

令和 2 年 6 月 8 日現在

機関番号：82401

研究種目：若手研究(B)

研究期間：2017～2019

課題番号：17K15132

研究課題名(和文)形態形成における全ゲノム転写制御ネットワークの高精度推定方法の開発

研究課題名(英文) Development of transcriptional regulatory network prediction methods in morphogenesis

研究代表者

鬼丸 洸 (Onimaru, Koh)

国立研究開発法人理化学研究所・生命機能科学研究センター・基礎科学特別研究員

研究者番号：30787065

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：本研究では深層学習を用いた形態形成における全ゲノム転写制御ネットワークの推定方法を目的とし、"四肢形成におけるATAC-seqを用いた制御配列候補の同定"による深層学習用訓練データの作成、"深層学習を応用した全ゲノム転写制御ネットワーク推定方法の確立"を試みた。前者においては四肢形成におけるATAC-seqデータの取得に成功し、形態形成に関わる制御配列の特徴を解析し、プレプリントとして公開した。後者では、遺伝子制御配列の特徴を深層学習を用いて解析するプログラムの開発を行い、先行研究のパフォーマンスを上回るものの開発に成功、論文をプレプリントとして、ソフトウェアをgithubにおいて公開した。

研究成果の学術的意義や社会的意義

本研究における学術的意義は、形態形成における転写制御配列について新たな特徴、傾向を発見し、転写制御と形態の多様性について理解を深めることに貢献したことにある。また、深層学習を用いた新たなゲノム配列の解析方法の提案が出来、ゲノム配列と生物の形態を結びつける研究が発展する上で礎となることが期待される。社会貢献としては、本研究は、ヒトの個々のゲノム配列に対する新たな解釈を行う上で、基礎的な知見が得られ、技術開発のさきがけとなる成果が得られたと考えている。

研究成果の概要(英文)：This project was aimed at development of transcriptional regulatory prediction methods in morphogenesis by applying deep learning. We set the following two tasks to achieve this goal: a) the genome-wide identification of morphogenic transcriptional enhancers using mouse limb buds; b) developing deep learning methods to analyze enhancer sequences and infer gene regulatory networks. We successfully determined limb-associated morphogenic enhancers and analyzed the characteristics of these sequences. Moreover, we developed a deep learning-based regulatory sequence classifier that outperformed previous studies. This software can extract information that is critical for transcriptional regulation from genomic sequences. As an output of this project, we have published one peer-reviewed original research paper and two original research papers as preprints and released the developed program as an open-source software.

研究分野：発生生物学

キーワード：深層学習 形態形成 転写制御配列

様式 C-19、F-19-1、Z-19（共通）

1. 研究開始当初の背景

ゲノム配列レベルにおける生物の形態形成メカニズムへの理解は、1次元的なゲノム情報から多様な3次元の形態がどのように生み出されるのかという基礎科学的に興味深いテーマであるとともに、再生医療などにおいて器官形成を人為的操作する上でも重要な土台となる。しかしながら、次の3つの問題点が存在する。**ゲノム配列の煩雑性**：制御配列は必ずしも、遺伝子座の近くにあるとは限らず、どの配列がどの遺伝子を制御しているのかを予測するのは困難である。**制御配列の定義の不透明性**：コーディング遺伝子と違い、制御配列には開始コドンや終始コドンといったわかりやすい配列がないため、制御配列の配列の特徴や、どんな転写因子が結合するかわからない。**時空間における転写制御ネットワークへの知識の欠如**：形態形成では時空間における遺伝子転写制御が必要となるが、現在多くの遺伝子制御関連の解析が培養細胞系で行われており、時空間における転写制御の知識が不足している。

本研究では、こうした課題を克服するため、①形態形成過程におけるオープンクロマチン領域（制御配列候補）をATAC-seq（後述）技術により網羅的に同定し、②ディープラーニングによるパターン認識アルゴリズムを開発することで、制御配列の特徴の同定および、ゲノム配列上における遺伝子発現と制御配列の関連付けを行う計画であった。

2. 研究の目的

本研究は、形態形成における全ゲノム転写制御ネットワークの高精度推定方法の確立を目的とした。遺伝子の転写制御関係を理解することは、最も基礎的かつ重要な分子生物学のテーマの一つであると同時に、困難なテーマでもある。特に形態形成過程における転写制御は、ダイナミックに変化し続けるので、転写制御の網羅的な同定は非常に挑戦的な課題である。本研究では、機械学習方法の一つであるディープラーニングを応用することにより、時系列のエピジェネティックデータ、全ゲノム配列比較解析から得たデータを元にした転写制御推定を網羅的に行う新規アルゴリズムを開発することで取り組んだ。

3. 研究の方法

四肢形成過程を主なモデルとして、A. 転写制御配列の網羅的同定、B. 転写制御配列の特徴の解析を行うことで、形態形成における転写制御ネットワーク推定、C. 軟骨魚類イヌザメの胸鰭形成でも同様の解析を行うことで、形態形成における転写制御のより一般的な原理に迫る。

（1）四肢形成過程における転写制御配列の網羅的同定

ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing) という次世代シーケンサーを応用した手法により、オープンクロマチン領域（遺伝子制御配列候補）を各発生段階で網羅的に同定する。転写活性のある遺伝子を同定するために、転写産物の網羅的解析方法であるRNA-seqも同時並行して行う。

（2）ディープラーニングを用いた、制御配列の特徴および制御関係の同定

ディープラーニングは、多層構造からなる人工的なニューラルネットワークの総称で、パターン認識などを行うための機械学習方法である。本研究では、特に空間パターン認識に優れた deep convolutional neural network (CNN) という手法を用いることで、ATAC-seqデータ特有のDNA配列パターンを抽出する。主に次の二つのパターン認識を行う。a) ローカルな制御配列の特徴を抽出することで、制御配列の特徴およびどの転写因子が結合しているかを同定する。b) グローバルなゲノムの組織性としてオープンクロマチン領域と転写活性と関係の特徴を抽出することで、転写制御ネットワークを推定する。

（3）他の種への拡大による転写制御の一般原理の抽出

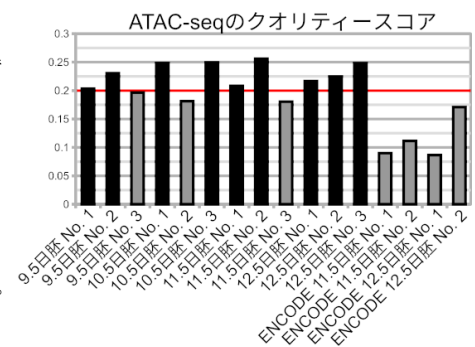
機械学習では、より多くのデータを学習サンプルとして取り入れることで、より精度の高いパ

ターン認識ができる。そこで、イヌザメという軟骨魚類の鰭を比較対象として取り入れることでより頑強な機械学習を行うことが期待される。

4. 研究成果

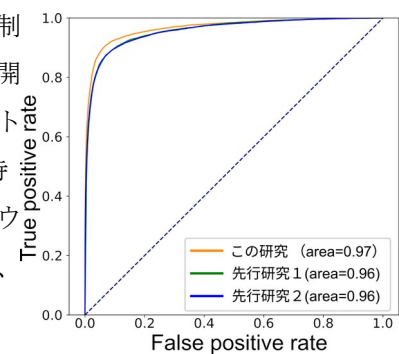
(1) 四肢形成過程における転写制御配列の網羅的同定について

9.5日胚、10.5日胚、11.5日胚、12.5日胚における前肢原基について、ATAC-seq、RNA-seqそれぞれ3個の反復データの取得に成功した。特に取得したATAC-seqデータは、ENCODEプロジェクトで公開されているものよりも、クオリティの高いものを得ることが出来た(右図)。イヌザメのヒレ原基においてもRNA-seqのデータは取得できたが、技術的問題から、ATAC-seqのデータ取得には至らなかった。また、予定していたディープラーニングへの応用を行う前に、既存のモチーフ解析プログラムや進化的な保存性の解析など様々な解析を行った。この結果、10.5日胚でにおいて、四肢形成に特異的な転写制御配列の活性頻度が高くなることが明らかとなった。また、これらの四肢特異的な転写制御配列は、進化的保存性が高いことや、gene ontology において”developmental process”、”multicellular organism development”などに分類される遺伝子を制御していることが示唆された。これらの解析結果をまとめ、論文化した[1]。

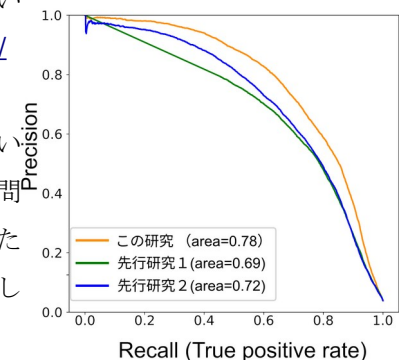


(2) ディープラーニングを用いた、制御配列の特徴および制御関係の同定について

ディープラーニングを応用することにより、ゲノム中の転写制御配列領域を予測し、制御配列の特徴を抽出するプログラムの開発に成功した。このプログラムでは、畳み込みニューラルネットワークと呼ばれる既存の手法を元に、DNA配列を取り込むのに特化した新奇の畳み込み層を開発することが出来た。右図は、マウスゲノム中のCTCF結合領域の予測精度を比較したものであるが、図のように、この研究で開発したプログラムが、先行研究を上回っていることが見て取れる。この成果について、論文化を行い[2]、プログラムについても、Githubにおいて公開した(<https://github.com/koonimaru/DeepGMAP>)。



転写配列と遺伝子の制御関係については、現在開発を行っている最中であるが、これについては先行研究結果が全て過学習の問題を抱えていたこと、テストデータに訓練データが混入していたことにより精度が過大評価されていたことなどがわかり、こうした問題を解決する必要がある。



(3) 他の種への拡大による転写制御の一般原理の抽出について

これについては、(2)で開発したプログラムが、非モデル動物にも応用可能であることを見出し、イヌザメとマウスのゲノム配列中の違いを抽出できる可能性が高く、今後の発展に期待したい。

【文献】

1. Onimaru K, Kaori Tatsumi, Chiharu Tanegashima, Mitsutaka Kadota, Osamu Nishimura SK (2020)

Preprint at bioRxiv, doi: 2020.01.10.901173.

2. Onimaru K, Nshimura O, Kuraku S 2018 Preprint at bioRxiv, doi: 10.1101/355974..

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 3件/うち国際共著 1件/うちオープンアクセス 5件）

1. 著者名 Onimaru Koh, Nishimura Osamu, Kuraku Shigehiro	4. 巻 -
2. 論文標題 A regulatory-sequence classifier with a neural network for genomic information processing	5. 発行年 2018年
3. 雑誌名 bioRxiv	6. 最初と最後の頁 355974
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1101/355974	査読の有無 無
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 Koh Onimaru, Fumio Motone, Itsuki Kiyatake, Kiyonori Nishida, Shigehiro Kuraku	4. 巻 247
2. 論文標題 A staging table for the embryonic development of the brownbanded bamboo shark (<i>Chiloscyllium punctatum</i>)	5. 発行年 2018年
3. 雑誌名 Developmental Dynamics	6. 最初と最後の頁 712 ~ 723
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1002/dvdy.24623	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 Koh Onimaru	4. 巻 -
2. 論文標題 The evolutionary origin of developmental enhancers in vertebrates: Insights from non model species	5. 発行年 2020年
3. 雑誌名 Development, Growth & Differentiation	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1111/dgd.12662	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 -
1. 著者名 Koh Onimaru, Shigehiro Kuraku	4. 巻 17
2. 論文標題 Inference of the ancestral vertebrate phenotype through vestiges of the whole-genome duplications	5. 発行年 2018年
3. 雑誌名 Briefings in Functional Genomics	6. 最初と最後の頁 352 ~ 361
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1093/bfpg/ely008	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 Koh Onimaru, Kaori Tatsumi, Chiharu Tanegashima, Mitsutaka Kadota, Osamu Nishimura, Shigehiro Kuraku	4. 巻 -
2. 論文標題 Developmental hourglass and heterochronic shifts in fin and limb development	5. 発行年 2020年
3. 雑誌名 bioRxiv	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1101/2020.01.10.901173	査読の有無 無
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 Koh Onimaru, Luciano Marcon	4. 巻 -
2. 論文標題 Systems biology approach to the origin of the tetrapod limb	5. 発行年 2019年
3. 雑誌名 arXiv	6. 最初と最後の頁 1907.0273
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
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

[学会発表] 計9件 (うち招待講演 3件 / うち国際学会 3件)

1. 発表者名 Koh Onimaru, Kaori Tatsumi, Chiharu Tanegashima, Mitsutaka Kadota, Osamu Nishimura, Shigehiro Kuraku
2. 発表標題 Systematic comparison of gene regulation in fin and limb development
3. 学会等名 JAPANESE SOCIETY of DEVELOPMENTAL BIOLOGISTS
4. 発表年 2019年

1. 発表者名 Koh Onimaru
2. 発表標題 Sequence-based prediction of regulatory genomic regions with an improved deep-learning method
3. 学会等名 International Society for Computational Biology (国際学会)
4. 発表年 2018年

1. 発表者名 鬼丸 洸
2. 発表標題 複合的アプローチによる鰭から四肢への形態進化における分子機序の研究
3. 学会等名 遺伝研研究集会「進化遺伝学における実験的研究と理論的研究の融合」(招待講演)
4. 発表年 2017年

1. 発表者名 Koh Onimaru, Luciano Marcon, Marco Musy, Mikiko Tanaka, James Sharpe
2. 発表標題 A deeply conserved Turing mechanism in the fin and limb development
3. 学会等名 the Pan-American Society for Evolutionary Developmental Biology (国際学会)
4. 発表年 2017年

1. 発表者名 Koh Onimaru, Kaori Tatsumi, Chiharu Tanegashima, Mitsutaka Kadota, Osamu Nishimura, Shigehiro Kuraku
2. 発表標題 Systematic comparison of gene regulation in fin and limb development
3. 学会等名 The 52nd annual meeting of the Japanese Society of Developmental Biologists
4. 発表年 2019年

1. 発表者名 Koh Onimaru, Kaori Tatsumi, Chiharu Tanegashima, Osamu Nishimura, Mitsukata Kadota, Shigehiro Kuraku
2. 発表標題 Omics approaches to the fin-to-limb evolution
3. 学会等名 EMBO Workshop: Limb Development and Regeneration: New Tools for a Classic Model System, Barcelona (国際学会)
4. 発表年 2019年

1. 発表者名 Koh Onimaru, Osamu Nishimura, Shigehiro Kuraku
2. 発表標題 Predicting gene regulatory regions with a convolutional neural network for processing double-strand genome sequence information
3. 学会等名 日本バイオインフォマティクス学会2019年年会 第8回生命医薬情報学連合大会 IIBMP2019
4. 発表年 2019年

1. 発表者名 鬼丸 洸
2. 発表標題 DNA上のタンパク質結合モチーフ配列予測を例に、機械学習を分子進化発生学に用いたご自身の研究について
3. 学会等名 Rinkai Hackathon 2019 with DDBJing (招待講演)
4. 発表年 2019年

1. 発表者名 鬼丸 洸
2. 発表標題 全ゲノム重複の痕跡を基にした外群に依存しない脊椎動物の祖先型推定
3. 学会等名 日本進化学会第20回大会 (招待講演)
4. 発表年 2018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
---------	---------------------------	-----------------------	----