

令和 5 年 5 月 12 日現在

機関番号：12601

研究種目：基盤研究(C) (一般)

研究期間：2020～2022

課題番号：20K06606

研究課題名(和文) 遺伝子転写制御領域の空間構造を定量的に解析する計算手法の開発

研究課題名(英文) Developing computational methods for quantitative analysis of 3D transcriptional domain

研究代表者

朴 聖俊 (Park, Sung-Joon)

東京大学・医科学研究所・准教授

研究者番号：40759411

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：ゲノムの長距離コンタクトを介した遺伝子転写制御は、正常ゲノム機能維持に必須であり、その破綻は異常細胞分化や異常発生などの原因となる。標的遺伝子の空間的近傍に存在する長距離コンタクトゲノム部位は、多種多様な転写調節関連因子の高度な相互作用のきっかけとなり、いわゆる、空間的転写ドメインの形成に貢献する。本研究では、このような転写ドメイン内におけるゲノム・エピゲノム転写調節因子の網羅的同定と、それらを定量的に解析する数理モデリング手法を開発した。これにより、ゲノム高次構造の機能的重要性に対するより深化した解析が可能となった。

研究成果の学術的意義や社会的意義

本研究では、ゲノムトポロジーと遺伝子発現の動的共役をより包括的に理解するための情報科学的手法を提案した。つまり、既存の発現制御に重要なコンタクトゲノム部位の同定に留まらず、プロモーターとコンタクトゲノム部位の配列情報から潜在的な転写因子間相互作用を推定し、さらに、コファクターやメディエーターのタンパク質間相互作用を取り入れ、転写ドメイン内の制御シグナルの伝達様相を可視化した。これは、いわば「ゲノム高次構造を介した転写制御シグナローム」という新しい学術的分野の展開に貢献するであろう。

研究成果の概要(英文)：Genomic long-range contacts mediate the regulation of gene expression, which is essential for maintaining normal genome functions and for avoiding abnormal cell differentiation. The long-range contacts in the spatial vicinity of target genes provide the 3D space where a wide variety of transcriptional regulators are highly interacting with each other and contribute to forming the 3D transcriptional domain. In this study, we developed a computational model to comprehensively identify genomic and epigenetic regulators from the transcriptional domain and analyze the impact of regulators quantitatively. This has enabled a more in-depth analysis of the functional importance of the higher-order chromatin structures.

研究分野：バイオインフォマティクス

キーワード：ゲノム高次構造 遺伝子転写制御 バイオインフォマティクス

## 様式 C-19、F-19-1、Z-19 (共通)

### 1. 研究開始当初の背景

ゲノムの三次元構造は遺伝子発現制御領域を物理的に制約し、染色体内・染色体間ゲノム配列の相互作用の空間場、すなわち、3D 転写ドメインを作り出す(図1)。この転写ドメインでは、多種多様な調節因子が標的遺伝子の空間近傍に集結して高度に複雑で巧妙な転写ファクトリーが機能しているが、その詳細については不明な点が多い。例えば、多様なシス・トランス制御因子はどうやって相互協調を実現しているのだろうか。このような問いに対して、先行研究ではもっぱらエンハンサー様非コード領域の接触様相の描像に注目しており、3D 転写ドメインにおける調節因子の定量的解析はなされていない。

### 2. 研究の目的

本研究では、大規模公共 NGS データを利用して、**転写制御に関わる 3D 転写ドメインを定義し、ゲノム・エピゲノム因子の網羅的検出とそれらの調節役割を定量的に解析する計算手法を開発する**。これにより、ゲノム高次構造の相互作用による遺伝子転写調節因子の協調様式の理解を目指す。

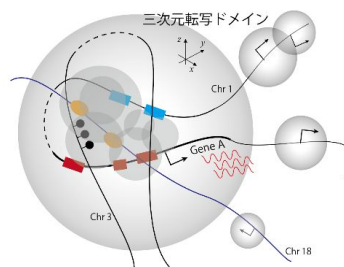


図1: 空間的転写制御ドメイン

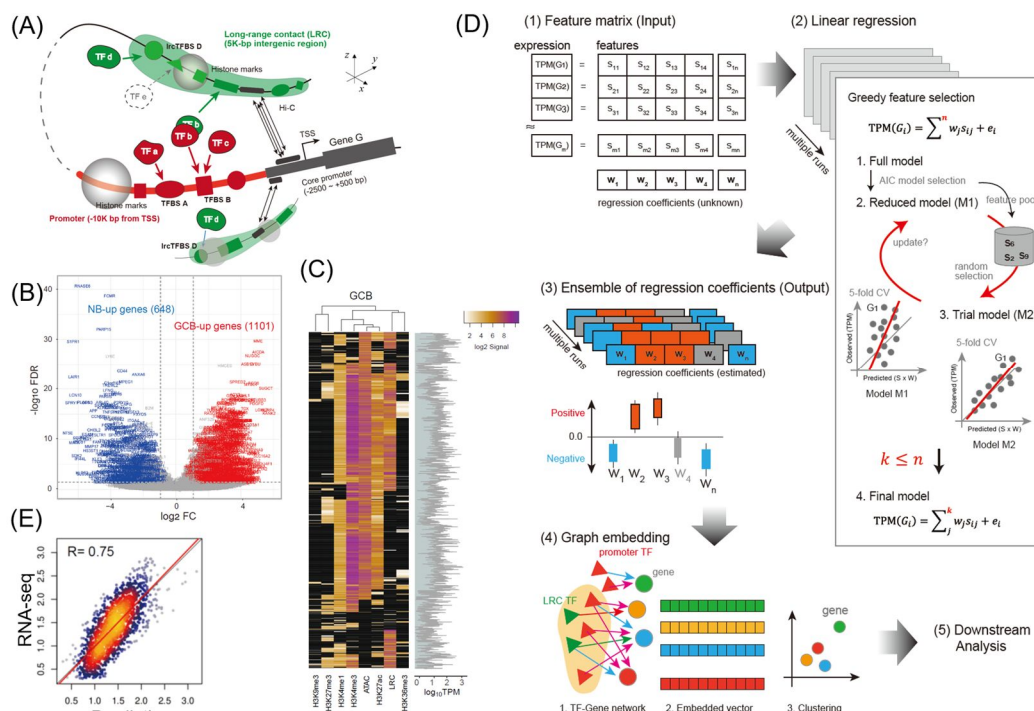


図2: 本研究で扱うデータとモデリング方法の概略

### 3. 研究の方法

#### (1) ゲノム・エピゲノム転写制御因子の網羅的検出

まず、関連 NGS データが豊富なヒト成熟ナイーブ B 細胞 (NB) と胚中心 B 細胞 (GCB) を題材に、公共 RNA-seq、Histone ChIP-seq、Hi-C データなどを用意した。そして、Hi-C データを用いて**遺伝子コアプロモーター部位に接触する非コードゲノム部位を検出し、これらの集合体をもって標的遺伝子に関わる 3D 転写ドメインを仮定した**。つぎに、ドメイン内の非コード領域に濃縮している転写因子結合部位 (TFBS) とそれらに潜在的に結合する転写因子 (TF) を推定した(図 2A)。このような転写ドメインを NB と GCB 間で発現が顕著に変化する遺伝子群 (DEG) に対して定義し(図 2B)、6 種類のヒストン修飾シグナルなどのエピゲノム情報も加えて(図 2C)、遺伝子転写制御の数理モデリングを行った(図 2D)。

#### (2) 転写制御因子の定量化

遺伝子転写制御の数理モデリングでは、**線形回帰による重要制御因子の定量化と、グラフ埋め込み法による類似相互作用の因子群を同定した**。まず、重要因子定量化のために、DEG の発現量を目的変数とし、重みづけした TFBS を説明変数とする特徴行列を構築した(図 2D-1)。ここで、

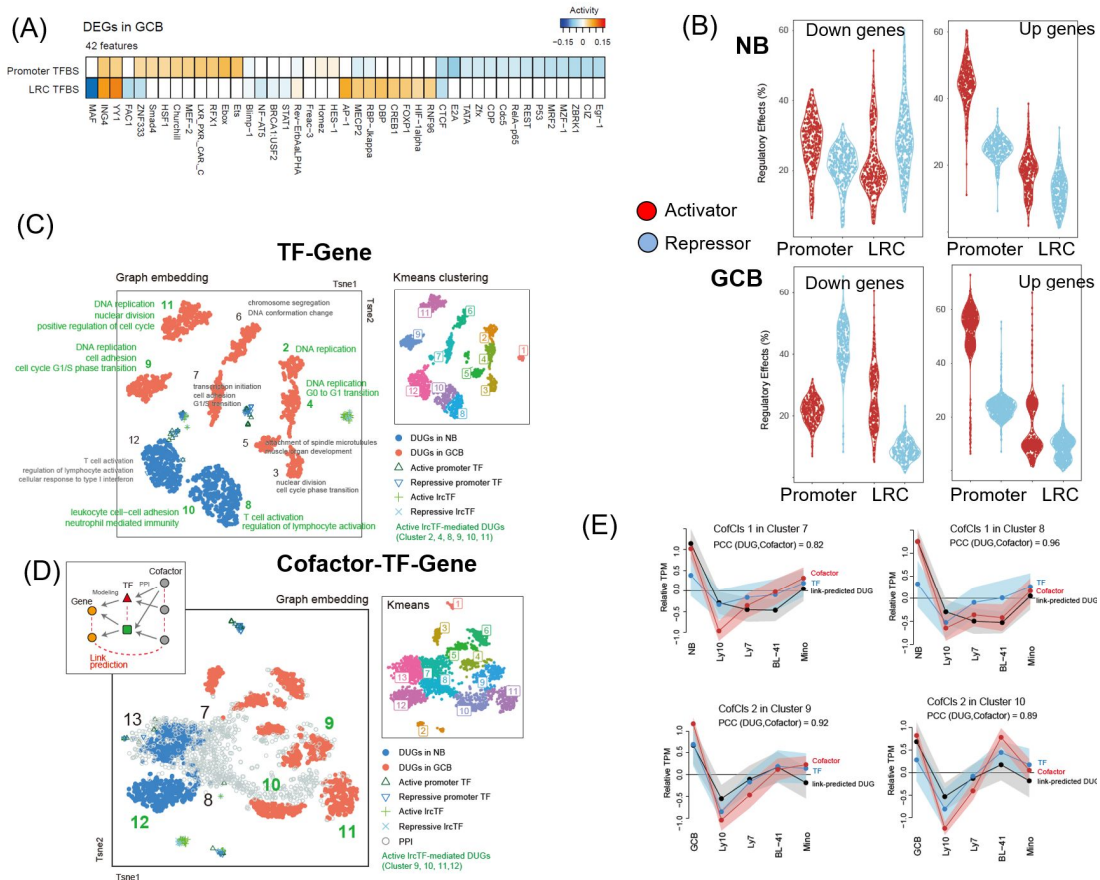


図 3：本研究での解析結果

TFBS はプロモーター由来 (prmTFBS) と長距離コンタクト由来 (lrcTFBS) に区別できるが、後者は H3K27ac、H3K4me1 のシグナルと Hi-C コンタクト確率で重みづけを行った。この特徴行列は反復変数選択法 (図 2D-2、2D-3) によって処理され、結果的に DEG 発現量を高精度で予測するモデルが構築できた (図 2E)。

### (3) 制御因子間相互作用ネットワークの解析

推定された重要 TFBS に潜在的に結合する TF 情報を用いて、TF - DEG ネットワークを構築した。さらに、プロモーター由来 TFBS に結合する TF と LRC 由来 TFBS に結合する TF のブリッジとなりうるコファクターやメディエーターなどをタンパク質相互作用 DB から検索することで Cofactor - TF - DEG ネットワークを構築した。このネットワークを用いて、転写ドメインに集結している TF や架橋タンパク質群のプロファイリングと相互作用を推定した。相互作用の推定では、この高度に複雑なグラフ構造をグラフ埋め込み技法で処理し、類似部分フラグ構造、すなわち、制御因子間相互作用が共通する DEG をシステムティックに検出して可視化した (図 2D-4)。

## 4. 研究成果

### (1) 重要転写制御因子の空間的組み合わせ

線形回帰モデリングによって DEG 発現制御に貢献する約 80 個の prmTFBS と lrcTFBS が同定された (図 3A)。これらの推定重要度 (回帰係数) は多様な組み合わせを呈していたが、遺伝子発現上昇には prmTFBS が、抑制には prmTFBS と lrcTFBS の細胞種特異的な抑制効果が存在することがわかった (図 3B)。すなわち、NB のような 静止状態の細胞における刺激に対する即答性には、長距離コンタクトによる制御効果が顕著に関係している可能性がある。

### (2) グラフ埋め込みによる共転写制御遺伝子の同定

同定された prmTFBS と lrcTFBS に結合する TF 情報を用いて TF - DEG ネットワークを構築した。そして各ノードをグラフ埋め込みベクトルに変換し、K-means でクラスタリングすると 12 個の DEG クラスタが検出された (図 3C)。各クラスタにおけるプロモーター由来・LRC 由来という TF 種類と、それらの組み合わせパターンは特異的であったが、濃縮している遺伝子オントロジーが互いに共通することから、B 細胞分化における機能的遺伝子の多様な制御機構の存在が示

された。

### (3) コファクターの重要性

最後に、プロモーターに結合する TF と LRC 部位に結合する TF に物理的に結合する補因子タンパク質を検索して Cofactor - TF - DEG ネットワークを構築し、(2)と同様なアプローチで解析した。その結果、DEG クラスと共存する Cofactor 群が同定された (図 3D)。面白いことに、疾患形成において、これらの Cofactor コーディング遺伝子の発現は、同じクラス内の DEG 発現と強く相関していた (図 3E)。すなわち、**機能的重要な遺伝子と共発現する補因子は、正常な 3D 転写ドメインの機能実現に重要**であることが示唆された。

### (4) 考察

ゲノムトポロジーと遺伝子発現の動的共役は、生命現象を司る重要な分子基盤を形成することが明らかになりつつある。本研究では、三次元転写制御領域における制御因子の定量化と相互作用を包括的に解析する情報科学的手法を提案した。その結果、近距離の遺伝子プロモーター構造の優先的な転写制御効果と、長距離コンタクト由来制御効果による発現のファインチューニングが示された。このような空間的制御シグナルの伝達には、DNA 結合 TF 間のクロストークに寄与するコファクターが必須であると推測された。今後、**「ゲノム高次構造を介した転写制御シグナローム」**の理解につながる研究成果であると考えている。

## 5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 Zeng Xin, Park Sung-Joon, Nakai Kenta	4. 巻 12
2. 論文標題 Characterizing Promoter and Enhancer Sequences by a Deep Learning Method	5. 発行年 2021年
3. 雑誌名 Frontiers in Genetics	6. 最初と最後の頁 681259
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/fgene.2021.681259	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Jia Hao, Park Sung-Joon, Nakai Kenta	4. 巻 22
2. 論文標題 A semi-supervised deep learning approach for predicting the functional effects of genomic non-coding variations	5. 発行年 2021年
3. 雑誌名 BMC Bioinformatics	6. 最初と最後の頁 128
掲載論文のDOI（デジタルオブジェクト識別子） 10.1186/s12859-021-03999-8	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Lee Jong-Hun, Saito Yutaka, Park Sung-Joon, Nakai Kenta	4. 巻 27
2. 論文標題 Existence and possible roles of independent non-CpG methylation in the mammalian brain	5. 発行年 2020年
3. 雑誌名 DNA Research	6. 最初と最後の頁 dsaa020
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/dnares/dsaa020	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計5件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 Taisuke Hori, Sung-Joon Park and Kenta Nakai
2. 発表標題 Meta co-expression analysis of genetic heterogeneity in GBM using single-cell RNA-seq data
3. 学会等名 生命医薬情報学連合大会（IIBMP 2021）
4. 発表年 2021年

1. 発表者名 Hao Jia, Sung-Joon Park, Kenta Nakai
2. 発表標題 A Semi-supervised Deep Learning Approach for Predicting the Functional Effects of Genomic Non-coding Variations
3. 学会等名 The 19th International Conference on Bioinformatics (InCoB 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Xin Zeng, Sung-Joon Park, Kenta Nakai
2. 発表標題 Characterizing Promoter and Enhancer Sequences by a Deep Learning Method
3. 学会等名 The 19th International Conference on Bioinformatics (InCoB 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Sung-Joon Park
2. 発表標題 Analyzing the impact of inter-chromosomal interactions on B cell development by linear regression modeling
3. 学会等名 生命医薬情報学連合大会 (IIBMP 2020)
4. 発表年 2020年

1. 発表者名 Darina Obukhova, Sung-Joon Park, Kenta Nakai
2. 発表標題 Co-expression network analysis of E3 ligase genes in embryonic stem cells and cell fate specification
3. 学会等名 生命医薬情報学連合大会 (IIBMP 2020)
4. 発表年 2020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
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------