

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 6 日現在

機関番号：14401

研究種目：若手研究(B)

研究期間：2012～2015

課題番号：24700294

研究課題名(和文) 高速シーケンサーを用いた遺伝子制御ネットワーク解析技術の開発

研究課題名(英文) A method for gene regulatory network estimation based on high-throughput sequencer data

研究代表者

瀬尾 茂人 (Seno, Shigeto)

大阪大学・情報科学研究科・助教

研究者番号：30432462

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：近年発展が目覚ましい高速シーケンサーによる遺伝子発現量の測定データと、従来のマイクロアレイによるデータとを組み合わせることで遺伝子制御ネットワーク解析の方法を開発した。実際にマウスの褐色脂肪細胞やベージュ脂肪細胞に対して寒冷曝露刺激を与えた際の遺伝子発現プロファイルにこの方法を適用し、脂肪細胞の褐色化に関連する候補遺伝子を抽出した。

研究成果の概要(英文)：We developed a method for the estimation of gene regulatory networks based on the gene expression data obtained by using both microarray technology and high throughput sequencer technology. Our method was applied to the data observing the effects of cold exposure against adipocytes, we successfully detected some candidate genes concerning browning mechanism.

研究分野：情報学

キーワード：バイオインフォマティクス 遺伝子発現解析 褐色脂肪細胞

1. 研究開始当初の背景

脂肪細胞は、生命活動のためのエネルギー貯蔵と供給といった重要な役割を果たす一方で、メタボリックシンドロームをはじめとする様々な深刻な症状を引き起こすことから、その分化の機序の解明が望まれている。脂肪細胞分化では、その過程で転写因子による遺伝子制御ネットワークが関与しており、Pparg 遺伝子をはじめ、いくつかの重要な遺伝子が報告されている。しかしながら、ある種の遺伝子は同一の遺伝子座から選択的スプライシングにより部分的に異なるアミノ酸配列をもつタンパク質群（アイソフォーム）を生成することが知られており、Pparg の場合ではこれまでに発現時期も機能も異なる3つのアイソフォームが同定されている。さらには、microRNAをはじめとするタンパク質に翻訳されない転写物、non-coding RNA (ncRNA) も脂肪細胞の分化において実際に重要な役割を果たしていることが報告されていた。

従来、遺伝子制御ネットワークを推定するためにはマイクロアレイによる発現プロファイルが広く用いられてきた。これらの従来手法は、1つの遺伝子座から1種類のタンパク質が生産されることを前提とした、いわば遺伝子を中心とする"gene-centric"なネットワーク推定法であり、これはマイクロアレイではアイソフォームや ncRNA といった転写物の量を網羅的に測定することが出来ないことにも起因していた。

2. 研究の目的

そこで本研究では、従来の推定法を拡張し、RNA-seq による発現データを元に遺伝子制御ネットワークを推定するための方法論の開発を行うことを目的とした。RNA-seq とは、高速シーケンサーを用いることで発現している RNA 配列を直接読み取って測定する方法であり、未知のアイソフォームや ncRNA の発見と定量化を行うことが出来る。本研究

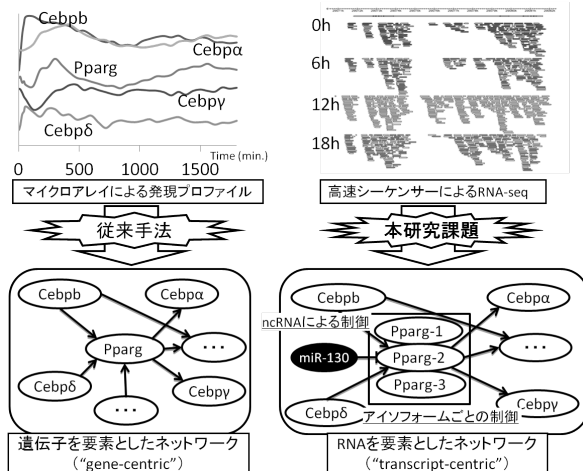


図 1：研究目的の概要

は、脂肪細胞分化の段階に応じて変化する動的なネットワークを描出するため、RNA-seq のデータを元に、アイソフォームや ncRNA といった転写物を要素として拡張した "transcript-centric" な制御ネットワークの推定法を開発することを目的とした (図 1：研究目的の概要)。

この大目的のために、本研究では以下の3点を小目的として推進した。

- (1) RNA-seq データからの遺伝子発現量の定量化手法の開発
- (2) 時系列遺伝子発現プロファイルのためのデータ解析手法の開発
- (3) RNA-seq や ncRNA の遺伝子発現プロファイルから遺伝子制御ネットワーク推定を行う方法論の開発

3. 研究の方法

- (1) RNA-seq データからの遺伝子発現量の定量化手法の開発

本研究では、正確なアイソフォーム推定を行うため、RNA-seq データのリードとゲノム配列との対応を調べるマッピング、アイソフォーム候補の推定、各アイソフォームの発現量の推定を反復的に行う方法を提案した。これによって、リードの長さやシーケンスエラー、ゲノム配列の相同性などの原因のために由来箇所を1ヶ所に特定できないリードも、正確にモデルに組み込むことが可能となり、従来法よりも高い精度でアイソフォームを推定できるようになった。

- (2) 時系列遺伝子発現プロファイルのためのデータ解析手法の開発

細胞分化に関するデータは時系列の遺伝子発現プロファイルとして測定されることが多い。時系列のデータから各変数の因果関係を推定する方法としてはダイナミックベイジアンネットワークを用いた方法がよく用いられるため、これを遺伝子発現プロファイルに対しても適用可能とするための技術開発を行った。

さらに、分化の各ステージにおける遺伝子の制御関係の変化や、高発現する遺伝子群の機能の変化を抽出するべく、スライディングウィンドウ方式を用いた時系列遺伝子発現解析手法の開発を行った。1つは従来のベイジアンネットワークとスライディングウィンドウ法を組み合わせたもので、またネットワークの変化を統計量として検定を行うことで、ウィンドウの幅も自動で決定する方法である。これにより時間局所的な遺伝子制御ネットワークとその変化を検出することが出来る。また遺伝子の機能注釈として Gene Ontology 等を用いて、遺伝子単体の変化ではなく機能を単位として解析を行う方法に GSEA (Gene Set Enrichment Analysis) があ

るが、この方法を時系列データへ適用可能なよう、アルゴリズムの拡張を行った。

(3) RNA-seq や ncRNA の遺伝子発現プロファイルから遺伝子制御ネットワーク推定を行う方法論の開発

ncRNA としては miRNA がよく研究されており、その発現量を測定する際には、高速シーケンサーによる方法の他にも、miRNA array といった方法が広く利用されている。miRNA は遺伝子の発現を主に抑制的に制御するため、遺伝子制御ネットワークの推定を行う際には、そのような特性を組み込んだアルゴリズムの開発を行った。

また、miRNA による遺伝子の転写調整を解明するために、遺伝子-miRNA 間の相互作用予測に関する研究を行った。従来の相互作用予測は、miRNA とそのターゲットの遺伝子における結合領域の配列相補性に基づいて行われていたが、偽陽性を大量に含んでいることが問題となっていた。そこで遺伝子と miRNA の発現量を入力として正準相関分析を行うことにより、偽陽性を大幅に減らすことに成功した。

4. 研究成果

(1) 本研究では遺伝子のアイソフォーム推定のための RNA-Seq 解析手法を提案した。各遺伝子の発現量を推定後、そのレベルに応じてマルチリードを再マッピングによって配分しなおすことでマッピングミス減らし、またアイソフォーム推定精度の向上を目的とした。結果として、アイソフォームの推定結果において、従来手法と比較して高い正解率を得ることができた。さらにエキソン/イントロンの推定精度も向上しており、新規アイソフォーム検出の可能性も示した

(2) 大規模な遺伝子群の遺伝子制御の変化を解析する手法としては、ノードセットセパレーション法があり、この方法は時系列発現プロファイルの全時点を使用する。細胞分化では時間経過に伴って制御関係の変化が起こっていると考えられ、時点数の多い時系列発現プロファイルを一括に用いると、制御関係の細かな変化を推定できない問題点があった。本研究ではノードセットセパレーション法を拡張子、スライディングウィンドウ法によって時間方向に分割して、複数の部分時系列発現プロファイルからそれぞれのダイナミックベイジアンネットワークモデルを推定することによって動的な遺伝子制御ネットワークを推定する手法を提案した。本手法を細胞分化の時系列発現プロファイルに適用した結果、既知の遺伝子制御関係の推定精度の向上を確認した。これにより、遺伝子制御ネットワーク中で、発現制御への寄与が高い遺伝子を抽出することができるように

なると考えられる。

(3) 寒冷曝露条件下でマウス生体から採取した、褐色脂肪細胞、白色脂肪細胞、ベージュ脂肪細胞のそれぞれについて、RNA-seq とマイクロアレイ、また miRNA アレイによって発現プロファイルを取得した遺伝子発現プロファイルを対象として解析を行った。結果としていくつかの注目すべき miRNA を特定し、また、これまで全く関係が報告されていなかったマクロファージ由来のサイトカインに熱産生の抑制効果が示唆されるなど、興味深い結果を得た。

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 6 件)

Tsuyoshi Goto, Supaporn Naknukool, Rieko Yoshitake, Yuki Hanafusa, Soshi Tokiwa, Yongjia Li, Tomoya Sakamoto, Takahiro Nitta, Minji Kim, Nobuyuki Takahashi, Rina Yu, Hiromi Daiyasu, Shigeto Seno, Hideo Matsuda, Teruo Kawada, Proinflammatory cytokine interleukin-1 β suppresses cold-induced thermogenesis in adipocytes. *Cytokine* 77 (2016): 107-114 (2016.1).

Yoichi Takenaka, Shigeto Seno, Hideo Matsuda, Detecting the shifts of gene regulatory networks during time-course experiments with a single time point temporal resolution, *Journal of Bioinformatics and Computational Biology*, DOI:10.1142/S0219720015430027 (2015.9).

渡邊之人, 瀬尾茂人, 竹中要一, 松田秀雄, 複数時系列遺伝子発現プロファイルを利用した遺伝子制御ネットワーク推定の精度向上手法, 情報処理学会論文誌数理モデル化と応用, Vol.6, No.3, pp.151-162 (2013.12).

Ryo Araki, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, An estimation method for a cellular-state-specific gene regulatory network along tree-structured gene expression profiles, *Gene*, Vol. 518, Iss. 1, pp. 17-25, (2013.3).

Tomoshige Ohno, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, A Method for Isoform Prediction from RNA-Seq Data by Iterative Mapping, *IPJSJ*

Transactions on Bioinformatics, Vol.5 , pp.27-33 (2012.4).

〔学会発表〕(計 12 件)

瀬尾茂人, イメージングデータの定量化とトランスオミクス解析, 生理学研究所研究会「生物学的階層構造をまたぐセルセンサー情報伝達に関する戦略的研究開発」(2015.9).

津田絢子, 瀬尾茂人, 竹中要一, 松田秀雄, ベイジアンネットワークによる遺伝子制御ネットワーク推定結果の反復構築のための計算速度向上手法, 第 98 回 MPS・第 38 回 BIO 合同研究発表会(2014.6).

Tomoshige Ohno, Hiromi Daiyasu, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda. Integrative prediction of miRNA-mRNA interactions from high-throughput sequencing data, RECOMB/ISCB Conference on Regulatory and Systems Genomics, with DREAM Challenges 2013 (2013.11).

Yuta Okuma, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda. Gene Set Enrichment Analysis for a Long Time Series Gene Expression Profile, BIOCOMP'13, pp.412-417 (2013.7).

Kensuke Suzuki, Daisuke Ueta, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda. A Method of Sequence Analysis for High-Throughput Sequencer Data Based on Shifted Short Read Clustering, BIOCOMP'13, pp.380-386 (2013.7).

Tomoyoshi Nakayama, Hiromi Daiyasu, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda. Reconstruction of Dynamic Gene Regulatory Networks for Cell Differentiation by Separation of Time-course Data, BIOCOMP'13, pp.375-379 (2013.7).

Ryo Araki, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, An estimation method for a cellular-state-specific gene regulatory network along tree-structured gene expression profiles, GIW2012, P13 (2012.12).

Yukito Watanabe, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, Bayes-based inference of gene regulatory network for multiple time series gene expression profile, Joint Conference on Informatics in Biology, Medicine and Pharmacology, (2012.10).

Tomoyoshi Nakayama, Yoshiyuki Kido, Hiromi Daiyasu, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, Estimation of Dynamic Gene Regulatory Networks for Cell Differentiation by Splitting Time Course Data, Joint Conference on Informatics in Biology, Medicine and Pharmacology (2012.10).

Tomoshige Ohno, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, Transcript Type Dependent Normalization of Expression Levels in RNA-Seq Data for Non-Coding RNA Analysis, Joint Conference on Informatics in Biology, Medicine and Pharmacology, (2012.10).

6 . 研究組織

(1)研究代表者

瀬尾 茂人 (SENO, Shigeto)

大阪大学・大学院情報科学研究科・助教

研究者番号 : 30432462