

## 先天性遺伝子疾患の機序解明のための時系列ゲノム解析

東京大学定量生命科学研究所  
大規模生命情報解析研究分野  
中戸 隆一郎

### 研究目的

コヒーシスはゲノム立体構造形成に重要な役割を担うタンパク複合体であり、エンハンサー・プロモーターループの形成または遮断などを通じて、遺伝子発現制御にも関与している。コヒーシスの機能欠失により引き起こされるコルネリア・デ・ランゲ症候群 (CdLS) は精神遅滞・特異顔貌・上肢の異常・心奇形など多様な表現型をもつ奇形症候群であり、コヒーシスによる転写制御機能の異常が原因であるとされるが、詳細な機序はわかっていない。一方、RNAポリメラーゼII (Po12)の伸長因子であるAFF4の変異によって引き起こされるCHOPS症候群もCdLSと極めてよく似た表現型を持ち、Po12伸長におけるコヒーシスの関与が示唆されるが、具体的なメカニズムはやはり不明である。

申請者は、遺伝子内に存在するコヒーシスのゲノム結合がRNA pol2の伸長と共に失われる現象に注目し、これがCdLSとCHOPSの類似性を説明するのではないかと考えた。そこで、エピゲノム・トランスクリプトーム・ゲノム立体構造を含む大規模なマルチオミクス解析を実施し、Po12伸長依存的に失われるコヒーシン部位では対応するループ構造も同時に失われること、これらの部位はPo12の一時停止機構と強い相関を示すこと、CdLS・CHOPS患者細胞ではこれらの結合部位において共にAFF4の結合が有意に増加していることなどを発見した [1]。これらの結果はこれまで謎であったCdLSとCHOPSの類似性の一端を説明するものであるが、コヒーシスをノックダウンした細胞でもこれらの遺伝子の発現量は有意な変動を示さなかったことから、これらの部位におけるコヒーシスの機能は依然解明されていない。本申請ではこの解析に数理モデリングを融合し、転写伸長におけるコヒーシスの機能モデルを網羅的に検証するシステムを構築し、これらの疾患の機序解明を目指した。

### 成果報告

#### 1. 高解像度時系列ChIP-seqデータのための解析手法の開発

未発表のRNA pol2の高解像度時系列ChIP-seqデータを東京大学アイソトープ総合センターの和田洋一郎教授のご厚意により供与いただいた。本データはTNF- $\alpha$ 刺激を与えたHUVEC細胞に対し、刺激後0分~30分後までを1分間隔で取得したChIP-seqデータであり、合計31タイムポイント、かつ複数回分のデータセット(複製)が利用可能である。本データはRNA Po12伸長の数理モデル化のための貴重なリソースとなる。一方、数理モデルのパラメータ推定に供するためには以下に示す2つの課題を解決する必要がある。

1) 各時点のChIP-seqデータにおいて、サンプルのS/N比(観測されるRNA pol2の強度)に技術的要因によるばらつきが認められるため、そのばらつきを補正する正規化手法が必要である。

2) TNF- $\alpha$ 刺激に対してRNA pol2伸長を開始・終了する時点は遺伝子ごとに異なるため、各遺伝子について、どの時点でRNA pol2伸長を開始・終了したかを同定する手法が必要である。

1) について、通常のChIP-seq比較解析では全マップリード数をサンプル間で等しくする全リード正規化

が一般的に用いられるが、この方法はバックグラウンド領域や反復配列領域へマップされるリードに依存する誤差が大きく、数理モデル化に必要な精度を満たさない。そこで、遺伝子発現量を元にピーク正規化を行う既存法であるCHIPIN (<https://github.com/BoevaLab/CHIPIN>) をまず検討した。CHIPINは別途与えられた遺伝子発現量データをもとに、サンプル間（本研究では異なる時点）で発現量が変動しない遺伝子セットを同定し、それらをコントロールとして、サンプル間での変動を最小とするようChIP-seqの正規化を行う。そこで、公開データベースよりTNF- $\alpha$  刺激前後のHUVEC細胞（0分及び30分時点）のトランクリプトームデータを取得し、我々のデータに適用したところ、全リード正規化よりも良好な結果となり、隣り合う時点間でより整合するRNA pol2強度分布を得ることができた。しかしながら、特にばらつきの大きいサンプルではうまく正規化できないことがあった（たとえばその時点のみPo12が全体的に減少して見えるなど）。そこで、この高解像度RNA pol2伸長データにより特化した独自の正規化手法を開発した。ここでは常染色体上に存在する全発現遺伝子（RNA pol2の濃縮で判定）を用いたうえで、各遺伝子におけるRNA pol2の濃縮レベルは広義単調増加（すなわち、隣り合う時点で大きく振動することはない）という仮定を置き、全発現遺伝子群が最もその仮定を満たすように線形回帰を用いて各サンプルの正規化係数を推定する。本手法をもとにデータ群を正規化したところ、CHIPINよりもばらつきに対して頑健にデータを正規化することが可能となった。しかしながら、全発現遺伝子について同じ正規化係数を用いることが困難であることも明らかになりつつあり、遺伝子ごとに個別に正規化係数を求める手法について現在検討している。

2) について、1) で正規化されたChIP-seqデータ群を元に、各遺伝子について、各時点での程度Po12が伸長しているのかを定量的に示す指標を開発した。ここでは各遺伝子を長さに応じて10のウィンドウに分割したうえで、各ウィンドウにおけるRNA pol2の濃縮度を時点間で比較評価することで濃縮・非濃縮の閾値を自動的に推定し、そのウィンドウで閾値を超えた時点を「RNA pol2が濃縮した時点」とする。この指標により、任意の遺伝子について、TNF- $\alpha$  刺激に対する応答性を持つか（伸長が開始するか）、またその時点を取得することが可能となった（図1に例を示す）。一方、現時点ではこの指標はまだ未完成であり、刺激前時点で転写開始点にRNA pol2の濃縮が認められるか否かや、複数のプロモーターを持つ遺伝子など、遺伝子内部に（転写伸長と別に）強いRNA pol2ピークが認められるような遺伝子を考慮できていない。特に、遺伝子内部におけるPo12ピークは研究目的で述べた遺伝子内コヒーシ結合部位を評価するうえでも重要であることから、今後この指標の改良を進める予定である。

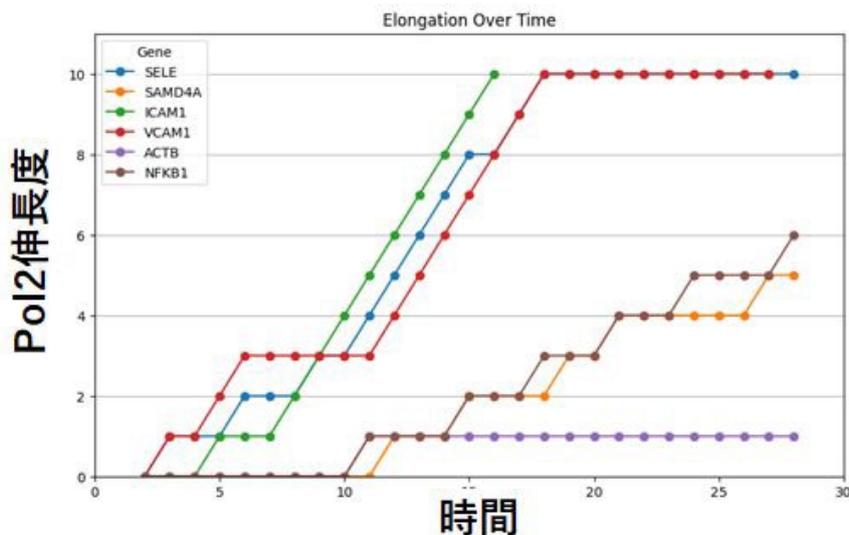


図1：考案した指標を用いた各遺伝子の伸長度合いの分布の可視化

## 2. RNA pol2伸長の数理モデルの構築

RNA pol2伸長のための数理モデルを構築した。これまでもPo12伸長のための数理モデルはいくつか提案されているが（文献[2]など）、これらのモデルは我々が目的とする遺伝子内でのタンパク質結合部位の影響を考慮していない。そこで、数理科学を専門とする北海道大学田崎創平先生のご協力により、非対称単純排除過程（asymmetric simple exclusion process; ASEP）を用いたPo12伸長のモデル化を試みた。このモデルは有限個の格子（lattice）を1次元上に接続したモデルであり、隣りあう格子間での値の輸送をモデル化する（図2）。各格子はChIP-seqデータのウィンドウ（100塩基）あたりの正規化されたマップリード数を保持する。任意の遺伝子についてRNA pol2の伸長は1方向にしか進まないことから、我々のモデルは格子の値（RNA pol2の濃縮）を順方向にのみ輸送するモデルとなる。ここで、実ChIP-seqデータに基づき特定の格子にタンパク質結合の変数を制御要因として与えると、タンパク質結合部位の格子でのみPo12伸長の速度を変動（低下）させることが可能である。NFkB1遺伝子など特定の遺伝子を対象にこの数理モデルの実験を実施したところ、遺伝子内コヒーシン結合部位に一部滞留しつつ伸長していくRNA pol2をシミュレートすることができた。現時点ではコヒーシン結合部位は単なる定性的な障害として定義されているが、コヒーシン結合そのものも時系列に沿って変動する変数としてモデルを拡張することもでき、最終的にはそのようなモデルによってPo12伸長とコヒーシン結合の連動性について評価する予定である。現状の課題として、項目1でも述べたように、刺激前時点で転写開始点にRNA pol2の濃縮が認められるか否かや、複数のプロモーターを持つ遺伝子など複数箇所から伸長が開始するような場合を考慮できていない。この点に関して、今後も改良を進めていく。

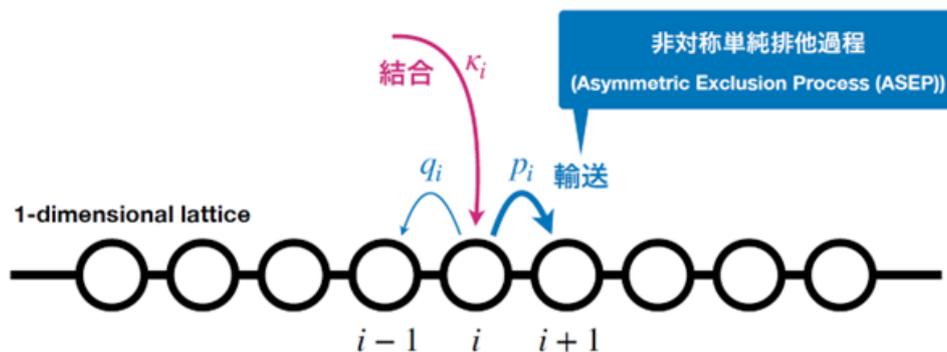


図2：非対称単純排除過程の模式図

[1] Wang J, Bando M, Shirahige K, Nakato R: Large-scale multi-omics analysis suggests specific roles for intragenic cohesin in transcriptional regulation, *Nature Communications*, vol 13, issue 1, 2022.

[2] Zhao et al, Model-based characterization of the equilibrium dynamics of transcription initiation and promoter-proximal pausing in human cells, *Nucleic Acids Research*, vol. 51, e106, 2023.