ACTA BIOMED 2007; 78; Suppl 1: 123-128
© MATTIOLI 1885

LECTURE

Serum peptide profiling: identifying novel cancer biomarkers for early disease detection

Andrew Martorella, Richard Robbins
Department of Medicine, Endocrinology Service, Memorial Sloan-Kettering Cancer Center, New York, USA

Abstract. Recent advances in mass spectrometry have enabled the identification of hundreds of low molecular weight (LMW) peptides that have previously been difficult to detect in human serum. Serum peptide patterns can now be analyzed using commercially available statistical programs to identify potential peptide patterns that may correlate with the presence or absence of specific diseases. A serum peptide profile (SPP), which is unique to each patient, can be created and compared to a known SPP from a specific disease. The SPP thus serves as a potential early stage biomarker prior to the clinical manifestation of disease. A unique and automated technology platform has been developed by members of the Protein Center at Memorial Sloan-Kettering Cancer Center (MSKCC). It involves a magnetic bead-based approach to extract LMW peptides from serum, placing them by robotic automation on a stainless steel MALDI-TOF target plate, subjecting them to mass spectrometric analysis, and using GeneSpring software to analyze the peptide patterns. Human serum from a cohort of 27 patients with metastatic thyroid cancer and 32 controls were analyzed on the MSKCC platform. 549 individual LMW peptides were identified. A SPP composed of 98 discriminatory LMW peptides was able to distinguish between the two groups of serum samples with high statistical accuracy. We believe that our automated system will serve as a model for future biotechnology laboratories in the quest for hidden diagnostic clues that may be detected by simply analyzing a drop of blood. (www.actabiomedica.it)

Key words: Proteomics, human proteome, peptide profiling, mass spectrometry, MALDI-TOF, biomarkers, biosignature profile, peack alignment, sample processing

Introduction

The elucidation of the human genome has provided investigators with extraordinary insights into human diseases. The final output of any gene is a protein, and it is proteins which confer the phenotypes of normal and pathological human conditions. Since multiple post-translational events that can modify the biological structure, function, and degradation of proteins (eg. glycosylation, proteolytic cleavage) the knowledge of the gene structures that encode proteins does not even begin to describe the full complexity of biological systems.

The "proteome" is a term that was coined in 1994 and refers to all the proteins in a cell, tissue, fluid, or organism. "Proteomics" is the study of proteins and can be subdivided into various disciplines related to protein 3D structure, function, and inter- and intracellular communication. Scientists using mass spectrometry to study proteins have recently revealed that human serum contains thousands of small, low molecular weight (LMW) peptides, most of which are fragments of larger precursor proteins. The role that LMW peptides serve in the overall biological human system, however, remains largely undetermined. The complex pattern that they create, however, may provide a novel and robust correlate of the biological events occurring in the entire organism. The development of cancer involves the transformation and proliferation of altered cell types which produce unique proteins

and enzymes that can significantly modify this pattern of serum peptides and proteins. Recent advances in mass spectrometry now enable clinical investigators to rapidly measure and sort these peptides based on mass (m) and charge (z). The serum peptide profile (SPP) that is generated appears to hold important information that may have direct clinical utility. Rather than searching for new proteins as specific tumor markers, this serum peptide profile itself becomes a new kind of potential tumor marker (1).

The first report to apply serum proteomics to the detection of cancer was published in the Lancet in 2002 by Petricoin et al. (2). They found that they could differentiate, with a high degree of accuracy, the serum of normal volunteers from those of women with early stage ovarian carcinoma. In this first report, 63 of 66 (95%) normal control samples were correctly classified as non-cancer, while 50 of 50 (100%) ovarian cancer samples were correctly classified as malignant. The experiment yielded a 100% sensitivity and 95% specificity with a positive predictive value of 94% compared to only 35% for the current ovarian cancer tumor marker, CA-125. This was the first proof of principle that the serum proteomic pattern itself could be used as a valid tumor marker, rather than one isolated specific protein such as CA-125. This approach was applied to prostate cancer and the same investigators were able to differentiate benign from malignant prostate disease in men with elevated PSA levels. The serum proteomic patterns were blindly tested against a training set and correctly predicted 36 of 38 (95%) men with prostate cancer and 177 of 228 (78%) men with benign prostate conditions (3). The authors concluded in both papers that complex proteomic patterns may reflect the underlying pathologic state of solid organs such as the ovary and prostate. Subsequently, another paper published by Adam et al. found similar results in distinguishing prostate cancer from benign prostatic hypertrophy (BPH) in healthy men using similar serum proteomic techniques (4).

These early reports indicate that important clinical information lies within the complex serum proteomic pattern. In order to accurately extract relevant clinical information from the multitude of data present in the serum peptide profile, sophisticated bioinformatic software is required. Machine learning algorithms can be created to search for discriminatory protein patterns that can exceed the diagnostic accuracy of conventional biomarkers or even tissue biopsies. (5). Such a high level of discrimination will be required to apply to individual cancer patients who are themselves a highly diverse population possessing unique peptide profiles that differ according to the type of tumor present. In the case of our study population, thyroid cancer was used as the model.

All thyroid cancers initially arise as thyroid nodules. In the US, there are approximately 23,000 new cases of thyroid cancer each year. There are over 300,000 thyroid cancer survivors in the US who are under surveillance. The MSKCC thyroid cancer group has made significant improvements to this standard testing approach (6-8) by focusing on a small number of tests with the highest diagnostic accuracy, to create a streamlined surveillance protocol (9-11). The availability of a single test that could predict no evidence of disease with high accuracy would prevent the vast majority of these individuals from having regular full evaluations. In addition, the serum proteome could be used to detect those patients with residual disease and provide a likelihood prediction that the lesions would respond to radioactive iodine, additional surgery, chemotherapy, or external beam radiotherapy.

The potential scenarios that serum proteomics could assist in making clinical decisions related to thyroid cancer patients is demonstrated in Figure 1. At locus (A), a SPP that could distinguish benign from malignant nodules may potentially save thousands of patients from undergoing unnecessary thyroid resections. At locus (B), a SPP that is able to identify those patients that completely respond to radioactive iodine remnant ablation therapy could then be used to place patients in a low risk group that would not need to undergo full surveillance testing one year later. Finally, at locus (C), a SPP would aid clinicians in identifying which high risk patients need further surveillance testing after high dose radioactive iodine therapy. The current methods (ie. thyroid ultrasound, whole body scanning, and thyroglobulin levels) used to assist clinicians all possess limitations in diagnostic accuracy that may be exceeded by an individual's unique serum peptide profile.

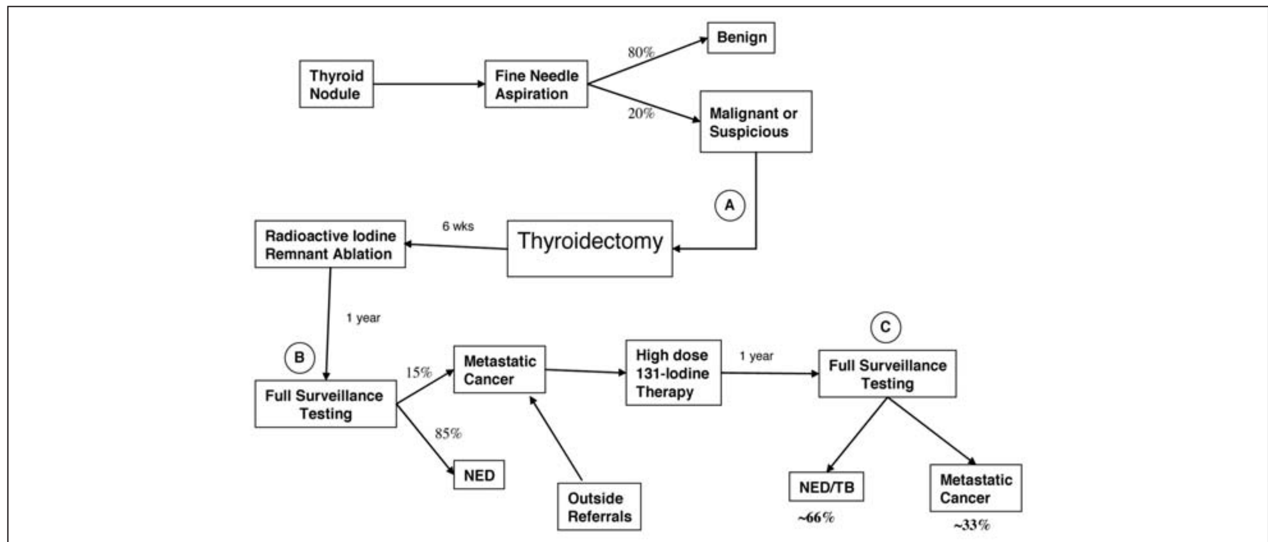

Figure 1. Serum proteomics and clinical decision making

Applying the basic science principles of mass spectrometry to translate into biomarker discovery and thus eventually altering bedside clinical medicine has only recently been even considered a possibility. The 2002 Nobel Prize in Chemistry was granted for the innovation of protein ionization methods which are crucial to measuring the mass-to-charge ratio (m/z) of ionized peptides residing in the gas phase (12). Mass spectrometry has also undergone a revolution in the speed and accuracy of the data that can be generated, and peptides at the femtomolar range can now be repeatedly identified (13). These LMW proteins make up only 1% of the serum proteome, whereas the remaining 99% of the protein content of plasma is represented by only 22 proteins. Detecting the remaining LMW proteins for use in biomarker discovery is thus extremely challenging (14). One recent hypothesis is that the LMW proteins may actually be the result of the complex tumor-host microenvironment, which sets off a cascade of events that generates an array of peptide cleavage products that are then released into the circulation and captured in the serum collection tube (15). We have undertaken a translational project utilizing the latest advancements in mass spectrometry to identify potential patterns of LMW peptides that distinguish the serum of patients with metastatic thyroid cancer from those of normal controls.

Methods

All serum was collected under a standard protocol approved by the IRB at Memorial Sloan Kettering Cancer Center (MSKCC). Serum samples were collected in SST tubes, allowed to clot for one hour, centrifuged for 10 minutes, and then the upper phase transferred to 4-ml cryovials and frozen at -80°C. Peptides were extracted by a reverse-phase method in combination with a magnetic particle-assisted processing technique developed in the Protein Center. A Genesis Freedom 100 automated robotic workstation was developed to carry out the serum peptide extraction and spotting on a Bruker stainless steel MALDI target plate. MALDI target plates were then analyzed with an Ultraflex MALDI-TOF mass spectrometer with an emphasis on the identification of LMW peptides with molecular mass of 0.8-4 kiloDaltons (kDa). Once all LMW peptides were identified, a spreadsheet was created with all the unique peptides and GeneSpring Software is used to evaluate proteomic data (16). Figure 2 is a flow sheet demonstrating the processing of serum samples from the time the blood is drawn to the printout generated by GeneSpring separating peptides of thyroid cancer patients and normal controls.

Once the data are acquired, each data point representing a unique peptide is stored with a naming con-

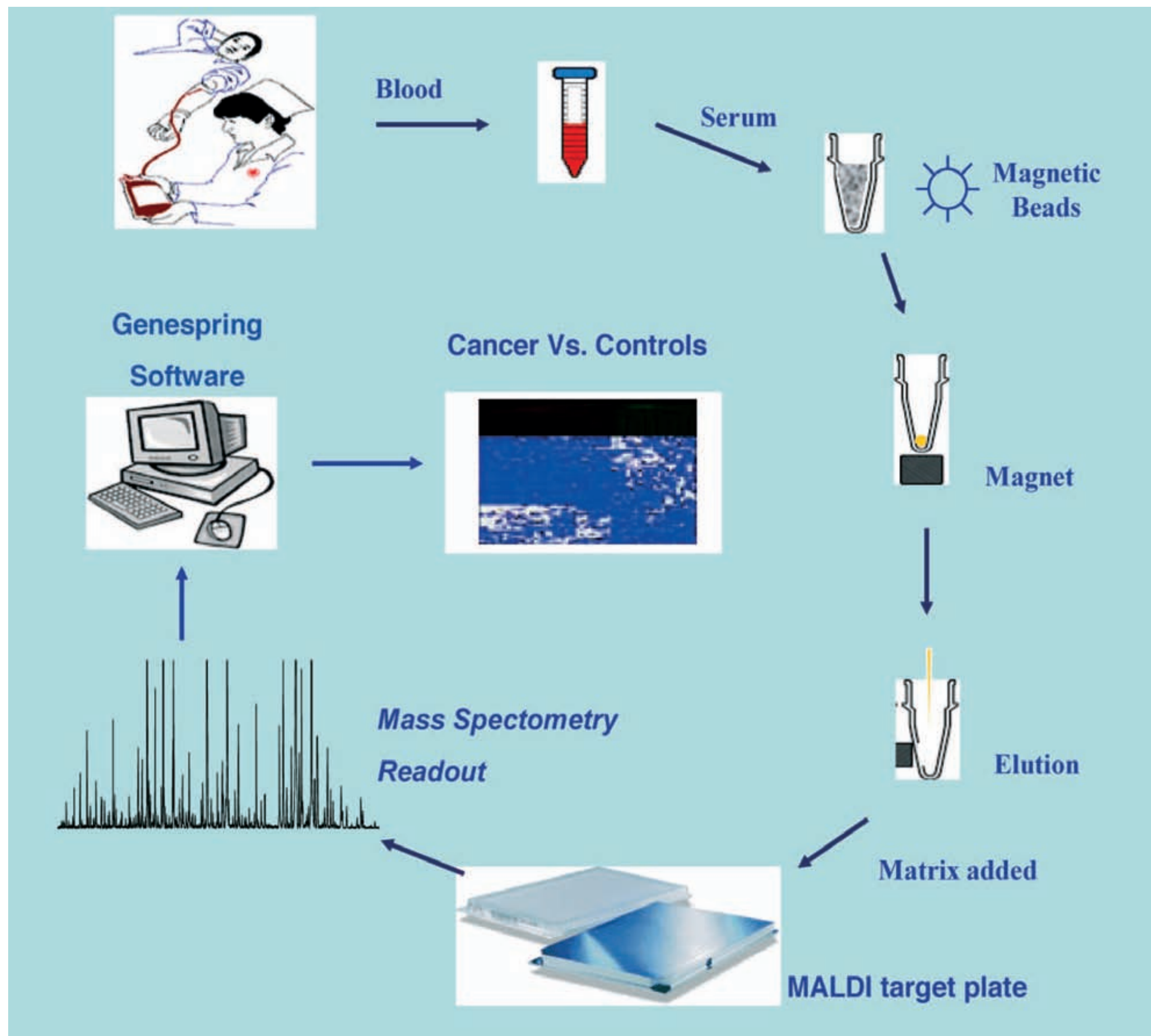

Figure 2. Processing of serum samples

vention that allows each sample to be associated with a unique calibrant. Further data processing was done in MATLAB with a custom script called 'Qcealign' (a software program that helps organize the original data), which creates a reference file to which all sample spectra are aligned using a custom alignment software algorithm called "entropycal" (17). The peptide peaks were filtered by using a non-parametric test (Mann-Whitney U test). The Benjamimi and Hochberg method was used to adjust p-values for multiple comparisons. These tests are meant to identify peptides which, when analyzed as a group, demonstrate a statistically significant difference between the two groups studied (17).

Results

From our cohort of 59 sample patients (32 healthy controls and 27 thyroid cancer patients) we identified a total of 549 unique peptides that were imported in-

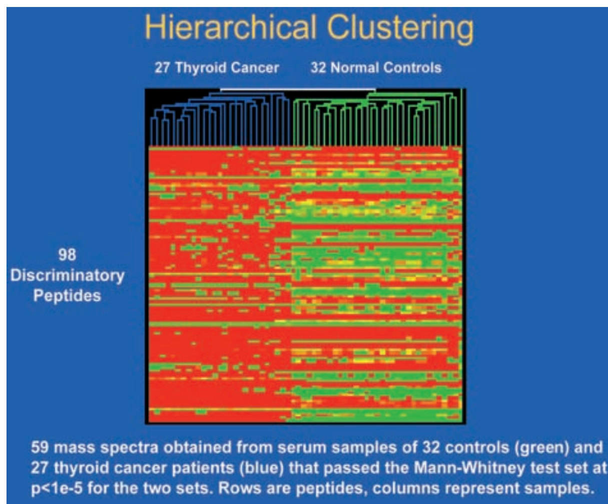

Figure 3. Peptide mapping diagram

to the GeneSpring program and analyzed using various statistical algorithms such as ANOVA, PCA, hierarchical clustering, K nearest neighbor, and support vector machines. Our group has determined that several pre-analytical variables strongly affect the outcome of the analyzed peptides. The type of tube, time set aside for clotting, and the number of freeze-thaw cycles, must all be standardized in order to ensure reproducibility of peptide profiles. We currently use SST tubes for serum collection, allow a clotting time of one hour, and limit the number for freeze-thaw cycles to only two. In addition, reproducibility of the system has been ensured after a vigorous, randomized testing protocol over a seven week period with seven samples taken from the same test batch of commercially available human serum. To ensure that data processing is done with minimal alteration of the original peptide profile, external calibration is done for each sample. Prior to statistical analysis by GeneSpring, the peptide profile undergoes a series of processing step including data smoothing, baseline correction, normalization, calibration/alignment, and peak labeling. All steps have been standardized and will be used in the future in a similar manner to ensure continuity of peptide profile data processing.

Thyroid cancer was chosen as a model for experimentation due to the availability of multiple found specimens in the clinical chemistry lab that were collected under the proper protocol. We collected serum from 27 thyroid cancer patients that were known to have widely metastatic disease and compared their serum to 32 healthy normal controls. All 59 serum samples generated 549 unique peptides that were used to separate serum samples in an unsupervised manner. Of the 549 peptides, 98 peptides passed a Mann-Whitney test at $p < 1 \times 10\text{-}5$. Clustering analysis using this highly discriminatory group of 98 peptides resulted in far superior separation of thyroid cancer patients from their corresponding normal control group. A peptide mapping diagram is available as Figure 3 which demonstrates how convincing the separation is between thyroid cancer patients and normal controls. All 98 peptides were individually analyzed and the sequencing and thus identification of their parent proteins is nearly complete. Their biomarker potential will be analyzed in future experiments.

Conclusions

The detection of low molecular weight peptides in human serum using mass spectrometry appears to have great clinical potential. Identifying peptide patterns that serve as new potential biomarkers in the early stage of disease or to monitor disease progress is now possible by combining mass spectrometry with high-powered statistical software packages. The Protein Center at MSKCC in collaboration with the Endocrinology Service has demonstrated that a group of 98 discriminatory peptides can be used to distinguish the sera of patients with metastatic thyroid cancer from healthy controls. We hope to use similar techniques in the future to identify new peptide profiles for patients with thyroid cancer at various stages of disease. Our goal is to develop a peptide profile unique to thyroid cancer that can then be used to follow patients throughout their entire course of treatment.

References

1. Simpson RJ. Proteins and proteomics: a laboratory manual. Cold Spring Harbor, Cold Spring Harbor Laboratory Press, 2002: chapter 1.
2. Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002; 359: 572-7.

3. Petricoin EF, Ornstein DK, Paweletz CP, et al. Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst* 2002; 94: 1576-8.
4. Adam BL, Qu Y, Davis JW, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* 2002; 62: 3609-14.
5. Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem* 2002; 48: 1296-304.
6. Robbins R, Drucker W, Hann L, Tuttle RM. Advances in the detection of residual thyroid carcinoma. *Adv Intern Med* 2001; 46: 277-294.
7. Wang W, Macapinlac H, Larson SM, et al. [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography localizes residual thyroid cancer in patients with negative diagnostic (131)I whole body scans and elevated serum thyroglobulin levels. *J Clin Endocrinol Metab* 1999; 84: 2291-2302.
8. Wang W, Larson SM, Fazzari M, et al. Prognostic value of [18F]fluorodeoxyglucose positron emission tomographic scanning in patients with thyroid cancer. *J Clin Endocrinol Metab* 2002; 85: 1107-13.
9. Mazzaferri EL, Robbins RJ, Spencer CA, et al. A consensus report of the role of serum thyroglobulin as a monitoring method for low-risk patients with papillary thyroid carcinoma. *J Clin Endocrinol Metab* 2003; 88: 1433-41.
10. Robbins RJ, Tuttle RM, Sharaf RN, et al. Preparation by recombinant human thyrotropin or thyroid hormone withdrawal are comparable for the detection of residual differentiated thyroid carcinoma. *J Clin Endocrinol Metab* 2001; 86: 619-625.
11. Robbins RJ, Chon JT, Fleisher M, et al. Is the serum thyroglobulin response to recombinant human thyrotropin sufficient, by itself, to monitor for residual thyroid carcinoma? *J Clin Endocrinol Metab* 2002; 87: 3242-324.
12. Aebersold R, Mann M. Mass Spectrometry-Based Proteomics. *Nature* 2003; 422: 198-207.
13. Petricoin E, Liotta L. Mass Spectrometry-based Diagnostics: The Upcoming Revolution in Disease Detection. Clin Chem 2003; 49: 533-4.
14. Tirumalai R, Chan K, Prieto D, Issaq H, Conrads T, Veenstra T. Characterization of the Low Molecular Weight Human Serum Proteome. *Mol Cell Proteomics* 2004; 3: 311-26.
15. Conrads T, Fusaro V, Ross S, et al. High-resolution serum proteomic features for ovarian cancer detection. *Endocrine-Related Cancer* 2004; 11: 163-78.
16. 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.
17. Villanueva J, Philip J, Chaparro C, et al. Correcting Common Errors in Identifying Cancer-Specific Serum Peptide Signatures. *J Proteome Res* 2005; 4: 1060-72.

Correspondence: Andrew Martorella,
Department of Medicine, Endocrinology Service,
Memorial Sloan-Kettering Cancer Center, 1275 York Avenue
New York, NY 10021,
Tel. 212-639-3109,
Fax 646-422-2245
E-mail: martorea@mskcc.org