

非標識サンプルの質量分析解析からタンパク質を定量する 簡易エクセルツールの作成とその応用

生 駒 俊 和

新潟大学大学院医歯学総合研究科生体機能調節医学専攻

(指導：山本 格教授)

An Excel VBA Tool Works on Protein Identification Results
by Using Label Free NanoLC - MS Analysis

Toshikazu IKOMA

Department of Structural Pathology Institute of Nephrology, Niigata University

Graduate School of Medical and Dental Sciences

(Director; Prof. Tadashi YAMAMOTO)

要 旨

質量分析計の進歩によりタンパク質を網羅的に解析するプロテオミクスではその情報が膨大となり、ある注目するタンパク質の定量情報を研究者が手作業で抽出することはほとんど不可能になった。この問題を解決するために高価な市販の解析ソフトもあるが、本研究では質量分析によるプロテオミクスで非標識サンプルのタンパク質を定量する指標として使われている Normalized Spectral Index (SI_N) をエクセルで算出する簡便なツールの作成を目指した。

それを用いた応用として、ヒト腎臓の皮質 (6 サンプル) と糸球体 (3 サンプル) の液体内トリプシン処理ペプチドサンプルを液体クロマトグラフィー質量分析装置で解析し (測定条件: MS target parameter: 732 m/z と 1,032 m/z の 2 つの条件で測定), Mascot サーチエンジンでタンパク質を同定した情報から各サンプル中のタンパク質の定量値, SI_N を算出した。また、タンパク質同定の信頼性を高めることを可能にするため、本ツールではペプチド同定の信頼性を示すスコア, peptide score が 0 以上の全てのペプチド情報を使う条件と 20 以上の信頼性の高いペプチド情報だけを使う条件で SI_N を算出できるようにした。さらに、このツールを用いて、サンプルの均質性、プロテオーム解析結果の信頼性、再現性を検証するために、同一皮質サンプルと糸球体サンプルを 3 回測定した結果と糸球体サンプルに人為的にヒトのフケ成分をいろいろな割合で添加したサンプルの測定結果で、いくつかのタンパク質の SI_N を求め、その変動係数率 (CV) を求めた。

その結果、皮質に有意に多いタンパク質が 55 種類、糸球体に有意に多いタンパク質が 23 種類を抽出できた。また、本ツールを利用し、CV を算定、比較することで各プロテオーム解析結果の信頼性を検証することが出来た。このツールは個人のパソコンでタンパク質の同定情報から注目するタンパク質の SI_N を算出し、発現レベルの比較を容易にできるものであることが示

Reprint requests to: Toshikazu IKOMA
Department of Structural Pathology Institute of
Nephrology, Niigata University Graduate School
of Medical and Dental Sciences,
1-757 Asahimachi - dori, Chuo - ku,
Niigata 951 - 8510, Japan.

別刷請求先: 〒951 - 8510 新潟市中央区旭町通 1 - 757
新潟大学大学院医歯学総合研究科
附属腎研究施設構造病理学分野 生駒 俊 和

された。

キーワード：プロテオーム、ラベルフリー、Normalized Spectral Index (SI_N), MASCOT, VBA

はじめに

ヒトゲノムの解析が進み、ヒト遺伝子の総数は約2.2万と推定されている¹⁾が、その遺伝子から約200万個のタンパク質が作り出されていると推定されている²⁾。病気の発症や予後はタンパク質の機能の変化が重要視されている。そのため、タンパク質を網羅的に解析し、その構造や発現量の変化から病態を解明することを目指すプロテオミクス研究が盛んに行われている³⁾⁴⁾。

これまで新潟大学大学院医歯学総合研究科構造病理学研究室ではヒト正常腎組織及び腎疾患組織を対象とした質量分析装置を用いたプロテオミクス解析を行い、それぞれ同定されたタンパク質情報を処理し、その質と量的な違いを検討してきた。しかし、その解析からのタンパク質関連情報のデータ量は大容量で、多数の標本から目的タンパク質の情報を調べるには、大容量のタンパク質関連情報のデータからタンパク質の数や発現量を整理し、容易に比較できるツールが必要となった⁵⁾⁶⁾。

質量分析プロテオミクスでタンパク質を定量するには、タンパク質やペプチドを安定同位体などで標識する標識法と標識物質を使わない非標識法があるが、前者は定量性が高いが、検査が煩雑で、解析コストが高いのが欠点である。一方、後者はその定量性が欠点と言われてきたが、近年、定量性が高い指標として Normalized Spectral Index (SI_N) が提唱されている。

本研究では質量分析装置で分析し、MASCOTサーチエンジンで同定されたタンパク質同定情報を、一定のフォーマットで出力したデータをMicrosoft社Excelに読み込み、目的とするタンパク質の定量的な情報であるSI_Nを算出するツールの作成を目指した。

材料及び方法

1. 材料

1) 正常ヒト腎の皮質と糸球体

全腎摘出手術を受けた患者さんから、インフォームドコンセントを得て、提供された腎臓から光学顕微鏡及び抗体組織検査で正常と判断された組織部分を得た。組織の取り扱いは新潟大学倫理委員会に定めた規定に従って行われた。糸球体の単離などは氷上にて、4℃の緩衝液を用い、篩法で行った。皮質と糸球体をホモジェネートし、尿素を含む緩衝液を用いて可溶化した。

2) ヒトフケ

ヒトフケはヒトの臓器、組織、体液などのタンパク質を電気泳動で分離し、そのタンパク質をゲル内トリプシン消化し、質量分析装置でタンパク質を同定する際に空中の浮遊物として、解析サンプルに混入する主な汚染物質で、タンパク質の同定の障害となるものである。今回のヒトフケを頭部から採取した。

2. ペプチドの調整

1) タンパク質抽出法

尿素を含む緩衝液(9.8 M尿素, 2% NP-40)を用いて⁷⁾、氷上で2-3mm³に切断したヒト腎皮質やシービング法で精製された糸球体を5-6 mg/ml (W/V)の濃度で調整し、さらにPrecellys 24 (Bertin. Tec. CO. French) ホモジナイザーで強化チューブ(MK28-R)にてメタルビーズを用いて、ホモジナイザした。6,000 rpm 15秒3回、毎回間30秒オンアイスで処理した。さらに37℃恒温槽にて2時間インキュベーションし、12,000 rpm 5分間遠心した後、上層をタンパク質標本とした。タンパク質定量をラマグリ法で行った。最終的に1-3 μg/μlになるように同じ緩衝液で希釈した。

ヒトフケは量が少なかったため、液内トリプシン消化を行った。3mg フケを500 μ l の50mM Tris-HCl 緩衝液で超音波ホモジナイザーを用いて15秒3回で可溶化した。さらに12,000 rpm で5分間遠心し、上清を還元とアルギル化を行った後トリプシン消化を行った。DTTとIAA反応濃度はそれぞれ20 mMと100 mMで行い、トリプシン最終濃度は20 ng/ μ lになるように調整した。消化は約16時間37℃で行った。

2) ペプチド精製

ヒト腎組織由来タンパク質は10% SDS-PAGEで展開し、その後ゲル内消化を行った。詳細操作は許波らの論文に従って行った⁸⁾。ペプチドの精製はStageTip[®]を用いて行った。操作はメーカーの取り扱い説明書通りで行った。さらに精製されたペプチドは凍結遠心乾燥を行い、-30℃で保存した。MS測定を行う前に超純水に溶かし、濃度を測定した。

3. MS測定

1) MS測定は主にAgilent社 nano LC-MS/MS (Agilent 6300 series LC-MSD)で行った。0.7-1 μ g/run ペプチドが使用した。基本的には同じグループの検体は同濃度で測定を行った。MS測定はウルトラスキャンモードで行い、exclusion timeは30秒に設定した。因みに30秒以内に同じペプチド由来したプリカサーイオンは一回しか測定しないようにMS機器の測定効率を上げていた。

測定条件は次の通りである。LC: Trap column; ZORBAX 300SB-C18, 5 μ m, 5mm \times 0.3mm, Germany. Separation (Analysis) column; ZORBAX 300SB-C18, 3.5 μ m, 150mm \times 0.075mm, Germany. Flow rate; 300nl/min. Mobile phase gradient: A solvent; 98% milli-Q water, 2.0% methanol, 0.1% formic acid and, B solvent; 98% methanol, 2.0% milli-Q water, 0.1% formic acid, (120min linear gradient method; 2-70% B solvent in 120min; 70% B isocratic run for 5 min; 100% B isocratic run for 10 min). MS: Instrument type; ESI-ion TRAP, MS range (m/z); 400-2,200, MS Threshold Absolution; 10,000, MS

Threshold Relative; 50%, MS Excluded after 2 Spectra (30 seconds), MS Target parameter; 732 m/z, MS Severy scan; top3 (z = 2+ or 3+), MS/MS Spectrum acquisition; Ultra Scan, Dynamic exclusion time; 30sec, Peptide mass tolerance; \pm 2.5Da, Fragment mass tolerance; \pm 0.7 Da.

2) 測定したデータとしてAgilent社独自のファイル方式で記録されていた。MSACOT検索を行うためにファイルの変換は欠かせない状況であった。データのコンバートはAgilent社が提供したDATAnalysis[®]で行った。変換されたmgf (Mascot Genetic File) ファイルは元ファイルに記録されたMSとMS/MS情報をすべて含まれるように変換条件を検討した。

3) 測定回数

正常な腎臓の腎皮質由来の同一ペプチドサンプルを6サンプル、糸球体由来の同一ペプチドサンプルを3サンプルの測定をおこなった。質量分析計の測定条件の違いによる影響を検証するために上記のサンプルをMS target parameterを732 m/zと1,032 m/zに変更して測定した。合計18回測定をおこなった。

また、再現性の検証には糸球体由来のペプチドサンプルにヒトフケ由来のペプチドサンプルを2%, 20%, 100%添加させた(フケ添加サンプル)3サンプルを各々3回測定し、その同定情報結果を実験結果再現性の検証ツールでの解析に利用した。

4. MASCOT検索

1) 今回のMSデータはすべてヒト腎組織やヒトフケ由来したペプチドと考えられた。同定した際に使われたデータベースはIPI Humanデータベース(ver.: 3.75)であった。検索条件としてMSトレランスは2.5DaでMS/MSトレランスは0.7Daと指定した。同時にCysのカルボルメチル化が固定した修飾条件で指定し、Met, His, Trpについては酸化反応が選択的修飾条件に指定した。

2) 同定結果は本研究で開発したツールを利用するために欠かせないパラメーターを指定し、決まったフォーマットで出力することが必要で、出力

ファイル形式と項目の指定は表1に示す通りである。

5. プロテオーム解析ツールの作成

MASCOT サーチエンジンにより同定されたタンパク質の同定情報データ (CSV ファイルでデータ出力) からタンパク質の定量的情報を算出するツールを作成した。また、算出された定量的情報から2群間のデータ比較ツール、実験結果再現性の検証ツールを作成した。データ処理の流れを図1に示す。

マスペクトラムの intensity 情報を処理することによって個々のタンパク質の定量的情報を満たすことが報告されている。その原理にしたがって、MASCOT サーチエンジンにより出力されたタンパク質同定結果から調べたいタンパク質の量的情報である Normalized Spectral Index (SI_N) を算出できる簡易ツールを作成した。計算式は以下のとおりである。

$SI_N = SI_{GI}/L$; SI: Spectral Index, GI: Global/Total intensity, $SI_{GI} = SI / \sum_{i=1}^n SI_i$, L: Protein Length [number of amino acids]⁹⁾。ただし、 SI_N の数値は小さい値になるのでこの簡易ツールでは SI_N に 10,000 を掛けた値で表示した。

また、MASCOT の解析データの信頼性は pep_score (Peptide score : 照合のイオンスコア、P 値から求めた値である。P 値とはピークリストに対するアミノ酸配列の照合の度合いを数値化したもので、MASCOT アルゴリズム独自の方法で算出されたもの。) が 20 以上であればタンパク質の信頼性が高いと報告されている¹⁰⁾。しかし、研究者によって意見が異なることがあるため pep_score 値を自由に変更できるようにした¹¹⁾。

タンパク質の定量的情報算出ツールの開発には Microsoft 社表計算ソフトエクセルの関数、及び VBA (Visual Basic for Application) を用いることで汎用性を考慮し、また、操作性を高めるため、ユーザはボタンクリック操作のみで使用できる部分を多くした。

この簡易ツールは、同定情報データ出力ファイル読み込み機能、Gene_Symbol 検索機能、 SI_N 出

力機能の3つからなっている。同定情報データ出力ファイル読み込み機能では、複数の MASCOT サーチエンジンにより出力されたタンパク質同定情報のファイルをまとめて処理するために、複数のファイルを一括で取り込めるようにしてある。また、出力された各タンパク質の同定情報に重複データ (各ファイルごと、同定されたタンパク質の Gene_Symbol と pep_query スコアが一致した場合、prot_score 値が大きいものをもとに) がある場合は自動的に計算に使用されない機能を組み込んだ。読み込み後、選択した範囲のみをコピーする機能や読み込んだデータファイルを削除する機能も組み込んだ。このことにより、多くのサンプルの同定情報ファイルを連続的な使用ができるようになる。読み込んだ各タンパク質の同定情報は各ファイルごとにワークシートに保存され、すべてのペプチドを同定される際に使われたスペクトラム (pep_query) から、レポートされたそれらのイオン強度 (intensity) を求める。この情報もとに同定されたタンパク質の SI_{GI} を算出し、同定されたタンパク質のアミノ酸配列の数 (Prot_len) で SI_{GI} を割ることによって SI_N が求められる。

この操作は Main ワークシート内のデータシート読込ボタンをクリックすると、ファイルを指定するダイアログボックスが開き、読み込み対象となる各タンパク質の同定情報ファイルを指定することが出来る。コントロールキーを押しながらすることで複数指定することも可能である。指定後、「開く」ボタンをクリックすることで指定された各タンパク質の同定情報ファイルのデータを新しいワークシート内に保存されるようにした。この時複数ファイル選択時の対応として各タンパク質の同定情報ファイル名をワークシート名とした。

各タンパク質の同定情報ファイルの読み込みについては行単位でデータの読み込みを行い、Protein hits の部分のみを出力するようにした。また、pep_query の番号と query の query_number を関連させ、intensity および、Intensity/len の値を最後の列に追加した。

表1 MASCOT レポートのファイル形式と項目の指定 (CSV出力)

Export search results	
Export format	CSV
Significance threshold P<	0.05
Ions score cut-off	0
Threshold type	Identity
Max. number of hits	AUTO
Protein scoring	MudPIT
include sub-set protein hits (additional proteins that span a sub-set of peptides)	0
Check items	
Search Information	
Header	
Decoy	
Modification deltas	
Search parameters	
Format parameters	
Protein Hit Information	
Score	
Description	
Mass(Da)	
Number of queries matched	
Percent coverage	
Length in residues	
pI	
Peptide Match Information	
Experimental Mr (Da)	
Experimental charge	
Score	
Identity threshold	
Expectation value	
Sequence	
Variable Modifications	
Query title	
Query Level Information	
Query title	

注：MASCOT は version 2.3.01 を使用した。

Gene_Symbol 検索機能では、まず、読み込んだ各タンパク質の同定情報結果ごとに prot_hit_num, prot_desc の項目を抽出し、ワークシート「Work_description」に格納した後、prot_desc の文字列を用いて Gene_Symbol を抽出し、prot_

hit_num ごとに自動で統合した。これにより同定されたタンパク質の個数を算出することが出来る。

ワークシート「Work_Gene」で Gene_Symbol 出力をクリックすることによりツールに保存され

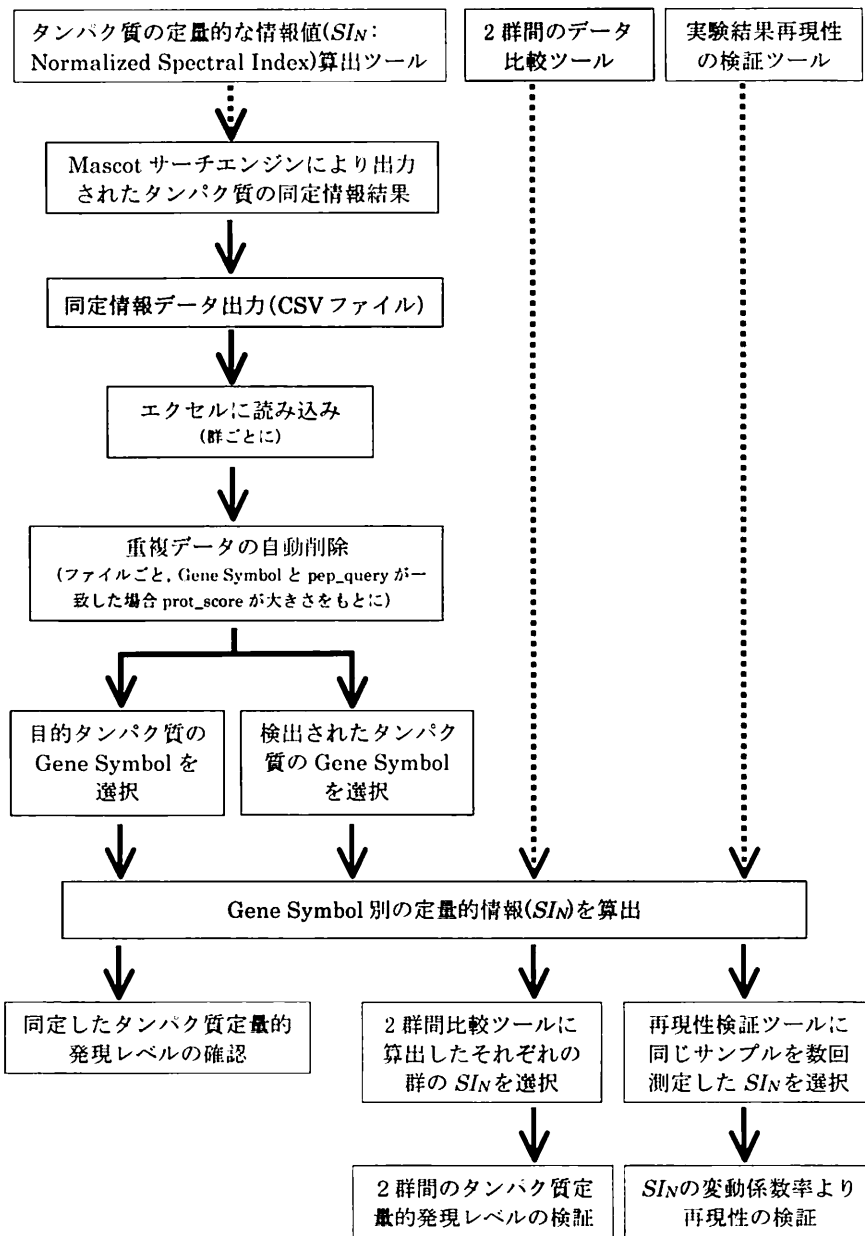


図1 データ処理の流れ

たすべてファイルで検出された Gene_Symbol が表示される。

SIN 算出機能はワークシート「Work_Gene_SIN」にワークシート「Work_Gene」で表示され

た Gene_Symbol を選択した後、計算を実行させることにより SIN が算出され、各タンパク質の同定情報で比較ができるよう一覧表にした。

SIN 出力機能では、エクセル関数の COUNTIFS,

SUMIFS, SUMIF, INDIRECT を組み合わせて算出できるようにした。また、データ処理量が多いのでエクセルの計算方法の設定は通常は自動であるが手動に変更する必要がある。

2 群間のデータ比較ツールでは、エクセル関数の IF, ISERROR, AVERAGE, STDEVP, TTEST, RANK, COUNTIF, ROW を、実験結果再現性の検証ツールでは IF, ISERROR, STDEVP, COUNTIF, COUNT を組み合わせて算出できるようにした。

結 果

1. タンパク質の定量的な情報 (SI_N) 算出ツールの結果

腎皮質由来のペプチド標本と、腎糸球体由来のペプチド標本を 18 回測定した同定結果からタンパク質の SI_N の算出を行った。

pep_score が 20 以上のときに検出されたタンパク質の数は 527 種類あった。そのうち腎皮質由来のタンパク質は 392 種類、糸球体由来のタンパク質は 205 種類で、腎皮質と糸球体ともにみられたタンパク質は 70 種類であった。

pep_score が 0 以上のときに検出されたタンパク質の数は 573 種類あった。そのうち腎皮質由来のタンパク質は 425 種類、糸球体由来のタンパク質は 218 種類で、腎皮質と糸球体ともにみられたタンパク質は 70 種類であった。

18 回測定した同定情報から SI_N の算出は短時間で行うことができた。また、pep_score の条件の変更による算出も容易にできた。

同定された 573 種類のタンパク質の中から発現レベルが大きいタンパク質 15 種類の SI_N を表 2 に示した。表からは腎皮質と糸球体ともに発現しているタンパク質、腎皮質または糸球体に発現しているタンパク質が見てとれる。

表 2 腎皮質と糸球体で同定された発現レベルが大きいタンパク質の SI_N

pep_score	Gene Symbol	腎皮質_1	腎皮質_2	腎皮質_3	腎皮質_4	腎皮質_5	腎皮質_6	糸球体_1	糸球体_2	糸球体_3
20 以上	HBB	11.049	9.932	12.969	5.978	9.174	13.451	4.658	4.614	4.213
	HBD	0.000	10.333	13.878	3.987	7.568	0.000	4.173	4.481	3.836
	ALB	4.697	5.916	3.549	6.091	4.911	4.792	0.025	0.029	0.050
	ACTB	0.331	0.406	0.312	0.380	0.536	0.323	3.296	3.873	3.380
	ACTA2	0.000	0.000	0.000	0.193	0.000	0.000	2.424	2.969	2.175
	ACTG2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	4.169	3.009
	HBA2;HBA1	2.390	2.400	2.589	2.773	2.233	2.886	2.185	2.086	3.170
	ACTG1	0.000	0.000	0.000	0.000	0.504	0.000	3.194	0.000	3.255
	ACTA1	0.079	0.000	0.255	0.229	0.131	0.211	2.653	3.154	0.000
	HBG1	6.657	4.877	0.000	0.000	0.000	0.000	0.000	0.000	1.001
	ACTBL2	0.000	0.152	0.117	0.000	0.000	0.000	1.674	1.701	0.989
	ACTC1	0.000	0.200	0.255	0.000	0.000	0.211	0.000	0.000	2.315
	GSTA2	0.225	0.381	0.292	0.368	0.581	0.696	0.000	0.000	0.000
	VIM	0.045	0.046	0.000	0.029	0.044	0.000	0.730	0.703	0.795
	GSTA1	0.225	0.381	0.292	0.368	0.581	0.696	0.000	0.000	0.000
0 以上	HBB	9.170	8.674	11.028	4.644	7.063	10.872	4.278	4.107	3.678
	HBD	0.000	8.912	11.717	3.230	5.770	0.000	4.006	3.785	3.558
	ALB	3.920	5.434	3.136	5.035	3.793	4.059	0.022	0.039	0.042
	ACTB	0.271	0.389	0.272	0.289	0.467	0.299	2.984	3.293	2.904
	ACTA2	0.000	0.000	0.000	0.301	0.000	0.000	2.555	2.867	2.263
	ACTG2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3.976	2.853
	HBA2;HBA1	1.957	2.049	2.159	2.433	2.021	2.327	1.923	1.884	2.742
	ACTG1	0.000	0.000	0.000	0.000	0.486	0.000	2.902	0.000	2.812
	ACTA1	0.072	0.000	0.244	0.307	0.470	0.309	2.587	2.882	0.000
	HBG1	5.537	4.226	0.000	0.000	0.000	0.000	0.000	0.000	0.845
	ACTBL2	0.000	0.205	0.112	0.000	0.000	0.000	1.572	1.511	0.944
	ACTC1	0.000	0.317	0.284	0.000	0.000	0.309	0.000	0.000	2.253
	GSTA2	0.542	0.392	0.850	0.740	0.965	0.787	0.000	0.000	0.000
	VIM	0.037	0.039	0.000	0.045	0.080	0.000	0.668	0.683	0.680
	GSTA1	0.398	0.392	0.428	0.777	0.690	0.726	0.000	0.000	0.000

表3 糸球体に特異的,あるいは比較的多く存在していることが示されている
4タンパク質を測定条件 MS target parameter を変更して3回測定したときの SI_N

pep score	Gene Symbol	糸球体 732		糸球体 1032		t検定
		平均	標準偏差	平均	標準偏差	
20 以上	ACTN4	0.5631	0.0514	0.4417	0.0451	0.0380
	NPHS2	0.0089	0.0093	0.0036	0.0062	0.4611
	PODXL	0.0141	0.0152	0.0039	0.0034	0.3658
	VIM	0.7426	0.0471	0.6816	0.0628	0.2550
0 以上	ACTN4	0.5510	0.0709	0.4219	0.0402	0.0670
	NPHS2	0.0150	0.0130	0.0057	0.0099	0.3844
	PODXL	0.0208	0.0126	0.0051	0.0055	0.1515
	VIM	0.6770	0.0079	0.6098	0.0342	0.0700

注: 732 MS target parameter を 732 m/z で測定したデータ

1032A MS target parameter を 1032 m/z で測定したデータ

表4-1 腎皮質と糸球体タンパク質の2群間の比較腎皮質に優位に検出されたタンパク質
腎皮質,糸球体ともにみられたタンパク質(51種類)の中で腎皮質に有意に
みられた8種類の SI_N の平均と標準偏差

Gene Symbol	腎皮質		糸球体		t検定
	平均	標準偏差	平均	標準偏差	
ATP5A1	0.2437	0.0397	0.0073	0.0061	0.0000
ATP5B	0.3932	0.0632	0.0364	0.0120	0.0000
ALB	4.9926	0.8442	0.0346	0.0109	0.0000
SLC25A5	0.2071	0.0522	0.0043	0.0060	0.0003
HBB	10.4254	2.5035	4.4949	0.2003	0.0030
MDH2	0.1125	0.0583	0.0053	0.0075	0.0086
GAPDH	0.2016	0.0916	0.0554	0.0141	0.0150
IGHV4-31;IGHG1	0.0805	0.0443	0.0173	0.0069	0.0232

2. 作成したツールの応用

1) 定量値の再現性の検討

糸球体の同一ペプチドサンプルを MS target parameter: 732 m/z と 1,032 m/z の二つの条件で各々3回測定し,糸球体に特異的,あるいは比較的多く存在していることが示されている4タンパ

ク質の SI_N を算出し再現性,変動定数を比較した。

573種類の中から糸球体に特異的,あるいは比較的多く存在していることが示されている糸球体上皮細胞のみみられるポドカリキシン (podxl), ビメンチン (VIM) とスリット膜関連分子のタンパク質であるポドシン (NPHS2), α -アクチニン4

表 4-2 腎皮質と糸球体タンパク質の 2 群間の比較腎皮質に優位に検出されたタンパク質
腎皮質にみられ、糸球体にみられなかったタンパク質 (243 種類) の中で
腎皮質に有意にみられた 47 種類の SI_{α} の平均と標準偏差

Gene Symbol	腎皮質		t検定
	平均	標準偏差	
DCXR	0.0669	0.0131	0.0001
ETFA	0.0913	0.0202	0.0002
HSPD1	0.5987	0.1488	0.0003
PRDX3	0.1107	0.0282	0.0003
CRYAB	0.1447	0.0376	0.0004
LDHB	0.2062	0.0543	0.0004
TUBB2C	0.1467	0.0410	0.0005
PCK2	0.1219	0.0398	0.0010
ACADSB	0.0250	0.0085	0.0012
ALDOB	0.2974	0.1019	0.0013
GATM	0.1861	0.0678	0.0017
KRT1	0.1227	0.0456	0.0018
BHMT	0.1904	0.0723	0.0020
GSTA1	0.4238	0.1636	0.0022
GSTA2	0.4238	0.1636	0.0022
TUBA1C	0.1017	0.0425	0.0031
FBP1	0.0461	0.0202	0.0037
HIBCH	0.0656	0.0300	0.0045
ALDH4A1	0.0669	0.0316	0.0051
ACY1;ABHD14A	0.0391	0.0186	0.0054
PRDX1	0.0962	0.0459	0.0054
PEBP1	0.1167	0.0575	0.0062
AK2	0.0782	0.0430	0.0097
HADH	0.0423	0.0241	0.0111
HSPE1	0.1575	0.0897	0.0111
GSTA3	0.3739	0.2151	0.0116
DPEP1	0.0607	0.0362	0.0134
ENO1	0.1308	0.0792	0.0141
ASS1	0.0940	0.0571	0.0143
ANPEP	0.0170	0.0104	0.0149
ECHS1	0.0501	0.0311	0.0156
ACSM2A	0.0741	0.0487	0.0192
EEF1A1	0.0705	0.0470	0.0202
ALDH2	0.1545	0.1044	0.0212
SOD2	0.1001	0.0688	0.0226
ACAT1	0.1272	0.0898	0.0249
TUBB4	0.0853	0.0610	0.0261
TF	0.0431	0.0313	0.0276
RPS27A	0.0412	0.0302	0.0282
ABAT	0.0897	0.0670	0.0304
RAP1B	0.0379	0.0290	0.0330
VDAC3	0.0196	0.0152	0.0345
ALDH1B1	0.0252	0.0197	0.0356
ALDH7A1	0.0306	0.0240	0.0357
DDR1	0.0368	0.0297	0.0393
LOC728638	0.0739	0.0600	0.0400
KRT8	0.0656	0.0542	0.0423

(ACTN4)の4タンパク質の SI_N の平均と標準偏差をもとめt検定を行った結果を表3に示した。

これらの4タンパク質で測定条件をMS target parameter: 732 m/zと1,032 m/zで測定した群で, pep_scoreを20以上では群間にACTN4は有意差($p = 0.038$)があり, 他の3タンパク質は有意差がみられなかった。pep_scoreを0以上では群間に有意差がみられなかった。糸球体に特異的, あるいは比較的多く存在しているタンパク質の発現は SI_N の算出法で求めた値ではサンプル間でバラツキが少ない傾向にあることを示された。

2) 2群間のプロテオームデータ比較

算出された SI_N を選択することにより2群間の平均値のデータ比較の検証—有効性の検証。

2群間の比較を行った結果を表4-1, 4-2, 4-3, 4-4に示した。腎皮質と糸球体ともに検出されたタンパク質は51種類, 腎皮質のみ検出されたタンパク質は243種類, 糸球体のみ検出されたタンパク質は87種類であった。さらにタンパク

質の SI_N の平均と標準偏差をもとめt検定を行ったところ, 腎皮質と糸球体ともに検出されたタンパク質の中で腎皮質に優位差のあったものが8種類, 糸球体に優位差のあったものが13種類, 腎皮質のみ検出されたタンパク質の中で優位差のあったものが47種類, 糸球体のみ検出されたタンパク質の中で優位差のあったものが10種類あった。

3) 実験結果の再現性の検証

同じサンプルを複数回測定した時に実験結果の再現性を検証するツール。算出された SI_N を選択することにより標準偏差と平均値の比(変動係数)を求め再現性の検証—有効性の検証。

表5に各ペプチドサンプルを3回測定して各々の SI_N から変動係数率(調査する値の変動係数の発現タンパク質数を全タンパク質発現数で割ったもの)を求めた結果を示した。変動係数率0.2以下は, 腎皮質でMS target parameter 732 m/zの時は9.0%(変動係数率: 0.2以下のタンパク質発現数20/タンパク質発現数222), 腎皮質でMS tar-

表4-3 腎皮質と糸球体タンパク質の2群間の比較糸球体に優位に検出されたタンパク質
腎皮質, 糸球体ともにみられたタンパク質(51種類)の中で糸球体に有意に
みられた13種類の SI_N の平均と標準偏差

Gene Symbol	腎皮質		糸球体		t検定
	平均	標準偏差	平均	標準偏差	
POTEE	0.0161	0.0175	0.4085	0.0183	0.0000
HSPG2	0.0013	0.0015	0.0067	0.0005	0.0003
VIM	0.0274	0.0202	0.7426	0.0385	0.0005
ACTN4	0.0280	0.0251	0.5631	0.0419	0.0010
LOC727848;POTEI	0.0080	0.0117	0.2984	0.0241	0.0017
ACTB	0.3814	0.0766	3.5163	0.2543	0.0025
ACTA2	0.0322	0.0721	2.5229	0.3316	0.0078
ACTN2	0.0217	0.0186	0.1577	0.0265	0.0084
ACTN3	0.0105	0.0083	0.0512	0.0104	0.0148
ACTN1	0.0226	0.0286	0.2888	0.0676	0.0243
ACTBL2	0.0449	0.0643	1.4546	0.3294	0.0248
MYH9	0.0003	0.0008	0.0245	0.0062	0.0306
COL6A1	0.0004	0.0008	0.0138	0.0042	0.0429

表 4-4 腎皮質と糸球体タンパク質の 2 群間の比較糸球体に優位に検出されたタンパク質糸球体にみられ、腎皮質にみられなかったタンパク質 (87 種類) の中で糸球体に有意にみられた 10 種類の SI_N の平均と標準偏差

Gene Symbol	糸球体		t検定
	平均	標準偏差	
LAMC1	0.0088	0.0006	0.0026
AGRN	0.0048	0.0006	0.0067
EHD3	0.0070	0.0011	0.0124
FN1	0.0079	0.0014	0.0153
FLNA	0.0086	0.0017	0.0202
LMNA	0.0516	0.0138	0.0338
LAMB2	0.0207	0.0057	0.0363
LAMA5	0.0089	0.0025	0.0368
NID1	0.0158	0.0049	0.0446
MYO1C	0.0039	0.0012	0.0469

表 5 3 回測定での変動係数 (CV) 率 (再現性の検証)

CV	腎皮質		糸球体		糸球体+フケ		
	サンプル 1	サンプル 2	サンプル 1	サンプル 2	添加サンプル 1	添加サンプル 2	添加サンプル 3
0.2 以下	9.0%	8.1%	12.3%	11.1%	0.5%	2.5%	0.8%
0.5 以下	20.7%	21.3%	28.3%	23.5%	4.3%	5.8%	5.8%
1.0 以下	48.6%	45.7%	55.8%	45.8%	25.5%	34.7%	28.9%

変動係数 (CV : Coefficient of Variation) = 標準偏差/平均値

get parameter 1,032 m/z の時は 8.1 % (18/221), 糸球体で MS target parameter 732 m/z の時は 12.3 % (17/138), 糸球体で MS target parameter 1,032 m/z の時は 11.1 % (17/153), フケ添加サンプルの 3 回測定結果は 0.5 % (1/184), 2.5 % (3/121), 0.8 % (1/121) であった。

変動係数率 0.2 以下が大きいほど再現性のよいデータであることが示されている。腎皮質と糸球体の数値とフケ添加サンプルの数値には大きな違いがあり、フケ添加サンプルの再現性が悪いことを示している。また、腎皮質と糸球体のサンプルで MS target parameter を変えて測定した結果も変動係数率はデータ間で大きな差は見られなかった。

考 察

本研究は各種腎疾患のプロテオーム解析のために用いられている MASCOT サーチエンジンで同定したタンパク質情報をエクセルで読み込み、検索したタンパク質の量的情報である SI_N を容易に算出できるツールを作成したことである。かつその有効性について検証したことである。

新潟大学医歯学総合研究科構造病理学研究室では今まで各種腎疾患のプロテオミクス解析を行い、それぞれ同定されたタンパク質情報を処理して、質的及び量的な違いを検討してきた。しかし同定されるタンパク質の数と関連情報が大容量となり、多数の標本から目的とするタンパク質同定

情報を検索することは困難になっていた。

この問題の解決のために、我々は腎疾患の SI_N を自動的に検出するツールを開発した。作成したツールの主な特徴は以下のようなものである。約600種類のタンパク質の分析データであったが容易に比較検討、データ処理を行うことができた。約2,000種類の大容量のタンパク質分析データでも SI_N を算出することができるようになっている。 SI_N の算出が容易にでき、 pep_score の条件による違いも算出できた。

pep_score の条件を変えられることで検出されたタンパク質の信頼性を容易に確認することができると考える。特異的なタンパク質の信頼性の確認に有用であることを示した。

また腎臓組織以外の成分のタンパク質の含有量も調べることができた。このことよりサンプルの採取の誤りによる検出結果の誤差の確認も調べることが可能であることを示した。

糸球体に特異的、あるいは比較的多く存在していることが示されているポドカリキシン (podxl)、ビメンチン (VIM) とポドシン (NPHS2)、 α -アクチニン4 (ACTN4) の4タンパク質の SI_N の結果からの発現状況の検証をおこなうことができた。

2群間の比較は正常ヒト腎組織の腎皮質と糸球体のペプチドサンプルで行った。2群間のタンパク質の発現状況の違いを容易に確認することができた。正常組織と疾患の組織や組織の部位別によるタンパク質の発現状況の比較検証に役立つことを示した。

実験結果再現性の検証ツールでは変動係数率0.2以下の値を求めることで再現性の検証を行えることを示した。また、測定条件 MS target parameter を変えた結果では大きな差が見られなかったが、測定条件を変更したときや実験の方法を改良したときの同定情報結果の比較検討に利用できる。

以上まとめると、このツールを利用することで実験の信頼性や質量分析系の安定性や信頼性を調べられることが明らかになった。また調べることのできるタンパク質データは2,000種類まで可能

であり、かつ容易に算出することができたことである。

次の課題として、疾患のある糸球体組織のプロテオーム分析を実施し、正常組織と比較することにより、疾患別のタンパク質の違いを明らかにすることである。さらに腎臓以外の組織においても本ツールが有用か否かの検討も行われるべきである。

結 論

このツールにより非標識サンプルの質量分析計の測定結果を MASCOT サーチエンジンで同定した情報の大容量データからタンパク質の種類と定量情報が効率的に取得でき、腎臓各部位のプロテオームのデータ比較を容易に行うことができた。質量分析計の安定性、再現性についても検証を行うことができた。このツールは通常よく使用されている表計算ソフトエクセルで作成してあるので、個人のパソコンでも質量分析計からのタンパク質同定情報結果からの定量的情報の発現レベルの比較検証を容易にできることを示した。

謝 辞

稿を終えるに際し、御指導賜りました新潟大学大学院医歯学総合研究科生体機能調節医学専攻構造病理学分野 山本 格教授、許 波先生、並びにサンプル作製の技術指導を頂きました構造病理学教室の皆様へ深謝いたします。

参 考 文 献

- 1) 山本 格：Human Proteome Organization (HUPO) のヒト腎臓・尿プロテオームプロジェクト。新潟医学会雑誌。124: 55-64, 2010。
- 2) 平野 久：プロテオーム解析概論。ぶんせき。7: 348-353, 2005。
- 3) 山本 格：腎糸球体タンパク質のデータベース。病理と臨床。28: 524-529, 2010。
- 4) 山本 格：腎臓病学におけるプロテオミクス。日本腎臓学会誌。52: 457-460, 2010。

- 5) 吉田 豊：腎組織プロテオームとそのデータベース. 日本腎臓学会誌. 52: 461 - 464, 2010.
- 6) 山本 格：腎糸球体タンパク質のデータベース. 病理と臨床. 28: 524 - 529, 2010.
- 7) Bo Xu: Usage of electrostatic eliminator reduces human keratin contamination significantly in gel-based proteomics analysis. Journal of Proteomics. 74: 1022 - 1029, 2011.
- 8) 許 波：プロテオミクスの試料調製. 日本腎臓学会誌. 52: 495 - 498, 2010.
- 9) Griffin NM, Yul J, Long F, Oh P, Shore S, Li Y, Koziol JA and Schnitzer JE: Label - free, normalized quantification of complex mass spectrometry data for proteomics analysis. Nature Biotechnol. 28: 83 - 89, 2010.
- 10) MATRIX SCIENCE: 「Summary Reports for MS/MS」 <http://www.matrixscience.com/help/msms_summaries_help.html> (2014/10/08 アクセス)
- 11) 川上隆雄：バイオデータベースとウェブツールの手とり足とり活用法改訂第2版. 東京：羊土社, 188 - 199, 2008.

(平成 27 年 1 月 19 日受付)