

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 6 日現在

機関番号：12601

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500360

研究課題名(和文) 超離散時空間解析と大規模配列解析の融合による遺伝子転写原理の解明

研究課題名(英文) Ultra-discrete model and ChIP-Seq analysis for gene transcription

研究代表者

大田 佳宏(Ohta, Yoshihiro)

東京大学・数理(科)学研究科(研究院)・特任教授

研究者番号：80436592

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：ヒト細胞を用いたRNAPII (RNA polymerase II) 実体の運動情報やエピジェネティック修飾情報の数理解析の研究を行った。これまでの我々の研究では、時間分解能 7.5 分間隔で細胞実験を行っていたが、本研究課題ではより高い時間分解能の転写におけるRNAデータの取得を行った。さらに、転写ダイナミクスを制御しその運動にも影響を与える因子として、各種タンパク質結合情報とエピジェネティック修飾情報も取得した。これらの細胞実験のデータをもとに、データマイニングの技術を応用し、転写の運動・修飾情報を染色体上の数値データとしてパスプリファレンスを含めた数理解析を行うプログラムを開発した。

研究成果の概要(英文)：Previous models of traffic flow of RNA polymerase II (RNAPII) during transcription were restricted to one dimension along the DNA template. Here we modeled an application of traffic flow that allows preferential paths of different dimensions only restricted to visit some transit points. According to its position, an RNAPII protein molecule prefers paths obeying two types of time-evolution rules. One is an asymmetric simple exclusion process along DNA, and the other is a three-dimensional jump between transit points in DNA where RNAPIIs are staying. Simulations based on our model, and comparison experimental results, reveal how RNAPII molecules are distributed at the DNA-loop-formation-related protein binding sites as well as CTCF insulator proteins. As time passes after the stimulation, the RNAPII density at these sites becomes higher. Apparent far-distance jumps in one dimension are realized by short-range three-dimensional jumps between DNA loops.

研究分野：数理生物学

キーワード：超離散 配列解析 遺伝子転写

1. 研究開始当初の背景

遺伝子の DNA 配列を鋳型に、RNAPII という酵素によって RNA が作られ、その RNA 配列からアミノ酸やタンパク質が作られることは、生命の基本原則と考えられている。

しかし、ヒトの細胞核内の DNA は遺伝子の情報を持つ部分が飛び飛びに存在し、しかも、DNA がヒストンというタンパク質に巻き付いて染色体を構成するため、DNA から RNA が作られる局面を実際に観察するのは非常に困難だった。一方で近年、ゲノム解読からヒト染色体の持つ DNA 配列が明らかになり、さらに大規模配列解析技術の進歩により、遺伝子転写によって RNA が作られて行く全体像を解読することが可能となってきた。

関連分野の国内外の研究動向としては、1998年に Jülicher らが polymerization に関する確率論的動力学モデルを提唱した。同時期に、Hippel らは配列依存性のある転写の熱力学解析についての論文を発表し、近年では、Chowdhury らが離散数学を用いた粒子交通量理論を、転写の RNAP 渋滞モデルに応用した論文を発表した。しかし、これらの研究はすべてアルゴリズムの特性上、膨大な計算量を必要とするため、原核生物における DNA 上の局所的な数理解析を行ったにすぎず、現実の細胞を用いた実験データに適用できるものではない。さらに、転写モデルとしても非常に単純化したモデルを扱っており、エクソン・イントロン構造やタンパク質結合部位なども考慮していなかった。

2. 研究の目的

我々は、若手研究(B)[H22.4~H24.3]の成果として、ヒトの長い遺伝子 (>100kbp) の転写運動の数理解析と超離散シミュレーションに世界で初めて成功した。ここでは、エクソン・イントロン構造とタンパク質結合部位などにおける RNAPII の粒子速度変化を転写モデルに取り入れた。さらに、この速度変化を起こす DNA 領域周辺においてはヒステリシス現象がおきることを発見し、RNAPII の運動状態遷移のソリトン系厳密解も導出した。この我々の新規転写モデルは、ヒトの長い遺伝子の実験データを非常によく再現している。本成果は、アメリカ物理学会誌 Physical Review に掲載されている。

さらに我々は、これらの遺伝子において生成されるエクソニック・イントロニック RNA の発現量シミュレーションにも成功した。パラメータとしては細胞の確率密度分布 $f_{cell}(t)$ に加えて、RNA の減衰係数 λ 、TSS 領域蓄積変数 (μ, ν) 、RNAPII の個数 M 、複数 RNAPII(k) の協調運動におけるセル (j) における時間間隔 $\{ \tau_{k,j} \}$ などを用いた。

学習に用いた SAMD4A 遺伝子に関する再現性が非常に高いのは当然であるが、この我々の新規転写モデルは、Validation を行っていない他の長い遺伝子 EXT1, ZFPM2,

ALCAM, NFKB1 などについても極めて高い再現性を示し、長い遺伝子についての Universality を証明することができた。

本研究では、この我々の最新研究成果をさらに応用して、平成 25 年度までに、短い遺伝子に特徴的な RNAPII 粒子動態の超離散渋滞モデルとヒステリシス現象の解明と、クロマチン相互作用など染色体 3 次元構造の変化による RNAPII 転写運動のゆらぎのモデル化の研究成果を発表し、平成 26 年度には、複数 RNAPII のソリトン系協調運動を制御する因子の分子生物学的意味の解明の研究成果を発表することで、3 つの独創的な新規転写制御の概念を発表する。本研究によって、癌細胞などの細胞特有の転写メカニズムの解明が可能になれば、難治疾患の新規治療法の開発など、さらなる大きな飛躍が期待できる。

3. 研究の方法

遺伝子の転写の原理解明の研究には、その高い時間変異性と微小不均一性から、多くの困難を伴うが、本研究では我々の最新研究成果を踏まえて、遺伝子転写の時間・空間における依存性を精密に測定し、そこで得られる膨大な観察事象の数理解析から原理のモデル化を行い、さらにそのモデルの数値シミュレーションによって遺伝子の転写原理を探求するという、実験と数理解析の綿密な連携によって、この困難を克服し、新しい転写原理を発見する。

研究計画・方法としては、ヒト細胞を用いた RNAPII 実体の運動情報・各種タンパク質結合情報・エピジェネティック修飾情報の大規模解析を行い、そこで得られた膨大で、かつ断片的な観察事象を有機的に結合し、新原理の発見へと導くために、遺伝子転写における RNAPII 実体の粒子動態モデリングと超離散シミュレーションを行い、さらに、この RNAPII 粒子動態の座標情報を引数として、動的 RNAPII から生成される RNA 発現量のソリトン系波動態コンピューティングを行う。これら 3 つの研究項目による生物医学実験解析と応用数理解析の融合研究を行い、ヒト遺伝子の転写原理を探求することで、それらを制御することも可能となる。最終的には、ここで得られた遺伝子転写原理を応用して、癌などの新しい治療法開発につなげることを目指す。

4. 研究成果

ヒト細胞を用いた RNAPII (RNA polymerase II) 実体の運動情報・各種タンパク質結合情報・エピジェネティック修飾情報の大規模データ数理解析の研究を行った。具体的には、タイリングアレイなどを用いた実験において、正常ヒト臍帯静脈内皮細胞 HUVEC の TNF 刺激における RNAPII 実体の運動情報を取得した。これまでの我々の研究では、時間分解能 7.5 分間隔で細胞実験を行っていたが、

本研究課題ではより高い時間分解能として3分間隔での転写におけるRNAデータの取得を行った。さらに、転写ダイナミクスを制御しその運動にも影響を与える因子として、各種タンパク質結合情報とエピジェネティック修飾情報も取得した。具体的な結合タンパク質としては、CTCF/Cohesin, HNF4A, GATA2などの結合データを取得した。エピジェネティック修飾情報に関しては、H3K4me3, H4K36me3などの修飾情報のデータを取得した。

これらの細胞実験のデータをもとに、データマイニングの技術を応用し、転写の運動・修飾情報を染色体上の数値データとして数理解析を行うプログラムを開発した。本技術によって、細胞のRNAPII実体の運動情報・各種タンパク質結合情報・エピジェネティック修飾情報を、染色体上に1塩基レベルの高精度な空間分解能として取得することが可能となった。例えば、エクソン・イントロン領域ごとの断片配列スコアを計算し、転写において重要な役割を果たすスプライシングバリエーション情報を自動抽出するなど、大規模データからの高度な時空間数理解析も可能とした。また、FOS, ICAM, VCAM 遺伝子などの特に重要と考えられる遺伝子についてはより詳細な転写解析を行った。

また、RNAPIIの速度変化、停止、バックトラック、相互作用などの、詳細なダイナミクスをより精細に再現するため、箱玉系に代表されるソリトン CA モデルの Totally Asymmetric Simple Exclusion Process (TASEP) をベースに、複数RNAPIIが相互作用をしながら移動していく蓄積排他モデルを改良した。転写運動の障害となりうる各種タンパク質の時空間情報を ChIP-Seq データから取得し、そこで得られた情報をもとに独自 CA モデルを用いて、短い遺伝子に特徴的な RNAPII 粒子動態の超離散渋滞モデルとヒステリシス現象の解明に関する研究を行った。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計4件)

Maejima T, Inoue T, Kanki Y, Kohro T, Guoliang Li, Ohta Y, et al. "Direct Evidence for Pitavastatin Induced Chromatin Structure Change in the KLF4 Gene in Endothelial Cells", PLoS One. 2014 May 05;9(5):e96005. doi: 10.1371/journal.pone.0096005.

Ohta Y and Ihara S. "Ultradiscrete Modeling and Simulation for Gene Transcription", RIMS Kokyuroku Bessatsu, The breadth and depth of nonlinear discrete integrable systems,

pp 101 - 124, 2013.

Ohta Y, Nishiyama A, Wada Y, Ruan Y, Kodama T, Tsuboi T, Tokihiro T, Ihara S. Path-preference cellular-automaton model for traffic flow through transit points and its application to the transcription process in human cells", Physical Review E 86, 021918, (2012) [11 pages].

Kawamura T, Ogawa Y, Nakamura Y, Nakamizo S, Ohta Y, et al. "Severe dermatitis with loss of epidermal Langerhans cells in human and mouse zinc deficiency", The Journal of Clinical Investigation, 122(2): 722 - 732, 2012.

[学会発表](計6件)

大田佳宏, 「転写機構解明のための時空間数理モデル」, 生命動態システム科学四拠点・CREST・PRESTO 合同シンポジウム「生命動態の分子メカニズムと数理」, 2015年3月16-17日 京都大学 芝蘭会館 稲盛ホール.

大田佳宏, 「遺伝子の転写機構解明のための数理モデル」, Seminar on Mathematics for various disciplines (諸分野のための数学研究会), 2015年1月27日, 東京大学.

大田佳宏, 「転写機構解明のための数理モデルとシミュレーション」, 日本数学会(2014年度年会)「数学連携ワークショップ -生命科学、材料科学における数理-」(主催:文部科学省、統計数理研究所) 2014年3月16日, 学習院大学.

大田佳宏, 「転写過程の情報処理と数理モデル」, 数学協働プログラム「生命ダイナミクスの数理とその応用」(主催:統計数理研究所), 2014年1月22日, 東京大学大学院数理科学研究科 大講義室.

大田佳宏, 「転写の CA モデルとシミュレーション」, 研究集会 [島根大学[数理生物]-東京大学 iBMath 合同研究会:生命動態の実験, 数理モデルおよびシミュレーションの現状と今後の課題], 2013年12月27日, 島根県 一畑東館.

大田佳宏, 「転写過程の超離散モデリングとシミュレーション」, 京都大学数理解析研究所研究集会 [非線形離散可積分系の拡がり], 益川ホール, Aug. 21, 2012.

[図書](計2件)

大田佳宏, 井原茂男, "遺伝子転写機構

の時空間シミュレーション",「生体の科学」(医学書院) Vol.65, No.5, p.460-461, 2014 Sep.-Oct.

井原茂男, 大田佳宏, "真核生物の転写機構の数理解析",「生体の科学」(医学書院) Vol.65, No.5, p.458-459, 2014 Sep.-Oct.

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

6. 研究組織

(1)研究代表者

大田 佳宏(OHTA, Yoshihiro)

東京大学・大学院数理科学研究科・特任教授

研究者番号：80436592

(2)研究分担者

井原 茂男(IHARA, Sigeo)

東京大学・先端科学技術研究センター・特任教授

研究者番号：30345136