様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２９　年　　６　月　１９　日現在

機関番号：　６２６０３
研究種目：　挑戦的萌芽研究
研究期間：　2015～2016
課題番号：　１５Ｋ１２１４５

研究課題名（和文）転写伸長過程の数理モデルとベイズ統計に基づく逆問題解法

研究課題名（英文）Inverse analysis of transcription elongation process using Bayesian Inference

研究代表者

吉田　亮（Yoshida, Ryo）

統計数理研究所・モデリング研究系・准教授

研究者番号：７０４０１２６３

交付決定額（研究期間全体）：（直接経費）　　2,800,000円

研究成果の概要（和文）：Total RNA-seqと呼ばれるRNAシーケンスの技術とデータ科学の解析技術を組み合わせ，観測データからRNAポリメラーゼと呼ばれる転写活性酵素がゲノム上を移動するプロセス（転写伸長速度）を再構成する問題に取り組んだ．これまで転写伸長速度の包括的測定を目的にいくつかの実験技術が開発されてきたが，実験の難しさ・精度・コストの問題があり，広く普及するに至っていない．本研究により，汎用的な実験手法であるTotal RNA-seqとデータ科学の解析手法を組み合わせることで転写伸長のプロセスを再構成できることが実証された．これにより転写伸長研究における新しい可能性が切り拓かれることが期待される．

研究成果の概要（英文）：Combining data science technologies and a widely used RNA sequencing technique called Total RNA-seq, we aimed to inversely predict a highly complex process of transcription elongation through which RNA polymerase II traverses the DNA template strand. Recently several experimental technologies, for instance, GRO-seq and NET-seq, have been developed for the genome-wide measurement of transcription elongation rates. However, such methods have not widely spread so far because of their experimental difficulties. Our study has shown that using the well-established RNA sequencing method coupled with a simple data science algorithm enables us to reconstruct the transcription elongation process by solving the inverse problem. This has opened a new possibility for transcription elongation studies.

研究分野：　統計科学

キーワード：　ベイズ統計　転写伸長　新生転写産物　Total RNA-seq　逆問題

１．研究開始当初の背景

転写制御の研究では，転写開始後の mRNA プロセッシングは，機能としてそれほど重要ではないと考えられてきた．しかしながら，近年の研究により，遺伝子発現の制御において，転写と共役して起こるスプライシング反応やヒストン修飾は重要な機能を持つと考えられるようになってきた．本研究のねらいは，近年再び脚光を浴びている転写伸長研究に対し，統計科学の観点から新たな切り口を提供することにある．現在，PolⅡや転写伸長因子のダイナミクスを捉えるために，１分子イメージング等の技術開発が推進されているが，本研究の提案手法はこれらと相補的に活用されるものである．

RNA-seq を用いた転写伸長解析は，これまでにほとんど研究がなされておらず，どの程度の精度あるいは解像度で速度分布を推定できるのか，現時点で予測できない部分が大きい．しかしながら，RNA-seq から転写速度を推定できることが実証されれば，転写伸長研究における新しい可能性が切り拓かれることになる．第一に，ゲノム全域の転写伸長速度のパターンを低いコストで入手できるようになる．さらに，速度分布と遺伝子発現量の比較，速度分布と ChIP-seq で計測されたヒストン修飾との比較等，提案手法をもとに様々な研究手法，発見が生まれることが期待される．

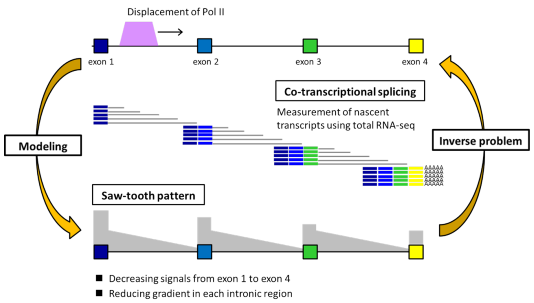

図 1: Total RNA-seq に基づく転写伸長プロセスの逆推定

２．研究の目的

RNA-seq のデータから，転写伸長の相対速度を統計的に推定できることを実証する．RNA-seq の本来の用途は，転写産物（成熟 mRNA）の発現量の計測である．しかしながら，近年の研究により，転写伸長とスプライシングは共役しながら段階的に起こることが明らかにされ，さらに RNA-seq のデータは転写伸長の途中段階にある新生転写産物の発現を捉えていることが分かってきた．このことから，データには転写伸長速度が生み出す鋸状のパターンが現れる．我々は，転写伸長プロセスの数理モデルを用いて，RNA-seq のデータのパターンから速度パラメータを推定し，ゲノム全域に渡る転写伸長の速度分布を予測する（図 1）．本研究では，提案手法の概念実証を行うと同時に，細胞種による転写伸長速度の違いや転写速度とヒストン修飾の関連性を解析する．

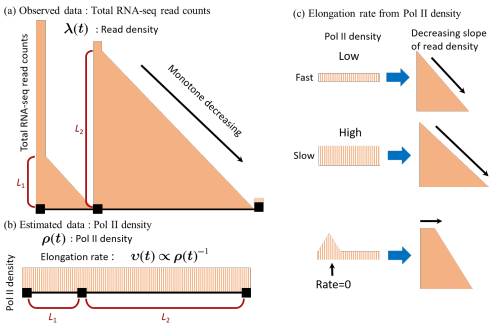

図 2: 逆問題の状態空間表現

３．研究の方法

本研究では，状態空間モデルに基づくベイズ推論による解法を示した．転写伸長とスプライシングの数理モデルを構築し，Pol II の存在確率とリード分布の変換式を用いて状態空間表現を導く．このもとで，粒子フィルタを適用してベイズ推定を行い，Pol II の存在確率とスプライスパターンを同時に推定する．

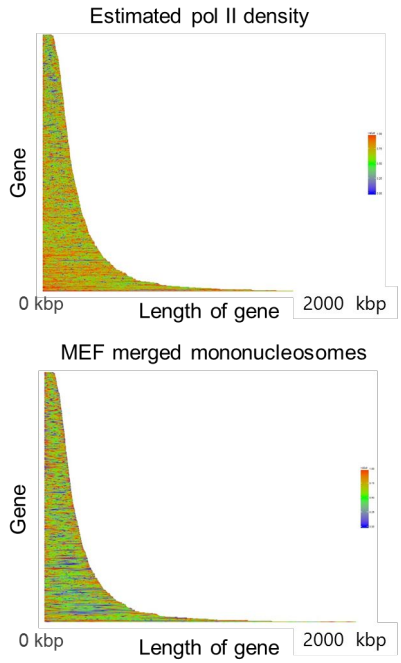

図 3: 推定された Pol-II の存在確率と ChIP-seq で観測されたヌクレオソーム占有率のパターン

４．研究成果

人工データや Total RNA-seq のデータに解析手法を適用し，Pol II の存在確率とスプライス部位の推定精度を検証した．SN（signal-to-noise）比が一定水準以上のリード分布を持つ遺伝子に対しては，現行手法は十分な推定性能を有することが確認され

た．しかしながら，Total RNA-seq の実験上の特性により，多数の短いイントロンから構成される遺伝子では SN 比及びリードカバー率が極端に低くなることが判明した．したがって，現行方法では全遺伝子規模の転写伸長速度を復元することは難しいと言わざるを得ない．しかしながら，専門家の協力を得て実験面のハードルを乗り越えることができれば，大きな科学的成果を達成する可能性もある．

マウス ES 細胞等，複数のデータに開発手法を適用し，推定された転写伸長の速度分布の妥当性を検証した．ChIP-seq 計測から導いたヌクレオソーム占有率やヒストン修飾の状態に速度分布を対応付け，両者の間に有意な相関性が確認された．現在，論文発表の準備を進めている．

これまで転写伸長速度の包括的測定を目的に様々な技術（GRO-seq，NET-seq など）が開発されてきたが，実験の難しさ・精度・コストの問題があり，広く普及するに至っていない．本研究により，Total RNA-seq という汎用技術とデータ科学を組み合わせることで転写伸長プロセスを再構成できることが実証された．これにより転写伸長研究における新しい可能性が切り拓かれることが期待される．

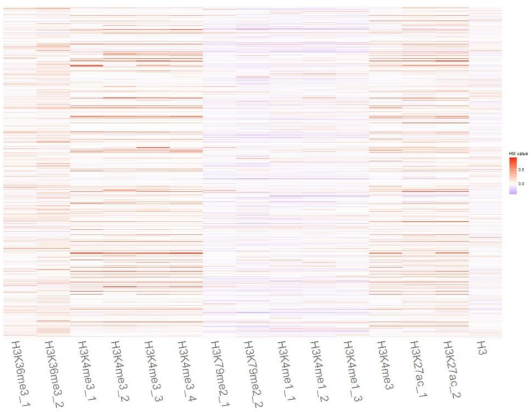

図4：推定された Pol-II 密度とヒストン修飾状態の相関

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　0　件）

〔学会発表〕（計　6　件）

1．河村優美，吉田亮，小山慎介，状態空間モデルを用いた転写伸長速度の予測，2016 年度統計関連学会連合大会，2016 年 9 月 7 日，金沢（金沢大学角間キャンパス）

2．吉田亮，生命科学におけるベイズ統計の先端応用，第 68 回日本細胞生物学会大会，2016 年 6 月 16 日，京都（京都テルサ）

3．吉田亮，統計的機械学習に基づくライフサイエンスの方法論，第 21 回 Statistical Bioinformatics Seminar （SBS），2015 年 10 月 21 日，京都（京都大学医学部付属病院）

4．吉田亮，ライフサイエンスにおけるベイズ統計の先端応用，名古屋大学大学院医学系研究科 基盤医学特論 オミクス解析学プログラム，2015 年 10 月 7 日，名古屋（名古屋大学大学院医学系研究科）

5．吉田亮，生命科学におけるデータサイエンス駆動型アプローチの開拓と実践，がんゲノムの情報と数理，2015 年 9 月 30 日，東京（東京大学医科学研究所）

6．吉田亮，ライフサイエンス分野における統計科学の先進応用，学友会セミナー，2015 年 8 月 31 日，東京（東京大学医科学研究所）

〔図書〕（計　0　件）

〔産業財産権〕

○出願状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者
吉田　亮（RYO YOSHIDA）
統計数理研究所・モデリング研究系・准教授
研究者番号：70401263

(2)研究分担者
（　　）

研究者番号：

(3)連携研究者
河岡 慎平（SHINPEI KAWAOKA）
株式会社国際電気通信基礎技術研究所・主任研究員・佐藤匠徳特別研究所
研究者番号： 70740009

小山 慎介（SHINSUKE KOYAMA）
統計数理研究所・モデリング研究系・准教授
研究者番号： 20589999

(4)研究協力者
（ ）