabolites (PGE-M) which was measured using LC-MS [72]. The patient cohort comprised of 420 healthy control patients, 130 low-risk adenoma patients and 290 high-risk adenoma patients. This finding is consistent with other proteomic studies that examined urinary PGE-M using the same analytical method [73, 74]. Nevertheless, further validation of the results is still required since PGE-M is implicated in other malignancies and is also associated with a number of other inflammatory conditions [75]. Furthermore, since the study by Bezawada and co-workers [72] did not include CRC samples, evaluation of PGE-M in CRC samples is needed.

### 3.4. Colorectal tumour models

Obtaining sequential clinical samples of tumour at intervals reflecting the progression of colorectal neoplasm is generally not considered ethical. Although analysis of other types of biological samples (e.g. blood, urine and faeces) obtained serially from patients with colorectal neoplasia maybe possible. Therefore, tumour models, especially *in vivo* models,

offer an opportunity for dynamic characterisation of the molecular changes that occur in various stages of tumour development.

The proteome and transcriptome profiles of fourteen organoids (7 colorectal tumours and 7 healthy controls) derived from seven patients were analysed using LC-MS and Affymetrix Human Gene 2.0 ST arrays [76]. Organoids were cultured in special medium after colonic crypts were isolated from surgically resected tissues of untreated colorectal cancer patients. Data analysis showed 78 proteins were upregulated and 227 were downregulated in tumour organoids compared to healthy ones, although only 22 proteins showed similar expression profiles at the transcript level (the proteins are listed in Table S1) [76]. In another study, quantitative LC-MS analysis of membrane-enriched protein fractions derived from colonic organoids identified tyrosine pseudokinase (PTK7) as a marker associated with self-renewal and re-seeding capacity of colonic stem cells (Table S1) [77]. This indicates that organoids could be a useful *in vitro* model which facilitates biomarker discovery through manipulation and analysis of tumour at different stages of development. Furthermore, a personalized patient-specific organoid proteome profile can be used to better understand the early molecular changes in CRC. Future studies may yield promising findings especially if a larger number of organoids representing different stages of CRC development (normal colonic epithelium, adenoma and early carcinoma) are included in the analysis. However, further verification of the suitability of this model is necessary considering the small number of organoids used. Moreover, validation of the results using clinical samples is needed as the organoids are cultured in a medium (rich in growth factors) different from the *in vivo* microenvironment of tumour. The laboratory processing time of colonic crypts is a key factor which can significantly change the RNA and protein expression profiles of tissues [78].

Protein markers associated with early stages of CRC progression have been also identified by proteomic analysis of a CRC mouse model (Apc $^{\text{multiple intestinal neoplasia (min)/+}}$, a nonsense mutation of the adenomatous polyposis coli (*APC)* gene) [79]. $APC^{min/+}$ mice and wild type mice of 8, 13, 18 and 22 weeks old were sacrificed for proteomic analysis. Tumour interstitial fluids and sera from the mice were analysed by iTRAQ-MS and verified by targeted MRM-MS [79]. The results indicated that the early stages of CRC development were associated with a significant increase in the levels of six serine proteases (chymotrypsin-like elastase 1 (CELA1), chymotrypsin-like elastase 2A (CEL2A), chymotrypsinogen B (CTRB1), trypsin 2 (TRY2), trypsin 4 (TRY4) and chymotrypsin like (CTRL)) [79]. The increased levels of these proteins in CRC was confirmed by MRM assay using sera of CRC patients (n=30) and healthy individuals (n=30). The combination of CELA1 and CTRL detected CRC with 90% sensitivity and 80% specificity. The overexpression of CELA1 and CTRL in CRC was also confirmed by immunohistochemistry on tissue microarray comprising 80 pairs of CRC tissues (majority of CRCs were stage II and stage III, Table S1) and corresponding normal tissues. Therefore, this study has presented a robust approach whereby novel protein markers associated with early CRC can be identified using tissue interstitial fluids. However, further investigation of the roles of serine proteases in early CRC is necessary since only a small number of clinical samples were used to validate the results. The serine proteases are members of a large family of proteolytic enzymes which have been implicated in tumour invasion and metastasis through their roles in digestion and cleavage activity of proteins such as matrix proteins [80, 81].

Tissue and faecal samples from CRC animal models could also be used to identify new biomarkers paving the way for subsequent validation on corresponding human samples [82]. For example, a number of proteins, including haemoglobin, haptoglobin, hemopexin, alpha-2-macroglobulin and cadherin-17, were identified by nanoflow reversed-phased LC–

MS/MS analysis of faecal samples from $APC^{min}$ mouse [83]. However, validation of the results using human samples remains essential.

## 4. Conclusions

There is still a need for sensitive, easily measured, reliable and cost-effective biomarkers for the early diagnosis of CRC. Proteomics is generating a rich database for potential biomarkers that are refining our understanding of important molecular pathways involved in the early stages of CRC development. However, it is still unclear when or if any of these targets will be translated into clinically useful tools for the early detection of CRC.

## 5. Expert Commentary

A major limitation in many proteomic studies is the small number of samples used in validating the results. The use of large, well-characterised and statistically adequately powered patient cohorts is essential for robust validation. Another potential limitation observed in a large number of studies was the composition of both the discovery and validation cohorts. Although the inclusion of advanced CRC cases can be useful, the focus of early biomarker studies should be on adenomas and stage I and stage II CRCs.

The analysis of controls *versus* early neoplasm samples is typically essential to the discovery and validation of early detection biomarkers. Therefore, the findings of biomarker studies can be influenced by the method of selecting and clinically classifying control samples (e.g. self-reported asymptomatic individuals or individuals with normal colonoscopy).

A significant proportion of studies have identified protein targets and recommended them as markers for early detection of CRC mainly based on two criteria; the markers were differentially expressed in CRC compared to normal colonic mucosa and they had a reasonable diagnostic accuracy. However, few studies have actually compared the performances of markers with existing screening tests (e.g. FOBT). For markers to have the potential of being introduced to clinical practice, their performance should be at least non-inferior and ideally superior to existing screening tools.

Although similar proteomic technologies are used, different proteins are frequently being identified as potential biomarkers for the same disease. This may be attributed to variations in processing of samples, size and quality of sample cohorts, type of samples, analytical platforms, data processing and interpretation methods [84, 85]. The reproducibility of proteomics could be enhanced if studies follow a standardised experimental approach and adhere to best practice guidelines. Moreover, the introduction of automated algorithms for data analysis and quality control will further improve the consistency of proteomic results [86, 87].

Although genomics and transcriptomics have been a major platform for biomarker discoveries, protein biomarkers are still necessary because they provide reflection on the physiological state of the cell and the phenotype of particular diseases [30]. Integration of genomics and proteomics data can provide better characterisation and understanding of the molecular events underlying CRC development and progression and this is reflected in the consensus molecular subtypes classification [7]. CRC develops through multiple pathways which contribute to the significant clinical variability between patients [7]. Therefore, a single biomarker is unlikely to have sufficient sensitivity or specificity for use as a screening tool for CRC [88]. Combining biomarkers could improve their clinical discriminative and

diagnostic value synergistically. This is reflected by the increasing number of proteomic studies which have focused on biomarker panels rather than a single marker.

The majority of proteomic technologies are still research-oriented and their precise relevance in clinical practice still needs to be established [89]. Nonetheless, new technologies such as SRM/MRM-MS and multiplex beads assays have shown a realistic translational capability [29]. For these assays to be incorporated in clinical practice, there is still a need for extensive assessment in multiple centres and where applicable these assays need to be compared to relevant clinically established assays (e.g. ELISA).

**Five-year view**

In the next few years, more protein biomarkers associated with early detection of CRC will be identified and validated by proteomics. However, the clinical potential of markers will not be fully determined without significant improvements in the validation process. New advancements in proteomic technologies may be integrated in studies of cancer-biomarkers. The quality of biomarker discoveries possibly will improve as more studies will try to address problems in study design, sample preparation, size and quality of patient cohort and protocol standardisation. Large and multidisciplinary research projects combining proteomics and other complementary technologies such as transcriptomics may become more common.

**Key issues**

- CRC is major disease with relatively high mortality rate.
- The early detection of CRC may significantly improve the survival rate of CRC.
- Protein biomarkers can be used as a screening tool to detect CRC at early stages.
- Proteomics technologies enable the identification of a large number of protein biomarkers for early CRC detection.

- Many protein biomarkers have been identified in blood-based samples, tissue samples and cell lines.
- Clinical and non-clinical samples (e.g. *in vitro* tumour models and animal tumour models) are used in proteomic analysis.
- There are weaknesses in the validation process in a large number of proteomic studies.
- Continued advancements in sample processing, detection technologies and computational analysis will gradually address the challenges in proteomics.

**References**

Papers of special note have been highlighted as:

* of interest

** of considerable interest

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin. 2016;66:7-30.

2. Marshall DC, Webb TE, Hall RA, et al. Trends in UK regional cancer mortality 1991–2007. Br J Cancer. 2016;114:340-347.

3. Siegel RL, Miller KD, Fedewa SA, et al. Colorectal cancer statistics, 2017. CA Cancer J Clin. 2017;67:177-193.

4. Kaminski MF, Wieszczy P, Rupinski M, et al. Increased rate of adenoma detection associates with reduced risk of colorectal cancer and death. Gastroenterology. 2017;153:98-105.

5. Strum WB. Colorectal adenomas. N Engl J Med. 2016;374:1065-1075.

6. Singh R, Zorron Cheng Tao Pu L, Koay D, et al. Sessile serrated adenoma/polyps: where are we at in 2016? World J Gastroenterol. 2016;22:7754-7759.

7. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21:1350-1356.

8. Simon K. Colorectal cancer development and advances in screening. Clin Interv Aging. 2016;11:967-976.

9.Todd KH, Thomas CR Jr, editors. Oncologic emergency medicine: principles and practice. Basel (Switzerland): Springer; 2016.

10. Jellema P, Van der Windt DA, Bruinvels DJ, et al. Value of symptoms and additional diagnostic tests for colorectal cancer in primary care: systematic review and meta-analysis. BMJ. 2010;340:c1269.

11. Atkin W, Wooldrage K, Brenner A, et al. Adenoma surveillance and colorectal cancer incidence: a retrospective, multicentre, cohort study. Lancet Oncol. 2017;18:823-834.

* This paper illustrated the potential significance of adenoma detection on reducing the risk of CRC.

12. Løberg M, Kalager M, Holme Ø, et al. Long-term colorectal-cancer mortality after adenoma removal. N Engl J Med. 2014;371:799-807.

13. Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet. 2014;383:1490-1502.

14. Welch HG, Robertson DJ. Colorectal cancer on the decline—why screening can't explain it all. N Engl J Med. 2016;374:1605-1607.

15. Løberg M, Holme Ø, Bretthauer M, et al. Colorectal adenomas, surveillance, and cancer. Lancet Oncol. 2017;18:e427.

16. Mayor S. One-off sigmoidoscopy screening cuts bowel cancer risk for 17 years, UK study finds. BMJ. 2017;356:j991.

17. Creeden J, Junker F, Vogel-Ziebolz S, et al. Serum tests for colorectal cancer screening. Mol Diagn Ther. 2011;15:129-141.

18. Ganepola GA, Nizin J, Rutledge JR, et al. Use of blood-based biomarkers for early diagnosis and surveillance of colorectal cancer. World J Gastrointest Oncol. 2014;6:83-97.

19. Leufkens AM, van Oijen MG, Vleggaar FP, et al. Factors influencing the miss rate of polyps in a back-to-back colonoscopy study. Endoscopy. 2012;44:470-475.

20. Ahn SB, Han DS, Bae JH, et al. The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies. Gut Liver. 2012;6:64-70.

21. Bujanda L, Cosme A, Gil I, et al. Malignant colorectal polyps. World J Gastroenterol. 2010;16:3103-3111.

22. Gibson JA, Odze RD. Pathology of premalignant colorectal neoplasia. Dig Endosc. 2016;28:312-323.

23. Elsafi SH, Alqahtani NI, Zakary NY, et al. The sensitivity, specificity, predictive values, and likelihood ratios of fecal occult blood test for the detection of colorectal cancer in hospital settings. Clin Exp Gastroenterol. 2015;8:279-284.

24. Schreuders EH, Grobbee EJ, Spaander MC, et al. Advances in fecal tests for colorectal cancer screening. Curr Treat Options Gastroenterol. 2016;14:152-162.

25. Symonds EL, Pedersen SK, Baker RT, et al. A blood test for methylated BCAT1 and IKZF1 vs. a fecal immunochemical test for detection of colorectal neoplasia. Clin Transl Gastroenterol. 2016;7:e137.

26. Wong CK, Fedorak RN, Prosser CI, et al. The sensitivity and specificity of guaiac and immunochemical fecal occult blood tests for the detection of advanced colonic adenomas and cancer. Int J Colorectal Dis. 2012;27:1657-1664.

27. Issa IA, Noureddine M. Colorectal cancer screening: an updated review of the available options. World J Gastroenterol. 2017;23:5086-5096.

28. Alvarez-Chaver P, Otero-Estevez O, Paez de la Cadena M, et al. Proteomics for discovery of candidate colorectal cancer biomarkers. World J Gastroenterol. 2014;20:3804-3824.

29. Timms JF, Hale OJ, Cramer R. Advances in mass spectrometry-based cancer research and analysis: from cancer proteomics to clinical diagnostics. Expert Rev Proteomics. 2016;13:593-607.

30. Wang H, Shi T, Qian WJ, et al. The clinical impact of recent advances in LC-MS for cancer biomarker discovery and verification. Expert Rev Proteomics. 2016;13:99-114.

31. Feist P, Hummon AB. Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. Int J Mol Sci. 2015;16:3537-3563.

32. Jacob T, Agarwal A, Ramunno-Johnson D, et al. Ultrasensitive proteomic quantitation of cellular signaling by digitized nanoparticle-protein counting. Sci Rep. 2016;6:28163.

33. Alnabulsi A, Murray GI. Integrative analysis of the colorectal cancer proteome: potential clinical impact. Expert Rev Proteomics. 2016;13:917-927.

34. Alnabulsi A, Swan R, Cash B, et al. The differential expression of omega-3 and omega-6 fatty acid metabolising enzymes in colorectal cancer and its prognostic significance. Br J Cancer. 2017;116:1612-1620.

35. Swan R, Alnabulsi A, Cash B, et al. Characterisation of the oxysterol metabolising enzyme pathway in mismatch repair proficient and deficient colorectal cancer. Oncotarget. 2016;7:46509-46527.

36. O'Dwyer D, Ralton LD, O'Shea A, et al. The proteomics of colorectal cancer: identification of a protein signature associated with prognosis. PLoS One. 2011;6:e27718.

37. Macdonald IK, Parsy-Kowalska CB, Chapman CJ. Autoantibodies: opportunities for early cancer detection. Trends Cancer. 2017;3:198-213.

38. Zaenker P, Ziman MR. Serologic autoantibodies as diagnostic cancer biomarkers--a review. Cancer Epidemiol Biomarkers Prev. 2013;22:2161-2181.

39. Villar-Vazquez R, Padilla G, Fernandez-Acenero MJ, et al. Development of a novel multiplex beads-based assay for autoantibody detection for colorectal cancer diagnosis. Proteomics. 2016;16:1280-1290.

** This study developed a non-invasive diagnostic assay for multiple tumour-associated antigens. The assay was examined on a well-characterised cohort and showed strong diagnostic performance.

40. Chen H, Werner S, Butt J, et al. Prospective evaluation of 64 serum autoantibodies as biomarkers for early detection of colorectal cancer in a true screening setting. Oncotarget. 2016;7:16420-16432.

** This study developed a multiplex bead-based serological assays for assessing a panel of tumour-associated autoantibodies simultaneously.

41. Jones JJ, Wilcox BE, Benz RW, et al. A plasma-based protein marker panel for colorectal cancer detection identified by multiplex targeted mass spectrometry. Clin Colorectal Cancer. 2016;15:186-194.

42. Surinova S, Choi M, Tao S, et al. Prediction of colorectal cancer diagnosis based on circulating plasma proteins. EMBO Mol Med. 2015;7:1166-1178.

** Excellent paper illustrating the benefits of targeted proteomics in the validation of proteins signature.

43. Jensen CD, Corley DA, Quinn VP, et al. Fecal immunochemical test program performance over 4 rounds of annual screening: a retrospective cohort study. Ann Intern Med. 2016;164:456-463.

44. Anjo SI, Santa C, Manadas B. SWATH-MS as a tool for biomarker discovery: from basic research to clinical applications. Proteomics. 2017;doi:10.1002/pmic.201600278.

45. Blume JE, Wilhelmsen M, Benz RW, et al. Discovery and validation of plasma-protein biomarker panels for the detection of colorectal cancer and advanced adenoma in a Danish collection of samples from patients referred for diagnostic colonoscopy. J Appl Lab Med AACC Publ. 2016;jalm-2016.020271.

46. Peltier J, Roperch JP, Audebert S, et al. Quantitative proteomic analysis exploring progression of colorectal cancer: modulation of the serpin family. J Proteomics. 2016;148:139-148.

47. Dimberg J, Strom K, Lofgren S, et al. Expression of the serine protease inhibitor serpinA3 in human colorectal adenocarcinomas. Oncol Lett. 2011;2:413-418.

48. Kwon CH, Park HJ, Lee JR, et al. Snail and serpinA1 promote tumor progression and predict prognosis in colorectal cancer. Oncotarget. 2015;6:20312-20326.

49. Dowling P, Hughes DJ, Larkin AM, et al. Elevated levels of 14-3-3 proteins, serotonin, gamma enolase and pyruvate kinase identified in clinical samples from patients diagnosed with colorectal cancer. Clin Chim Acta. 2015;441:133-141.

* This paper used a combination of technologies and followed a well-designed approach to identify potentially useful biomarkers.

50. Meng W, Zhu HH, Xu ZF, et al. Serum M2-pyruvate kinase: A promising non-invasive biomarker for colorectal cancer mass screening. World J Gastrointest Oncol. 2012;4:145-151.

51. Zhou CF, Li XB, Sun H, et al. Pyruvate kinase type M2 is upregulated in colorectal cancer and promotes proliferation and migration of colon cancer cells. IUBMB Life. 2012;64:775-782.

52. Taguchi A, Rho JH, Yan Q, et al. MAPRE1 as a plasma biomarker for early-stage colorectal cancer and adenomas. Cancer Prev Res. 2015;8:1112-1119.

53. Ladd JJ, Busald T, Johnson MM, et al. Increased plasma levels of the APC-interacting protein MAPRE1, LRG1, and IGFBP2 preceding a diagnosis of colorectal cancer in women. Cancer Prev Res. 2012;5:655-664.

54. Stypula-Cyrus Y, Mutyal NN, Cruz MD, et al. End-binding protein 1 (EB1) up-regulation is an early event in colorectal carcinogenesis. FEBS Lett. 2014;588:829-835.

55. Sugihara Y, Taniguchi H, Kushima R, et al. Proteomic-based identification of the APC-binding protein EB1 as a candidate of novel tissue biomarker and therapeutic target for colorectal cancer. J Proteomics. 2012;75:5342-5355.

56. Brown GT, Cash BG, Blihoghe D, et al. The expression and prognostic significance of retinoic acid metabolising enzymes in colorectal cancer. PLoS One. 2014;9:e90776.

57. Vincenti DC, Murray GI. The proteomics of formalin-fixed wax-embedded tissue. Clin Biochem. 2013;46:546-551.

58. Quesada-Calvo F, Massot C, Bertrand V, et al. OLFM4, KNG1 and Sec24C identified by proteomics and immunohistochemistry as potential markers of early colorectal cancer stages. Clin Proteomics. 2017;14:9.

* This study demonstrated the benefits of tissue samples as a platform for discovery and initial validation of protein biomarkers.

59. Besson D, Pavageau AH, Valo I, et al. A quantitative proteomic approach of the different stages of colorectal cancer establishes OLFM4 as a new nonmetastatic tumor marker. Mol Cell Proteomics. 2011;10:M111.009712.

60. Wang J, Wang X, Lin S, et al. Identification of kininogen-1 as a serum biomarker for the early detection of advanced colorectal adenoma and colorectal cancer. PLoS One. 2013;8:e70519.

61. Guette C, Valo I, Vétillard A, et al. Olfactomedin-4 is a candidate biomarker of solid gastric, colorectal, pancreatic, head and neck, and prostate cancers. Proteomics Clin Appl. 2015;9:58-63.

62. Qiu Y, Patwa TH, Xu L, et al. Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot. J Proteome Res. 2008;7:1693-1703.

63. Yu J, Li X, Zhong C, et al. High-throughput proteomics integrated with gene microarray for discovery of colorectal cancer potential biomarkers. Oncotarget. 2016;7:75279-75292.

64. Peng F, Huang Y, Li MY, et al. Dissecting characteristics and dynamics of differentially expressed proteins during multistage carcinogenesis of human colorectal cancer. World J Gastroenterol. 2016;22:4515-4528.

65. Barderas R, Mendes M, Torres S, et al. In-depth characterization of the secretome of colorectal cancer metastatic cells identifies key proteins in cell adhesion, migration, and invasion. Mol Cell Proteomics. 2013;12:1602-1620.

66. Alfonso P, Canamero M, Fernandez-Carbonie F, et al. Proteome analysis of membrane fractions in colorectal carcinomas by using 2D-DIGE saturation labeling. J Proteome Res. 2008;7:4247-4255.

67. Wu N, Liu S, Guo C, et al. The role of annexin A3 playing in cancers. Clin Transl Oncol. 2013;15:106-110.

68. Mussunoor S, Murray GI. The role of annexins in tumour development and progression. J Pathol. 2008;216:131-140.

69. Torres S, Garcia-Palmero I, Herrera M, et al. LOXL2 is highly expressed in cancer-associated fibroblasts and associates to poor colon cancer survival. Clin Cancer Res. 2015;21:4892-4902.

70. Uppara M, Adaba F, Askari A, et al. A systematic review and meta-analysis of the diagnostic accuracy of pyruvate kinase M2 isoenzymatic assay in diagnosing colorectal cancer. World J Surg Oncol. 2015;13:48.

71. Tonus C, Sellinger M, Koss K, et al. Faecal pyruvate kinase isoenzyme type M2 for colorectal cancer screening: a meta-analysis. World J Gastroenterol. 2012;18:4004-4011.

72. Bezawada N, Song M, Wu K, et al. Urinary PGE-M levels are associated with risk of colorectal adenomas and chemopreventive response to anti-inflammatory drugs. Cancer Prev Res. 2014;7:758-765.

*Excellent paper which compared large number samples collected from prospective cohorts. The paper presented a promising analytical approach for the identification of urinary markers for CRC.

73. Shrubsole MJ, Cai Q, Wen W, et al. Urinary prostaglandin E2 metabolite and risk for colorectal adenoma. Cancer Prev Res. 2012;5:336-342.

74. Johnson JC, Schmidt CR, Shrubsole MJ, et al. Urine PGE-M: a metabolite of prostaglandin E2 as a potential biomarker of advanced colorectal neoplasia. Clin Gastroenterol Hepatol. 2006;4:1358-1365.

75. Colbert Maresso K, Vilar E, Hawk ET. Urinary PGE-M in colorectal cancer: predicting more than risk? Cancer Prev Res. 2014;7:969-972.

76. Cristobal A, van den Toorn HW, van de Wetering M, et al. Personalized proteome profiles of healthy and tumor human colon organoids reveal both individual diversity and basic features of colorectal cancer. Cell Rep. 2017;18:263-274.

77. Jung P, Sommer C, Barriga FM, et al. Isolation of human colon stem cells using surface expression of PTK7. Stem Cell Reports. 2015;5:979-987.

78. David KA, Unger FT, Uhlig P, et al. Surgical procedures and postsurgical tissue processing significantly affect expression of genes and EGFR-pathway proteins in colorectal cancer tissue. Oncotarget. 2014;5:11017-11028.

79. Xie Y, Chen L, Lv X, et al. The levels of serine proteases in colon tissue interstitial fluid and serum serve as an indicator of colorectal cancer progression. Oncotarget. 2016;7:32592-32606.

* Excellent paper illustrating the benefits of tumour models in the discovery of protein biomarkers.

80. Brown GT, Murray GI. Current mechanistic insights into the roles of matrix metalloproteinases in tumour invasion and metastasis. J Pathol. 2015;237:273-281.

81. Herszenyi L, Barabas L, Hritz I, et al. Impact of proteolytic enzymes in colorectal cancer development and progression. World J Gastroenterol. 2014;20:13246-13257.

82. Ma H, Chen G, Guo M. Mass spectrometry based translational proteomics for biomarker discovery and application in colorectal cancer. Proteomics Clin Appl. 2016;10:503-515.

83. Ang CS, Rothacker J, Patsiouras H, et al. Murine fecal proteomics: a model system for the detection of potential biomarkers for colorectal cancer. J Chromatogr A. 2010;1217:3330-3340.

84. Coghlin C, Murray GI. Progress in the development of protein biomarkers of oesophageal and gastric cancers. Proteomics Clin Appl. 2016;10:532-545.

85. Borrebaeck CA. Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer. Nat Rev Cancer. 2017;17:199-204.

86. Bielow C, Mastrobuoni G, Kempa S. Proteomics quality control: quality control software for MaxQuant results. J Proteome Res. 2015;15:777-787.

87. Rost HL, Liu Y, D'Agostino G, et al. TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics. Nat Methods. 2016;13:777-783.

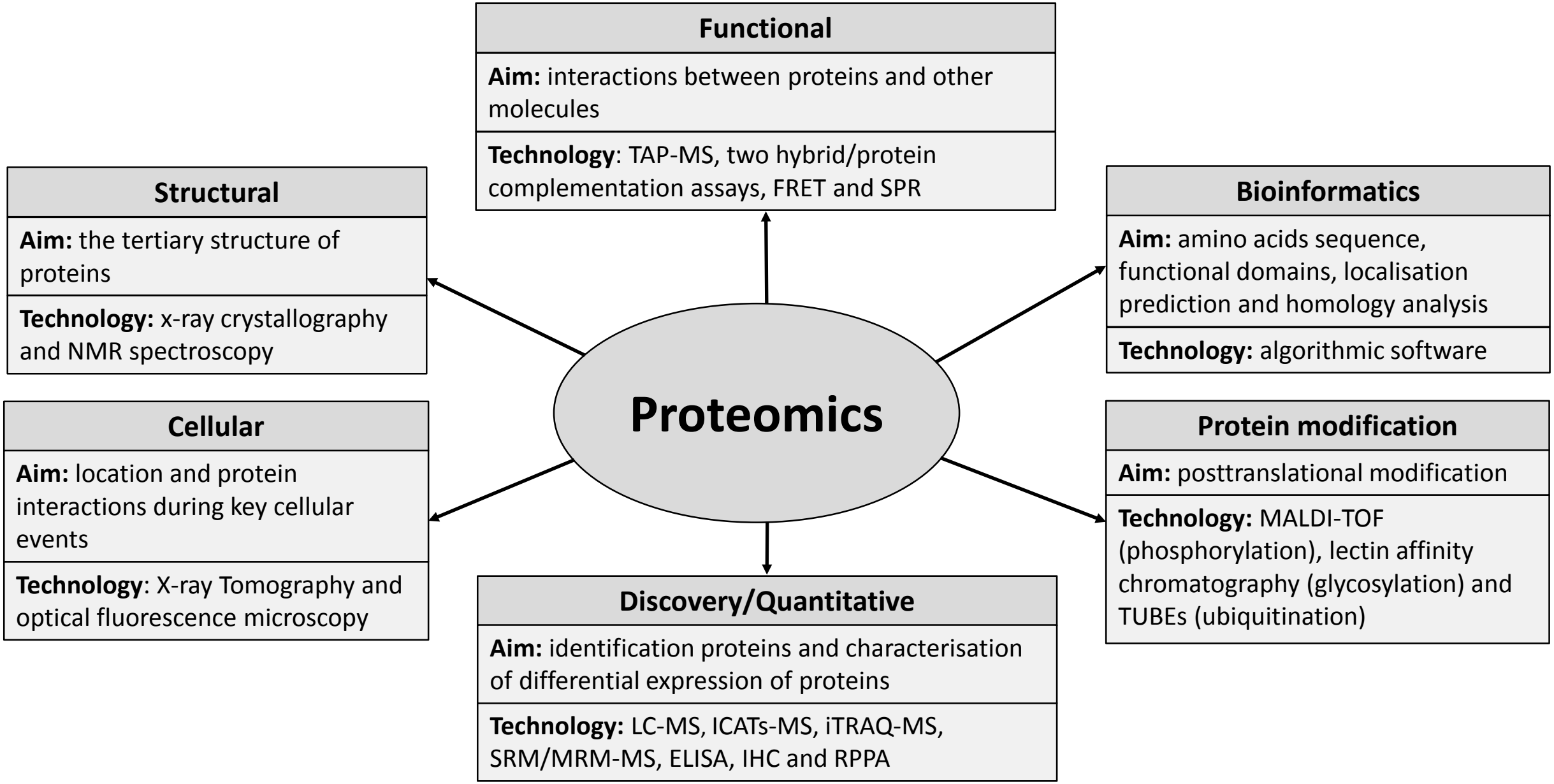
88. Fung KY, Nice E, Priebe I, et al. Colorectal cancer biomarkers: to be or not to be? cautionary tales from a road well travelled. World J Gastroenterol. 2014;20:888-898.

89. de Wit M, Fijneman RJ, Verheul HM, et al. Proteomics in colorectal cancer translational research: biomarker discovery for clinical applications. Clin Biochem. 2013;46:466-479.

**Figure legends**

Figure 1: Overview of proteomic technologies. Abbreviations: FRET, fluorescence resonance energy transfer; SPR, surface plasmon resonance; TAP-MS, tandem affinity purification-mass spectrometry; MALDI-TOF, matrix assisted laser desorption ionization time-of-flight; LC-MS, liquid chromatography–mass spectrometry; S/MRM-MS, single/multiple reaction monitoring tandem mass spectrometry; iTRAQ, isobaric tags for relative and absolute quantitation; ICATs, isotope-coded affinity tags; TUBEs, tandem-repeated ubiquitin-binding entities; RPPA, reverse phase protein array; IHC, immunohistochemistry; ELISA, enzyme-linked immunosorbent assay.

Figure 2: Overview of screening biomarkers and their main aims in CRC, types of biomarker, methods of detection and types of bio-specimen. Abbreviations: IHC, immunohistochemistry; FISH, fluorescence in situ hybridization; HPLC, high-performance liquid chromatography; MALDI-TOF, matrix assisted laser desorption ionization time-of-flight; LC-MS, liquid chromatography-mass spectrometry; S/MRM-MS, single/multiple reaction monitoring tandem-mass spectrometry; DIGE, difference gel electrophoresis; PCR, polymerase chain reaction; ELISA, enzyme-linked immunosorbent assay.


Functional
Aim: interactions between proteins and other molecules
Technology: TAP-MS, two hybrid/protein complementation assays, FRET and SPR
Structural
Aim: the tertiary structure of proteins
Technology: x-ray crystallography and NMR spectroscopy
Bioinformatics
Aim: amino acids sequence, functional domains, localisation prediction and homology analysis
Technology: algorithmic software
Proteomics
Cellular
Aim: location and protein interactions during key cellular events
Technology: X-ray Tomography and optical fluorescence microscopy
Protein modification
Aim: posttranslational modification
Technology: MALDI-TOF (phosphorylation), lectin affinity chromatography (glycosylation) and TUBEs (ubiquitination)
Discovery/Quantitative
Aim: identification proteins and characterisation of differential expression of proteins
Technology: LC-MS, ICATs-MS, iTRAQ-MS, SRM/MRM-MS, ELISA, IHC and RPPA

# Early CRC-Biomarkers

| Type | Detection method | Bio-specimen |
| --- | --- | --- |
| ►Proteins<br>►DNAs and RNAs<br>►Carbohydrates and Metabolites | ►Proteomic: DIGE, HPLC, LC-MS, MALDI-TOF, S/MRM-MS and antibody based (IHC, ELISA and protein microarrays)<br>►Genomic: PCR, FISH, mRNA expression profiling and whole genome sequencing | ►Blood, sera and plasma<br>►Tissue, faeces and urine<br>►*In vitro* tumour model or animal model |

**Aims**

►Screen healthy (high risk) individuals

►Detection of adenomas with high risk of malignancy

►Detection of early carcinoma (stage I and stage II)

**Supplementary Information**

**Methods S1.**

**Research criteria**

PubMed and Google Scholar searches were performed using key words “biomarker, colorectal cancer, early colorectal cancer, early diagnosis, proteomics, proteome, screening”, within the title, abstract and/or text. Only English language publications from 2014 onwards were selected (Table S1). However, to discuss the findings of some of these studies reference has been made pre 2014 studies. References identified in retrieved articles were further screened for potentially relevant studies. Only studies utilising clinical human samples in the discovery and/or validation phase were selected (the only exception to this was studies using organoid model, to highlight the potential of this new technology). The full texts of selected articles were reviewed, and a decision on their eligibility for inclusion was then made based on; biomarker targets, proteomic technologies, study design, analytical approach, validation process, limitations and potential clinical impact. Although there was no specific criterion for the size of patient cohorts, the focus was on studies with larger samples and better characterised .

**Table S1.** Summary of proteomic studies, biomarker targets and their patient cohorts.

| Target(s) | Discovery | | Validation | | Commentary | Ref |
|---|---|---|---|---|---|---|
| | Method(s) | Sample | Method(s) | Sample | | |
| General transcription factor IIB, EGF-like repeats and discoidin I-like domains 3, HCK proto-oncogene, pim-1 proto-oncogene, serine/threonine kinase 4 and tumour protein P53 | Targets were identified by previous studies using protein microarray-based methods | | Multiplex beads assay and ELISA | Sera: 135 CRC (stage I=35, stage II=25, stage III=46 and stage IV=29), 65 other cancer types, 14 inflammatory bowel diseases and 93 healthy controls | Illustrated the potential of autoantibodies as CRC markers, presented a useful assay for assessing biomarker panel and used a large and well-characterised patient cohort | [39] |
| Autoantibodies against tumour-associated antigens: inosine monophosphate dehydrogenase 2, MAGE family member A4, MDM2 proto-oncogene and tumour protein P53 | Multiplex serology, a fluorescent bead-based GST capture immunosorbent assay | Sera: 124 normal and 352 CRC (stage I=96, stage II=102, stage III=105 and stage IV=49) | Multiplex serology, a fluorescent bead-based GST capture immunosorbent assay | Sera: 49 CRC (high-grade dysplasia=4, stage I=18, stage II=5, stage III=19 and stage IV=3), 100 normal, 29 non-advanced adenomas and 99 advanced adenomas | Illustrated the potential of autoantibodies as CRC markers, presented a useful assay for assessing biomarker panel and used a large and well-characterised patient cohort | [40] |
| Alpha-1-acid glycoprotein 1, alpha-1 antitrypsin, amylase alpha 2b, clusterin, complement c9, enoyl-coa hydratase 1, ferritin light chain, gelsolin, osteopontin, selenium binding protein 1, seprase, spondin 2 and tissue inhibitor of metalloproteinases 1 | Targeted multiplex MRM-MS assay | Plasma: 69 healthy controls and 69 CRC cases (stage I=13, stage II=35, stage III=15 and stage IV=6) | Targeted multiplex MRM-MS assay | Plasma: 68 controls and 68 CRC cases (stage I=16, stage II=35, stage III=14 and stage IV=3) | Demonstrated the benefit of targeted MS as a validation technology for protein biomarker panel | [41] |

| | | | | | | |
|---|---|---|---|---|---|---|
| Caeruloplasmin, serum paraoxonase/arylesterase 1, serpin peptidase inhibitor clade A, leucine-rich alpha-2-glycoprotein and tissue inhibitor of metalloproteinases 1 | LC MS/MS<br><br>Targeted LC-MS (SRM) | Discovery: tissues, 16 CRC (stage I=10, stage II=2, stage III=2 and stage IV=2) with adjacent normal mucosa<br><br>Screening: plasma, 19 CRC (stage I=12, stage II=3, stage III=3 and stage IV=1) | Targeted LC-MS (SRM) | Training: plasma: 23 non-advanced adenomas, 11 hyperplastic polyps, 66 normal and 100 CRC (missing =3, stage I=32, stage II=26, stage III=31 and stage IV=8)<br><br>Validation: plasma: 4 advanced adenomas, 2 benign adenomas, 1 dysplastic polyp, 6 diverticular disease, 4 Crohn, 50 healthy and 202 CRC (stage I=43, stage II=58, stage III=49 and stage IV=52) | Demonstrated the potential of targeted MS as a validation tool and used a robust study design with patient cohorts reflecting different stages of CRC development | [42] |
| CRC detection panel: carcinoembryonic antigen, seprase, serpin A3, macrophage migration inhibitory factor, complement component 3, complement component 9, p-selectin glycoprotein ligand 1 and cathepsin D<br><br>Adenoma detection panel: cathepsin D, clusterin, growth differentiation factor 15 and serum amyloid A1 | ELISA | Plasma: 75 CRC (stage I=17, stage II=30, stage III=16 and stage IV=12), 75 advanced adenomas and 150 healthy controls | ELISA | Plasma: 75 CRC (stage I=17, stage II=21, stage III=18 and stage IV=19), 76 advanced adenomas and 151 healthy controls | Validated markers (identified by study above [41]) using clinically established assay (ELISA) | [45] |
| Serpin A1, serpin A3 and serpin C1 | iTRAQ-MS | Serum: 15 CRC (stage I=2, stage II=2, stage III=4 and stage IV=7) and 15 adenomas and 15 healthy controls | ELISA | Serum: 21 healthy controls and 19 CRC (stage I=2, stage II=5, stage III=5 and stage IV=7) | Used a representative cohort, however small number of samples was used especially for stage I CRC | [46] |
| Pyruvate kinase isoenzyme type M2, gamma enolase, serotonin and 14-3-3 family members | LC-MS/MS, UHPLC-MS and GC-MS | Plasma: 16 CRC (stage III=8 and stage IV=8) and 10 healthy controls | ELISA<br><br>IHC (14-3-3 epsilon) | Plasma: 40 CRC (10 for each stage), adenomas (n=20) and healthy controls (n=20)<br><br>24 CRC tissue cores with corresponding normal (tumour stage not provided) | Multiple technologies were used to evaluate and validate promising CRC biomarkers, however patient cohorts were small | [49] |

| | | | | | | |
|---|---|---|---|---|---|---|
| Microtubule-associated protein, RP/EB family, member 1 (MAPRE1) | LC/MS-MS | Plasma: 60 adenomas, 60 CRC (stage I= 11, stage II= 19, stage III= 21 and stage IV=9) and 60 healthy controls.<br>Mouse model and cell lines | Antibody array<br><br>IHC | Plasma: 60 adenomas, 60 CRC (stage I=11, stage II=19, stage III=21 and stage IV=9) and 60 healthy<br><br>20 normal colonic tissues, 10 adenomas, and 66 CRC (tumour stage not provided) | Interesting protein target with high potential, multiple technologies were used to evaluate and validate the results, although patient cohorts were small | [52] |
| **Tissues-Based markers** | | | | | | |
| Kininogen-1, transport protein Sec24C and olfactomedin-4 | LC-MS | Fixed tissues of 36 early CRC (stage pT1N0=16 and stage pT2N0=20), 20 controls and 20 diverticulitis inflammatory | IHC | Fixed tissues of 20 healthy controls, 20 diverticulitis controls, 20 low grade adenomas, 20 high grade adenomas and 112 CRC (high-grade dysplasia=12, stage pT1N0=20, stage pT2N0=20, stage pT3=20 and stage pT4=20) | Kininogen is a promising target, highlighted the applicability of proteomics on fixed tissue and used patient cohorts reflecting different stages of CRC | [58] |
| S100 calcium-binding protein A9 (S100A9), annexin A3, nicotinamide phosphoribosyltransferase, carboxylesterase 2 and calcium activated chloride channel A1 | iTRAQ-LC-MS<br><br>Gene microarray, Affymetrix U133plus2.0 | 24 pairs of fresh-frozen CRC (stage I=6, stage II=6, stage III=6 and stage IV=6) and normal tissues<br><br>52 pairs of fresh-frozen CRC (stage I=4, stage II=17, stage III=27 and stage IV=4) and normal tissues | ELISA<br><br>IHC | Serum:76 healthy controls and 100 CRC (stage I=12, stage II=38, stage III=25 and stage IV=25)<br><br>18 pairs of CRC (stage I=2, stage II=6, stage III=9 and stage IV=1) and normal tissues | Used multiple technologies and different sample types, but number of stage I CRCs was small | [63] |

| | | | | | | |
|---|---|---|---|---|---|---|
| Lysyl oxidase-like 2 | iTRAQ –LC-MS | Cell lines: SW480, SW620, KM12C, and KM12SM<br><br>Tissues: 12 matched colon cancer (stage: II=5 and stage III=7) | PCR<br><br>IHC<br><br>Gene expression database | Tissues: 70 colon cancer (stage I=8, stage II=26, stage III=22 and IV=14)<br><br>Tissues: 121 colon cancer (stage I=31, stage II=53, stage III=9 and stage IV=28)<br><br>Tissues: three external cohorts: 232 colon cancer (tumour stage was not provided), 90 stage II colon cancers and 21 stage III colon cancers) | Presented a robust model for assessing fibroblast-associated proteins using multiple technologies | [69] |
| **Faecal and urine-based biomarkers** | | | | | | |
| Faecal pyruvate kinase isoenzyme type M2 (M2-PK) | Meta-analysis of eight clinical studies including 2,654 participants | | | | Highlighted the potential of faecal M2-PK as a screening marker for early CRC | [70] |
| Prostaglandin $E_2$ ($PGE_2$) metabolite: 11 alpha-hydroxy,9,15-dioxo-2,3,4,5-tetranor-prostane-1,20-dioic acid | LC/MS | Urine: control=420, low risk adenoma=130 and high-risk adenoma=290 | NA | NA | One of the few proteomic studies that used urine samples from a large and well characterised cohort | [72] |
| **Colorectal tumour models** | | | | | | |
| Synaptotagmin 7, ras-related protein rab-27b, coagulation factor iii, chloride intracellular channel 5, kin of IRRE like, dual oxidase 2, carcinoembryonic antigen related cell adhesion molecule 7, mucin 12, v-set and immunoglobulin domain containing 2, microtubule associated protein 2, mucin 4, calpain 8, beta-1,3-galactosyltransferase 5, macrophage stimulating 1 receptor, myosin 1C , shroom family member 3, AHNAK, plastin 1, heparan sulfate proteoglycan 2, filamin binding lim protein 1 and dedicator of cytokinesis 5 and gelsolin | LC-MS | 14 organoids: 7 CRC (stage not provided) and 7 healthy controls | Affymetrix Human Gene 2.0 ST arrays | 14 organoids: 7 CRC (stage not provided) and 7 healthy controls | Illustrated the benefits of organoids as a CRC model for proteomic analysis and biomarker discovery in CRC | [76] |

| | | | | | | |
|---|---|---|---|---|---|---|
| Protein tyrosine pseudokinase PTK7 | Quantitative LC-MS | 3 organoids cultured in stem cell supporting medium versus 3 organoids cultured in differentiation supporting medium (epidermal growth factor and Noggin [EN])<br><br>CRC patients (stage not provided) | IHC<br><br>Quantitative real-time PCR | 1 normal human colonic mucosa<br><br>2 organoids cultured in stem cell medium versus 2 organoids cultured in differentiation medium (epidermal growth factor and Noggin [EN]) | Highlighted the benefits of organoids in biomarker discovery | [77] |
| Chymotrypsin-like elastase 1 (CELA1), chymotrypsin-like elastase 2A (CEL2A), chymotrypsinogen B (CTRB1), trypsin 2 (TRY2), trypsin 4 (TRY4) and chymotrypsin like (CTRL) | iTRAQ-MS | $APC^{min/+}$ mice and wild type mice of 8, 13, 18 and 22 weeks old | Targeted MRM-MS<br><br>IHC (CELA1 and CTRL) | Sera 30 CRC (tumour stage: T2=4, T3=15 and T4=11, nodal stage: N0=15 and N1-2=15) and 30 healthy individuals<br><br>80 pairs of CRC tissues (tumour stage: T1=2, T2=12 and T3=39 and T4=27, nodal stage: N0=38 and N1-2=42) and corresponding normal tissues | Used multiple technologies on different types of sample and presented a proteomic model for analysing tumour interstitial fluids | [79] |

References are listed in the main manuscript. Abbreviations: IHC, immunohistochemistry; ELISA, enzyme-linked immunosorbent assay; UHPLC-MS, Ultra high-performance liquid chromatography tandem mass spectrometry; LC-MS, liquid chromatography–mass spectrometry; iTRAQ, isobaric tags for relative and absolute quantitation; GC-MS, Gas Chromatography Mass Spectrometry; Targeted LC-MS (SRM/MRM), Targeted mass spectrometry based on selected/multiple reaction monitoring.