REVIEW

Mass spectrometry based translational proteomics for biomarker discovery and application in colorectal cancer

Hong Ma[1,2], Guilin Chen[1,3] and Mingquan Guo[1]

[1] Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Sino-Africa Joint Research Center, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, P. R. China
[2] Haematology and Oncology Division, Children's Hospital Los Angeles, Los Angeles, CA, USA
[3] University of Chinese Academy of Sciences, Beijing, P. R. China

Received: July 28, 2015
Revised: November 17, 2015
Accepted: November 25, 2015

Colorectal cancer (CRC) is a leading cause of cancer-related death in the world. Clinically, early detection of the disease is the most effective approach to tackle this tough challenge. Discovery and development of reliable and effective diagnostic tools for the assessment of prognosis and prediction of response to drug therapy are urgently needed for personalized therapies and better treatment outcomes. Among many ongoing efforts in search for potential CRC biomarkers, MS-based translational proteomics provides a unique opportunity for the discovery and application of protein biomarkers toward better CRC early detection and treatment. This review updates most recent studies that use preclinical models and clinical materials for the identification of CRC-related protein markers. Some new advances in the development of CRC protein markers such as CRC stem cell related protein markers, SRM/MRM-MS and MS cytometry approaches are also discussed in order to address future directions and challenges from bench translational research to bedside clinical application of CRC biomarkers.

Keywords:
Biomarker / Colorectal cancer / MS / MRM / Translational proteomics

1 Introduction

Colorectal cancer (CRC) is the fourth most commonly diagnosed cancer and the third leading cause of cancer death in both men and women worldwide, with approximately 1 million new cases and 500 000 deaths each year [1]. It develops in the colon or the rectum which are parts of the digestive system, also called the gastrointestinal system. In the United States alone, the number of new cases of colon and rectum cancer was 55.8 per 100 000 men and women per year. The number of adjusted deaths was 22.7 per 100 000 men and women per year from 1975 to 2012. Interestingly, after 1980s, the number of new cases and deaths was steadily declined in a rate of 3.5 and 2.8 to 38.5 and 14.7 per 100 000 men and women per year. In contrast to the trend of incidence and death rate, the 5-year relative survival rate, however, had reached plateau to ~65% in 1999 from as low as 48.6% in 1975, and has been staggering even since (http://seer.cancer.gov/ statfacts/html/colorect.html). From the analysis of these data, a central tenet is that, since 1990, people have adopted more healthy life style (diets, exercise, and body weight control) and advanced diagnostic technologies have become more accessible (namely Computed Tomography Scan (CT scan), Magnetic Resonance Imaging (MRI), and Positron Emission Computed Tomography (PET)), but early detection and subsequent treatment options have been a bottleneck even though some progress has been made.

1.1 Biology of CRC

CRC develops from a benign precursor lesion (polyps) called an adenoma, fewer than 10% of polyps progress into cancer and some become metastasis at the late stage. Most polyps remain benign for a long time and are often termed as hyperplastic polyps. This process can easily take more than 10 years providing an excellent window of opportunity for screening. Disease outcome is highly dependent upon the stage at

Correspondence: Professor Mingquan Guo, Medicinal Chemical Biology Group, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, P. R. China
E-mail: zhaoguo2000@yahoo.com

Abbreviations: **APC**, adenomatous polyposis coli; **CB**, chromatin binding; **CCAP**, CRC-associated protein; **CRC**, colorectal cancer; **CSC**, cancer stem cell; **DIGE**, difference gel electrophoresis; **EGFR**, epidermal growth factor receptor; **FOBT**, fecal occult blood test; **GDF**, growth/differentiation factor; **IMS**, imaging MS; **MALDI-MSI**, matrix assisted laser desorption ionization imaging mass spectrometry; **UPLC**, ultra performance LC

Colour Online: See the article online to view Fig. 1 in colour.

which the disease is detected. Patients with localized disease (stage I) have an excellent outcome with 5-year survival rate up to 90% whereas the 5-year survival rate for patients with metastasized disease (stage IV) is as low as 8% [2]. Unfortunately, clinical symptoms mostly arise at a late stage, when the disease has already spread outside of the colon.

During the progression from hyperplastic polyps to cancerous, multiple steps of genetic mutations accumulate. Recent genome-wide analyses of solid tumors, including CRC, have shown mutations in between ten and 100 protein encoding genes [3–6]. Activation of the canonical Wnt signaling pathway is an early event that occurs in about 80% of all colon adenomas, and is present in 93% of all CRCs. This often results from a mutation in the adenomatous polyposis coli (APC) tumor suppressor gene or a gain-of-function mutation in the regulatory domain of β-catenin, a transcription factor (TCF) transcription signaling activator [7]. Subsequent genomic alternations including activation of mutations in KRAS and BRAF, inactivation of TP53, alterations of the PI3K/Akt, PTEN, TGF-β signaling pathways [8], specific gene methylation (namely vimentin), CpG island methylation, chromosomal instability (namely a loss of chromosome 18q or a gain of chromosome 20q), microsatellite instability, and global hypomethylation all contribute to the disease progression and metastasis [9]. Continuous effort to discovery molecular traits with more advanced technologies such as MS-based translational proteomics that help explore the underlying mechanism of adenoma-to-carcinoma progression can lead to the detection of novel molecular biomarkers that potentially have added value to the currently applied clinical parameters and existing biomarkers.

Clinically, there are five well-defined and established stages of CRC development from stages 0 to IV [10]. The US Centre for Disease and Control Prevention (CDC) reports that the 5-year survival rate for persons who received a diagnosis of localized CRC (stages I and IIA) is ∼91% compared with 12.5% for distant-stage cancer (stage III or IV), but only ∼40% of patients were diagnosed when the disease is at the localized stage [11]. A study registered at the National Cancer Institute's SEER database, conducted with more than 28 000 people diagnosed with colon cancer between 1998 and 2000, found that the observed 5-year survival rates related to the stage of the disease at diagnosis were the following: I-74%, IIA-67%, IIB-59%, IIC-37%, IIIA-73%, IIIB-46%, and IIIC-28% (source: American Cancer Society). These findings have led to the conclusion that the earlier the neoplastic lesion in the CRC developmental sequence is detected, the better clinical outcome is. However, about 20–30% of stage II patients will still get a disease relapse after adjuvant, while on the other hand only about 15% of stage III patients actually benefit from adjuvant chemotherapy. This illustrates the unmet clinical need for (protein based) biomarkers to optimize clinical decision making in stages II and III CRC.

Recent introduction of personal medicine and targeted therapies has made biomarkers and molecular diagnosis for predicting responses to therapies more prominent in CRC. As an example, treatment with drugs directed against epidermal growth factor receptor (EGFR) is restricted to patients with KRAS wild-type tumors only. About 35–45% CRC patients with KRAS have mutations on codons 12 and 13. CRC patients with KRAS mutations do not normally benefit from EGFR-based therapy, and to detect KRAS mutation prior to appropriate drug regimens becomes an important step.

1.2 Current clinical screening methods for early detection of CRC

The common clinically used screening methods for early detection of CRC are the fecal occult blood test (FOBT) and the endoscopy. FOBT is simple, inexpensive, and the least invasive method and has been proven to effectively reduce CRC mortality, but it presents relatively high false-negative and false-positive rates, and particularly poor sensitivity for detection of early-stage lesions [12]. Therefore, it requires a follow-up detection by endoscope. The introduction of immunochemical-FOBT has increase the sensitivity of stool-based screening in detecting advanced adenoma and early CRC. In contrast, Barium Enema, flexible sigmoidoscopy and colonoscopy offer significant improvements in detection rates, but they also have important disadvantages associated as inconvenience, invasiveness, and relatively high economic burden.

Recent advances in genomic and proteomic technologies have contributed to our molecular understanding of pathogenesis of CRC by identifying the genetic defects in genome, evaluating the expression profiles of genes and their encoded proteins in cancerous and noncancerous surrounding tissues and body fluids. The identification of genes and/or proteins that are characteristic of the development of CRC can render potential biomarkers that will facilitate the early detection of CRC. For examples, the isolation of DNA derived from patient stool samples, and the subsequent identification of CRC-associated rat sarcoma (RAS) mutations have been reported [13, 14]. These open up a possibility of screening for specific genetic mutations associated with CRC. Based on these research progresses, a commercial test was developed, which detects KRAS, APC, and TP53 gene mutations [15]. Other nucleotide markers such as the microsatellite instability marker BAT-26, miRNA29a, and miRNA92a also showed some clinical potential specifically discriminating CRC from normal tissues, but mostly for patients with advanced adenoma [16]. Nevertheless, one of the challenges for proteomics-driven biomarker discovery research is that the bulk of secreted mutant proteins cannot be identified and quantified directly by MS due to the lack of mutated peptide information in extant proteomics databases. Mathivanan et al. [17] utilized an integrated genomics and proteomics strategy (referred to identification of mutated and secreted proteins) to identify 112 putative mutated tryptic peptides (corresponding to 57 proteins) in the collective secretomes derived from

a panel of 18 human CRC cell lines. The identification of mutated and secreted proteins technology has great promises to improve the link between proteomics and genomic mutation data thereby providing an effective tool for targeting tryptic peptides with mutated amino acids as potential cancer biomarker candidates. Likewise, Zhang et al. [18] performed LC-MS/MS-based shotgun proteomic analyses on 95 TCGA colon cancer samples, they identified a total of 124 823 distinct peptides among the 95 samples, corresponding to 6 299 756 spectra in an assembly of 7526 protein groups with a protein-level false discovery rate of 2.64%, which is corresponding to 7211 genes. With searching customized sequence databases from matched RNA-Seq data for individual samples, 108 somatic protein variants mapped to 105 genes, including known cancer genes in the Cancer Gene Census database such as *KRAS, CTNNB1, SF3B1, ALDH2,* and *FH*. Intriguingly, this approach revealed abnormal amplification of HNF4α on chromosome 20 associated with CRC. Upon reanalysis of the HNF4A shRNA knockdown data for CRC cell lines, they found that the dependency of CRC cells on HNF4α correlated significantly with its amplification level. Other interesting candidates included TOMM34, which is overexpressed frequently in CRC tumors and is involved in the growth of CRC cells. In reality, compliance to early screening for CRC is low for high-risk age group (age > 50 years) [19]. It was hopeful that the development of blood-based tests would increase compliance, be less invasive, and be more universally available for routine screening. This contention has led to study on serum tumor markers such as carcinoembryonic antigen, which has been widely used in the surveillance of patients following primary surgical resection of CRC [20]. The other heavily studied serum marker is carbohydrate antigen 19-9. However, the levels of both tumor markers are generally low in plasma of patients with in situ carcinoma excised. Needless to say, these initial evaluations of serum markers for CRC screening demonstrated low sensitivity and poor specificity, especially when considering early adenomas.

The proteins involved in metabolism, protein folding, and signaling pathways have been analyzed using a variety of proteomics workflows to answer a variety of biological questions underlying CRC. It is gratifying that numerous biomarkers have been continuously conformed by immunohistochemical analysis in microarrayed tissue samples which have been relevant to the existing data [21]. In more detail, a comprehensive overview of differential candidate protein biomarkers in CRC related to adenoma-to-carcinoma progression and clinically significant outcomes has been summarized and shown in Table 1.

"Translational proteomics" focuses on the translation of basic proteomics science, the processes and platforms that facilitate the delivery of applications derived from proteomic analysis, offers opportunities to define protein expression profiles that reflect phenotypic change, and contributes to clinical application and utility [68]. To this end, the two excellent reviews on CRC and proteomics by Jimenez et al. [69] and De Wit et al. [70] are highly recommended. Given the plenty

Table 1. Differential candidate protein biomarkers in colorectal cancer

Colorectal cancer biomarker	References
Kininogen-1	[22]
GRP78, fructose-bisphosphate aldolase A	[23]
Carbonic anhydrase I, peptidyl-prolyl *cis–trans* isomerase A	
OLFM4	[24]
FXYD3, S100A11, GSTM3	[25]
MX1	[26]
NADH-dependent oxidative enzymes	[27]
Glutathione *S*-transferase pi	[28, 29]
cfDNA	[30]
PKC delta	[31]
ColoGuide Pro	[32]
Coloprint™	[33, 34]
Chaperonin t-complex proteins	[35]
Matrix metalloproteinases	[36, 37]
Cytochrome P450	[38]
14-3-3 Beta, aldehyde dehydrogenase 1	[39]
Annexins	[40]
Proline, glutamine, *N*-acetyl lysine *N*1-acetylspermidine, *N*8-acetylspermidine, spermine	[41]
Xanthosine, inosine, deoxyuridine, thymidine	
Carnitine, symmetricdimethylarginine, asymmetric-dimethylarginine	
GLUT1,PrP^C and 42 other cell surface candidate biomarkers	[42]
Hemoglobin, Myeloperoxidase	[43]
S100A9, Filamin A, L-plastin	
Actin, cytoplasmic 2	[44]
α-1-Antitrypsin 1	[44–47]
α-2-Macroglobulin precursor	[46, 48, 49]
Hemoglobin subunit alpha, hemoglobin subunit beta-1	[46, 50]
Complement C3 precursor	[45, 51]
Serotransferrin precursor	[45, 52, 53]
Haptoglobin, hemopexin precursor	[48, 54]
Vitamin D-binding protein precursor	[47, 48]
Sucrase-isomaltase, maltase-glucoamylase, antithrombin-III precursor, histone H2B type 1-B	[49]
Carbonic anhydrase 2	[55]
Muc2 protein	[56]
Protocadherin 24	[57]
Transthyretin precursor	[58]
TFF3, GDF15	[59]
A1AT, CTSD	[60]
C9, ApoAI	[61]
STOML2	[62]
HSP 60	[63]
Gelsolin	[64]
CHI3L1	[65]
Mucin, CEA, MCM2, CAF, M2-PK	[66]
α_1 Antitrypsin, hemoglobin, serotransferin, AAG1/2	[67]
Myeloblastin, albumin, CEA5, CEACAM5	

CAFs, cancer-associated fibroblasts; CEA, carcino-embryonic antigen.

of information emerged and applications to some extent, the aim of this review is to provide an up-to-date overview on MS-based translational proteomics which may lead to the identification of novel protein biomarkers suitable for addressing the CRC clinical needs outlined above.

1.3 General strategies and considerations for the identification of protein biomarkers

Molecular biomarkers can be categorized by their ability to discriminate disease from health states, to promote early detection, and to establish prognosis and predict response of patient to specific therapies. The discovery of biomarkers will also help in the understanding of the biological mechanisms underlying disease pathogenesis.

Typically, almost all CRC clinical studies begin to compare tumor tissues and matched normal tissues collected from various disease stages (namely stages I vs. IV) and prognostic natures (relapsed vs. nonrelapsed). Based on great advances in genomic, transcriptomic, and proteomic technologies which have contributed to our molecular understanding of pathogenesis of CRC by identifying the genetic defects in genome, evaluating the transcriptional, and expressional profiles of genes and their encoded proteins in cancerous and noncancerous surrounding tissues and body fluids, a workflow has been proposed to identify genes and/or proteins that are characteristic of the development of CRC (as shown in Fig. 1). The workflow can render three-in-one analysis of genome/transcriptome/proteome from a single biological samples for the discovery of potential biomarkers. In this workflow, tissue samples or body fluids are lysed and total proteins are directly used for proteomic analysis. More often than not, this limits the analyses to the more abundant proteins. Identifying low-abundant proteins that are potential CRC-associated biomarkers is a challenge and could be improved by sample enrichments through fractioning cell components based on subcellular locations (membrane proteins, cytoplasmic proteins, or nuclear proteins), physical-chemistry properties (*pI* and hydrophobicity) or antibody-capture and targeted proteomics biomarkers of CRCs, or by means of high-abundance protein depletion kit. Even with these enrichment methods, the natures of tumor mass genetic heterogeneity may pose some new challenges in identifying low-abundant CRC-associated protein (CCAP) biomarkers. Thus, specific cellular subpopulations (namely colon cancer stem cell (CSC)) could also be isolated for studies through cell surface immunophenotype-based fluorescence-activated cell sorting (FACS)-sorting or laser-capture microdissection followed by quantitative-based MS and bioinformatics data analyses [71].

In fact, many putative biomarkers discovered poise differential expression through quantitative comparison between tumor and health tissues, which requires a systematic and quantitative analysis of protein expression or quantitative proteomics. Currently, there are several popular peptide tagging methods for quantitative analysis in mass spectrum studies: (i) stable isotope labeling by amino acids in cell culture, the entire proteome is labeled by metabolic incorporation of isotopically stable and nonradioactive forms of amino acids. Usually, up to three biological conditions can be compared directly with each other using different isotopic forms of the amino acids arginine and lysine; (ii) the isotope-coded affinity tags strategy is a technology for relative protein quantification through a reactive group that reacts with the free thiol functionality of cysteine residues and a linker in which stable isotopes have been incorporated; (iii) Isobaric tagging (iTRAQ and TMT), peptides or proteins are labeled with various chemical groups that are isobaric, or the same in mass, through the amine-reactive group of the tag (*N*-hydroxylsuccinimideester-activated compounds) covalently binding to the peptide amino terminus or free amino termini of lysine residues of peptides. The relative abundances of the isobarically tagged peptides are revealed when the moieties fragment during MS/MS experiments to release reporter ions with different masses. There are two isobaric tag families commercially available TMT and iTRAQ. Other tagging methods such as nonselective isotope-coded protein labeling and isotope-differentiated binding energy shift tag have also been used for similar purposes.

After protein extraction, the 2D-PAGE separation coupled with an MS/MS technology is a traditional way of performing protein profiling. The 2D-DIGE (differential gel electrophoresis) technology is a step forward providing not only improved protein separation resolution and reproducibility, but also reducing gel-to-gel variation by allowing simultaneous separation and comparison of multiple samples on one gel [72]. However, proteins with extreme sizes and hydrophobic properties are often missing. Shotgun proteomics using LC coupled with MS/MS-based approach detection allows high-resolution separation of modified proteins and hundreds of proteins to be quantified simultaneously with improving confidence for accurately determining differential protein expression profiles. Gradually, LC-MS/MS shotgun proteomics practically replaced the 2D-PAGE MS/MS approach. This is optimally suited for the discovery of protein functions and widely used in novel biomarker identification. In contrast to shotgun proteomic approach, targeted proteomics methods by means of SRM or MRM-MS strategies, which restrict the MS measurement to a predefined set of peptides of interest, have recently been adopted for quantitative proteomics.

Contemporarily, bottom-up, middle-down, and top-down MS approaches, along with CID and electron-capture dissociation/electron-transfer dissociation based techniques for the characterization of histones and their PTMs have been proven to be powerful tools in this field [73]. These methods offer several means of combinations to evaluate proteomic compositions under changing conditions and thereby allow the characterization of cellular processes in an unbiased and increasingly comprehensive manner. In principle, the quantitative technologies use a complex mixture of tryptic peptides that can be selectively detected by LC coupled to, e.g. an

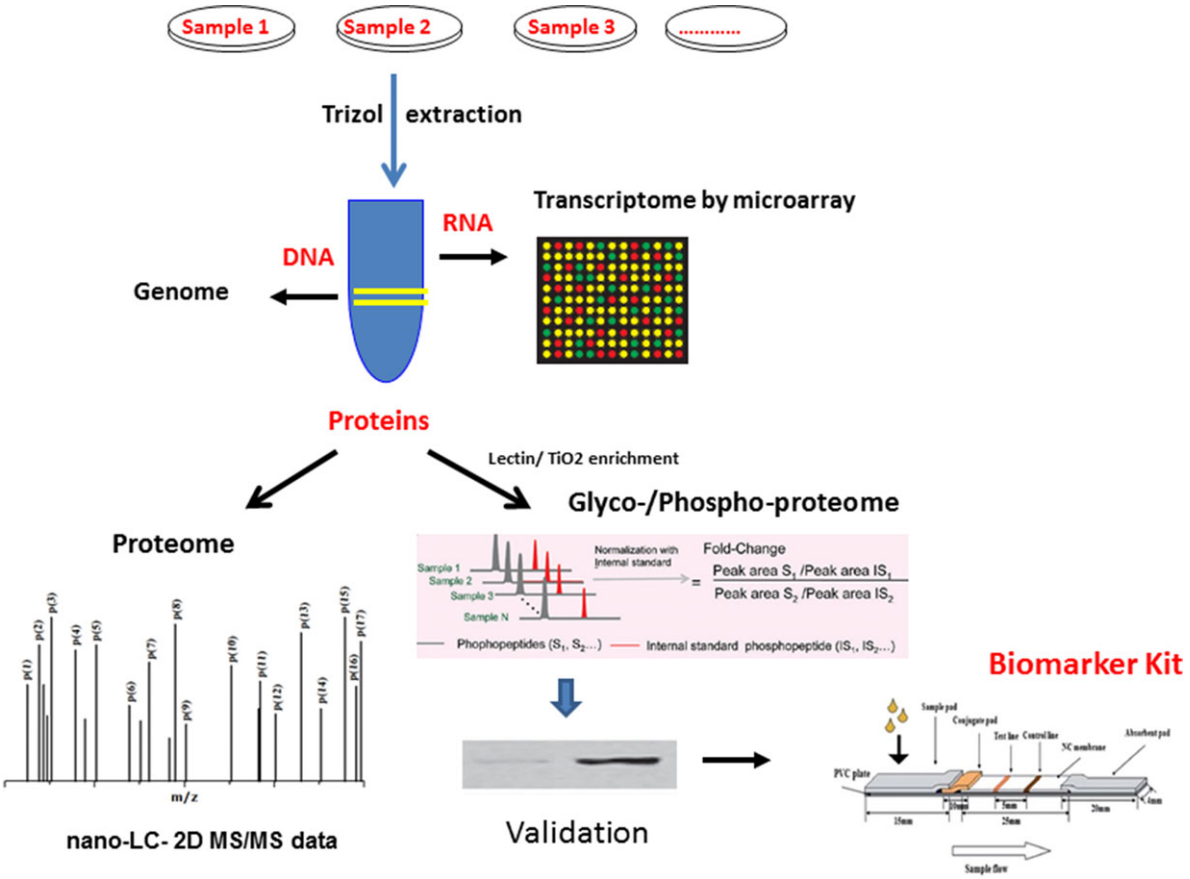

Figure 1. A proposed workflow for three-in-one analysis of genome/transcriptome/proteome from a single CRC cellular/patient sample for the discovery of potential CRC biomarkers.

electrospray triple-quadrupole MS; this system can select precursor ions in combination with their correspondent product ions during CID to produce specific detection related to a particular protein. Both SIM and MRM are highly sensitive techniques. They are possible to detect peptide not normally seen by a typical MS approach. In theory, peptides can be accurately quantitated at femto-mole concentration. In complex samples such as plasma or serum, abundance proteins may be depleted first by immunoaffinity followed by peptide fractionation through cation exchange chromatography prior to LC-SIM/MRM analysis. This system enables quantification of plasma proteins at low nanogram per milligram levels with as little as 10 μL of human plasma [74] or even at the subnanogram per milligram range if larger volumes (e.g. 1 mL) are used [75].

Human tumor mass often shows a high degree of cellular heterogeneity that is due to the presence of various cell types belonging to the tumor microenvironment such as immune cells and fibroblasts. Clonal evolution and/or cancer "stem" cell abnormal differentiation act in concert leading to genetic diversity within the tumors themselves, resulting in the existence of complicated cellular subpopulations with distinct functional properties. In addition, most available solid tumor samples are biopsies that are obtained on frozen block or formalin-fixed states, which complicate normal MS-based proteomic analyses. For such intact tissues analysis, imaging MS (IMS) approach would be suited because it allows simultaneous localization, distribution, and quantification of specific biomolecules in different histological regions of interest. With differences in ion sources, spatial resolution, and mass range, there are three major IMS techniques used for different analytes: MALDI, secondary ion mass spectrometry (SIMS), and desorption electrospray ionization (DESI). Among these, MALDI-IMS is used for analyzing peptides or proteins, and SIMS-IMS or DESI more for small molecules. In practical terms, which strategy is used often depends upon experimental objectives, expertise of selected technologies and availability of materials and equipment.

2 Preclinical studies on CRC

2.1 Cell lines in vitro models

Despite the fact that colon cancer cell lines have been widely used material sources in proteomic research addressing various scientific questions because of advantages of more

homogeneous cell population (e.g. cells of clonal origin, lack of tumor stromal or immune cells) and "limitless" supply. In the past 5 years, there are at least 20 studies published on global proteomic comparisons using popular colon cancer cell lines with following objectives: (i) isogenic lines [76–80]; (ii) metastatic versus poor metastatic comparison [81]; (iii) drug (namely 5-FU) treated versus nontreated [82]; (iv) secretomes [83]. However, the "unique" biomarkers identified from these proteomic comparison studies were validated not beyond immunoblotting or immunocytochemistry, in some cases, SRM/MRM-MS, while reports of clinical validation of the protein biomarkers identified from cell lines are generally lacking. However, the disadvantages of using cell lines for the identification of clinically applicable biomarkers are that: (i) they are generally derived from malignant tumors and do not represent the premalignant precursor stage (i.e. adenoma), (ii) they are grown in vitro on plastic under high serum and high oxygen conditions, and do not represent in vivo situation in which nonneoplastic cells (such as fibroblasts, endothelial cells, and immune cells) may also play a critical role in tumor development. With these taken into consideration, validation of the findings in clinical materials and settings is pivotal.

In an attempt to overcome, or at least partially, these shortcomings and mimic in vivo tumor microenvironment settings, Zeng et al. co-cultured colon cancer cell line HT29 with normal human colon mucosal epithelial cell line NCM460 in vitro to investigate the differential expression pattern of secretome [84]. In this study, a quantitative proteomics approach based on stable isotope labeling by amino acids in cell culture and LC-MS was used for secretome analysis, and a total of 45 proteins were altered over twofold in co-cultured cellular supernatants between equal amounts of NCM460 and HT29 cells, compared with mono-cultured conditions.

2.2 Animal in vivo models

The application of genetically engineered mouse models has widely been used to monitor cancer development and metastasis in vivo. In CRC, the most widely used model is the APC Min mouse that was generated by introducing mutations into the mouse *Apc* gene, a known suppressor in the Wnt/β-catenin signaling, whose human counterpart has implicated in the onset and progression of CRC [85]. However, the most frequent form of CRC in human is sporadic in origin (~85%). The frequency of familial adenomatous polyposis or *Apc* mutation only accounts for less than 0.1%. In addition, adenomas are often found in small intestine but not colon in Apc Min mice. Therefore, the representation of Apc Min mouse model for human CRC in search for novel biomarkers is becoming questionable. However, a handful of studies have applied such mouse model to study CRC-related protein biomarkers from either serum [86], adenoma tissues, or fecal samples [49] of the Apc Min/+ mice that developed spontaneously multiple intestinal neoplasia, and SDS-PAGE separation followed by nanoflow reversed-phased LC-MS/MS, and proximal fluid samples from $Apc^{15lox/+}$ mouse model collected from normal colon and colon tumors [65] were used to identify resultant proteins in these cases. Among the proteins identified in each dataset, only haptoglobin, hemoglobin, and hemopexin were uniquely present in both samples. In another study, cell adhesion protein cadherin-17 found in the fecal sample of Apc Min/+ was later confirmed to play a critical role in CRC-caused liver metastasis in human [87]. Xu et al. [88] compared protein profiles of samples collected from adenomas of Apc Min/+ versus normal top and bottom crypts of wild-type mice using laser capture microdissection combined with MALDI-MS, and found a set of proteins were differentially expressed. Among those, S100A8 was validated in human CRC suggesting a potential biomarker for early CRC detection. A similar study was conducted by Zhu et al. [89] in which adenoma from Apc Min/+ were collected together with adjacent normal mucosa instead of from normal crypts of wild-type mice, and the proteins in the lysates of these adenoma and normal tissues were identified and compared using 2D nano-LC-LTQ MS. In this study, 52 proteins were found to be differentially expressed: 27 upregulated and 25 downregulated in tumor subjects. To explore the expression network in which genes are co-expressed with the 27 upregulated genes, a co-expression network of 45 genes were revealed that is upregulated in colon tumors; many of them are associated with innate immunity and inflammation, but there are substantial differences between this study and other studies [49,90]. In an attempt to identify potential blood-based biomarkers indicative of early-stage intestinal cancers using the Apc Min/+ mouse model of intestinal cancer, Ivancic et al. [91] utilized serum proteins from tumor-bearing Apc Min/+ mice quantitatively comparing to tumor-free Apc +/+ wild-type mice via in animal metabolic labeling with 14N/15N-labeled Spirulina algae and an LTQ Orbitrap mass spectrometer. Out of 1116 serum proteins quantified, 40 proteins were determined to be differentially expressed and correlated with the increase in intestinal neoplasms. In addition, a subset of these 40 differentially expressed proteins underwent a secondary quantitative screen using SRM-MS with stable isotope-labeled peptides. Using both quantitative techniques, MGAM and COL1A1 as downregulated and ITIH3 and F5 as upregulated in serum were identified as potential biomarkers for CRC. With similar methods, the levels of proteins EGFR, LRG1, ITIH4, and F5 in serum were also identified to be highly correlated with the number of colonic adenomas in Apc^{Pirc}/+ rats [41] . Encouragingly, these four proteins resulted in a sensitivity of 100%, a specificity of 80%, and an AUC of 0.93 at 135 days of age, when the Pirc rats bore an average of 19 tumors in the colon and seven in the small intestine. These studies demonstrate that the quantitative analysis of a panel of serum proteins can detect the presence of early intestinal tumors in a rat model, and provide support for future measurements in human subjects.

Azoxymethane/dextran sodium sulfate or AOM-DSS-induced mouse model can also be used for mimicking sporadic, colitis-associated CRC in human [92]. Torres et al. [93]

used this mouse model to identify novel cancer-associated fibroblasts markers that might contribute to the invasion and the prognosis of CRC, and conducted in-depth quantitative proteomic analysis of both whole-cell extracts and supernatants, and further validated the upregulated proteins in cancer-associated fibroblasts by chemokine microarray and immunohistochemical analyses of mouse and human tissues. This study concludes that LTBP2, CDH11, OLFML3, and FSTL1 as selective biomarkers of cancer stroma, and CALU and CDH11 as candidate stromal biomarkers of prognostic significance to colon cancer.

3 Clinical studies on CRC

3.1 Clinical samples with patient tissue materials

Protein biomarkers that can be detected in blood or stool have been the primary focus for CRC early detection since they could be applied in a standard clinical setting alongside the routinely used markers such as carcino-embryonic antigen and carbohydrate antigen 19-9 (blood test) or hemoglobin (stool test). However, the challenge of identifying tumor-derived protein biomarkers in plasma is that their concentration is much lower than that of endogenous blood proteins such as albumin. One approach to overcome this challenge is to focus on tumor-excreted proteins, whose concentrations are highest in fluids in close proximity to the tumor. In this way, De Wit et al. [94] determined secretomes of four pairs of human CRC tissue and patient-matched normal colon tissue samples using GeLC-MS/MS, and identified 2703 proteins in the tissue secretomes, of which 76 candidates were based on consistent and abundant overrepresentation in cancer compared to control-secretomes, and presumed neoplastic origin; 21 biomarkers were revealed to be suited for early detection of CRC through overlap analysis with previously obtained datasets from CRC mice model. Further, one candidate marker (MCM5) was confirmed using immunohistochemistry after overexpression in CRC, which is also one of three critical genes (ALDH6A1 and TFF2) in two independent colon cancer cohorts (49 and 76 patients) for predicting overall responses to Bevacizumab and progression-free survival [95]. The FOBT is currently the first line method for CRC screening in clinics, but has an unacceptably low sensitivity and specificity. Improved screening tests are therefore urgently required for early-stage CRC screening. Using a hypothesis-driven approach for a rapid biomarker discovery process whereby 60 selected proteins previously implicated as CCAPs and 1D-SDS-PAGE analysis followed by direct identification and relative quantification using MRM, Ang and Nice [43] found 19 of these proteins were detected in the feces from a patient with CRC. Relative quantitation of these 19 CCAP across five CRC patients and five healthy volunteers were carried out, revealing hemoglobin, myeloperoxidase, S100A9, filamin A, and L-plastin to be present only in the feces of CRC patients. In a separate experiment using similar approach but with multidimensional fractionation (1D SDS-PAGE, RP-HPLC, size exclusion chromatography) process, the same group of scientists [67] confirmed that hemoglobin is a unique CCAP in human feces.

Screening patients at high risk of recurrence of cancer would allow for more accurate and personalized treatment. In this context, Kim et al. [96] determined the prognosis-related protein profile by two different quantitative proteomic techniques, differential in-gel electrophoresis (2D-DIGE) and cleavable isotope-coded affinity tag method with LC-MS/MS. In this study, six tumor tissues were obtained from stage IV CRC patients, of which three have survived more than 5 years (good prognostic group) and the other three died within 25 months (poor prognostic group) after palliative surgery and subsequent chemotherapy treatment. Among the 175 proteins identified with abundance ratios greater than twofold, five proteins, fatty acid binding protein 1, intelectin 1, transitional ER ATPase, transgelin, and tropomyosin 2, were revealed to be significantly different between the two prognostic groups within 95% confidence, while no single protein could completely distinguish the two groups from each other. However, a combination of the five proteins effectively distinguished poor prognostic group from good prognostic group patients (AUC = 1). In a different study conducted by O'Dwyer et al. [97], fresh frozen sections of paired Dukes B CRC and normal colorectal mucosa ($n = 28$) were analyzed, and 45 proteins were identified as at least 1.5 times increased expression in CRC, among which, 15 were validated by immunohistochemistry using a well-characterized CRC tissue microarray containing 515 primary CRC, 224 lymph node metastasis, and 50 normal colonic mucosal samples. The proteins that showed the greatest degree of overexpression in primary CRC compared with normal colonic mucosa were HSP 60 ($p < 0.001$), S100A9 ($p < 0.001$). Protein 14-3-3b was identified as a prognostic biomarker ($x^2 = 6.218$, $p = 0.013$, hazard ratio (HR) = 0.639, 95% confidence interval (CI): 0.448–0.913). In another effort, Morita et al. [98] found that HSP 40 family member such as DNAJB8 was highly expressed in CRC, and overexpression of DNAJB8 enhanced the expression in tumorigenicity indicating that DNAJB8 played a major role in CRC prognosis. More recently, Mornon et al. [99] quantified two putative tumor markers DcR3 and growth/differentiation factor GDF15 from sera samples obtained from 100 patients undergoing standard surgical procedures for primarily diagnosed with CRC, and found serum levels were significantly elevated in tumor patients for DcR3 (116.94 ± 57.37 fmol/mL) and GDF15 (164.44 ± 79.31 fmol/mL) in comparison to three healthy human serum samples (DcR3: 27.23 ± 2.49 fmol/mL; GDF15: 98.11 ± 0.49 fmol/mL), which were in good agreement with ELISA and qPCR measurements.

As for large-scale identification of biomarkers, label-free quantitative proteomics has often been used besides labeling methods. In comparison of nonmalignant and malignant colorectal tissues from patients, Knol et al. [100] examined tissue fractions containing chromatin-binding (CB) proteins

using label-free LC-MS/MS. Over 1700 proteins were identified in the CB fraction from colonic tissues, among which 169 proteins were differential between adenomas and carcinomas indicating that there is a specific shift in the chromatin landscape going from adenomas to carcinomas. Unlike Knol et al.'s study focusing on nuclear CB proteins, De Wit et al. [42] studied differential expression of cell surface proteins in colon adenoma-to-carcinoma progression. Their study set out to examine cell surface proteins of five CRC cell lines. Membrane-bound proteins were biotinylated, isolated, and analyzed using gel electrophoresis and nano-LC coupled to MS/MS. In total, 2609 proteins were identified in the cell surface fractions. Of these, 44 proteins were selected as promising cell surface candidate biomarkers for adenoma-to-carcinoma progression based on the following criteria: protein identification in at least four out of five cell lines, a predicted transmembrane location and increased mRNA expression in CRCs compared to adenomas. High expression of glucose transporter type 1 protein (gene symbol *SLC2A1*; $p < 0.00001$, 66 adenomas vs. 56 carcinomas) and prion protein (gene symbol *PRNP*; $p < 0.005$, 75 adenomas vs. 68 carcinomas) was later confirmed to be associated with high-risk adenomas and carcinomas by immunohistochemistry in patient tissue samples.

Lipid profiling representing an attractive avenue for novel cancer biomarker discovery has also been explored. To date, there are a limited number of studies specifically focusing on lipid signatures with respect to CRC biomarkers. Using MS imaging approach, Thomas et al. found a panel of lipid-based biomarkers to be up- and downregulated in CRC liver metastases [101]. In a recent study by Mirnezami et al. [102], lipid profiles of fresh frozen sections of CRC tissue and adjacent healthy mucosa obtained from 12 consenting patients were mapped for CRC microenvironment via partial least squares pattern recognition-based MALDI imaging MS (MALDI-MSI). These results demonstrated that CRC tissue harbors characteristic phospholipids signatures compared with healthy tissue, and different tissue regions within the CRC tumor microenvironments exhibited distinct biochemical profiles, which revealed novel cancer-associated field effects.

3.2 Colon CSC

The CSC hypothesis postulates that CSCs are a small and unique subset of tumor-initiating cells present in tumor mass responsible for tumor initiation, propagation, metastases, and resistance to treatment leading to disease relapse following surgery and/or chemotherapy and radiotherapy. Different from the traditional, stochastic model of tumorigenesis or the clonal evolution model, the CSCs are in a hierarchical organization for sustaining tumorigenesis and establishing the cellular heterogeneity inherent in the primary tumor. There is increased evidence that the nature of tumor cells is highly complex and dynamic suggesting the clonal evolution and CSC models could act in concert. Nevertheless, the CSC hypothesis therefore raises questions regarding current diagnostic and therapeutic modalities, suggesting that the CSC could be a rational target for the development of more efficacious screening, early detection, prevention, treatment and surveillance modalities, and interventions. In CRC, several putative CSC markers have been reported [103]: CD133, CD24, CD26, CD29, CD44, CD166 (ALCAM), EpCAM, Lgr5, ALDH1A1, and ALDH1B1, which are either cell surface glycoproteins responsible for cell adhesion and migration, or cytoplasmic proteins involved in oxidation of intracellular aldehydes (detoxification). However, the problems are that these biomarkers are not colon CSC specific, they can also be found in other solid tumor stem cell populations. For examples, CD133 was initially described as a CSC marker for glioblastoma multiforme [104]. CD44 and CD24 have been extensively used to isolate CSCs in breast cancer [105]. Regardless of their stemless properties, some of these markers, CD26 [106], CD44, EpCAM, and ALDH, are likely associated with metastasis and prognostics of the CRC in patients. All these putative CSC surface markers could be used to fractionate specific CSC populations for MS analysis.

Nevertheless, in proteomic stem cell research, extensive fractionation comes at the expense of sample losses, hampering the analysis of very limited materials. Di Palma et al. [107] described a highly sensitive multidimensional chromatographic strategy based on a combination of hydrophilic interaction LC and RP chromatography, which allows proteomic analysis with minimal sample losses. They applied this strategy to the analysis of a limited number of Lgr5+ FACS-sorted colon stem cells extracted from mouse intestine, obtaining proteome coverage comparable to current methods that generally require 100-fold more starting material. This multidimensional chromatographic technology will find ample applications such as in the analysis of distinct cellular populations obtained by laser microdissection.

4 Concluding remarks

In the past decade, there are numerous claimed biomarker candidates in the scientific literatures for CRC alone [69, 108], but none of them has been approved by FDA for clinical application. In turn, this discrepancy inevitably raises the question in the research community: What causes the congestion between discovery and validation? One of the obvious explanations is the lack of the large scale of validation which requires alternative and robust assays including multiplex immunoassay. Antibodies for novel protein candidates identified from shotgun proteomics do not always exist in many cases, especially for PTM markers, or are difficult to develop without having significant interferences and cross-reactivity. In addition, many identified biomarkers have never been validated or verified on independent cohorts. These are the major reasons why many candidate biomarkers do not get beyond a proof of concept phase. In addition, other gap barriers prevent

transition of laboratory discovery to clinical practices including: (i) measurement inconsistency and a lack of reproducibility within/across proteomic platforms; (ii) a lack of knowledge by the research community regarding the analytical evaluation criteria required by the regulatory agency to guide the processes through the approval pipeline; and (iii) a lack of publicly accessible, high-quality affinity reagents, reference materials, and datasets for data mining, hypotheses generation, and experimental validation prior to clinical validation. Fortunately, the establishment of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) aims to close the gap barriers and facilitate the transition.

The limitations of appropriate antibody-based immunoassays in biomarker verification and validation have spurred the development of alternative approaches for more sensitive, accurate, and target-descriptive and cost-effective assays. Two main targeted MS-based technologies currently applied by the proteomics community for validation potentially replacing immunoassays are SIM-MS and MRM-MS. Both methodologies enable MS-based strategies to focus on a particular peptide or a set of peptides and allow it to monitor one or more of its sequence-specific fragment ions, or equivalently peptide transitions [109]. They are particularly powerful when coupled to ultra performance LC (UPLC) because UPLC-SIM/MRM-MS could potentially monitor hundreds of peptide transitions combined with their quantitative analysis in a single experiment. However, a main limitation of SIM/MRM-based targeted quantification is the lack of sufficient sensitivity such as antibodies for quantification of very low-abundance proteins or protein modifications in a very complex biomedical matrix [110]. UPLC-SRM-MS can typically detect low to moderately abundant proteins in human blood plasma/serum with concentrations at the low microgram per milliliter or high nanogram per milliliter levels without the application of front-end fractionation and/or enrichment [110–112]. Particularly, stable isotope standards and capture by antipeptide antibodies will conduce to the enrichment of lower abundance proteins in plasma for increasing sensitivity of MS-based assays, and finally allow the replacement of a rapid bind/elute process for the time-consuming reverse phase separation applied as a prelusion to online MS peptide assays [113].

Although the detection of selected plasma proteins below 100 ng/mL levels has been reported in several recent multiplexed targeted quantification efforts in plasma without depletion [114, 115], the majority of plasma proteins in the low nanogram per milliliter range are still not detectable in nondepleted plasma by direct LC-SRM-MS. Many candidate protein biomarkers of proven clinical importance are present at the picogram per milliliter to low nanogram per milliliter levels in human plasma/serum; thus, well below the LOD for conventional LC-SRM.

It is worthy to mention that more sensitive and powerful methods in translational proteomics analysis are still ongoing efforts to overcome the challenges in the analysis of very low amount of specific protein markers of interest like some very low abundant protein-kinases in vivo. An effective and practical combination of the optimized enrichment/depletion methods for targeted proteins and peptides, the improved fractionation techniques using various chromatography approaches, and the tailored quantitative methods based on UPLC-SRM-MS, a novel quadrupole-orbitrap instrument operated in PRM-MS/MS mode (parallel reaction monitoring) [116]with newer, narrower, and variable window protocols for SWATH-DIA (Serial Window Acquisition Strategy-Data-Independent Acquisition) [117] as well as the associated bioinformatics tools will greatly boost large-scale and high-throughput quantitative analysis of these biomarkers, and thus significantly facilitate the achievement of our ultimate goals of the discovery of improved biomarker candidates that can be verified preclinically, and ultimately used in clinical studies.

More recently, the utility of MS-based proteomics and its translational clinical applications has been increasingly recognized due to their high sensitivity, specificity, and throughput. MS-based translational proteomics has been applied in a wide range of biological and biomedical investigations from early disease diagnosis to new drug targets for drug development and therapeutic intervention. In this regard, we strongly believe that the subtle integration of translational proteomics and most recent multidisciplinary advances in biological and biomedical fields has made decisive contribution to and will continue to promote and revive the proteomic studies on CRC protein biomarkers or to discover and develop new therapeutic targets in the CRC biomedical context based on a better understanding of key mechanisms underlying normal and diseased states of cells in the body.

This work was jointly supported by The Sino-Africa Joint Research Center project (Grant No. SAJC201328 to M. Guo) and "the Hundred Talents Program" from Chinese Academy of Sciences (Grant No. 29Y429291a0129 to M. Guo). These funding agents played no direct roles in the study design, data collection and analysis, and decision to publish, or preparation of the manuscript.

The authors have declared no conflict of interest.

5 References

[1] Bandrés, E., Zárate, R., Ramirez, N., Abajo, A. et al., Pharmacogenomics in colorectal cancer: the first step for individualized-therapy. *World J. Gastroenterol.* 2007, *13*, 5888–5901.

[2] Jemal, A., Siegel, R., Ward, E., Hao, Y. et al., Cancer statistics, 2009. *CA Cancer J. Clin.* 2009, *59*, 225–249.

[3] Wood, L. D., Parsons, D. W., Jones, S., Lin, J. et al., The genomic landscapes of human breast and colorectal cancers. *Science* 2007, *318*, 1108–1113.

[4] Jones, S., Zhang, X., Parsons, D. W., Lin, J. et al., Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 2008, *321*, 1801–1806.

[5] Parsons, D. W., Jones, S., Zhang, X., Lin, J. et al., An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008, *321*, 1807–1812.

[6] Goel, A., Boland, C. R., Recent insights into the pathogenesis of colorectal cancer. *Curr. Opin. Gastroenterol.* 2010, *26*, 47–52.

[7] Cancer Genome Atlas Network, Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012, *487*, 330–337.

[8] Jones, S., Chen, W. D., Parmigiani, G., Diehl, F. et al., Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. USA* 2008, *105*, 4283–4288.

[9] Vecchione, L., Saridaki, Z., Tejpar, S., Clinical implications and quality assurance of molecular testing for EGFR-targeting agents in colorectal cancer. *Curr. Colorectal. Cancer Rep.* 2011, *8*, 42–50.

[10] Sithambaram, S., Hilmi, I., Goh, K. L., The diagnostic accuracy of the M2 pyruvate kinase quick stool test—a rapid office based assay test for the detection of colorectal cancer. *PLoS One* 2015, *10*, e0131616.

[11] Siegel, R., Desantis, C., Jemal, A., Colorectal cancer statistics, 2014. *CA Cancer J. Clin.* 2014, *64*, 104–117.

[12] Greenberg, P. D., Bertario, L., Gnauck, R., Kronborg, O. et al., A prospective multicenter evaluation of new fecal occult blood tests in patients undergoing colonoscopy. *Am. J. Gastroenterol.* 2000, *95*, 1331–1338.

[13] Sidransky, D., Tokino, T., Hamilton, SR, Kenneth, W. K. et al., Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science* 1992, *256*, 102–105.

[14] Dong, S. M., Traverso, G., Johnson, C., Li, G. et al., Detecting colorectal cancer in stool with the use of multiple genetic targets. *J. Natl. Cancer Inst.* 2001, *93*, 858–865.

[15] Imperiale, T. F., Ransohoff, D., Itzkowitz, S. H., Turnbull, B. A. et al., Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. *New Engl. J. Med.* 2004, *351*, 2704–2714.

[16] Markovic, S., Antic, J., Dimitrijevic, I., Zogovic, B. et al., Microsatellite instability affecting the T17 repeats in intron 8 of HSP110, as well as five mononucleotide repeats in patients with colorectal carcinoma. *Biomarkers Med.* 2013, *7*, 613–621.

[17] Mathivanan, S., Ji, H., Tauro, B. J., Chen, YS et al., Identifying mutated proteins secreted by colon cancer cell lines using mass spectrometry. *J. Proteomics* 2012, *76*, 141–149.

[18] Zhang, B., Wang, J., Wang, X., Zhu, J. et al., Proteogenomic characterization of human colon and rectal cancer. *Nature* 2014, *513*, 382–387.

[19] Courtney, R. J., Pall, C. L., Fisher, R. W. S., Macrae, F. A. et al., Individual- and provider-level factors associated with colorectal cancer screening in accordance with guideline recommendation: a community-level perspective across varying levels of risk. *BMC Public Health* 2013, *13*, 248–257.

[20] Locker, G. Y., Hamilton, S., Harris, J., Jessup, J. M. et al., ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J. Clin. Oncol.* 2006, *24*, 5313–5327.

[21] Coghlin, C., Murray, G. I., Biomarkers of colorectal cancer: recent advances and future challenges. *Proteom Clin. App.* 2015, *9*, 64–71.

[22] Qiu, Y. H., Patwa, T. H., Xu, L., Shedden, K. et al., Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot. *J. Proteome Res.* 2008, *7*, 1693–1703.

[23] Peng, Y., Li, X. Y., Wu, M. H., Yang, J. et al., New prognosis biomarkers identified by dynamic proteomic analysis of colorectal cancer. *Mol. Biosyst.* 2012, *8*, 3077–3088.

[24] Besson, D., Pavageau, A. H., Valo, I., Bourreau, A. 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*, 1–14.

[25] Meding, S., Balluff, B., Elsner, M., Schone, C. et al., Tissue-based proteomics reveals FXYD3, S100A11 and GSTM3 as novel markers for regional lymph node metastasis in colon cancer. *J. Pathol.* 2012, *228*, 459–470.

[26] Croner, R. S., Sturzl, M., Rau, T. T., Metodieva, G. et al., Quantitative proteome profiling of lymph node-positive vs. -negative colorectal carcinomas pinpoints MX1 as a marker for lymph node metastasis. *Int. J. Cancer* 2014, *135*, 2878–2886.

[27] Jankova, L., Chan, C., Fung, C. L., Song, X. et al., Proteomic comparison of colorectal tumours and non-neoplastic mucosa from paired patient samples using iTRAQ mass spectrometry. *Mol. Biosyst.* 2011, *7*, 2997–3005.

[28] Tan, K. L., Jankova, L., Chan, C., Fung, C. L. et al., Clinicopathological correlates and prognostic significance of glutathione S-transferase Pi expression in 468 patients after potentially curative resection of node-positive colonic cancer. *Histopathology* 2011, *59*, 1057–1070.

[29] Jankova, L., Robertson, G., Chan, C., Tan, K. L. et al., Glutathione S-transferase Pi expression predicts response to adjuvant chemotherapy for stage C colon cancer: a matched historical control study. *BMC Cancer* 2012, *12*, 196–204.

[30] Schwarzenbach, H., Hoon, D. S., Pantel, K., Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer* 2011, *11*, 426–437.

[31] McKinley, E. T., Liu, H., McDonald, W. H., Luo, W. et al., Global phosphotyrosine proteomics identifies PKCdelta as a marker of responsiveness to Src inhibition in colorectal cancer. *PLoS ONE* 2013, *8*, e80207.

[32] Sveen, A., Agesen, T. H., Nesbakken, A., Meling, G. I. et al., ColoGuidePro: a prognostic 7-gene expression signature for stage III colorectal cancer patients. *Clin. Cancer Res.* 2012, *18*, 6001–6010.

[33] Salazar, R., Roepman, P., Capella, G., Moreno, V. et al., Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. *J. Clin. Oncol.* 2010, *29*, 17–24.

[34] Maak, M., Simon, I., Nitsche, U., Roepman, P. et al., Independent validation of a prognostic genomic signature (ColoPrint) for patients with stage II colon cancer. *Ann. Surg.* 2013, *257*, 1053–1058.

[35] Coghlin, C., Carpenter, B., Dundas, S. R., Lawrie, L. C. et al., Characterization and over-expression of chaperonin t-complex proteins in colorectal cancer. *J. Pathol.* 2006, *210*, 351–357.

[36] Curran, S., Dundas, S. R., Buxton, J., Leeman, M. F. et al., Matrix metalloproteinase/tissue inhibitors of matrix metalloproteinase phenotype identifies poor prognosis colorectal cancers. *Clin. Cancer Res.* 2004, *10*, 8229–8234.

[37] Murray, G. I., Duncan, M. E., O'Neil, Pa., Melvin, W. T. et al., Matrix metalloproteinase−1 is associated with poor prognosis in colorectal cancer. *Nat. Med.* 1996, *2*, 461–462.

[38] Kumarakulasingham, M., Rooney, P. H., Dundas, S.R., Telfer, C. et al., Cytochrome p450 profile of colorectal cancer: identification of markers of prognosis. *Clin. Cancer Res.* 2005, *11*, 3758–3765.

[39] O'Dwyer, D., Ralton, L. D., O'Shea, A., Murray, G. I., The proteomics of colorectal cancer: identification of a protein signature associated with prognosis. *PLoS ONE* 2011, *6*, e27718.

[40] Duncan, R., Carpenter, B., Main, L. C., Telfer, C., Murray, G. I., Characterisation and protein expression profiling of annexins in colorectal cancer. *Brit. J. Cancer* 2008, *98*, 426–433.

[41] Manna, S. K., Tanaka, N., Krausz, K. W., Haznadar, M. et al., Biomarkers of coordinate metabolic reprogramming in colorectal tumors in mice and humans. *Gastroenterology* 2014, *146*, 1313–1324.

[42] De Wit, M., Jimenez, C. R., Carvalho, B., Belien, J.A. et al., Cell surface proteomics identifies glucose transporter type 1 and prion protein as candidate biomarkers for colorectal adenoma-to-carcinoma progression. *Gut* 2012, *61*, 855–864.

[43] Ang, C. S., Nice, E. C., Targeted in-gel MRM a hypothesis driven approach for colorectal cancer biomarker discovery in human feces. *J. Proteome Res.* 2010, *9*, 4346–4355.

[44] Friedman, D. B., Hill, S., Keller, J. W., Merchant, N. B. et al., Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics* 2004, *4*, 793–811.

[45] Ward, D. G., Suggett, N., Cheng, Y., Wei, W. et al., Identification of serum biomarkers for colon cancer by proteomic analysis. *Brit. J. Cancer* 2006, *94*, 1898–1905.

[46] Poullis, A., Foster, R., Northfliield, T. C., Mendall, M.A., Faecal markers in the assessment of activity in inflammatory bowel disease. *Aliment Pharmacol. Ther.* 2002, *16*, 675–681.

[47] Polley, A. C., Mulholland, F., Pin, C., Williams, E. A. et al., Proteomic analysis reveals field-wide changes in protein expression in the morphologically normal mucosa of patients with colorectal neoplasia. *Cancer Res.* 2006, *66*, 6553–6562.

[48] Quaresima, B., Crugliano, T., Gaspari, M., Faniello, M. C. et al., A proteomics approach to identify changes in protein profiles in serum of Familial Adenomatous Polyposis patients. *Cancer Lett.* 2008, *272*, 40–52.

[49] Ang, C. S., Rothacker, J., Patsiouras, H., Burgess, A. W. et al., Murine fecal proteomics: a model system for the detection of potential biomarkers for colorectal cancer. *J. Chromatogr. A* 2010, *1217*, 3330–3340.

[50] Osborn, N. K., Ahlquist, D. A., Stool screening for colorectal cancer: molecular approaches. *Gastroenterology* 2005, *128*, 192–206.

[51] Habermann, J. K., Roblick, U. J., Luke, B. T., Prieto, D. A. et al., Increased serum levels of complement C3a anaphylatoxin indicate the presence of colorectal tumors. *Gastroenterology* 2006, *131*, 1020–1029; quiz 1284.

[52] Ciordia, S., de Los Rios, V., Albar, J. P., Contributions of advanced proteomics technologies to cancer diagnosis. *Clin. Transl. Oncol.* 2006, *8*, 566–580.

[53] Pei, H. P., Zhu, H., Zeng, S., Li, Y. X. et al., Proteome analysis and tissue microarray for profiling protein. *J. Proteome Res.* 2007, *6*, 2495–2501.

[54] Rodriguez-Pineiro, A. M., Ayude, D., Rodriguez-Berrocal, F. J., Paez de la Cadena, M., Concanavalin A chromatography coupled to two-dimensional gel electrophoresis improves protein expression studies of the serum proteome. *J. Chromatogr. B* 2004, *803*, 337–343.

[55] Drew, J. E., Rucklidge, G. J., Duncan, G., Lufty, A. et al., A proteomics approach to identify changes in protein profiles in pre-cancerous colon. *Biochem. Biophys. Res. Commun.* 2005, *330*, 81–87.

[56] Velcich, A., Yang, W. C., Heyer, J., Fragale, A. et al., Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science* 2002, *295*, 1726–1729.

[57] Okazaki, N., Takahashi, N., Kojima, S. I., Masuho Y. et al., Protocadherin LKC, a new candidate for a tumor suppressor of colon and liver cancers, its association with contact inhibition of cell proliferation. *Carcinogenesis* 2002, *23*, 1139–1148.

[58] Fentz, A. K., Sporl, M., Spangenberg, J., List, H. J. et al., Detection of colorectal adenoma and cancer based on transthyretin and C3a-desArg serum levels. *Proteom. Clin. Appl.* 2007, *1*, 536–544.

[59] Xue, H., Lu, B., Zhang, J., Wu, M. L et al., Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. *J. Proteome Res.* 2010, *9*, 545–555.

[60] Xie, L. Q., Zhao, C., Cai, S. J., Xu, Y. et al., Novel proteomic strategy reveal combined α1 antitrypsin and cathepsin D as biomarkers for colorectal cancer early screening. *J. Proteome Res.* 2010, *9*, 4701–4709.

[61] Murakoshi, Y., Honda, K., Sasazuki, S., Ono, M. et al., Plasma biomarker discovery and validation for colorectal cancer by quantitative shotgun mass spectrometry and protein microarray. *Cancer Sci.* 2011, *102*, 630–638.

[62] Han, C. L., Chen, J. S., Chan, E. C., Wu, C. P. et al., An informatics-assisted label-free approach for personalized tissue membrane proteomics: case study on colorectal cancer. *Mol. Cell. Proteomics* 2011, *10*, 1–15.

[63] Hamelin, C., Cornut, E., Poirier, F., Pons, S. et al., Identification and verification of heat shock protein 60 as a

potential serum marker for colorectal cancer. *FEBS J.* 2011, *278*, 4845–4859.

[64] Tsai, M. H., Wu, C. C., Peng, P. H., Liang, Y. et al., Identification of secretory gelsolin as a plasma biomarker associated with distant organ metastasis of colorectal cancer. *J. Mol. Med.* 2012, *90*, 187–200.

[65] Fijneman, R. J., De Wit, M., Pourghiasian, M., Piersma, S. R. et al., Proximal fluid proteome profiling of mouse colon tumors reveals biomarkers for early diagnosis of human colorectal cancer. *Clin. Cancer Res.* 2012, *18*, 2613–2624.

[66] Ang, C. S., Phung, J., Nice, E. C., The discovery and validation of colorectal cancer biomarkers. *Bio. Chromatogr. BMC* 2011, *25*, 82–99.

[67] Ang, C. S., Rothacker, J., Patsiouras, H., Gibbs, P. et al., Use of multiple reaction monitoring for multiplex analysis of colorectal cancer-associated proteins in human feces. *Electrophoresis* 2011, *32*, 1926–1938.

[68] Rice, G. E., Georgiou, H. M., Ahmed, N., Shi, G. et al., Translational proteomics: developing a predictive capacity—a review. *Placenta* 2006, *27*(Suppl. A), S76–S86.

[69] Jimenez, C. R., Knol, J. C., Meijer, G. A., Fijneman, R. J., Proteomics of colorectal cancer: overview of discovery studies and identification of commonly identified cancer-associated proteins and candidate CRC serum markers. *J. Proteomics* 2010, *73*, 1873–1895.

[70] De Wit, M., Fijneman, R. J., Verheul, H. M., Meijer, G. A. et al., Proteomics in colorectal cancer translational research: biomarker discovery for clinical applications. *Clin. Biochem.* 2013, *46*, 466–479.

[71] Glaros, T. G., Blancett, C. D., Bell, T. M, Natesan, M. et al., Serum biomarkers of *Burkholderia mallei* infection elucidated by proteomic imaging of skin and lung abscesses. *Clin. Proteom.* 2015, *12*, 7. doi:10.1186/s12014-015-9079-4

[72] Poli, G., Ceni, E., Armignacco, R., Ercolino, T. et al., 2D-DIGE proteomic analysis identifies new potential therapeutic targets for adrenocortical carcinoma. *Oncotarget* 2015, *6*, 1–11.

[73] Moradian, A., Kalli, A., Sweredoski, M. J., Hess, S., The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics* 2014, *14*, 489–497.

[74] Keshishian, H., Addona, T., Burgess, M., Mani, D. R. et al.,Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* 2009, *8*, 2339–2349.

[75] Whiteaker, J. R., Zhao, L., Anderson, L, Paulovich, A. G., An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers. *Mol. Cell. Proteomics* 2010, *9*, 184–196.

[76] Ali, N. A., McKay, M. J., Molloy, M. P., Proteomics of Smad4 regulated transforming growth factor-beta signalling in colon cancer cells. *Mol. Biosyst.* 2010, *6*, 2332–2338.

[77] Flavin, P., Redmond, A., McBryan, J., Cocchiglia, S. et al., RuvBl2 cooperates with Ets2 to transcriptionally regulate hTERT in colon cancer. *FEBS Lett.* 2011, *585*, 2537–2544.

[78] Halvey, P. J., Zhang, B., Coffey, R. J., Liebler, D. C. et al., Proteomic consequences of a single gene mutation in a colorectal cancer model. *J. Proteome Res.* 2012, *11*, 1184–1195.

[79] Ikonomou, G., Kostourou, V., Shirasawa, S., Sasazuki, T. et al., Interplay between oncogenic K-Ras and wild-type H-Ras in Caco2 cell transformation. *J. Proteomics* 2012, *75*, 5356–5369.

[80] Pocsfalvi, G., Votta, G., De Vincenzo, A., Fiume, I. et al., Analysis of secretome changes uncovers an autocrine/paracrine component in the modulation of cell proliferation and motility by c-Myc. *J. Proteome Res.* 2011, *10*, 5326–5337.

[81] Luque-Garcia, J. L., Martinez-Torrecuadrada, J. L., Epifano, C., Canamero, M. et al., Differential protein expression on the cell surface of colorectal cancer cells associated to tumor metastasis. *Proteomics* 2010, *10*, 940–952.

[82] Bauer, K. M., Lambert, P. A., Hummon, A. B., Comparative label-free LC-MS/MS analysis of colorectal adenocarcinoma and metastatic cells treated with 5-fluorouracil. *Proteomics* 2012, *12*, 1928–1937.

[83] Wu, C. C., Hsu, C. W., Chen, C. D., Yu, C. J. et al., Candidate serological biomarkers for cancer identified from the secretomes of 23 cancer cell lines and the human protein atlas. *Mol. Cell. Proteomics* 2010, *9*, 1100–1117.

[84] Zeng, X., Yang, P., Chen, B., Jin, X. et al., Quantitative secretome analysis reveals the interactions between epithelia and tumor cells by in vitro modulating colon cancer microenvironment. *J. Proteomics* 2013, *89*, 51–70.

[85] Guo, H., Nagy, T., Pierce, M., Post-translational glycoprotein modifications regulate colon cancer stem cells and colon adenoma progression in Apc(min/+) mice through altered Wnt receptor signaling. *J. Biol. Chem.* 2014, *289*, 31534–31549.

[86] Hung, K. E., Kho, A. T., Sarracino, D., Richard, L. G. et al., Mass spectrometry-based study of the plasma proteome in a mouse intestinal tumor model. *J. Proteome Res.* 2006, *5*, 1866–1878.

[87] Bartolome, R. A., Barderas, R., Torres, S., Fernandez-Acenero, M. J. et al., Cadherin-17 interacts with alpha2beta1 integrin to regulate cell proliferation and adhesion in colorectal cancer cells causing liver metastasis. *Oncogene* 2014, *33*, 1658–1669.

[88] Xu, B. J., Li, J., Beauchamp, R. D., Shyr, Y. et al., Identification of early intestinal neoplasia protein biomarkers using laser capture microdissection and MALDI MS. *Mol. Cell. Proteomics* 2009, *8*, 936–945.

[89] Zhu, W., Fang, C., Gramatikoff, K., Niemeyer, C. C. et al., Proteins and an inflammatory network expressed in colon tumors. *J. Proteome Res.* 2011, *10*, 2129–2139.

[90] Huttlin, E. L., Chen, X., Barrett-Wilt, G. A., Hegeman, A. D. et al., Discovery and validation of colonic tumor-associated proteins via metabolic labeling and stable isotopic dilution. *Proc. Natl. Acad. Sci. USA* 2009, *106*, 17235–17240.

[91] Ivancic, M. M., Huttlin, E. L., Chen, X., Pleiman, J. K. et al., Candidate serum biomarkers for early intestinal cancer using 15N metabolic labeling and quantitative proteomics in

the ApcMin/+ mouse. *J. Proteome Res.* 2013, *12*, 4152–4166.

[92] Kang, Y. James, *Methods in Pharmacology and Toxicology*, Humana Press, New Jersey USA 2014.

[93] Torres, S., Bartolome, R. A., Mendes, M., Barderas, R. et al., Proteome profiling of cancer-associated fibroblasts identifies novel proinflammatory signatures and prognostic markers for colorectal cancer. *Clin. Cancer Res.* 2013, *19*, 6006–6019.

[94] De Wit, M., Kant, H., Piersma, S. R., Pham, T. V. et al., Colorectal cancer candidate biomarkers identified by tissue secretome proteome profiling. *J. Proteomics* 2014, *99*, 26–39.

[95] Pentheroudakis, G., Kotoula, V., Fountzilas, E., Kouvatseas, G. et al., A study of gene expression markers for predictive significance for bevacizumab benefit in patients with metastatic colon cancer: a translational research study of the Hellenic Cooperative Oncology Group (HeCOG). *BMC Cancer* 2014, *14*, 111–120.

[96] Kim, H. J., Kang, U. B., Lee, H., Jung, J. H. et al., Profiling of differentially expressed proteins in stage IV colorectal cancers with good and poor outcomes. *J. Proteomics* 2012, *75*, 2983–2997.

[97] O'Dwyer, D., Ralton, L. D., O'Shea, A., Murray, G. I., The proteomics of colorectal cancer: identification of a protein signature associated with prognosis. *PLoS ONE* 2011, *6*, e27718.

[98] Schuijers, J., van der Flier, L. G., van Es, J., Clevers, H., Robust cre-mediated recombination in small intestinal stem cells utilizing the olfm4 locus. *Stem Cell Reports* 2014, *3*, 234–241.

[99] Mornon, J. P., Hoffmann, B., Jonic, S., Lehn, P. et al., Full-open and closed CFTR channels, with lateral tunnels from the cytoplasm and an alternative position of the F508 region, as revealed by molecular dynamics. *Cell. Mol. Life Sci.* 2015, *72*, 1377–1403.

[100] Knol, J. C., De Wit, M., Albrethsen, J., Piersma, S. R. et al., Proteomics of differential extraction fractions enriched for chromatin-binding proteins from colon adenoma and carcinoma tissues. *Bioch. Bioph. Acta* 2014, *1844*, 1034–1043.

[101] Thomas, A., Patterson, N. H., Marcinkiewicz, M. M., Lazaris, A. et al., Histology-driven data mining of lipid signatures from multiple imaging mass spectrometry analyses: application to human colorectal cancer liver metastasis biopsies. *Anal. Chem.* 2013, *85*, 2860–2866.

[102] Mirnezami, R., Spagou, K., Vorkas, P. A., Lewis, M. R. et al., Chemical mapping of the colorectal cancer microenvironment via MALDI imaging mass spectrometry (MALDI-MSI) reveals novel cancer-associated field effects. *Mol. Oncol.* 2014, *8*, 39–49.

[103] Muraro, M. G., Mele, V., Daster, S., Han, J. et al., CD133+, CD166+CD44+, and CD24+CD44+ phenotypes fail to reliably identify cell populations with cancer stem cell functional features in established human colorectal cancer cell lines. *Stem Cells Transl. Med.* 2012, *1*, 592–603.

[104] Wilson, R. J., Thomas, C. D., Fox, R., Roy, D. B. et al., Spatial patterns in species distributions reveal biodiversity change. *Nature* 2004, *432*, 393–396.

[105] Chekhun, S.V., Zadvorny, T. V., Tymovska, Yu.O., Anikusko, M. F. et al., CD44+/CD24- markers of cancer stem cells in patients with breast cancer of different molecular subtypes. *Exp. Oncol.* 2015, *37*, 58–63.

[106] Pang, R., Law, W. L., Chu, A. C., Poon, J. T. et al., A subpopulation of CD26+ cancer stem cells with metastatic capacity in human colorectal cancer. *Cell Stem Cell* 2010, *6*, 603–615.

[107] Di Palma, S., Stange, D., van de Wetering, M., Clevers, H. et al., Highly sensitive proteome analysis of FACS-sorted adult colon stem cells. *J. Proteome Res.* 2011, *10*, 3814–3819.

[108] Yin, H. R., Zhang, L., Xie, L. Q., Huang, L. Y. et al., Hyperplex-MRM: a hybrid multiple reaction monitoring method using mTRAQ/iTRAQ labeling for multiplex absolute quantification of human colorectal cancer biomarker. *J. Proteome Res.* 2013, *12*, 3912–3919.

[109] Lange, V., Picotti, P., Domon, B., Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 2008, *4*, 222–235.

[110] Shi, T., Su, D., Liu, T., Tang, K. et al., Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. *Proteomics* 2012, *12*, 1074–1092.

[111] Anderson, L., Hunter, C. L., Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* 2006, *5*, 573–588.

[112] Whiteaker, J. R., Zhao, L., Anderson, L., Paulovich, A. G., An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers. *Mol. Cell. Proteomics* 2010, *9*, 184–196.

[113] Andeson, N. L., Andeson, N. G., Haines, L. R., Hardie, D. B. et al., Mass spectrometric quantitation of peptides and proteins using stable isotope standards and capture by anti-peptide antibodies (SISCAPA). *J. Proteome Res.* 2004, *3*, 235–244.

[114] Domanski, D., Percy, A. J., Yang, J., Chambers, A. G. et al., MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. *Proteomics* 2012, *12*, 1222–1243.

[115] Huttenhain, R., Soste, M., Selevsek, N., Rost, H. et al., Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci. Transl. Med.* 2012, *4*, 142–166.

[116] Gallien, S., Duriez, E., Demeure, K., Domon, B., Selectivity of LC-MS/MS analysis: implication for proteomics experiments. *J. Proteome Res.* 2013, *81*, 148–158.

[117] Parker, S. J., Raedschelders, K., Van Eyk, J. E., Emerging proteomic technologies for elucidating context-dependent cellular signaling events: a big challenge of tiny proportions. *Proteomics* 2015, *15*, 1486–1502.

