

第7章 ビッグデータの活用事例と今後の狙いどころ

第2節 新規事業, 研究開発テーマ創出への活用

[3] ヒト細胞からのビッグデータの情報管理と情報解析技術

加藤 有己 京都大学 iPS 細胞研究所 特定拠点助教

桜井 都衣 京都大学 iPS 細胞研究所 研究員

藤渕 航 京都大学 iPS 細胞研究所 教授

(株)技術情報協会
「ビッグデータの収集、調査、分析と活用事例」

[3] ヒト細胞からのビッグデータの情報管理と情報解析技術

京都大学 加藤有己, 桜井都衣, 藤瀬航

はじめに

ライフサイエンスの分野では、近年の目覚ましいシーケンシング技術の進展により、膨大な個数の配列データを代表とするビッグデータが産生されている。ビッグデータが与えられたとき、大量の個々の事象を解析できるだけでなく、データ全体を包括的に取り扱うことが可能となる。ビッグデータを有効利用するためには、それが誰にでも使える状態として保持されている必要があり、データベースが果たす役割は非常に大きい。また、ビッグデータから有用な情報を取り出すために、コンピューターを用いた情報解析技術が必要不可欠である。本節では、ヒト細胞からのビッグデータに焦点を当て、主として研究が進展している遺伝子発現データと、それを利用した細胞解析について言及する。

1. ライフサイエンス研究分野におけるビッグデータ産出の現状

1985年にアメリカのエネルギー省によって開始されたヒトゲノムプロジェクトは、多くの国際的な研究機関の協力の下行われ、2003年にアメリカのホワイトハウスにてヒトゲノムの全塩基配列解読完了の報告がなされた。このときに用いられた技術がシーケンシング技術であり、現在では各種のゲノムプロジェクトによって、さらに多くのモデル生物種のゲノム配列解読が完了、もしくは進行中である。ゲノム配列データはアメリカ、ヨーロッパ、そして日本を拠点として運営されている国際DNAバンクに登録されている。2012年の時点でそのデータサイズは約540 GBにまで上り、さらに指数的な増加を続けている。また、2000年以降はゲノムプロジェクトの成果として生命システムの構成要素が明らかになってきたことで、それまでに行われていた断片的な材料を用いて現象の一部のみを観察する方法よりも、生命システム全体を包括的に解析する方法(オミクス解析)の有用性が重視された研究が多くなったのではないだろうか。そこで以下では、現在のライフサイエンス分野研究におけるビッグデータの産出技術の一例と、それらを取り巻く状況について述べる。

1.1 現代のライフサイエンス研究とオミクス解析技術

ゲノム配列情報の取得は生命現象の研究において非常に重要なものであるが、生物の複雑かつ精密な機能制御の理解のためには、ゲノム配列解読だけでは不十分である。現在までの研究で体内位置情報を区別して計算すると約2,600種類あるとされているヒト細胞¹⁾のうち、細胞1種類の機能解明のためにも、ゲノム配列情報に加えて以下のような詳細な階層、構成成分の網羅的な解析が必要になる。

- ・分子単体レベルでの情報(遺伝子・タンパク質発現情報:配列・構造・時空間的制御)
- ・分子ネットワークレベルでの機能情報(分子間相互作用, パスウェイ, リアクトーム)
- ・細胞レベルの機能情報(代謝制御, フェノタイプ発現)
- ・組織・器官の一部としての機能

これらのデータを効率よく取得するため、現在ではシーケンシング技術だけではなく、オミクス研究のための様々な技術の開発がなされている。また、検出やデータ解析技術のハイテク化、アメリカの国家プロジェクト「1000ドルゲノム計画」に代表されるように、ハイスループットなゲノムシーケンシング技術が安価になっており、ビッグデータを産出する研究が以前よりも盛んに行われる環境になっている。表1において、細胞の機能解析に用いられる手法の代表的なものを示す。これらを組み合わせることで、研究対象となる細胞固有のシステムや生体におけるその機能の網羅的な解析が可能となる。表1の中でも、近年は特に、RNAシーケンシング(RNA-seq)やバイサルファイトシーケンシング等の技術を用いて、個々の細胞の持つ疾病関連状況や、iPS細胞、幹細胞の分化の程度を網羅的に調べる研究が盛んに行われている。

表1 細胞機能解析に用いられる手法の一例

| 実験デザイン | 解析技術の一例 |
|--------------|-----------------------------------|
| ゲノム解析 | ゲノムシーケンシング |
| エピゲノム解析 | マイクロアレイ, ゲノムシーケンシング, ChIP シーケンシング |
| トランスクリプトーム解析 | マイクロアレイ, RNAシーケンシング |
| プロテオーム解析 | タンパク質マイクロアレイ, 2D-MS, LC-MS |
| インタラクトーム解析 | LC-MS, 酵母-two hybrid 法, FRET/BRET |
| メタボローム解析 | イメージング質量分析法 |
| フェノーム解析 | イメージング |

1.2 ライフサイエンス研究ビッグデータと情報統合技術

ハイスループットな解析技術は、以前まで特定の研究機関や創薬研究開発分野など、限られた施設でのみ行うことが可能であったが、現在はそれらの技術の確立やコスト面での問題が解決されつつあるため、様々な研究で用いられるようになった。そのため、ライフサイエンス研究分野全体のアウトプットとしてビッグデータの産出が一般的になりつつある。また、データ量だけでなく、例えばトランスクリプトーム解析のようにRNA-seqとマイクロアレイなど、データを得るために用いる実験、計測技術が異なる場合もあるため、これらの実験デザインの情報も保持した上で、ビッグデータを効率的に取り扱うツールをどう設計するかが現在の大きな課題である。ライフサイエンス研究分野におけるデータベースシステムはまさに、あるカテゴリーの実験系で得られるビッグデータの収集、管理、情報の検索、そしてその利用という作業を容易に実現させるためのツールであり、そのビッグデータ処理における重要性は日に日に高まっている。例えば、遺伝子発現解析データのための公共データリポジトリ・データベースとして、NCBIのGene Expression Omnibus、EBIのArrayExpress、国立遺伝学研究所のOmics Archiveがあり、これらの登録実験データ数は合計で140万件以上にものぼる(2014年5月現在)。これら以外にも、遺伝子発現解析研究に関する大量のデータを取り扱うものに限定しても、世界中で数多くのデータベースが開発されている(表2参照)。

表2 ヒト細胞における遺伝子発現データを取り扱うデータベースの一例

| 名称 | 公開状況 | 開発機関 | URL |
|------------------------------|--------------------|----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| Gene Expression Omnibus | Public | NIH | http://www.ncbi.nlm.nih.gov/geo/ |
| ArrayTrack | Public and private | FDA | http://www.fda.gov/ScienceResearch/BioinformaticsTools/Arraytrack/ |
| ArrayExpress | Public | EMBL-EBI | https://www.ebi.ac.uk/arrayexpress/ |
| GeneInvestigator | Private | NEBION | https://www.geneinvestigator.com/gv/ |
| Stanford Microarray Database | Public and private | Stanford University School of Medicine | http://smd.princeton.edu/index.shtml |
| GeneNetwork system | Public | University of Tennessee | http://www.genenetwork.org/webqtl/main.py |
| GeneCards | Public | Weizmann Institute of Science | http://www.genecards.org |
| DDBJ Omics Archive | Public | 国立遺伝学研究所 | http://trace.ddbj.nig.ac.jp/dor/index.html |
| RefExA | Public | 東京大学先端科学技術研究センター | http://157.82.78.238/refexa/main_search.jsp |

しかしながら、多くのデータベースは異なる組織によって運営されているため、登録されているプロジェクトデータの重複の有無がある。また、データベースごとに異なるIDやフォーマットを用いてデータを登録するという方式のため、登録項目や使用されている語彙が統一されていないという場合が多い。このようなことから、異なるデータベース間での内容の比較やデータ解析が困難という問題も生じている。この問題を解決し、データの共有や再利用を促進し付加的な情報を引き出すために、上記3つの遺伝子発現解析データベースを始めとした様々なデータベースの統合プロジェクトが、既に国家規模で立ち上げられ取り組まれている。さらに、データベースの統合化の必要性が増すにつれて、実験データ報告における必要最小限の情報項目チェックリスト(The minimum information standards, MI standards)²⁾や制御さ

れた語彙を用いるオントロジー分野も注目されるようになってきた。

MI standards は、実験の全体像の形式的な記述や報告を可能にする、また実験条件を形式的に管理し実験データの再現性を高める必要最小限の情報を統一する、という目的で提案された報告用チェックリストである。最初の MI standards の例としては、マイクロアレイデータの報告のために作成された Minimum Information about a Microarray Experiment (MIAME) チェックリスト³⁾があり、異なる研究機関で行われた実験データやメタデータを統合する際のテンプレートとして利用されている。また、MIAME を始めとした MI standards に基づいた形式でのデータ、メタデータの投稿を推奨する学術系雑誌もある⁴⁾。2008 年には生命医科学や生物学の分野で必要最小限の情報を統一する Minimum Information about a Biomedical or Biological Investigation (MIBBI) プロジェクト (<http://www.biosharing.org/standards/mibbi>) が国際協力の下発足され、現在までに様々な MI standards が作成されている。MI standards では実験データに付随するメタデータの報告項目を対象となる研究分野や実験技術ごとに定めており、特にビッグデータを産出するハイスループット実験技術を用いて行う研究の情報管理に有用であるとされている。さらに、MI standards はある特定の研究分野や手法でデータを得るための重要な要素を項目として含むので、データベースのフレームワーク構築にも有用であると考えられている。

オントロジーは、ある知識領域における様々な概念を構成する語彙の体系的な分類や、それらの語彙の共有を可能とする。例えば、以前まで新たに発見された遺伝子は、シーケンシングの結果で得られた配列や構造の類似性から命名または分類されることが多かった。しかし、ゲノムプロジェクト以降、研究開発技術の発達に伴い膨大な数の遺伝子配列およびその機能が明らかになり、その蓄積された知識に基づいて遺伝子を分類し直すために構築されたものが、よく知られている Gene Ontology (GO) である⁵⁾。すなわち、GO は様々な生物種の遺伝子機能に関する知識やその記述に必要な共通語彙 (GO term) を統一された形式で分類し有している、遺伝子機能の辞書のようなものである。さらに、現在のライフサイエンス分野のデータ解析ソフトウェアやデータベースには、共通語彙を用いることで異なるツール間の知識表現の統一化や比較化を容易にするという目的で、解析や検索のアウトプットに GO term を使用するものも多い。このことから、オントロジーの有用性に期待し、近年では他のライフサイエンス研究分野のオントロジーの作成が進められている。

ビッグデータは活用されてこそ価値がある。今後は次々と新しいデータを生み出してだけでなく、既にあるビッグデータを生物学的な意味のある知識として蓄積し、応用研究や新たな基礎研究のアイデア取得のためのリソースとしてリサイクルしていくことが重要である。そのためには、情報や知識が統一された形式で表示され、誰もが共有できるシステムが必要である。現在は、上記の MI standards やオントロジーを組み合わせた MAGE-TAB (<http://www.mged.org/mage-tab/>) や ISA-TAB (<http://isatab.sourceforge.net/index.html>) のようなアッセイノテーションフォーマットも作成されている。このような環境を整備し利用することができれば、多くの研究グループにより膨大な時間とコストをかけて産出されたビッグデータおよび大量の知識が様々な研究グループにより再度活用されることが可能になり、ライフサイエンス研究の促進的な発展が期待できる。

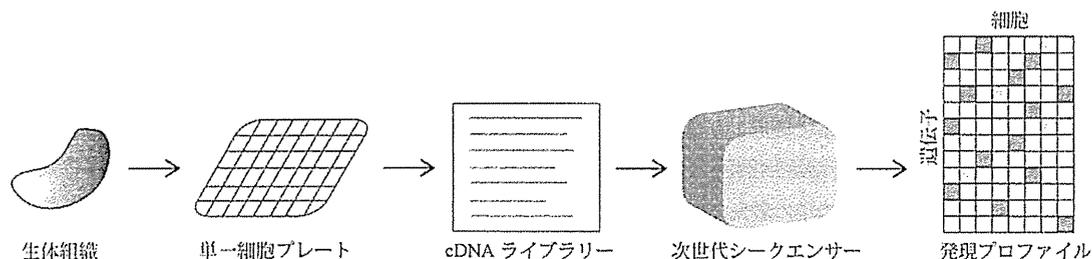
2. 細胞の自動分類技術

前述のように、ヒト細胞からはトランスクリプトームなどの大量のデータが得られる。これらのビッグデータから意味のある情報を抽出し、細胞から構成される生命の謎に迫るためには、情報科学、統計科学などに裏打ちされた方法論およびコンピューターの能力を駆使して網羅的な解析を行うことが重要である。以下では、近年注目を浴び始めている単一細胞解析を用いた細胞分類問題に焦点を当て、用いられている実験技術および情報技術の具体例を紹介し、今後の展開について述べる。

2.1 RNA-seq を用いた単一細胞解析

近年では細胞解析を行う際、細胞の集団レベルで遺伝子発現量を考察するよりも、より解像度の高い単一細胞レベルで遺伝子発現量解析を行うことが有効と考えられている。生体の組織から単一細胞レベルでトランスクリプター

ム解析を行う手法として、RNA-seq⁶⁾を利用するアプローチが知られている（図1参照）。まず、対象とする生体の組織から、FACS (fluorescence-activated cell sorting) などで行われる細胞分取法を用い、組織の構成要素である単一細胞に分離する。次に、分離された各細胞から発現している RNA を RNA-seq により配列解析するが、複数の単一細胞からの RNA を同時に大量解析するため、どの細胞から得られた RNA なのかを記憶するために、核酸配列にバーコードを埋め込むなどの工夫が必要である⁷⁾。RNA-seq で転写配列断片の定量的情報により得られた遺伝子発現の全体像を発現プロファイルと呼んでいる。

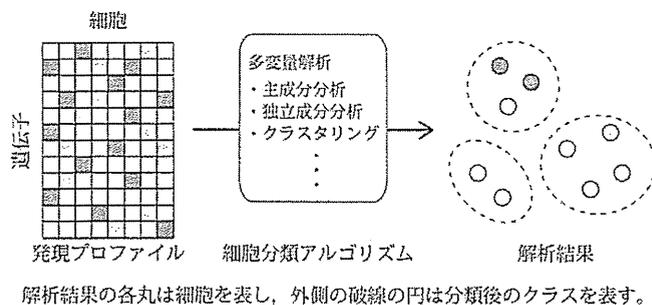


図の cDNA ライブラリーの各横線は各細胞内で発現している RNA から作成した cDNA を表す。遺伝子発現プロファイルでは、行と列がそれぞれ遺伝子と細胞を表し、各マスの色が濃く塗りつぶされているほど発現量が多いことを示す。

図1 単一細胞 RNA-seq 解析

2.2 発現プロファイルからの細胞の自動分類

発現プロファイルにおいて、各細胞は遺伝子の個数だけ次元を持つ実数値ベクトル（多変量）とみなすことができる。例えばヒトの場合、約 22,000 個の遺伝子があると推定されているため、その全てを考慮した場合の発現空間における各細胞は 22,000 次元のデータとなる。したがって、細胞情報解析を行うに当たって、高次元データの集合からデータ間の隠れた構造の導出をめざす多変量解析が有効であると考えられる（図2参照）。ただし、細胞（入力）に対する出力値（教師出力）が与えられていないことが多いため、入力群から必要な情報だけを抽出する必要がある。このとき、低次元の実数値ベクトルだけを出力として考えると良い場合があり、代表的な手法として主成分分析がある。また、出力としてラベルなどの離散値を考えて分類するクラスタリングもしばしば行われる。しかしながら、現実には入力データには様々な因子が混在しており、主成分分析などのみで隠れた構造をあぶり出せることは稀である。以下では、ごく最近発表された2件の論文で行われている組織特異的な細胞分類手法を紹介する。両者は、解析の最初の段階で細胞データの次元圧縮を行って扱いやすくしている点において共通しているが、その後の処理において、前者は確率モデルを扱い、後者は離散的なグラフモデルを扱っている点異なる。



解析結果の各丸は細胞を表し、外側の破線の円は分類後のクラスを表す。

図2 細胞の自動分類

2.2.1 脾臓における細胞の分類

Jaitin らは大量並列単一細胞 RNA-seq 解析と教師なし分類アルゴリズムを利用して、マウスの脾臓中の血液細胞をその種類によって分類することに成功している⁸⁾。分類アルゴリズムでは、発現プロファイルから発現量の分散の大きな

遺伝子のみを考慮した細胞データに対し、あらかじめ階層的クラスタリングで推定した個数の細胞クラスごとに異なる多項分布を仮定する。次に、そのクラスの下で各細胞を出力する確率分布のパラメータを発現プロファイルから推定する。その後、各クラスが各細胞を出力する確率の対数（対数尤度）をもとに、事後確率が最大となるクラスタリングを求めることによって細胞分類を行っている。

2.2.2 筋芽細胞における細胞の分類

Trapnellらは単一細胞 RNA-seq 解析を応用してヒトの筋芽細胞の時系列遺伝子発現解析を行っている⁹⁾。ここでは、複数時点での RNA-seq データを用いることで、トランスクリプトームのダイナミクスを考慮していることに注意されたい。具体的に、分化による細胞の発展時系列を以下のような教師なしアルゴリズムで計算している。まず、発現プロファイルから独立成分分析を用いて細胞の遺伝子発現データの次元圧縮を行う。次に、圧縮された発現空間上において、各細胞をグラフの頂点、各頂点間の距離を辺の重みとみなしたときの重みつき完全グラフを考え、その最小全域木を計算する。そして、最小全域木の中の最長パスを探索し、パスを構成する頂点系列（細胞群）を転写に関連する類似の細胞とみなし、それらは分化を通じて発展したものであると判定する。

2.3 Non-coding RNA と細胞情報解析

これまで RNA の全体を表すトランスクリプトームの解析に主眼を置いたが、これらは原理的にタンパク質に翻訳されない non-coding RNA も含んでいる。ヒトのような高等生物では多種多様な non-coding RNA の制御機構が複雑な生命システムを維持していると考えられており、これまで述べてきた遺伝子発現量は non-coding RNA の遺伝子発現制御機構による影響を強く受けていると思われる。そのため、細胞情報解析を行うにあたり、RNA-seq などにより得られた遺伝子発現量という情報のみを考慮するのではなく、遺伝子発現現象の背後にある non-coding RNA の分子機能や他の生体分子との相互作用を併せて多角的に考察することが、今後精度の良い解析結果を得るために必要になってくると思われる。具体的に、現在までに大量配列解析に適した高速 RNA 配列情報解析ツールセット¹⁰⁾などが提供されているので、それらとの連携を模索することが新たな知見を得る手がかりとなるかもしれない。

2.4 細胞の周期表

単一細胞解析と細胞分類から得られる結果により、そもそも細胞は何種類存在するのか、という問題を再考することができる。一般に、細胞は 200 ~ 400 種類あると言われているが、遺伝子発現量といった細胞の内部状態を考慮すると、より精緻な細胞のカテゴリーを示すことができる可能性がある。化学における構成単位が原子であり、その周期表が存在するように、生物の構成単位である細胞にも周期表に相当するものが存在するかもしれない。周期表の存在の有無に対する初等的なアプローチとして、当研究室では、マイクロアレイで取得したヒトの約 3,000 細胞の遺伝子発現量から得られるヒートマップを用いて細胞を自動的に分類し、分類後の細胞を系統的に並べ表にした結果、組織ごとに細胞のクラスターが形成されていることを確認している (<http://shogoindb.stemcellinformatics.org/>)。これは細胞間の関係において周期表のような細胞間の規則を予想させる結果である。

細胞の辞書ともみなせる周期表を完成させるためには、細胞の法則などを記述する数理モデルを推定する必要がある。細胞を支配する法則の解明のために、ビッグデータを利用した網羅的な細胞情報解析が鍵を握るのは想像に難くない。細胞周期表が完成した暁には、工学的に効率良く細胞を設計することが可能になるなど、その応用範囲は広いと予想される。

2.5 現状の問題点と今後の展開

上で紹介した内容は発展途上の研究分野であり、現状では様々な問題点がある。まず、単一細胞解析から得られる RNA の収量が非常に小さく、配列解読に至るまでに技術的なノイズが混入する可能性が高い。また、細胞の中で起きる遺伝子発現は環境によっても変化するもので、生物学的な揺らぎのある確率的な事象と考えるべきである。したがって、これらの因子を柔軟に扱える数理モデルを開発することが必要であろう。また、本節では主にトランスクリプトームに

よる解析について言及したが、DNAのメチル化やヒストン修飾を代表とするエピジェネティクスも細胞を理解する上で重要な要素となるため、これらを扱えるように解析手法を発展させていくことが求められる。さらに、これまでの解析法は組織ごとに着目して行われていたが、将来的には生物そのものをまるごと細胞解析することで、細胞レベルで見た生物の設計図を一度に作るができるようになるかもしれない。

おわりに

遺伝子発現量などのビッグデータを用いて細胞解析を行うと、これまで見えてこなかった細胞間の関係性が見えるようになる。すると、我々はまだ細胞のごく一部を知っているに過ぎないのではないかと認識させられる。今後、多種多様のビッグデータが量産されていくと考えられるので、既成の枠にとらわれないデータベースの柔軟な構築とともに、それに連携する高速解析ツールの開発が活発に行われるであろう。これらの技術の創出により、細胞の詳細、ひいては周期表といった細胞の規則が分かるだけでなく、コンピューターを用いて細胞を設計する時代の到来を予感させる。

文 献

- 1) Hatano A et al., Database, 2011, bar046 (2011)
- 2) Taylor FC et al., Nat. Biotech., 8, 889-896 (2008)
- 3) Brazma A et al., Nature Genetics, 29, 365-371 (2001)
- 4) Opinion: Microarray standards at last. Nature, 419, 323 (2002)
- 5) Gene Ontology consortium. Nat. Genetics, 25, 25-29 (2000)
- 6) Tang F et al., Nat. Methods, 6, 377-382 (2009)
- 7) Islam S et al., Nat. Protoc., 7, 813-828 (2012)
- 8) Jaitin DA et al., Science, 343, 776-779 (2014)
- 9) Trapnell C et al., Nat. Biotechnol., 32, 381-386 (2014)
- 10) Kato Y et al., Nucleic Acids Res., 40, W29-W34 (2012)

単細胞技術に基づく iPS 細胞の標準化

山根 順子 丸山 徹 藤 洸

これまでの生物学の常識を大きく覆した人工多能性幹細胞 (iPS 細胞; induced pluripotent stem cell) の発見がなされたのが 2006 年のことである¹⁾。iPS 細胞は ES 細胞と異なり作製段階にヒト胚を破壊する必要がないことから、ES 細胞を用いた研究において大きな障壁となっていた倫理問題が生じず、再生医療を一気に加速させる夢の細胞として登場した。体細胞にわずか数因子を導入するのみで多能性を持った細胞を生み出すことができるという報告はあまりにセンセーショナルであり、それ以降様々な細胞種由来の iPS 細胞の樹立や、より安全かつ効率的な樹立法が次々に見いだされた。また、iPS 細胞を用いた幹細胞生物学としての基礎研究や再生医療、創薬へ向けた応用研究など、多数の報告がなされている。世界中で樹立が試みられ報告されている iPS 細胞は、樹立された数だけ質の異なる細胞になっている可能性が指摘され、今度は質の良い iPS 細胞を選別する手法を開発するという新たな研究の方向性も生まれた。

iPS 細胞は通常コロニー (細胞集団) として維持培養される。しかしながら、コロニーのなかでも均一な状態ではなく細胞の個性があることがわかっており、集団レベルで iPS 細胞の解析を続けていくだけでは標準化を目指すことは難しい。そこでわれわれはより解像度を上げた解析が必要に

なると考え、従来のような“細胞集団”として遺伝子発現レベルを調べるのではなく、“個”としての細胞、つまり“シングルセルレベル”での遺伝子発現を調べ、標準化に向けた試みを行った。

I. シングルセルトランスクリプトーム解析

2009 年に Tang ら²⁾によってシングルセルレベルで RNA-seq を行う方法が発表されてから、これまでに幾つかの手法が報告されてきた。われわれが行ったシングルセルトランスクリプトーム解析の手法は各細胞を判別するための DNA バーコードを template switching により導入する STRT 法³⁾であり、同時に大量のシングルセルを解析できる利点がある。この方法では、mRNA の 5' 末端に DNA バーコードを導入するため、次世代シーケンサーで読まれたリードは 5' 末端側の配列であるという特徴がある。その他の手法で最近報告されたものでは、SMART-seq⁴⁾や CEL-seq⁵⁾、Quartz-seq⁶⁾などがあり、そのうち CEL-seq、Quartz-seq は 3' 末端側のバイアスがあることが知られている。加えて、これまで問題とされてきたデータの精度についても、Quartz-seq では細胞周期が区別できるまで改善されており、この先シングルセルトランスクリプトーム解析が大幅に普及することが予想される。

Standardization of iPS cells by single-cell transcriptome analysis

Yamane Junko : 京都大学 iPS 細胞研究所 (特定研究員 : 〒 606-8507 京都市左京区聖護院川原町 53)

Maruyama Toru : 早稲田大学大学院 先進理工学研究科 生命医科学専攻 (修士課程), 京都大学 iPS 細胞研究所 (特別研究学生)

Fujibuchi Wataru : 京都大学 iPS 細胞研究所 (教授)

0370-9531/04/ ¥500/論文/JCOPY

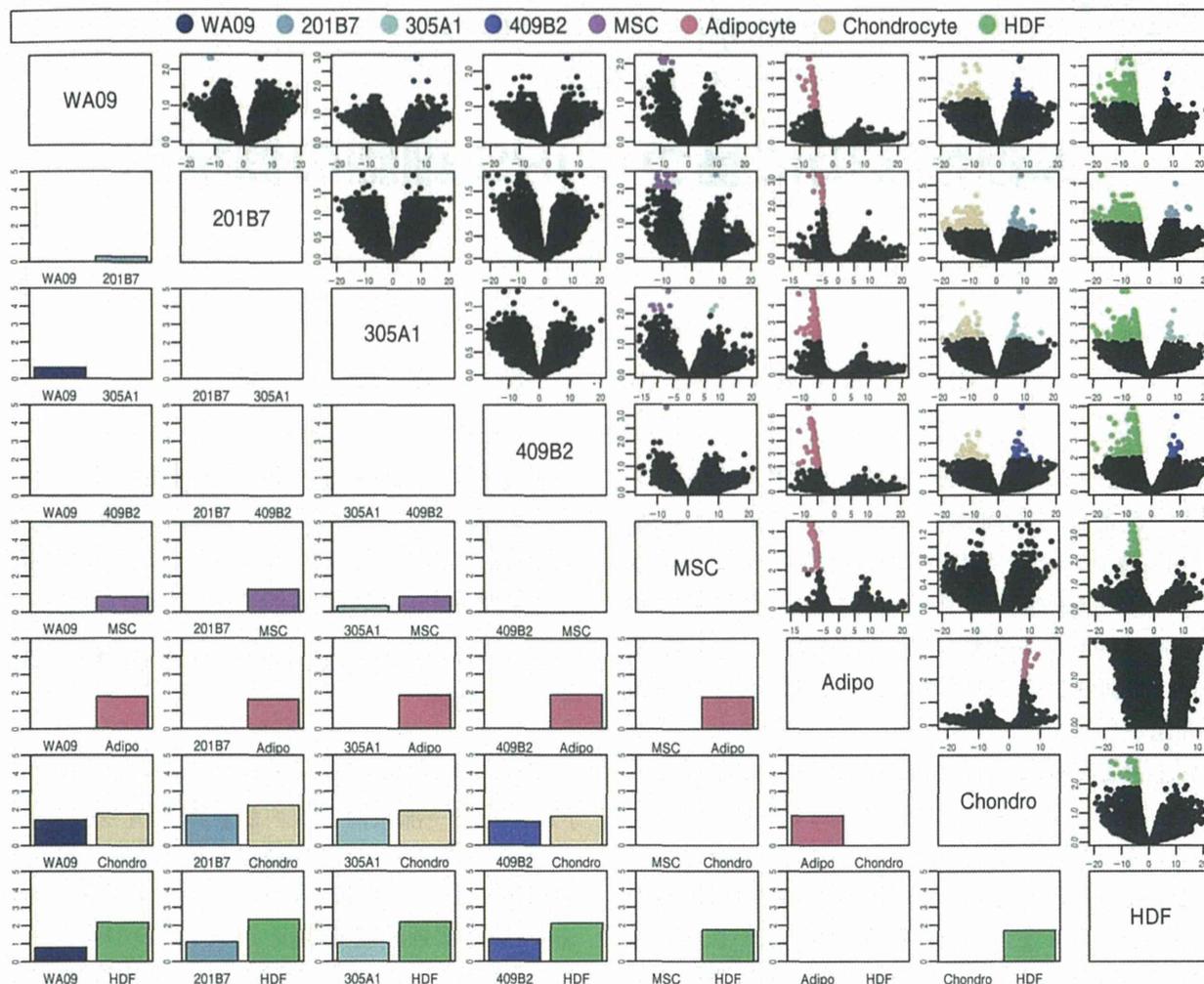


図 1 DEG 解析

対角線より上：Volcano plot (横軸： $\log_2(FC)$ ，縦軸： $\log_{10}(FDR)$)，色がついたプロットは有意差($FDR < 0.01$)があった遺伝子。対角線より下：DEG の数を示す棒グラフ。プロット・棒の色は細胞種に対応している。例えば赤いプロットは Adipocyte で発現が高い遺伝子を，赤い棒は Adipocyte で発現が高い遺伝子の数を表す。

II. シングルセルトランスクリプトーム解析により見えてきた特徴

われわれは異なる手法によって樹立された iPS 細胞 3 種類 (201B7, 305A1, 409B2) に加えて ES 細胞 1 種類 (WA09)，ヒトの体細胞 4 種類 (MSC：間葉系幹細胞，Chondro：軟骨細胞，adipo (誘導させた) 脂肪細胞，HDF：皮膚線維芽細胞) の計 8 細胞種を対象にしてシングルセルレベルのトランスクリプトーム解析を行った。

まず，Bioconductor の edgeR パッケージを用いて 2 細胞種間で真の発現変動遺伝子 (DEG) の数を調べたところ，ES 細胞と iPS 細胞間で発現変動がある遺伝子はほとんど見つからなかった

(図 1)。しかし，本解析は細胞集団を対象にした手法を適用しているため，シングルセルレベルでの解析とは結果が異なる可能性がある。

そこでシングルセルレベルでの各種細胞における遺伝子発現パターンの特徴をより詳細に調べるため，8 細胞種の主成分分析 (PCA) を行ったところ，おおよそ細胞種ごとにクラスターを形成していることが確認できた (図 2)。この結果から軟骨細胞と HDF のクラスターが重なっていることがわかったが，これらの細胞種間では遷移が起こることが知られている⁷⁾。よって，単一細胞のトランスクリプトームには細胞系譜における細胞間の関係性や，細胞間で遷移が起こりうる可能性といった新たな情報が含まれている可能性が示唆

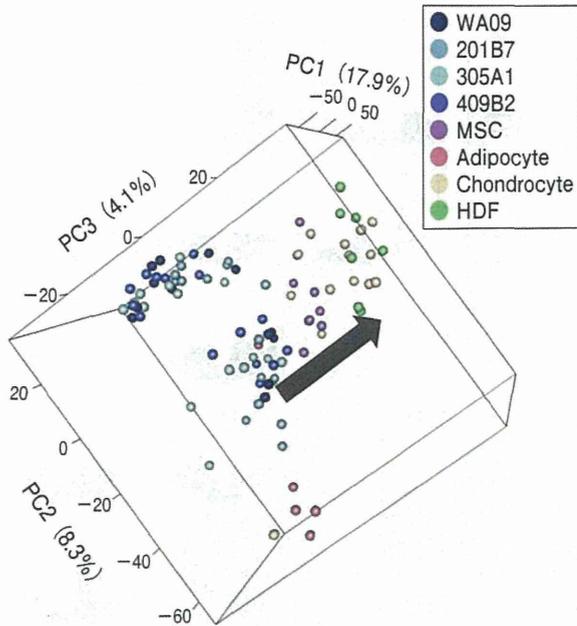


図 2 8細胞種の主成分解析(PCA)

細胞ごとにある程度クラスターができています。括弧内は各種成分の寄与率を示す。HDFとChondrocyteが混在しているが、これはHDFとChondrocyteの間でトランジションが起りやすいことを示唆していると考えられる。また、矢印の方向に相転移が起りうることを示唆している。

された。これまでの集団として解析してきた方法では平均化されてしまい見えていなかった細胞本来のばらつきの様相の一端が明らかになった。

また、遺伝子発現の揺らぎの大きさを評価するために、細胞種ごとに全遺伝子の変動係数(CV)を調べた(図3)。細胞集団での解析と一致して、すべての細胞でハウスキーピング遺伝子のCVは他の遺伝子と比較して遺伝子発現の揺らぎが小さいという結果が得られた。微量なサンプルを対象とするシングルセル解析を行ううえで、得られた結果が細胞由来の揺らぎによるものか、テクニカルな問題により生じる差を見ているのかを見極めることは非常に重要である。よって、指標の一つとなるハウスキーピング遺伝子の発現の揺らぎが他の遺伝子と比べて小さいという結果はシングルセルを扱う研究を進めるにあたり精度の良さを確認するための大きな情報となり得る。

最後に、全遺伝子の変動係数の解析において今回調べた細胞間でヒストグラム分布の異なる遺伝子について一部報告する。まず、DNAメチル化にかかわるDNMT3Bは既報のように⁸⁾、ES細胞、iPS細胞の多能性幹細胞群ではほぼ共通し発現レ

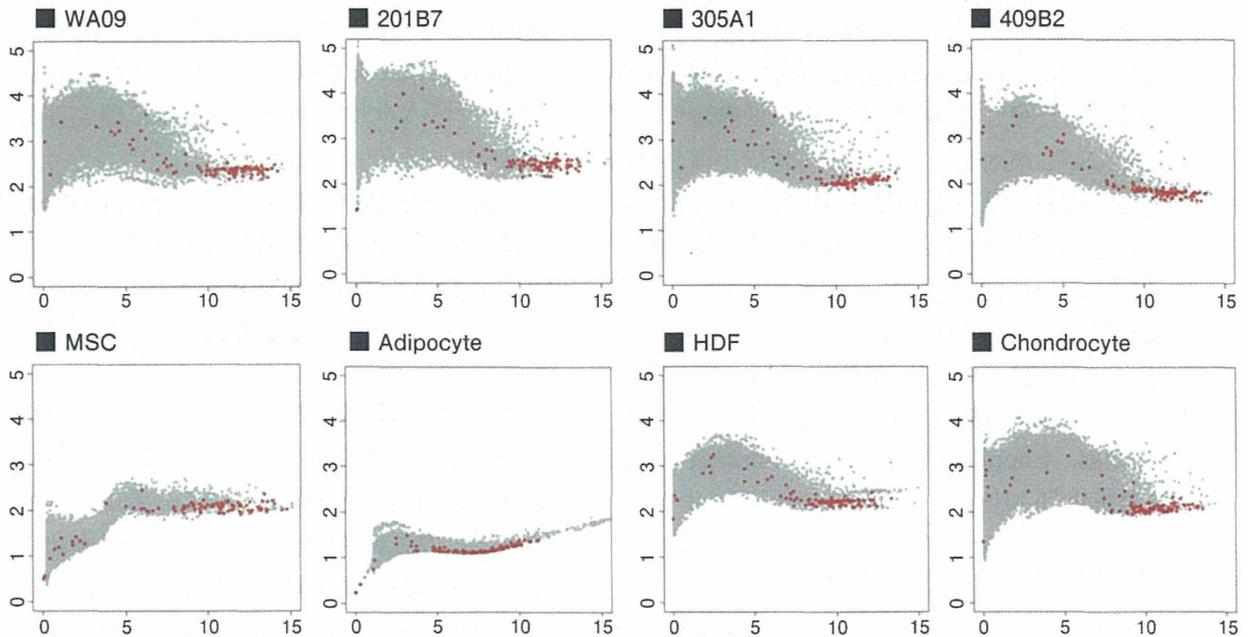


図 3 全遺伝子の変動係数(CV)

各点は遺伝子に対応しており、赤点はハウスキーピング遺伝子に対応している。横軸は $\log_2(\text{RPM})$ の平均、縦軸はedgeRによって算出した発現の変動(biological coefficient of variation)を表している。下に位置する遺伝子ほど安定に発現していることを意味する。

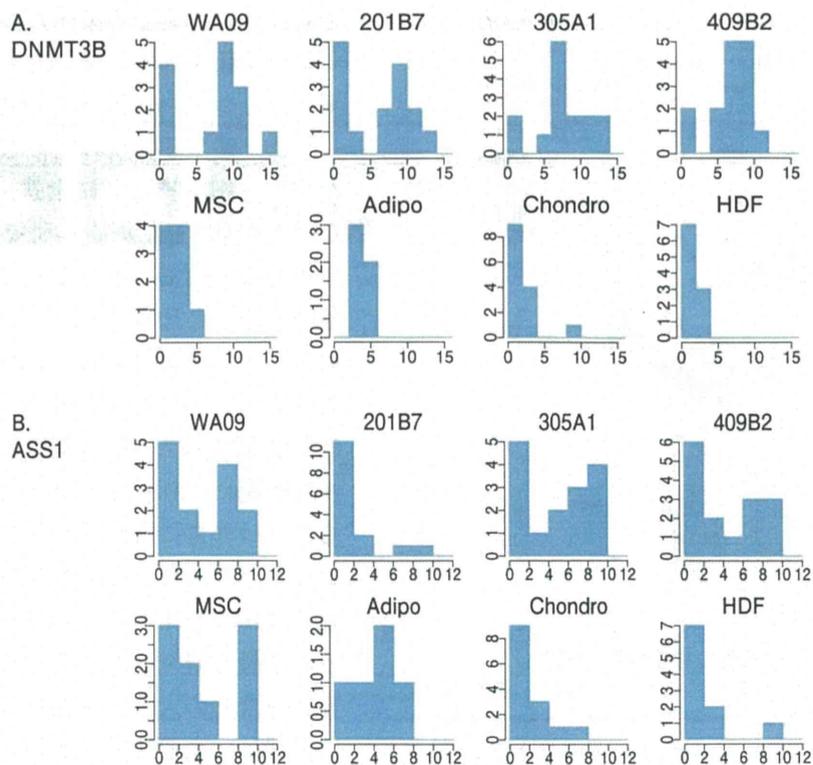


図 4 ヒストグラムの分布で差が見られた遺伝子例(横軸は $\log_2(\text{RPM} + 1)$)

A. は多能性幹細胞と分化細胞の間で顕著な差が見られた。B. は多能性幹細胞の中の一部の株(201B7)で異なる分布が見られた。

ベルの高い分布が見られ、逆に分化細胞ではいずれも発現レベルの低い分布が傾向として見られた(図 4A)。また、アルギニン生合成経路にかかわる酵素の一つである ASS1 は多能性幹細胞群の中で比較をした場合、201B7 株でのみ低発現の分布を示している傾向があった。この遺伝子は分化細胞でも比較的低発現の分布を示している(図 4B)。このように平均で解析していると見えてこなかった遺伝子発現の分布が、個々のレベルで見ると差が見えてくるケースがある。

以上のように、われわれはシングルセルトランスクリプトーム解析によって、これまでの細胞集団を対象にする解析では見えなかった事象を確認することができた。

おわりに

シングルセル RNA-seq 解析が普及するに伴い、これまで 250 種類程度と言われてきた細胞がより細かく分類されるようになる可能性がある。そこで、われわれの研究室では細胞のデータベ

スである“SHOGoin”を開発しており、細胞の遺伝子発現プロファイル、画像、形態計測情報、実験条件などのメタデータ、文献情報を貯蔵している(図 5)。ユーザーはウェブブラウザを通して本データベースにアクセスし、細胞に関する情報を検索することができる。

今後、われわれはシングルセル RNA-seq 解析と体系的に貯蔵された細胞のデータを用いて、細胞に普遍的に存在する理論を見つけることを目指している。その一つとして次のような問題に取り組んでいる。

iPS 細胞を作製する際には、元となる体細胞に *Oct4*, *Sox2*, *Klf4*, *c-Myc* などの遺伝子を導入し、強制的に発現させる必要がある。しかし、すべての細胞に同じ刺激を与えているにもかかわらず、なぜ iPS 細胞の作製効率は悪く、質の良し悪しがあるのだろうか。われわれはこの問題に対して以下のようなモデルを考えている。遺伝子の発現は同じ種類の細胞であっても均一でなく、揺らぎがある。揺らぎは転写因子の結合や転写反応、

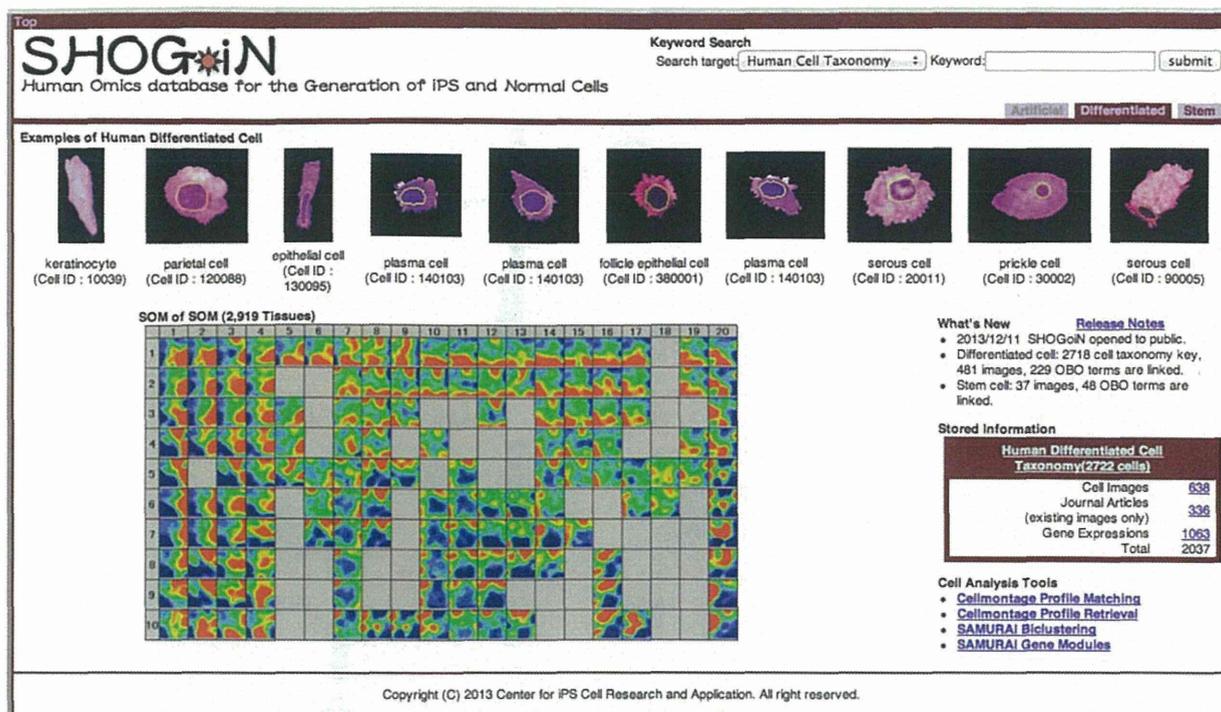


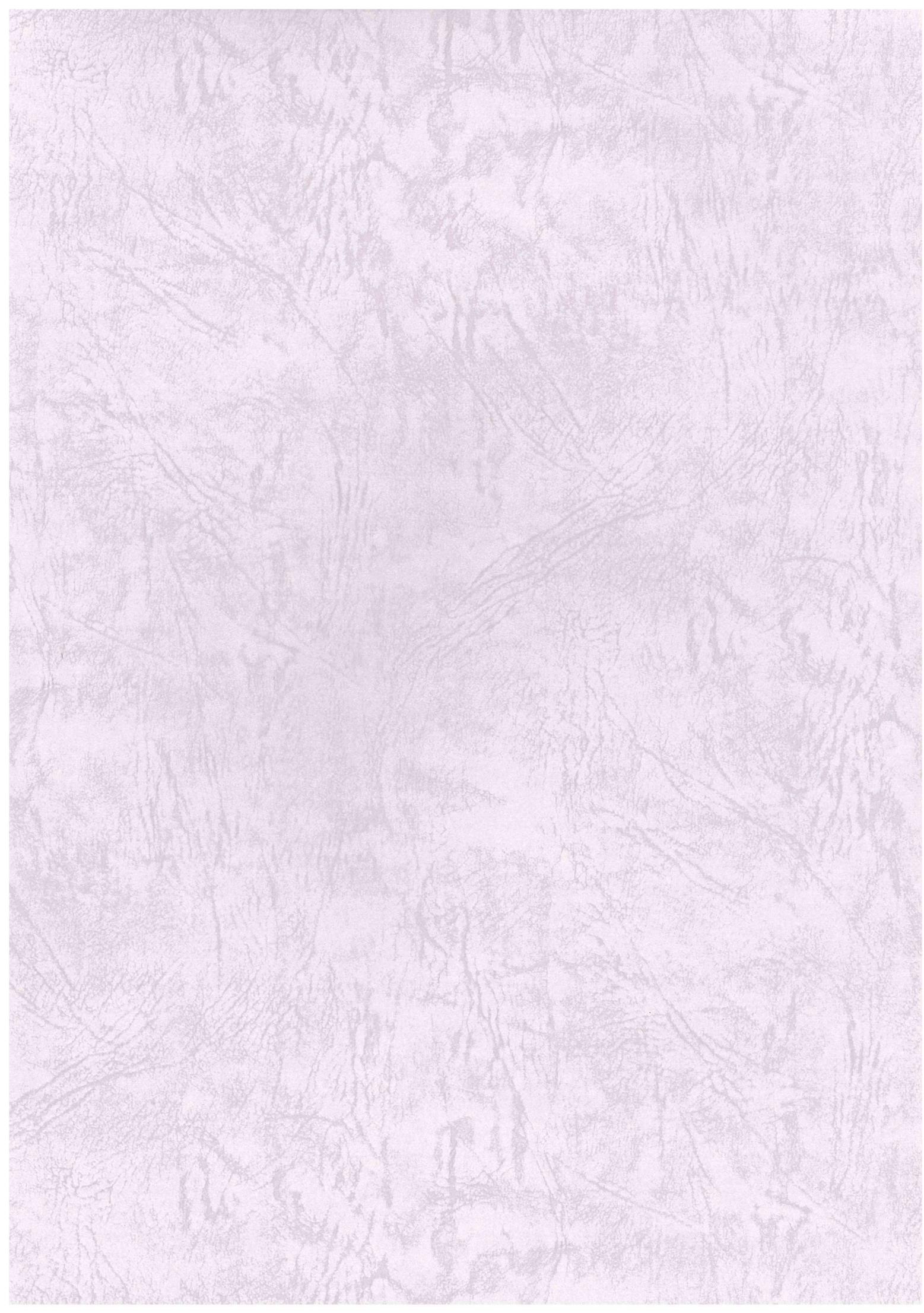
図 5 ヒト細胞情報統合データベース “SHOGoin”
<http://shogoindb.cira.kyoto-u.ac.jp>

RNA の分解といったイベントが確率的であることに起因している⁹⁾。Huang らの論文¹⁰⁾にもあるように、われわれはこういった遺伝子発現の揺らぎが細胞状態の変化に影響を与えているため、細胞分化や iPS 細胞の作製が確率的なものになっているのではないかと考えている。よって、揺らぎが細胞分化に与える影響を解き明かすことができれば、質の良い細胞をより効率的に作製する手法が明らかになるかもしれない。また、シングルセル解析により得られた詳細な情報をデータベースに集約し、広く研究者らが利用することで、iPS 細胞の標準化や、そこから分化させた各種分化細胞の標準化に向けた足がかりとなることを期待している。

謝辞 本研究は『内閣府最先端研究開発支援プログラム (FIRST プログラム)』の支援のもと行いました。

●文献

- 1) Takahashi K, Yamanaka S : *Cell* 126 : 663-676, 2006
- 2) Tang F, Barbacioru C, Wang Y et al : *Nat Methods* 6 : 377-382, 2009
- 3) Islam S, Kjällquist U, Moliner A et al : *Genome Res* 21 : 1160-1167, 2011
- 4) Ramsköld D, Luo S, Wang Y-C et al : *Nat Biotech* 30 : 777-782, 2012
- 5) Hashimshony T, Wagner F, Sher N et al : *Cell Rep* 2 : 666-673, 2012
- 6) Sasagawa Y, Nikaido I, Hayashi T et al : *Genome Biol* 14 : R31, 2013
- 7) Outani H, Okada M, Yamashita A et al : *ProS One* 8 : e77365, 2013
- 8) Takahashi K, Tanabe K, Ohnuki M et al : *Cell* 131 : 861-872, 2007
- 9) Wills QF, Livak KJ, Tipping AJ et al : *Nat Biotechnol* 31 : 748-752, 2013
- 10) Chang HH, Hemberg M, Barahona M et al : *Nature* 453 : 544-547, 2008



201406021A (2/2)

厚生労働科学研究費補助金

再生医療実用化研究事業

ヒト幹細胞を用いた再生医療の
臨床実用化のための基盤構築に関する研究

平成26年度 総括・分担研究報告書

研究代表者 中井 謙太

平成 27 (2015) 年 3 月

2 / 2 冊

目 次
(2 / 2冊)

V. 資料

| | | |
|-----------------------------------------|-------|-----|
| 1. 中核拠点 基盤システム運用活動報告 | ----- | 163 |
| 2. サーバ移設関連資料（札幌データセンター→中核機関） | ----- | 169 |
| 3. 研究班会議資料 | ----- | 177 |
| 4. システム説明会資料 | ----- | 181 |
| 5. 開発ソフト資料 | ----- | 185 |
| 1) 実験データの共有とアプリケーションの集約に向けた統合システムの開発 | | |
| 2) 再生医療における電子標準作業手順書作成・管理・運用支援システムの次世代化 | | |
| 3) ユニバーサル実験画像ビューワの開発 | | |
| 6. システム運用・保守資料 | ----- | 413 |
| 1) データ共有、ログ収集・管理システム及びメール環境の構築 | | |
| 7. セキュリティ対策資料 | ----- | 469 |
| 8. 倫理申請手続資料 | ----- | 517 |

V. 資料

1. 中核拠点 基盤システム運用活動報告

- 1) 基盤システム利用 wiki (画面キャプチャ)
- 2) 基盤システム管理者用 wiki (画面キャプチャ)
- 3) セキュリティ対策状態監視 (画面キャプチャ)
- 4) サーバ・ネットワーク状態監視 (画面キャプチャ)



幹細胞関連情報の基盤システム wiki

http://.../wiki/index.php?%B4%B4%BA%D9%CB%A6%B4%D8%CF%A2%BE%F0%CA%F3%A4%CE%B4%F0%C8%D7%A5%B7%A5%B9%A5%C6%A5%E0%20wiki

[トップ] [編集 | 凍結 | 差分 | バックアップ | 添付 | リロード] [新規 | 一覧 | 単語検索 | 最終更新 | ヘルプ]

[Menu]

- 幹細胞関連情報の基盤システム wiki

[全般]

- 基盤システムを使う
- デジタルペン
- VFPIにファイルをアップロードする
- 各種ツール
- 解析ソフト

[解析サーバ]

- VNCを使う
- 解析サーバからのFTP利用
- 文字化けを直す
- ウイルス対策ソフト
- 各種操作手順
- プロキシの設定
- 各種ドキュメント
- システム説明会

幹細胞関連情報の基盤システム wiki +

再生医療HWシステムの一般利用者向けの情報を公開し、互いに情報を共有することを目的としています。

ウイルス対策ソフトSymantec Endpoint Protectionの設定更新について +

各拠点のWindowsスレート端末にインストールされているウイルス対策ソフトSymantec Endoint Protection(SEP)を中核拠点で一元管理することとなりました。各拠点とも順次移行していく予定です。移行方法を以下に記します。

- ダウンロード、解凍して、SylinkDrop.exeを実行してください。
 [SepCommunicationUpdater.zip](#)
- VPRemoteExecutionStatus.xmlというファイルが作成されます。
中を見て、

```
<?xml version="1.0" encoding="UTF-8"?>
<Status Description="Sylink file has been successfully replaced." ExitCode="0"/>
```

となっていれば設定ファイルが更新されました。
SEPを立ち上げ、LiveUpdate...の文字が黒く活性化されていれば成功です。

なお、設定が反映されるまで数分～10分程度かかります。
しばらくの間はPCの電源をつけたまま置いておいてください。

お知らせ +

- 2015年2月24日(火)VFPIにファイルをアップロードするにWinSCPがうまく実行できない場合を追記しました。
- 2014年9月8日(月)トップページにWindowsスレート端末のSEPの設定変更方法をアップしました。



幹細胞関連情報の基盤システム管理者用 wiki

http://wikiadmin/index.php?%B4%B4%BA%D9%CB%A6%B4%D8%CF%A2%BE%F0%CA%F3%A4%CE%B4%F0%C8%D7%A5%B7%A5%B9%A5%C6%A5%E0%B4%C9%CD%FD%BC%D4%CD%D1%20wiki

[トップ] [編集 | 凍結 | 差分 | バックアップ | 添付 | リロード] [新規 | 一覧 | 単語検索 | 最終更新 | ヘルプ]

アクセス数
since 2013/06/05

- 総計:183
- 今日:1
- 昨日:1

Menu

- 幹細胞関連情報の基盤システム管理者用 wiki
- 各種操作手順
 - Open Grid Engine環境構築手順
- マニュアル
- 運用資料
- バックアップ
- Fortigate200B/100Dの時間設定について
- ネットワークドライブの割り当て
- デンペン利用記録
- 解析ソフト
- プロキシについて
- 研究分担者連絡先
- Zabbixによる監視について
- Subversionの運用について
 - Subversion環境構築手順
 - TortoiseSVNインストール手順
- FTPS環境構築
- 定期点検に関する段取りについて
- 稼働状態確認

幹細胞関連情報の基盤システム管理者用 wiki +

再生医療HWシステムの管理者向けの情報を公開し、互いに情報を共有することを目的としています。

最新のお知らせ +

- [2015-02-09]
 - Subversionの運用についてを追記しました。
- [2015-01-27]
 - 定期点検に関する段取りについてを追記しました。
- [2014-03-17]
 - Windows と Adobe Flash Playerの脆弱性が報告されています。修正プログラムの適用と拠点の周知徹底をお願いいたします。
- [2014-03-05] 拠点が追加されました。
 - 神戸理研
 - 京都大学iPS研
 - 京都大学再生研
 - 東大医科研病院

過去のお知らせはこちら → お知らせ

運用・保守情報 +

- [2015-01-29]
 - 理研、iPS研、再生研、慶應大の定期点検を実施しました。
- [2015-01-28]
 - 中核拠点、医科研病院、成育医療研究センターの定期点検を実施しました。
- [2015-01-27]

Symantec Endpoint Protection Manager

Symantec™ Endpoint Protection Manager 更新 ヘルプ ログオフ

クライアント

My Company
Default Group

Default Group ポリシーシリアル番号: 060F-12/02/2014 01:33:50 588

クライアント | **ポリシー** | 詳細 | インストールパッケージ

表示: 保護技術 全てのユーザーとコンピュータ フィルタ

| 名前 | 健全性の状態 | ログオンユーザ... | 前回のスキャン | ウイルス対策の状態 | ファイアウォールの状 |
|----|---------|---------------|----------------------|-----------|------------|
| | ▲ オフライン | reghw | ... 2015/02/26 12:10 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/03/10 11:59 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/03/03 16:34 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/03/12 0:30 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/03/10 11:49 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/03/11 3:19 | 有効 | 有効 |
| | ● オンライン | Administrator | ... 2015/03/13 8:35 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/03/13 8:47 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/02/27 8:43 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/03/12 15:01 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/03/12 0:30 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/02/12 13:13 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/02/16 11:30 | 有効 | 有効 |
| | ▲ オフライン | re | ... 2015/03/12 11:42 | 有効 | 有効 |
| | ▲ オフライン | reghw | ... 2015/03/13 9:39 | 有効 | 有効 |
| | ● オンライン | Administrator | ... 2015/03/13 7:35 | 有効 | 有効 |
| | ● オンライン | Administrator | ... 2015/03/13 6:37 | 有効 | 有効 |
| | ● オンライン | reghw | ... 2015/03/13 8:30 | 有効 | 有効 |
| | ● オンライン | Administrator | ... 2015/03/13 7:31 | 有効 | 有効 |
| | ● オンライン | Administrator | ... 2015/03/13 6:50 | 有効 | 有効 |
| | ● オンライン | reghw | ... 2015/03/13 0:30 | 有効 | 有効 |
| | ● オンライン | Administrator | ... 2015/03/13 6:30 | 有効 | 有効 |
| | ● オンライン | Administrator | ... 2015/03/13 8:08 | 有効 | 有効 |
| | ● オンライン | Administrator | ... 2015/03/13 6:29 | 有効 | 有効 |

タスク

- クライアントを追加する
- コンピュータアカウントの追加
- ユーザーアカウントの追加
- Active Directory または LDAP ユ
- グループでコマンドを実行
- クライアントの検索
- 表示フィルタの設定

管理(A)

1/1