

## II. 分担研究報告

### 厚生労働科学研究費補助金（難治性疾患克服研究事業） （分担）研究報告書

ヒト ES/iPS 細胞遺伝子発現プロファイルのバイオインフォマティクス解析  
研究分担者：水口 賢司

#### 独立行政法人医薬基盤研究所 創薬基盤研究部 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：** 現在は、多くのヒト ES/iPS 細胞が国内外の様々な研究所において樹立され、様々な培養方法により培養され、研究に使用されている。これらの細胞を治療目的あるいは創薬研究目的のために利用するには、詳細な細胞特性分析や培養手順の標準化が必要である。これらの細胞株の差異、及び培養条件の差異を調査するために、我々は、NIBIO ヒト幹細胞応用開発室において出されたヒト ES/iPS 細胞の遺伝子発現解析結果について、バイオインフォマティクス解析を行った。

#### A. 研究目的

ゲノム解析技術や各種ハイスループット実験技術の進展に伴い、医学生物学のいずれの分野においても大規模データの取り扱いが日常的なものとなり、コンピュータ解析が必須となっている。扱われるデータの種類は、遺伝子発現、相互作用ネットワーク、タンパク質立体構造など多種多様だが、それらの解析には、データベース技術や統計学、機械学習（コンピュータがデータから自動的にルールを抽出し学習する技術の総称）などの共通技術が用いられ、これらの基本技術に

支えられ、生物情報からの知識抽出を目指す分野一般を広い意味でバイオインフォマティクスと呼ぶことが多い。

ヒト胚性幹（ES）細胞は、1998年にウィスコンシン大学でトムソンらによって樹立された。それ以来、多能性をもつ多くのヒト ES 細胞株が報告されている。やがて、2006年に人工多能性幹細胞(iPS)は日本で高橋と山中により開発された。これらのヒト ES/iPS 細胞の最大の特徴は未分化状態が維持できることに加え、多能性と分化能であり、それらは一般に組織形態学的解析及びマーカー遺伝子発現

解析によって同定される。これまで、一般的に新しくヒト ES/iPS 細胞株を樹立した際には、特定の幹細胞遺伝子マーカー発現や分化させた際の分化マーカーにより細胞特性評価がなされてきたが、他の細胞株との比較などはほとんど行われてこなかった。結果として、新規細胞株の樹立論文では多くの場合、わずかな細胞特性評価しか発表されていない。

一方で、幹細胞の再生医療及び創薬研究への応用は徹底的な細胞特性評価を必要とする。これらの細胞特性のわずかな差異は、実際に応用する際に大きな差異となって研究結果に影響を与えうる。これを考慮して、国際幹細胞イニシアチブは、世界中の研究機関から集められた 59 の胚性幹細胞株の詳細な特性評価を実施し、その結果を 2007 年に報告した。このプロジェクトの詳細な経緯やデータセットは、ウェブ上で公開され (<http://www.stem-cell-forum.net/initiatives/isci/isci-overview-and-structure/isci-1/>)、閲覧可能である。これらの細胞株の幹細胞(未分化細胞)マーカー及び分化細胞マーカー等の表面抗原プロフィール及び遺伝子発現プロフィールの両方が調査された。この研究により、特性評価に使われた幹細胞株は、同じ細胞株間では異なる条件下及び異なる樹立機関から作られたとしてもマーカー遺伝子のセットは非常に似通った発現パターンを有していたことがわかった。しかし、異なる細胞株間では、

遺伝子発現及び抗原プロフィールにおいて差異が認められ、それが個々の細胞株を特徴づけていた。各種細胞株間の差異に関するその他の報告も発表されている。これらの研究結果より、ヒト ES/iPS 細胞株の多様性及び類似点に関する初期的洞察が可能であるが、表面抗原プロフィール及び遺伝子発現プロフィールに関する生物学的意味は解明されていない。そこで、我々は分化及び未分化状態のヒト ES/iPS 細胞株に関して、広く研究されているマーカー遺伝子の発現プロフィールの系統的分析を行い、細胞特性評価を行うことを本研究の目的とした。

未分化状態で維持培養されたヒト ES/iPS 細胞は、Oct3/4、Nanog、Tra1-60、Tra1-81、Tra2-54、SSEA3、SSEA4 等の未分化マーカーを発現し、SSEA1 等の分化マーカーを発現しない。これら表面抗原マーカー発現を判定する免疫染色やフローサイトメトリー(FCM)解析は、細胞の培養状態、すなわち未分化/分化状態の比較的解釈が簡単な評価方法であり、世界中で採用されている。NIBIO ヒト幹細胞応用開発室においても、この評価方法が細胞の品質管理のひとつとして採用され、改良・改善を加えつつ、継続的に運用されている。しかし、免疫染色や FCM 解析では、基本的には蛋白質や糖脂質など抗体の抗原となりうる遺伝子の発現解析が可能であるが、mRNA レベルの遺伝子発現の解析が困難である。また、アフィニティーの

高い抗体がある場合にのみ解析が可能であり、一度に解析できる蛋白質や糖脂質の数はある程度限られているため、網羅的に解析をすることはできない。しかし、ヒト ES/iPS 細胞を治療目的あるいは創薬研究目的のために利用するには、詳細な細胞特性分析や培養手順の標準化が求められている。そこで、NIBIO ヒト幹細胞応用開発室では、免疫染色、FCM 解析と平行して、未分化状態で維持培養されたヒト ES/iPS 細胞の mRNA レベルの遺伝子発現を網羅的・定量的に測定するために、細胞株毎、培養条件ごとに RNA サンプルを採取し、PCR アレイ測定をおこなってきた。

本研究で我々は、NIBIO ヒト幹細胞応用開発室において蓄積された PCR アレイ測定結果をより詳細に解析し、細胞株の差異、及び培養条件の差異を調査することを第一の目的とし、バイオインフォマティクス解析を進めた。さらに、より迅速且つ正確で詳細な遺伝子発現プロフィール解析を進めるための解析手順を策定することを目的とし、アレイデータの取り扱い方法についての検討を行った。

## B. 研究方法

### 未分化維持培養及び胚様態形成時のヒト ES/iPS 細胞株における幹細胞遺伝子アレイ

NIBIO ヒト幹細胞応用開発室(当時は培養資源研究室)において、種々のヒト ES/iPS 細胞株の未分化状態ならびに胚様体を形成させて分化させた状態の RNA を抽出して、幹細胞遺伝子 PCR アレイを行った結果を解析した。

これらの実験結果はそれぞれ CT 値(ある一定の核酸量が測定されるまでに必要な PCR の増幅回数; 遺伝子発現量が高いと CT 値は低い)で表される。本解析の目的は、CT 値に生物学的解釈を付与して妥当な統計量を定義することであり、特に様々な細胞株における類似点と相違点を強調し、遺伝子発現の全体のパターンを要約することである。

実験結果の統計学的特徴を理解するため、結果を視覚化するために最適な方法を検討した。ヒートマップとクラスタリング図は発表文献で広く用いられており、これらの方法やその他の簡単で広く使われているアプローチを適用することを目指した。可視化ツールと統計学的ツールの両方を用いることで、細胞株及び各種細胞状態のバイオマーカーとしての遺伝子の発現プロフィールと、これまでに知られている生物学的知識との間で一致する結論を導き出すことを目標とした。PCR アレイで用いられた細胞株のリストは下記の表に示した。

アレイ名	細胞株
PCR 多能性 アレイ (未分化状態)	EC_NCR.G3, EC_2102Ep.c1.2A6, ES_KhES.1 (three experiments), ES_KhES.3 (three experiments), ES_H9 (three experiments), ES_HES (three experiments), iPS_Tic (three experiments), iPS_Dotcom (three experiments), iPS_Squeaky (three experiments), iPS_Toe (three experiments), iPS_Lollipop, iPS_201B7 (three experiments), iPS_201B2subclone2 (three experiments), iPS_201B2subclone4 (three experiments), iPS_ForESkin, iPS_iDMC.03, iPS_iDMC.13,
多能性 アレイ (胚様体状態)	ES_KhES.1, ES_KhES.3 (three experiments), ES_H9 three experiments), ES_HES3 (two experiments), iPS_Tic (two experiments), iPS_Dotcom (three experiments), iPS_Squeaky (three experiments), iPS_Toe (three experiments), iPS_201B7 (three experiments), iPS_201B2subclone2 (three experiments), iPS_201B2subclone4 (three experiments), iPS_ForESkin (two experiments), iPS_iDMC.03,

### 未分化状態のヒト ES/iPS 細胞株の 特性解析

平成 24～25 年度には、NIBIO ヒト幹細胞応用開発室において、未分化状態でのヒト ES/iPS 細胞の特性を明らかにするため、未分化維持培養をおこなったヒト ES/iPS 細胞から RNA を抽出し、Human Stem Cell Pluripotency Array (Applied Biosystems) を用い、定量 RT-PCR による遺伝子発現プロファイル解析が重点的に行われた。この研究で使用し

た細胞株の詳細については、古江の項を参照されたい。その検査方法は、平成 23 年度に策定を行った方法に従った。これは、国際幹細胞イニシヤティブ (ISCI) にて策定された方法と同様のものである。

この PCR アレイ (Human Stem Cell Pluripotency Array) において発現量が測定される遺伝子は 96 遺伝子であり、このうち 6 遺伝子 (ATCB、RAF1、CTNNB1、GAPD、EEFA1、18S) はハウスキーピング遺伝子と呼ばれる一般に発現量が変化しにくい

とされる遺伝子群、残る 90 遺伝子が本来の解析対象の遺伝子群である (表 2)。各遺伝子の発現量の測定結果は CT 値 (ある一定の核酸量が測定されるまでに必要な PCR の増幅回数; 遺伝子発現量が高いと CT 値は低い) として表される。ハウスキーピング遺伝子をコントロールとしてこの CT 値に補正をかけ、サンプル間で遺伝子発現量を直接比較できるように算出したものが  $\Delta$ CT 値である。我々は、まず、 $\Delta$ CT 値を算出する際のコントロールとして適切なハウスキーピング遺伝

子 3 遺伝子 (ATCB、RAF1、GAPD) を選抜し、これらをもとに算出された  $\Delta$ CT 値を用いて解析を進めた。

階層的クラスタリングは、R 統計ソフトウェア (<http://www.r-project.org>) の `hclust` 関数で実行し、距離としてユークリッド距離を使用、`average-linkage` のクラスタリングを行った。ヒートマップは、R による `gplot` パッケージ中の `heatmap2` 関数により作成した。

本研究で使用した PCR アレイ Human Stem Cell Pluripotency Array (Applied Biosystems) に含まれる遺伝子のリストは次の表に示した。

コントロール遺伝子	解析対象遺伝子
ACTB, 18S, CTNNB1, RAF1, EEF1A1, GAPD	BRIX, CD9, COMMD3, CRABP2, DNMT3B, EBAF, FGF5, FOXD3, GABRB3, GAL, GBX2, GDF3, GRB7, IFITM1, IFITM2, IL6ST, IMP2, KIT, LEFTB, LIFR, LIN28, Nanog, NODAL, NR5A2, NR6A1, PODXL, POU5F1, PTEN, REST, SEMA3A, SFRP2, SOX2, TDGF1, TERT, TFPC2L1, UTF1, Xist, ZFP42, FGF4, NOG, CDX2, CGB, EOMES, KRT1, GCM1, DDX4, SYCP3, COL1A1, COL2A1, RUNX2, HBB, HBZ, ACTC, NPPA, DES, MYF5, MYOD1, T, WT1, CD34, CDH5, FLT1, PECAM1, GCG, IAPP, INS, IPF1, PAX4, SST, TAT, AFP, FN1, FOXA2, GATA4, LAMA1, LAMB1, LAMC1, PTF1A, SERPINA1, SOX17, GFAP, HLXB9, ISL1, NES, NEUROD1, OLIG2, PAX6, SYP, TH

## C. 研究結果

### 未分化維持培養及び胚様態形成時のヒト ES/iPS 細胞株における幹細胞遺伝子アレイ

CT 値に生物学的解釈を付与して妥当な統計量を定義すること、特に様々な細胞株における類似点と相違点を強調し、遺伝子発現の全体のパターンを要約することを目的として、可視化ツールの開発と、適切な統計量の定義を行った。

#### プレ処理：

CT データの品質管理のための処理を行った。まず、その値を越えると信頼のおけるデータ分析が難しいと考えられる CT の最大値を定義した。次に、同じ細胞株の分化状態で複数実験のデータが存在する場合、全ての実験の CT 値の平均を求めて解析に用いた。いくつかのケースでは、最大数の増幅周期まで全くシグナルが観察されなかったため、CT 値は“不定”とした。このようなケースでは不定値を除き、残りのデータから平均値を算出した。

#### データスケーリングと標準化：

全ての CT 値は実験内の特定のコントロール遺伝子に対する相対値として扱う必要がある。各実験に複数のコ

ントロール遺伝子が使われているため、注目する遺伝子の CT 値はその実験に使われた全てのコントロール遺伝子の平均 CT 値からの差 ( $\Delta$  CT 値) として表現した。  $\Delta$  CT 値はコントロール遺伝子と比較して特定の条件下での遺伝子の発現を表しており、負の  $\Delta$  CT 値は遺伝子の過剰発現を意味する。特定の細胞株の単独の (あるいは平均の) PCR アレイ中の各遺伝子に  $\Delta$  CT 値が割り当てられる。同じ細胞株のアレイの中の二つの遺伝子間の、或いは同じ遺伝子の二つの細胞株間の比較は、これらの  $\Delta$  CT 値の差異を取ることにより行われ、  $\Delta \Delta$  CT と呼ばれる (Livak KJ, Schmittgen TD, 2001)。CT 値は指数関数的発現パターンを表すため、真の発現量は  $2^{-\Delta \Delta \text{CT}}$  を使って得ることができる。これらの値は広範囲である可能性があるため、プロットでは対数目盛を用いた。

#### クラスタリング：

多くの細胞株についてのアレイ中の全ての遺伝子の発現値の比較のため、クラスタリングによって、似た発現パターンを持つ遺伝子をグループ化した。このアプローチは発表文献で広く使われている。多能性アレイの様々な細胞株の発現レベルを示すクラスタリング図の例を下記に示した (図 1、2)。遺伝子は縦軸に、細胞株は横軸に示され、two-way クラスタリングを実施した。未分化状態の遺伝子

発現レベルが、高発現量と低発現量の二つのグループに明確に分けられた。概して、遺伝子のグループは 2007 年の ISCI による研究結果とかなり一致している (図 1, 2)。

#### 散布図:

これは二つの変数の関係を見る標準的方法である。二つの条件での多数の遺伝子の発現パターン間の一対一の比較は単純な散布図から得られ、それは回帰直線を引くことで更に解釈ができる。個々の遺伝子の一般的傾向からのずれは、回帰直線からの距離を見ることでわかる。適切に距離をスケールリングして、発現量のずれが閾値外である遺伝子のラベリングをすることにより、特異的な発現パターンを示す遺伝子を同定できる (図 3)。H9 細胞株に対する比較評価を行った際に Dotcom 細胞や Toe 細胞、Foreskin 細胞において特異的に過剰に、或いは過小に発現している遺伝子が二つの破線で示される散布図の閾値ラインの外に表されている。

特定の状態 (EB 或いは未分化) での各細胞株で同じ遺伝子群が発現したように、本研究で使用した全てのヒト ES/iPS 細胞株の全体的な遺伝子発現パターンは大変類似している。例えば、発生において重要な役割を果たす遺伝子 SOX2 の EB と未分化細胞状態での様々な細胞株における発現パターンを図から見て取れる。全ての細胞株において EB 状態よりも未分化状態で、SOX2 がより高い発現レベルを有することがわかる。しかし、SOX2 は全ての細胞株で双方の状態において、いくらかの正の発現レベルを有している。これは、発現レベルが大幅に変化しても、SOX2 が分化細胞と同様に未分化細胞でも発現し続けるという報告と一致する。また、全ての細胞株の発現プロファイル間の全体的な類似にもかかわらず、わずかな差異が存在しうるということも発見した。これまでに行ったこのような差異の解析は、個々の ES 細胞株の細胞特性評価に役立つと期待される。

# Pluripotency UD array

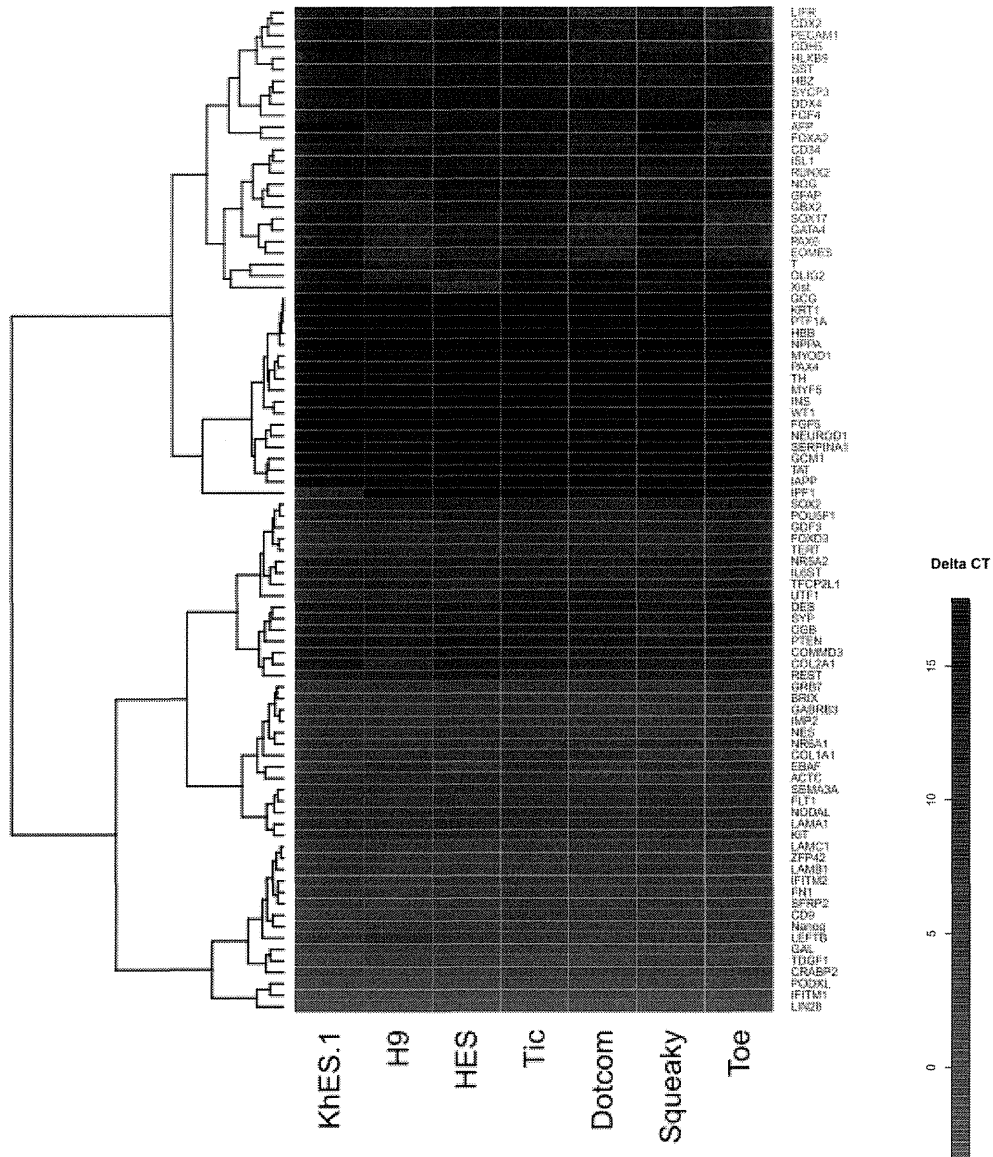


図1. ヒト ES/iPS 細胞の未分化状態における幹細胞遺伝子アレイ解析結果のヒートマップによる分析



## Pluripotency EB array

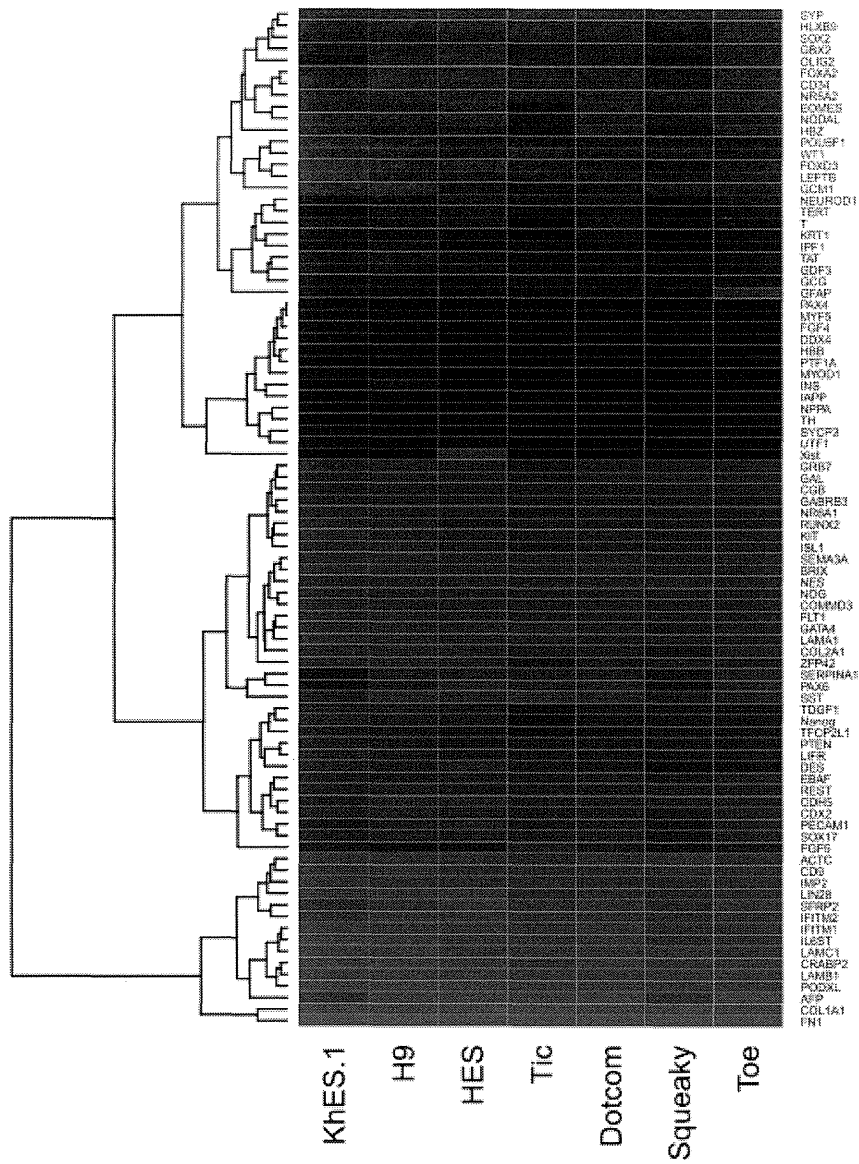


図2. EB 形成により分化させたヒト ES/iPS 細胞における幹細胞遺伝子アレイ解析結果のヒートマップによる分析



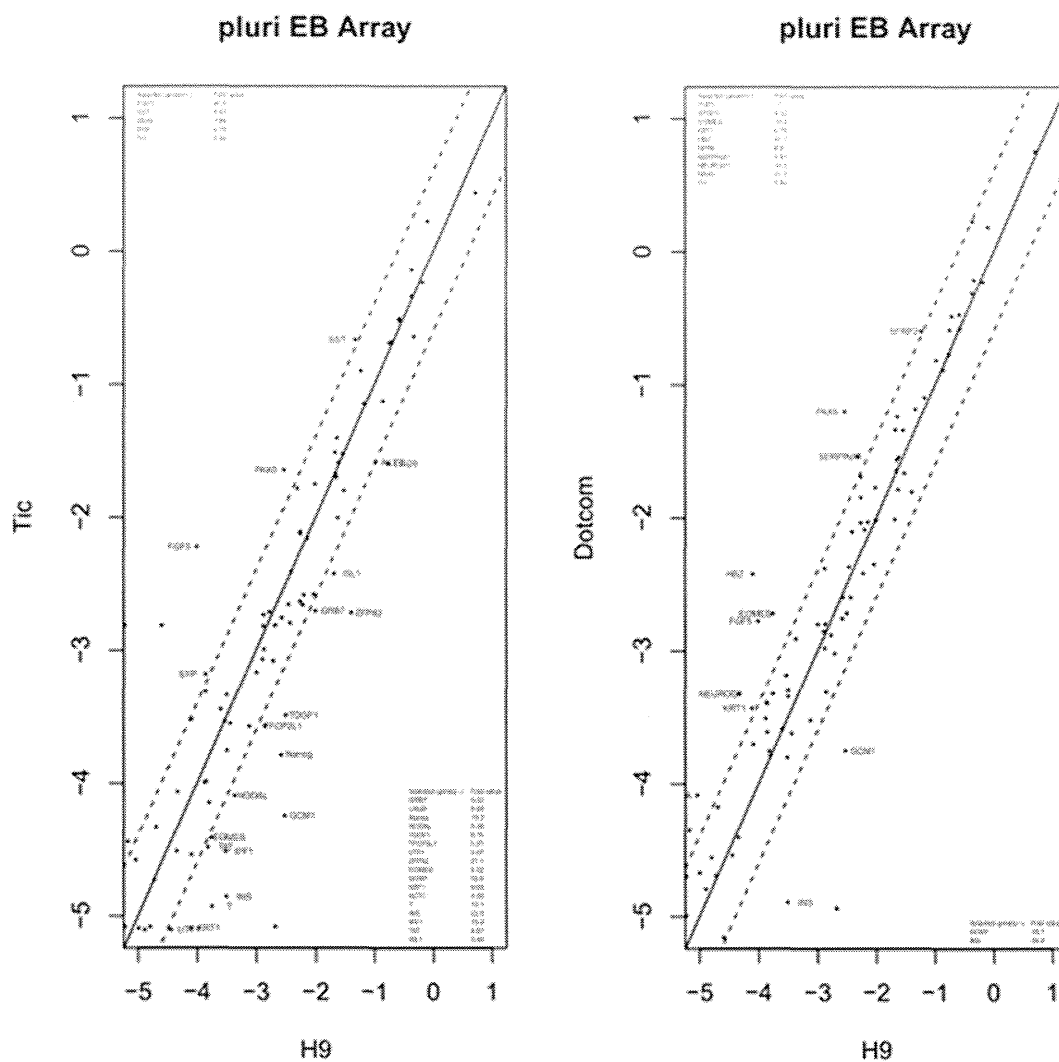


図 3-2. ヒト ES/iPS 細胞株における幹細胞遺伝子アレイ解析結果の散布図による分析 (Tic, Dotcom)

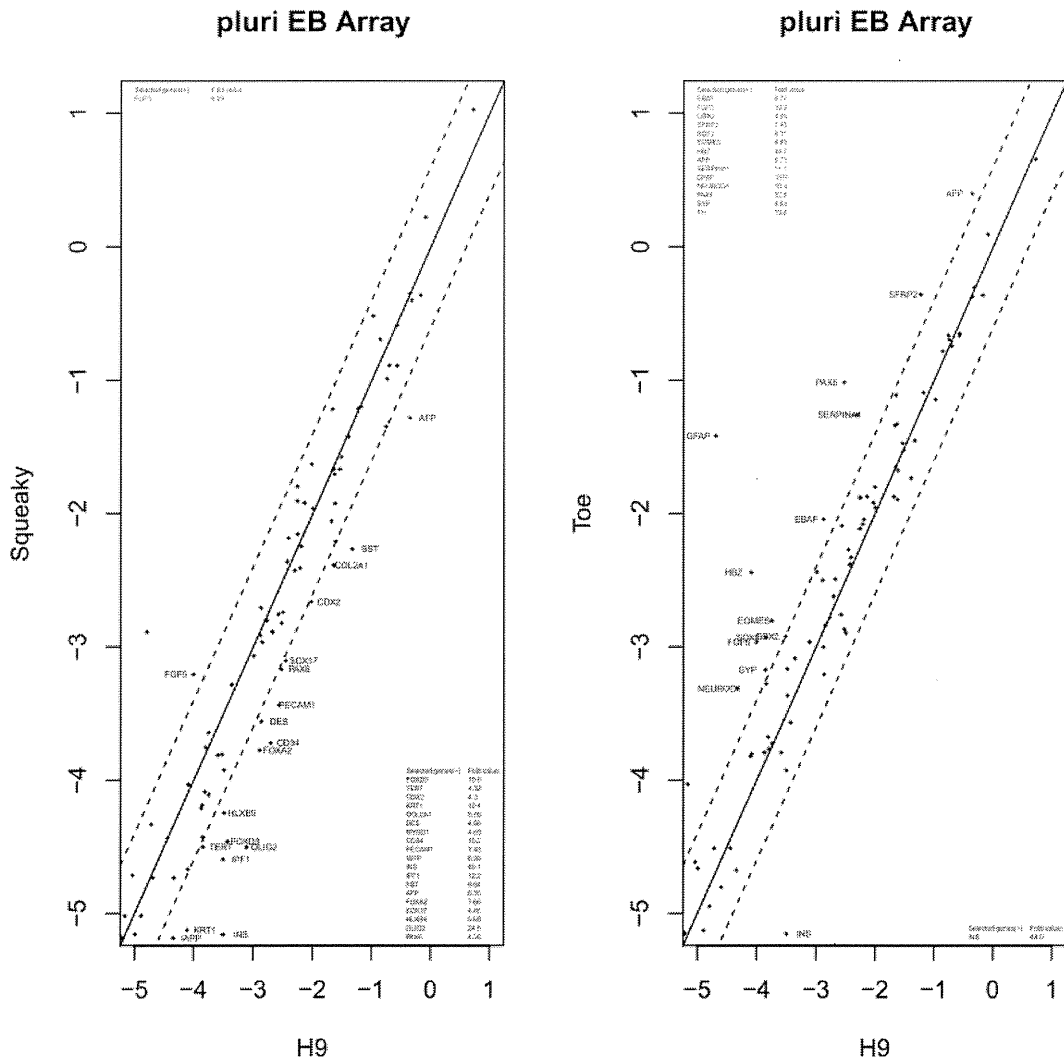


図 3-3. ヒト ES/iPS 細胞株における幹細胞遺伝子アレイ解析結果の散布図による分析 (Squeaky, Toe)

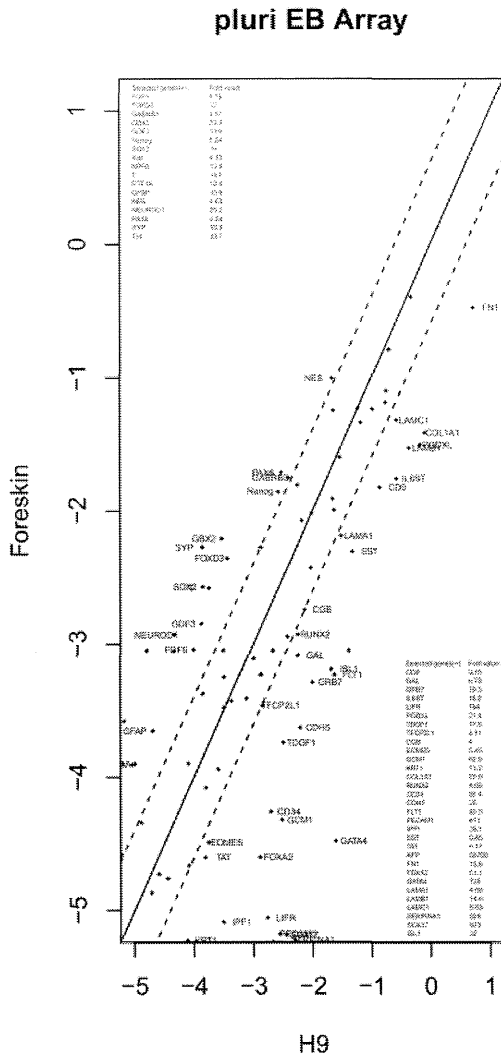


図 3-4. ヒト ES/iPS 細胞株における幹細胞遺伝子アレイ解析結果の散布図による分析 (Foreskin-iPS)

## 未分化状態でのヒト iPS 細胞の特性解析

未分化状態でのヒト iPS 細胞の特性を明らかにするため、遺伝子発現プロフィール解析を行った。平成 24 年度には、厚生労働省研究資源バンク・細胞バンク (JCRB Cell Bank) に登録されたヒト iPS 細胞 JCRB1331 Tic、JCRB1327 Dotocom、JCRB1437 iPS-TIG-114-4f1、コントロールとして京都大学 iPS 細胞研究所より供与されたヒト iPS 細胞 201B7、京都大学ヒト ES 細胞 KhES-1(京都大学より分与)、ウィスコンシン大学ヒト ES 細胞 H9(Wicell Bank より分譲)を解析し、平成 25 年度には、3 株のヒト iPS 細胞株 ; JCRB1341 iPS-TIG120-3f7 、 JCRB1363 iPS-TIG120-4f1 及び NIHS0693 UTA-SF2-2 (以上 JCRB Cell Bank)、2 株のヒト ES 細胞株 ; KhES-4 及び KhES-5 (京都大学より分与)を追加し、平成 24 年度までに蓄積したデータと合わせて遺伝子発現プロフィールの細胞株間の相違や個々の細胞株の特性を示す遺伝子等について検討した。

### ハウスキーピング遺伝子群 :

前述したとおり、ハウスキーピング遺伝子群は一般的に発現量の変化しにくいものとされている。しかしながら、ヒト ES/iPS 細胞においてはハウスキーピング遺伝子群の発現の挙動

が一般細胞と異なり、必ずしも一定でないということが我々の解析によっても明らかとなった(図 4)。そこで、このバイオインフォマティクス解析においては、 $\Delta CT$  値による解析をより正確なものにするために、ATCB、RAF1、および GAPD の 3 遺伝子のみをコントロール遺伝子として取り扱うことと決定した。

### プレ解析 :

NIBIO ヒト幹細胞応用開発室において、ヒト ES/iPS 細胞の未分化維持培養過程におけるヒト幹細胞の未分化マーカーや早期分化マーカーなどの遺伝子群の発現の安定性を確認するために、Human Stem Cell Pluripotency Array を用いて、対象株について異なる継代数で複数回、測定された。その結果について、我々はまず、上記のコントロール遺伝子を用いて  $\Delta CT$  値の算出、および遺伝子プロフィールのプレ解析(図 5)を行った。

### 遺伝子発現プロフィール解析 :

次に、対象株毎に、異なる継代数のサンプルの  $\Delta CT$  値を比較した。解析結果を視覚的にとらえ理解しやすくするために、二次元スキャッタープロット(散布図)を作成した(図 6、7)。このサンプルごとの  $\Delta CT$  値比較解析により、同一の細胞株、同一の培養方法によってえられた細胞サンプルであっても、継代数が異なる場合、発現

量が一定している遺伝子がある一方、ある程度発現量の異なる遺伝子が存在することが明らかとなった。つまり、未分化維持培養を続けていたとしても、継代ごとに遺伝子発現プロフィールが異なるということである。我々は、このことに注目をして、より詳細にサンプル間の遺伝子発現プロフィール解析を行った。その例を図 8、9 に挙げる。

UTA-SF2-2 は、東京大学浅島研究室にて無フィーダー・無血清にて樹立され寄託されたヒト iPS 細胞である。NIBIO ヒト幹細胞応用開発室では、UTA-SF2-2 の資源化にあたり、ウシファイブロネクチンならびに hESF9a 培地を用いた無フィーダー・無血清の培養工程における品質管理法を策定し、実際に資源化をおこなっている。その過程で、品質管理の一環として、RNA サンプルを 3 回、異なる継代数で採取し、遺伝子発現プロフィール解析が行われた。遺伝子発現プロフィールは 3 サンプル間で非常によく一致し、このサンプルにおいて遺伝子発現が安定していることが明らかとなった。

一方、Dotcom は分化誘導に適しているため資源化細胞の需要も高いが、一般には安定して未分化に維持して培養することが難しいとされている。この Dotcom は未分化維持培養の過程においても、たびたび分化する傾向があるため、資源化や研究で品質管理を行った場合に、免疫染色や FCM 解析の結果、未分化細胞の割合が通常より低い（全体の 80%以下）と判定されることがある。今回、このバイオインフォマティクス解析をおこなったサンプルは、少なくとも 80%以上未分化細胞が含まれるという細胞集団のサンプルから RNA を抽出し、遺伝子発現を測定したものであるが、個々のサンプル間の遺伝子発現プロフィールは大きく異なる場合があることが判明した。つまり、免疫染色や FCM 解析の未分化/分化マーカー発現解析のみでは判明しなかった遺伝子発現の変化が、PCR アレイによる網羅的遺伝子発現測定とバイオインフォマティクス解析を組み合わせることによって詳細に解析できたということを示している。







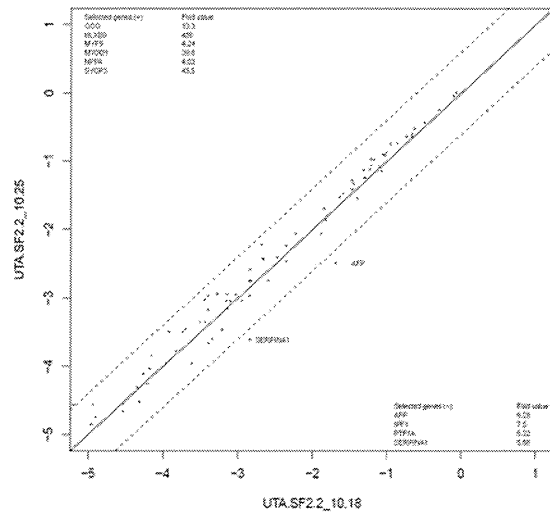
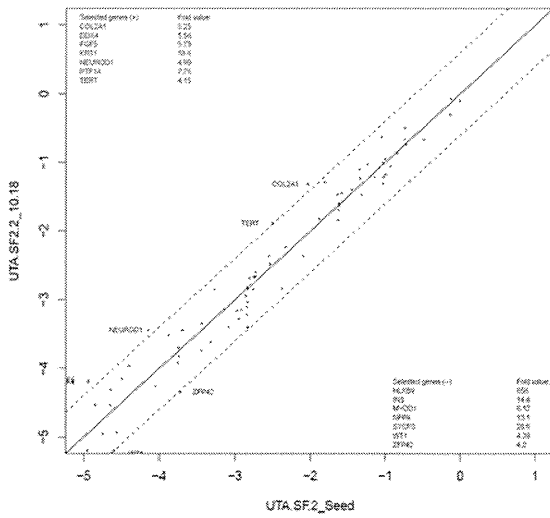
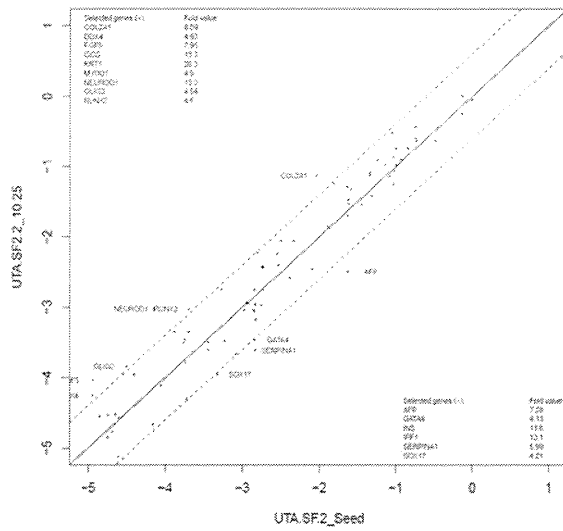


図6 UTA-SF2-2における遺伝子発現プロファイル解析例（スキャタープロット）



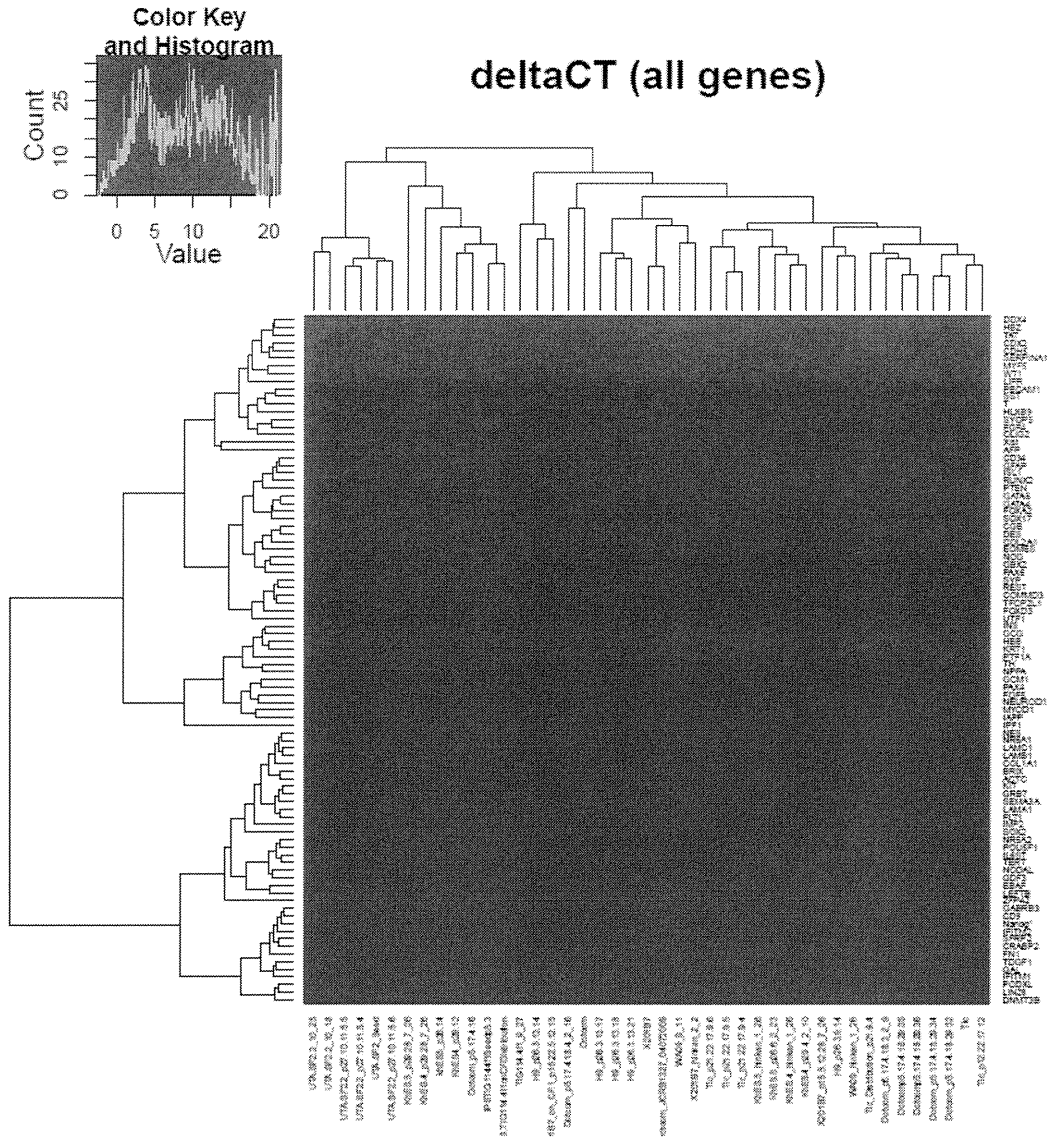


図 8 遺伝子発現プロフィール解析例 (ヒートマップ)

32 サンプル(8 細胞株)について 90 遺伝子の発現量を  $\Delta CT$  値で示したもの。各サンプル (横軸) について、遺伝子発現プロフィールの類似度をユークリッド距離によって定義し、階層的クラスタリングを行なうと共に、各遺伝子 (縦軸) についてサンプル群による発現パターンで階層的クラスタリングを行なった。