

令和 2 年 6 月 1 日現在

機関番号：12601

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K00397

研究課題名(和文) パーソナルゲノム時代の組織特異的エンハンサー予測アルゴリズム研究

研究課題名(英文) Development of algorithms for the prediction of tissue-specific enhancers in the era of personalized genomics

研究代表者

中井 謙太 (NAKAI, KENTA)

東京大学・医科学研究所・教授

研究者番号：60217643

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：本研究では真核生物の組織特異的発現に重要と考えられているエンハンサーについて、その特異性制御機構と標的遺伝子群を、主にゲノム配列情報から理解することを目指した。異なる細胞種に対する Hi-C データを再解析・相互比較することで、細胞の種類によって TAD や A/B コンパートメント等が変化した領域の遺伝子発現の特異性を調べた。その結果、B細胞とそのリンパ腫などの間に、興味深い機能的相関を観察した。

研究成果の学術的意義や社会的意義

塩基配列決定技術の進歩により、個人間のゲノムの差(多型)やがん組織中の個々の細胞内ゲノム変異分布等の情報が次々に産出されており、いわゆるパーソナルゲノム医療への応用が期待されている。しかし、転写制御領域に関する我々の理解が不十分なため、これらを含む非コード領域中の多型や変異の及ぼす影響を見積もることは難しい。本研究は染色体構造の細胞による変化などをもとに転写制御の仕組みを研究し、いくつかの新しい報告を行った。

研究成果の概要(英文)：Enhancers play a major role in the regulation of tissue-specific gene expression in eukaryotes. In this study, we tried to understand how enhancers work to their target genes mainly from the analyses of genome sequence data. We reanalyzed different sources of Hi-C data and figured out where in the genome shows differential TAD and/or A/B compartment structures among cell types. Then, we checked if the expression pattern of genes in these regions is correlated with the changes of their cell types. Indeed, we found some interesting correlations in their expression and function between B-cells and their lymphomas, for example.

研究分野：ゲノム情報科学

キーワード：細胞種特異的エンハンサー Hi-Cデータ解析 A/Bコンパートメント グラフ畳み込みネットワーク B細胞リンパ腫

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

ヒトなどのゲノムの圧倒的部分はいわゆる非コード領域で占められている。非コード領域中で特に転写制御情報を担っている領域の重要性は広く認識されており、進化を促すような変異は、タンパク質コード領域中よりもむしろ転写制御領域中に多く起こると言われている。転写制御の仕組みについては、領域内のシスエレメントに転写因子と呼ばれるタンパク質が特異的に結合することが基本的なステップであることが知られているが、その詳細は未だ不明で、たとえばある遺伝子がどのような制御(筋肉特異的発現など)を受けているかをその塩基配列から読み取ることはまだまだ難しい。さらには、エンハンサーやインシュレーターなど、遺伝子から離れて存在する領域については、ループモデルなどが以前から知られているものの、その詳細はいまだ謎と言ってもよい。近年ゲノム塩基配列決定技術の進歩により、個々人のゲノム配列の差や、疾病患者におけるゲノムの変異、さらには一細胞毎のもつ変異の差などのデータが洪水のように産出されつつある。しかし、ENCODE プロジェクトなどで組織的なデータが出揃った現在でも、転写制御領域内における変異の効果を見積もることは非常に困難であり、特にエンハンサー領域については未だお手上げ状態であると言え、その合理的解釈法が広く待ち望まれている。エンハンサーがどのようにして、ターゲット遺伝子を制御しているかについては、大雑把に言って、2つの(必ずしも対立しない)考え方がある。一つは、プロモーター同様、個々のエンハンサーを形成するシスエレメントの組み合わせなどによって規定されているという考え方で、もう一つは、CTCF 等に基づくインシュレーターによって規定されるクロマチンのループ構造が変化することで、内部のエンハンサーの作用範囲が変わるという考え方である(。これに対して、申請者らのグループは、まず、組織特異的・発生時期特異的プロモーター領域におけるシスエレメント(未知のものを含む)の相対位置などの共通アーキテクチャーを自動的に抽出する研究で実績をあげてきた。エンハンサーの場合は位置と特異性の同定の問題があるが、原理的には同じアプローチが可能であると考えられる。一方、クロマチンループ構造については、Hi-C などのデータベースづくりを通して、エンハンサーの作用についての情報を扱ってきた。さらに、遺伝子の転写開始点データベース DBTSS の構築・維持やそのデータ解析にも長年かかわってきた。これらの経験をいかして、新たなアプローチに挑戦する着想を得た。

2. 研究の目的

塩基配列決定技術の進歩により、個人間のゲノムの差(多型)やがん組織中の個々の細胞内ゲノム変異分布等の情報が次々に産出されているが、転写制御領域に関する我々の理解が不十分なため、これらの領域中の多型や変異の及ぼす影響を見積もることは難しい。本研究では真核生物の組織特異的発現などに重要な役割を果たすエンハンサーについて、その組織特異的発現を制御する仕組みとターゲット遺伝子群を、ゲノム配列情報から理解するための研究を行う。具体的には、組織特異的エンハンサーに共通するシスエレメントの相対配置を抽出・カタログ化するとともに、Hi-C 実験等から推測されるクロマチンループ構造を規定するエレメント候補の探索を行う。その動的変化による組織特異性実現をモデル化し、知見を統合した予測アルゴリズムを開発することを目的とした。

3. 研究の方法

エンハンサーの作用について、主に2つの観点から研究を行い、最終的にはそれらから得られ

た知見を統合したアルゴリズムを構築する。第 1 の観点は、既知もしくは共発現データなどから推定された組織特異的エンハンサー内部に存在するシスエレメントの頻出する組み合わせを自動的に抽出する方法である。第 2 の観点は、ChIP-seq/Hi-C 実験等から示唆される、インシュレーターによって区切られたクロマチンのループ構造によってエンハンサーの作用範囲が規定されるという考え方である。ゲノム上で近隣遺伝子間の共発現状況を調べ、実験で示唆されたループ構造との相関を調べる。ターゲット遺伝子の予測可能性や多型データへの適用性も検証することを計画した。

4 . 研究成果

4.1.1

まず、第一の観点として、modENCODE プロジェクトのデータを用いて、ショウジョウバエの発生時期特異的シス制御モジュールのシステマティックなエレメント解析を行った。ハエのゲノムの場合は、サイズが小さく、遺伝子間領域が比較的狭いので、シス制御モジュールの候補位置は特に事前の知識がなくても推定できる。その結果、多数の共発現遺伝子に対する興味深い共通構造を抽出し、論文として報告した (López et al. PeerJ 2017)。

4.1.2

次に、既知のエンハンサー情報を文献から収集し、データベース化することを試みたが、海外で類似データベースが発表されたため、我々が以前構築したホヤの転写データベース DBTGR に関してエンハンサー情報を中心に更新する作業を行った (現在も継続中)。その一部の努力は国際ホヤ遺伝子発現データベース ANISEED として発表された (Brozovic et al. Nucl. Acids Res. 2018)。一方、培養細胞などを別とすると、ヒトやマウスなどの高等生物の組織別エンハンサーデータは、特にターゲット遺伝子まで含めた情報は十分とは言えず、予備的な研究の結果は思わしくなかった。その原因の一つとして、今回のようにデータ数や質に問題がある場合、組織特異的発現という定義がやや大雑把すぎる可能性を考えた。そこで、最後に、細胞種類特異的な遺伝子発現に関与しているエレメントを抽出するために、シングルセルトランスクリプトームデータを用いて、従来の方法より精密な特異的発現情報を抽出する試みを現在も続けている。現在は、エンハンサーの位置情報の問題と切り分けるために、線虫のデータを試みている。

4.1.3

一方、第二の観点からの研究では、異なる細胞種に対する Hi-C データを収集 / 再解析し、相互比較することで、細胞の種類によって、TAD や A/B コンパートメントなどの構造領域がどのように変化するかを調べ、それらが変化した領域が細胞種特異的遺伝子発現と相関しているかどうかを調べた。その結果、データ数の関係で統計性有意性は限定的ではあるが、B 細胞とそのリンパ腫などの間に、興味深い機能的相関が観察された (Nagai et al. BMC Med. Genomics 2019)。

4.1.4

最後に、複数の細胞種に対して、塩基配列と各種ヒストン修飾情報などを入力情報として、深層学習の枠組みでエンハンサー予測アルゴリズムを構築した。方法論的には、今のところ、

GCN (グラフ畳み込みネットワーク) という方法が良い成績を示すことを見出している。また、データの入力時に、単純に各々のヌクレオチド位置に対応する ChIP-seq ピーク強度を入力するのではなく、DNA 塩基配列を bin にわけて、その中での最大強度を用いると、予測能力が向上することを見出している。現在、さらに、当初からの目標である、上記二つの観点を融合させたアルゴリズムの構築を試みている。具体的には ATAC-seq のいろいろな細胞における値を比較して、近隣ピークの共起などをもとにして、エンハンサーとプロモーターの相互作用を推定する情報を予測に組み込む可能性を試みている。いくつかの理由のため、研究を設定期間中に終わらせることができなかったが、今後もできる範囲で研究を継続して、結果を出したいと考えている。なお、これらの研究と関連した perspective を、中国の研究者の依頼により、論文にまとめて発表した (Nakai, Engineering 2019)。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 1件/うちオープンアクセス 3件）

1. 著者名 Kenta Nakai	4. 巻 5
2. 論文標題 Information Science Should Take a Lead in Future Biomedical Research	5. 発行年 2019年
3. 雑誌名 Engineering	6. 最初と最後の頁 1155-1158
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1016/j.eng.2019.07.023	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 Brozovic M, Dantec C, Dardaillon J, et al.	4. 巻 46
2. 論文標題 ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets	5. 発行年 2018年
3. 雑誌名 Nucleic Acids Research	6. 最初と最後の頁 718-725
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1093/nar/gkx1108	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 該当する
1. 著者名 Luis A. E. Nagai, Sung-Joon Park, and Kenta Nakai	4. 巻 11
2. 論文標題 Analyzing the 3D chromatin organization coordinating with gene expression regulation in B-cell lymphom	5. 発行年 2019年
3. 雑誌名 BMC Medical Genomics	6. 最初と最後の頁 127
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1186/s12920-018-0437-8	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 Lopez Yosvany, Vandenbon Alexis, Nose Akinao, Nakai Kenta	4. 巻 5
2. 論文標題 Modeling the cis-regulatory modules of genes expressed in developmental stages of Drosophila melanogaster	5. 発行年 2017年
3. 雑誌名 PeerJ	6. 最初と最後の頁 e3389 ~ e3389
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.7717/peerj.3389	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計9件（うち招待講演 3件 / うち国際学会 6件）

1. 発表者名 Kenta Nakai
2. 発表標題 Deciphering the architecture of genetic switches encoded in non-coding DNA from massive sequence data
3. 学会等名 APSCIT(Asia Pacific Society for Computing and Information Technology) (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Zhuo Qin and Kenta Nakai
2. 発表標題 High-Dimensional Outlier Detection Algorithm via Locality Sensitive Hashing
3. 学会等名 APSCIT(Asia Pacific Society for Computing and Information Technology) (国際学会)
4. 発表年 2019年

1. 発表者名 Hao Jia and Kenta Nakai
2. 発表標題 Graph-based Semi-Supervised Learning for Off-target Predictions in CRISPR/ Cas9 System
3. 学会等名 APSCIT(Asia Pacific Society for Computing and Information Technology) (国際学会)
4. 発表年 2019年

1. 発表者名 PhitLing Tang, Ashwini Patil and Kenta Nakai
2. 発表標題 統合シングルセル造血トランスクリプトームにおける解析E Pluribus Cellulam: A Study of Aggregated Single-Cell Haematopoietic Transcriptomes
3. 学会等名 第42回日本分子生物学会
4. 発表年 2019年

1. 発表者名 山谷恭代、朴聖俊、中井謙太
2. 発表標題 脂肪由来幹細胞の運命決定と分化制御に関する時空間特異的ヒストン修飾のin-silico特定 In-silico Identification of Spatio-Temporal Histone Modifications Associated with Regulation for Fate Decision and Differentiation of Adipose-derived Stem Cells
3. 学会等名 第42回日本分子生物学会
4. 発表年 2019年

1. 発表者名 Kenta Nakai
2. 発表標題 How DNA sequencers are useful in various ways
3. 学会等名 The 19th KIAS Conference on Protein Structure and Function (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 川上 泰治、横森 類、坂本 倫紀、野口 大樹、原田 瑞輝、行者 蒔、鈴木 穰、中井 謙太、大道 裕、日下部 岳広
2. 発表標題 錐体視細胞特異的miRNA欠損メダカの網膜トランスクリプトームと視細胞モザイクパターンの解析
3. 学会等名 第42回日本分子生物学会
4. 発表年 2019年

1. 発表者名 Luis A.E. Nagai, Sung-Joon Park and Kenta Nakai
2. 発表標題 Analyzing the 3D Chromatin Organization Coordinating with Gene Expression Regulation in B-cell Lymphoma
3. 学会等名 The 17th International Conference On Bioinformatics(InCoB-2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Kenta Nakai
2. 発表標題 Analysis of Transcriptional Regulation through a Variety of OMICs Data
3. 学会等名 OMICs 2017 (招待講演) (国際学会)
4. 発表年 2017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
--	---------------------------	-----------------------	----