Editorial

Plasma Protein Profiling by Mass Spectrometry for Cancer Diagnosis: Opportunities and Limitations
☐☐*Commentary on Koomen JM et al., p. 1110*

Eleftherios P. Diamandis and Da-Elene van der Merwe

Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada and Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

There is now solid scientific evidence suggesting that early detection of various forms of cancer can lead to improved clinical outcomes (1). It has thus been suggested that early cancer diagnosis and administration of definitive therapy is probably the most promising way to reduce the burden of cancer in the shortest period of time. The National Cancer Institute has created the Early Detection Research Network which is focusing on discovery and validation of biomarkers for early cancer detection. In addition to classic serum biomarkers, other techniques such as imaging, cytology, and serology can also play a major role in early cancer diagnosis or for identifying precancerous lesions. However, at the moment, neither serum biomarkers nor imaging is sensitive and/or specific enough to diagnose human cancers early. For this reason, there is an urgent need to discover and validate novel biomarkers or other diagnostic modalities.

How could putative new biomarkers be discovered? The sequence of the human genome has provided us with a list of all human genes. Potentially, this knowledge can lead to the development of specific reagents which will allow testing of thousands of proteins as potential biomarkers for human diseases. The focus in cancer biomarker discovery is driven by the following approaches: (*a*) The secreted protein hypothesis assumes that the most promising serum biomarkers will be secreted proteins (2). (*b*) With the candidate protein approach, a particular protein is tested in sets of samples from normal individuals and patients with cancer to determine its discriminatory value. (*c*) Bioinformatics compare the expression of various genes between cDNA libraries that have been constructed from either normal or cancerous tissues (3). This analysis can identify highly overexpressed genes which may

reveal worthwhile candidate biomarkers. (*d*) cDNA microarrays applied to normal and tumor tissues may be able to identify overexpressed genes which can then be examined for candidate biomarkers (4). (*e*) Comparative multiparametric analysis of serum can be done by quantitative mass spectrometry to differentiate health and diseased states.

To date, despite extensive experimentation with all these technologies, no major cancer biomarkers have as yet been discovered or validated.

In our quest to discover novel serum-based biomarkers for cancer, it is instructive to examine the classic cancer biomarkers such as carcinoembryonic antigen, α-fetoprotein, prostate-specific antigen, cancer antigen 125, cancer antigen 15.3, etc. and note their concentrations in serum and the requirements for their quantification. These biomarkers are present in serum at the low nanogram per milliliter concentration ranges and therefore require highly sensitive immunologic techniques for their quantification. In order for these molecules to be useful in the clinical setting, the between-run analytic imprecision should be less than 10%. These assay characteristics would allow longitudinal measurements for early cancer relapse and improved discrimination between normal subjects and individuals with cancer by using well-defined cutoff levels. At present, although these classic biomarkers are used clinically to assess therapeutic response and early detection of relapse, they are not recommended for population screening. Their lack of diagnostic specificity would yield too many false-positive results, which could lead to unnecessary and potentially harmful interventions in many patients who do not have cancer (5).

Biological mass spectrometry currently represents the most important analytic proteomic tool (6). This method is capable of positively identifying proteins and peptides with relative ease and for performing multiparametric analysis of complex biological fluids such as serum. Mass spectrometry has been used in two different settings in the area of cancer diagnostics. First, for novel cancer biomarker discovery, where biological fluids such as serum, urine, cerebrospinal fluid, etc. are fractionated by chromatographic techniques and analyzed by mass spectrometry to identify new protein markers. In a second approach, introduced originally by Petricoin, Liotta, and co-investigators, mass spectrometry is used to generate a profile of peaks from serum, which is first treated with a chromatographic surface (a protein chip) to allow immobilization of a subpopulation of proteins or peptides. Without knowledge of the identity of these peaks, these authors have shown, through powerful bioinformatic algorithms, that they could discriminate between health and diseased states with unprecedented sensitivity and specificity (7). This approach has already been used for diagnosis of ovarian, prostate, breast, bladder, pancreatic, and many other cancers (8). If these findings are

reproduced and validated, they could represent a major scientific breakthrough with immediate clinical applicability.

Recently, important concerns were raised on the validity of serum proteomic pattern analysis by mass spectrometry for early cancer diagnosis (9–14). Based on the published methodology, it was predicted that this approach would identify high-abundance proteins in the circulation which are not released by the tumor, likely representing nonspecific epiphenomena of cancer presence (15). Initially, published papers using this technology were unable to positively identify the discriminatory peaks and it was therefore impossible to determine whether these peaks represent novel biomarkers or high abundance non-cancer-specific proteins. More recent reports do reveal the identity of these peptides/proteins and examine their pathophysiologic connection to cancer. A list of positively identified candidate biomarkers by mass spectrometry for various forms of cancer is shown in Table 1. The table includes biomarkers previously described by other investigators as well as by Koomen et al. in a paper published in this journal (16). It is clear that most, if not all, identified proteins thus far represent acute-phase reactants produced by the liver in response to inflammation. These proteins are present in extremely high abundance in serum, precluding their release from small tumor tissues, as exemplified elsewhere (9–11). Moreover, close examination of the concentration differences of these candidate biomarkers between normal subjects and patients with cancer, in comparison to classic cancer biomarkers, reveals that such differences are extremely small and of doubtful clinical value (17). In fact, in the paper by Koomen et al. (16), haptoglobin, which was identified as a candidate biomarker for pancreatic carcinoma with mass spectrometry, was not shown to be different between patients with or without cancer, when analyzed by a biochemical test. Furthermore, the ELISA results for serum amyloid A confirmed that this marker was marginally useful for identifying pancreatic carcinoma, adding approximately only another 5% of patients to those

already detected by the classic pancreatic cancer biomarker, cancer antigen 19.9. Koomen et al. also reported the lowest concentration of analyte that could be measured with their technology to be around 20 μg/mL, a level that is more than 1,000-fold higher than levels of classic biomarkers found in serum. Although Koomen et al. found reasonable sensitivity for pancreatic cancer diagnosis (88%), the specificity was unacceptably low (75%), precluding use in clinical practice.

Where do we go from here? The original papers on serum proteomic profiling for diagnosis of various forms of cancer reported impressive results (7). As yet, these results have not been reproduced by other laboratories and the method has not been validated. Others tried to refine the methodology with prepurification steps to isolate informative peptides, presumably released by the proteolytic activity of proteases around the tumor microenvironment (18). These approaches merit further investigation. Using peaks of unknown identity for diagnostic purposes should not be a reason to invalidate the method; instead, as Ranshoff points out, it will be important to examine "if this technology does work" and leave the question of "how it works" for investigation at a later time (19). The "does it work" question could be addressed quickly by using simple, retrospective studies. Precautionary measures about sample collection, processing, and patient selection must be seriously considered to avoid biases. The same applies to the use the bioinformatic tools (12, 13). It is possible that the inappropriate use of bioinformatic algorithms can lead to overfitting of data, which could not be reproduced in a different experimental setting, as described by Rogers et al. (20).

In conclusion, the study by Koomen et al. (16) confirms some of the initial concerns regarding this technology by showing that the discriminatory peaks identified for pancreatic cancer represent acute-phase reactants which are present in serum at extremely high concentrations. Furthermore, these authors have shown that the current approach of using unfractionated or minimally fractionated serum, in association

Table 1 Serum concentration of some abundant proteins, classic cancer biomarkers, and putative new cancer biomarkers identified by mass spectrometry

Compound	Approximate concentration (pmol/L)	Biomarker for cancer type	Reference
Serum proteins			
Albumin	600,000,000	–	23
Immunoglobulins	30,000,000	–	23
C-reactive protein	40,000	–	23
Classic tumor markers			
α-Fetoprotein	150	Hepatoma, testicular	23
Prostate-specific antigen	140	Prostate	23
Carcinoembryonic antigen	30	Colon, pancreas, lung, breast	23
Choriogonadotropin	20	Testicular, choriocarcinoma	23
β-Subunit of choriogonadotropin	2	Testicular, choriocarcinoma	23
Mass spectrometry-identified proteins			
Apolipoprotein A1	40,000,000	Ovarian, pancreatic	16[*], 17
Transthyretin fragment	6,000,000	Ovarian	17
Inter-α-trypsin inhibitor fragment	4,000,000	Ovarian, pancreatic	16[*], 17
Haptoglobin α-subunit	1,000,000	Ovarian, pancreatic	16[*], 24
Vitamin D-binding protein	10,000,000	Prostate	25
Serum amyloid A	20,000,000	Nasopharyngeal, pancreatic	16[*], 26
α_1-Antitrypsin	10,000,000	Pancreatic	16[*]
α_1-Antichymotrypsin	5,000,000	Pancreatic	16[*]

*Found by Koomen et al.

with mass spectrometry, is not sensitive enough to identify molecules in the sub-nanogram per milliliter range. It remains to be seen whether further refinements, such as more powerful fractionation techniques and isolation of low molecular weight peptides (21, 22), combined with bioinformatic analysis and mass spectrometry, will yield clinically useful diagnostic methods for cancer. Until such methods are published and thoroughly validated, the initial claims that this technology could revolutionize cancer diagnostics should remain speculative.

REFERENCES

1. Etzioni R, Urban N, Ramsey S, et al. The case for early detection. Nature Rev Cancer 2003;3:243–52.

2. Welsh JB, Sapinoso LM, Kern SG, et al. Large-sscale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. PNAS 2003;100:3410–5.

3. Yousef GM, Polymeris ME, Yacoub GM, et al. Parallel overexpression of seven kallikrein genes in ovarian cancer. Cancer Res 2003;63:2223–7.

4. Welsh JB, Zarrinkar PP, Sapinoso LM, et al. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Proc Natl Acad Sci U S A 2001;98:1176–81.

5. Menon U, Jacobs I. Screening for ovarian cancer. Best Pract Res Clin Obstet Gynaecol 2002;16:469–82.

6. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003;422:198–207.

7. Petricoin EF III, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359:572–5.

8. Isaaq HJ, Conrads TP, Prieto DA, Tirumalai R, Veenstra TD. SELDI-TOF MS for diagnostic proteomics. Anal Chem 2003;75:149–55A.

9. Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool. Mol Cell Proteomics 2004;3:367–78.

10. Diamandis EP. Analysis of serum proteomic patterns for early cancer diagnosis: Drawing attention to potential problems. J Natl Cancer Inst 2004;96:353–6.

11. Diamandis EP. Proteomic patterns in biological fluids: Do they represent the future of cancer diagnostics? Clin Chem 2003;49: 1271–8.

12. Sorace JM, Zhan M. A data review and reassessment of ovarian cancer serum proteomic profiling. B.M.C. Bioinformatics 2003;4:24 [http://www.biomedcentral.com/1471–2105/4/24].

13. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing data sets from different experiments. Bioinformatics Adv Access 1–9, Oxford University Press 2004.

14. Check E. Running before we can walk? Nature 2004;429:496–7.

15. Diamandis EP. Identification of serum amyloid a protein as a potentially useful biomarker for nasopharyngeal carcinoma. Clin Cancer Res 2004;10:5293.

16. Koomen JM, Shih LN, Coombes KR, et al. Plasma protein profiling for diagnosis of pancreatic cancer reveals the presence of host response proteins. Clin Cancer Res. In press.

17. Zhang Z, Bast RC Jr, Yu Y, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. Cancer Res 2004;64:5882–90.

18. Villanueva J, Philip J, Entenberg D, et al. Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. Anal Chem 2004;76:1560–70.

19. Ransohoff DF. Evaluating discovery-based research: When biologic reasoning cannot work. Gastroenterol 2004;127:1028.

20. Rogers MA, Clarke P, Noble J, et al. Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization (SELDI) and neural-network analysis: Identification of key issues affecting potential clinical utility. Cancer Res 2003;63:6971–83.

21. Radhakrishna S, Tirumalai S, Chan KC, et al. Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics 2003;2:1096–103.

22. Liotta LA, Ferrari M, Petricoin E. Written in blood. Nature 2003;425:905.

23. Burtis CA, Ashwood ER. Tietz textbook of clinical chemistry. 2nd ed. Philadelphia: W.B. Saunders Co.; 1994.

24. Ye B, Cramer DW, Skates SJ, et al. Haptoglobin-α subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry. Clin Cancer Res 2003;9:2904–11.

25. Hlavaty JJ, Partin AW, Shue MJ, et al. Identification and preliminary clinical evaluation of a 50.8-kDa serum marker for prostate cancer. Urology 2003;61:1261–5.

26. Cho WSC, Yip TTC, Yip C, et al. Identification of serum amyloid A protein as a potentially useful biomarker to monitor relapse of nasopharyngeal cancer by serum proteomic profiling. Clin Cancer Res 2004;10:43–52.