

厚生労働科学補助金（第3次対がん総合戦略研究事業）分担研究報告  
がん検診に有用な新しい腫瘍マーカーの開発

「新規プロテオーム解析技術 2DICAL 法による血中腫瘍マーカーの開発」

氏名	所属	職名
分担研究者 尾野雅哉	国立がんセンター研究所化学療法部	室長

研究要旨

われわれが開発した 2DICAL (2 Dimensional Image Converted Analysis of Liquid Chromatography and Mass Spectrometry(LC/MS))法はハイスループットなショットガンプロテオミクス解析手法である。この手法を用い新規の血中腫瘍マーカーを開発することを目的として、膵がん血漿腫瘍マーカー、及び、子宮体がん血清腫瘍マーカーを開発することに成功した。

A. 研究目的

膵がんは主要な固形がんの中では最も予後が不良で、その5年生存率は6%と低率である（地域がん登録共同調査）。わが国で1年間に22260人の方が膵がんのために死亡し、がんによる死亡原因の第5位を占めている（平成16年度厚生労働省「人口動態統計」）。膵がんが難治性である理由は、早期から周囲組織に浸潤性増殖をきたし、肝臓やリンパ節などに転移巣を形成することにより外科的な根治的切除術施行が困難なだけでなく、臨床症状が乏しいため先に述べた根治切除術を施行可能な患者を有効に検出する方法が存在しないことなどが理由にあげられる。事実、全国膵がん登録20年間の集計によると、膵がん症例の約95%が発見時、臨床病期 III または IV の進行がんであると報告されている。このような膵がんの生命予後を改善

させるためには、切除可能な膵がんを有効に検出できる診断法が開発が望まれており、もしこのような診断法が開発されれば、予後成績は飛躍的に向上することが期待される。

現在、膵がんで臨床的に使用される腫瘍マーカーはCEA、CA19-9、エラスターゼIがあげられるが、早期膵がんに対する検出感度は十分ではなく、腫瘍の最大径別でみた陽性率はそれぞれT1で0%、0%、30%、T2で0%、25%、8%でしかない。また、胆石や膵炎のような良性疾患にも高値をしめすことが多く、膵がんを有効的に検出するマーカーとしては、満足できるものではない。

また、子宮体がんは年々増加傾向にあり、年間約18000人が発病しているといわれている。子宮がんに占める割合は1970年代では10%程度であったが2000年で約40%と著明に増加して

いる。好発年齢は50～60歳代で、危険因子として未婚・不妊・出産児数の減少、高血圧・肥満・高血糖、乳癌術後Tamoxifen内服などが挙げられ、生活習慣の変化とともに、今後更に増加すると考えられている。

しかし、現在の子宮体がん検診法としては、子宮内腔に細いスティック状の器具を挿入して細胞を採取する子宮内膜細胞診が主に行われているが、進行初期では診断にいたらないことも多く、また本検査は閉経後の被検者では子宮口が閉鎖して検査自体が行えない場合があること、検査時の強い痛み、出血、感染、子宮穿孔などの合併症の可能性があり幾多の問題点を有している。別の診断法としての経膈超音波法は子宮内膜の厚さ、形状をみる侵襲度の低い検査ではあるが、早期のがん診断における感度は10～30%程度とされており、検診における有用性は十分とはいえない。子宮体がんを用いられている腫瘍マーカーは、CA125、CA602、CA19-9などであるが、いずれもI期の早期がんでの陽性率は20%前後であり、治療効果の判定と治療後の再発指標として用いられているのが現状である。このような現況において、子宮体がんも膵がん同様に、非侵襲的で有効な診断法の開発が強く望まれている。

近年、プロテオームの研究手法が急速な進歩をとげ、血液中に含まれる微量なタンパク質やペプチドが網羅的に解析できるようになってきた。本研究班において本田らは、トップダウン

プロテオミクス解析手法として、SELDI-QqTOF-MS法を開発し、膵がん腫瘍マーカーの開発をすすめてきたが、われわれはショットガンプロテオミクス解析手法として2DICAL (2 Dimensional Image Converted Analysis of Liquid Chromatography and Mass Spectrometry (LC/MS))法を新規に開発し、LC/MSから得られる膨大なデータを効率よくハイスループットに解析することを可能とした(Ono, M. et al. 2006)。われわれはこの新規技術を応用し、前述の如く開発が望まれている、膵がん、子宮体がんの血中腫瘍マーカー開発を目的として研究を行った。

## B. 研究方法

### 1) 2DICAL法によるプロテオーム解析

われわれが開発した2DICAL法を用いて、腫瘍マーカー探索を行った。2DICAL法は、複数のスペクトラムからなるLC/MSデータを、各スペクトラムの相関係数からLCの時間変動を補正して、質量電荷比( $m/z$ )、保持時間(RT)の2軸を持つ平面に描出する手法であり、これにより同一ペプチド由来のピークが、強度(Intensity)を変数に持つ $m/z$ 、RT座標に変換され、複数サンプル間での無標識定量比較を可能にしたショットガンプロテオミクス解析手法である。さらに、サンプル(S)をもうひとつの次元にとらえ、同一 $m/z$ のピークをRT、サンプル

の2軸の平面で描出することにより、多数検体間の定量比較を可能としている。

この技術の開発により、LC/MSで作成される膨大なピークデータから効率よくかつ定量的に多数検体解析が可能となり、ショットガンプロテオミクスにおける血中腫瘍マーカー開発が可能となった。

## 2) 膵がん血漿腫瘍マーカーの開発

倫理要件を満たす膵がん患者血漿および健常者血漿を国立がんセンター中央病院にてそれぞれ38症例、39症例、東京医科大学病院にてそれぞれ5症例、4症例集積し、計86症例を解析対象とした。アルブミンなどの血中に存在する大量のたんぱく質の影響を除外するために、特異糖鎖を認識するコンカナバリンA (ConA) にて血漿を前処理し、ConAに吸着する分画を抽出した。同分画にトリプシン3.3 $\mu$ g 5M尿素10 $\mu$ l 1M炭酸水素アンモニウム2.5 $\mu$ lを加え、MilliQで計30 $\mu$ lに調整した後、20時間摂氏37度でたんぱく質を消化させた。消化産物にアセトニトリルを加え遠沈し、上清を抽出、真空乾燥させた後、0.1%ギ酸を80 $\mu$ l加え、質量分析測定用試料とした。一回の質量分析には10 $\mu$ lを用い、同一試料に対して3回の測定を行った。残りの試料は同定用試料として摂氏-80度で保存した。

LC/MS測定に当たっては、ナノLC(超低流量液体クロマトグラフィー)を用い、C18逆相カラム(内径0.15mm、

長さ50mm)で、毎分200nlの超低流量の流速でアセトニトリル濃度を0%から80%まで60分かけて直線勾配で上昇させ、測定試料を分離した。MS(質量分析計)はQTOF-ULTIMA(Waters, Milford, MA)を用い、質量測定範囲を250-1600 $m/z$ とし、1秒毎の積算でスペクトラムデータを60分間採取した。採取したLC/MSデータは、独自に開発した方法で、2DICAL用の解析データに変換した。

2DICALでは、各測定における液体クロマトグラフィーの時間変動を補正し、同一ペプチド由来のピークを検出し、(質量電荷比( $m/z$ ), 保持時間(RT))の座標を与え、症例毎に3回測定したピーク強度の中央値を各症例のピーク強度代表値とした。国立がんセンター中央病院症例において膵がん患者群と健常者群間で有意差のあるピークをU検定で $p < 0.0005$ 、ピーク強度 $\geq 10$ 、強度比 $\geq 2$ で拾い出し、目視にてピークの有効性を確認したものをマーカー候補とした。東京医科大学病院症例にてマーカー候補の検証を行い、同時にペプチド配列を決定した。

これらのピークが示すペプチド配列を用い、特異抗体を産生し、2DICALで解析した血漿でその発現を確認した。さらにこの抗体を用いELISAキットを作製し、本研究班で収集した血液の発現を検討した。

## 3) 子宮体がん血清腫瘍マーカーの開発

倫理要件を満たす子宮体癌 40 症例、対照者 30 症例、子宮筋腫 30 症例、子宮肉腫 25 症例の血清を北海道がんセンターで集積し、計 125 例を解析対象とした。アルブミン、IgG などの血中に大量に存在する 12 種のタンパク質を除去するカラム (ProteomeLab IgY-12 SC, Beckman Coulter Inc., Fullerton, CA) で前処理し、前記と同様のトリプシン消化処理を施し、質量分析測定用試料を作成した。

前記と同様に 2DICAL で拾い出したピークから、子宮体がん患者群と対照者群間で、 $t$  検定で  $p < 0.01$  のピークを拾い出し、それらのピークから強度差の大きいピークを抽出するために、IQR (四分位範囲) 20 以上のピークを選別した。これらのピークに対し、ペプチド、蛋白質同定を行い、それら同定蛋白質の発現差をウェスタンブロットにて確認した。

#### (倫理面への配慮)

国立がんセンター、東京医科大学病院、北海道がんセンターの倫理委員会による審査で承認された方法で採取保管され、検体の個人情報が出ることが無いように匿名化が厳重に行われるように配慮したがん患者、健常人より得られた血漿血清検体を用いた。

### C. 研究結果

#### 1) 膵がん血漿腫瘍マーカー

2DICAL 法で検出したピークは 115,325 ピークあり、そのピークから

上記方法で絞り込んだマーカー候補は 6 ピークであった。それぞれピークの (質量電荷比 ( $m/z$ ), 保持時間 (RT)) は、(412  $m/z$ , RT 13.7min)、(546  $m/z$ , RT 8.3min)、(552  $m/z$ , RT 8.3min)、(827  $m/z$ , RT 8.3min) (1141  $m/z$ , RT 29.0min) (1185  $m/z$ , RT 9.2min)、であった。これらのピークは東京医科大学病院症例にても差が認められ、検証された。

これらのピークのなかでペプチド配列が決定されたものに、水酸化プロリンを含む  $\alpha$ -フィブリノゲンのペプチド配列が認められたため、この配列に対する特異抗体を作成した。この抗体は、 $\alpha$ -フィブリノゲンがタンパク質の状態でも反応したため、2DICAL で解析した血漿でその発現を解析したところ、膵がんで水酸化プロリンを含む  $\alpha$ -フィブリノゲンの発現が上昇していることが確認された。

さらに、本研究班で収集した血液での発現を解析したところ健常者に比べ膵がんで有意に上昇していることが確認された。年齢性別をあわせ、各種疾患での発現を調べたところ、膵がんでより上昇が認められた。

#### 2) 子宮体がん血漿腫瘍マーカー

2DICAL 法で検出したピークは 154,992 ピークであり、そのピークから上記方法で絞り込んだマーカー候補は 100 ピークであった。これら 100 ピークに対して、ペプチド、蛋白質同定を行ったところ、53 ピークがペプチドとして同定され、それらは 14 のた

たんぱく質を構成していることが明らかとなった。

同定されたたんぱく質には、Apolipoprotein A-IV、Complement component C4A、Complement component C3が含まれており、これらの血中での発現差をウェスタンブロットで確認した。これらの知見をまとめ論文発表した。

#### D. 考察

ショットガンプロテオミクスにおいては、たんぱく質がトリブシン処理された膨大なペプチドデータとなるため、同位体標識等することにより、小数のサンプル比較を精密に行うことが解析の主体であった。事実、培養細胞などで洗練された比較系を構築することで、新規のシグナル伝達物質やたんぱく質相互作用を見出している。しかし、多数検体の処理を必要とする臨床検体の解析はショットガンプロテオミクスにおいては不得手の領域であった。われわれが開発した2DICAL法はいくつかの斬新的な手法を構築することで、ショットガンプロテオミクスにおける多数検体処理を可能にし、膵がんや子宮体がんの血中腫瘍マーカー候補の発見を可能とした。

今回発見された膵がんにおける腫瘍マーカーは、肝臓で主に産生され、血中に豊富に認められるたんぱく質の2つのペプチド配列であった。腫瘍マーカー候補となったのは、このペプチド配列の中のひとつのアミノ酸が翻

訳後修飾を受けたものであり、膵がん患者と健常者の間で大きな差が認められた。しかし、修飾のない同ペプチド配列も同時に血中に存在し、それは膵がん患者と健常者の間で差が認められなかった。このような変化は、たんぱく質レベルでの検討では十分な差が見出せない可能性が高く、ペプチドレベルまで分解するショットガンプロテオミクスの中で初めて見つかる変化であると考えられる。このように血中の豊富なたんぱく質でもペプチドまで分解して解析すれば新しい腫瘍マーカーとなりうる可能性があり、2DICAL法による解析は、そのような腫瘍マーカーを発見する上で、大きな強み(advantage)を持っているシステムであるといえる。

血中腫瘍マーカーの開発に当たっては、アルブミンなどの血中に存在する大量のたんぱく質の影響を除外する前処理が重要である。前処理としては、必要なものを集めてくる方法(濃縮)と、不要な物質を除く方法(除去)があるが、われわれは膵がん血漿腫瘍マーカー探索に当たっては、糖ペプチドを濃縮するConAを用い、子宮体がん血清腫瘍マーカー探索に当たっては12種の主要たんぱく質を除去するカラムを用いて検討した。ConAを用いることにより、マンノースを糖鎖に持つ糖たんぱく質が濃縮され、がん患者と健常人間で差のあるたんぱく質の修飾変化を見出すことが出来た。豊富な上位12種のたんぱく質を取り除くことにより、がんで量的変化が起こ

ることが知られているたんぱく質の定量的変化を 2DICAL 法で捉えられることが確認され、さらに新規の腫瘍マーカー候補を見出せる可能性が示された。

発見されたマーカーを検証するためには、抗体を用いることが簡便である。今回のマーカーは新規のもので、新しい抗体を作成しなければならなかったが、いったん作製できると、ELISA キットなどの既存の検査法が構築可能となり、多数検体で新規のマーカーの検証が容易に行えた。2DICAL 法は候補マーカーの構造を決定することが容易に行えるシステムであり、このような抗体による検査法を開発していく上にも有用な手法であることが示された。

#### E. 結論

- 1、 2DICAL 法を用い、膵がん血漿腫瘍マーカーとなる可能性のある水酸化プロリンを含む $\alpha$ -フィブリノゲンを同定することができた。
- 2、 血中に豊富に存在するたんぱく質の修飾変化が、膵がん血漿腫瘍マーカーとなる可能性が示唆された。
- 3、 2DICAL 法は定量性に優れ、がんにおける血中たんぱく質の量的変化を正確に反映するだけでなく、翻訳後修飾による新規の血中腫瘍マーカーを発見できる可能性が示唆された。

#### F. 健康危険情報

特になし

#### G. 研究発表

(2008年4月1日から2009年3月31日まで)

##### 1、 論文発表

- 1) Matsubara J, Ono M, Honda K, Yamada T et.al. Identification of a Predictive Biomarker for Hematological Toxicities of Gemcitabine. *Journal of Clinical Oncology*. (in press)
- 2) Negishi A, Ono M, Honda K, Yamada T et.al. Large-scale quantitative clinical proteomics by label-free liquid chromatography and mass spectrometry. *Cancer Science*. 100(3):514-9. 2009.
- 3) Yamaguchi U, Honda K, Ono M, Yamada T et.al. Distinct gene expression-defined classes of gastrointestinal stromal tumor. *Journal of Clinical Oncology*. 26(25):4100-8. 2008.
- 4) Shitashige M, Honda K, Ono M, Yamada T et.al. Regulation of Wnt signaling by the nuclear pore complex. *Gastroenterology*.

134(7):1961-71. 2008.

- 5) Kikuchi S, Honda K, Ono M, Yamada T et.al. Expression and gene amplification of actinin-4 in invasive ductal carcinoma of the pancreas. Clin Cancer Res. 14(17):5348-56. 2008.
- 6) 尾野雅哉 LCMS 解析手法 2 DICAL を用いた糖鎖疾病マーカーの探索 遺伝子医学 MOOK11 号 66-72. 2008
- 7) 尾野雅哉、本田一文、山田哲司 他 がん診断治療のバイオマーカー Cancer Frontier Vol.10 14-20. 2008
- 8) 尾野雅哉、本田一文、山田哲司 他 癌の診断における最近の proteomics 研究について教えて下さい サージェリーフロンティア (印刷中)
- 2、 学会発表
- 1) Matsubara J, Ono M, Yamada T et.al. Matsubara J, Ono M, Honda K, Yamada T et.al. Identification of a Predictive Biomarker for Hematological Toxicities of Gemcitabine. ASCO -GI 2009, San Francisco USA 2009.
- 2) 尾野雅哉、根岸 綾子、本田 一文、山田哲司、他 2 DICAL による修飾タンパク質バイオマーカーの発見、第 4 回日本臨床プロテオーム研究会、シンポジウム、2008 年 5 月、大阪
- 3) 根岸 綾子、尾野雅哉、本田 一文、山田哲司、他 ホルマリン固定パラフィン包埋組織を用いた舌扁平上皮癌のバイオマーカー探索、第 4 回日本臨床プロテオーム研究会、口演、2008 年 5 月、大阪
- 4) 松原淳一、尾野雅哉、本田 一文、山田哲司、他 切除不能膀胱がんに対するゲムシタピン単剤療法の副作用・効果・予後の各予測マーカーの開発、第 6 回日本ヒトプロテオーム機構大会、シンポジウム、2008 年 7 月、大阪
- 5) 根岸 綾子、尾野雅哉、本田 一文、山田哲司、他 ホルマリン固定パラフィン包埋組織を用いた舌扁平上皮癌のバイオマーカー探索、第 67 回日本癌学会学術総会、口演、2008 年 10 月、名古屋
- 6) 松原淳一、尾野雅哉、本田 一文、山田哲司、他 切除不能膀胱がんに対するゲムシタピン単剤療法の血液毒性予測マーカーの開発、第 67 回日本癌学会学術総会、口演、

2008年10月、名古屋

7) 村越雄介、尾野雅哉、本田一文、山田哲司、他 大腸癌の大規模血漿プロテオミクス、第67回日本癌学会学術総会、ポスター、2008年10月、名古屋

8) 尾野雅哉、がんのプロテオーム(2)、第67回日本癌学会学術総会、座長、2008年10月、名古屋

H, 知的財産権の出願・登録状況(予定を含む)

1、特許取得

「酸化修飾タンパク質またはポリペプチドに対する高特異性モノクローナル抗体」として平成20年10月14日株式会社トランスジェニックより特許出願(特願2008-330745)した。

2、実用新案登録

なし

3、その他

なし

がん検診に有用な新しい腫瘍マーカーの開発

近藤格

プロテオーム バイオインフォマティクス プロジェクト  
プロジェクトリーダー

研究要旨

【研究1：特異抗体を用いた炎症性サイトカインの網羅的解析】がんの病態の背景には慢性炎症が存在する。慢性炎症に関連して変動する個々のサイトカインの濃度の組み合わせを診断マーカーとして使用できる可能性を検討する目的で、肺癌患者、胃癌患者、大腸癌患者、健常者の血清をそれぞれ25症例分ずつ（合計100症例分）調べた。特異抗体を用いて50種類のサイトカインの発現を健常者と悪性腫瘍患者の血清間で比較し、統計的な有意差を示すサイトカインとして、肺癌で25種類、胃癌で25種類、大腸癌で24種類を同定した。肺癌患者、胃癌患者、大腸癌患者に共通していたサイトカインは11種類あり、それぞれの悪性腫瘍に特異的なサイトカインは肺癌で5種類、胃がんで4種類、大腸癌で6種類だった。悪性腫瘍患者に特有の炎症性サイトカインの発現パターンを同定し、もって早期診断のためのバイオマーカーとして使用することができる可能性が示唆された。【研究2：血清腫瘍マーカーの候補となる腫瘍組織に特異的タンパク質の解析】腫瘍細胞が特異的に産生するタンパク質を網羅的に調べることで、血清腫瘍マーカーとしての有用性のあるタンパク質を同定することを試みた。食道がん、肝細胞癌の手術検体を用いたプロテオーム解析の結果、既知の血清腫瘍マーカーに加え、構造的に血清腫瘍マーカーの候補になりえるタンパク質を多数同定した。

研究1：特異抗体を用いた炎症性サイトカインの網羅的解析

A. 研究目的

悪性腫瘍に関連する炎症性サイトカインの発現を網羅的に探索し、がん検診に有用なサイトカインを同定すること。

B. 研究方法

B-1. 血清サンプル

肺癌患者、胃癌患者、大腸癌患者、健常者、それぞれ25名ずつより採取した血清サンプル（合計100サンプル）を使用した。

B-2. BioPlexサスペンションアレイシステム

測定にはバイオラド社のBioPlexサスペンションアレイシステムを使用した。本システムでは特異抗体を固相化したビーズを用いて血清中に含まれる微量なタンパク質の同時測定を行うことができる。メーカー推奨の方法にしたがって合計50種類のサイトカインを調べた。

B-3. 数値解析

それぞれのサイトカインについて、健常者と悪性腫瘍患者の血清間で統計的な有意差を示すものを調べた。Z検定のp値に対して、BH法にてFalse Discovery Rate (FDR)を推定した。

C. 研究結果

C-1. 炎症性サイトカインの発現レベル

ほとんどのサイトカインについて有効レンジに測定値がきており、BioPlexサスペンションアレイシステムが実験系として有効に使用できると結論した。

C-2. 発現解析

健常者と悪性腫瘍患者の血清間で統計的な有意差（Z検定のp値に対して、BH法にてFalse Discovery Rateが5%未満）を示したサイトカインとして、肺癌で25種類、胃癌で25種類、大腸癌で24種類を特定した。3種類の悪性腫瘍に共通していたものは11種類あり、それぞれの悪性腫瘍に特異的なものは肺癌で5種類、胃がんで4種類、大腸癌で6種類だった。

D. 考察

がんの病態の背景には慢性炎症が存在する。過去の血清プロテオーム解析では炎症性タンパク質が血清腫瘍マーカーとしてたびたび報告されている。しかし、炎症性タンパク質は個別には診断能に乏しいことが判明している。本研究では炎症性のサイトカインを一度に網羅的に調べ、個々の炎症性サイトカインの組み合わせによって診断精度の高い血清腫瘍マーカーを開発することを試みた。肺癌で5種類、胃がんで4種類、大腸癌で6種類の腫瘍特異的なサイトカインを同定した。

E. 結論

本研究において、悪性腫瘍症例の血清中に存在する炎症性サイトカインを検診に使用できる可能性が

あることが示唆された。しかしながら、用いた検体の数に比べて発現差を示す炎症性サイトカインの数が多く、組み合わせによって高い診断精度が得られるかどうかは今回の実験からは分からない。同定されたサイトカインについて、今後は多数の症例において再現性のある結果が得られるかどうかを検証する。

## 研究2：血清腫瘍マーカーの候補となる腫瘍組織に特異的タンパク質の解析

### A. 研究目的

腫瘍組織で発現亢進しているタンパク質を網羅的に調べることで、血清腫瘍マーカー候補を同定すること。

### B. 研究方法

#### B-1. 組織サンプル

国立がんセンター中央病院において外科切除を受けた食道がんおよび肝細胞癌の腫瘍組織を研究対象とした。食道癌の正常組織は食道腫瘍組織より離れた正常食道粘膜上皮組織を使用した。また、肝細胞癌については、転移性肝癌の症例から得られる正常の肝臓組織、および肝細胞癌症例の非腫瘍組織を使用した。腫瘍組織は気相窒素のタンクに凍結保存されているものを使用した。

#### B-2. サンプルング

凍結検体をいったんOCTコンパウンドに包埋し、レーザーマイクロダイセクションで目的の細胞を回収した。次に、回収した細胞から高濃度のウレアを含むタンパク質可溶化液でタンパク質を抽出した。

#### B-3. 蛍光二次元電気泳動法

回収したタンパク質を超高感度の蛍光色素で標識し、内部標準となるサンプルと同時に大型の二次元電気泳動ゲルで分離した。電気泳動後のゲルをレーザーキャナーにかけ、タンパク質スポットを蛍光シグナルとして読み取った。ゲル間のばらつきを内部標準サンプルで補正してタンパク質発現情報を取得した。

#### B-4. 数値解析

Wilcoxon検定で $p$ 値0.05以下、群間平均値の差が4倍以上（食道がん）あるいは3倍以上（肝細胞癌）という基準でタンパク質スポットを選択した。

#### B-5. タンパク質同定実験

同定したタンパク質スポットを自動分取ロボットを用いてゲルから回収し、トリブシンにてペプチド化し、質量分析装置（LTQ、サーモフィニガン社）に

タンパク質を決定した。同定にはMasCotソフトウェアを使用した。

### C. 研究結果

正常組織と腫瘍組織の間で上記の基準で有意な発現差を示すタンパク質スポットとして、食道がん20種類、肝細胞癌で41種類を特定し、対応するタンパク質をすべて同定した。同定されたタンパク質の中には、DNA修復酵素、シグナル伝達関連タンパク質、転写因子、細胞周期制御因子などの細胞内タンパク質に加え、分泌型のタンパク質で悪性腫瘍の発生や悪性度に関係することがすでに報告されているタンパク質、肝臓線維化の既知の血清マーカー、などが含まれていた。また、DNA配列から予測される分子量よりもかなり小さい分子量をもつタンパク質も多数同定された。

### D. 考察

本研究では、既存の血清腫瘍マーカーの多くは腫瘍細胞が産生しているタンパク質であることに着目し、正常組織と腫瘍組織の間で著しく発現差を示すタンパク質を網羅的に同定することを試みた。今回同定されたタンパク質の中には既存の血清腫瘍マーカーも含まれていたが、このようなタンパク質は血清中では極少量しか存在していないため、血清サンプルを直接プロテオーム解析の技術で調べたのでは同定できなかったと考えられる。また、DNA配列から予測されるよりずいぶん小さい分子量で存在しているタンパク質については、切断されることで細胞外に出やすくなっている可能性があり、過去の報告では実際に血中に分泌される場合があることが判明している。本実験では、組織検体には血清腫瘍マーカー候補が含まれていること、既存のプロテオーム解析の技術はそのようなタンパク質を検出するために必要な感度を有していることが分かった。

### E. 結論

同定したタンパク質のうちどれが血中に放出されているかを推定することが今後の課題である。考えられる方法は三つある。一つは、既存のデータベースの活用である。アミノ酸配列の情報に基づく水溶性の検討、血清タンパク質データベースの情報に基づく血中での安定性などの情報に基づき、血中で存在する可能性を検討する。もう一つの方法は、細胞内での存在様式に着目する方法である。腫瘍細胞で発現が高くしかも切断されているタンパク質が、切断されることによって腫瘍組織から放出されて血中で検出されるようになっていたという報告がある。この報告にならって、予測される分子量と実測される分子量に著しい差異のあるタンパク質を探索するという手法が考えられる。この方法がもっとも現実的

な方法である。最後に、網羅的に抗体を用いて血清サンプルを調べるという方法がある。抗体のライブラリーが必要であり容易には実行できない。炎症細胞由来のタンパク質が同定されたことは、「研究1 特異抗体を用いた炎症性サイトカインの網羅的解析」の結果とも関連すると思われる。今回の実験データからは解釈が困難だったが、データを蓄積することで血中のサイトカインと組織中の炎症細胞の関連性を見出すことが今後可能になる可能性がある。

## F. 健康危険情報

特記事項なし

## G. 研究発表

### 1. 論文発表

1. Kikuta K, Tochigi N, Shimoda T, Yabe H, Morioka H, Toyama Y, Hosono A, Beppu Y, Kawai A, Hirohashi S, Kondo T, Nucleophosmin as a candidate prognostic biomarker of Ewing sarcoma revealed by proteomics. Clin Cancer Res, in press
2. Yamada M, Fujii K, Koyama K, Hirohashi S, Kondo T, The proteomic profile of pancreatic cancer cell lines corresponding to carcinogenesis and metastasis. J Proteomics Bioinform, 2, 1-18, 2009
3. Kondo T, Tissue proteomics for cancer biomarker development: laser microdissection and 2D-DIGE. BMB Rep, 41, 626-34, 2008
4. Kondo T, Cancer proteomics for biomarker development. J Proteomics Bioinform, 1, 477-84, 2008
5. Uemura N, Nakanishi Y, Kato H, Saito S, Nagino M, Hirohashi S, Kondo T, Transglutaminase 3 as a prognostic biomarker in esophageal cancer revealed by proteomics. Int J Cancer, 124, 2106-15, 2009
6. Saito S, Ojima H, Ichikawa H, Hirohashi S, Kondo T, Molecular background of alpha-fetoprotein in liver cancer cells as revealed by global RNA expression analysis. Cancer Sci, 99, 2402-9, 2008
7. Orimo T, Ojima H, Hiraoka N, Saito S, Kosuge T, Kakisaka T, Yokoo H, Nakanishi K, Kamiyama T, Todo S, Hirohashi S, Kondo T, Proteomic profiling reveals the prognostic value of adenomatous polyposis coli-end-binding protein 1 in hepatocellular carcinoma. Hepatology, 48, 1851-63, 2008
8. Kawai A, Kondo T, Suehara Y, Kikuta K, Hirohashi S, Global protein-expression analysis of bone and soft tissue sarcomas. Clin Orthop Relat Res, 466, 2099-106, 2008
9. Pftin as a prognostic biomarker of gastrointestinal stromal tumors revealed by proteomics. Clin Cancer Res, 14, 1707-17, 2008

## 2. 学会発表

1. Human Proteome Organization年会、オランダアムステルダム市、Tadashi Kondo et al. Cancer Proteomics for Personalized Medicine
2. 日本プロテオーム機構年会、大阪、Tadashi Kondo, Cancer Biomarker Discovery for Personalized Medicine
3. 韓国プロテオーム機構年会、韓国済州島、Tadashi Kondo, Cancer Proteomics for Personalized Medicine
4. Asia-Oceania Human Proteome Organization年会、オーストラリアケアンズ市、Tadashi Kondo, Cancer Proteomics for Personalized Medicine
5. HongKong Human Proteome Organization年会、中華人民共和国香港、Tadashi Kondo, Cancer Proteomics for Biomarker Development
6. Taiwan-Japan Joint Symposium, Taiwan Proteome Society年会、Tadashi Kondo, Cancer Proteomics for Biomarker Development

## H. 知的財産権の出願・登録状況（予定を含む）

### 1. 特許出願

#### 国内出願

「食道癌の予後予測方法」特願2008-278554

#### 国際出願

1. 「消化管間質腫瘍（GIST）を処置するための医薬組成物、ならびに消化管間質腫瘍を患う患者の予後を予測するためのキットおよび方法」PCT/JP2008/71495
2. 「肝細胞癌のマーカーおよび肝細胞癌の検査方法」PCT/JP2008/71495

### 2. 実用新案登録

特記事項なし

### 3. その他

特記事項なし

厚生労働科学研究補助金（第3次対がん総合戦略研究事業）  
分担研究報告書

がん検診に有用な新しい腫瘍マーカーの開発  
分担研究者 西村俊秀 東京医科大学外科第一講座 客員教授

研究要旨

ホルマリン固定肺癌保存組織を用いた肺癌タイプ識別マーカーの  
探索と検証を行った。

A. 研究目的

肺癌は非小細胞癌(non-small cell lung cancer: NSCLC)と小細胞癌(small-cell lung cancer: SCLC)とに大きく分けられる。非小細胞癌中にはいわゆる大細胞癌(large cell cancer: LCC)以外に肺大細胞神経内分泌系癌(large cell neuroendocrine cancer: LCNEC)がここ数年議論されてきた。LCNECは、これまで大細胞癌の亜流として位置づけられてきたが、その予後の悪さと再発率の高さから、できるだけ正確にリスクの高いLCNECであるかどうかを判定する方法が望まれている。また、LCNECの治療戦略についても非小細胞癌の方法がよいか、また小細胞癌の方法が良いかにつき検討が行われているのが現状である。本年度は、ホルマリン固定パラフィン包埋(Formalin-fixed Paraffin Embedded: FFPE)組織切片で、LCNEC 4症例、LCC 5症例、及びSCLC 5症例につき詳細解析でマーカー候補を探索し、質量分析法におけるターゲットした定量法により検証を行った。

B. 研究方法

ホルマリン固定組織は過去の症例で収集されたものであり、患者の病態、投与歴など情報が付随していることから、レトロスペクティブな探索的プロテオーム解析が可能となる。

レーザーマイクロダイセクション(LMD)によりホルマリン固定組織切片における当該細胞箇所を切り出し、Liquid Tissue™(Expression Pathology社)により可溶化後、低流速液体クロマトグラフィーとナノエレクトロスプレーイオン化インターフェースをもつタンデム質量分析計により解析する。測定は各試料あたり繰り返し数3で行った。得られたLC-MS/MSデータにつき、非ラベル化法のひとつであるスペクトラル・カウント法(spectral counting method)に基づいて、候補タンパク質群を見出し、候補タンパク質のペプチドレベルでの定量ができるmultiple-reaction monitoring (MRM) 質量分析法を開発し、これを用いて検証を行った。

(倫理面への配慮)

本研究にあたってインフォームド・コンセントが  
取られたホルマリン固定パラフィン包埋(FFPE)組織切  
片を東京医科大学倫理委員会の承認を得て使用した。  
C. 研究結果

FFPE組織切片(厚さ10 $\mu$ m、面積 2mm x 4 mm)を、脱パラ後、ヘマトキシリン染色を行い、マイクロダイセクションで各試料当たり 30,000セル相当を収集した。収集癌細胞群からタンパク質可溶化技術(Liquid Tissue®)を用いて抽出し、トリプシンによる酵素特異的消化を行いペプチド混合物試料とした。ペプチド混合物は高感度液体クロマトグラフィー-質量分析計(LC-MS)(高分解能質量分析計LTQ-Orbitrap™ MS)によりプロテオーム解析を行い、タンパク質同定には検索ソフト MASCOT v2.1.04を用いた。その結果、LCC、SCLC、LCNECの各々1,000種類以上のタンパク質が同定された。同定されたタンパク質群におけるスペクトラル・カウント法に基づく半定量比較解析から各群に有意と考えられるタンパク質を各々100種類以上見出した。

関心のある44個のタンパク質群についてMRMアッセイをデザインし(182 MRM transitions)、各試料繰り返し数3回で、ハウスキーピングタンパク質由来特定配列をもつペプチドを(試料内)内標として、(相対)定量を行った。

D. 考察

MRMアッセイに基づく検証の結果、20以上の有望マーカータンパク質を同定することができた。その中には、secretagoin (SEGN)、neuron specific enolase (NSE)、neural cell adhesion molecule (NCAM) 及び chromogranin A (CgA) など神経内分泌系(NE)のマーカーといわれている分子が定量され、LCNECとSCLCに特異的に発現していることも確認された。SCLCは、もともと神経内分泌系である。

現在、神経内分泌系(NE)肺癌に特異性を示す3マーカータンパク質、LCC特異的なマーカー候補8タンパク質、また大きい癌に共通に高発現するマーカー2種類、SCLCに特異的なマーカー1種類を同定している。これらのマーカータンパク質はMRMアッセ

イによる特異性の検証を経ているために、准マーカーと呼ぶべきものである。今後、再度検証を行うとともにさらに他の候補を含めたMRMアッセイをデザインして、実施する予定である。

## E. 結論

MRMアッセイに基づく定量により、単なる半定量比較解析による仮説形成ではなく、信頼できる准マーカーが検証できることを明らかにした。20個以上の准マーカーが見出され、FFPE組織を用いるDirect Tissue ProteomicsとMRMアッセイの有用性が確認された。特に、定量を行うMRMアッセイによるバイオマーカー検証法は、新規バイオマーカーの確立に大きな推進力となることが結論された。

## F. 健康危険情報

ホルマリン固定組織試料はごく小さな切片を用い、また蛋白質はすでに架橋・変性しており、本研究のために危険が加わることは無いと考えられる。提供を受けたホルマリン固定組織試料の解析はすべて試料提供者の体外で行うため、試料提供者が被る身体的危険度および副作用はまったくない。

## G. 研究発表

### 1. 論文発表

Toshihide Nishimura, Masaharu Nomura, Tomoyo Nakano, Kiyonaga Fujii, Takeshi Kawamura, Hiroko Hamasaki, Hiroko Endo, Kouichi Yoshida, Tatsu o Ohira, Yasuhiko Bando, Harubumi Kato, Clinical Tissue Proteomics for Personalized Medicine of Early Lung Adenocarcinoma: (1) Candidate discovery from FFPE tissues. Journal of Proteome Res. (Revised version was submitted for publication).

### 2. 学会発表

Kiyonaga Fujii, Tomoyo Nakano, Yasuhiko Bando, Toshihide Nishimura, Fuyuhiko Inagaki, Evidence-based Protein Identification by LC-MS3 Linear Ion Trap Mass Spectrometry, The 56<sup>th</sup> Annual Conference for American Society on Mass Spectrometry, June 1-5, 2008, Denver, Co. USA. ThP22\_472.

Toshihide Nishimura, Masaharu Nomura, Hiroko Endo, Tetsuya Fukuda, Sumie Ando, Kiyonaga Fujii, Hiromasa Tojo, Yasuhiko Bando, Ryutaro Nishiyama, David Krizman, Harubumi Kato, Global MRM MS-Based Assays for FFPE Cancer Tissues Using In-Sample Internal Standards. The 56th Annual C

onference for American Society on Mass Spectrometry, June 1-5, 2008, Denver, Co. USA.

西村 俊秀, 野村 将春, 福田 哲也, 遠藤 洋子, 三上 沙弥香, 西山 隆太郎, 安東 澄江, 藤井 清永, 板東 泰彦, 池田 徳彦, 加藤 治文, 「病理組織プロテオーム研究のための試料調製」, 日本ヒトプロテオーム学会大会要旨集, Vol. 2008 (2008), pp. 91., 7月29日-30日, ホテル阪急エキスポパーク (吹田市).

Toshihide Nishimura, Formalin-Fixed Paraffin-Embedded (FFPE) Clinical Proteome Initiative of Lung Cancer: Toward Construction of Large-Scale Database for Personalized Best-Medicine. AOHUPO & PRICPS JOINT CONFERENCE, Cairns Convention Centre, Australia, June 22 - 26, 2008.

Toshihide Nishimura, Masaharu Nomura, Sayaka Mikami, Hiroko Endo, Yasuhiko Bando, Ryutaro Nishiyama, Kiyonaga Fujii, Sumie Ando, Hiroko Hamazaki, Harubumi Kato, Formalin-fixed paraffin embedded (FFPE) clinical proteome initiative of lung cancer. HUPO 7th Annual World Congress, Amsterdam The Netherlands, August 16-20, 2008. 13775

三上紗弥香, 遠藤洋子, 塩田晃久, 西山隆太郎, 前田めぐみ, 山田誠子, 板東泰彦, 西村俊秀, 「ホルマリン固定パラフィン包埋組織を用いたタンパク質バイオマーカー探索」, BMB2008 (第31回日本分子生物学会年会・第81回日本生化学会大会 合同大会) 3P-0159, 2008年12月9日-12日, 神戸ポートアイランド.

遠藤洋子, 三上紗弥香, 板東泰彦, 西村俊秀, 「リン酸化ペプチドの新規濃縮方法および超高感度LC/MSによる網羅的解析」, BMB2008 (第31回日本分子生物学会年会・第81回日本生化学会大会 合同大会), 3P-0132, 2008年12月9日-12日, 神戸ポートアイランド.

福田哲也, 遠藤洋子, 野村将春, 安藤純江, 藤井清永, 東城博雅, 板東泰彦, 西山隆太郎, Krizman David, 加藤治文, 西村俊秀, 「初期肺腺がんに対する病期・予後・悪性度マーカー検証の新規アッセイ法」, BMB2008 (第31回日本分子生物学会年会・第81回日本生化学会大会 合同大会), 23T19-10及び3P-1068, 2008年12月9日-12日, 神戸ポートアイラ

ンド。

H. 知的財産権の出願・登録状況（予定を含む）

1. 特許取得  
なし
2. 実用新案登録  
なし
3. その他  
なし

別添5 研究成果の刊行に関する一覧表

**Large-scale quantitative clinical proteomics by label-free liquid chromatography and mass spectrometry**

Ayako Negishi, Masaya Ono, Yasushi Handa, Hidenori Kato, Kohki Yamashita, Kazufumi Honda, Miki Shitashige, Reiko Satow, Tomohiro Sakuma, Hideya Kuwabara, Ken Omura, Setsuo Hirohashi, and Tesshi Yamada  
*Cancer Sci*, 100:514-519, 2009.

**p53 gene mutation and p53 protein overexpression in a patient with simultaneous double cancer of the gallbladder and bile duct associated with pancreaticobiliary maljunction**

Kazuhiko Kasuya, Yuichi Nagakawa, Takaaki Matsudo, Takashi Ozawa, Akihiko Tsuchida, Tatsuya Aoki, Takao Itoi, and Fumihide Itokawa  
*J Hepatobiliary Pancreat Surg*, in press.

**Detection of circulating tumor cells in patients with pancreatic cancer: a preliminary result**

Toshio Kurihara, Takao Itoi, Atsushi Sofuni, Fumihide Itokawa, Takayoshi Tsuchiya, Shujiro Tsuji, Kentaro Ishii, Nobuhito Ikeuchi, Akihiko Tsuchida, Kazuhiko Kasuya, Takashi Kawai, Yoshihiro Sakai, and Fuminori Moriyasu  
*J Hepatobiliary Pancreat Surg*, 15:189-195, 2008.

**Personality and Colorectal Cancer: The Fukuoka Colorectal Cancer Study**

Jun Nagano, Suminori Kono, Kengo Toyomura, Tetsuya Mizoue, Guang Yin, Ryuichi Mibu, Masao Tanaka, Yoshihiro Kakeji, Yoshihiko Maehara, Takeshi Okamura, Koji Ikejiri, Kitaroh Futami, Yohichi Yasunami, Takafumi Maekawa, Kenji Takenaka, Hitoshi Ichimiya, and Nobutoshi Imaizumi  
*Jpn J Clin Oncol*, 38:553-561, 2008.

**Current status of chemoradiotherapy for locally advanced pancreatic cancer in Japan**

Takuji Okusaka, Yoshinori Ito, Junji Furuse, Shigeru Yamada, Hiroshi Ishii, and Keiko Shibuya, Tatsuya Ioka, and Hiroyuki Shinchi  
*Int J Clin Oncol*, 13:127-131, 2008.

**Ultrasound-guided percutaneous pancreatic tumor biopsy in pancreatic cancer: a comparison with metastatic liver tumor biopsy, including sensitivity, specificity, and complications**

Junichi Matsubara, Takuji Okusaka, Chigusa Morizane, Masafumi Ikeda, and Hideki Ueno

*J Gastroenterol*, 43:225-232, 2008.

**Epidemiological study of malignant tumors in the oral and maxillofacial region: survey of member institutions of the Japanese Society of Oral and Maxillofacial Surgeons, 2002**

Yasunori Ariyoshi, Masashi Shimahara, Ken Omura, Etsuhide Yamamoto, Harumi Mizuki, Hiroshige Chiba, Yutaka Imai, Shigeyuki Fujita, Masanori Shinohara, and Kanichi Seto

*Int J Clin Oncol*, 13:220-228, 2008.

**Transglutaminase 3 as a prognostic biomarker in esophageal cancer revealed by proteomics**

Norihisa Uemura, Yukihiro Nakanishi, Hoichi Kato, Shigeru Saito, Masato Nagino, Setsuo Hirohashi, and Tadashi Kondo

*Int J Cancer*, 124:2106-2115, 2009.

**Molecular background of  $\alpha$ -fetoprotein in liver cancer cells as revealed by global RNA expression analysis**

Shigeru Saito, Hidenori Ojima, Hitoshi Ichikawa, Setsuo Hirohashi, and Tadashi Kondo  
*Cancer Sci*, 13:127-131, 2008.

# Large-scale quantitative clinical proteomics by label-free liquid chromatography and mass spectrometry

Ayako Negishi,<sup>1,2</sup> Masaya Ono,<sup>1</sup> Yasushi Handa,<sup>3</sup> Hidenori Kato,<sup>3</sup> Kohki Yamashita,<sup>3</sup> Kazufumi Honda,<sup>1</sup> Miki Shitashige,<sup>1</sup> Reiko Satow,<sup>1</sup> Tomohiro Sakuma,<sup>4</sup> Hideya Kuwabara,<sup>4</sup> Ken Omura,<sup>2</sup> Setsuo Hirohashi<sup>1</sup> and Teshi Yamada<sup>1,5</sup>

<sup>1</sup>Chemotherapy Division and Cancer Proteomics Project, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045; <sup>2</sup>Oral and Maxillofacial Surgery, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo, 113-8549; <sup>3</sup>Department of Gynecology, National Hospital Organization Hokkaido Cancer Center, 4-2 Kikusui, Shiroishi-ku, Sapporo 003-0804; <sup>4</sup>BioBusiness Group, Mitsui Knowledge Industry, 3-3-2 Nihonbashihama, Chuo-ku, Tokyo 103-0007, Japan

(Received August 16, 2008/Revised November 1, 2008/Accepted November 4, 2008/Online publication December 22, 2008)

We previously reported the development of an integrated proteome platform, namely 2-Dimensional Image Converted Analysis of Liquid chromatography and mass spectrometry (2DICAL), for quantitative comparison of large peptide datasets generated by nano-flow liquid chromatography (LC) and mass spectrometry (MS). The key technology of 2DICAL was the precise adjustment of the retention time of LC by dynamic programming. In order to apply 2DICAL to clinical studies that require comparison of a large number of patient samples we further refined the calculation algorithm and increased the accuracy and speed of the peptide peak alignment using a greedy algorithm, which had been used for fast DNA sequence alignment. The peptide peaks of each sample with the same *m/z* were extracted every 1 *m/z* and displayed with along the horizontal axis. Here we report a precise comparison of more than 150 000 typical peptide ion peaks derived from 70 serum samples (40 patients with uterine endometrial cancer and 30 controls). The levels of 49 MS peaks were found to differ significantly between cancer patients and controls ( $P < 0.01$ , Welch's *t*-test and interquartile range [IQR] of  $>40$ ), and the differential expression and identification of selected three proteins was validated by immunoblotting. 2DICAL was highly advantageous for large-scale clinical proteomics because of its simple procedure, high throughput, and quantification accuracy. (*Cancer Sci* 2009; 100: 514–519)

The incidence and morbidity of endometrial cancer have been rising in Western countries and Japan. In 1970, endometrial cancer constituted only 3% of all uterine cancers, but the proportion had increased to 40% by 1998.<sup>(1)</sup> The contemporary human lifestyle is characterized by excessive fat consumption, obesity, physical inactivity, high energy intake, hypertension, and a high serum glucose concentration, which are risk factors for endometrial cancer,<sup>(2–6)</sup> and therefore its incidence is predicted to increase further. Abnormal uterine bleeding is the most frequent initial symptom of endometrial cancer, but many other disorders including endometriosis, metropothesis, myoma uteri, and uterine sarcoma can also produce this symptom. Endometrial cancer is usually diagnosed by histological examination of endometrial tissue. However, endometrial biopsy is often associated with complications such as infection, bleeding and perforation of the uterus, and the development of an alternative safe test is therefore necessary for diagnosis of endometrial cancer.

The circulating serum proteome holds great promise as a reservoir of information that will be applicable for diagnosis of various diseases. Various gel-based and MS-based quantitative analyses of plasma/serum proteins have been conducted actively to identify serum biomarkers that can be applied to blood tests.<sup>(7–10)</sup> HPLC with a flow rate of the nanoliter-per-minute order coupled with high-speed MS/MS scanning has especially attracted considerable attention because of its comprehensive protein identification capacity.<sup>(11–13)</sup> However, MS/MS is essentially

not a method for quantification, and several attempts have been made to provide a quantitative dimension to nano-LC-MS/MS. Semi-quantification is possible to some extent by counting the number of peptides sequenced.<sup>(14–16)</sup> To achieve more accurate quantification, several labeling methods have been developed,<sup>(17–19)</sup> but the number of samples that can be compared with these isotope labeling methods is limited, and the efficacy of labeling cannot be controlled from one experiment to another. Furthermore, the long-term stability of nano-LC-MS can be troublesome. The sensitivity, resolution, and profile of MS instruments change over time, and more seriously the flow condition of LC may fluctuate even from sample to sample. All of these factors prevent the use of nano-LC-MS/MS for quantitative clinical studies in which comparison of a large number of samples is necessary.

We previously developed a labeling-free MS-based quantitative proteome platform known as 2DICAL.<sup>(20)</sup> Unlabeled protein samples were digested with trypsin and separated by nano-flow (200 nL/min) HPLC. MS spectra were obtained every 1 s (3600 scans/h) by using the fast scan mode of quadrupole time-of-flight (QTOF)-MS. Because we eliminated time-consuming MS/MS in the biomarker discovery phase, 60 000–160 000 MS peaks could be detected and quantified during a 1-h run of LC-MS. Using 2DICAL we reported comprehensive analyses of proteins that were expressed differently in non-clinical samples between metastatic and non-metastatic cancer cell lines. In order to apply 2DICAL to clinical studies that require the comparison of a large number of patient samples, we further refined the calculation algorithm and improved the accuracy and speed of the peak alignment of 2DICAL (Ono *et al.* manuscript in preparation). In this short technical report, we describe a large-scale quantitative comparison of the serum proteome of endometrial cancer patients using 2DICAL. This method can be applied to clinical studies of any kind, especially those in which isotope labeling is not applicable.

## Materials and Methods

**Clinical samples.** Serum samples ( $n = 70$ ) were collected at the National Hospital Organization Hokkaido Cancer Center Hospital (Sapporo, Japan) between 2000 and 2004 with the informed consent of the donors.<sup>(21)</sup> The subjects included 40 patients with untreated endometrial cancer (age  $54.0 \pm 10.6$  years) and 30 controls ( $50.3 \pm 16.3$  years) comprising healthy volunteers ( $n = 15$ ) and patients with metropothesis ( $n = 15$ ). The samples were collected

<sup>5</sup>To whom correspondence should be addressed. E-mail: tyamada@ncc.go.jp  
Abbreviations: 2DICAL, 2-dimensional image converted analysis of liquid chromatography and mass spectrometry; CC, correlation coefficient; HPLC, high-performance liquid chromatography; IHRP, inter- $\alpha$ -trypsin inhibitor family heavy chain-related protein; IQR, interquartile range; LC-MS, liquid chromatography and mass spectrometry; MS/MS, tandem mass spectrometry; *m/z* (M/Z), mass-to-charge ratio; nano-ESI, nano-electrospray ionization; QTOF-MS, quadrupole time-of-flight mass spectrometry; RT, retention time.

in glass tubes, and after allowing them to clot, the serum was separated and cryopreserved at  $-80^{\circ}\text{C}$  until analysis. The protocol of the study was reviewed and approved by the ethics committees of the National Hospital Organization Hokkaido Cancer Center (Sapporo, Japan) and the National Cancer Center (Tokyo, Japan).

**Sample preparation.** Twelve highly abundant serum proteins (albumin, IgG,  $\alpha$ 1-antitrypsin, IgA, IgM, transferrin, haptoglobin,  $\alpha$ 1-acid glycoprotein,  $\alpha$ 2-macroglobulin, apolipoprotein A-I, apolipoprotein A-II, and fibrinogen) that constitute ~95% of the total protein mass of human serum, were depleted using an IgY-12 spin column (Beckman Coulter, Fullerton, CA, USA) as instructed by the supplier. Briefly, a 10- $\mu\text{L}$  serum sample was diluted with 490  $\mu\text{L}$  dilution buffer (1:50 dilution) and applied to an IgY-12 spin column. After 15 min of incubation with constant rotation on an end-to-end rotator, the flow-through fraction was recovered into 2-mL microcentrifuge tubes by centrifugation at 300 g for 1 min. The half volume (250  $\mu\text{L}$ ) was precipitated with 250  $\mu\text{L}$  of acetonitrile, dried with a SpeedVac concentrator (ThermoFisher Scientific, Waltham, MA, USA), and dissolved in 30  $\mu\text{L}$  of milliQ water, to which 10  $\mu\text{L}$  of 5 M urea, 2.5  $\mu\text{L}$  of 1 M  $\text{NH}_4\text{HCO}_3$ , and 3.3  $\mu\text{g}$  of Sequencing Grade Modified trypsin (Promega, Madison, WI, USA) were added. After 20 h of digestion at  $37^{\circ}\text{C}$ , peptides were precipitated with 50- $\mu\text{L}$  acetonitrile, dried with a SpeedVac concentrator, and then dissolved in 80  $\mu\text{L}$  of 0.1% formic acid.

**LC-MS.** LC-MS and data acquisition were performed as reported previously.<sup>(20)</sup> Briefly, the peptide samples were separated in triplicate with a linear gradient from 0 to 80% acetonitrile in 0.1% formic acid at a speed of 200 nL/min for 60 min using a splitless nano-flow HPLC system (DiNa; KYA Technologies, Tokyo, Japan, or NanoFrontier; Hitachi High-technologies, Tokyo, Japan). MS spectra were acquired with a nano-ESI-QTOF-MS (QTOF Ultima; Waters, Milford, MA, USA) every second for 60 min in the 250–1600  $m/z$  range.

**Peak display and statistical analyses.** MS peaks of each sample with the same  $m/z$  were extracted every 1  $m/z$  and aligned with the retention time of LC along the horizontal axis. A proper identification (ID) number was applied to each MS peak. Statistically significant differences were detected using Welch's  $t$ -test and IQR. The statistical tests were performed using tools available in the R statistical package (Version 2.0.1) (<http://www.r-project.org/>).

**Protein identification by MS/MS.** Peak lists were created using the MassNavigator software package (Version 1.2; Mitsui Knowledge Industry, Tokyo, Japan). Target MS/MS data acquisition for candidate peaks was performed on the same samples. MS/MS data were aligned with the reference data, and the target peak was extracted with a tolerance of  $\pm 0.5$   $m/z$  and  $\pm 0.4$  min of RT. The MS/MS data were analyzed with the Mascot software package (Version 2.2.1; Matrix Sciences, London, UK) against the NCBI database (NCBIInr\_20070419.fast) using the following parameters. Initial peptide tolerances in MS and MS/MS modes were  $\pm 0.05$  Da and  $\pm 0.1$  Da, respectively. Trypsin was designated as the enzyme, and up to one missed cleavage was allowed. The score threshold to achieve  $P < 0.05$  is set by the Mascot algorithm, and is based on the size of the database used in the search.

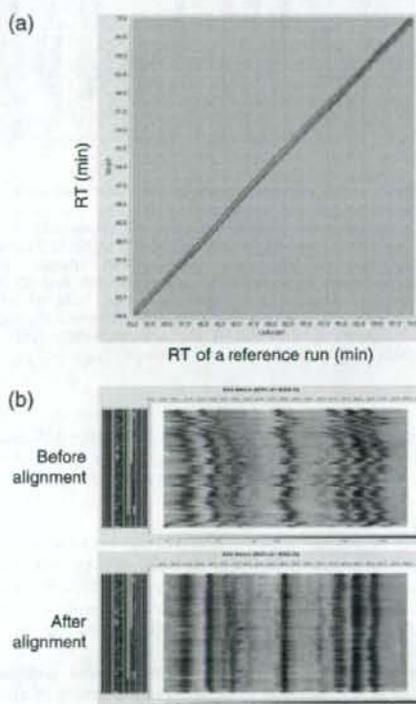
**Immunoblot analysis.** Anti-human apolipoprotein A-IV mouse monoclonal antibody (A4-18A3) was purchased from BML (Saitama, Japan). Anti-complement component C3 mouse monoclonal antibody (GAU 017-01) and anticomplement component C4A mouse monoclonal antibody (HYB 162-02) were purchased from AntibodyShop (Gentofte, Denmark). Anti-human  $\alpha$ 2-macroglobulin mouse monoclonal antibody was purchased from R&D Systems (Minneapolis, MN, USA) and used as a loading control.

Serum samples were separated by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and electroblotted onto polyvinylidene difluoride (PVDF) membranes. The blots were incubated with one of the above primary antibodies and

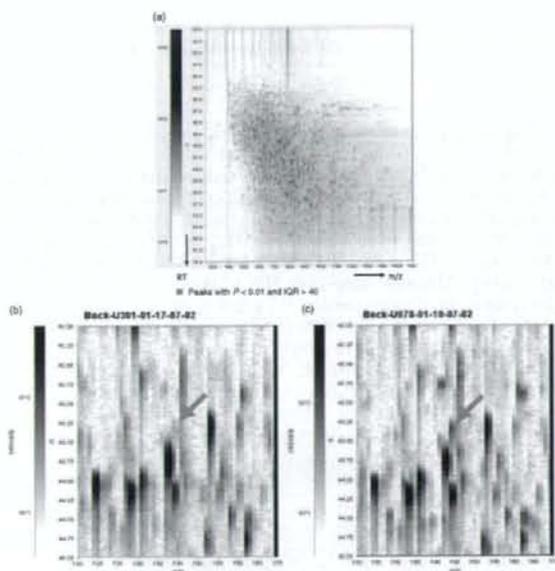
relevant horseradish peroxidase-conjugated secondary antibody and detected using an enhanced chemiluminescence kit.<sup>(22)</sup>

## Results

**High-throughput LC-MS analysis of clinical samples.** Seventy serum samples (from 40 uterine endometrial cancer patients and 30 controls) were blinded, randomized, and measured in triplicate. We performed an average of 10.5 LC-MS runs every day (3.5 samples/day), and the entire 210 runs were finished in a month without any serious problems. Fig. 1(a) is a plot depicting the difference of RT in the 209 runs from a reference run selected at random. The new nano-flow HPLC system with a degassing function showed much smaller RT jittering (0.32 min on average) than our previous HPLC system (1.68 min). However, if MS spectra of particular  $m/z$  were displayed along the RT, there was still significant time jittering among runs (before alignment, Fig. 1b). However, the small RT jittering was readily adjustable (after



**Fig. 1.** Fine alignment of mass spectrometry (MS) peaks by retention time (RT) adjustment of 210 LC-MS runs. (a) Comparison of RT (in the range 30.0–70.0 min) between a reference liquid chromatography and mass spectrometry (LC-MS) run (plotted on the horizontal axis) and the other 209 LC-MS runs (plotted on the vertical axis). Seventy serum samples (from 40 uterine endometrial cancer patients and 30 controls) were run in triplicate. One of the 210 LC-MS runs was selected at random as the reference. MS peaks having the same  $m/z$  but different LC-flows were selected using a greedy algorithm and aligned by adjustment of RT.<sup>(20)</sup> The differences of RT from the reference run could be assessed by the deviation from the central 45-degree line, but no significant RT difference was apparent among the 210 LC-MS runs. (b) 622.0 (621.5–622.5)- $m/z$  MS peaks of the 210 runs (70 samples in triplicate) displayed with RT (39.5–50.5 min) along the horizontal axes. Although no significant RT variation was observed in the global analysis (a), it was difficult to compare MS peaks among runs due to small variations of RT before RT alignment (top). A comparison became possible only after RT adjustment (bottom).



**Fig. 2.** Location of the selected 49 peaks on the 2-dimensional gel-like view. (a) Two-dimensional display of >150 000 mass spectrometry (MS) peaks of a representative sample with the *m/z*-values (250–1600 *m/z*) along the horizontal axis and retention time (RT) (20.0–70.0 min) along the vertical axis. 49 MS peaks whose mean intensity differed significantly between 40 endometrial cancer patients and 30 controls ( $P < 0.01$ , Welch's *t*-test and interquartile range (IQR) of >40) are highlighted in red. (b, c) A representative peak differentially expressed between a healthy control (U301) (b) and an endometrial cancer patient (U078) (c) is highlighted by red arrows.

alignment, Fig. 1b). The CC value for the entire MS peaks of triplicate runs was at least 0.96 throughout the experiment (Supporting Fig. S1a–c). We also ran a standard serum mixture at least once a week to ensure consistency of the LC-MS during the same period (Supporting Fig. S1d).

After RT alignment, a total of 154 992 MS peaks were detected across the 70 serum samples in the 250 *m/z* to 1600 *m/z* and 20–70 min range, and these were numbered serially from ID 1 to ID 154992. The entire MS peaks detected in a representative run are displayed in Fig. 2(a), with *m/z*-values along the horizontal axis and RT along the vertical axis.

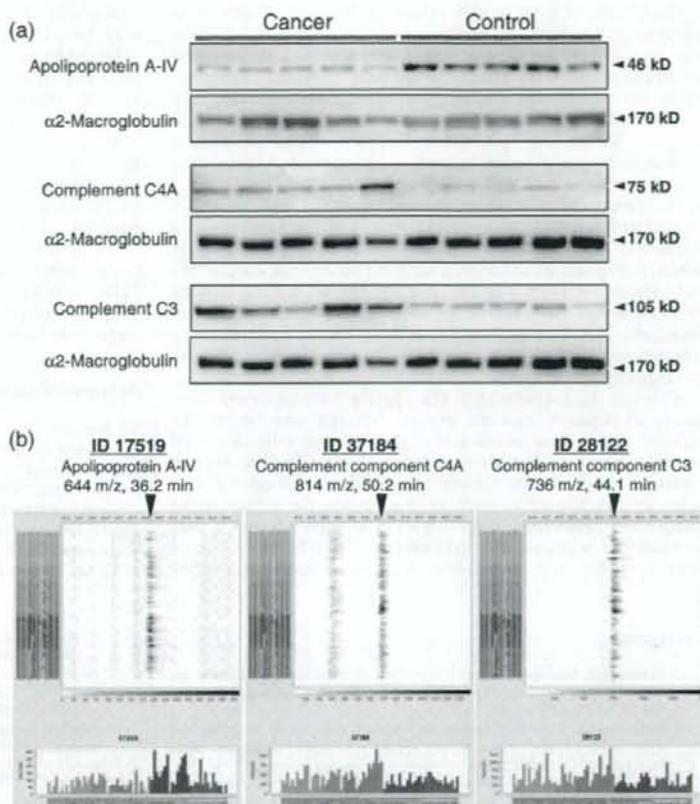
**Difference between uterine endometrial cancer patients and controls.** After normalizing to the total ion intensity of all of the peaks, the maximum intensity of each peak was averaged. To eliminate peaks that had little or no variation across samples (peaks that were not working well), we selected 1043 peaks whose intensity had an IQR of >40 across the 70 samples. There were 49 MS peaks showing a statistically significant difference between the cancer patients and controls ( $P < 0.01$ , Welch's *t*-test) (highlighted in red, Fig. 2a). Fig. 2(b,c) depicts a representative MS peak (indicated by an arrow) differentially expressed between a cancer patient (Fig. 2b) and a control (Fig. 2c).

**Validation of differential expression.** The 49 MS peaks were subjected to MS/MS for protein identification. A database search revealed that the MS/MS spectra of 11 peaks matched the amino acid sequences of proteins deposited in the NCBI database (NCBIInr\_20070419.fast) with significant confidence ( $P < 0.05$ ), including apolipoprotein A-IV, IHRP, and complement components C3, C4A, and C4B (Table 1 and Supporting Figs. S2–5). It is

**Table 1.** Detected peptides that differed between patients with uterine endometrial cancer and controls

ID	M/Z	Charge	Accession number	Protein description	Peptide sequence	Mascot score	Average peak intensity		P-values*
							Cancer (n = 40)	Control (n = 30)	
8627	552	2	gi 178757	Apolipoprotein A-IV precursor	IDQVVEELR	76.09	101.3 ± 41	136.1 ± 58.6	0.005815
9150	558	2	gi 179674	Complement component C4A	VGDTLNILNR	51.60	204.4 ± 102.6	142.3 ± 55.6	0.002147
10500	573	2	gi 1041907	GP120, IHRP = ITI heavy chain-related protein (internal fragment)	GPDVLIATVSGK	51.33	131.3 ± 68.6	86.7 ± 44.8	0.002911
13516	605	2	gi 179665	Complement component C3	YYTYLIMNK	40.95	201.1 ± 87.2	157.4 ± 45.8	0.009564
14166	611	2	gi 1483187	Inter- $\alpha$ -trypsin inhibitor family heavy chain-related protein (IHRP)	ETLFSVMPGLK	49.85	155.0 ± 50.1	123.4 ± 31.1	0.002124
14947	619	2	gi 179674	Complement component C4A	KYVLPNFEVK	37.50	90.6 ± 42.0	68.7 ± 21.0	0.006617
17519	644	2	gi 178757	Apolipoprotein A-IV precursor	TQVNTQAEQLR	59.77	71.5 ± 44.9	127.9 ± 81.3	0.001617
25565	715	3	gi 179674	Complement component C4A	MRPSTDTITVAVNESHGLR	48.54	301.4 ± 147.6	225.7 ± 77.1	0.008078
23827	717	2	gi 87191	Complement C4B-human (fragment)	AEMADQAAAALTR	47.75	134.2 ± 83.9	93.5 ± 38.6	0.009756
28122	736	2	gi 179665	Complement component C3	IPIEDSGSEVILSR	74.91	582.2 ± 293.0	405.6 ± 185.3	0.003440
37184	814	2	gi 179674	Complement component C4A	EELVYELNPLDHR	89.65	477.2 ± 228.5	348.2 ± 122.3	0.003940

\*Welch's *t*-test.



**Fig. 3.** Confirmation of differential expression. (a) Confirmation of differential expression of apolipoprotein A-IV, complement component C4A, and complement component C3 in serum samples from five endometrial cancer patients and five controls by immunoblotting. (b) Gel-like view with retention time (RT) along the horizontal axis (top) and distribution of peak intensity (bottom) of apolipoprotein A-IV (left), complement component C4A (middle), and complement component C3 (right) in 40 cancer patients (red) and 30 controls (blue).

noteworthy that apolipoprotein A-IV, complement component C4A, complement component C3, and IHRP were repeatedly detected two, four, two, and two times, respectively (Table 1). The identification and differential expression were confirmed by immunoblotting using sera from five endometrial cancer patients and five healthy controls (Fig. 3a). Apolipoprotein A-IV was down-regulated in the cancer patients, whereas complement components C4A and C3 were up-regulated, confirming the results of 2DICAL (Fig. 3b).

## Discussion

The entire protein content (or proteome) of each human sample varies significantly from individual to individual, and 100 biomarkers with  $P$ -values of  $<0.01$  might be "discovered" in a large data set consisting of 10 000 variables. The accidental identification of mere personal heterogeneity as a biomarker candidate should be avoided by comparing a sufficient number of cases and controls. "Biomarkers" identified in a small sample set rarely pass through the pipeline connecting the biomarker discovery phase and the preclinical phase. According to our calculations, when the experiment involves a comparison of 10 cases and 10 controls, a factor that has a 90% discrimination rate can have a confidence interval that ranges from 64% to 98%, and comparison of at least 40 cases and 40 controls would be necessary to discover a biomarker with a diagnostic accuracy of  $>90\%$  even with the use of high generalization capability algorithms such as support vector machine (Kuwabara *et al.* manuscript in preparation).

In order to apply 2DICAL to clinical studies that require comparison of such a scale of patient samples, we refined the calculation algorithm and increased the speed and accuracy of the MS peak alignment. We exploited the fact that the order of appearance of the peptide peaks is constant even though the LC RT fluctuates. We first selected MS peaks having the same  $m/z$  every 1  $m/z$  and displayed their MS spectra along the LC RT. Peptide peaks having the same  $m/z$  but different LC-flows were aligned by adjusting RT, and the corresponding peaks were selected using a greedy algorithm that had been used for fast DNA sequence alignment. Using the new algorithm we compared over 150 000 MS peaks of 40 uterine cancer patients and 30 controls by aligning more than 10 000 000 data points, and identified 49 MS peaks whose intensity differed significantly between cancer patients and controls ( $P < 0.01$ , Welch's  $t$ -test and IQR of  $>40$ ). We performed MS/MS analysis for protein identification (Supporting Figs. S2–5), and the differential expression of representative proteins was confirmed by immunoblotting (Fig. 3a).

Direct mass spectrometric analysis of plasma/serum samples is considered to be most immediate to the discovery of protein biomarkers of various diseases, but it has been hampered by the marked dominance of a handful of particularly abundant plasma/serum proteins, including albumin, immunoglobulins, and transferrins, and thus removal of these abundant proteins and reduction of sample complexity are necessary for comprehensive protein profiling.<sup>(23,24)</sup> The highly abundant proteins present in quantities of mg/mL make up less than 0.1% of the total number of proteins, yet they constitute more than 95% of