

表4 日本人由来hMSCsにおけるA3B遺伝子型の解析

A3B genotype		A3B genotype		A3B genotype	
Yub621	Ins/Del	PL505	Ins/Ins	UC701	Ins/Ins
Yub622	Ins/Ins	PL507	Ins/Ins	UC702	Ins/Del
Yub623	Ins/Ins	PL508	Ins/Ins	UC704	Ins/Del
Yub625	Ins/Del	PL518	Ins/Ins		
Yub631	Ins/Del	PL512	Ins/Ins		
Yub632	Ins/Del	PL514	Ins/Ins		
Yub633	Ins/Ins	PL521	Ins/Del		
Yub634	Ins/Del	PL523	Ins/Ins		
Yub635	Ins/Ins	PL532	Del/Del		
Yub636	Ins/Ins				
Yub637b	Del/Del				
Yub642p	Ins/Ins				
Yub 10F	Ins/Del				

Genotype (No. of human)

	Ins/Ins	Ins/Del	Del/Del
Yub	6	6	1
PL	7	1	1
UC	1	2	0
total	14	9	2

Allele frequency (%)

Ins	Del
74	26

厚生労働科学研究費補助金（医薬品・医療機器等レギュラトリーサイエンス総合研究事業）  
「再生医療製品の品質・安全性評価のための新たな指標に関する研究」  
分担研究報告書

次世代シーケンサーを用いた細胞の遺伝的安定性評価指標の開発

研究分担者：鈴木 孝昌 国立医薬品食品衛生研究所 遺伝子細胞医薬部・第3室・室長

## 研究要旨

再生医療の実用化に向けた流れが加速する中、治療に用いる細胞の品質、特に安全性が重要な懸念となっており、レギュラトリーサイエンス的立場からその評価に必要な技術や情報を提供することが我々の責務となっている。こうした要望を受け、我々はこれまでに主に細胞の遺伝的安定性の評価という観点から、近年進歩の著しい次世代型シーケンサーの活用に関して検討を行ってきた。まず、ホールゲノムシーケンシングに関しては、バンク化されるマスター細胞の品質保証としての利用が考えられているが、そのデータの信頼性評価と解析が課題となっている。我々は、既に遺伝子増幅や染色体のリアレンジメントが起きていることが確認されている細胞をモデルとして用いて、次世代シーケンサーを用いたホールゲノム解析を進めた。その結果、既に SNP の検出を含めて、シーケンス解析法としての有用性ととも、コピー数変化の検出法としても有用であることを示すとともに、今回はさらに詳細な検討を加え、遺伝子レベルでの欠失および増幅の検出に関する知見を得た。さらに、遺伝的不安定性のモデルとして、染色体の安定性に寄与する遺伝子（BLM）を破壊した細胞を用い、次世代シーケンサーによる変異の検出に関する検討を開始した。

一方で、我々は以前より細胞の表現形質としてのタンパク質発現プロファイルに着目し、網羅的プロテオーム解析による細胞の品質評価と標準化に関する検討を行ってきたが、今回新たに、LC-MS/MS を用いたプロテオミクスデータの可視化により各種細胞より得られるタンパク発現プロファイルに関する質的および量的情報をリファレンスデータとして提供するためのツールとして ProteomeMap というソフトウェアを開発した。

## 研究協力者

本間 正充

国立医薬品食品衛生研究所

変異遺伝部 部長

スレッシュ ティルパッティ

国立医薬品食品衛生研究所

遺伝子細胞医薬部

山田 雅巳

国立医薬品食品衛生研究所

変異遺伝部 室長

## A. 研究目的

ES 細胞、ヒト人工多能性幹細胞 (iPS) および体性幹細胞等を利用した再生医療、細胞治療法の開発、および、実用化に向けた流れが各方面で加速されつつある。画期的な治療法の開発に期待が高まる反面、その安全性、主に腫瘍原性が問題となっており、これを如何に克服して、実用化へのハードルをクリアするかは、我々に課せられた大きな課題である。

細胞治療において最も懸念される問題は、移植後の細胞の癌化であり、事前にその安全性を担保しうる試験法の開発が求められている。癌化形質を持った細胞の存在を高感度に検出する手法の開発も進められているが、より一般的に細胞の遺伝的安定性を調べる試験法の開発も重要である。従来型の試験法として、染色体異常試験や SNP アレイ解析、アレイ CGH などの方法が存在するが、より網羅性の高い、高感度な試験法の開発が望まれている。そのため新たな手法として、近年発展の著しい次世代型のシーケンサーを用いたゲノム解析法の応用が注目されている。従来はゲノム解析には多くの費用と多大な努力が必要であったため、日常的に用いる評価方法としては非現実的であったが、ここ数年のハード面を中心とした目覚ましい技術革新の結果、かつて言われていた“1000 ドルゲノム”、即ち全ゲノムを 1000 ドルで読むというのが既に現実的なものとなっている。このレベルまでコストダウンが進めば、標準的試験法として、細胞治療医薬品の品質評価に用いることも可能になると考えられるため、こうした応用を意識した次世代シーケンサーによるシーケンス解析データの評価とその応用に向けた基礎的知見を得るための検討を開始した。ゲノムシーケンスに関するアプローチとしては、ホールゲノムシーケンスに加えて、より効率化した方法としての遺伝子のエクソン部分のみを読むエクソンシーケ

ンスの手法も注目されている。我々は、これまでの検討により、すでに変異部位がある程度確かめられている HL60 細胞およびヒト骨髄由来間葉系幹細胞 (hMSC) 細胞をモデルとして、これらの遺伝子異常をどの程度次世代シーケンサーによるシーケンス解析で検出できるかについて検討を行った。そして、あわせて次世代シーケンサーから得られる配列情報の信頼性およびその評価基準の設定に向けた基礎検討を行った。

一方で、細胞のタンパク質発現プロファイルは、細胞の遺伝子情報の反映のみならず、環境因子により変化するエピゲノムや翻訳後修飾によるシグナル伝達を反映する細胞の動的な表現形質の変化の指標として捉えることができる。細胞の動的な品質評価としてのタンパク質発現プロファイル情報の利用に向け、我々は LC-MS/MS によるショットガンプロテオミクスにより得られた網羅的ペプチド発現情報を可視化し、比較可能な形としてリファレンスデータを提供するための手段としてのデータビューワーとなるソフトウェアの開発に関する取り組みを開始した。まず、スタンドアローンのソフトウェアとしてのデータ解析、可視化ツールを開発し、将来的にはウェブ上でリファレンスデータを提供するとともに、ユーザーからのデータをアップロードおよび比較を可能とすることを目指す。

## B. 研究方法

### 1 使用した細胞株

#### ・ヒト白血病細胞株 HL60 (RG)

国立医薬品食品衛生研究所、細胞バンク (当時) より入手したヒト前骨髄球系白血病細胞株 HL60 細胞およびその増殖性変異株である HL60-RG 株を使用した。この細胞は既に染色体解析および CGH 法にて解析が行われ、myc 遺伝子の増幅を Double Minute (DM) 染

色体および Homologous Staining Region (HSR)として持つことが知られている。

HL60 細胞は、10% 牛胎児血清添加 RPMI1640 培地にて培養をした。

#### ・ヒト骨髄由来間葉系幹細胞株 (hMSC)

Cambrex 社より入手した正常ヒト骨髄由来間葉系幹細胞株 (hMSC) のうち、以前の検討において、異常が認められたロット 4 F1560 と同一ロットを使用した。hMSC は、間葉系幹細胞培養用基本培地培地 Mesenchymal Stem Cell Basal Medium (MSCBM) に間葉系幹細胞添加因子セット (MSCGM SingleQuots、TAKARA) および 10%FBS を添加し培養を行い、70-80%コンフルエントの状態に継代を続けた。18-19 世代再培養し、凍結保存した細胞を使用した。

#### ・ヒトリンパ芽球細胞株 TK6

国立医薬品食品衛生研究所変異遺伝部より入手。正常 p53 遺伝子を持ち、EB ウイルスにて不死化された細胞株で、thymidine kinase (TK) 遺伝子の変異をヘテロに持つことで、この遺伝子を指標とした突然変異の検出が可能となっている。細胞は 10% (v/v)馬血清を含む RPMI-1640 培地 (Nacalai Tesque)にて培養。

## 2. ゲノム DNA の抽出

次世代シーケンサー解析用のサンプル調整を行うため、DNA Extractor WB キット (和光純薬工業)を用いて DNA 抽出を行った。本キットは、フェノールやクロロホルムといった有毒な有機溶媒を用いず、ヨウ化ナトリウムとイソプロパノールにて、細胞より DNA のみを抽出する簡便な方法である。また、核分離を行った後 DNA の抽出を行うため、比較的純度の高い DNA を得ることができる。以下の操作にしたがって、細胞よりゲノム DNA を抽出

した。

(細胞の溶解と核分離)

1) 凍結保存細胞  $1-2 \times 10^6$  個に溶解液を 0.5ml 加えて、チューブを数回転倒混和した。

2) 遠心分離 ( $10K \times g$ ,  $4^\circ C$ , 20 秒間)した後、上清を除いた。

3) 再び溶解液を 1ml 加えて、30 秒間激しく攪拌し、遠心分離 ( $10K \times g$ ,  $4^\circ C$ , 20 秒間)した後、上清を除いた。

4) ステップ 3) をもう一度繰り返した。

(核膜の破壊とタンパク変性)

5) 酵素反応液  $200 \mu l$  とタンパク質分解酵素  $10 \mu l$  (使用前に酵素 10mg を 0.6ml の滅菌蒸留水に溶解)を加えて混合した。

6)  $37^\circ C$ で 1 時間反応させた。(途中 2~3 回軽く振り混ぜた)

7) よう化ナトリウム溶液を  $300 \mu l$  加えて混合した。

(DNA の精製)

8) イソプロパノールを 0.5ml 加えて、白い綿状の DNA が完全に見えてくるまで混合した。

9) 遠心分離 ( $10K \times g$ , 室温, 10 分間)した後、上清をゆっくり除き、容器をろ紙の上に逆さに置き、器壁に残った溶液を十分に除いた。

10) 洗浄液 A を 1ml 加えて混合し、遠心分離 ( $10K \times g$ , 室温, 5 分間)した後、上清を除いた。

11) 洗浄液 B を 1ml 加えて混合し、遠心分離 ( $10K \times g$ , 室温, 5 分間)した後、上清を除いた。

12) DNA 沈殿を風乾し、TE バッファーに溶解させた。

## 3. 次世代シーケンサーを用いたホールゲノムシーケンス解析

シーケンス用のサンプルは、Illumina TruSeq DNA sample preparation guide に従

い、1  $\mu$ g のゲノム DNA を covaris system にて断片化し 300-400bp のインサートサイズを持つライブラリーを作成した。3'または 5'エンドにオーバーハングを持つ二本鎖 DNA フラグメントを End Repair Mix にて blunt エンドに変換して、3'末端に A を一塩基追加し、T を 3'末端に一塩基追加したアダプター配列とライゲーションした。ライゲーションプロダクトのうち約 300-400bp のインサートサイズを持つものを選択し、次のクラスター生成に用いた。こうしてエンリッチした DNA ライブラリーを用い、アダプター配列に相補的プライマーによる PCR にて増幅しシーケンス解析用サンプルとした。Illumina HiSeq2000 シーケンサーにて、Sequencing-by-Synthesis 法にて、数百万のクラスターを持つフローセル内での独自の架橋増幅反応と一塩基伸長ごとのイメージングにより、各クラスターごとの配列情報を読み取った。

読み取ったデータを、BWA ソフトウェアにてヒトファレンスゲノム UCSC hg19 に対してマッピングした。そして、SNP 等のリファレンスゲノムに対する変化を SAMTOOLS ソフトウェアを用いて解析した。

なお、シーケンス解析に関しては、株式会社アプロサイエンスに委託した。

## 5. 遺伝的不安定性モデルとしての BLM 欠損 TK6 細胞の利用

ブルーム症候群の原因遺伝子である BLM を、相同的組み換えを利用して薬剤耐性遺伝子と置き換えることにより破壊した TK6 BLM-TSCER2 株を国立医薬品食品衛生研究所、変異遺伝部より入手し、ホールゲノムシーケンス解析データを親株の TK6 と比較することにより、遺伝的不安定性の検出の可能性を検討した。シーケンス解析に関しては、株式

会社アプロサイエンスに委託した。

また、発現タンパク質の変化についても、LC-MS/MS を用いたショットガンプロテオミクスの手法により比較した。

## 6. タンパク質プロファイル情報提供のための可視化ツールとしての Proteome Map ソフトウェアの開発

LC-MS/MS を用いたショットガンプロテオミクス解析より得られた生データを用い、検出されたペプチドをイメージデータに変換して 2 次元マップ上に記載するとともに、MS/MS データやタンパク同定結果に関する情報をマップ上にて提供するためのソフトウェアとして「ProteomeMap」の開発を行った。ソフトウェアプログラムのプログラミングに関しては、インド Rushmore 社に委託した。

### (倫理面への配慮)

Lonza 社の hMSC はヒト由来細胞であるが、提供者の同意を取り適切に細胞を採取していることが確認されている。このため国立医薬品食品衛生研究所研究倫理審査委員会規程にある倫理審査該当品目ではなかった。

## C. 研究結果

### 1. ホールゲノムシーケンス解析による細胞の品質評価

遺伝子変異や欠失と増幅を含めたコピー数変化、さらには遺伝子転座等の染色体異常といったゲノムの異常を、どの程度シーケンス解析により検出できるかを確かめる目的で、昨年度に報告した HL60-RG 細胞に続き、ヒト骨髄由来間葉系幹細胞 (hMSC) に関しても、ホールゲノムシーケンス解析を行った。

イルミナ HiSeq2000 シーケンサーにて

101bp リード4回のランにて読み取られたリード数は約 18.5 億で、塩基数としては 1871 億 bp に達した。ヒトのゲノムは約 30 億塩基対であるため、平均重複度は約 60 となった。重複度 (sequence depth) の分布を図 1 に示す。

得られたシーケンスデータをヒトリファレンスゲノム hg19 に対してマッピングした結果、最終的にマップ可能であったリードの割合は全体の 91.7% と良好であり、リファレンスゲノム上の 99.3% をカバーできた。ベースコールにある程度の信頼度を持つ重複度を 10 とした場合のカバー率は 99.1% であった。よって十分な情報が得られなかったのはゲノム上わずか 1% 未満であり、遺伝子上のエクソ配列に関してはほぼ網羅できていると考えられる。

**SAMTOOLS** ソフトウェアを用いてリファレンスゲノムからの変異を検出した結果、SNP の総数として 4,475,876 個、Indel (インサージョンおよびデリージョン) の総数 687,240 個、合計約 500 万箇所が抽出された。

今回解析に用いた hMSC 細胞は、クロソナルな染色体異常を持つことが既に確認されている細胞であり、今回得られたホールゲノムシーケンスデータの各染色体部位のリード数 (冗長度) を用いた擬似的比較ゲノムハイブリダイゼーション (CGH) 法による解析の結果、図 2 に示すように、既に得られている 7 番および 17 番染色体の部分的コピー数の増減が確認できた。コピー数変化に関しては、塩基配列レベルで高精細な情報が得られるため、今後さらに詳細なりアレンジメントの解析に応用できると期待できる。さらに、従来のアレイ CGH 法等によっても検出されていなかった微細なコピー数変化も検出され、その確認が必要となった。

次に、すでに得られた HL60 細胞のホール

ゲノムシーケンスデータについて、遺伝子レベルでの詳細なコピー数変化の検討を加えた。表 1 に、HL60 細胞の各染色体におけるコピー数変化の見られた領域と、そこに存在する遺伝子をリストアップしたものである。コピー数変化の解析は、10kb を単位として冗長度の平均を取ったデータを使用したため、コピー数変化の最小単位は 10kb となっているが、実際にはさらに細かい単位での検討も可能である。

全体としては、欠失に比べて増加の方が多く、領域としては短いものから長いものまで幅が見られた。一つの特徴としては、比較的近傍に増加または減少の同じ方向の変化がまとまって見られた点であり、代表的なものが 8 番染色体の *c-myc* 領域 (8q24) における複雑な増幅であり、合計 7 つの領域が約 16 コピーという類似した増幅度で変化していた。このうち一つの領域に関してはコピー数が 33 と他の約二倍であり、この領域を 2 回含んだ増幅単位が 1 ユニットとして増幅したことを示している。

欠失については主にコピー数は 1 と片側のアレルのロスを示したが、部分的にはホモ欠失の領域も認められた。注目すべきは 17 番染色体の欠失領域であり、表 2 に示したように、ヘテロ欠失として短腕全体 (17p) のロスが見られるが、さらにこのうちの二つの領域、それぞれ癌抑制遺伝子である *TP53* と *ARHGAP44* (Rho GTPase activating protein 44) を含む短い領域はホモに欠失していることがわかった。即ち染色体の FISH 解析により 7p のヘテロ欠失は簡単に観察されていたが、正常に見えた片側のアレルにも欠失領域が存在し、短いため染色体解析においては検出不可能であった変化が、シーケンス解析データにより検出できたことになる。

また増幅している領域には、*MAP2K3* や *CDC27* という細胞周期に関連した遺伝子が存在し、癌化との関わりが示唆された。17 番染

色体以外においても、コピー数変化の見られた領域には重要と思われる遺伝子が多く含まれており、HL60 細胞の癌化のヒストリーを反映している可能性が考えられる。

さらに、コピー数変化のあった領域について見てみると、レトロトランスポソンの挿入サイトを含めた繰り返し（相同）配列が特徴的に観察され、これらが染色体異常の生成に関与している可能性が示唆された。

## 2. 次世代シーケンサーを用いた細胞の遺伝的不安定性の検出

細胞の遺伝的不安定性を検出するためのモデル細胞として、国立医薬品食品衛生研究所変異遺伝部において開発された BLM 遺伝子ノックアウト細胞を使用した。BLM 遺伝子は染色体不安定性を示すブルーム症候群の原因遺伝子であり、DNA 二本鎖切断の修復酵素である DNA ヘリカーゼをコードしている。この遺伝子を破壊した TK6 細胞株は、親株に比べて高い染色体異常および突然変異誘発性を持つことが確かめられている。

この TK6 細胞 BLM 欠損株の遺伝子配列を、親株の TK6 細胞と比較して、突然変異およびコピー数変化を検出するため、ホールゲノムシーケンシング解析を行った。また同時にミトコンドリアのシーケンシング解析も行った。さらに、LC-MS/MS を用いたショットガンプロテオミクス解析による発現タンパク質プロファイルの比較を行った。得られたデータの解析結果については、次年度の報告書にて報告する。

## 3. タンパク質プロファイル情報提供のための可視化ツールとしての Proteome Map ソフトウェアの開発

我々はこれまでに、LC-MS/MS を用いたシ

ョットガンプロテオーム解析により得られたデータの可視化に関する検討を行ってきたが、その経験を生かして、各種細胞のタンパク発現プロファイルに関するリファレンス情報の提供と細胞間のデータ比較を可能とするためのソフトウェア「ProteomeMap」の開発に着手した。

通常質量分析装置から得られるデータは、膨大な量の数値データであり、このままではその全貌および詳細をつかみにくいことから、リテンションタイムと質量数(m/z)を各軸に取った 2 次元マップ上にイメージデータとして変換して可視化を行うことにした。この際、各ペプチドピークに対してタンデムマス (MS/MS) 測定が行われていた場合には、そのスペクトル情報が付随してくるが、これらも合わせて情報提供できるよう、クリックابلマップとして、対応するペプチドピークをマップ上でクリックした際に、ピーク情報がグラフとして表示される機能を加えた。また、MS/MS 測定がされたピークに対しては、MASCOT によるデータベース検索でのタンパク質同定結果の取り込みを行い、MASCOT 検索結果を表示させる機能も開発した。実際のプログラミングに関しては外部の専門家に委託し、修正と改良を加えつつプロトタイプソフトウェアが完成した。その内容を図 3 に示す。

現在のところ使用している質量分析器が Thermo Scientific 社のものなので、".raw"形式の生データを取り込めるよう設計した。ここでは、細胞より得られたデータの例として TK6 細胞のプロテオームデータを使用した。得られたペプチドピークの総数は 48,950 であり、これらを横軸にリテンションタイム、縦軸に質量数 (m/z) を取りプロットした。(2) 得られたペプチドピークの総数は 48950 であり、全てのピークを 2 次元マップ上にスポットした。この 5 万近いペプチドの

うち、実際に MS/MS 測定が成されたのは約 2 割の 10,000 個であり、そのうち MASCOT 検索にて同定されたペプチドの総数は 2805 と全体の約 6 パーセントであり、ここからタンパク質が同定されたのは 931 個であった。今後はさらに同定数を増やしていきたい。

2次元マップ上 MS/MS データを持つペプチドピークは青または赤の印がマークされるが、前者は MASCOT 検索にて同定結果が得られたピーク、後者は未同定のピークをあらわす。画面はズームイン機能を有し、それぞれのピークをクリックすることにより、MS/MS のスペクトルデータを表示させることができる。

本ソフトウェアは複数のサンプルのデータを取り込み、相互に比較することが可能であり、各ピークの濃度からおよその定量比較が可能であるが、今後より定量的な比較が可能となるよう改良を加えてゆきたい。

以上のように、現段階ではスタンドアローンソフトウェアとして、生データの取り込みと可視化表示が可能であることが動作確認できたが、今後各種幹細胞のプロテオーム解析から得られたデータに関して同様に 2次元マップ化を行うとともに、同定数を増やし、ウェブ上でリファレンスデータとして提供できるように改良を加えてゆきたい。将来的には、ウェブ上でのデータデポジットを可能にし、多種の細胞種やそれらの異なる状態におけるタンパク発現プロファイルに関するデータベースの構築を目標としている。

#### D. 考察

これまでの検討より、次世代シーケンサーを用いたホールゲノム解析により、遺伝的同一性の評価のみならず、CGH 法の代替としてゲノムのコピー数変化の検出に有用であることが

示された。特に、他の CGH 法に比べて塩基配列レベルでの高解像度な変化の検出が可能となる。今回癌細胞である HL60 細胞のホールゲノムシーケンスデータを詳しく解析することにより、かなり小さな領域でのコピー数変化まで検出できたが、特筆すべきはその多くの領域に細胞の増殖とかかわる遺伝子など重要だと考えられる遺伝子が存在していたことである。既知の癌遺伝子 *c-myc* の増幅に加え、神経芽細胞腫にて異常の見つかった *neuroblastoma breakpoint family member 8&9* (NBPF8&9) の増加、発がん物質処理により湯洗される染色体転座の共通した切断点となることが知られている *fragile histidine triad* (FHIT) 遺伝子座位の変化、*myeloid/lymphoid or mixed-lineage leukemia (MLL) family* の遺伝子である *lysine (K)-specific methyltransferase 2C* (KMT2C) の増幅、その過剰発現が急性骨髄性白血病と関連する *annexin A8 (ANXA8)* 遺伝子の増加、低酸素ストレスにより誘導され腫瘍との関連性が注目されている *hypoxia induced factor 1* (HIF1A) 遺伝子の欠失、細胞分裂に関与する *cell division cycle27 (CDC27)* の増加、TGFβ シグナルパスウェイに関与し、細胞の増殖分化とアポトーシスを制御する因子である *SMAD2* 遺伝子の増幅、遺伝子多型と癌との関連性が指摘されている *glutathione S-transferase T1 (GSTT1)* の欠失など、腫瘍との関連性が示唆される多くの遺伝子がコピー数変化した領域に見いだされた。さらに、17 番染色体の短腕(7p)は片側のアレルが完全にロスしているが、正常に残っていると思われるアレルにも微小な欠失領域があることが分かった。その一つは有名な癌抑制遺伝子である *TP53* 遺伝子であり、この遺伝子の完全欠失が HL60 細胞の癌化に大きく寄与していることを物語っている。7p のもう一つのホモ欠失領



域には Rho GTPase-activating protein 44 (ARHGAP44) 遺伝子が存在し、この遺伝子の機能と癌との関連性が注目される。

また染色体全体に目を向けると、同様にホモ欠失している短い領域があり、この中には TGF $\beta$  シグナルパスウェイに関与し、細胞の増殖分化やアポトーシスと関連する signal-regulatory protein beta 1 (SIRPB1) 遺伝子が含まれていた。また、大腸癌の予後との相関が報告されている Cytoplasmic poly(A) binding protein 4 (PABPC4) 遺伝子もホモ欠失しており、その機能との関連性が注目される。

以上のように、ホールゲノムシーケンス解析にて得られたコピー数変化領域を詳しく見てみることで、そこに存在する遺伝子が何らかの癌の発生との関連性を示唆しており、複数の遺伝子の異常が複雑に絡み合っただけで癌を発生させた、いわば細胞癌化のヒストリー(プロファイル) を物語っていると考えられる。各種の癌細胞に関してこうした癌化のプロファイルを蓄積していくことで、細胞癌化の共通したメカニズムおよび癌の個性に寄与する遺伝子を浮き彫りにできる可能性が示された。

コピー数変化の検討から得られたもう一つの有益な情報として、増減のおこっている領域には、 $\alpha$ -サテライトのような繰り返し配列、およびレトロトランスポゾンによる挿入配列が多く存在する点が見受けられた。特にレトロトランスポゾンの配列に関しては、CDC27 遺伝子のように、元の遺伝子の部位も同様にコピー数の変化が起きている例もあり、染色体上の相同配列が引き金となって、遺伝子および染色体領域のコピー数の増減が起きていることが示唆された。

細胞の発現するタンパク質のプロファイルは、遺伝子異常のプロファイルと同様にその品質を規定する重要な情報となりうる。遺伝子解析に関しては、ホールゲノムシーケンスとい

う強力な解析ツールが存在するが、タンパク質発現解析に関しては、すべての発現タンパクを網羅できる完璧な手法はまだ存在しない。これはタンパク発現のダイナミックレンジがかなり大きいことに由来するが、近年の質量分析装置の進歩により、高感度化が図られ、ペプチドレベルでは数万、タンパクレベルでは数千のタンパク質の発現情報が得られるようになった。タンパク発現は最終的な細胞の表現形質につながるものであり、細胞のおかれた環境や状態の変化を反映して動的に変化する指標となる。細胞の標準化および品質評価においては、この動的に変化するプロテオームの情報をリファレンスと比較することにより品質の評価が可能であり、その意味で比較可能な形として細胞のタンパク発現プロファイル情報を提供することは有益だと考えられる。

我々はこれまでに LC-MS/MS を用いたショットガンプロテオミクス解析を通じた質量分析データの取り扱いに関する経験を生かし、プロテオームの発現プロファイルを視覚的に分かりやすい形で情報提供し、リファレンス情報の提供と比較解析を実現するためのツールとして、ProteomeMap というオリジナルなソフトウェアの開発に着手した。そして、今年度は質量分析装置から得られた生データの取り込み、イメージデータへの変換、2次元マップとしてのデータ可視化、およびペプチドピークの認識と付随する MS/MS スペクトルおよびタンパク質同定結果の提供を同一マップ上で実現することができた。また、2次元マップ上でのペプチドの存在量はバンドの濃さとして表現されるため、量的変化も視覚的にとらえることが可能となった。

現在 ProteomeMap ソフトウェアはスタンダードアローンで、Thermo Scientific 社の質量分析装置 Q-Exactive より得られた raw 形式のデータに対して動作確認をしているが、今後は各種

細胞から得られたプロテオーム情報を Web 上にてオンラインで情報提供できるように開発を進めていきたい。オンラインで各種細胞のリファレンスデータとなる情報を提供できれば、同定結果に対する情報を MS/MS スペクトルとともに同時提供でき、各ユーザーが得たデータに対して有益なリファレンス情報を与えることが可能になる。さらに将来的には、各ユーザーからのデータのアップロードを可能とし、リファレンスデータベースの充実を図るとともに、オンラインでのリファレンスデータとの定量比較を可能とすることにより、さらに有益な研究ツールを提供することを目標とする。

## E. 結論

次世代シーケンサーによるホールゲノム解析のパフォーマンスは高く、通常の解析でゲノム上の 99 パーセント以上の領域について、信頼性のある配列情報が得られた。さらに、ゲノム上のコピー数変化領域に関しても高精細なデータが得られ、がん細胞株では、その領域に多くの増殖関連遺伝子等が存在していることが明らかになり、コピー数異常のプロファイルが、細胞癌化のプロセスおよび癌の個性の理解に役立つことが示唆された。さらに、コピー数変化の見られた領域には、繰り返し配列やレトロトランスポゾンによる挿入配列が散見され、これらがゲノムのリアレンジメントの引き金となったことが示唆された。

一方で、細胞のタンパク発現プロファイルを細胞の品質評価および標準化に役立てるため、プロテオーム解析によって得られた網羅的なペプチド発現情報を可視化し、わかりやすい形で情報提供するためのツールとして ProteomeMap ソフトウェアを開発した。このソフトウェアにより、LC-MS/MS を用いたショットガンプロテオミクス解析により得られ

た質量分析データを二次元マップ化し、ピーク強度、MS/MS スペクトル、タンパク質同定結果の情報を同時に提供するリファレンスデータの作成が可能となった。今後さらに Web 上での公開を目指したソフトウェアの開発を進めたい。

## F. 健康危険情報

なし

## G. 研究発表

### 1. 論文発表

1) 佐藤陽治, 堤秀樹, 澤田留美, 鈴木孝昌, 安田智 細胞・組織加工製品の開発環境整備に向けたレギュラトリーサイエンス研究. 国立医薬品食品衛生研究所報告, 131, 16-19. 2013

2) 斎藤嘉朗, 前川京子, 齊藤公亮, 佐藤陽治, 鈴木孝昌 タンパク質・内在性代謝物バイオマーカーを利用した医薬品開発の活性化にむけて. 国立医薬品食品衛生研究所報告, 131, 20-24. 2013

3) 中村里香, 酒井信夫, 藪島由二, 福井千恵, 鈴木孝昌, 中村亮介, 蜂須賀暁子, 安達玲子, 手島玲子 ショットガンプロテオミクスによる加水分解小麦とその原料であるグルテンに含まれるタンパク質の網羅的解析. 国立医薬品食品衛生研究所報告, 131, 50-57. 2013

4) T. Suzuki "Scientific Considerations Regarding Radiation Risk" JEMS Open Symposium 2012. Genes and Environment, 35, 57-62. 2013

5) T. Suzuki Unconscious Exposure to Radiation. Genes and Environment, 35, 63-68. 2013

6) 鈴木孝昌 「網羅的な発現をみる マイクロアレイ解析との比較を例に」 実験医学別冊 原理からよくわかるリアルタイム PCR 完全実験ガイド (羊土社) 111-121. 2013

### 2. 学会発表

1) Suresh T., Oshizawa T., Maekawa K., Saito Y., Sato Y., Suzuki T. Improvement of Rat Urinary Proteomics by a Differential Precipitation of Proteins. Human Proteome Organization 12th World Congress (2013.9) (横浜)

2) Suzuki T., Suresh T., Oshizawa T., Maekawa K., Saito Y., Sato Y. Basic factors that influence the rat urinary proteome. 第13回国際毒性学会 (ソウル)

3) 鈴木孝昌, Suresh Thiruppathi, 本間正充, 鈴木和博, 佐藤陽治 次世代 DNA シークエンサーの染色体異常解析への応用 日本環境変異原学会第42回大会 (2013. 11) (岡山)

4) スレッシュ テイルパッティ, 斎藤嘉朗, 本間正充, 佐藤陽治, 鈴木孝昌 ヘモグロビンアダクトーム; 環境変異原に対する暴露マーカーとしての新しいアプローチ 日本環境変異原学会第42回大会 (2013. 11) (岡山)

5) 降旗千恵, 櫻井幹也, 渡辺貴志, 鈴木孝昌 Toxicogenomics/JEMS・MMS V: クリセン投与48時間後までのマウス肝臓における遺伝子発現変化 日本環境変異原学会第42回大会 (2013. 11) (岡山)

6) Suzuki T., Suresh T., Yamada M., Honma M., Suzuki K., Sato Y. Use of the next generation sequencers for the evaluation of genomic integrity of cellular therapy products. 11th International Conference on Environmental Mutagens (2013.11) (Foz do Iguaçu)

H. 知的財産権の出願・登録状況

H-1. 特許取得 なし

H-2. 実用新案登録 なし

H-3. その他 特記事項なし

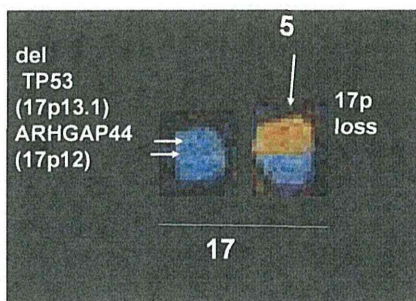
表 1 HL60 細胞におけるコピー数変化領域の詳細

chromosome	start_position	end_position	Length(kb)	copy #	type	Gene	description
1	5730000	5740000		10	5	gain	none
1	40020000	40030000		10	0	loss	PABPC4 poly(A) binding protein, cytoplasmic 4 (inducible form)
1	85980000	86010000		30	4	gain	DDAH1 dimethylarginine dimethylaminohydrolase 1
1	144920000	144950000		30	5	gain	NBPF9 neuroblastoma breakpoint family, member 9
1	144990000	145090000		100	7	gain	PDE4DIP phosphodiesterase 4D interacting protein
1	145090000	145110000		20	6	gain	SEC22B SEC22 vesicle trafficking protein homolog B
1	147710000	147730000		20	4	gain	NBPF8 neuroblastoma breakpoint family, member 8
1	156180000	156190000		10	6	gain	SLC25A44 solute carrier PMF1
1	229810000	229820000		10	1	loss	none
1	234910000	234920000		10	5	gain	none
1	248570000	248580000		10	5	gain	OR2T1 olfactory receptor, family 2, subfamily T, member 1
2	89830000	89890000		60	6	gain	none
2	133020000	133040000		20	16	gain	retro-CDC27 Retroposed Genes V5, Including Pseudogenes
2	133050000	133060000		10	4	gain	none
3	4720000	4810000		90	1	loss	ITPR1 inositol 1,4,5-trisphosphate receptor, type 1
3	59820000	59910000		90	4	gain	FHIT fragile histidine triad
3	60180000	60370000		190	1	loss	FHIT fragile histidine triad
3	75760000	75790000		30	5	gain	ZNF717 zinc finger protein 717
3	196620000	196630000		10	5	gain	SEN5 SUMO1/sentrin specific peptidase 5
4	50000	70000		20	9	gain	ZNF595 zinc finger protein 595
4	3580000	3640000		60	3	gain	LINC00955 long intergenic non-protein coding RNA 955
4	10210000	10230000		20	0	loss	none
4	34780000	34830000		50	0	loss	none
4	68260000	68270000		10	15	gain	none
4	190470000	190680000		210	6	gain	none
5	21530000	21570000		40	4	gain	GUSBP1 glucuronidase, beta pseudogene 1
5	53670000	127680000		74010	1	loss	many
5	128170000	135610000		7440	1	loss	many
5	135840000	139580000		3740	1	loss	many
5	156820000	156900000		80	1	loss	ADAM19 ADAM metalloproteinase domain 19
6	260000	370000		110	3	gain	DUSP22 dual specificity phosphatase 22
6	31280000	31290000		10	0	loss	HLA-B major histocompatibility complex, class I, B
6	32450000	32660000		210	1	loss	HLA-DRB1 major histocompatibility complex, class II, DR beta 1
6	35750000	35770000		20	4	gain	CLPS colipase, pancreatic
6	57210000	57370000		160	4	gain	PRIM2 primase, DNA, polypeptide 2 (58kDa)
6	57370000	57560000		190	6	gain	PRIM2 primase, DNA, polypeptide 2 (58kDa)
6	74590000	74600000		10	0	loss	none
6	126100000	126160000		60	1	loss	NCOA7 nuclear receptor coactivator 7
6	135280000	135810000		530	3	gain	MYB v-myb myeloblastosis viral oncogene homolog (avian)
6	157730000	157740000		10	4	gain	TMEM242 transmembrane protein 242
6	162720000	162910000		190	3	gain	PARK2 parkin RBR E3 ubiquitin protein ligase
7	12840000	13510000		670	3	gain	retro-CDC27 CTAGE family, member 5
7	57940000	58000000		60	3	gain	none
7	61540000	61560000		20	4	gain	none
7	61740000	61980000		240	5	gain	none
7	61980000	62040000		60	3	gain	none
7	130810000	131000000		190	1	loss	MKLN1 muskelin 1, intracellular mediator containing kelch motifs
7	133780000	133800000		20	1	loss	none
7	152100000	152120000		20	7	gain	KMT2C lysine (K)-specific methyltransferase 2C
7	155120000	155130000		10	4	gain	none
8	2200000	2280000		80	4	gain	none
8	32680000	32690000		10	0	loss	none
8	43090000	43100000		10	4	gain	none
8	123280000	123310000		30	1	loss	none
8	126220000	126550000		330	16	gain	NSMCE2 non-SMC el: TRIB1 tribbles pseudokinase 1
8	126710000	127390000		680	16	gain	LINC00861 long intergenic non-protein coding RNA 861
8	128060000	128070000		10	4	gain	none
8	128070000	128350000		280	16	gain	retro-SRFM1 serine/arginine repetitive matrix 1
8	128680000	128690000		10	4	gain	none
8	128690000	128770000		80	16	gain	MYC v-myc myelocytomatosis viral oncogene homolog (avian)
8	128770000	128780000		10	5	gain	retro-RPL39 ribosomal protein L39
8	130000000	130090000		90	33	gain	none
8	130090000	130220000		130	17	gain	none
8	130360000	130700000		340	18	gain	CCDC26 coiled-coil domain containing 26
9	11780000	32400000		20620	1	loss	many
9	66840000	66850000		10	7	gain	none
9	104630000	106950000		2320	1	loss	CYLC2 cyclin, basic protein of sperm head cytoskeleton
9	107060000	107100000		40	1	loss	retro-WDR45B WD repeat domain 45B
9	130890000	130900000		10	4	gain	PTGES2 prostaglandin E synthase 2
10	0	37870000		37870	1	loss	many
10	38780000	38820000		40	3	gain	none
10	38870000	39150000		280	3	gain	retro-PABPC1 poly(A) binding protein, cytoplasmic 1
10	42400000	42410000		10	63	gain	highly repeated sequences
10	46970000	47150000		180	4	gain	ANXA8 annexin A8

表 1 (つづき)

chromosome	start_position	end_position	Length(kb)	copy #	type	Gene	description
11	0	190000	190	1	loss	AL137655	cDNA DKFZp434B2016
11	200000	45180000	44980	1	loss		almost whole short arm
11	48340000	48390000	50	3	gain	OR4C45	olfactory receptor, family 4, subfamily C, member 45
11	48810000	48820000	10	4	gain	none	
11	50720000	50770000	50	3	gain	none	
11	51570000	51580000	10	24	gain	none	alpha satellite
11	54950000	54990000	40	1	loss	none	
11	55010000	55020000	10	3	gain	none	
11	85190000	85200000	10	4	gain	DLG2	discs, large homolog 2 (Drosophila)
12	2360000	2370000	10	4	gain	CACNA1C	calcium channel, voltage-dependent, L-type, alpha 1C subunit
12	11220000	11240000	20	1	loss	PRH1-PRR4	PRH1-PRR4 readthrough
12	34830000	34840000	10	4	gain	none	alpha satellite
13	55460000	63620000	8160	3	gain	PODH17 他	protocadherin 17
13	63620000	63630000	10	6	gain	none	
13	63630000	63650000	20	5	gain	none	
13	63650000	115169878	51519.878	3	gain	many	half distal long arm
14	0	20430000	20430	3	gain	many	whole short arm +alpha
14	62180000	62630000	450	1	loss	HIF1A 他	hypoxia inducible factor 1
14	64720000	81910000	17190	1	loss	many	middle of long arm
14	106540000	106560000	20	1	loss	abParts	immunoglobulin G heavy chain variable region
14	106710000	106730000	20	3	gain	abParts	immunoglobulin G heavy chain variable region
16	15050000	15060000	10	4	gain	none	
16	21570000	21590000	20	3	gain	LOC23117	cDNA FLJ59829 complete cd
16	22630000	22710000	80	3	gain	none	
16	32520000	32560000	40	5	gain	none	
16	32560000	33910000	1350	3	gain	TP53TG3 他	TP53 target 3
16	33950000	33960000	10	21	gain	none	
16	34190000	34200000	10	4	gain	none	alpha satellite
16	34470000	34760000	290	3	gain	many	retroposed genes
16	35200000	35240000	40	3	gain	none	alpha satellite
16	70890000	71200000	310	4	gain	HYDIN	axonemal central pair apparatus protein
16	78270000	78320000	50	3	gain	WWOX	WW domain containing oxidoreductase
16	81290000	81900000	620	1	loss	CMP1 他	c-Maf inducing protein
16	85460000	90354753	4894.753	1	loss	many	
17	0	7490000	7490	1	loss	many	
17	7490000	7580000	90	0	loss	TP53	tumor protein p53
17	7580000	12820000	5240	1	loss	many	middle of short arm
17	12820000	12850000	30	0	loss	ARHGAP44	Rho GTPase activating protein 44
17	12850000	20680000	7830	1	loss	many	middle of short arm
17	21200000	21330000	130	4	gain	MAP2K3	mitogen-activated protein kinase kinase 3
17	34450000	34480000	30	3	gain	retro-GOLM1	golgi membrane protein 1
17	41390000	41400000	10	4	gain	none	
17	41400000	41410000	10	8	gain	none	
17	44210000	44280000	70	4	gain	KANSL1	KAT8 regulatory NSL complex subunit 1(histon acetylation)
17	44340000	44350000	10	6	gain	none	
17	45210000	45220000	10	4	gain	CDC27	cell division cycle 27
17	45220000	45250000	30	3	gain	CDC27	cell division cycle 27
18	2350000	2810000	460	3	gain	many	
18	9080000	9950000	870	3	gain	many	
18	45340000	45500000	160	3	gain	SMAD2	SMAD family member 2
18	46560000	46960000	420	3	gain	ZBTB7C	zinc finger and BTB domain containing 7C - also CTIF & SMAD7
18	61430000	61800000	370	3	gain	SERPINF	serpin peptidase inhibitor, clade B (ovalbumin
18	69200000	69310000	110	1	loss	LOC100505771	uncharacterized LOC100505776
19	20600000	20690000	90	1	loss	ZNF826P	zinc finger protein 826, pseudogene
19	24510000	24570000	60	4	gain	none	alpha satellite
19	24570000	27860000	3290	3	gain	none	centromere
19	27870000	27890000	20	7	gain	none	alpha satellite
19	27890000	27930000	40	3	gain	none	alpha satellite
20	1560000	1580000	20	0	loss	SIRPB1	signal-regulatory protein beta 1
20	14360000	14390000	30	1	loss	MACROD2	MACRO domain containing 2 (O-acetyl-ADP-ribose deacetylase)
20	14840000	15110000	270	1	loss	MACROD2	MACRO domain containing 2 (O-acetyl-ADP-ribose deacetylase)
20	16740000	16750000	10	1	loss	none	
20	26260000	26320000	60	3	gain	none	alpha satellite
20	29810000	29830000	20	4	gain	none	(GAATG)n
20	58420000	58500000	80	3	gain	SYCP2	synaptonemal complex protein 2
21	0	10710000	10710	7	gain	many	almost short arm
21	10710000	10760000	50	20	gain	none	L1PA3 repeat
21	10760000	10860000	100	7	gain	none	L1PA3 repeat
22	16850000	16860000	10	3	gain	none	HSATIII repeat
22	24380000	24390000	10	1	loss	GSTT1	glutathione S-transferase theta 1
22	25750000	25900000	150	3	gain	LRP5L	low density lipoprotein receptor-related protein 5-like

表 2 17 番染色体上の遺伝子欠失領域



Invisible deletions were detected!

chromosome	start	end	size(kb)	copy #	type	Gene	description
17	0	7490000	7490	1	loss	many	
17	7490000	7580000	90	0	loss	<b>TP53</b>	tumor protein p53
17	7580000	12820000	5240	1	loss	many	middle of short arm
17	12820000	12850000	30	0	loss	<b>ARHGAP44</b>	Rho GTPase activating protein 44
17	12850000	20680000	7830	1	loss	many	middle of short arm
17	21200000	21330000	130	4	gain	<b>MAP2K3</b>	mitogen-activated protein kinase kinase 3
17	34450000	34480000	30	3	gain	<b>retro-GOLM1</b>	golgi membrane protein 1
17	41390000	41400000	10	4	gain	none	
17	41400000	41410000	10	8	gain	none	
17	44210000	44280000	70	4	gain	<b>KANSL1</b>	KAT8 regulatory NSL complex subunit 1
17	44340000	44350000	10	6	gain	none	
17	45210000	45220000	10	4	gain	<b>CDC27</b>	cell division cycle 27
17	45220000	45250000	30	3	gain	<b>CDC27</b>	cell division cycle 27

図 1 ホールゲノムシーケンスデータの冗長度

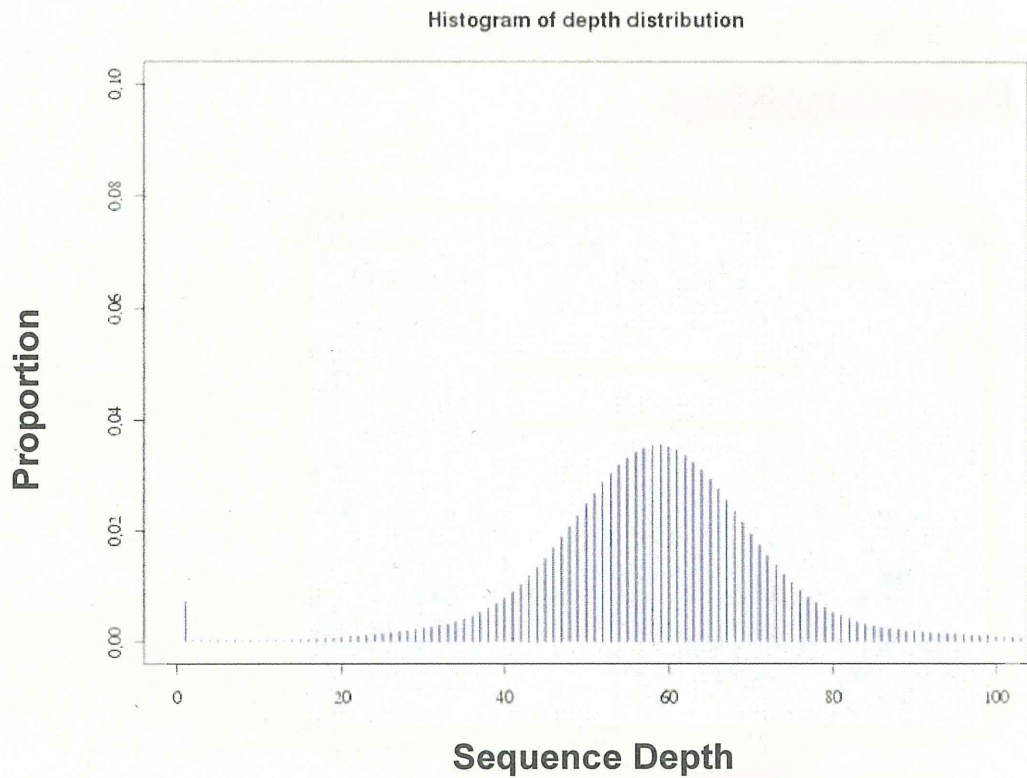


図 2 h MSC 細胞のシーケンスデータを用いた CGH (コピー数異常) 解析

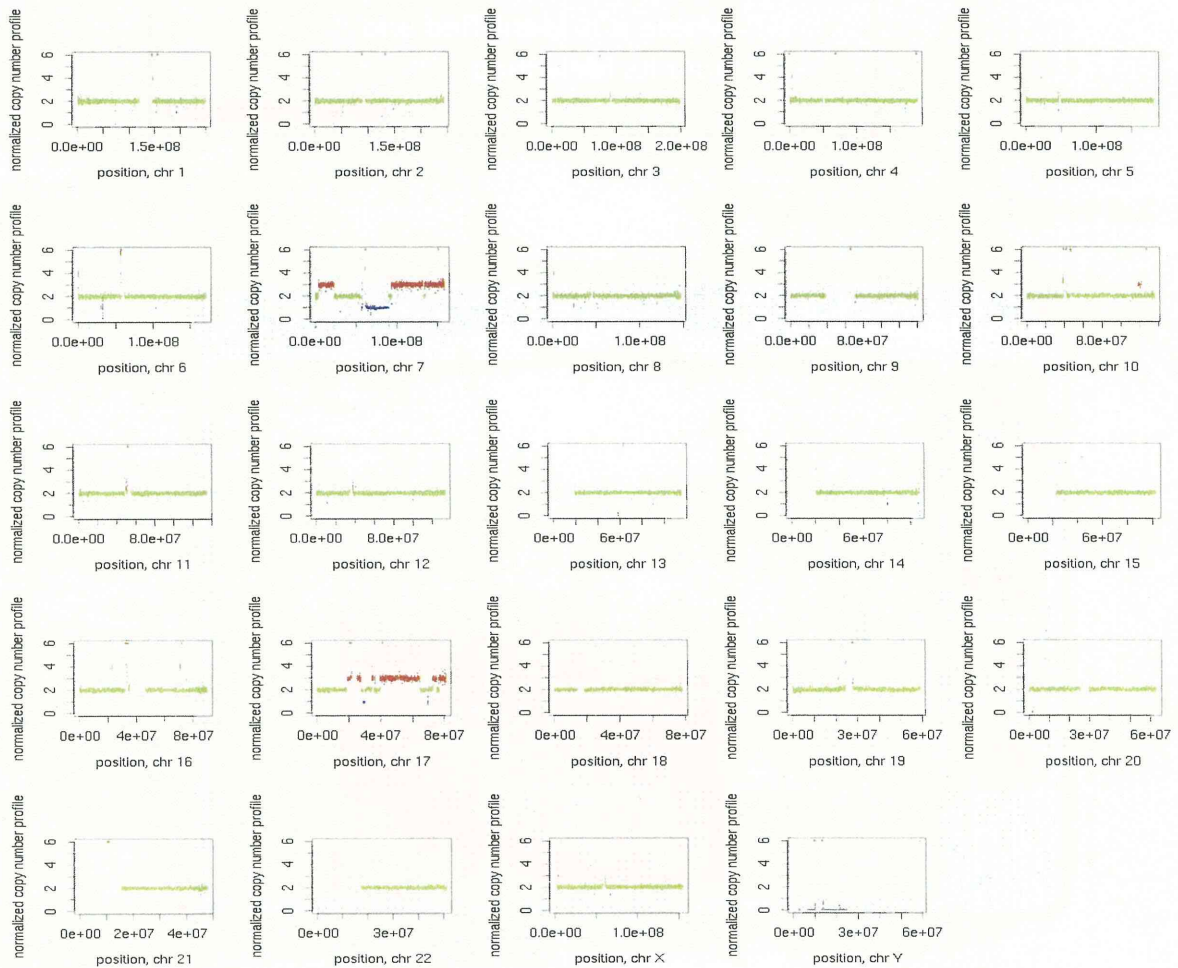


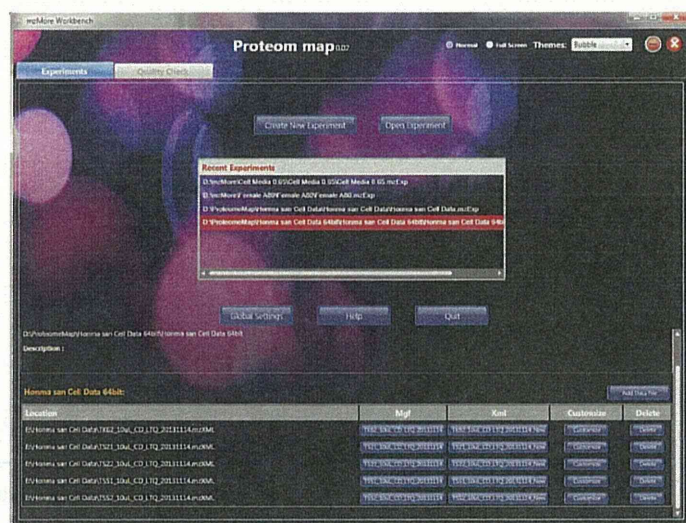


図3 ProteomeMapソフトウェアの概要

1) 生データの取り込み画面

# ProteomeMap

- put your proteins on the map...

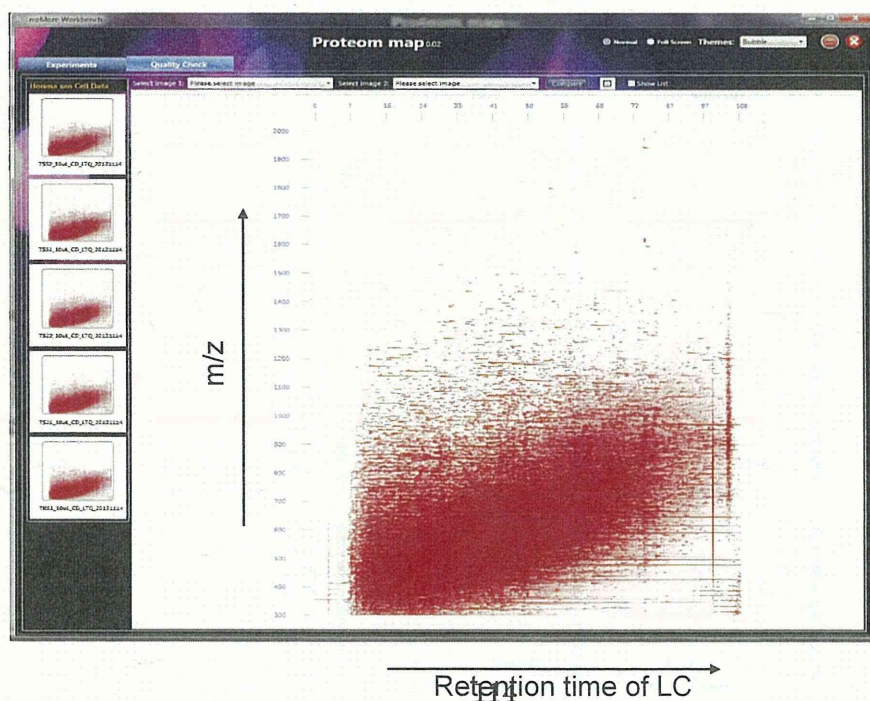


## Compare

what proteins are identified,  
where it is identified and  
how its identified

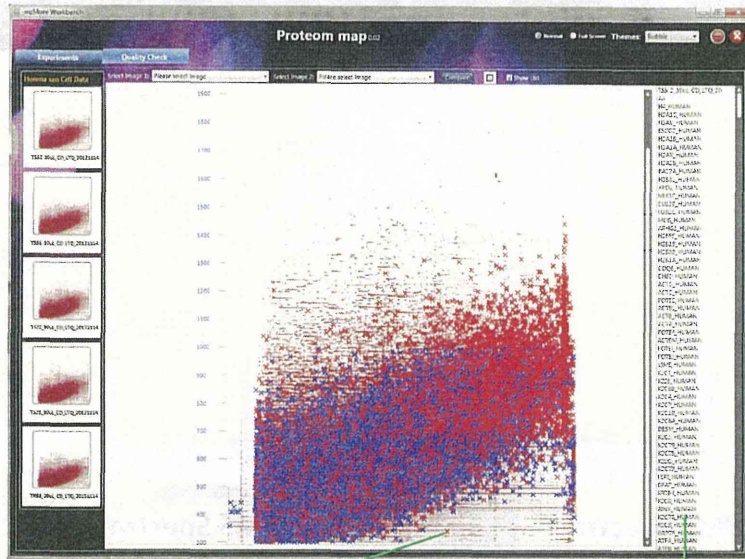
*all in one screen*

2) LC-MSデータの2次元マップ表示



3) MS/MSデータの取り込みと表示

### Mapping of MS/MS Data on 3D Map & List of Identified Proteins

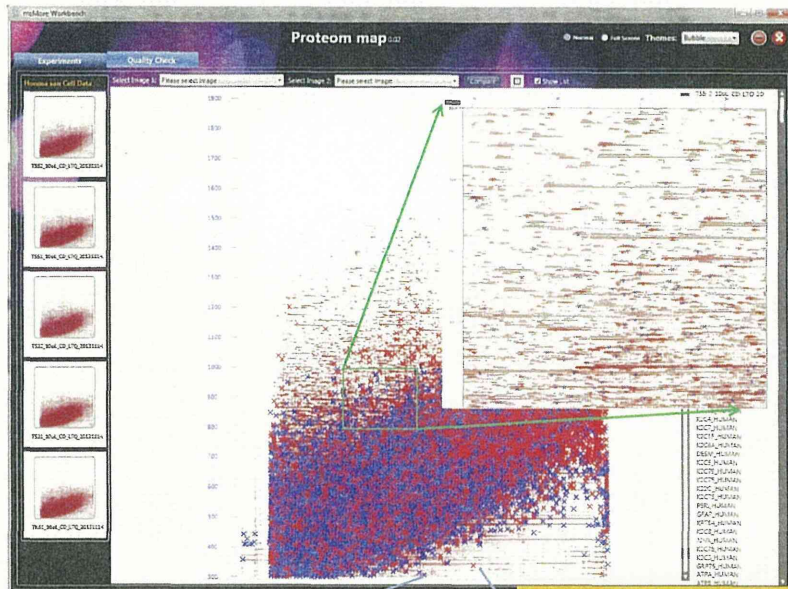


Location of MS/MS and identified proteins

List of identified proteins

4) ズームイン機能

### Zoom IN Option of ProteomeMap



Blue X marks identified peptides

Red X marks un-identified MS/MS data



厚生労働科学研究費補助金(医薬品・医療機器レギュラトリーサイエンス総合研究事業)  
「細胞・組織加工製品の開発環境整備に向けたレギュラトリーサイエンス研究」  
分担研究報告書

分化プロペンシティを指標とした細胞特性解析法の開発

研究代表者：佐藤 陽治 国立医薬品食品衛生研究所 遺伝子細胞医薬部・部長  
研究分担者：安田 智 国立医薬品食品衛生研究所 遺伝子細胞医薬部・第4室・室長

**研究要旨**

iPS細胞といった多能性幹細胞は、多分化能と自己増殖能を併せ持つため、再生医療を目的とした細胞・組織加工製品の原材料として大きな期待を担っている。多能性幹細胞利用製品の实用化促進には、多能性幹細胞から目的細胞への分化効率を上昇させ、目的外細胞の製品への混在をより低減する必要がある。多能性幹細胞は、分化における指向性(分化プロペンシティ)を持つことから、目的細胞への分化に最適な株を選定して製品化を行うことが重要である。本研究では、多能性幹細胞株の目的細胞への分化を適切に予測することを目標とし、分化プロペンシティを指標とした細胞特性解析法の開発を行った。ヒト iPS 細胞 10 株から胚葉体を形成させ、三胚葉系細胞のマーカー遺伝子を定量し、主成分分析を行うことにより、各々の細胞株の内胚葉、中胚葉および外胚葉系細胞への分化プロペンシティを数値化した。さらに未分化状態でのヒト iPS 細胞株の網羅的なトランスクリプトーム解析を行い、発現量と分化プロペンシティとの相関のある mRNA と miRNA の同定を試みた。今回の結果は、未分化状態のヒト iPS 細胞株における内胚葉、中胚葉および外胚葉系細胞への分化プロペンシティ予測マーカーの同定に繋がるものと期待される。

**研究協力者**

黒田 拓也 国立医薬品食品衛生研究所 遺伝子細胞医薬部 研究員  
城 しおり 名古屋市立大学大学院薬学研究科 医薬品質保証学講座 修士課程2年  
草川 森士 先端医療振興財団 細胞療法開発事業部門 研究員  
田埜 慶子 国立成育医療研究センター 生殖・細胞医療研究部 研究員  
中島 啓行 先端医療振興財団 再生医療研究開発部門 研究員  
高田 のぞみ 先端医療振興財団 再生医療研究開発部門 研究員