

科学研究費助成事業 研究成果報告書

平成 29 年 8 月 14 日現在

機関番号：82612

研究種目：挑戦的萌芽研究

研究期間：2014～2016

課題番号：26670181

研究課題名(和文) オンライン分散機械学習を活用した細胞評価システムの構築

研究課題名(英文) Construction of cell evaluation system utilizing online distributed machine learning

研究代表者

梅澤 明弘 (Umezawa, Akihiro)

国立研究開発法人国立成育医療研究センター・再生医療センター・副所長/再生医療センター長

研究者番号：70213486

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：本研究では機械学習技術・Information Retrieval技術のバイオビッグデータへ応用として細胞評価技術を確立し、実証検証により評価技術の製品化・サービス化への課題検討を行った。ゲノム・トランスクリプトーム・エピゲノムデータを利用したiPS細胞、ES細胞、体性幹細胞の分類・評価技術について、大規模分散オンライン機械学習フレームワークを活用し、ゲノムワイドにデータを活用する。異なる組織由来、異なる継代数のヒト多能性幹細胞、および異なる方法で樹立したヒトES細胞を用いた網羅的DNAメチル化解析を行った。

研究成果の概要(英文)：In this research, we established cell evaluation technology as applied to bio-big data of machine learning technology, information Retrieval technology, and examined issues for commercialization and service of evaluation technology by verification verification. We utilize genome wide data utilizing genome, transcriptome, epigenome data to classify and evaluate iPS cells, ES cells and somatic stem cells using a large-scale distributed online machine learning framework. Comprehensive DNA methylation analysis was performed using different tissue origin, different human stem cells, and human ES cells established by different methods.

研究分野：病理学

キーワード：移植病理 エピゲノム 機械学習 再生医療 幹細胞

1. 研究開始当初の背景

近年自動車・監視カメラなどのセンサーデータ、ウェブサービスの行動履歴・監視情報などのログデータなどのビッグデータの活用が研究・ビジネス双方で注目されている。バイオ分野でも生命情報の解析技術が飛躍的に向上し、タンパク質構造、アミノ酸配列、遺伝子発現、ゲノム・エピゲノム情報等の生データは膨大な蓄積がなされている。ヒト iPS 細胞、ES 細胞および体性幹細胞にかかる DNA メチル化を含むエピジェネティクス研究は、主に網羅的 DNA メチル化解析、網羅的クロマチン解析として多数報告されてきている。申請者は、ヒト羊膜、胎児肺線維芽細胞、子宮内膜、胎盤動脈、月経血、網膜、指皮膚細胞、耳軟骨など様々な由来の異なるヒト組織から iPS 細胞を多数樹立してきた (Nishino et al., *PLoS Genetics*, 7:e1002085, 2011; Nishino et al., *PLoS One*, 5:e13017, 2010)。これまでの研究では、限られたヒト iPS 細胞を用いて、限られた継代数および亜株で解析がされてきた。申請者らは、ヒト iPS 細胞株とこれらを長期培養した際の経時的なサンプルを得ることが可能であり、これら豊富なヒト iPS 細胞サンプルを用いて、バイオインフォマティクス手法による解析から、ヒト多能性幹細胞におけるリプログラミング機構の一端を解明した (Nishino et al., *PLoS Genetics*, 7:e1002085, 2011)。情報工学分野のバイオロジーへの活用は今に始まった事ではなく、BLAST を始めとするシーケンシングデータ解析ツールはバイオインフォマティクス分野の大きな成果の一つである。しかし最近 10 年ほどの機械学習分野の発展は目覚ましく、既存アルゴリズムを大規模・オンラインに改良した手法が次々に提案されており、これらの

異分野への活用が与える恩恵は現状では未知数な部分が多い。本課題は最新の情報工学技術を用いてオミックスデータをフル活用する点で、再生医療分野・情報工学分野双方に対し大きなインパクトを与える可能性を有すると確信している。本開発課題は、ES 細胞・iPS 細胞における生命現象・生命情報に関するビッグデータと機械学習を基本とした IT 技術の融合により、これらの iPS 細胞評価のスタンダードを確立するための一翼を担うことが可能である。日本発の iPS 細胞と日本発の解析手法の確立は世界に大きなインパクトを与えるのみならず、情報分野におけるグローバルスタンダードとなる可能性を有している。

2. 研究の目的

iPS 細胞の樹立成功以降、多能性幹細胞を用いた高度再生医療の実現に大きな期待が寄せられているが、この医療が日常的な治療法として普及するためには、治療に用いる細胞群について再現性と安全性を保証するための「基準」が必要不可欠である。なぜなら、iPS 細胞や ES 細胞等の多能性幹細胞の持つ高い増殖能により、安全面において腫瘍形成の可能性が懸念され、また iPS 細胞株間における分化指向性や分化抵抗性など細胞株間の性質の違いが存在するからである。自家細胞移植によるオーダーメイド医療が基本となる iPS 細胞再生医療においては、各個人ごとに異なる細胞株を利用する必要があるため、厳密な評価基準を確立しなければならない。各 iPS 細胞株の性質を正確に理解、評価し、安全性を保証する評価スタンダードを確立する。

再生医療に用いる細胞の有効性・安全性についての検証技術となる。iPS 細胞

の樹立成功以降、多能性幹細胞を用いた高度再生医療の実現に大きな期待が寄せられているが、この医療が日常的な治療法として普及するためには、治療に用いる細胞群について再現性と安全性を保証するための「基準」が必要不可欠である。なぜなら、iPS 細胞や ES 細胞等の多能性幹細胞の持つ高い増殖能により、安全面において腫瘍形成の可能性が懸念され、また iPS 細胞株間における分化指向性や分化抵抗性など細胞株間の性質の違いが存在するからである。自家細胞移植によるオーダーメイド医療が基本となる iPS 細胞再生医療においては、各個人ごとに異なる細胞株を利用する必要があるため、厳密な評価基準を確立しなければならない。しかし、各 iPS 細胞株の性質を正確に理解、評価し、安全性を確保する評価スタダードは未だ確立していないのが現状である。本開発課題は、これらのスタダードを確立するための一翼を担うことが可能である。日本発の iPS 細胞と日本発の解析手法の確立は世界に大きなインパクトを与えるのみならず、情報分野における持続可能な社会の形成に不可欠な課題である。

3. 研究の方法

ヒト胚性幹(ES)細胞及び体性幹細胞(Multipotent Stem Cell)の調製

ヒト幹細胞は、個体や組織ごとにそれぞれ特徴をもち、クローナルな集団とはいえない。たとえば、拒絶などの問題のない自分の細胞で自分の組織再生をおこなう細胞として注目される体性幹細胞であるが、これを移植源として使うという研究の歴史は比較的浅く、単離法やその定義に統一された概念がまだ存在せず、研究者により異なる細胞を扱っている可能性がある。「エピジェネティッ

ク・プロファイル」を用いた再生医療に適した細胞の選別を行うためのシステムを構築し、臨床への実用化を見極める。研究開発を進める上で、再現性・信頼性のあるデータを蓄積する必要がある。さらにヒト胚性幹(ES)細胞やヒト多能性幹細胞は、培養条件や培養期間に応じて変化していくという課題がある。したがって、どの細胞を用いるかの選択は何より重要なポイントとなる。国立成育医療研究センターでは様々なヒト ES 細胞やヒト体性幹細胞を樹立保有している。本提案においては、細胞数が十分確保しやすくかつ再現的に細胞を得やすいヒト幹細胞全般に関して実験を進めた。

ヒト胚性幹(ES)細胞、ヒト多能性幹細胞のエピジェネティック・プロファイリングによる解析

上記調製したヒト幹細胞、並びに分化誘導した細胞群に対し、illumina HumanMethylation450K を使用した網羅的 DNA メチル化解析、ChIP-on-chip を使用した網羅的ヒストンメチル化解析、網羅的遺伝子発現解析を行う。そのための適切なプロトコールを確立し、必要に応じて他の手法による確認を行なう。得られた解析結果のシグナル値を元にエピジェネティック・プロファイル評価システムにより解析するとともに、高品質バイオインフォマティクス解析系を構築する。様々なエピゲノム解析によるプロファイリングの結果、細胞に発現するエピゲノム修飾が推定され、これに対応した分化指向性、腫瘍リスク率の発現を決定する。

大規模オンライン機械学習を用いた細胞品質評価

PFI 社製 Jubatus を用いた大規模オンラ

イン機械学習による細胞品質評価を行う。既存技術では、品質に sensitive に反応し比較的検知しやすい情報のみを拾い上げ、細胞マーカーとしている。これは細胞から得られる ゲノム情報のうち、ごく一部のみを利用しているに過ぎない。数千カ所のミクロな特異性を積み重ねた結果によるマクロな特性の発見は細胞マーカーを用いた指標では限界がある。また、既存の機械学習ツールは、何万個のオーダーの特徴を利用したタスクを想定していない。Jubatus の搭載する最新の機械学習アルゴリズムでは数万個の特徴を用いた分類問題を解く事で、ゲノムワイドな情報をフルに活用して品質評価が可能である。

従来までのバッチ形式を基盤とする解析手法では、バッチ毎に解析の値が変化するため、データセットが固定され、1回の解析ですべてを完結する必要があり、データの追加による精度の向上は不可能である。また、解析の度に全データを一から処理するバッチ形式では対応できないほど解析対象のデータが膨大になり、データを一部のみを利用するなどの次善策が講じられた。結果として供出されるデータの全体像を把握することは困難であった。この問題の解決にはデータの追加時の処理が軽いオンライン処理が必要である。株式会社 Preferred Infrastructure は日本において大規模データにおける機械学習・検索技術の研究開発において、プラットフォーム Jubatus の開発実績を有している。この技術を応用しリアルタイムで生物情報の解析が可能な機械学習プラットフォームを開発しており、本研究においてもプラットフォームとして利用した。

4. 研究成果

異なる組織由来、異なる継代数のヒト多能性幹細胞、および異なる作成方法で樹立したヒト ES 細胞を用いた網羅的 DNA メチル化解析を行う。ヒト幹細胞は、個体や組織ごとにそれぞれ特徴をもち、クローナルな集団とはいえない。たとえば、拒絶などの問題のない自分の細胞で自分の組織再生をおこなう細胞として注目される体性幹細胞であるが、これを移植元として使うという研究の歴史は比較的浅く、単離法やその定義に統一された概念がまだ存在せず、研究者により異なる細胞を扱っている可能性がある。「エピジェネティック・プロファイル」を用いた再生医療に適した細胞の選別を行うための革新的システムを構築し、臨床への実用化を見極める研究開発を進める上で、再現性・信頼性のあるデータを蓄積できた。

生命現象・生命情報に関するビッグデータを効率的に活用するシステムを開発する。生命現象・生命情報に関するビッグデータの効率的活用は、活気ある持続可能な (Active Sustainability) 社会を構築していくための大きな役割を果たすことになる。「持続可能である」ということは、使用エネルギーや解析コスト、解析効率に関して、従来までの手法を凌駕するものでなければならない。すなわち、本研究開発を遂行することで蓄積されたビッグデータに対するアプローチであり、情報リソースの有効利用を実現することが可能 (使用エネルギーの節約) リアルタイム・ストリームでデータを処理、解析、蓄積できるオンライン解析を基盤技術とするため、データの追加により精度が向上し、既存の解析が無駄になることはない (解析コストの低減) 従来までのバッチ形式を基盤と

する解析手法では、バッチ毎に解析の値が変化するため、データセットが固定され、1回の解析ですべてを完結する必要があり、データの追加による精度の向上は不可である（解析効率の向上）のすべてが解決可能であり、生命情報の取り扱いに関するイノベーションとなりうる。

5．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕(計 0 件)

〔学会発表〕(計 0 件)

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕
ホームページ等

6．研究組織

(1)研究代表者

梅澤明弘 (UMEZAWA, Akihiro)
国立成育医療研究センター・再生医療センター・センター長
研究者番号：70213486

(2)研究分担者

神崎誠一 (KANZAKI, Seiichi)
国立成育医療研究センター・細胞医療研究部・研究員
研究者番号：20589741

平成 27 年 3 月に研究組織より削除