Integrative analysis of the colorectal cancer proteome: potential clinical impact

Abdo Alnabulsi\textsuperscript{1,2} and Graeme I Murray\textsuperscript{1}

\textsuperscript{1}Pathology, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK. \textsuperscript{2}Vertebrate Antibodies, Zoology Building, Tillydrone Avenue, 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

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 one of the commonest types of cancer that affects a significant proportion of the population and is a major contributor to cancer related mortality. The relatively poor survival rate of CRC could be improved through the identification of clinically useful biomarkers.

**Area covered:** This review highlights the need for biomarkers and discusses recent proteomics discoveries in the aspects of CRC clinical practice including diagnosis, prognosis, therapy, screening, and molecular pathological epidemiology (MPE). Studies have been evaluated in relation to biomarker target, methodology, sample selection, limitations, and potential impact. Finally, the progress in proteomic approaches is briefly discussed, and the main difficulties facing the translation of proteomics biomarkers into the clinical practice are highlighted.

**Expert opinion:** The establishment of specific guidelines, best practice recommendations and the improvement in proteomic strategies will significantly improve the prospects for developing clinically useful biomarkers.

Keywords: biomarkers, colorectal cancer, diagnosis, prognosis, proteomics, predictive screening
1. Introduction

1.1 Colorectal cancer background

Colorectal cancer (CRC) is one of the commonest types of cancer and a major cause of cancer related death [1]. The survival rate is still relatively poor particularly for patients presenting with distant metastases [2]. Established primary CRC can generally be diagnosed based on the histopathological characteristics of tissue biopsies obtained at colonoscopy [3]. However, it can be more difficult to diagnose early CRC or CRC that presents as metastatic disease [4, 5]. Screening programmes for CRC using either colonoscopy or faecal occult blood testing have shown that they may reduce the mortality rate from CRC [6, 7]. Nevertheless, current screening methods suffer from several drawbacks including lack of sensitivity and poor participation rates that impede their potential benefits [8].

Prognosis using the current staging system which is based on the histopathological examination of resected CRC does not necessarily reflect the biological heterogeneity of CRC and thus patients with the same tumour stage often have variation in clinical outcome [9]. This staging system is also the main method whereby therapeutic options are determined, yet patients with the same stage often respond differently to the same treatment [10, 11]. Therefore, reliable and easily measurable biomarkers are urgently required to assist clinicians to overcome current difficulties in clinical practice (Figure 1).

1.2. Proteomics and genomics perspective on biomarker discoveries

A biomarker is defined as a “characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” [12]. In recent years, there have been a noticeable shift in the biomarker literature towards genomics and transcriptomics technologies, which
undoubtedly have increased our understanding of cancer biology and have led to a variety of biomarker discoveries [13, 14]. However, genotype is not necessarily reflected in phenotype because of the influence of a range of factors including epigenetic changes, alternative splicing, non-coding RNAs (including microRNAs), post translation modifications and protein-protein interactions [15, 16]. Moreover, the analysis of large genomic data sets is a challenging task that demands complex and sophisticated analytical tools and software [17, 18]. This can make the interpretation of data generated using different samples, array types, sequencing platforms difficult [19, 20].

However, proteomics can assess proteins which directly reflect a pathogenic phenotype and potentially is more likely to provide accurate information on disease state and clinical outcome [15, 21, 22]. Current proteomics technologies are able to assess protein modifications such as post-translation modifications and sequence variants [23]. Advances in proteomic technologies have enabled more accurate and in-depth identification of individual proteins within complex protein mixtures [24, 25]. Furthermore, improvements in protein extraction and separation have made proteomics analysis suitable for use on formalin fixed paraffin embedded (FFPE) tissues, thus potentially exploiting a larger number of archival samples necessary for protein biomarker validation [26]. However, current proteomics technologies are still potentially lacking in terms of their ability to detect very low abundance proteins [27].

2. Biomarkers for CRC diagnosis

2.1. Diagnosis of malignant polyps

Currently, the risk of malignant transformation of polyps is determined through the pathological analysis of polyp characteristics such as increasing size, degree of epithelial cell
dysplasia and greater “villousness” [28]. Patient management guidelines (e.g. time interval between surveillance colonoscopies) are based on the identification of these histopathological characteristics [29]. Since only less than 5% of polyps proceed to malignancy, it is imperative to identify novel protein biomarkers to further assist in identifying high-risk adenomas and therefore allow rational use of endoscopic resources [5]. Proteomic studies have identified several candidates proteins associated with the malignant transformation of adenomatous polyps (Table 1).

Kininogen-1 was identified as a marker for the diagnosis and/or screening of the malignant transformation of adenomas [30]. This study analysed the serum samples of 110 participants using matrix assisted laser desorption ionization time-of-flight (MALDI-TOF/TOF-MS). Kininogen-1 levels were significantly lower in normal mucosa compared with CRC and advanced colorectal adenoma, which is consistent with a previous finding [36]. Nevertheless, the exact role of kininogen-1 in CRC pathogenesis remains unclear. Moreover, it is not clear whether the protein levels vary between different types of adenomas. Therefore, there is a need for further experimental validation of the results especially in the presence of contradictory results regarding the levels of kininogen-1 detected in adenoma and carcinoma [37].

In another study, the analysis of plasma samples using MALDI-TOF/TOF-MS has enabled the identification of a peptide signature to monitor and predict the malignant transformation of polyps in familial adenomatous polyposis (FAP) [31]. This study was the first to show that peptide profiling can be used to monitor CRC development in FAP patients. However, a significant concern is the specificity of the peptide signature for CRC because some of the identified peptides are derived from proteins (e.g. complement C3 and C4) involved in inflammation. Moreover, enzyme linked immunosorbent assay (ELISA) validation of complement proteins revealed contradictory results to the proteomic findings.
This highlights the need for further evaluation of the results in an independent and much larger cohort since the ELISA was performed only on a small cohort (10 FAP, 8 adenoma, and 36 CRC). The contradictory ELISA results were attributed to the increased activity of proteases rather than the circulating level of precursor proteins. The inability to use ELISA for assessing and monitoring the peptide signature makes it much more difficult to validate the findings and may compromise the potential clinical utility of this proposed biomarker signature. For the detection of peptides using ELISA, it may be possible to use monoclonal antibodies produced against short synthetic peptides [38].

A diagnostic portfolio of proteins that could be used to distinguish adenomas from CRC or healthy controls was identified using MALDI-TOF/TOF-MS [32]. This research investigated plasma samples from healthy controls, and from patients with colorectal adenomas or invasive disease. Blood-based biomarkers which can accurately diagnose potentially malignant adenomas could have great clinical utility. However, the proteins identified in this study included inflammatory cytokines [39]. Moreover, the validation was only conducted on a small cohort (30 adenomas and 30 carcinomas). Hence, additional studies are required before these proteins can be considered as potential diagnostic markers.

Proteomic technologies can also be utilised to assess urine for the purpose of identifying biomarkers associated with high-risk adenomas [33]. This study has found that urinary levels of prostaglandin metabolites (PGE-M) measured using liquid chromatography–mass spectrometry (LC/MS) are associated with high-risk adenomas. The study cohort contained a relatively large number of controls and adenomas that were from a prospective cohort. There are limitations in the design and the composition of the cohort used in this study. Firstly, the cohort only included females. Secondly, and more importantly, there were no colorectal carcinomas in the samples analysed. Nevertheless, the findings are consistent with other proteomics studies that have measured urinary PGE-M using the same method [40,
The study by Shrubsole and colleagues [40] used 224 cases with at least one advanced adenoma, 152 small tubular adenomas, 300 single small tubular adenoma, and 364 controls. Whereas the study by Johnson et al. [41] assessed PGE-M in 58 CRC, 70 polyps, 28 Crohn’s disease, and 72 healthy controls. In addition, Nakanishi et al. [42] found that the inhibition of PGE₂ production suppresses intestinal carcinogenesis in an APC-mutant mouse model. A significant concern is that PGE-M are also involved in several inflammatory pathways and also other malignancies, hence there is a need to assess the specificity of urinary PGE-M in CRC versus other cancers and inflammatory diseases (e.g. Crohn’s disease, ulcerative colitis) [43]. It is therefore difficult to interpret the association between PGE-M levels and development of CRC without further study of well-defined cohorts.

In another study, nuclear magnetic resonance (NMR) spectrum analysis of urinary samples of 988 high-risk individuals who required colonoscopy identified metabolomics signatures associated with CRC [34]. Using an algorithmic classifier of metabolic signature (4 metabolites and 4 clinical questions), the study showed it was possible to predict individuals who required colonoscopy with an accuracy better than faecal occult blood test. This is potentially a useful addition to the clinical practice because it could ensure patients avoid colonoscopy. Additionally, the learning algorithmic classifier may represent a novel method of transforming complex proteomics data into clinically relevant parameters. However, the study classified both participants with hyperplastic polyps or adenomas into one group requiring colonoscopic follow-up even though the risk of progression to CRC significantly differs between the two groups. The risk of malignant transformation should be the main determinant of who needs colonoscopy. Furthermore, it is difficult to interpret these findings without further study since only two carcinomas were included in the study. The lack of early CRC from the samples may results in a failure to identify important changes in proteins associated with the early stages of malignant transformation.
2.2. Diagnosis of metastatic CRC

The identification of the primary origin of an unknown metastatic tumour is still challenging in spite of the availability of a variety of diagnostic methods including histopathology, molecular analysis, imaging, and endoscopy [4]. The majority of cancers of unknown primary are metastatic adenocarcinomas of which around 7% are of colo-rectal origin [44]. Failure to identify the primary origin of a tumour is a significant problem since the clinical management of patients particularly the selection of appropriate treatment regimens depends on the identification of the specific cancer type. Histopathological assessment using a combination of the immunohistochemical markers cytokeratin 20, cytokeratin 7 and CDX2 is often used to identify CRC, although the typical cytokeratin 20+/CDX2+/cytokeratin 7-ve phenotype is not expressed by all colorectal carcinomas [45]. Recent studies have shown that the assessment of a combination of Stabilin-2 (STAB2) with cytokeratin 7 and cytokeratin 20 can provide a highly sensitive and specific test for CRC diagnosis [46, 47]. Both studies used immunohistochemistry (IHC) to evaluate STAB2, cytokeratin 20 and cytokeratin 7 expression in large cohorts (n = 840 and n = 2696) which included CRC, benign tumours, normal tissues, and other common malignancies. While few studies have focused on the identification of new markers that help differentiate metastatic CRC from other malignancies there is a clear requirement for such biomarkers.

3. Biomarkers for CRC prognosis

3.1. The need for prognostic markers in the clinical practice

A prognostic biomarker can be defined as a (biological) variable that provides prospective information on patient outcome which is complementary to the data obtained by
the pathologist from histopathology and on which therapeutic decisions can be guided [48]. There is unjustifiable scepticism towards prognostic marker studies as they are generally perceived as unnecessary in CRC since histopathological examination of the resected colorectal cancers provide key prognostic information (tumour stage, lymph node stage, extramural venous invasion). However, the clinical outcome can vary considerably between patients who are diagnosed with the same tumour stage especially for patients with stage II and III CRC [3]. Thus, the identification of protein biomarkers on both biopsies of CRC and surgically resected CRC, which reflect the heterogeneity of CRC, will help in providing accurate prediction of the clinical outcome of patients.

3.2. Recent proteomics discoveries

Large numbers of potential prognostic markers have been identified using proteomic-based approaches (supplementary information Table S1). For example, a combination of Nano LC-MS and gene expression analysis of stage IV CRC patients (n = 46) was used to identify metastasis associated markers [49]. Maspin was found to vary between the two groups of patients divided based on time to recurrence (Table S1). Immunohistochemical analysis of maspin expression in a tissue microarray containing 419 stage II and III CRC samples indicated it was an independent prognostic factor of time to recurrence and disease specific survival in stage III CRC. The finding was validated using three transcriptomics data sets (75 stage II, 78 stage III, and 53 stage IV). Consistent with this study, high maspin expression was linked to increased apoptosis resistance in a colon cancer cell line (HCT-116RC) [50]. The discovery of maspin highlights the fact that some proteins might have a stage-specific function and thus have stage-related expression profiles that could be detected through proteomics. Stage-specific markers carry important prognostic information that
could complement or even replace the current staging system when surgical specimens are not available. Several other studies have demonstrated that it is possible to identify protein markers associated with tumour stage using comparative proteomics (studies 1, 2, and 9 described in Table S1). However, it is worth noting that the size of the samples used in those studies were small and therefore further validation using larger and independent cohorts is required.

Stomatin-like 2 (STOML2) protein was identified using high-performance LC-MS analysis of membrane proteome in 28 pairs of normal and CRC tissues [51]. The subsequent assessment of the protein expression by IHC showed a strong association between STOML2 and disease-specific survival. High expression of STOML2 was associated with decreased CRC-related survival. Furthermore, the plasma levels of STOML2 as measured by ELISA were higher in early stage CRC compared with healthy individuals, which suggest STOML2 could potentially be used as a screening marker. The main drawback of this study is the small numbers of CRC samples included in the proteomics analysis (n = 28), ELISA (n = 70) and IHC (n = 205). Additional validation using a large and independent cohort is still needed as well as further investigation of the role of STOML2 in CRC pathogenesis.

Another interesting study has revealed a prognostic protein signature using tailored computational analysis of proteomics data generated via a combination of LC-MS and targeted LC-MS (SRM) assessment of plasma samples [52]. The protein signature (major histocompatibility complex class I-A, complement factor H, CD44, protein tyrosine phosphatase, receptor type J, haptoglobin, and cadherin 5 type 2) was associated with different prognostic parameters and could stratify patients to distinct prognostic subgroups. The study also used data from three external transcriptomics cohorts for additional validation. The findings of this study are encouraging and may have considerable implication on the management of CRC patients because, unlike the key pathological prognostic factors which
require examination of a resected CRC specimen, this protein signature can be evaluated
noninvasively in the plasma. Nevertheless, additional assessment of the findings on a larger
cohort is needed because only 202 CRC samples were used. Moreover, there may be a need
to evaluate whether this protein signature is detectable by ELISA since this technology is
generally easily implemented in laboratories where quality assurances and best practice
guidelines are in place. Other studies which have also utilised proteomics analysis of plasma
for prognostic marker identification are described in Table S1.

The analysis of preclinical models (cell lines and xenograft tumours) using 2-D
difference gel electrophoresis (DIGE) and MALDI-TOF/TOF MS followed by validation
with human samples provides another approach to the identification of prognostic markers
[53]. Comparative proteome analysis found stathmin 1 (STMN1) levels to be lower in colon
cancer cell line (HCT-116) compared to its metastatic derivative E1. Both knockdown and
overexpression of STMN1 in HCT-116 and E1 showed it was associated with significant
changes in cell migration, invasion, adhesion, and colony formation. This study also
performed IHC staining on a tissue microarray containing 324 primary CRC. The expression
of STMN1 was higher in CRC compared to the adjacent normal mucosa, and increasing
intensity of expression was associated with poorer CRC specific survival. These results are
consistent with a recent study which showed that the silencing of STMN1 inhibited
metastasis in (E1) and (HCT116) colon cancer cell lines [54]. However, in contrast to the
finding of that study, the expression of STMN1 assessed by IHC using 546 CRC cases from
two independent cohorts showed that the overexpression of STMN1 was associated with
improved survival [55]. Therefore, although STMN1 has shown a promising potential as a
prognostic marker for CRC, there is a need for further research of STMN1 in CRC.

Although some studies have used large and well-characterised cohorts for the
validation of their proteomics results [56, 57], the majority of studies have used relatively
small cohorts. Another limitation is the insufficient reporting in a variety of aspects including the collection and processing procedure of specimens, inclusion criteria of patients, clinicopathological characteristics of the cohort and clearly defined endpoints. Many biomarker studies still suffer from lack of adherence to the Reporting Recommendations for Tumour Marker (REMARK) guidelines [58]. Compliance with the REMARK guidelines should help to standardise and improve the quality of biomarker studies [59].

4. Biomarkers for predicting the outcome of CRC therapy

4.1. The need for predictive markers in the clinical practice

A predictive biomarker is defined as a variable that indicates the outcome of a specific type of therapy and therefore aids in making treatment decisions [60]. Predictive markers are needed in CRC management because the benefit of neoadjuvant and/or adjuvant therapy is not clear for a significant proportion of patients [10, 11]. The increasing range of therapeutic options have further highlighted the need for predictive biomarkers. One of the few predictive markers to be in current clinical practice is the identification of KRAS mutations as KRAS mutant tumours do not respond to anti-epidermal growth factor receptor drugs [61]. In addition, the assessment of mismatch repair proteins seems to offer valuable information on the potential benefit of fluorouracil based adjuvant therapy and immune checkpoint inhibitors [62, 63].

4.2. Recent proteomics studies

Several research groups have used proteomics analysis to identify putative predictive markers for CRC (Table 2). Using a combination of 2D-DIGE and LC-MS/MS to assess
serum samples from a group of patients with CRC who had received chemotherapy and bevacizumab, sixty-four differentially expressed proteins were identified between responders and non-responders [64]. The study also used ELISA and IHC to validate three proteins (apolipoprotein E, angiotensinogen and vitamin D binding protein) which were significantly associated with overall survival and or progression free survival in metastatic CRC patients treated with chemotherapy and bevacizumab. This could be useful as less than 50% of patients showed a response to this therapy [69]. Nevertheless, the number of CRC samples in the validation cohort was relatively small (ELISA: n = 68 and IHC: n = 95), therefore further validation of the results is still required.

A panel of 32 proteins associated with CRC was identified using isobaric tags for relative and absolute quantitation (iTRAQ –LC-MS) analysis of cancer-associated fibroblasts obtained from colon cancer and normal tissue [65]. This study presented strong and well-designed discovery model whereby proteins derived from colon-associated fibroblast can be assessed for biomarker discoveries. High expression of lysyl oxidase-like 2 (LOXL2) was associated with poor overall survival and high recurrence, and demonstrated predictive value for adjuvant therapy in stage II colon cancer. The results were validated on a number of independent cohorts using different methods (IHC, gene expression profiling and polymerase chain reaction (PCR)). Still, the number of colon cancer cases in the validation cohorts (IHC: n = 121 and PCR: n= 70) was relatively small and hence further validation is required. A previous study found LOX, a family member and paralog of LOXL2, to play an important role in promoting CRC angiogenesis using in in vitro (SW480 and SW620 cell lines) and mouse models (LS174T human CRC cell lines grown as subcutaneous tumours in nude mice) [70]. The results of the preclinical models were further validated by IHC (on a CRC tissues microarray (n=515)) which showed the expression of LOX correlated with VEGF expression and blood vessel formation in patients [70].
Dasatinib, an inhibitor of Src tyrosine kinases, is currently being evaluated for use in CRC. However, this therapy is expensive and can cause side effects. Therefore, there is a need for predictive biomarkers that can accurately select patients based on the potential response to this drug. In a recent study, delta-type protein kinase C (PKCδ) was identified as predictive marker for dasatinib in CRC [67]. Shotgun phosphotyrosine proteomics was used to obtain a global view of tyrosine phosphorylation in HCT-116 colon cancer cell lines and HCT-116 xenograft tumour. The results showed that the measurement of PKCδ pY313 as a promising method for assessing the response to dasatinib.

However, there is still an apparent lack of research focused on the identification of predictive markers. This can be attributed in part to the lack of readily available, large, well-characterised cohorts.

5. Biomarkers for CRC screening

5.1. The need for screening markers in clinical practice

The five-year survival rate of CRC patients is significantly worse for those patients diagnosed with metastatic disease compared with early stage disease [1]. Considering CRC is often asymptomatic at early stages of development, sensitive screening methods may reduce CRC associated mortality through early diagnosis when treatment is more effective. Flexible sigmoidoscopy and faecal occult blood test based screening programmes have shown some success in reducing the mortality rate of CRC [6, 7]. However, faecal occult blood tests generally suffer from a lack of sensitivity and a significant false positive rate, while endoscopic examination of the colon is invasive, expensive and involves a degree of risk [8]. Therefore, there is a need for accurate, cost effective, reliable, and non-invasive
biomarkers. Blood-based markers are ideal for screening because samples can be obtained with minimal trauma.

The plasma levels of adiponectin measured using ELISA seem to be inversely associated with CRC risk in men in a large prospective cohort (n=616 CRC and n=1205 controls) [71]. Further validation studies are needed before adiponectin can be used as screening marker. MicroRNAs, cell-free DNA and circulating tumour cells are available in the peripheral blood and they have shown encouraging results as useful screening markers for CRC, however further optimisation is still required [3, 72].

5.2. Recent proteomics studies

Proteomics studies enabled the identification of large number of proteins that may potentially be used as screening biomarkers for CRC (supplementary material Table S2). The use of a multiple autoantibody-based assay as a screening tool for CRC has shown promising results [73, 74]. In a recent study, 64 autoantibodies were assessed using multiplex serology assay, and a panel of four autoantibodies showed combined strong diagnostic ability in detecting early CRC [75]. The study conducted following a robust approach using technology that allows simultaneous assessment of a large number of potential markers in blood samples selected from the target screening population. Nonetheless, the relatively small cohort used in validation (49 CRC, 29 non-advanced adenoma, and 99 advanced adenoma) and the dissimilarities in the clinico-pathological characteristics between cohorts used for training and validation are limitations which could influence the outcome of this study. Moreover, there are concerns regarding the determination of the appropriate cut-off values.

A panel of proteins including APC-binding protein EB1 (MAPRE1) were elevated in CRC compared to healthy controls in a study that used LC-MS analysis of plasma samples
and CRC cell lines [76]. The study used one cohort (90 CRC) for discovery phase and two independent cohorts for ELISA validation. The samples were collected 17 months prior to CRC diagnosis in the discovery cohort, 7 months prior to diagnosis in one validation cohort and from newly diagnosed patients in the second validation cohort. The inclusion of prior-diagnosis samples can help in detecting important protein changes that happen very early in the process of CRC development. However, one validation cohort included only 32 CRC (with only four stage I cases) and the second cohort included 58 CRC without providing further characteristics of the samples. Furthermore, the study used two cohorts that consisted entirely of women while the third cohort was composed of both women and men. Therefore, further assessment of the results with a large cohort in a well-designed study is imperative.

In another study, the diagnostic potential of MAPRE1 in CRC was assessed using a combination of LC-MS and antibody array analysis of plasma samples, followed by IHC validation on fixed tissue samples [77]. The level of MARPE1 was higher in adenoma and CRC compared to normal healthy samples, which is consistent with the previous study [76]. In this study, the combination of MAPRE1 with carcinoembryonic antigen and adenylate kinase 1 has shown promising results in differentiating adenoma and early CRC, respectively, from healthy controls. Nevertheless, the relatively small number of samples (antibody array: 60 adenomas and 60 CRC/ IHC: 10 adenomas and 66 CRC) makes it difficult to determine the clinical usefulness of this marker combination. MAPRE1 knockdown in APC mutant (HT-29) and APC wildtype (HCT-116) showed an anti-proliferative effect which maybe dependent on APC status [78]. This indicates the importance of further research of the role of MAPRE1 in CRC and the need for further validation.

A diagnostic protein signature (ceruloplasmin, serum paraoxonase/arylesterase 1, serpin peptidase inhibitor, clade A, leucine-rich alpha-2-glycoprotein, and tissue inhibitor of metalloproteinases 1) was identified using proteomics and computational analysis [79]. The
study used robust methodology utilising LC-MS for discovery and targeted LC-MS for validation on a cohort that reflected different stages of CRC development and included other relevant diseases. Comprehensive analysis of the data revealed a five-protein panel with 72% diagnostic accuracy. The protein signature could be utilised in clinical practice as an independent screening test or in combination with existing diagnostic tests. However, there is a need for further validation using larger cohorts and it may be necessary to investigate the finding using alternative technique such as ELISA.

Proteomics analysis of blood proteome is not the only approach exploited to identify screening markers. Other studies have shown promising results using proteomics to assess urine, faecal, tissues and cell lines (Table S2). Follow up validation is required in larger cohorts for all these biomarkers.

6. The role of proteomics in molecular pathological epidemiology (MPE)

The rise in the incidence of CRC in developed countries have been linked to classic epidemiological factors such as diet, physical exercise, smoking, alcohol intake and an ageing population and these factors can be associated with specific molecular abnormalities [80]. Such epidemiological factors can be integrated by MPE to provide a more comprehensive understanding of CRC [81]. The phenotype of a disease can be better defined within the paradigm of MPE, which interpret specific molecular signatures within the context of recognised aetiological factors [82, 83]. For example, both CpG island methylator phenotype and microsatellite status in CRC can be linked to a variety of aetiological factors [84, 85]. Therefore, proteomics can provide another dimension for MPE. The challenges, opportunities, and recommendations of this multidisciplinary approach have been recently discussed at the second international MPE meeting [86].
7. Progress and difficulties

Potentially significant limitations observed in many of proteomics studies are the sample size, annotation and the composition of the cohorts. For example, there is a lack of early stage CRC in samples used for investigating potential screening markers. The presence of advanced CRC instead may distort the findings because changes in the profile of plasma proteome are likely to be greater than in early invasive lesions. Moreover, there is a lack of detailed clinicopathological characteristics of many cohorts (e.g., tumour differentiation, presence of extramural venous invasion, stage).

Another potential problem is the low levels of individual protein markers in early tumours which might not be reliably detected in serum with current technologies since it can be difficult to detect low abundance proteins especially in complex protein mixtures [87]. New strategies such as enrichment technologies (e.g. enrich for N- or C-terminal peptides), labelling approaches (e.g. neutron encoding (NeuCode)) may help overcome the inconsistency and lack of sensitivity of unlabelled MS, particularly when dealing with post-translation modifications [88]. Targeted MS is also gaining popularity because it is highly specific, accurate, and even applicable when there are problems with the antibodies [89]. However, there are difficulties when targeting several biomarkers simultaneously in multiple samples, and the method requires exhaustive and challenging optimisation process. Some of the challenges might potentially be addressed using a wider MS/MS window termed Sequential Window Acquisition of all THeoretical Mass Spectra (SWATH) strategy [90]. Finally, new approaches that are increasingly used include immunocapture strategies such as reverse phase protein microarray and immunocapture coupled to mass spectrometry [91]. Some of the advantages using such methods include the ability to measure multiple targets with highly sensitivity, requiring only small volume of serum or plasma, and with no need for
albumin depletion. However, the main weakness of this method is the dependency on the quality of the antibodies.

8. Expert commentary

“Precision medicine refers to the tailoring of medical treatment to the individual characteristics of each patient” [92]. The ongoing improvements in proteomic technologies should enable a comprehensive profiling of the proteome to provide a platform from which specific biomarkers necessary for precision medicine can be identified. In colorectal cancer there is still an urgent need for sensitive, reliable, and cost-effective biomarkers to complement the current methods of diagnosis, prognosis, therapy determination, and screening. Proteomics studies have generated a large number of potentially useful biomarkers. However, no protein biomarkers appear to have been successfully translated into clinical practice. This is attributed in part to the lack of reproducibility of results and the limitations of the validation studies. The reproducibility of proteomics studies is compromised by deficiencies in the studies design; small sample size, variations in the sample preparation and storage protocols, and complexities of data analysis and interpretation [93]. The lack of standardisation between different laboratories regarding quality assurance in the analytical techniques makes the results difficult to replicate and interpret [94]. The lack of reproducibility of the results could be minimised if the studies adopted strong experimental design and adhered to best practice guidelines. In addition, the introduction of automated quality control might significantly improve the reproducibility of proteomics results [95].

Lack of follow-up validation studies and deficiencies in validation assays also contribute to the lack of biomarkers success [96]. The assessment of biomarkers studies and their subsequent validation can seriously be hindered by the lack of large collaborative
projects, shortage of well-characterised samples, inconsistency of proteomics results, and insufficient reporting. Consequently, it makes it difficult to interpret, analyse, and validate the findings. The characteristics of samples should be carefully selected, justified, and clearly stated because they effect the results of biomarker studies. Finally, the validation methods (often antibody based) also suffer from absence of standardisation, absence of reliable antibodies, lack of best practice and quality controls [97, 98]. Rigorous and standardised characterisation process is needed to validate the antibodies used for immunoassays [99].

9. Five-year view

Further advancement in proteomics technologies will result in a more accurate assessment of plasma, serum, tissues, urine, saliva, and faeces proteome. Moreover, the introduction of sophisticated computational software should lead to improved and consistent data generation and analysis. There will also be significant improvement in study design, quality of samples, quality of antibodies, and adherence to best practice guidelines. It is also expected that there will be more collaborative projects with pooling of a wide range of expertise and resources. Thus, more biomarker targets will be identified using proteomics but their potential impact on the clinical practice is largely determined by the amount of progress made in addressing current limitations.

10. Key issues

- CRC is a common cancer with significant mortality.
- Biomarkers offer a solution to some of the current problems in CRC clinical management.
- Major advancements in proteomics coupled with innovative computational analysis have allowed the analysis of complex protein samples.
- Large numbers of potential CRC biomarkers have already been identified for use in diagnosis, prognosis, therapy determination, and screening.
• The majority of proteomics studies have focused on prognosis and screening.
• The failure of biomarkers is the consequence of three main factors; lack of validation, limitations in proteomics technologies and methodology deficiencies.
• More biomarkers will be identified considering the ongoing advancement in proteomics strategies and computational analysis.
• Addressing the current limitations will give future biomarkers discoveries a higher chance of success.

References

Papers of special note have been highlighted as:

* of interest
** of considerable interest


*Excellent paper which compared large number of samples collected from prospective cohorts. The paper presented promising urinary- based marker and confirmed the importance of prostaglandin pathway in CRC.


* Excellent paper which used different techniques and followed well-designed approach to identify possible stage specific markers.


* Good paper illustrating interesting method whereby membrane proteome can be analysed for biomarkers discovery.

** Excellent paper illustrating the benefits of using targeted proteomic in plasma proteome analysis. Strong study design was used for the identification and validation of proteins signature


*The study described the purification and analysis of proteome in colon cancer associated fibroblast using multiple techniques.


** This study developed non-invasive diagnostic assay for different antigens in sera samples. The assay showed strong diagnostic performance and was validated on a large size cohort.


** Excellent paper which described innovative proteomic approach for biomarker discovery and validation. The study also validated their results by transcriptomics validation on external cohorts.


Table 1. Summary of recent diagnostic biomarkers discoveries and their potential applications in CRC.

<table>
<thead>
<tr>
<th>Target(s)</th>
<th>Proteomics</th>
<th>Validation</th>
<th>Findings</th>
<th>Potential utility</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kininogen-1</td>
<td>Clinprot-MALDI-TOF/TOF-MS</td>
<td>Sera: 35 healthy, 35 advanced colorectal adenoma (ACA) and 40 preoperative CRC</td>
<td>ELISA Sera: 85 healthy, 80 ACA, 143 preoperative CRC (AJCC stage: I = 14, II = 63, III = 37 and IV = 29) and 58 postoperative CRC</td>
<td>The sensitivity, specificity, and accuracy are 70.13%, 65.88% and 67.90%, respectively</td>
<td>Diagnosis and screening [30]</td>
</tr>
<tr>
<td>Peptide signature</td>
<td>MALDI-TOF/TOF-MS</td>
<td>Plasma: 38 healthy, 13 FAP, 26 adenoma and 58 sporadic CRC (29 early stages and 29 late stages)</td>
<td>ELISA Plasma: 22 healthy, 10 FAP, 8 adenoma and 36 CRC</td>
<td>Associated with malignant transformation of adenomas in FAP patients</td>
<td>Diagnosis and screening [31]</td>
</tr>
<tr>
<td>Panel of proteins including; IL8, IP-10 and TNF-alpha</td>
<td>MALDI-TOF/TOF-MS</td>
<td>Plasma: 30 adenomas and 30 carcinomas (AJCC Stage: I = 4, II = 15 and III = 11)</td>
<td>ELISA and multiplex array Plasma: 30 adenomas and 30 carcinomas (AJCC stage: I = 4, II = 15 and III = 11)</td>
<td>A significant increase in the levels of proteins in carcinoma compared to adenomas</td>
<td>Diagnosis and screening [32]</td>
</tr>
<tr>
<td>PGE(_2) metabolites (PGE-M)</td>
<td>LC-MS</td>
<td>Urine: 420 control, 130 low risk adenoma and 290 high risk adenoma</td>
<td>None</td>
<td>PGE-M level is indicator of an increased risk for advanced adenoma and identifies patients who might benefit from NSAID chemoprevention</td>
<td>Diagnosis, screening and predictive [33]</td>
</tr>
<tr>
<td>Metabolomics profiles</td>
<td>NMR spectrum and learning algorithm</td>
<td>Urine: 633 healthy, 110 hyperplastic polyps 243 adenoma and 2 CRC</td>
<td>None</td>
<td>Sensitivity of 64% and a specificity of 65%</td>
<td>Diagnosis [34]</td>
</tr>
<tr>
<td>Olfactomedin 4 (OLFM4)</td>
<td>iTRAQ - MALDI-TOF/TOF-MS</td>
<td>Tissues: 4 adenoma and 24 CRC</td>
<td>IHC Tissues: 30 adenomas, 12 intramucosal carcinoma and 84 CRC (AJCC stage: I = 26, II = 14, III = 25 and IV = 19)</td>
<td>OLFM4 increases in adenomas and in early stage CRC before dropping significantly in stages (II-IV)</td>
<td>Diagnosis and prognosis [35]</td>
</tr>
</tbody>
</table>
Table 2. Summary of recent predictive biomarkers discoveries and their potential applications in CRC.

<table>
<thead>
<tr>
<th>Target(s)</th>
<th>Proteomics</th>
<th>Validation</th>
<th>Findings</th>
<th>Potential utility</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGT, APOE and DBP</td>
<td>2D-DIGE and LC-MS/MS</td>
<td>Sera: 23 CRC responders to treatment (AJCC stage: II = 1 and III = 10) and 12 non-responders (stage IV)</td>
<td>ELISA</td>
<td>Sera: 68 CRC (AJCC stage: I = 3, II = 8, III = 14 and IV = 43)</td>
<td>Proteins are associated with survival outcomes in metastatic CRC patients treated with chemotherapy and bevacizumab</td>
</tr>
<tr>
<td>LOXL2</td>
<td>iTRAQ – LC-MS</td>
<td>Cell lines Tissues: 12 matched colon cancer (AJCC stage: II = 5 and III = 7)</td>
<td>PCR</td>
<td>Tissues: 70 colon cancer (AJCC stage: I = 8, II = 26, III = 22 and IV = 14)</td>
<td>LOXL2 identifies a subgroup of patients (stage II and III CRC) who can benefit from adjuvant chemotherapy</td>
</tr>
<tr>
<td>Protein panel (n=14), validated (n=4): HADHA, PLEC1, TAGLN and TKT</td>
<td>Isotope coded protein label</td>
<td>Tissues: 20 rectal carcinoma (AJCC stage: II = 10 and III = 10)</td>
<td>IHC</td>
<td>Tissues: 10 good responders and 10 bad responders to neoadjuvant chemoradiation after surgery</td>
<td>This protein panel predicts the response for neoadjuvant chemoradiation in rectal carcinomas</td>
</tr>
<tr>
<td>Delta-type protein kinase C (PKCδ)</td>
<td>LC-MS/MS</td>
<td>Cell line and animal xenograft tumour treated with dasatinib</td>
<td>IHC and western blot</td>
<td>Cell lines and animal xenograft tumour treated with dasatinib</td>
<td>PKCδ pY313 assessment can determine the benefit of dasatinib in a subset of CRC patients</td>
</tr>
<tr>
<td>Phosphorylated epidermal growth factor receptor (pEGFR)</td>
<td>LC-MS/MS</td>
<td>3D secretomes of CRC isogenic cells treated with cetuximab Sera: 18 metastatic CRC with the KRAS (exon 2) WT status, treated with cetuximab plus FOLFIRI</td>
<td>ELISA</td>
<td>Plasma: 18 metastatic CRC with the KRAS (exon 2) WT status, treated with cetuximab plus FOLFIRI</td>
<td>pEGFR is associated with CRC cells sensitivity to cetuximab and therefore patients’ response to this drug</td>
</tr>
</tbody>
</table>
Figure legend

Figure 1.
An overview of colorectal cancer biomarkers types, methods of assessment and potential utilities of biomarkers in clinical practice.
CRC biomarkers

Type

Proteins

Genomics

Lipids, metabolites and carbohydrate

Method

Proteomics including antibody arrays, immunoassays

DNA/RNA microarray, miRNA expression profiling, whole genome sequencing

Metabolomics

Potential utilities

- **Screening**: diagnosis at asymptomatic, early stage and at precancerous stage (high risk population).
- **Diagnostic**: diagnosis of metastatic CRC of unknown origin and classifications of polyps.
- **Prognostic**: risk stratification based on natural outcome and guide treatment decisions.
- **Predictive**: patients stratification based the outcome of particular drug therapy (response and side affect).
- **Monitoring**: monitor progression and detection of recurrence.
- **Others**: therapeutic targets and molecular understanding of cancer biology.
<table>
<thead>
<tr>
<th>Target(s)</th>
<th>Proteomics</th>
<th>Validation</th>
<th>Findings</th>
<th>Potential utility</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGF1-R, IRF2BP1 and MX1</td>
<td>LC-MS</td>
<td>Tissues: 19 CRC (10 lymph node metastatic and 9 non-metastatic)</td>
<td>IHC</td>
<td>Tissues: 40 CRC (UICC stage: II = 20 and III = 20)</td>
<td>Expression of these proteins is associated with lymph node metastasis</td>
</tr>
<tr>
<td>ALDOA, CA1, GRP78 and PPIA</td>
<td>MALDI-TOF-MS and 2D-DIGE</td>
<td>Tissues: 5 CRC for each stage</td>
<td>IHC and Western blot (WB)</td>
<td>Tissues: 103 CRC (AJCC stage: I = 3, II = 45, III = 30 and IV = 25)</td>
<td>Dynamic patterns of proteins expression are associated with CRC prognosis especially for stage III and IV</td>
</tr>
<tr>
<td>Mapsin</td>
<td>Nano LC-MS</td>
<td>Tissues: 5 stage IV patients (time to recurrence&lt;6 months) and 5 patients longer time to recurrence</td>
<td>IHC</td>
<td>Tissues: 1 control and 1 CRC sample for each stage</td>
<td>Maspin expression is independent predictor of time to recurrence and is associated with diseases specific survival in stage III CRC</td>
</tr>
<tr>
<td>STOML2</td>
<td>LC-MS/MS</td>
<td>Tissues: 28 pair of normal and CRC (Dukes’ stage: A = 4, B = 7, C = 11 and D = 6)</td>
<td>ELISA and IHC</td>
<td>Tissues: 184 adenoma and 205 CRC matched with normal mucosa (AJCC stage: I + II 33 and III + IV = 172)</td>
<td>Overexpression of STOML2 is associated with poor survival. Plasma concentrations of STOML2 were higher in early-stage CRC compared with healthy individuals</td>
</tr>
<tr>
<td>CDH17, DEFA1, EZR, FN1, ITGB2, MLEC and TNC</td>
<td>LC–MS/MS</td>
<td>Tissues: 8 primary CRC (2 for each stage) and their corresponding adjacent normal mucosa</td>
<td>IHC and WB</td>
<td>Tissues: 8 primary CRC (2 for each stage) and their corresponding adjacent normal mucosa</td>
<td>Proteins signature is associated with CRC stage and epidermal growth factor receptor expression</td>
</tr>
<tr>
<td>Protein Marker</td>
<td>Analytical Methods</td>
<td>Tissues</td>
<td>Immunohistochemistry (IHC)</td>
<td>Metabolomics</td>
<td>Metabolite Profile</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------</td>
<td>---------</td>
<td>--------------------------</td>
<td>-------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>ALDH1 and 14-3-3β</td>
<td>LC-MS and 2D-DIGE</td>
<td>Tissues: 28 normal and 28 stage B CRC</td>
<td>IHC</td>
<td>Tissues: 515 CRC (Dukes’ stage: A = 90, B = 201 and C = 224) and 50 normal mucosa</td>
<td>ALDH1 and 14-3-3β negative tumours have a better prognosis than tumours showing either 14-3-3β or ALDH1 positivity</td>
</tr>
<tr>
<td>RAI3</td>
<td>LC-MS/MS</td>
<td>Tissues: 4 colon cancer and 4 normal</td>
<td>IHC</td>
<td>367 colon cancer (Dukes’ stage: A = 49, B = 122, C = 144 and D = 52) and 51 normal mucosa</td>
<td>High RAI3 expression is associated with colon cancer recurrence in small subgroup of patients</td>
</tr>
<tr>
<td>Cytokeratin 17 and Moesin</td>
<td>2D-DIGE and MALDI-TOF/TOF-MS</td>
<td>Tissues: 4 CRC (AJCC stage: II = 1 and III = 3) and 4 normal</td>
<td>IHC</td>
<td>Tissues: 166 CRC (AJCC stage: I = 33, II = 59, III = 65 and IV = 19) and 4 CRC (AJCC stage: II = 1 and III = 3) and 4 normal</td>
<td>Moesin and KRT17 were not expressed in normal mucosa and their expression increased as tumour (pT) stage advanced</td>
</tr>
<tr>
<td>FXYD3, GSTM3 and S100A11</td>
<td>MALDI-TOF MS and LC-MS</td>
<td>Tissues: 54 colon cancer (UICC stage: II = 21 and III = 33) and 6 colon cancer (II = 3 and III = 3)</td>
<td>IHC</td>
<td>Tissues: 168 colon cancer (UICC stage: II = 87 and III = 81)</td>
<td>Protein expressions correlate with the presence of nodal metastases</td>
</tr>
<tr>
<td>HSP27</td>
<td>2D-DIGE and LC-MS/MS</td>
<td>Tissues: 9 colon and 3 rectal cancer (stage III)</td>
<td>IHC</td>
<td>Tissues: 199 colon cancer and 205 rectal cancer (AJCC stage I + II = 108 and III + IV = 97)</td>
<td>HSP27 expression is associated with poor outcome in rectal cancer</td>
</tr>
<tr>
<td>Metabolomic profile</td>
<td>(H NMR) spectrometry and GC-MS</td>
<td>Sera: 42 stages II and III loco-regional CRC, 45 liver-only metastases and 25 extrahepatic metastases</td>
<td>None</td>
<td>None</td>
<td>Metabolome profile is different in locoregional CRC, in liver-only metastases and in extrahepatic metastases</td>
</tr>
<tr>
<td>CEA, IL-8 and prolactin</td>
<td>Multiplex immunoassay platforms</td>
<td>Plasma: 75 CRC (15 for each Dukes stage A-D) and 15 healthy</td>
<td>None</td>
<td>None</td>
<td>Protein signature is associated with increased CRC progression and correlates with Dukes’ stage</td>
</tr>
<tr>
<td>Protein</td>
<td>Labeling Method</td>
<td>Plasma Sample Details</td>
<td>Tissues Sample Details</td>
<td>Prognostic Information</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Gelsolin</td>
<td>Cy-dye labelled proteins (MALDI-TOF MS, LC-MS)</td>
<td>Plasma: 32 CRC, collected before surgery and one closest to distal metastasis diagnosis (AJCC stage at first diagnosis: I = 2, II = 7 and III = 23)</td>
<td>Plasma: 149 CRC (I + II = 74 and III+IV=75) and 25 normal tissues: 148 CRC (I = 10, II = 64, III = 48 and IV = 26) and 133 normal mucosa</td>
<td>Plasma levels of secretory gelsolin are higher in patients with distal metastases (stage IV versus stages I–III CRC before treatment)</td>
<td></td>
</tr>
<tr>
<td>CD44, CDH5, CFH, HLA-A, HP and PTPRJ</td>
<td>LC-MS/MS</td>
<td>Plasma: 202 CRC (AJCC stage: I = 43, II = 58, III = 49 and IV = 52)</td>
<td>Plasma: 202 CRC (AJCC stage: I = 43, II = 58, III = 49 and IV = 52)</td>
<td>This panel provides a prognostic information on survival and other prognostic parameters</td>
<td></td>
</tr>
<tr>
<td>MIC1 and PTGS2</td>
<td>ELISA</td>
<td>Plasma: 618 CRC (AJCC stage: I + II + III = 533 and IV = 85)</td>
<td>Tissues: 245 CRC (stage not stated)</td>
<td>MIC1 level measured prior diagnosis is associated with disease specific mortality, mainly in PTGS2-positive tumours</td>
<td></td>
</tr>
<tr>
<td>STMN1</td>
<td>2-D DIGE, MALDI-TOF/TOF MS</td>
<td>CRC cell lines (HCT-116 and its metastatic derivative E1)</td>
<td>Tissues: 324 CRC (AJCC stage: I = 22, II = 120, III = 97 and IV = 85)</td>
<td>Higher expression of STMN1 correlates with poorer prognosis. STMN1 expression is higher in primary and metastatic CRC compared with normal mucosa</td>
<td></td>
</tr>
<tr>
<td>COL6A3</td>
<td>iTRAQ-LC-MS</td>
<td>Cell lines</td>
<td>Tissues: 90 matched CRC (AJCC stage: I = 9, II = 47, III = 31 and IV = 2)</td>
<td>Expression of COL6A3 is higher in CRC and it is associated with Dukes stage, T stage and recurrence</td>
<td></td>
</tr>
</tbody>
</table>
Table S2. Summary of recent screening biomarkers discoveries and their potential applications in CRC

<table>
<thead>
<tr>
<th>Target(s)</th>
<th>Proteomics</th>
<th>Validation</th>
<th>Findings</th>
<th>Potential utility</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP, LRG1, PON1, SERPINA3 and TIMP1</td>
<td>LC MS/MS</td>
<td>Plasma: 23 non advanced adenoma, 11 hyperplastic polyp, 66 normal and 97 CRC (AJCC stage: I = 32, II = 26, III = 31 and IV = 8)</td>
<td>Targeted LC-MS (SRM)</td>
<td>Plasma: 4 advanced adenoma, 2 benign adenoma, 1 dysplastic polyp, 6 diverticular disease, 4 Crohn, 50 healthy and 202 CRC (AJCC stage: I = 43, II = 58, III = 49 and IV = 52)</td>
<td>This panel detect CRC at 72% accuracy compared with 49% for CEA</td>
</tr>
<tr>
<td>Autoantibodies: IMPDH2, MAGEA4, MDM2 and TP53</td>
<td>Multiplex serology, a fluorescent bead-based GST capture immunosorbent assay</td>
<td>Sera: 124 normal and 352 CRC (AJCC stage: I = 96, II = 102, III = 105 and IV = 49)</td>
<td>The same method</td>
<td>Sera: 49 CRC (AJCC stage: 0 = 4, I = 18, II = 5, III = 19 and IV = 3); 100 normal, 29 non-advanced adenoma, and 99 advanced adenoma</td>
<td>Sensitivity of autoantibodies is 26% for early stage CRC at a specificity of 90%. Detected 20% of advanced adenomas</td>
</tr>
<tr>
<td>CEA, IGFBP2, LRG1 and MAPRE1</td>
<td>2D-HPLC and LC-MS/MS</td>
<td>Cell lines and plasma: 18 months pre diagnosis, 90 CRC (AJCC stage: I = 8, II = 29, III = 37 and IV = 16) and 90 controls</td>
<td>ELISA</td>
<td>Plasma: 58 newly diagnosed CRC (stage not provided) and 58 age-matched controls Plasma: 7 months prior diagnosis, 32 CRC (AJCC stage: I = 4, II = 13, III = 12 and IV = 3) and 32 controls</td>
<td>Predictive value in pre-diagnostic CRC plasmas (41% sensitivity at 95% specificity)</td>
</tr>
<tr>
<td>Anti-p53, CEA, CYFRA 21-1, OPN and seprase</td>
<td>ELISA</td>
<td>Sera: 301 CRC (UICC stage: 0 = 6, I = 53, II = 68, III = 76 and IV = 68), 14 hyperplastic polyps, 143 advanced adenoma, 135 healthy, 176 other cancers (prostate, liver, lung, breast, kidney, bladder, ovary, and endometrium) 258 disease and other controls (Diverticulitis, inflammatory bowel disease, infection-related diarrhoea)</td>
<td></td>
<td>Diagnostic power: 69.6% sensitivity at 95% specificity and 58.7% at 98% specificity</td>
<td>Screening</td>
</tr>
<tr>
<td>Protein/Marker</td>
<td>Assay Method</td>
<td>Study Details</td>
<td>Performance Notes</td>
<td>Stage/Screening</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>--------------</td>
<td>---------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>Anti-p53, CEA, ferritin, osteopontin and seprase</td>
<td>Electrochemiluminescence cobas e601 assay</td>
<td>Sera: 1,200 controls, 420 advanced adenoma, 4 carcinoma in situ, and 36 CRC (UICC stage: I = 13, II = 5, III=12 = 6 and IV = 2)</td>
<td>Performance is inferior to FIT, but comparable with the faecal occult blood test (FOBT)</td>
<td>Screening</td>
<td></td>
</tr>
<tr>
<td>Complement component 9 (C9)</td>
<td>2DICAL and LC-MS</td>
<td>Plasma: 59 healthy and 31 CRC (AJCC stage: 0 = 5, I = 10, II = 7, III = 6 and IV = 3)</td>
<td>RPPM</td>
<td>Plasma: 115 CRC (0=17, I = 35, II = 28, III = 25 and IV = 10) and 230 healthy</td>
<td>C9 was elevated in patients with early stages of CRC</td>
</tr>
<tr>
<td>C3, C9, GSN, HABP2, ORM1 and SAA2</td>
<td>HPLC and MRM LC/MS</td>
<td>Sera: 259 healthy and 172 CRC (AJCC stage: I = 19, II = 53, III = 71 and IV = 27)</td>
<td>None</td>
<td>NA</td>
<td>Diagnostic assay showed promising results in detecting CRC (sensitivity of 93.75%, a specificity of 82.89%)</td>
</tr>
<tr>
<td>ORM2</td>
<td>iTRAQ coupled with micro Q-TOF/MS</td>
<td>Plasma: 10 CRC and 10 healthy</td>
<td>ELISA WB</td>
<td>Plasma: 65 control, 59 hyperplastic polyp, 62 inflammatory bowel disease, 53 adenoma and 180 CRC (AJCC stage: I = 49, II = 31, III = 62 and IV = 38)</td>
<td>ORM2 level in plasma and tissue was higher in CRC compared with the healthy samples</td>
</tr>
<tr>
<td>Clusterin</td>
<td>LC-ESI-MS/MS</td>
<td>Plasma: 10 CRC and 10 healthy in each of first two phases</td>
<td>Targeted LC-MS (SRM)</td>
<td>Plasma: 48 CRC and 48 healthy (Stage of CRC not stated)</td>
<td>Plasma levels of clusterin were higher in CRC compared with control and protein was associated with risk of CRC (only in men)</td>
</tr>
<tr>
<td></td>
<td>MALDI-TOF-MS</td>
<td>Proteins expressed in E.coli</td>
<td>Multiplex beads assay and ELISA</td>
<td>Sera: 135 CRC (AJCC stage: I = 35, II = 25, III = 46 and IV = 29), 65 other cancer types, 14 inflammatory bowel disease and 93 healthy</td>
<td>Combination of autoantibodies achieved diagnostic accuracy of 89.7%, with 66% sensitivity at 90.0% fixed specificity</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------</td>
<td>-----------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Autoantibodies</strong></td>
<td><strong>against: EDIL3, GTF2B, HCK, P53, PIM1 and STK4</strong></td>
<td><strong>Proteins expressed in E.coli</strong></td>
<td><strong>Multiplex beads assay and ELISA</strong></td>
<td><strong>Sera: same cohort used in proteomic</strong></td>
<td><strong>The expression of collagen I may be an early event in CRC tumorigenesis and could provide prognostic information</strong></td>
</tr>
<tr>
<td><strong>Collagen I</strong></td>
<td><strong>LC–MS</strong></td>
<td><strong>Sera: 91 CRC (UICC stage: I = 21, II = 41, III = 22 and IV = 7) and 33 healthy</strong></td>
<td><strong>ELISA</strong></td>
<td><strong>Tissues: 26 pair of normal and CRC (UICC stage: I = 7, II = 7, III = 7 and IV = 5)</strong></td>
<td><strong>Sera: same cohort used in proteomic</strong></td>
</tr>
<tr>
<td><strong>Adipophilin</strong></td>
<td><strong>LC/MS (2DICAL)</strong></td>
<td><strong>Plasma: 21 healthy and 22 CRC (AJCC stage: I = 3, II = 6, III = 8 and IV = 5)</strong></td>
<td><strong>RPPM</strong></td>
<td><strong>Plasma: 109 healthy and 101 CRC (AJCC stage: I = 19, II = 31, III = 32 and IV = 17)</strong></td>
<td><strong>Plasma: 87 healthy and 26 CRC (AJCC stage: I = 12, II = 5, III = 8 and IV = 1)</strong></td>
</tr>
<tr>
<td><strong>MRC1 and S100A9</strong></td>
<td><strong>LC-MS</strong></td>
<td><strong>Sera: 3 healthy and 3 CRC</strong></td>
<td><strong>WB</strong></td>
<td><strong>Cell lines and sera: 3 healthy and 3 CRC</strong></td>
<td><strong>Sera: 96 healthy and 112 CRC (AJCC stage: I = 21, II = 50 and III = 41)</strong></td>
</tr>
<tr>
<td><strong>MAPRE1</strong></td>
<td><strong>LC/MS</strong></td>
<td><strong>Mouse model, cell lines and plasma: 60 adenomas, 60 CRC (AJCC stage: I = 11, II = 19, III = 21 and IV = 9) and 60 healthy</strong></td>
<td><strong>Antibody array</strong></td>
<td><strong>Plasma: 60 adenomas, 60 CRC (AJCC stage: I = 11, II = 19, III = 21 and IV = 9) and 60 healthy</strong></td>
<td><strong>Tissues: 20 normal tissues, 10 adenomas, and 66 CRC (stage not provided)</strong></td>
</tr>
<tr>
<td>Proteins (A2M, APOH, IGL@, MACF1 and VDB) and metabolite signature</td>
<td>2DIGE, Finnigan LTQ-MS and GC-MS</td>
<td>Sera: 30 CRC (Dukes’ stage: A = 3, B = 13, C = 8 and D = 6) and 30 healthy</td>
<td>ELISA</td>
<td>Sera: same cohort as used in proteomics</td>
<td>Differential expression of proteins in CRC compared with healthy. 93.5% of CRC patients are identified using the 6 metabolites</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
<td>--------</td>
<td>---------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Volatile organic compounds signature</td>
<td>FAIMS and GC-MS</td>
<td>Urine: 83 CRC (AJCC stage: I = 9, II = 24, III = 32, IV = 9 and no stage = 9) and 50 control</td>
<td>None</td>
<td>NA</td>
<td>Sensitivity and specificity for CRC detection were 88% and 60% respectively</td>
</tr>
<tr>
<td>Angiopoietin-2, calprotectin, FGF-23, IL-13, M2-PK, MMP-10 and TPO</td>
<td>Biotin label-based protein array</td>
<td>Faeces: 20 CRC (AJCC stage: I = 2, II = 7, III = 9 and IV = 2) and 20 healthy</td>
<td>ELISA and multiplex faecal protein biochip</td>
<td>Faeces: same cohort as used in proteomics</td>
<td>Proteins levels are significantly higher in CRC compared with healthy controls</td>
</tr>
<tr>
<td>A1AT and CTSD</td>
<td>Gel-enhanced LC−MS</td>
<td>Tissues: 37 CRC (AJCC stage: I = 13 and II = 24) and 37 normal</td>
<td>IHC</td>
<td>Tissues: 93 CRC (AJCC stage: I = 4, II = 86 and III = 3)</td>
<td>Less A1AT and more CTSD in CRC compared with healthy samples. Combination of both proteins identified 96.77% of CRC</td>
</tr>
<tr>
<td>EFEMP2</td>
<td>Nano LC-MS/MS</td>
<td>Tissues: 9 CRC (AJCC stage: I = 7 and II = 2) paired with normal mucosa</td>
<td>IHC and ELISA</td>
<td>Tissues: 88 CRC (UICC stage: I = 23, II = 29, III = 26 and IV = 10), 19 adenoma and 16 normal colon Sera: 79 healthy, 14 adenoma, and 122 CRC, stage not stated, but smallest proportion of cases is in stage I (figure 7B) Tissues and sera: 9 pairs of CRC and normal mucosa</td>
<td>The expression level of EFEMP2 increases in early stages CRC. Diagnostic accuracy significantly better than CEA</td>
</tr>
<tr>
<td>Protein Panel</td>
<td>Method</td>
<td>Tissues</td>
<td>Assay</td>
<td>Screening</td>
<td>References</td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>---------</td>
<td>-------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>CAMP, ERp29, HSPA8 and TPM3</td>
<td>2D LC-MS/MS</td>
<td>Tissues: 3 CRC and 3 normal</td>
<td>IHC, WB</td>
<td>Tissues: 69 CRC matched with normal (AJCC stage: I = 15, II = 21 and III = 33)</td>
<td>The protein panel can detect CRC via IHC (accuracy of 73.2%)</td>
</tr>
<tr>
<td>SORD</td>
<td>iTRAQ 8-plex labelled LC-MS/MS</td>
<td>Cell lines and tissues: 30 adenomas and 30 normal</td>
<td>IHC, WB</td>
<td>Cell lines and tissue: normal colon, colorectal adenomas, and adenocarcinomas (numbers not provided)</td>
<td>Significant increase in SORD expression in adenomas and cancer cell lines</td>
</tr>
<tr>
<td>TRFM</td>
<td>LC-MS/MS</td>
<td>Cell lines</td>
<td>ELISA, WB</td>
<td>Plasma: 77 healthy and 228 CRC (I = 68, II = 68, III = 65 and IV = 27)</td>
<td>TRFM expression increases in stages I and II compared with stages III and IV</td>
</tr>
</tbody>
</table>
References


