

令和 3 年 6 月 21 日現在

機関番号：82609

研究種目：基盤研究(B)（一般）

研究期間：2016～2020

課題番号：16H02902

研究課題名（和文）ゲノムDNA転写開始活性の数理的構造を用いた制御領域リバーエンジニアリング

研究課題名（英文）Reverse engineering of transcriptional regulatory regions through data modeling of transcription initiation activities

研究代表者

川路 英哉（KAWAJI, Hideya）

公益財団法人東京都医学総合研究所・ゲノム医学研究センター・副センター長

研究者番号：20525406

交付決定額（研究期間全体）：（直接経費） 13,100,000円

研究成果の概要（和文）：塩基対単位での転写開始活性データから転写を近位・遠位より制御する非コードゲノム領域（プロモーター、エンハンサー）を予測する技術の改良を実施した。転写開始活性データ、ヌクレオソーム構造やヒストン修飾をはじめとする多階層のゲノム機能データを統合的に可視化する技術を開発した上で、転写の方向性（片側方向、両側方向）、様式（発散性、収斂性）を中心とした解析を実施し、主に様式を中心とした数理モデルを構築した。これに基づく予測技術を開発し、精度・感度共にそのゲノム位置に関わらず高い感度を達成した結果、これまでに見過ごされてきた転写制御領域までも同定することができた。

研究成果の学術的意義や社会的意義

転写制御領域は多くの場合、複数のエピジェネティック測定結果の組み合わせ（オープンクロマチン領域、ヒストン修飾など）によって同定されることが一般的であるが、本研究で取り組んでいるのは、転写開始活性という単一のデータセットからこれを同定する手法である。臨床研究などへ展開していく際には試料が限定されることから、数少ない測定で多くの情報が得られることが望ましい。本手法はそのような応用においても、高精度・高感度の転写制御領域同定に道を開くものである。

研究成果の概要（英文）：We successfully improved a method to predict promoters and enhancers, non-coding genomic region to regulate gene expression from neighboring and distal regions, based on transcription initiation activities. Firstly we performed integration of multiple layers of functional genomic data, such as transcription initiation activities, nucleosome structure, and histone modification. Next we revisited the feature of transcription initiation activities from the perspectives of directionalities (unidirectional, bidirectional) and architecture (convergent and divergent). We developed a new prediction method based on the examined feature, in particular the architecture, and succeeded to increase precision and recall of the predicted results. Finally we identified a set of overlooked regulatory regions in the human genome.

研究分野：ゲノミクス

キーワード：転写制御 転写開始点 プロモーター エンハンサー

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

生命の遺伝情報はゲノム DNA にコードされており、生物を構成する最小単位である細胞はそれが必要とする分子を適切なタイミングで、かつ適切な量だけ合成することで、その生命を維持し自律的活動を行なっている。DNA を鋳型として RNA を合成する転写は、その合成の第一段階であり、細胞における転写機構の解明は、現在においても中心的命題である。転写の契機となる現象を実験的に測定するため、日本でこれまでに培われてきた完全長 cDNA 作成技術と次世代シーケンサーを組み合わせ、転写開始活性を一塩基単位でゲノムワイドに定量する CAGE 法の開発に申請者らはこれまで携わってきた。哺乳類ゲノムの機能解明を目指す国際共同研究プロジェクト FNATOM における CAGE 解析を通じて、転写開始活性パターンの分類や一塩基単位での変化(ずれ)などを明らかにした他、遺伝子近傍から転写を制御する領域(プロモーター)、遠位より制御する領域(エンハンサー)などを網羅的に同定した。

2. 研究の目的

細胞の分子的基盤を形成する上での基本プロセスである転写について、これを統御する転写制御領域を、高解像度かつゲノムワイドに定量された転写開始活性から推定する技術を改良する。近位・遠位より転写に影響を与えるプロモーター、エンハンサーのサブクラスを始めとする様々な転写制御領域における転写開始活性の構造を抽象化した数理モデルを構築することで、転写の結果である転写活性プロファイルより転写の要因である転写制御領域を推定(リバースエンジニアリング)し、精度向上を実現する。

3. 研究の方法

(1) ヌクレオソーム構造データの取得

転写が起きない状態のゲノム DNA はヒストン・タンパク質に巻きついたヌクレオソーム構造をとるなど、転写反応とその物理的基盤は密接に関係している。そこで、トランスポザゼ Tn5 を用いたヌクレオソーム構造の実験的測定(ATAC-seq)を実施する。

(2) 転写開始活性データの取得

転写開始活性を一塩基対単位で測定する方法として、メッセンジャーRNAや長鎖非コードRNAの5'末端を選択的に配列決定する CAGE 法がこれまで効果的に用いられてきた。細胞に含まれる全ての RNA (トータル RNA) を対象とする当該手法において、細胞内で安定的に保持される RNA は転写活性に比して高頻度で観測される一方、転写後すぐに分解される RNA は相対的に低頻度で観測される。つまり転写後調節によるバイアスが観測結果に含まれ、転写開始活性そのものはダイレクトに反映されていないという懸案があった。一方、エンハンサーに由来する RNA (eRNA) を高感度に補足することを目的として、転写直後の新生 RNA を実験的に分画しこれを CAGE 法により測定する手法(NET-CAGE 法)を村川らと開発した¹。新生 RNA は転写後の分解や細胞質での安定化を受ける前の状態であることから、当該測定結果は転写開始活性をダイレクトに反映することが期待される。そこで本手法を用いた転写開始活性データの測定を実施する。

(3) 転写開始活性やエピゲノム等といった機能ゲノミクスデータの統合

主たる解析対象である転写開始活性データを解釈する上では、転写反応の物理的基盤であるヌクレオソーム構造のみならず、ヒストンや DNA 等の化学修飾(エピジェネティック・マーク)、これらの統合解析処理した結果、等との塩基対レベルでの精査が必要である。そこで、広く利用されているデータが数多く編纂されている UCSC ゲノムブラウザ・データベースを用い、独自に取得したデータをも扱うことで、データ可視化、抽出、再計算を円滑に行う機能ゲノミクスデータの統合解析環境を整備する。

(4) 転写開始の数理的構造抽出

転写開始を特徴付けるパラメータとして、転写の向き(センス鎖、アンチセンス鎖)、転写制御領域の幅、転写開始活性のレベルなどが考えられる。そこで、これらと他のエピジェネティックマーク等との相関を検討した。特に、ゲノム各領域の機能を低い解像度ながらも多面的に特徴づけるゲノムセグメンテーション、DNA-タンパク質複合体のオープン・クローズを 1 ヌクレオソーム単位で測定した DNaseI-seq、ATAC-seq データとの相関関係を精査する。また研究の進行中、新しいデータセットとして、ヌクレオソーム単位でプロモーター・エンハンサーを判別したデータ(ENCODE cCREs)が公開されたことから、これとの相関関係解析も追加的に実施する。その他、刺激応答等によって遺伝子発現が変化する際には、エンハンサーの活性化が先行するという制御モデルが知られていることから、これと転写開始活性シグナルの変化に着目した精査を行う。これらによって上記で得られた転写開始活性の数理的構造を基に、転写制御領域を予測するモデルを構築し、さらにこれを実データに対して適用する。

4. 研究成果

本研究で主たる解析対象とするモデルとして、これまで既に蓄積されてきたデータなどを考慮し、乳がん由来細胞株 MCF-7 を選定した。この細胞より測定したヌクレオソーム構造、転写開始活性データに加え、公開されている同細胞株への測定結果、機能アノテーション結果を統合的に精査する環境を整えた。世界中で試みられている測定や解析結果をタイムリーに取り込んだ解析や評価を実施するため、trackhub と呼ばれるデータ共有フレームワークの利用を検討した。これは公開データと自身のデータを統合的に扱うことを可能とする強力なフレームワークである一方、その環境設定が煩雑であるという困難があった。そこで、統合対象となるデータファイルのディレクトリ構造を参照することで環境を自動的に整える技術を開発した²。これによって主たる解析対象である MCF-7 細胞のみならず、他細胞由来のデータの統合的な扱いも極めて円滑にできるようになった他、研究開始時点では予想されていなかったデータが公開された折にも、これを用いた統合解析環境を速やかに具現化することができた。当該プログラムは本研究に用いられた他、一般の研究者も利用可能な形として公開しており、データベースの構築などにも活用されている。

統合化された複数階層データを用い、エピジェネティックマーク等を参照しつつ転写活性の数理的構造の抽出を試みた。双方向に転写されている領域の同定がエンハンサー領域の予測につながることを確認する一方、エンハンサー領域にも片側方向にしか転写が起きていない領域が存在することも見出された。また、エンハンサーと対比されることの多いプロモーターについても、双方向に転写されるものも存在する。そこで、転写開始の方向性が片側方向(unidirectional)、両側方向(bidirectional)を対立軸として捉えるのではなく、これら転写制御領域はいずれも発散的(divergent)な転写が起きている領域と位置づけることで、プロモーター・エンハンサーの統一的な枠組みを構築した。発散的な転写が起きている領域はおよそヌクレオソームと対応していることも見出し、これは転写反応の物理的基盤であるヌクレオソーム構造であるという知見と一致するものであった。

併せて、遺伝子発現の動的な変化に際して、エンハンサーとそれに対応するプロモーターの活性化タイミングに関する検討を行った。トータル RNA を対象とする CAGE で測定された転写開始活性では、活性がピークに達する時期がプロモーターではエンハンサーと比して遅くなることが過去に報告されており、同様の現象が確認できた。一方、新生 RNA を対象とする NET-CAGE を用いて測定された転写開始活性では、プロモーター、エンハンサー活性がピークに達する時期はほぼ同時であることが見出された。エンハンサーとそれが制御するプロモーターの関係が逐一明らかになっていない現在においては一般化することは難しいものの、少なくとも観測されたペアにおいてはプロモーターとエンハンサーのタイミングに関する階層性はみられなかった。これら二つのクラスの転写制御領域の活性化に関する厳密な差は認められなかったことから、プロモーターとエンハンサーを同列に扱う枠組みに変更は必要とならないと判断した。

これら数理的枠組みの下、CAGE データより転写制御領域を予測する技術(プログラム)の開発に取り組んだ。既存手法では予測される転写制御領域が遺伝子より離れた場所に位置するものと、イントロンを含む遺伝子内に位置するものいずれかに偏る傾向があったが、新規に開発した手法ではその双方を高感度に検出することに成功した。更にエクソン領域によって誘発されるノイズをデータ処理によって一定程度抑えることにも成功し、イントロンなどの遺伝子内に存在するエンハンサー領域を積極的に同定することができるようになった。最終段階として、この結果を確認する実験測定に着手した。数百以上も存在する対象領域の活性を体系的に測定するため、超並列レポーターアッセイの準備を行った。これらの測定を実施することで、新たに同定したエンハンサー領域候補の機能的検証が期待される。

<引用文献>

1. Hirabayashi, S., Bhagat, S., Matsuki, Y., Takegami, Y., Uehata, T., Kanemaru, A., Itoh, M., Shirakawa, K., Takaori-Kondo, A., Takeuchi, O., Carninci, P., Katayama, S., Hayashizaki, Y., Kere, J., Kawaji, H., Murakawa, Y., 2019. NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements. *Nat. Genet.* 51, 1369-1379.
2. Kawaji, H., 2018. dirHub: a trackHub configurator with directory structure projection. *bioRxiv*. <https://doi.org/10.1101/314807>

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 5件/うち国際共著 4件/うちオープンアクセス 5件）

1. 著者名 Hideya Kawaji	4. 巻 May
2. 論文標題 dirHub: a trackHub configurator with directory structure projection	5. 発行年 2018年
3. 雑誌名 bioRxiv	6. 最初と最後の頁 4
掲載論文のDOI（デジタルオブジェクト識別子） 10.1101/314807	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 Lizio M, Harshbarger J, Abugessaisa I, Noguchi S, Kondo A, Severin J, Mungall C, Arenillas D, Mathelier A, Medvedeva Y, Lennartsson A, Drabls F, Ramiłowski J, Rackham O, Gough J, Andersson R, Sandelin A, Ienasescu H, Ono H, Bono H, Hayashizaki Y, Carninci P, Forrest A, Kasukawa T, Kawaji H	4. 巻 45
2. 論文標題 Update of the FANTOM web resource: high resolution transcriptome of diverse cell types in mammals	5. 発行年 2016年
3. 雑誌名 Nucleic Acids Research	6. 最初と最後の頁 D737 ~ D743
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/nar/gkw995	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 Abugessaisa I, Ramiłowski J, Lizio M, Severin J, Hasegawa A, Harshbarger J, Kondo A, Noguchi S, Yip C, Ooi J, Li C, Tagami M, Hori F, Agrawal S, Hon Chung C, Cardon M, Ikeda S, Ono H, Bono H, Kato M, Hashimoto K, Bonetti A, Kato M, Kobayashi N, Shin J, de Hoon M, Hayashizaki Y, Carninci P, Kawaji H, Kasukawa T.	4. 巻 49
2. 論文標題 FANTOM enters 20th year: expansion of transcriptomic atlases and functional annotation of non-coding RNAs	5. 発行年 2020年
3. 雑誌名 Nucleic Acids Research	6. 最初と最後の頁 D892 ~ D898
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/nar/gkaa1054	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 Hirabayashi Shigeki, Bhagat Shruti, Matsuki Yu, Takegami Yujiro, Uehata Takuya, Kanemaru Ai, Itoh Masayoshi, Shirakawa Kotaro, Takaori-Kondo Akifumi, Takeuchi Osamu, Carninci Piero, Katayama Shintaro, Hayashizaki Yoshihide, Kere Juha, Kawaji Hideya, Murakawa Yasuhiro	4. 巻 51
2. 論文標題 NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements	5. 発行年 2019年
3. 雑誌名 Nature Genetics	6. 最初と最後の頁 1369 ~ 1379
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41588-019-0485-9	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 該当する

1. 著者名 Oki Shinya, Ohta Tazro, Shioi Go, Hatanaka Hideki, Ogasawara Osamu, Okuda Yoshihiro, Kawaji Hideya, Nakaki Ryo, Sese Jun, Meno Chikara	4. 巻 19
2. 論文標題 ChIP Atlas: a data mining suite powered by full integration of public ChIP-seq data	5. 発行年 2018年
3. 雑誌名 EMBO reports	6. 最初と最後の頁 e46255
掲載論文のDOI (デジタルオブジェクト識別子) 10.15252/embr.201846255	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 Lizio Marina, Abugessaisa Imad, Noguchi Shuhei, Kondo Atsushi, Hasegawa Akira, Hon Chung Chau, deHoon Michiel, Severin Jessica, Oki Shinya, Hayashizaki Yoshihide, Carninci Piero, Kasukawa Takeya, Kawaji Hideya	4. 巻 47
2. 論文標題 Update of the FANTOM web resource: expansion to provide additional transcriptome atlases	5. 発行年 2018年
3. 雑誌名 Nucleic Acids Research	6. 最初と最後の頁 D752 ~ D758
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/nar/gky1099	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

[学会発表] 計6件 (うち招待講演 2件 / うち国際学会 4件)

1. 発表者名 Hideya Kawaji
2. 発表標題 dirHub: a trackHub configurator with directory structure projection
3. 学会等名 Biology of Genome (国際学会)
4. 発表年 2018年 ~ 2019年

1. 発表者名 Bhagat S, Hirabayashi S, Matsuki Y, Takegami Y, Uehata T, Kanemaru A, Itoh M, Shirakawa K, Takaori-Kondo A, Takeuchi O, Carninci P, Katayama S, Hayashizaki Y, Kere J, Kawaji H, Murakawa Y
2. 発表標題 5' nascent RNA profiling of cis-regulatory elements
3. 学会等名 Takeda Symposium (国際学会)
4. 発表年 2018年 ~ 2019年

1. 発表者名 川路英哉
2. 発表標題 転写開始イベントの高精度・高解像度プロファイルが明らかにするタンパク質非コード領域とその複雑な制御
3. 学会等名 京都大学ウイルス・再生医科学研究所第3回生命情報研究会（招待講演）
4. 発表年 2017年

1. 発表者名 Kawaji H.
2. 発表標題 “dogma (model)” conscious and agnostic data sharing
3. 学会等名 Biohackathon 2016 Symposium (国際学会)
4. 発表年 2016年

1. 発表者名 Kawaji H.
2. 発表標題 Precise annotation of TSS peaks uncovering complexity in state-dependent activation of transcription initiation
3. 学会等名 Genome Informatics (国際学会)
4. 発表年 2016年

1. 発表者名 Kawaji H.
2. 発表標題 Precise annotation of TSS peaks uncovering complexity in state-dependent activation of transcription initiation
3. 学会等名 Study Sessions on Bioinformatics and Related Topics (at IFRc, Osaka Univ.) (招待講演)
4. 発表年 2016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
---------	---------