

令和 6 年 5 月 16 日現在

機関番号：14401

研究種目：若手研究

研究期間：2020～2023

課題番号：20K14361

研究課題名（和文）一細胞データから遺伝子の制御構造を定量推定するための確率・統計理論の構築

研究課題名（英文）Quantifying and clarifying gene regulatory structure from single-cell level data

研究代表者

飯田 溪太 (Iida, Keita)

大阪大学・蛋白質研究所・准教授

研究者番号：10709653

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、1細胞レベルで観察されている遺伝子発現のゆらぎ・ばらつきの確率ダイナミクスを数学的に定式化することで、実験では直接観察することの難しい転写制御パラメータをスナップショットの1細胞データ（mRNA量や蛋白質量）から定量的に推定する方法を開発した。この中で、遺伝子産物の確率的な生成・消滅を表す方程式を導出し、さらにその定常解（確率モデル）を数学的に厳密に導出する方法を発見した。この理論を大腸菌ラクトース代謝系と出芽酵母ガラクトース代謝系のデータにそれぞれ適用することで、転写制御に関する生物パラメータを定量的に推定し、さらに両者に共通した転写制御の機構を見出すことができた。

研究成果の学術的意義や社会的意義

近年、一細胞RNAシーケンスなどの技術進展により、全遺伝子の発現情報から発生段階などの細胞の長期的な状態変容を予測する研究が進展している。しかし、こうした一細胞データには遺伝子に特有のノイズやばらつきが含まれるため、決定論にもとづく従来のデータ解析手法では、転写制御などの生物パラメータを定量的に推定することは困難であった。そこで、本研究では確率過程論を用いたアプローチにより一細胞のスナップショットデータから転写制御パラメータを推定する方法を確立し、原核生物、真核生物の実データに適用可能であることを実証した。この手法を正常とがん細胞の比較に用いることで臨床応用にもつながることが期待される。

研究成果の概要（英文）：In this study, we mathematically formulated the stochastic dynamics of fluctuations and cell-to-cell variations in gene expression observed at the single-cell level. Using snapshot single-cell data, we developed a method to quantitatively estimate transcriptional control parameters that are difficult to observe directly in experiments. This includes estimating the amounts of mRNA and protein.

During this process, we derived an equation that models the stochastic production and degradation of gene products, and we discovered a method to rigorously derive its steady-state solution (probabilistic model). By applying this theory to single-cell data from the *Escherichia coli* lactose metabolic system and the *Saccharomyces cerevisiae* galactose metabolic system, we quantitatively estimated the kinetic parameters in gene regulation and uncovered a transcriptional control mechanism common to both systems.

研究分野：応用数学

キーワード：確率過程論 数理モデル論 超幾何関数 ベイズ推定 一細胞解析 遺伝子発現

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

近年、単一細胞 RNA シークエンスなどの技術が進展し、すべての遺伝子に対する一細胞レベルのスナップショットデータから転写速度などの生物パラメータを網羅的に推定することで、細胞の長期的な状態遷移を予測する研究が進展している (La Manno et al., Nature, 2018)。しかしながら、一細胞データには生物種や遺伝子に特有のノイズやばらつきが含まれ、かつデータの多くは絶対定量でないため、レート方程式 (決定論) や離散確率過程にもとづく従来の解析手法では、こうした速度パラメータを精度よく推定することは困難であった。特に、研究の律速となっている問題は、遺伝子制御の定量的な推定が実験と理論のいずれにおいても困難な点であり、このことが遺伝子発現解析による細胞内反応過程の予測を困難なものにしている。

こうした中、申請者は連続変数の確率過程を用いた数理モデリング手法を研究することで、一細胞由来の遺伝子産物量の計測値から、転写のオン・オフ状態の遷移確率を推定するための理論を構築してきた。その結果、転写のオン・オフ遷移確率が、大腸菌のラクトース代謝系遺伝子の発現分布に重要な役割を果たすことを明らかにした。この結果から、申請者らの解析パイプラインを真核生物の遺伝子発現系に適用できるよう一般化すれば、これまで以上に多くの遺伝子に対して、一細胞データに隠された遺伝子制御の機構を解明できると考えた。

2. 研究の目的

本研究では、申請者らの開発した転写のオン・オフ状態つき確率遺伝子発現モデルを、真核生物の遺伝子発現に適用できるよう一般化し、出芽酵母代謝系などの転写制御に適用することで、原核生物と真核生物のいずれにも適用可能な手法を確立する。また、この手法を原核生物と真核生物の一細胞データ解析に適用し、両者に共通の転写制御のキネティックパラメータを定量推定することで、より普遍的な転写制御の機構解明を目指す。

3. 研究の方法

本研究では、生物の遺伝子発現制御に共通の機構を一細胞データから推定するため、以下の3点を明らかにする：

1. 【確率モデルの導出】

先行研究により、高等生物の遺伝子制御は非マルコフ的な状態遷移を持つことが示唆されている (Suter et al., Science, 2011)。これに対して、準マルコフ過程を用いた定式化が提案されているが、その方程式を解くことは困難である (Zhou and Liu, Quant. Biol., 2015)。そこで、カテゴリカル過程を用いて真核生物の遺伝子発現に現れる不応期を表現し、その確率モデルを導出する。

2. 【パラメータ推定法の開発】

予備検討により、上述の確率モデルは数値的評価が難しいことが分かっている。そのため、マルコフ連鎖モンテカルロ法 (MCMC) などのベイズ推定の枠組みは直接的には適用が難しい。そこで、数値的評価が難しい確率モデルに対する効率的なパラメータ推定法として古典的な MCMC に代わる手法を開発する。

3. 【出芽酵母のデータを用いたパラメータ推定】

出芽酵母は真核生物であり、一部、大腸菌に類似した転写制御構造を持つが、その機構はより複雑である (Gnugge et al., ACS Synth. Biol., 2016)。したがって、出芽酵母の遺伝子発現データは、本研究で開発する確率・統計理論の性能評価に有用である。そこで、「2.」で開発したパラメータ推定法をこのデータに適用し、転写のオン・オフ状態の遷移機構を明らかにする。

4. 研究成果

2020-2021年度は、真核生物に見られる蛋白質の生成崩壊について、連続変数の確率過程を用いた定式化を行った。転写のオン・オフ状態遷移については「OFF OFF ... OFF ON」がサイクルする確率過程 (Schwabe et al., Sci Rep, 2014) を採用し、転写の状態がオンの時に蛋白質分子が生成する連続変数マルコフモデル (Friedman et al., Phys Rev Lett, 2006) と連結させることで独自の数理モデルを構築した。さらに、この確率過程から一般化フォッカー・プランク方程式を導出することで定常確率密度関数 (確率モデル) の解析を行った。確率モデルは積分空間において一般超幾何関数によって記述されたが、これを逆変換する初等的な方法は知られていなかった。そこで、報告者は複素関数論とポストの逆変換公式 (Post, 1930) を組み合わせた新たな逆変換手法を発見し、当初の問題を解決した。これにより、目標の一つである確率モデルの数学的妥当性の証明が完了した。この研究成果は、国内会議で報告した。

2022-2023年度は、前年度までに導出した確率モデル (細胞ごとの遺伝子産物量のばらつき) のパラメータ推定の問題に取り組み、原核生物および真核生物の一細胞遺伝子発現データから、分子の崩壊係数以外のすべてのモデルパラメータ (転写のオンオフ速度、最大転写効率、転写バ

ーラストサイズ)を同時かつロバストに推定する方法を開発した。これまでは、転写のオン・オフ確率のみが推定の対象であったが、この方法では転写効率などが公開データなどから事前にわかっている遺伝子にしか適用できなかった。そこで、ベイズ事前確率を導入し、さらにモデルパラメータに細胞間での多様性を考慮することで、事前知識の必要性を大幅に緩和することができた。その結果、従来以上の推定精度を実現することができ、確率モデルとパラメータ推定法の汎用性が向上した。ただし、この結果は転写のオン・オフ状態の遷移が2状態間を遷移する場合にのみ適用可能であり、より一般の場合(OFFが長く続く場合)には、確率モデルの表示の導出まではできるものの、その数値的評価は困難である。そこで、低次の条件付きモーメントを利用したパラメータ推定法の開発も並行して行った。この方法は、計算時間の面からまだ実用化には至っていないが継続して開発を続ける予定である。他方、近年注目されている多様体上のマルコフ連鎖モンテカルロ法の理論を応用することで、より効率的な計算法の開発を行う。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Iida Keita, Kondo Jumpei, Wibisana Johannes Nicolaus, Inoue Masahiro, Okada Mariko	4. 巻 38
2. 論文標題 ASURAT: functional annotation-driven unsupervised clustering of single-cell transcriptomes	5. 発行年 2022年
3. 雑誌名 Bioinformatics	6. 最初と最後の頁 4330 ~ 4336
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/btac541	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9件（うち招待講演 4件/うち国際学会 1件）

1. 発表者名 飯田 溪太
2. 発表標題 遺伝子発現調節の確率モデリングとパラメータ推定
3. 学会等名 若手数学者交流会2023（招待講演）
4. 発表年 2023年

1. 発表者名 飯田 溪太
2. 発表標題 遺伝子発現の1細胞データ解析
3. 学会等名 日本数学会（招待講演）
4. 発表年 2022年

1. 発表者名 Keita Iida
2. 発表標題 Clustering single-cell and spatial transcriptomes through multifaceted biological aspects
3. 学会等名 1st International Symposium on Aihara Moonshot Project（国際学会）
4. 発表年 2022年

1. 発表者名 飯田 亶太
2. 発表標題 遺伝子発現調節の確率モデリングとパラメータ推定
3. 学会等名 若手数学者交流会2023 (招待講演)
4. 発表年 2023年

1. 発表者名 飯田 亶太
2. 発表標題 生物データの解釈問題に対する 記号学・数学的アプローチ
3. 学会等名 第1回 合原ムーンショットプロジェクト全体会議
4. 発表年 2021年

1. 発表者名 飯田 亶太
2. 発表標題 Mathematical and semiological approach to single-cell and spatial transcriptomics
3. 学会等名 第16回生命医科学研究所ネットワーク国際シンポジウム
4. 発表年 2021年

1. 発表者名 飯田 亶太
2. 発表標題 遺伝子発現の1細胞データ解析
3. 学会等名 日本数学会
4. 発表年 2022年

1. 発表者名 飯田 溪太
2. 発表標題 一般超幾何関数を用いた遺伝子発現モデルの解析
3. 学会等名 大阪大学MDSモデリング部門主催ワークショップ「工学と数学の接点を求めて」(招待講演)
4. 発表年 2020年

1. 発表者名 飯田 溪太
2. 発表標題 記号学と数理情報科学にもとづく遺伝子発現データの多層解析
3. 学会等名 IPRセミナー
4. 発表年 2021年

〔図書〕 計3件

1. 著者名 飯田 溪太, 岡田 眞里子	4. 発行年 2022年
2. 出版社 羊土社	5. 総ページ数 9
3. 書名 空間オミクス解析スタートアップ実践ガイド	

1. 著者名 飯田 溪太	4. 発行年 2023年
2. 出版社 医学書院	5. 総ページ数 4
3. 書名 生体の科学 2023年 4月号 特集 未病の科学	

1. 著者名 飯田 深太, 粕川 雄也	4. 発行年 2023年
2. 出版社 羊土社	5. 総ページ数 15
3. 書名 改訂版RNA-Seqデータ解析 WETラボのための超鉄板レシピ	

〔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
---------	---------