

令和 2 年 5 月 27 日現在

機関番号：17104

研究種目：挑戦的研究（萌芽）

研究期間：2018～2019

課題番号：18K19930

研究課題名（和文）低分子化合物による転写因子制御と細胞直接変換のためのインシリコ予測法の開発

研究課題名（英文）Development of computational methods for transcription factor regulation and direct cell conversion by small compounds

研究代表者

山西 芳裕（Yamanishi, Yoshihiro）

九州工業大学・大学院情報工学研究院・教授

研究者番号：60437267

交付決定額（研究期間全体）：（直接経費） 4,800,000円

研究成果の概要（和文）：iPS細胞のような未分化細胞を介さずに、すでに分化した細胞を別の種類の細胞に直接変換する「ダイレクトリプログラミング」が、再生医療のための新しい革新的技術として注目されている。本研究では、ダイレクトリプログラミングのための情報技術を開発した。シミュレーションでの予測精度で、既存手法を上回ることを確認した。提案手法をヒトゲノムにコードされている約千個の転写因子に対して適用し、分化誘導する転写因子や低分子化合物を大規模推定した。皮膚線維芽細胞から様々な細胞（肝細胞など）への変換に関与すると予測された転写因子や低分子化合物の詳細について生物学的な考察を行なった。

研究成果の学術的意義や社会的意義  
再生医療への社会的な期待は大きい。ダイレクトリプログラミングはiPS細胞を介さないため腫瘍形成の可能性が低く、分化誘導率や再現性が高いなど、医療応用における利点が多い。ダイレクトリプログラミングにより生体内で目的の細胞を作成できれば、細胞移植の必要がなくなり、全く新しい再生医療に繋がる。本研究で提案する低分子化合物での分化誘導を支援する情報技術は、生体内リプログラミングを更に容易にできるため大きな意義がある。

研究成果の概要（英文）：Direct reprogramming is a research field on direct conversion of fully differentiated mature cells into a variety of other cell types while bypassing an intermediate pluripotent state (e.g., iPS cell). Direct reprogramming is known to have several advantages over the other reprogramming methods in terms of high efficacy and safety, thus, it has been receiving much attentions in regenerative medicine. In this study, we develop novel computational methods to predict TFs and compounds that induce direct reprogramming for a variety of human cells. The prediction is performed by an integrative analysis of genomic data, transcriptome data, and epigenome data. We show the usefulness of the proposed methods on several direct cell conversions.

研究分野：バイオインフォマティクス

キーワード：ダイレクトリプログラミング 低分子化合物 転写因子 細胞直接変換 インシリコ

## 様式 C-19、F-19-1、Z-19 (共通)

### 1. 研究開始当初の背景

iPS 細胞のような未分化細胞を介さずに、すでに分化した細胞を別の種類の細胞に直接変換する「ダイレクトリプログラミング」が、再生医療のための新しい革新的技術として注目されている。ダイレクトリプログラミングは、iPS 細胞を介さないため腫瘍形成の可能性が低く、分化誘導率や再現性が高いなど、医療応用における利点が多い。その細胞運命転換を決めるのは、転写因子 (DNA 上の制御領域に特異的に結合するタンパク質) の組み合わせである。例えば、マウスの線維芽細胞に、Hnf4a (hepatocyte nuclear factor 4 $\alpha$ ) と Foxa1 (forkhead box A1)、Foxa2 (forkhead box A2)、Foxa3 (forkhead box A3) のいずれか 1 つの転写因子を組み合わせることで導入することにより、マウスの誘導肝細胞様 (iHep) 細胞を樹立できることが報告されている。他にも、ヒト膵臓の細胞である  $\beta$  細胞を標的細胞とした場合、ヒト膵外分泌細胞に PDX1 (pancreatic and duodenal homeobox 1)、NGN3 (neurogenin 3)、MAFA (MAF bZIP transcription factor A) という 3 つの転写因子を組み合わせることで樹立することができる。これまでのダイレクトリプログラミングの標準的な方法は、レトロウィルスを用いた転写因子の遺伝子導入であり、ウィルスに起因する発がんリスクという深刻な問題があった。

### 2. 研究の目的

本研究では、転写因子だけでなく低分子化合物で制御するダイレクトリプログラミングを提案する。ダイレクトリプログラミングを担う転写因子を制御する低分子化合物を同定し、細胞運命転換を低分子化合物で行うダイレクトリプログラミングを支援するためのインシリコ手法を開発するのが目的である。

### 3. 研究の方法

まず、細胞分化誘導に関与する転写因子や低分子化合物のデータを整備した。転写因子や低分子化合物のオミックスデータや化学構造を世界中の関連データベースから収集して情報解析可能な形に整備した。過去に転換の成功が報告されている既知の細胞転換パターンについても、文献から 35 種類の細胞転換 (皮膚細胞から肝細胞への変換、皮膚細胞から腸前駆細胞への直接変換など) をデータとして整備した。

次に、転写因子を制御する低分子化合物を予測する手法を開発した。細胞分化誘導に関与する転写因子や低分子化合物のオミックスデータを整備した。新たに、細胞分化誘導に関与するパスイエのデータを整備した。化合物に関するデータを表す特徴ベクトルを入力とし、低分子化合物が転写因子を活性化するかまたは不活性化するかどうか機械学習で予測する手法を実装した。

### 4. 研究成果

ゲノム情報、エピゲノム情報、トランスクリプトーム情報、分子ネットワーク情報、化学構造情報を融合解析し、ダイレクトリプログラミングを誘導する転写因子と低分子化合物を予測する手法を開発した。提案手法をヒトゲノムにコードされている約千個の転写因子に対して適用し、各細胞直接変換に対して転写因子セットを予測し、それを制御する低分子化合物を大規模推定するためのモデルのプロトタイプを実装した。提案手法は、オミックスデータが入手できるさまざまな細胞に対して適用できる。

ここでは例として線維芽細胞から 5 つの標的細胞 (肝細胞、軟骨細胞、神経細胞、心筋細胞、 $\beta$  細胞) への直接変換を担う転写因子を予測した。クロスバリデーション実験を行い、シミュレーションでの予測精度で、既存手法を上回ることを確認した。評価指標として、AUC (area under the receiver operator characteristics curve) と AUPR (area under the precision-recall curve) を用いた。AUC とは、ROC (receiver operator characteristics) 曲線の下部分の面積のことであり、ランダムな予測の時 0.5 を示し、完全に予測されたとき 1.0 を示す。AUPR は、予測の適合率と再現率から算出される指標である。どちらの指標も、予測性能が高ければ、高い数値を示す。

以下の表は、AUC と AUPR を用いて、標的細胞ごとにそれぞれの予測手法の性能を評価した結果を示す。これらの数値が高いほど、その手法の予測性能が高いことを示す。それぞれの組織について、最も高い数値を示した手法を太字で示している。NA は、スコアが算出できなかったことを示す。AUC と AUPR のどちらの指標についても、「オミックス情報を融合した手法」が高い予測性能を示す。この結果から、オミックス情報を融合することで、性能の向上ができることが明らかとなった。

標的細胞	トランスクリプトーム情報を用いた手法	エピゲノム情報を用いた手法	ゲノム情報を用いた手法	分子ネットワーク情報を用いた手法	オミックス情報を融合した手法
<b>AUC</b>					
肝細胞	0.794	0.854	<b>0.990</b>	0.821	0.890
軟骨細胞	NA	0.689	0.808	NA	<b>0.901</b>
神経細胞	0.704	0.640	0.699	0.724	<b>0.801</b>

心筋細胞	<b>0.831</b>	0.663	NA	0.804	0.808
β細胞	0.605	0.554	NA	0.645	<b>0.672</b>
<b>AUPR</b>					
肝細胞	0.072	0.224	0.103	0.639	<b>0.670</b>
軟骨細胞	NA	0.027	0.025	NA	<b>0.029</b>
神経細胞	0.150	0.085	0.130	0.392	<b>0.475</b>
心筋細胞	<b>0.347</b>	0.021	NA	0.042	0.110
β細胞	<b>0.011</b>	0.007	NA	<b>0.011</b>	<b>0.011</b>

提案手法をヒトゲノムにコードされている約千個の転写因子に対して適用し、各細胞直接変換に対して分化誘導する転写因子や低分子化合物を大規模推定した。線維芽細胞から5つの標的細胞（肝細胞、軟骨細胞、神経細胞、心筋細胞、β細胞）への直接変換を例にして、提案手法がさまざまな標的細胞を樹立する際に有望な転写因子や、それを制御する化合物を網羅的に予測できることを示した。例えば、皮膚線維芽細胞から肝細胞への変換に関与すると予測された転写因子や低分子化合物の詳細について生物学的な考察を行なった。予測上位の転写因子や化合物について、文献報告と比較し、その妥当性を検証できた。

元細胞から目的細胞に変換される際の細胞状態を表す遺伝子発現データを実験的に測定して解析し、妥当性を検証した。化合物応答の遺伝子発現プロファイルと比較して、その妥当性を検証できた。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 中村透, 岩田通夫, 濱野桃子, 江口凌平, 山西芳裕
2. 発表標題 低分子化合物によるデータ駆動型ダイレクトリプログラミング
3. 学会等名 第8回生命医薬情報学連合大会
4. 発表年 2019年

1. 発表者名 江口凌平, 濱野桃子, 岩田通夫, 中村透, 沖真弥, 山西芳裕
2. 発表標題 トランスクリプトーム情報とエピゲノム情報の融合解析によるインシリコ・ダイレクトリプログラミング
3. 学会等名 第8回生命医薬情報学連合大会
4. 発表年 2019年

1. 発表者名 中村透, 岩田通夫, 濱野桃子, 江口凌平, 山西芳裕
2. 発表標題 Small Compound-based Direct Reprogramming Using Large-scale Omics Data
3. 学会等名 情報計算法学生物学会2019年大会
4. 発表年 2019年

1. 発表者名 江口凌平, 濱野桃子, 岩田通夫, 中村透, 沖真弥, 山西芳裕
2. 発表標題 Computational direct reprogramming by integrating genome, transcriptome and epigenome data
3. 学会等名 情報計算法学生物学会2019年大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担 者	鈴木 淳史  (Suzuki Atsushi)  (30415195)	九州大学・生体防御医学研究所・教授    (17102)	