

FIG. 4. Detection of differentially expressed proteins in poorly motile Capan-1 and highly motile BxPC3 pancreatic cells. *A*, overlap of spots appearing in Capan-1 cells (*red*) and BxPC3 cells (*green*) in the 400–16,000 *m/z* and 0–40-min range. *B*, differentially expressed spots in Capan-1 cells (*red*) and BxPC3 cells (*green*) in the 400–16,000 *m/z* and 0–40-min range. The levels of expression of 15,407 peaks (*spots*) were concluded to differ significantly between the two cell lines ($p < 0.01$, Student's *t* test, between triplicates). *C*, a peptide differentially expressed in Capan-1 cells (*left*) and BxPC3 cells (*right*) at 614.0 *m/z* and 13.0 min. This peptide was subsequently identified as being derived from the product of the BCSG1 by MS/MS. *D*, immunoblot confirmation of the differential expression of BCSG1, cytoke-
 18 in Capan-1 cells (*left*) and BxPC3 cells (*right*).

We currently use a high performance computer cluster consisting of four processors with a clock speed of 3.6 GHz connected in parallel to overcome this limitation. We conclude that 2DICAL is a simple, high throughput, and highly reproducible platform that will provide a new paradigm of quantitative proteomics. The software will be made available to the scientific community.

* This work was supported by the "Program for Promotion of Fundamental Studies in Health Sciences" conducted by the National Institute of Biomedical Innovation of Japan and the "Third-Term Comprehensive Control Research for Cancer" conducted by the Ministry of Health, Labor and Welfare of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Chemotherapy Division, National Cancer Center Research Inst., Tokyo 104-0045, Japan. Tel.: 81-3-3542-2511; Fax: 81-3-3547-6045; E-mail: tyamada@gan2.res.ncc.go.jp.

REFERENCES

- Herbert, B. (1999) Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis* **20**, 660–663
- Oh-Ishi, M., Satoh, M., and Maeda, T. (2000) Preparative two-dimensional gel electrophoresis with agarose gels in the first dimension for high molecular mass proteins. *Electrophoresis* **21**, 1653–1669
- Swanson, S. K., and Washburn, M. P. (2005) The continuing evolution of shotgun proteomics. *Drug Discov. Today* **10**, 719–725
- Wolters, D. A., Washburn, M. P., and Yates, J. R., III (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73**, 5683–5690
- Li, X. J., Pedrioli, P. G., Eng, J., Martin, D., Yi, E. C., Lee, H., and Aebersold, R. (2004) A tool to visualize and evaluate data obtained by liquid chromatography-electrospray ionization-mass spectrometry. *Anal. Chem.* **76**, 3856–3860
- Natsume, T., Yamauchi, Y., Nakayama, H., Shinkawa, T., Yanagida, M., Takahashi, N., and Isobe, T. (2002) A direct nanoflow liquid chromatography-tandem mass spectrometry system for interaction proteomics. *Anal. Chem.* **74**, 4725–4733
- Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207
- Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B. B., Simpson, R. J., Eddes, J. S., Kapp, E. A., Moritz, R. L., Chan, D. W., Rai, A. J., Admon, A., Aebersold, R., Eng, J., Hancock, W. S., Hefta, S. A., Meyer, H., Paik, Y. K., Yoo, J. S., Ping, P., Pounds, J., Adkins, J., Qian, X., Wang, R., Wasinger, V., Wu, C. Y., Zhao, X., Zeng, R., Archakov, A., Tsugita, A., Beer, I., Pandey, A., Pisano, M., Andrews, P., Tammen, H., Speicher, D. W., and Hanash, S. M. (2005) Overview of the HUPO Plasma Proteome Project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **5**, 3226–3245
- Fujii, K., Nakano, T., Kanazawa, M., Akimoto, S., Hirano, T., Kato, H., and Nishimura, T. (2005) Clinical-scale high-throughput human plasma proteome analysis: lung adenocarcinoma. *Proteomics* **5**, 1150–1159
- Li, X. J., Yi, E. C., Kemp, C. J., Zhang, H., and Aebersold, R. (2005) A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry. *Mol. Cell. Proteomics* **4**, 1328–1340
- Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999
- DeSouza, L., Diehl, G., Rodrigues, M. J., Guo, J., Romaschin, A. D., Colgan, T. J., and Siu, K. W. (2005) Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. *J. Proteome Res.* **4**, 377–386
- Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386
- Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T. A., Hill, L. R., Norton, S., Kumar, P., Anderle, M., and Becker, C. H. (2003) Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* **75**, 4818–4826
- Radulovic, D., Jelveh, S., Ryu, S., Hamilton, T. G., Foss, E., Mao, Y., and Emili, A. (2004) Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. *Mol. Cell. Proteomics* **3**, 984–997
- Silva, J. C., Denny, R., Dorschel, C. A., Gorenstein, M., Kass, I. J., Li, G. Z., McKenna, T., Nold, M. J., Richardson, K., Young, P., and Geromanos, S. (2005) Quantitative proteomic analysis by accurate mass retention time pairs. *Anal. Chem.* **77**, 2187–2200
- Anderle, M., Roy, S., Lin, H., Becker, C., and Joho, K. (2004) Quantifying reproducibility for differential proteomics: noise analysis for protein liquid chromatography-mass spectrometry of human serum. *Bioinformatics* **20**, 3575–3582
- Weir, B. S. (1988) Statistical analysis of DNA sequences. *J. Natl. Cancer Inst.* **80**, 395–406
- Honda, K., Yamada, T., Hayashida, Y., Idogawa, M., Sato, S., Hasegawa, F., Ino, Y., Ono, M., and Hirohashi, S. (2005) Actinin-4 increases cell motility and promotes lymph node metastasis of colorectal cancer. *Gastroenterology* **128**, 51–62
- Shimamura, T., Yasuda, J., Ino, Y., Gotoh, M., Tsuchiya, A., Nakajima, A., Sakamoto, M., Kanai, Y., and Hirohashi, S. (2004) Dysadherin expression facilitates cell motility and metastatic potential of human pancreatic cancer cells. *Cancer Res.* **64**, 6989–6995
- Honda, K., Hayashida, Y., Umaki, T., Okusaka, T., Kosuge, T., Kikuchi, S., Endo, M., Tsuchida, A., Aoki, T., Itoi, T., Moriyasu, F., Hirohashi, S., and Yamada, T. (2005) Possible detection of pancreatic cancer by plasma protein profiling. *Cancer Res.* **65**, 10613–10622
- Breen, E. J., Hopwood, F. G., Williams, K. L., and Wilkins, M. R. (2000) Automatic Poisson peak harvesting for high throughput protein identification. *Electrophoresis* **21**, 2243–2251
- Gras, R., Muller, M., Gasteiger, E., Gay, S., Binz, P. A., Bienvenu, W., Hoogland, C., Sanchez, J. C., Bairoch, A., Hochstrasser, D. F., and Appel, R. D. (1999) Improving protein identification from peptide mass fingerprinting through a parameterized multi-level scoring algorithm and an optimized peak detection. *Electrophoresis* **20**, 3535–3550
- Giegerich, R. (2000) A systematic approach to dynamic programming in bioinformatics. *Bioinformatics* **16**, 665–677
- Seike, M., Kondo, T., Mori, Y., Gemma, A., Kudoh, S., Sakamoto, M., Yamada, T., and Hirohashi, S. (2003) Proteomic analysis of intestinal epithelial cells expressing stabilized beta-catenin. *Cancer Res.* **63**, 4641–4647
- Reinders, J., Lewandrowski, U., Moebius, J., Wagner, Y., and Sickmann, A. (2004) Challenges in mass spectrometry-based proteomics. *Proteomics* **4**, 3686–3703
- Ji, H., Liu, Y. E., Jia, T., Wang, M., Liu, J., Xiao, G., Joseph, B. K., Rosen, C., and Shi, Y. E. (1997) Identification of a breast cancer-specific gene, BCSG1, by direct differential cDNA sequencing. *Cancer Res.* **57**, 759–764
- Zhang, H., Yi, E. C., Li, X. J., Mallick, P., Kelly-Spratt, K. S., Masselon, C. D., Camp, D. G., 2nd, Smith, R. D., Kemp, C. J., and Aebersold, R. (2005) High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* **4**, 144–155
- Bogdanov, B., and Smith, R. D. (2005) Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom. Rev.* **24**, 168–200
- Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867

日本臨牀 第64巻・第9号（平成18年9月号）別刷

‘膵癌’早期発見の新技术

本田一文 尾野雅哉 下重美紀 山田哲司

‘膵癌’ 早期発見の新技术

本田一文 尾野雅哉 下重美紀 山田哲司

Early detection of pancreatic cancer by novel proteomic technique

Kazufumi Honda, Masaya Ono, Miki Shitashige, Tesshi Yamada
Chemotherapy Division and Cancer Proteomics Project,
National Cancer Center Research Institute

Abstract

Pancreatic cancer is the fifth leading cause of cancer-related mortality in Japan. Early detection in pancreatic cancer is one of the most feasible strategies to improve outcome. We compared plasma proteome between pancreatic cancer patients and healthy controls using surface-enhanced laser desorption/ionization coupled with hybrid quadrupole time-of-flight mass spectrometry. Proteomic spectra were generated from a total of 245 plasma samples obtained from two institutes. A discriminating proteomic pattern was built from training cohort using machine learning algorithm and was applied two validation cohorts. This set discriminating cancer patients in the first validation cohort with a sensitivity of 90.9% and a specificity of 91.9%, and was further validated in an independent cohort at a second institution. When combined with CA19-9, 100% tumor of pancreatic cancers, including early stage tumors, were detected. In this report, we describe a possible detection of early pancreatic cancer using novel proteomic technique.

Key words: plasma proteome, pancreatic cancer, plasma biomarker, peptide profile

はじめに

膵癌の5年生存率は主な固形腫瘍の中では最も低率であり、我が国では年間約19,000人以上が膵癌で死亡する¹⁾。膵癌は間質浸潤性が強い腫瘍で、早期から肝臓や所属リンパ節への転移をみる。膵癌に対する最も信頼性の高い治療法に外科切除があげられるが、膵臓は解剖学的に後腹膜に位置し、臨床症状が乏しく、更に超音波診断や臨床理学的な所見を取りづら

いため診断が難しく、外来を訪れる膵癌患者のうち多くは、病期の進行により外科治療をあきらめざるを得ないというのが現状である。事実、膵癌と診断された患者のうち、手術適応があったのは20-40%程度といわれている^{1,2)}。CT(computed tomography)やMRI(magnetic resonance imaging)による集団検診に注目が集まっているが、医療経済学的には必ずしも効率的であるとは言い難く、PET(positron emission tomography)による検診の有用性につ

いては、いまだ意見の分かれるところであり³⁾、新たな効率の良い検診法が求められている。

治療可能な膵癌を効率的に検出できる血漿・血清診断マーカーがあれば予後の改善に著しく貢献するものと思われる。近年、ゲノム異常の蓄積が発がん過程に関与することが明らかになり、2001年には国際ヒトゲノムプロジェクトにより、約30億個の遺伝暗号の解読が発表された⁴⁾。発がんの原因はゲノム異常の蓄積であるとしても、生体内の事象に対して直接的な機能を有するものは、そこから転写、翻訳された後の蛋白質である。ゲノム情報は、蛋白質情報に比べれば静的であり、時事刻々変化する病態に伴い鋭敏に変化していくのもまた蛋白質や糖鎖である。非侵襲的に病態診断を診断できる血漿・血清マーカーを開発するうえで血液中に循環している蛋白質や糖鎖変化を網羅的に探索することは、有効な戦略になり得るわけである。

昨今のバイオテクノロジーの発展は、個々の事象による特定の蛋白質の発現変動を解析するのではなく、蛋白質発現情報を全体として鳥瞰的にとらえて解析するプロテオーム解析を可能にした。なかでも質量分析 (mass spectrometry: MS) の発達は目覚ましく、現在では血液中に含まれる蛋白質の質量と、そのイオン化率をラベルやタグを付加することなく再現性良くハイスループットに解析することが可能になっている。この情報を利用した診断マーカー探索の試みが世界的になされ始め、なかでも米国国立がん研究所 (National Cancer Institute: NCI) と米国食品医薬品局 (Food and Drug Administration: FDA) の研究グループから SELDI-TOF-MS 法 (surface enhanced laser desorption/ionization time of flight mass spectrometry) と情報工学の手法を組み合わせた解析から、早期卵巣癌を感度 100%、特異度 95% の正確性で診断できる血清ペプチドプロファイルが報告されると⁵⁾、同様なプラットフォームを用いて、世界中の様々なグループによって、多様ながんでの腫瘍マーカーの開発がなされている。

国立がんセンター研究所化学療法部・腫瘍プロテオミクスプロジェクトでは、膵癌に対する

信頼性の高い血漿腫瘍マーカー開発を目指して、多くの膵癌患者の血漿からペプチドプロファイルを取得している。ペプチドプロファイル取得には、質量精度や分解能の高い質量分析計を用い、更に独自に開発したバイオインフォマティクスソフトウェアを用いて、マーカー探索を行っている。

本稿では、国際的に行われているプロテオーム解析を使った診断マーカー探索の実情と、国立がんセンター研究所で行われている血漿ペプチドプロファイルを用いた膵癌患者検出の試みならびに今後の実用化に向けた取り組みについて概説する。

1. 質量分析を用いた腫瘍マーカー探索の試み

a. 国際コンソーシアムによる血漿・血清プロテオーム解析

2001年にヒトプロテオーム解析に対する国際諮問会議 HUPO (Human Proteome Organization) が結成され、国際協力体制の中で、‘ヒト・ゲノム・プロジェクト’に続く国際事業として‘ヒト・プロテオーム・プロジェクト’構想の促進が図れるようになった。その一環として HUPO 主体の ‘Plasma Proteome Project’ が 2002 年から開始され、血漿に含まれる蛋白質の網羅的解析とデータベース化ならびに解析技術の開発が進められている⁶⁾。血清や血漿には数千種類の蛋白質が含まれていると考えられているが、その 99% がアルブミンをはじめとする 22 種類のメジャー蛋白質に占められているため、少ない含有量の蛋白質を網羅的に解析するためには、アルブミンや免疫グロブリンなどの蛋白質を除いて解析する必要がある。現在まで 3,000 種類強の蛋白質が血漿中から同定されているが、このような試みはあくまで血漿蛋白質のリスト化と研究手法の整備であり、直接的に腫瘍マーカー探索を行っているわけではない。更に最近ではがん診断マーカーの探索に特化した国際コンソーシアムとして ICBC (International Cancer Biomarker Consortium; http://www.fhcr.org/science/international_biomarker/) が立ち

上がり、国際的な協力関係のもとマーカー探索が進められつつある。

b. SELDI-TOF-MS法, MALDI-TOF-MS (matrix assisted laser desorption/ionization time of flight mass spectrometry) 法を用いた血漿・血清ペプチドプロファイルによる診断マーカー探索

先に述べたが、PetricoinらがSELDI-TOF-MS法と情報工学手法を用いて、早期卵巣癌も含めて正診率の高い血清ペプチドプロファイルの同定に成功すると⁹⁾、腫瘍マーカーを探索していた多くの研究者に強い衝撃を与えた。その後、前立腺癌での血清ペプチドプロファイルを利用した決定木アルゴリズムによる診断法などの報告が続き⁷⁾、SELDI-TOF-MS法によるマーカー探索の有用性が確立されていく。SELDI-TOF-MS法は、ラベルやタグ、酵素消化などを必要としないため、生体に循環しているネイティブな状態の蛋白質を解析できるという利点をもつ。更に、多検体処理を目的として開発されたツールであるため、バイオロボットによる自動化が容易であり、わずか20 μ lの血液から網羅的な蛋白質情報を引き出すことができる⁸⁾。

しかしながら、初期のSELDI-TOF-MS法による血清・血漿ペプチドプロファイル法による診断法の報告は、十分な数の検体数が検討されていなかったり、検討されていたとしても、診断プロファイルの統計学的な導出のみで、その後の診断精度の検証が導出された検体群(学習セット)とは別の検体(検証セット)を用いて、検証がなされていないなどの問題点があるものも多く見受けられた。最近では、これら問題を回避するために、学習セットは診断モデルの導出のみにとどめ、その診断モデルの信頼性に関しては、学習セットには使用されなかった検証セットを用いて検討することが推奨されている。更にそのモデルが多施設での検証に耐えられることが重要な要件にあげられる。後に詳述するが、現在著者らは、本方法を用いて膵癌の診断マーカーの探索を行っている。

また、SELDI-TOF-MS法のみならずMALDI-TOF-MS法による診断マーカー探索が試みら

れている。血清、血漿中に含まれる低分子蛋白質はアルブミンのようなキャリア蛋白質に結合することによって、腎臓からの排出を回避すると考えられているため、アルブミンに結合しているペプチドを効率良く集め、そのプロファイルを用いて診断モデルを構築するという試みがなされている^{9,10)}。著者らは、子宮体癌患者に特徴的なアルブミン結合ペプチドプロファイルの中から早期子宮体癌に反応するペプチドマーカーの探索を行ったところ、早期子宮体癌に反応するペプチドピークの存在を明らかにした(Kikuchiら、論文投稿準備中)。

c. ショットガンプロテオミクス法による血漿・血清診断マーカー探索

SELDI-やMALDI-TOF-MS法では、酵素消化を行わない蛋白質を解析の対象とすることが多いため、現在の質量分析計の性能からいうと、観察範囲が30kDa以下のペプチド領域の低分子量蛋白質に制限されている。より大きな分子量をもつ蛋白質を解析するためには、トリプシンなどで、血清・血漿蛋白質を酵素消化した後、質量分析計で解析するショットガンプロテオミクスと呼ばれる方法が有利である。このような解析には、液体クロマトグラフィ(liquid chromatography: LC)と質量分析(MS)を組み合わせ使用するLC/MSと呼ばれる方法で解析されることが多い¹¹⁾。LC/MSによるショットガン解析で多検体比較を行う場合は困難を伴うことがある。なぜならば、LCによる蛋白質の溶出時間の同期を制御しづらいためである。これら問題を回避し単純な2検体間でのディフュゼンシャル解析を行うために、ICAT(isotope-coded affinity tag)やiTRAQなどの*in vitro*での安定同位体蛋白質標識法を用いることで、定量再現性を向上させる方法も考案されているが¹²⁾、多検体に対する同時比較にはいまだ様々な問題点が残っている。著者らは、無標識検体を低速流量(50-200 μ l/分)での液体クロマトグラフィと、ESI(electrospray ionization)型質量分析計をオンラインで接続して計測し、質量と分離時間とイオン化量を2次元画像として描出し解析する2D ICAL(two dimensional image

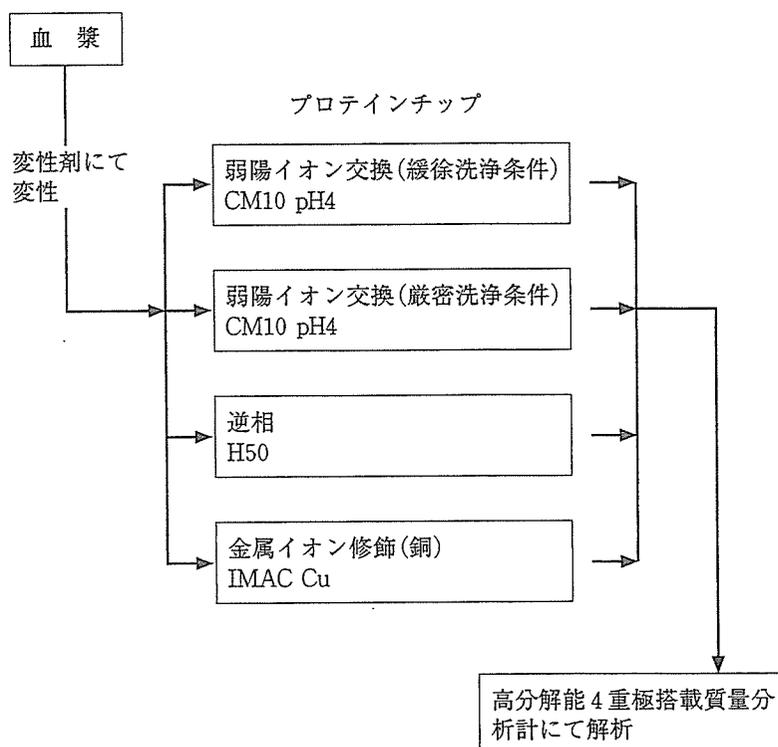


図1 プロテインチップ®を使ったマーカー探索方法

converted analysis of nano-flow liquid chromatography)を開発した。本方法は、ダイナミックアルゴリズムによって液体クロマトグラフィーの溶出時間を補正しながら質量とイオン化量を表すことができるため、多検体処理や解析が可能になるという利点を有している¹³⁾。

2. 国立がんセンター研究所による膵癌血漿ペプチドプロファイルによる診断マーカー開発¹⁴⁾

国立がんセンター研究所化学療法部・腫瘍プロテオミクスプロジェクトでは、検診に使用できる膵癌診断マーカーの開発を目的に SELDI-TOF-MS法を利用して、血漿ペプチドプロファイルを行っている。以下にその詳細を述べる¹⁴⁾。

a. 対象検体と解析方法

本研究に使用したのは、2002年8月から2003年10月までに国立がんセンター中央病院で採取された未治療の膵癌患者(104例)と健常者(116例)の血漿と、東京医科大学病院で2004年2月から2005年2月までに採取された未治療の膵

癌患者(9例)、良性膵臓疾患(11例)と健常者(5例)の計245例分の血漿である。これら、血漿はすべて同一条件で採取され、解析まで -80°C で凍結保存された。これら、血漿検体を変性剤を用いて蛋白変性した後、図1に示す4条件のCIPHERGEN社製のProteinChip®を用いてアフィニティ精製し、4重極搭載高分解能質量分析計Q-Star XL PCI 1000を用いてTOF-MS解析を行った(図1)。

b. 機械学習法を用いた学習セットから膵癌診断モデルの導出

血漿ペプチドプロファイルから診断モデルを導出するために、国立がんセンター中央病院で採血された検体のうち、がん患者と健常者の間に、年齢、性差が偏らないように学習セットを選別した。選別された集団は、膵癌患者が71例、健常者が71例分の計142例分の血漿とした(表1)。血漿プロファイルを質量分析計を用いて取得したのち、すべてのデータをサーバーコンピュータに保存し、膵癌患者と健常者を感度94.4%、特異度97.2%で判別する診断モデルを人工知能などに使われる判別アルゴリズム

表1 国立がんセンター中央病院にて採取された検体の臨床背景(文献¹⁴⁾より改変)

	学習セット			検証セット		
	がん (n=71)	健常者 (n=71)	p-value	がん (n=33)	健常者 (n=45)	p-value
年齢(mean±SD)	61.3±9.06	62.1±10.0	0.6*	62.0±9.06	63.2±11.7	0.6*
性別						
男性	37	33	0.5 [†]	18	24	0.92 [†]
女性	34	38		15	21	
腫瘍部位						
膵頭部	34			17		
膵臓体尾部	37			10		
不明	0			6		
臨床病期						
I	1			1		
II	6			4		
III	10			1		
IV	54			27		

* Fisher's exact probability *t*-test

[†] Student's *t*-test

表2 SELDI 診断モデルの正診率(文献¹⁴⁾より改変)

	学習セット		検証セット	
	症例数	正診率(%)	症例数	正診率(%)
健常者	71	67(94.4%)	45	41(91.1%)
がん	71	69(97.2%)	33	30(90.9%)
腫瘍部位				
膵頭部	34	33(97.1%)	17	14(82.4%)
膵体尾部	37	36(97.3%)	10	10(100%)
不明	0	0	6	6(100%)
臨床病期				
I	1	0(0%)	1	1(100%)
II	6	6(100%)	4	3(75%)
III	10	9(90%)	1	1(100%)
IV	54	54(100%)	27	25(92.6%)

ムを用いて導出した(表2)。導出された診断モデルは、CM10 pH4で観察される8,766m/z, 17,272m/z, 28,080m/z, ならびにH50で観察される14,779m/zの4本のペプチドピークで構成されていた(図2)。これら4本のピーク情報から、膵癌と健常者を分離する分離境界面を機械学習法を用いて設定し、各検体ごとの分離境界面までの距離と方向を計算した。計算された距離と方向を図3-aに示す。この場合、(+)の方向は膵癌、(-)の方向を健常者と定義する。これら4本のペプチドピークを使った場合の

ROC(receiver operation characteristics)曲線を図4に示す。4本のピークを使った診断モデルによる曲線下面積は0.978と非常に高い値を示していた。学習セットでは、7例の臨床病期I, II期の早期膵癌のうち6例を膵癌と診断することが可能であった(表2)。

c. 検証セットを用いた診断モデルの検証

学習セットによって導出された診断モデルの正確性を、国立がんセンター中央病院で採取された血漿検体のうち、診断モデルの導出に使用しなかった検体、検証セットを使って検証した。

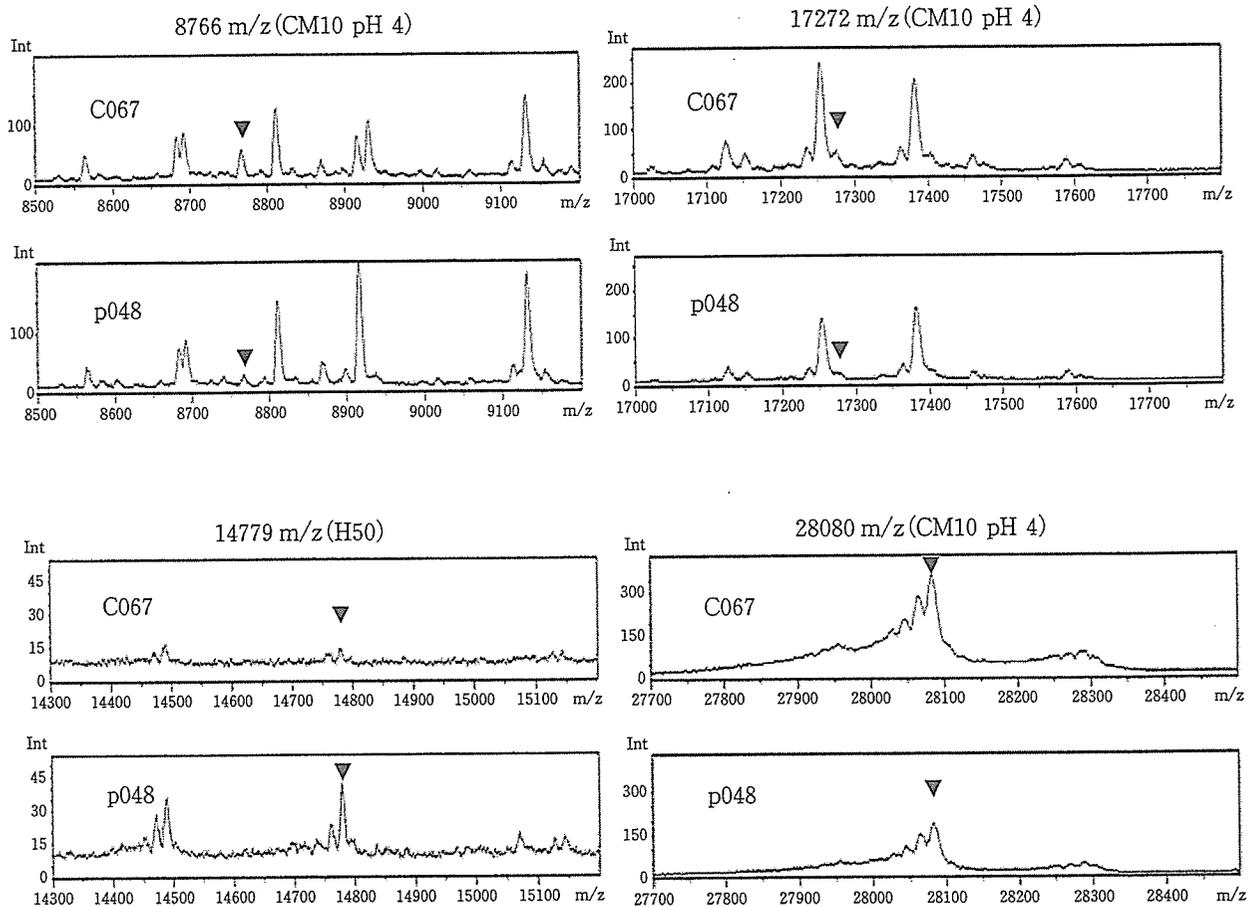


図2 代表的ながん患者と健常者のマーカーピークのスペクトル(文献¹⁴⁾より改変)
C067: 健常者, p048: 膵癌患者

検証セットには、膵癌患者血漿 33 例、健常者 45 例分を使用した。先に述べた 4 本のピークを使って学習セットから定義された分離境界面を固定し、その分離境界面に対する検証セットの各検体ごとに距離と方向を計算した。計算結果を図 3-b に示す。33 例の膵癌患者のうち膵臓癌と判定された症例は 30 例 (90.9%)、健常者 45 例のうち健常者と判定された症例は 41 例 (91.1%) であった。国立がんセンター中央病院の検証セットのうち、臨床病期 I, II 期の早期膵癌が 5 例含まれていたが、4 例を膵癌と診断することが可能だった (表 2)。

d. 多施設検体による診断モデルの検証

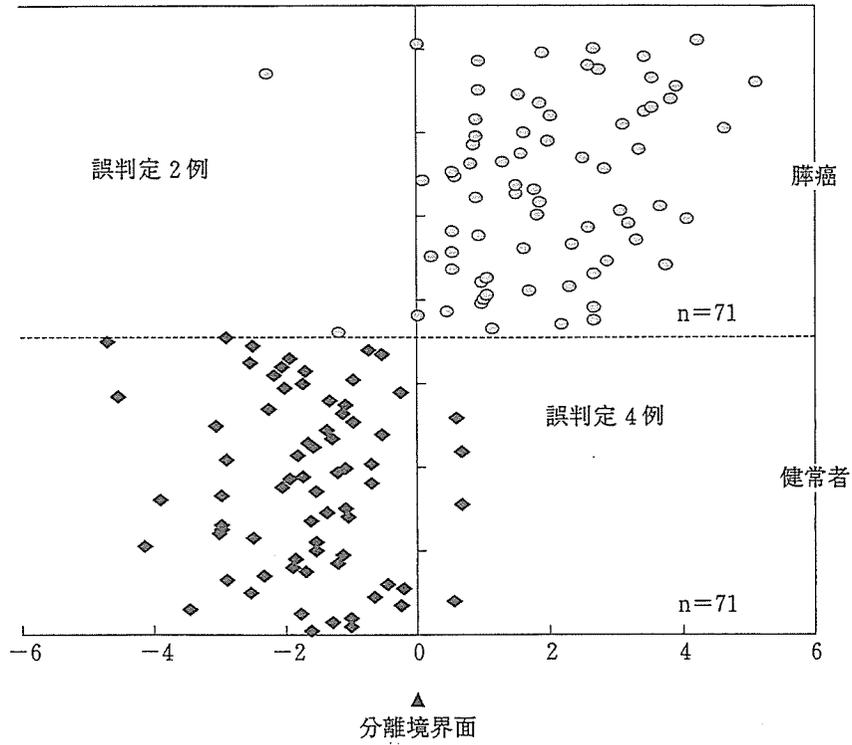
診断モデルの正確性について、単施設で採取された検体を用いて検証したところ、90% 以上の正確性をもって診断できたわけだが、モデル検証に使われた検体は、あくまで国立がんセンター中央病院で採取された検体であるため、

採血条件や保存方法による差異を検出している可能性も否定できない。更に、国立がんセンター中央病院という特殊性から、検体に良性膵疾患が含まれていなかった。そこで、本診断モデルが、他施設から採取された検体でも正確に診断できるかどうかと膵癌と良性膵臓疾患の鑑別への有効性について、東京医科大学病院で採取された血漿を用いて検証を試みた。図 5 に各検体ごとに対する学習セットで定義された分離境界面までの距離と方向を示す。膵癌患者 9 例中 8 例を膵癌と診断することが可能であり、健常者 5 例中の 4 例を正確に診断することができた。良性膵疾患については、良性膵腫瘍の 6 例中 4 例、慢性膵炎 5 例中 1 例を膵癌と判定した。

e. CA19-9 と SELDI 診断モデルとのコンビネーションによる膵癌検出

既存の膵癌腫瘍マーカーとして CA19-9 が知られている。CA19-9 と SELDI 診断モデルのコ

a. 学習セット



b. 検証セット

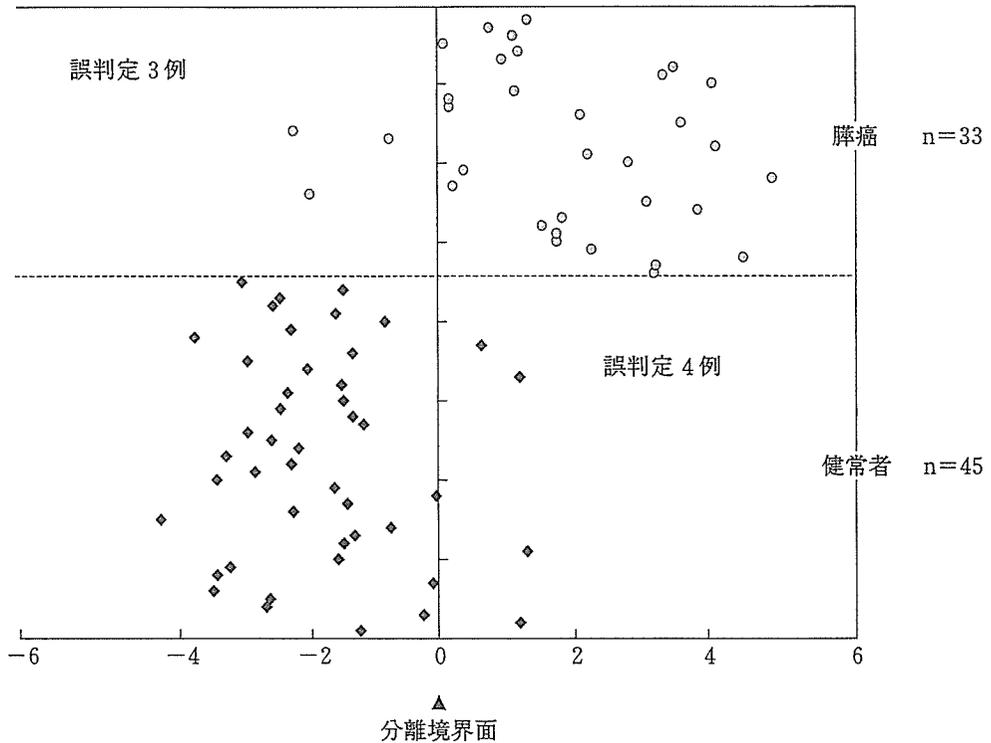


図3 機械学習法により計算された分離境界面からの各症例ごとの距離と方向
(文献¹⁴⁾より改変)

(-)方向は健常者, (+)方向は肺癌と診断する.

a: 学習セット(国立がんセンター中央病院採取検体)

b: 検証セット(国立がんセンター中央病院採取検体)

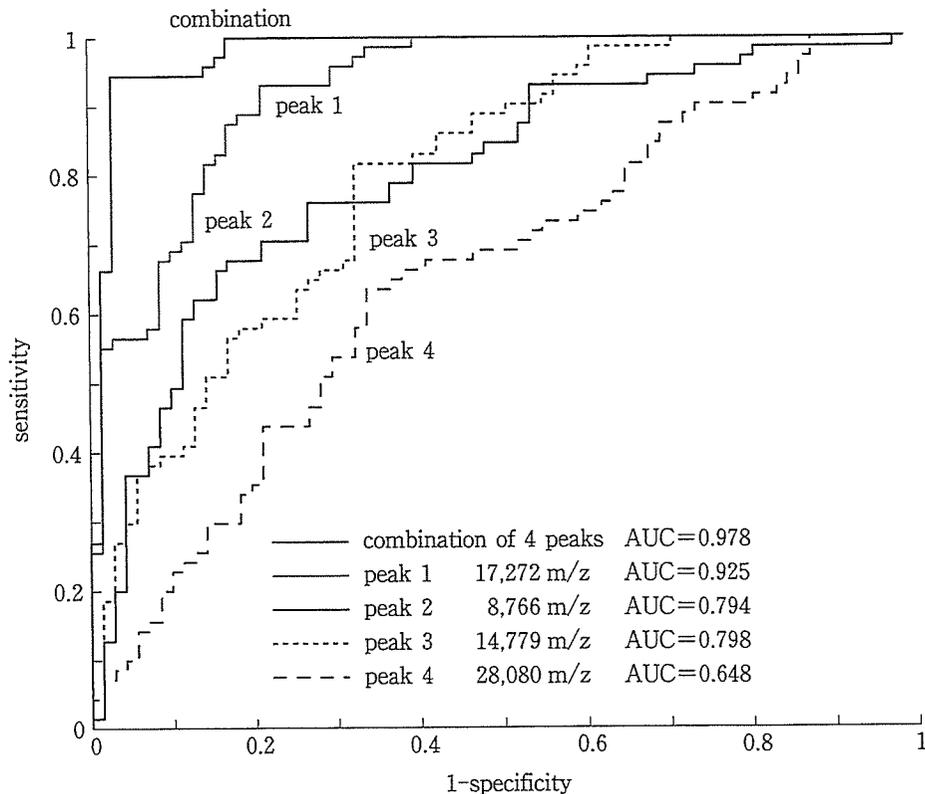


図4 学習セットから計算された各ピークならびに組み合わせた場合のROC曲線
(文献¹⁴⁾より改変)
AUC: ROC曲線下面積

コンビネーションによる診断率の向上を試みた。コンビネーションによる陽性とは、CA19-9を37U/mlをカットオフと定義し、CA19-9もしくはSELDI診断モデルのいずれかが陽性であった場合を、膵癌と判断した。コンビネーションの診断法の検証には、CA19-9が測定可能であり、検証セットとして使用された78例のうち、健常者39例、膵癌患者29例の計68例の血漿を使用した。表3にその結果を示す。CA19-9単独による膵癌患者の検出率は86.2%(25/29)で偽陽性率は7.7%(3/39)であったのに対して、SELDI診断モデルの検出率は89.7%(26/29)でその偽陽性率は10.3%(4/39)であった。SELDI診断モデルは検出率では、CA19-9に勝っていたが、偽陽性率は劣っていた。この2つのコンビネーション診断法では、CA19-9とSELDI診断法の陽性者は相補的であり、臨床病期I、II期の膵癌を含めて100%(39/39)を検出することが可能だった。その偽陽性率は15.4%(6/39)

であった(表3)。

おわりに —実用化に向けた試み—

以上述べてきたとおり、現在までに著者らは国立がんセンター中央病院と東京医科大学病院の2施設から採取された245例の血漿検体を用いて膵癌診断法の開発を行ってきた。しかしながら、本研究で用いられた検体は、2施設から集められた検体であり、十分な検証が行われたとは言い難い。また、用いた質量分析計も研究用に開発された非常に高価で高性能な装置を1台使用して行われたにすぎない。次に著者がなすべきことは、いかにして本診断法を実用化していくかが重要なポイントになってくる。今後、実用化に向けた戦略として以下の2点があげられると思う。

(1) できるだけ多くの施設から膵癌、良性膵疾患、健常者血漿を多く集め、更なる検証とマ

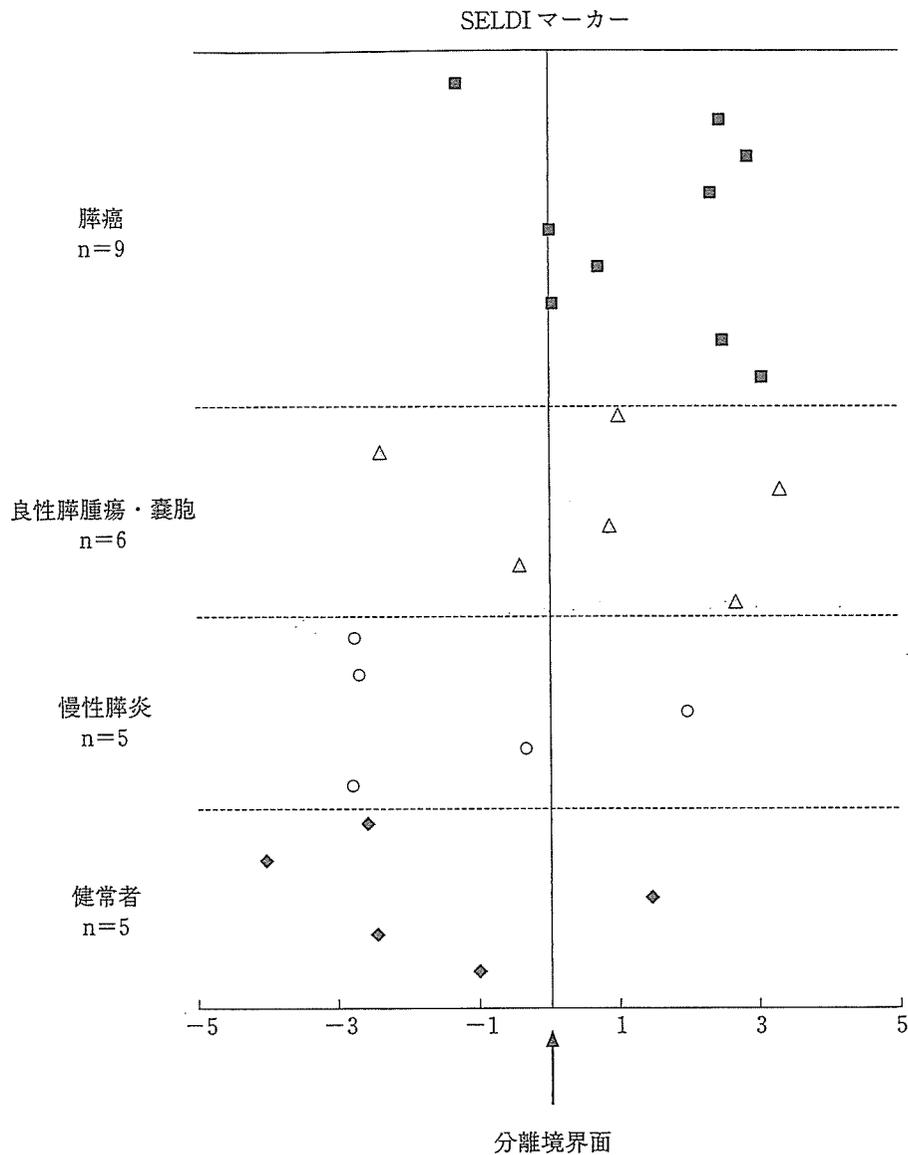


図5 東京医科大学病院の症例を用いた検証(文献¹⁰より改変)

ーカー候補の探索を行う。

(2) 臨床検査として十分に信頼性と再現性を保つクオリティを確保し、できるだけ低コストで、多検体を高速に処理するシステムを構築する。

(1)に関しては本年度から国内6施設の協力を得てシステムチックに血漿が集められ、解析する多施設共同研究がスタートする。本血漿検体を使って、更なる検証とより有効なプロファイルの再探索を行おうと考えている。また、(2)に関していえば、現在使用されている機械、試薬、ProteinChip[®]などすべて研究用クオリティのものであり、必ずしも薬事法に対応した体

外診断法レベルまで確立されたものではない。

また、現在使用している高分解能4重極搭載直行型質量分析計はアミノ酸配列を決定するための機械であり、非常に多くのパラメーターを有している。そのため、複数台の質量分析計を使って再現性良くプロファイルを取得できるように調整するのは簡単ではない。また、研究用に特化された装置のため計測の自動化がなされておらず、高速度な解析は困難である。そのため、できるだけ候補ピークのアミノ酸配列を決定し、免疫生化学検査を構築すべきであるという意見も多い。しかしながら、これらピークの同定はそれほど容易ではなく、また決定したとしても、

表 3 CA19-9 と SELDI 診断モデルのコンビネーションによる正診率
(文献¹⁰より改変)

	症例数*	CA19-9	SELDI 診断モデル	コンビネーション
健常者	39	3(7.7%)	4(10.3%)	6(15.4%)
がん	29	25(86.2%)	26(89.7%)	29(100%)
臨床病期				
I	1	1(100%)	1(100%)	1(100%)
II	3	3(100%)	2(66.7%)	3(100%)
III	1	0(0%)	1(100%)	1(100%)
IV	24	21(87.5%)	22(91.7%)	24(100%)

多くの場合が生体に豊富に存在する蛋白質の特殊な断片であるという報告が多い。であるならば、ELISA 法などによる免疫性生化学検査を構築しても、肝心の質量情報が抜け落ちてしまうために、必ずしも検査として成り立たない可能性がある。すなわち、マーカー自体の蛋白質構造を明らかにしても質量情報が検査のためには必須となるわけである。やはり質量分析を真剣に臨床検査に応用していく時期が来ているのだと思う。ペプチドプロファイルを正確に取得するためだけであれば、質量精度と分解能に優れ

ていれば必ずしもアミノ酸同定のための機構を必要とするものではない。実用化に向けた取り組みとしては、高速度で処理できる機能をもち、機械間の再現性に優れ、高い分解能と質量精度をもち、汎用性のある質量分析計の開発と、薬事法を踏まえた臨床治験を行うための試薬、装置や検査手順などの基盤整備を、研究者、臨床医のみならずシステムエンジニア、メーカー技術者、検査技術者などが一体となって取り組んでいく必要があると考えられる。

■ 文 献

- 1) Yamamoto M, et al: Japan Pancreatic Cancer Registry: current status. *Pancreas* 16: 238-242, 1998.
- 2) Shimamura T, et al: Dysadherin overexpression in pancreatic ductal adenocarcinoma reflects tumor aggressiveness relationship to e-cadherin expression. *J Clin Oncol* 21: 659-667, 2003.
- 3) Berrington de Gonzalez A, Darby S: Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. *Lancet* 363: 345-351, 2004.
- 4) Lander ES, et al: Initial sequencing and analysis of the human genome. *Nature* 409: 860-921, 2001.
- 5) Petricoin EF, et al: Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359: 572-577, 2002.
- 6) Omenn GS, et al: Overview of the HUPO Plasma Proteome Project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* 5: 3326-3345, 2005.
- 7) Adam BL, et al: Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy man. *Cancer Res* 62: 3609-3614, 2002.
- 8) Issaq HJ, et al: SELDI-TOF-MS approach to proteomics: protein profiling and biomarker identification. *Biochem Biophys Res Commun* 292: 587-592, 2002.
- 9) Petricoin EF, Liotta LA: SELDI-TOF-MS-based serum proteomic pattern diagnostics for early detection of cancer. *Curr Opin Biotechnol* 15: 24-30, 2004.

- 10) Lowenthal MS, et al: Analysis of albumin-associated peptides and proteins from ovarian cancer patients. *Clin Chem* **51**: 1933-1945, 2005.
- 11) Fujii K, et al: Fully automated online multidimensional protein profiling system for complex mixture. *J Chromatogr A* **1057**: 107-113, 2004.
- 12) Gygi SP, et al: Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**: 994-999, 1999.
- 13) Ono M, et al: Label-free quantitative proteomics using large peptide data sets generated by nano-flow liquid chromatography and mass spectrometry. *Mol Cell Proteomics*. (in press)
- 14) Honda K, et al: Possible detection of pancreatic cancer by plasma profiling. *Cancer Res* **65**: 10613-10622, 2005.

BIO

バイオ
テクノロジー
ジャーナル

ブレイクスルーを生む、注目の新技術と実験メソッド

隔月刊

別刷

株式会社 羊土社

〒101-0052

東京都千代田区神田小川町2-5-1神田三和ビル

TEL 03-5282-1211 FAX 03-5282-1212

E-mail : btjournal@yodosha.co.jp

HP : <http://www.yodosha.co.jp/btjournal/>

【バイオマーカー探索①】

血液試料の最適な調製法 と解析

下重美紀, 本田一文, 山田哲司

血清・血漿タンパク質には少なくとも 10^3 以上の発現量格差があるといわれており、アルブミン、イムノグロブリンなどの含有量の多いタンパク質の分離除去やゲル濾過、イオン交換などにより分子量、等電点の違いによる分画、レクチンによる糖タンパク質の精製操作が必要である。金属チップ表面を修飾したプロテインチップ (ProteinChip) は特定の血清・血漿タンパク質を定量的に捕捉・濃縮するのに威力を発揮する。

はじめに

近年の質量分析技術の進歩により、疾患の発症ならびに進行に伴う体液や組織、細胞でのタンパク質やペプチドの発現レベルの変動を高感度かつ網羅的に解析できるようになり、疾患の診断に応用する試みがなされている。われわれは膵癌などの難治性癌患者と健常対照者の血清、血漿のプロテオーム解析を行い、新しい診断マーカーを探索している。本稿では大規模な血液プロテオーム解析において必要な血液検体の要件、血液検体の調製法、質量分析方法、得られたデータの処理などについて、その意義とともに紹介する。

1. 試料調製のストラテジー

i) 試料調製

研究を始める前に、①十分な検体量や統計学的な解析が可能な症例数があり、②対照となる健常者や良性疾患患者の検体が入手可能であり、③施設の倫理審査を受け、④提供者の同意を得る方法に倫理的問題がないこと、⑤個人情報厳密に管理されていることを充分確認することは言うまでもない。検体の処理開始までの取り扱い、可能なかぎり一定化することにより変動因子を極力取り除く必要がある。特に採血方法 (採血管のメーカー・種類)、保存条件 (温

度・期間)、凍結融解の回数などは比較する検体群間で一致していることが必須である。さらには採血前の治療の有無、比較する検体群間での年齢や性別の分布などもよく検討しておかなければ、研究そのものが無駄になってしまう可能性がある。

血清、血漿には数千種類以上のタンパク質が含まれていると考えられており、Tirumalaiらの報告によると血漿はその成分の約90%をアルブミンをはじめとする10種類の構成タンパク質で占め、さらに全体の99%までを22種類のタンパク質で構成されている¹⁾。バイオマーカー候補となりうるタンパク質群は残りのわずか1%に含まれていると考えられている。さらに、タンパク質には少なくとも 10^3 以上の発現量格差があるといわれており、試料からのアルブミン、イムノグロブリンなどの含有量の多いタンパク質の分離除去が重要である。検体の50%以上を占めるアルブミンは固定化 cibacron blue (シバクロンブルー) によってそのリガンドに静電的および、もしくは疎水的に結合させることにより除くことができる (図1, SDS-PAGE lane2)。より特異的に構成タンパク質の除去を行うには、これらを特異的に認識する抗体アフィニティーなどが有効である。最近では同時に6種類 (図1, SDS-PAGE lane3) あるいは12種類の高発現タン

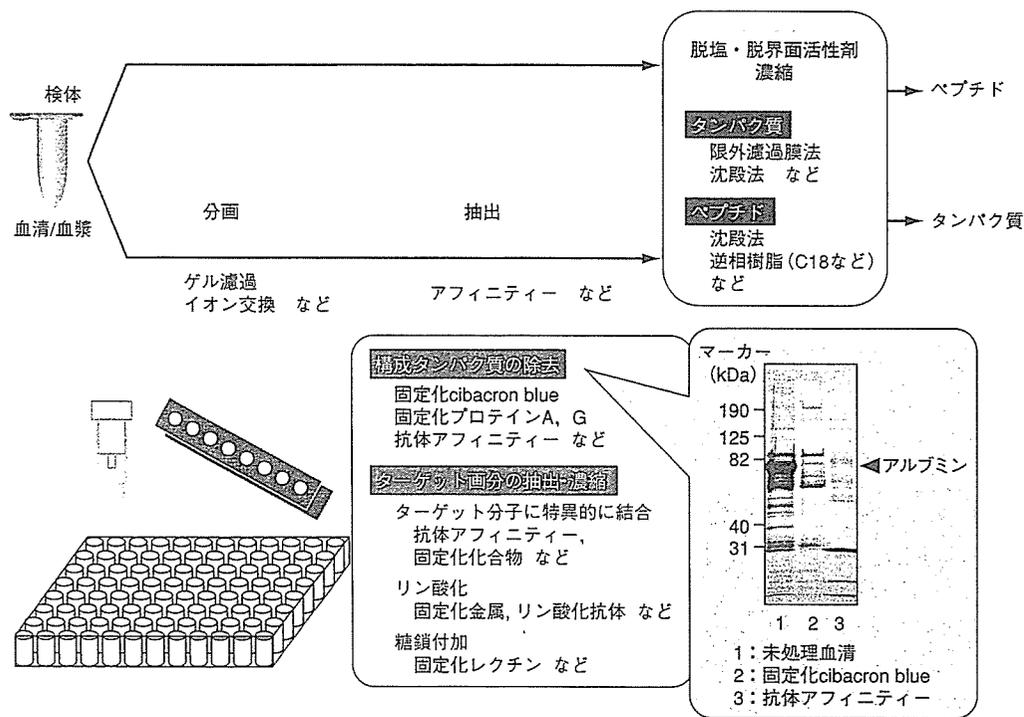


図1 ● 体液試料調製のストラテジー

検体は必要に応じて尿素などで変性後、ゲル濾過、イオン交換などにより分子量、等電点の違いにより分画を行う。さらに、親和性の高い抗体や低分子化合物によるアフィニティーなどにより構成タンパク質の除去 (SDS-PAGE: lane 1: 未処理, lane 2: 固定化 cibacron blue, lane 3: 抗体 (6種) アフィニティー) やターゲット画分の抽出・濃縮により特定タンパク質 (群) の抽出を行う。質量解析に供する前には限外濾過膜法、沈殿法などで可能な限り塩および界面活性剤は除く。最近では、分画・抽出などのタンパク質精製ツールがキット化され、スピナラムやマルチウェルプレートなど微量検体を安定的に処理できる工夫が施されたデバイスとして市販されている

タンパク質を除くことができるカラムも市販されている。

検体は必要に応じて尿素などで変性後、ゲル濾過、イオン交換などにより分子量、等電点の違いによる分画を行い、低発現量のタンパク質の検出を可能にする。さらに、特定のタンパク質 (群) を積極的に捕捉・濃縮するには、親和性の高い抗体や低分子化合物によるアフィニティーが有効である。アルブミン以外の多くの血清・血漿タンパク質は糖鎖付加による翻訳後修飾を受けていることから、特異的な糖残基と可逆的に結合する各種レクチンタンパク質を用いて濃縮・精製ができる。さらに特異性の異なる種々のレクチンを組み合わせて使うことで網羅性を上げることが可能である。

金属チップ表面に化学官能基や固定化された分子を装着し、試料中から特定の性質をもつ分子をチップ上で捕捉・精製するプロテインチップも同様の目的に使用されるものであり、質量分析と組み合わせた実験系は表面エンハンス型レーザー脱離イオン化 (surface-enhanced laser desorption/ionization time-of-flight mass spectrometry: SELDI-TOF-MS) 法と呼ばれている。

Petricoinらは診断に有用なバイオマーカー検索には、血流中に存在する癌細胞が分泌するタンパク質のみならず、癌の発症および病態の進行に伴う病変局所のプロテアーゼ活性の亢進によって産生されるペプチドの解析も重要であることを指摘している²⁾。流血中のペプチドはアルブミンなどのキャリアータンパク質と結合し、腎

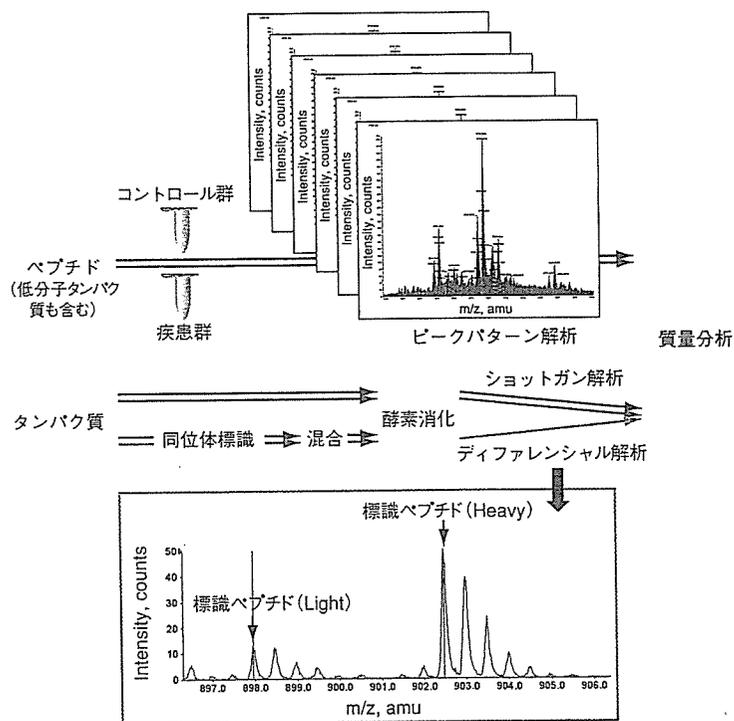


図2 ●体液試料の質量分析

質量分析による定量解析には、SELDI-TOF-MSやMALDI (matrix-assisted laser desorption ionization)-MSのように低分子タンパク質を含むペプチドをそのまま解析する方法に加えて、タンパク質をトリプシンなどの基質特異性の高い酵素で消化し、ペプチドとして物理化学的な性状を均一にして解析するショットガン法がある。ショットガン法ではICAT (isotope-coded affinity tag) などの *in vitro* での安定同位体タンパク質標識法を応用することで、ディファレンシャル解析が可能である

ionization)-MSのようにタンパク質・ペプチドをそのまま解析する方法に加えて、先にトリプシンなどの基質特異性の高い酵素で消化し、ペプチドとして物理化学的な性状を均一にして解析するショットガン法がある (図2)。ショットガンはICAT (isotope-coded affinity tag), iTRAQ などの *in vitro* での安定同位体タンパク質標識法を応用することで、比較定量が可能である。ICAT法では比較したい2検体を、分子量の異なる ^{13}C (Heavy) もしくは ^{12}C (Light) の同位体を含有する試薬で標識し、酵素消化後、標識部に導入されたビオチンタグで精製することで検体の複雑性を下げ、図2に示すようにペプチドの質量スペクトルから、由来するタンパク質の発現量を比較することができる。

われわれは無標識の検体を低流速 (50 ~ 200 nL/min) でかつ長時間流速が一定の液体クロマトグラフィーで分離し、質量分析器で定量し、ペプチドをその質量/価数 (m/z) と量を二 (ないし三) 次元画像として表す2D-ICAL-MS (two dimensional Image converted analysis of nano-flow liquid chromatography and mass spectrometry) 法を開発した (日本ヒトプロテオーム機構第3回大会要旨集)。

のクリアランスを回避していると考えられ、アルブミン結合ペプチドの検索もバイオマーカー候補を捕捉・濃縮する方法として有用であると考えられる。

ii) 血液試料の質量分析

サンプルを質量解析に供する前には可能なかぎり塩および界面活性剤は除く必要がある (図1)。最近では、分画・抽出などのタンパク質精製ツールがキット化され、スピンカラムやマルチウェルプレートなど微量検体を安定的に処理できる工夫が施されたデバイスとして市販されており便利である³⁾。

質量分析による定量解析には、SELDI-TOF-MSやMALDI (matrix-assisted laser desorption

2. マーカー分子の絞り込みやバリデーション

プロテオーム解析では莫大な量の情報が発生するため、生物学的、臨床的意味のある情報を抽出するためには、バイオインフォマティクスの手段を用いる必要がある。われわれの研究室で行っている機械学習法の手順を示す (図3)。年齢、性別、採血法、採血時期に偏りがないような疾患群とコントロール群のサンプルを学習セットとして用意する。このセットの質量解析を行い、得られたスペクトルからピーク情報を定量的に検出する。単一のピークのみから最終判別をすることは判別率が十分でないことが多いため、人工知能などに応用される機械学習アルゴリズム [サポートベクタマシン (線形, 非

線形)、ニューラルネットワーク、ファジーニューラルネットワーク]を用いて、癌と正常を区別するマルチマーカーからなるデータセット(分類器)を抽出する。

しかし、この段階では数学的に強制的に疾患群とコントロール群を分類しただけに過ぎず、生物学的な意味のない分類をしただけの可能性もある。そこで、学習セットにおける分類結果が正しいかどうかを、検証セット(臨床情報が未知の全く別の検体)を用いて評価する。検証セットを盲検してはじめて、学習セットの分類器が診断的意義のあるものであると確認できる。このような流れでマーカー分子の絞り込みやバリデーションを行うことにより、その後に臨床情報未知検体サンプルが持ち込まれたとしても高い判別率で癌か否かを判定することが可能である。

3. 応用例および課題

われわれの研究室では SELDI-TOF MS 法を用いたアプローチで微量の検体を再現性よくハイスループットに解析することによって、有効な癌のマーカー候補があがってきているので紹介する。腎腫瘍の9割を占める腎細胞癌の早期診断マーカーを検索することを目的とした検討⁴⁾では、解析に供した血清では構成タンパクの除去は行わず、陰イオン交換により分画した。さらに各分画を疎水性、金属親和性そしてイオン交換の各プロテインチップ上で特定の性質をもつ分子の捕捉、精製を行い、サイファージェン社の飛行時間型質量分析計である PBS II c を用いて分析を行った。検出ピークは未分画では数百であったものが、分画操作により約3,000に増加していた。これらのピークを図3に示す機械学習法により解析を行ったところ、2つのピークを組み合わせることにより、感度、特異度共に良好に腎細胞癌を判別できるマーカーピークを得ることができた。さらにマーカー検索の再現性、安定性を向上させるためには、質量分析の精度を向上させることが必要であると考えられ

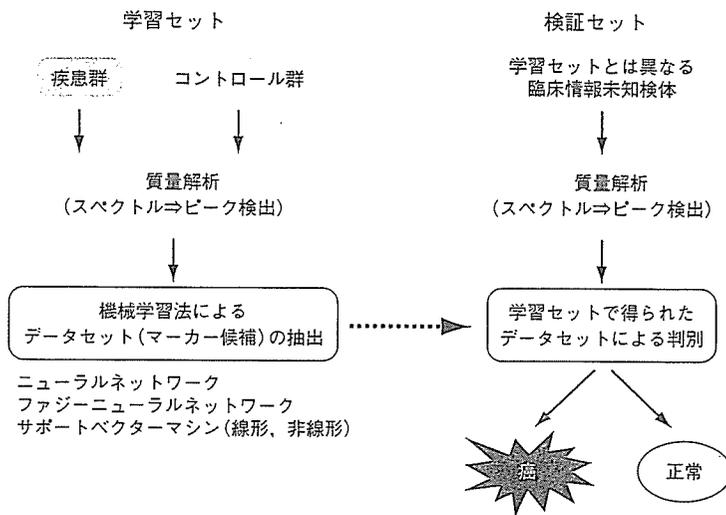


図3 ●マーカー分子の絞り込みやバリデーション

年齢、性別、採血法、採血法時期に偏りがないような疾患群とコントロール群のサンプルを学習セットとして用意する。このセットの質量解析を行い、得られたスペクトルからピーク情報を定量的に検出する。次に、機械学習アルゴリズム(サポートベクターマシン(線形、非線形)、ニューラルネットワーク、ファジーニューラルネットワーク)を用いて、癌と正常を区別するマルチマーカーからなるデータセット(分類器)を抽出する。そこで、学習セットにおける分類結果が正しいかどうかを、検証セット(臨床情報が未知の全く別の検体)を用いて評価する

たので、高分解能型四重極飛行時間型(quadropole time-of flight: QTOF)質量分析計を用い、膵臓癌の血漿腫瘍マーカーの探索を行った⁵⁾。血漿サンプルを未分画でプロテインチップに捕捉・濃縮し、質量分析を行い、得られたピークを機械学習法により解析を行ったところ、感度、特異度ともに高い値を有する4つのピークを検出した。このマーカーピークは既存の腫瘍マーカーである CA19-9 と相補的で、両者の組み合わせにより膵臓癌患者の100%の検出が可能であった。膵臓癌は早期診断が非常に困難であるために、この血漿腫瘍マーカー開発は多くの期待が寄せられており、実用化を進めている。

バイオマーカー検索アプローチは、癌の診断だけでなく、個別化治療選択のための診断法開発にも用いられる。われわれは進行食道癌の術前化学放射線療法(5-フルオロウラシル+シスプラチン+放射線照射)の奏効性を左右しうるマーカー分子の検索を行った⁶⁾。化学放射線療法前

に採取した血清を各種プロテインチップ上での捕捉、精製後、高分解能質量分析計にて分析を行った。得られたピークの解析結果から化学放射線療法の奏効性を予測することが可能な4つのピークを抽出することに成功している。

いたことがある。しかし最近では試料調製の重要性がさまざまな場面で認識されはじめてきている。今後、試料調製技術のさらなる進展により臨床に役立つバイオマーカー検索に役立てたいと考えている。

おわりに

質量解析によるプロテオーム解析は、機械およびアプリケーションの技術開発が先行して、実体は何年か遅れてついてきている現状がある。その要因として、生体内で発現しているタンパク質を発現量格差がある状態にもかかわらず高感度な質量分析計で網羅的という言葉とともに一網打尽に分析できるかのような幻想を抱いて

参考文献

- 1) Tirumalai, R. S. et al : Mol. Cell Proteomics, 2 : 1096-1103, 2003
- 2) Petricoin, E. F. & Liotta, L. A. : Curr. Opin. Biotechnol., 15 : 24-30, 2004
- 3) Smith, C. et al : Nature Methods, 2 : 71-77, 2005
- 4) Hara, T. et al : J. Urol, 174 : 1213-1217, 2005
- 5) Honda, K. et al : Cancer Res, 65 : 10613-10622, 2005
- 6) Hayashida, Y. et al : Clin. Cancer Res., 11 : 8042-8047, 2005



下重美紀 (Miki Shitashige)

国立がんセンター研究所 化学療法部 研究員。

1998年東京医科歯科大学大学院博士課程を修了後、タンパク質精製、分子細胞生物学を中心に研究に従事し、2002年より質量分析技術に取り組んでいる。2004年8月より現職。癌の早期診断・病態診断に有用な腫瘍マーカーの開発を見出すことを目的とした研究を行っている。

本田一文 (Kazufumi Honda)

国立がんセンター研究所化学療法部 室長。

山田哲司 (Tesshi Yamada)

国立がんセンター研究所 化学療法部 部長、腫瘍プロテオミクス プロジェクトリーダー併任。

実験医学別冊 注目バイオ実験シリーズ

羊土社

決定版! プロテオーム 解析マニュアル

絶賛発売中

発現解析・機能解析の最新プロトコルから
データ整理、トラブル対処法まで

特典 データ整理に必ず
役立つソフトウェア

編集 / 磯辺俊明 (東京理科大学大学院理学研究科化学専攻生物化学)
高橋信弘 (東京農工大学農学部応用生物科学科構造生化学)



B5判 281頁 定価(本体6,200円+税) ISBN4-89706-415-5