PO Box 2345, Beijing 100023, China
www.wjgnet.com
wjg@wjgnet.com

World J Gastroenterol 2006 November 7; 12(41): 6594-6601
World Journal of Gastroenterology ISSN 1007-9327

EDITORIAL

Current status and prospects of clinical proteomics studies on detection of colorectal cancer: Hopes and fears

ME de Noo, RAEM Tollenaar, AM Deelder, LH Bouwman

ME de Noo, RAEM Tollenaar, LH Bouwman, Department of Surgery, Leiden University Medical Center, Leiden, The Netherlands
AM Deelder, Biomolecular Mass Spectrometry Unit, Department of Parasitology, Leiden University Medical Center, Leiden, The Netherlands
Correspondence to: ME de Noo, MD, Department of Surgery, K6-R, Leiden University Medical Center, PO Box 9600, Leiden 2300 RC, The Netherlands. m.e.de_noo@lumc.nl
Telephone: +31-71-5261278 Fax: +31-71-5266750
Received: 2006-07-13 Accepted: 2006-08-21

Abstract

Colorectal adenocarcinoma (CRC) is the third most common type of cancer and the fourth most frequent cause of death due to cancer worldwide. Given the natural history of CRC, early diagnosis appears to be the most appropriate tool to reduce disease-related mortality. A field of recent interest is clinical proteomics, which was reported to lead to high sensitivity and specificities for early detection of several solid tumors. This emerging field uses mass spectrometry-based protein profiles/patterns of easy accessible body fluids to distinguish cancer from non-cancer patients. These discrepancies may be a result of: (1) proteins being abnormally produced or shed and added to the serum proteome, (2) proteins clipped or modified as a consequence of the disease process, or (3) proteins subtracted from the proteome owing to disease-related proteolytic degradation pathways. Therefore, protein pattern diagnostics would provide easy and reliable tools for detection of cancer. This paper focuses on the current status of clinical proteomics research in oncology and in colorectal cancer especially, and will reflect on pitfalls and fears in this relatively new area of clinical medicine, which are reproducibility issues and pre-analytical factors, statistical issues, and identification and nature of discriminating proteins/peptides.

Key words: Proteomics; Colorectal cancer; Protein profiling; Diagnosis; Biomarkers; Statistical issues

de Noo ME, Tollenaar RAEM, Deelder AM, Bouwman LH. Current status and prospects of clinical proteomics studies on detection of colorectal cancer: Hopes and fears. *World J Gastroenterol* 2006; 12(41): 6594-6601

http://www.wjgnet.com/1007-9327/12/6594.asp

BACKGROUND

In order to improve early detection, monitor disease outcome and find targets for more individualised therapy, there is an urgent demand for new biomarkers in oncology. A field of recent interest is clinical proteomics, which has been reported to lead to high sensitivity and specificities for early detection of several solid tumors[1,2]. This emerging field uses mass spectrometry-based protein profiles/patterns of easy accessible body fluids to distinguish cancer from non-cancer patients. This would offer a solution to the problem of cancer often being diagnosed in late stages, when curative resection of the diseased organ is no longer possible and the disease has already metastasized, thus dropping survival rates dramatically. However, after the initial positive announcements in early 2002, criticisms were raised about several aspects of serum proteomics. In this paper we describe the hopes and fears for the introduction of clinical proteomics for early detection of Colorectal adenocarcinoma (CRC).

COLORECTAL CANCER

Colorectal adenocarcinoma is the third most common type of cancer and the fourth most frequent cause of death due to cancer worldwide. Worldwide almost one million new cases occur annually, amounting to 492 000 related deaths[3]. In developed countries it is the second most common type of tumor, with a lifetime risk of 5%, but its incidence and mortality is now decreasing[4,5]. Surgery is the cornerstone of therapy when the disease is confined to the bowel wall. This results in 70%-80% of patients who have tumors that, when diagnosed, can be resected with curative intent[6]. After curative surgery the five-year survival rate for patients with localised disease is 90%, decreasing to 65% in the case of metastasized disease to the lymph nodes. Adjuvant radiation therapy, chemotherapy, or both are useful in selected patients. Classification of tumors into pathogenetical subtypes with distinct clinical courses enables clinicians to target therapy. For CRC, the TNM staging system remains the gold standard and relies entirely on the morphological appearance of the tumor. However, tumors with similar histopathological characteristics may have different clinical outcomes and responsiveness to therapy[7]. Therefore, more individualised treatment would benefit the patient and may avoid unnecessary morbidity. Nonetheless, early detection of CRC will increase survival the most, in view of the fact that it is well recognized that CRC arises from a multistep sequence of genetic alterations that result

in the transformation of normal mucosa to a precursor adenoma and ultimately to carcinoma. Given the natural history of CRC, early diagnosis appears to be the most appropriate tool to reduce disease-related mortality[8-10].

BIOMARKERS

In cancer research, biomarkers are molecules that indicate the presence of cancer in the body. Most biomarkers are based on mutations in genes or abnormal changes in RNA, proteins and metabolites. Since the molecular changes that occur during tumour development can take place over a number of years, some biomarkers can potentially be used to detect early colorectal cancer. Furthermore, they might be used to predict prognosis, monitor disease progression and therapeutic response. Gion *et al* classified different circulating biomarkers according to their clinical application[11]. These candidate biomarkers, however, are frequently found in relatively low concentrations amid a sea of other biomolecules. Therefore biomarker research and possible diagnostic tests depend critically on the ability to make highly sensitive and accurate biochemical measurements. Ideally, such biomarkers should be specific to the disease and easy accessible, such as in serum, plasma or urine, increasing their clinical applicability.

Carcinoembryonic antigen (CEA) is the best-characterised serologic tumor marker for CRC. However, its use as a population-based screening tool for early detection and diagnosis of CRC is hindered by its low sensitivity and specificity. Fletcher demonstrated that for screening purposes in a normal population, a cut-off concentration of 2.5 μg/L CEA would yield a sensitivity of 30%-40%. Based on these data, he calculated that there would be 250 false positive tests for every true positive test; i.e. a patient with cancer. Furthermore, 60% of the cancers would not be detected. The same poor sensitivity applies for diagnosis of CRC. In addition, as CEA can be elevated in the absence of malignancy, specificity is also impaired [12-15].

Faecal occult-blood testing (FOBT) is another biomarker for which clinical trials have shown evidence for a decreased risk of death. This approach is a non-invasive option that limits the need for follow-up colonoscopy to patients with evidence of bleeding. However, neoplasms bleed intermittently, allowing many to escape detection with faecal occult-blood testing. Annual retesting is therefore necessary but is still insufficient, detecting only 25 to 50% of colorectal cancers and 10% of adenomas. The specificity of FOBT is also limited by frequent false positive reactions to dietary compounds, medications, and gastrointestinal bleeding from causes other than colorectal cancer[16-18].

A NEW DIAGNOSTIC PARADIGM: CLINICAL PROTEOMICS

In 2002, several studies demonstrated discrimination of patients with various cancers from healthy subjects on the basis of presence/absence of multiple low-molecular-weight serum proteins using SELDI-TOF mass spectrometry technologies[19-22]. The authors hypothesised that proteomic patterns are correlated to biological events occurring in the entire organism and are likely to change in the presence of disease. New types of bioinformatic pattern recognition algorithms were used to identify patterns of protein changes in order to discriminate cancer patients from healthy individuals with promising results.

Petricoin and his co-workers stated that finding a single disease-related biomarker is like searching for a needle in a haystack as each entity has to be separated and identified individually[23,24]. Moreover, they postulated that the blood proteome constantly changes as a consequence of the perfusion of the diseased organ adding, subtracting, or modifying the circulating proteome. These differences might be the result of proteins being abnormally produced or shed and added to the serum proteome, proteins being clipped or modified as a consequence of the disease process, or proteins being subtracted from the proteome owing to disease-related proteolytic degradation pathways. Therefore, protein pattern diagnostics would provide easier and more reliable tools for detection of cancer. The advantages of the SELDI proteomic pattern approach were stressed in several papers. In addition to the high sensitivity and specificity, cost-effectiveness, easy accessibility of body fluid and especially the high-throughput, ultimately allowing application in future screening studies, were discussed[20,25]. After these hopeful voices, soon critical notes were made on analytical reproducibility and the use of the so-called black box approach, lacking identification of discriminating proteins.

In the next paragraphs, this paper will focus on the current status of clinical proteomics research in oncology and will reflect on pitfalls and fears in this relatively new area of clinical medicine, including reproducibility issues and pre-analytical factors, statistical issues, and the identification and nature of discriminating proteins/peptides.

REPRODUCIBILITY ISSUES AND PRE-ANALYTICAL FACTORS

Boguski and McIntosh were among the first to argue that serum proteomics may be susceptible to observational biases. They stated that any confounding factor could conceivably cause a phenotypic response that might be confused with a specific characteristic of the disease process under study[26]. Confounding factors, which not only include smoking, diet and preoperative stress, but also logistic factors such as sample collection and sample quality, make a reliable and clear differentiation between a normal or malignant status hazy. Another cause for concern, mentioned in this study, is the sample quality and number. The authors favoured use of homogeneous groups with sufficient sample size and stringent standard procedures for serum collection, which is an aspect that is also advocated in other studies[27,28]. Another critical study questioned the reliability of the presence of statistically significant signals at M/Z values less than 500, as used in one of the first studies. Sorace *et al* claimed that the presence of statistically significant bands of low M/Z includes degradation products of higher molecular weight macromolecules or

Table 1 Recommendations of various pre-analytical variations from three MALDI-TOF based reproducibility studies

	Blood component	Peptide isolation	Temperature before sample handling (℃)	Time before centrifugation	Storage of serum	Freeze/ thaw cycles	Circadian rhythm effect
Baumann *et al*	Serum plasma	C3, C8, C18 beads	21	< 30 min	-80 ℃	1	NA[1]
de Noo *et al*	Serum	C8 beads	21	Ideally < 30 min, practically < 2-4 h	NA[1]	2	No effect
West-Nielsen *et al*	Serum plasma	C8 beads	21	< 8 h	-20/-80℃	1	NA[1]

[1]Not applicable for this study.

a matrix effect. Furthermore, this study raised caution regarding poor reproducibility of experimental conditions of chip-based mass spectrometry[29]. This is also reported by another group, which described the poor reproducibility of the SELDI-TOF ovarian cancer data. Baggerly and colleagues postulated that this could partly be contributed to baseline correction, poor sample features in noise regions and even a change of protocol mid-experiment[30]. Most importantly, the promising results that were reported earlier could not be reproduced, therefore stressing the importance of standardised approaches and stringent experimental design. Furthermore, their study pointed out that strong pre-processing of the protein spectra is required in order to obtain reliable classification results in the search for new biomarkers.

Possible confounding factors can be categorized into three sources of variation and bias: biological variation, pre-analytical variation and analytical reproducibility. Biological variation, consisting of both environmental and individual factors, such as race, age, diet, smoking, stress, general physical condition and the use of drugs, may also influence serum protein profiles. However, at the present time no data have been published on this source of variation. Nevertheless, in a previous study our group analyzed pre-analytical and reproducibility issues of our MALDI-TOF approach[31]. The pre-analytical variations corresponded to the logistical conditions in the routine clinical setting; i.e. the effects of sample handling and storage. So far, only a few other studies have reported on the effects of different serum sample preparations and the use of a magnetic-beads-based approach to capture and concentrate serum proteins for MALDI-TOF mass spectrometry[32-34]. Villanueva *et al* mostly focused on influences of different magnetic bead capturing and automation on the reproducibility of serum protein profiles, while Baumann and co-workers mainly studied pre-analytical variation of sample handling.

In Table 1, results for sample handling experiments of the above mentioned studies are summarized. For clinical studies, the use of two freeze/thaw cycles is recommended in three manuscripts. This is mainly due to logistical reasons, such as the ´standard´ for centralized sample collection in large hospitals. The point all authors agreed on is the influence of sample handling; i.e. the time venous blood is left to stand before serum centrifugation. This aspect appears to account for the largest effect on serum or plasma protein profiles. Consequently, standardized sample collection and a well-documented population are recommended in all studies. Standardized protocols should be used from the point of sample collection, sample handling, and storage and freezing of the samples. Although the importance of homogeneity and uniformity within sample groups must once again be stressed, variation of such factors cannot totally be excluded in a clinical setting. In all, when these recommendations are strictly followed and both clinical and analytical factors are controlled, we are confident that the methodology can be standardized to a level that allows application as a tool in biomarker discovery.

STATISTICAL ISSUES

As in all research with high dimensional data, two practical realities constrain the analysis of mass spectra in proteomics. The first is the 'curse of dimensionality', which means the number of features characterizing these data is in the thousands or tens of thousands. The second is the 'curse of dataset sparsity', which means the number of samples is limited. Somorjai *et al* showed the influences of these two curses on classification outcomes. Both the sample per feature ratio, which should ideally be 5 to 10, and feature selection are of pivotal importance for reliable classification and biological optimal relevance[35,36].

Previous to any feature selection or classification, raw mass spectra have to be submitted to so-called pre-processing. During pre-processing, the noise of protein/peptide mass spectra is reduced and the spectra are normalised. Furthermore, smoothing, binning and baseline correction are also performed during pre-processing of the data. Currently, there is a lot of discussion among several groups on how to establish the best method, because data pre-processing is extremely important. There are complex interactions between baseline subtraction, normalization, noise estimation, and peak identification, and therefore these steps should not be considered in isolation[31,37-40].

Another recurring topic for debate is the bioinformatic approach and statistical analysis of protein spectra. Clinically, the most relevant is the issue of an independent validation set for the classification of diseased versus healthy individuals. This is primarily based on a specific problem in the discovery-based research field of clinical proteomics, namely overfitting. Overfitting may occur in the analysis of large datasets when multivariate models show apparent discrimination that is actually caused by data over-interpretation, and hence give rise to results that are not reproducible[30,41,42]. The chance of overfitting, however, can be reduced by appropriate application of validatory estimation and assessment, such as through application of double cross-validation, when properly implemented[43]. Although we have shown

this in a previous study, the general opinion is in favor of performing a classification study with independent validation. In addition, feature selection is also given a lot of attention by statisticians in the field. Several experimental investigations have been made with different peak-feature selection methods. A common approach so far is analysing the data in two phases. First, the peaks in the spectra are extracted and quantified. Secondly, a resulting matrix of peak quantifications is created. For more detailed information on this statistical matter, we refer the reader to the literature[37,44-46].

IDENTIFICATION AND NATURE OF DISCRIMINATING PROTEINS

The controversy regarding the use of protein profiles as a pattern diagnostic, without identification of the individual diagnostic biomarkers, remains to be solved before its clinical application. Whereas the first clinical proteomics studies published their classification method mainly as a black box study, at present identification of the most discriminating proteins or peptides is required for publication in most scientific journals. Identification and functional analysis of these discriminating proteins/peptides might render new insights on tumour development and environmental responsiveness, which could eventually be translated into new diagnostic and prognostic insights for the clinician. Unfortunately, little success has been achieved to date in assigning reproducible discriminating biomarkers[35,42].

Furthermore, several studies have identified their discriminating peaks as components of the coagulation cascade or complement system[47-51]. Therefore, in contrast to the original reflection that discriminating proteomic patterns would identify cancer-specific proteins, it appears that these potential markers belong to the normal serum and plasma proteome. Consequently, some investigators have argued that low molecular weight proteins in serum and the serum peptidome are biologically aspecific and therefore do not yield any reliable biomarkers in the currently technically available mass range[29,52]. Others have proposed that the discriminatory protein peaks represent acute phase reactants that are present in serum in extremely high concentrations[49,53]. Conversely, a study recently reported that although discriminating peptides do indeed belong to the well-known coagulation and complement pathways, their patterns or signatures can nevertheless indicate the presence of cancer. Villanueva *et al* showed that most of the cancer-type specific biomarker fragments were generated in patient serum by enzymatic cleavage at previously known endoprotease cleavage sites after the blood sample was collected[54,55]. They postulated that the discriminating peptides originated after *ex vivo* proteolysis by tumor specific proteases of high abundance protein fragments primarily generated by the coagulation and complement enzymatic cascades. In view of this, they consider these cancer-specific low molecular weight proteins in the serum peptidome to be an indirect snapshot of the enzyme activity in tumor cells. We support their hypothesis that proteolytic process profiles in the serum peptidome hold important information that may have direct clinical utility as a surrogate marker for the detection and classification of certain types of tumors. Unique proteases may be shed by tumor cells or reflect activity of the host immune response, which may contribute to new proteins such as chemokines and lymphokines. These processes result in subtle changes in low molecular proteomic signatures, which may ultimately be used for classification methods in various cancers and diseases in the future[54]. Proteases have been extensively implicated in the development and progression of cancer[56,57]. Song *et al* recently stated that proteolytic processing of high abundance host-response proteins actually amplifies the signal of potentially low-abundance biologically active disease markers such as proteases. Therefore, it might be expected that more convenient and reliable blood proteins and peptides simply serve as an endogenous substrate pool for proteases as surrogate markers for the detection and classification of cancer[58].

Another recurrent topic of debate concerns the type of blood component that is best for protein profiling and peptidome analysis. Some investigators favour the use of plasma because they presume that, in serum, ongoing enzymatic activity occurring during clotting is likely to cleave even proteins that are not involved in biologically relevant pathways[53,59]. Others, however, advocate the use of serum. We support the hypothesis that since the kidneys rapidly clear peptides smaller than 4 kDa, which are generated *in vivo* in the circulation, the majority of peptides in blood samples exist from *ex vivo* proteolysis. This explains the possibility that low abundance proteins, including possible tumor markers, may be totally obscured and not retraceable during direct mass spectrometry. However, it has recently been shown that exogenous proteases are functionally measurable in serum, however in higher concentrations than in plasma[54].

Functional proteomics studies allow the investigation of environmental factors over time, rendering the monitoring of metabolic responses to various stimuli. Hence, post-translational modifications can be studied, whereas they cannot be detected by genomic studies. Post-translational modification changes, like glycosylation of proteins and lipids, are a common feature in colorectal cancer influencing cancer cell behaviour and can be detected using mass spectrometry due to characteristic mass shifts[60]. We expect that both phosphoproteomics and/or glycoproteomics, enabling study of crucial post-translational modifications of proteins in the cancer pathway, will revolutionize our understanding of the function of these proteins and hence render new insights for monitoring and therapy.

CLINICAL PROTEOMICS IN CRC

Until present, few protein profiling studies have been published on the detection of CRC, two being based on SELDI-TOF and one on MALDI-TOF mass spectrometry. The first SELDI-TOF study showed seven potential biomarkers that could differentiate CRC patients from those with colorectal adenoma with a sensitivity of 89% and specificity of 83%. The seven potential biomarkers have a large range in mass values, differing from 4654 to 21 742 Da[61]. A more recently published study found 5 possible bi-

omarkers to differentiate between healthy control subjects and CRC patients. The study consisted of a training set of samples from 40 patients with colorectal cancer (all Dukes' D) and 49 healthy controls. The second set included samples from 37 patients with colorectal cancer (1 Dukes' A, 2 Dukes' B, 12 Dukes' C, 17 Dukes' D, 5 unknown) and 31 healthy controls. For three of these potential markers, they found a sensitivity and specificity between 65% and 90%. They reported that m/z 3.100, 3.300, 4.500, 6.600 and 28.000 were the most important biomarkers. The total sample set showed that 1 of 1 Dukes' A, 1 of 2 Dukes' B, 11 of 12 (91.7%) Dukes' C, and 47 of 57 (82.5%) Dukes' D were correctly classified. Stratification by Dukes' stages showed a significantly better sensitivity of the classification trees (91.7%, 11/12) compared to CEA (25.0%, 3/12) in Dukes' C colorectal cancer, although at stage D CEA performed better. No conclusions can be drawn on the performance of our classification trees at earlier stages of colorectal cancer due to limited samples, but 2 of 3 patient samples from stage A and B were correctly classified by the trees and none when using the clinical cut-off for CEA[59]. Our group used MALDI-TOF mass spectrometry to differentiate CRC patients from healthy controls. In a randomized block design, pre-operative serum samples obtained from 66 colorectal cancer patients and 50 controls were used to generate high-resolution MALDI-TOF protein profiles[42]. After pre-processing of the spectra, linear discriminant analysis with double cross-validation, based on principal component analysis, was used to classify the protein profiles. Thirty-four patients out of thirty-seven with early stage disease (stage 1 and 2) and all patients with stage 3 or 4 disease were correctly classified as having cancer. For the misclassified control subjects, it was not possible to retrieve the current physical state as it concerned anonymous healthy controls.

A total recognition rate of 92.6%, a sensitivity of 95.2% and a specificity of 90.0% for the detection of CRC were shown. In our study two first principal components accounted for most of the between-group separation, both with a m/z between 1 and 2 kDa.

Although much research has been done using 2D gel electrophoresis to detect possible biomarkers and targets for CRC, this falls outside the scope of this review since this technique cannot be scaled up to a directly applicable diagnostic test. On the other hand, a screening assay based on an APC protein truncation test has recently been proposed and other studies mention the potential use of protein microarrays[2,60,62,63]. However, studies linking large protein expression patterns with clinical outcome in colorectal cancer are still in their infancy. To be able to predict occurrence of disease and treatment outcome, more studies on genotype-phenotype correlations are needed both in sporadic and in hereditary colorectal cancer.

FUTURE PERSPECTIVES

The best anticancer strategies still rely on early detection followed by close monitoring for early relapse so that therapies can be appropriately adjusted[64]. In addition, new targets for therapy are a constant subject of study in oncology. In fact, increased understanding of the molecular mechanisms of cancer progression may refine treatment and management of patients. Advances in genomics and proteomics may lead to earlier detection of cancer thus enabling a more precise classification of (smaller subsets of) patients based on their predicted response to individual therapies. Conceptually, proteomics is more suitable than genomics for novel targeted therapies, since most protein biomarkers are based on aberrant protein signalling circuits represented by post-translational modifications. The dynamic range of the proteome allows more insight in the functional state of a cell, tissue or organ over a period of time. Besides, protein profiling and classification of several components of multiple aberrant cell signalling cascades would be expected to predict disease behaviour better than just single pathways in isolation[62]. Therefore, proteomics could be expected to render better insight in pathogenetic mechanisms, disease progression and treatment response. This is of paramount importance as cancer advances dynamically and affects heterogeneous cell populations, either as a part of cancer or as a part of a tumor-host reaction[49,65].

Further refinement of serum protein profiles is needed before these mass spectrometry-based techniques become part of clinical routines. Currently, several studies have carefully evaluated reproducibility, automation, sample throughput and sensitivity of serum proteomic techniques. The first problems related to these factors seem to have been overcome due to stringent standardised approaches as described earlier. However, proteomics studies still have several drawbacks: (a) current tools only allow narrow-range analyses, (b) identification of proteins of interest remains cumbersome, (c) protein studies address mixtures of high complexity. Hence, due to the dynamic ranges of the human proteome and the lack of amplification methods in protein studies, targeted proteomics techniques for (quantitive) identification of low-abundant proteins have to be investigated further[66]. Another approach to studying proteins at a functional level might be the use of array-based proteomics platforms. This technique offers the potential for highly multiplex and sensitive analysis of serum or tumor proteins[62]. Using this direct approach to study the proteomic circuitry would theoretically allow for the creation of functional signalling maps of cancers, even at the level of the individual patient. Regarding identification of potential biomarkers, limitations of direct MS/MS have been stressed previously as well as the fact that antibody-approaches may yield higher sensitivity[53,54].

In the next era, research in oncology will drift to more individualised medicine. In this view, molecular profiling forms a welcome addition to the pathology report of cancer. Until now, histopathological staging and demographics have been used to predict disease outcome. However, we believe that protein profiling and other proteomics techniques may lead to more individualised medicine and tailor made therapy[67,68]. At first, both approaches should be used complementarily instead of competitively.

It is unlikely that in the next decade, serum protein profiles will replace the current gold standard, which is colonoscopy, for the diagnosis of CRC. Nevertheless, we hypothesise that MALDI-TOF based serum protein profiles, once validated in independent studies, could be used

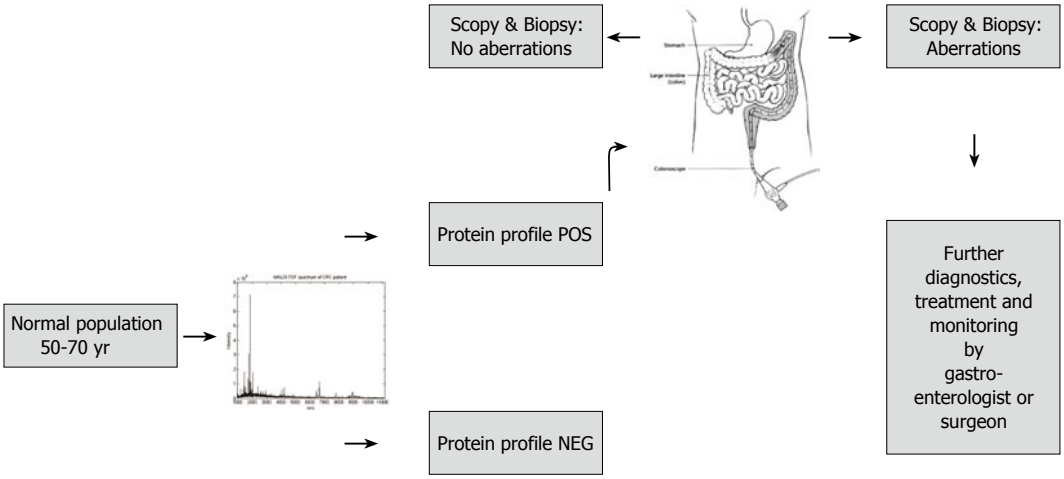

Figure 1 Flow chart of possible clinical application of MALDI-TOF.

as selection criteria for the more invasive and time-consuming diagnostic colonoscopy (Figure 1). Eventually, with the present debate on screening programs for colorectal cancer in several countries, clinical proteomics may replace and surpass the use of faecal occult-blood testing (FOBT). When used in independent validation studies, sensitivity and specificity remain about 90%. Protein profiling might even replace FOBT, since this approach has a lower specificity and a number of disadvantages. Non-bleeding tumors and, more relevantly, polyps and adenomas cannot be detected using FOBT, whereas we expect to realise this detection with serum protein profiling within the next decade[17,18].

Although the current reality may not have kept pace with previous expectations and the translation from bench to bedside is more laborious than initially thought, there is supporting evidence for the potential vast use of clinical proteomics in oncology. Particularly, this potential will be realized when technical innovations to further increase sensitivity and specificity of proteomic techniques are implemented and more sensitive methods for protein identification on alternations are developed. In combination with the use and set-up of well-defined cases together with well-documented serum banks, including not only for CRC samples but also inflammatory disease and polyps, serum protein profiling may propel diagnostic research in CRC in the right direction.

REFERENCES

1 **Chambers G**, Lawrie L, Cash P, Murray GI. Proteomics: a new approach to the study of disease. *J Pathol* 2000; **192**: 280-288
2 **Posadas EM**, Simpkins F, Liotta LA, MacDonald C, Kohn EC. Proteomic analysis for the early detection and rational treatment of cancer--realistic hope? *Ann Oncol* 2005; **16**: 16-22
3 **Weitz J**, Koch M, Debus J, Hohler T, Galle PR, Buchler MW. Colorectal cancer. *Lancet* 2005; **365**: 153-165
4 **Russo MW**, Wei JT, Thiny MT, Gangarosa LM, Brown A, Ringel Y, Shaheen NJ, Sandler RS. Digestive and liver diseases statistics, 2004. *Gastroenterology* 2004; **126**: 1448-1453
5 **Jemal A**, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, Feuer EJ, Thun MJ. Cancer statistics, 2004. *CA Cancer J Clin* 2004; **54**: 8-29
6 **Pfister DG**, Benson AB 3rd, Somerfield MR. Clinical practice. Surveillance strategies after curative treatment of colorectal cancer. *N Engl J Med* 2004; **350**: 2375-2382
7 **Liefers GJ**, Tollenaar RA. Cancer genetics and their application to individualised medicine. *Eur J Cancer* 2002; **38**: 872-879
8 **Ruo L**, Gougoutas C, Paty PB, Guillem JG, Cohen AM, Wong WD. Elective bowel resection for incurable stage IV colorectal cancer: prognostic variables for asymptomatic patients. *J Am Coll Surg* 2003; **196**: 722-728
9 **Gill S**, Sinicrope FA. Colorectal cancer prevention: is an ounce of prevention worth a pound of cure? *Semin Oncol* 2005; **32**: 24-34
10 **Hawk ET**, Levin B. Colorectal cancer prevention. *J Clin Oncol* 2005; **23**: 378-391
11 **Gion M**, Daidone MG. Circulating biomarkers from tumour bulk to tumour machinery: promises and pitfalls. *Eur J Cancer* 2004; **40**: 2613-2622
12 **Duffy MJ**, van Dalen A, Haglund C, Hansson L, Klapdor R, Lamerz R, Nilsson O, Sturgeon C, Topolcan O. Clinical utility of biochemical markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines. *Eur J Cancer* 2003; **39**: 718-727
13 **Fletcher RH**. Rationale for combining different screening strategies. *Gastrointest Endosc Clin N Am* 2002; **12**: 53-63
14 **Winawer S**, Fletcher R, Rex D, Bond J, Burt R, Ferrucci J, Ganiats T, Levin T, Woolf S, Johnson D, Kirk L, Litin S, Simmang C. Colorectal cancer screening and surveillance: clinical guidelines and rationale-Update based on new evidence. *Gastroenterology* 2003; **124**: 544-560
15 **Ouyang DL**, Chen JJ, Getzenberg RH, Schoen RE. Noninvasive testing for colorectal cancer: a review. *Am J Gastroenterol* 2005; **100**: 1393-1403
16 **Pignone M**, Rich M, Teutsch SM, Berg AO, Lohr KN. Screening for colorectal cancer in adults at average risk: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2002; **137**: 132-141
17 **Ransohoff DF**, Lang CA. Screening for colorectal cancer with the fecal occult blood test: a background paper. American College of Physicians. *Ann Intern Med* 1997; **126**: 811-822
18 **Ransohoff DF**. Colon cancer screening in 2005: status and challenges. *Gastroenterology* 2005; **128**: 1685-1695
19 **Adam BL**, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, Semmes OJ, Schellhammer PF, Yasui Y, Feng Z, Wright GL Jr. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* 2002; **62**: 3609-3614
20 **Petricoin EF**, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002; **359**: 572-577
21 **Rai AJ**, Zhang Z, Rosenzweig J, Shih IeM, Pham T, Fung ET, Sokoll LJ, Chan DW. Proteomic approaches to tumor marker discovery. *Arch Pathol Lab Med* 2002; **126**: 1518-1526
22 **Yanagisawa K**, Shyr Y, Xu BJ, Massion PP, Larsen PH, White BC, Roberts JR, Edgerton M, Gonzalez A, Nadaf S, Moore

JH, Caprioli RM, Carbone DP. Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* 2003; **362**: 433-439
23 **Petricoin EF**, Liotta LA. Proteomic analysis at the bedside: early detection of cancer. *Trends Biotechnol* 2002; **20**: S30-S34
24 **Wulfkuhle JD**, Liotta LA, Petricoin EF. Proteomic applications for the early detection of cancer. *Nat Rev Cancer* 2003; **3**: 267-275
25 **Petricoin EE**, Paweletz CP, Liotta LA. Clinical applications of proteomics: proteomic pattern diagnostics. *J Mammary Gland Biol Neoplasia* 2002; **7**: 433-440
26 **Boguski MS**, McIntosh MW. Biomedical informatics for proteomics. *Nature* 2003; **422**: 233-237
27 **Diamandis EP**. Re: Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst* 2003; **95**: 489-490; author reply 490-491
28 **Diamandis EP**. Proteomic patterns in serum and identification of ovarian cancer. *Lancet* 2002; **360**: 170; author reply 170-171
29 **Sorace JM**, Zhan M. A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinformatics* 2003; **4**: 24
30 **Baggerly KA**, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* 2004; **20**: 777-785
31 **de Noo ME**, Tollenaar RAEM, Ozalp A, Kuppen PJK, Bladergroen MR, and Deelder AM. Reliability of human serum protein profiles generated with C8 magnetic beads assisted MALDI-TOF mass spectrometry. *Anal Chem* 2005; **77**: 7232-7241
32 **Villanueva J**, Philip J, Entenberg D, Chaparro CA, Tanwar MK, Holland EC, Tempst P. Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. *Anal Chem* 2004; **76**: 1560-1570
33 **Baumann S**, Ceglarek U, Fiedler GM, Lembcke J, Leichtle A, Thiery J. Standardized approach to proteome profiling of human serum based on magnetic bead separation and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin Chem* 2005; **51**: 973-980
34 **West-Nielsen M**, Hogdall EV, Marchiori E, Hogdall CK, Schou C, Heegaard NH. Sample handling for mass spectrometric proteomic investigations of human sera. *Anal Chem* 2005; **77**: 5114-5123
35 **Baggerly KA**, Morris JS, Wang J, Gold D, Xiao LC, Coombes KR. A comprehensive approach to the analysis of matrix-assisted laser desorption/ionization-time of flight proteomics spectra from serum samples. *Proteomics* 2003; **3**: 1667-1672
36 **Yasui Y**, Pepe M, Thompson ML, Adam BL, Wright GL Jr, Qu Y, Potter JD, Winget M, Thornquist M, Feng Z. A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics* 2003; **4**: 449-463
37 **Eilers PH**. A perfect smoother. *Anal Chem* 2003; **75**: 3631-3636
38 **Coombes KR**, Morris JS, Hu J, Edmonson SR, Baggerly KA. Serum proteomics profiling--a young technology begins to mature. *Nat Biotechnol* 2005; **23**: 291-292
39 **Ransohoff DF**. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* 2004; **4**: 309-314
40 **Diamandis EP**. Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J Natl Cancer Inst* 2004; **96**: 353-356
41 **de Noo ME**, Mertens BJ, Ozalp A, Bladergroen MR, van der Werff MP, van de Velde CJ, Deelder AM, Tollenaar RA. Detection of colorectal cancer using MALDI-TOF serum protein profiling. *Eur J Cancer* 2006; **42**: 1068-1076
42 **Hu J**, Coombes KR, Morris JS, Baggerly KA. The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales. *Brief Funct Genomic Proteomic* 2005; **3**: 322-331
43 **Wang X**, Zhu W, Pradhan K, Ji C, Ma Y, Semmes OJ, Glimm J, Mitchell J. Feature extraction in the analysis of proteomic mass spectra. *Proteomics* 2006; **6**: 2095-2100
44 **Levner I**. Feature selection and nearest centroid classification for protein mass spectrometry. *BMC Bioinformatics* 2005; **6**: 68
45 **Somorjai RL**, Dolenko B, Baumgartner R. Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. *Bioinformatics* 2003; **19**: 1484-1491
46 **Koomen JM**, Zhao H, Li D, Nasser W, Hawke DH, Abbruzzese JL, Baggerly KA, Kobayashi R. Diagnostic protein discovery using liquid chromatography/mass spectrometry for proteolytic peptide targeting. *Rapid Commun Mass Spectrom* 2005; **19**: 1624-1636
47 **Ye B**, Cramer DW, Skates SJ, Gygi SP, Pratomo V, Fu L, Horick NK, Licklider LJ, Schorge JO, Berkowitz RS, Mok SC. Haptoglobin-alpha subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry. *Clin Cancer Res* 2003; **9**: 2904-2911
48 **Koomen JM**, Shih LN, Coombes KR, Li D, Xiao LC, Fidler IJ, Abbruzzese JL, Kobayashi R. Plasma protein profiling for diagnosis of pancreatic cancer reveals the presence of host response proteins. *Clin Cancer Res* 2005; **11**: 1110-1118
49 **Wang X**, Wang E, Kavanagh JJ, Freedman RS. Ovarian cancer, the coagulation pathway, and inflammation. *J Transl Med* 2005; **3**: 25
50 **Li J**, Orlandi R, White CN, Rosenzweig J, Zhao J, Seregni E, Morelli D, Yu Y, Meng XY, Zhang Z, Davidson NE, Fung ET, Chan DW. Independent validation of candidate breast cancer serum biomarkers identified by mass spectrometry. *Clin Chem* 2005; **51**: 2229-2235
51 **Diamandis EP**, van der Merwe DE. Plasma protein profiling by mass spectrometry for cancer diagnosis: opportunities and limitations. *Clin Cancer Res* 2005; **11**: 963-965
52 **Koomen JM**, Li D, Xiao LC, Liu TC, Coombes KR, Abbruzzese J, Kobayashi R. Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker discovery. *J Proteome Res* 2005; **4**: 972-981
53 **Villanueva J**, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, Fleisher M, Lilja H, Brogi E, Boyd J, Sanchez-Carbayo M, Holland EC, Cordon-Cardo C, Scher HI, Tempst P. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* 2006; **116**: 271-284
54 **Liotta LA**, Petricoin EF. Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. *J Clin Invest* 2006; **116**: 26-30
55 **Matrisian LM**, Sledge GW Jr, Mohla S. Extracellular proteolysis and cancer: meeting summary and future directions. *Cancer Res* 2003; **63**: 6105-6109
56 **Bank U**, Kruger S, Langner J, Roessner A. Review: peptidases and peptidase inhibitors in the pathogenesis of diseases. Disturbances in the ubiquitin-mediated proteolytic system. Protease-antiprotease imbalance in inflammatory reactions. Role of cathepsins in tumour progression. *Adv Exp Med Biol* 2000; **477**: 349-378
57 **Song J**, Patel M, Rosenzweig CN, Chan-Li Y, Sokoll LJ, Fung ET, Choi-Miura NH, Goggins M, Chan DW, Zhang Z. Quantification of fragments of human serum inter-alpha-trypsin inhibitor heavy chain 4 by a surface-enhanced laser desorption/ionization-based immunoassay. *Clin Chem* 2006; **52**: 1045-1053
58 **Marshall J**, Kupchak P, Zhu W, Yantha J, Vrees T, Furesz S, Jacks K, Smith C, Kireeva I, Zhang R, Takahashi M, Stanton E, Jackowski G. Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. *J Proteome Res* 2003; **2**: 361-372
59 **Steinert R**, Buschmann T, van der Linden M, Fels LM, Lippert H, Reymond MA. The role of proteomics in the diagnosis and outcome prediction in colorectal cancer. *Technol Cancer Res Treat* 2002; **1**: 297-304
60 **Yu JK**, Chen YD, Zheng S. An integrated approach to the detection of colorectal cancer utilizing proteomics and bioinformatics. *World J Gastroenterol* 2004; **10**: 3127-3131
61 **Engwegen JY**, Helgason HH, Cats A, Harris N, Bonfrer JM, Schellens JH, Beijnen JH. Identification of serum proteins discriminating colorectal cancer patients and healthy controls

using surface-enhanced laser desorption ionisation-time of flight mass spectrometry. *World J Gastroenterol* 2006; **12**: 1536-1544

62 **Gulmann C**, Sheehan KM, Kay EW, Liotta LA, Petricoin EF 3rd. Array-based proteomics: mapping of protein circuitries for diagnostics, prognostics, and therapy guidance in cancer. *J Pathol* 2006; **208**: 595-606

63 **Miller JC**, Zhou H, Kwekel J, Cavallo R, Burke J, Butler EB, Teh BS, Haab BB. Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. *Proteomics* 2003; **3**: 56-63

64 **Etzioni R**, Urban N, Ramsey S, McIntosh M, Schwartz S, Reid B, Radich J, Anderson G, Hartwell L. The case for early detection. *Nat Rev Cancer* 2003; **3**: 243-252

65 **Kolch W**, Mischak H, Pitt AR. The molecular make-up of a tumour: proteomics in cancer research. *Clin Sci* (Lond) 2005; **108**: 369-383

66 **Srinivas PR**, Verma M, Zhao Y, Srivastava S. Proteomics for cancer biomarker discovery. *Clin Chem* 2002; **48**: 1160-1169

67 **de Noo ME**, Liefers GJ, Tollenaar RA. Translational research in prognostic profiling in colorectal cancer. *Dig Surg* 2005; **22**: 276-281

68 **Ludwig JA**, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* 2005; **5**: 845-856

S- Editor Wang GP **L- Editor** Lutze M **E- Editor** Liu WF