

## 論文の内容の要旨

論文題目 : Spatial-temporal Modeling and Simulation of Transcription

(転写の時空間モデリングとシミュレーション)

氏名 : 大田 佳宏

遺伝子の DNA 配列を鋳型に、RNA ポリメラーゼ II (RNAPII) という酵素が遺伝子上を 5' 側から 3' 側の方向へ転写を行いながら移動することで RNA が作られ、その RNA の配列を元に蛋白質が作られることは、生命の基本原則と考えられている。近年、ゲノム解読からヒト染色体のもつ DNA の配列が明らかになり、RNA が生成されていく全体像を解読することが可能となってきた。しかし、この DNA から RNA が生成される過程を、時間と空間において高分解能で精密に解析するのは困難だった。

本論文では、ヒトの血管の細胞が炎症刺激を受けた後、7.5 分おきに RNA が作られていく様子を染色体上で観測し、RNAPII が 3100 塩基 (約 1 ミクロン) /分の速度で動き、さらに CTCF/Cohesin において RNAPII の速度が遅くなることを、時空間上において高い分解能で解明した。

具体的には、空間分解能の高いカスタムタイリングアレイを独自に設計し、細胞刺激後に RNAPII が作る RNA を 7.5 分という短い時間間隔で計測し、その結果得られる大量データの解析を行い、遺伝子上の転写産物の時間変化を特定した。ここでは、遺伝子上のプロベ特性を考慮した  $d(\text{probe}(Gx))/dt$  を用いて転写速度を導出し、イントロンについては通常 Premature termination ( $Ck(ik-xk/v)$ ) から逸脱した箇所を網羅的に解析した結果、転写開始点 (TSS) 近傍に特異点を発見した。

従来ポリメラーゼは確率的に動いたり止まったりすると考えられていたが、今回 ChIP-Seq Intensity 評価法に基づき、RNAPII の密度分布の強度  $S(x,t) > \zeta$  を満たす全ての  $(x,t)$  について  $\sum S(x,t)/N$  を計算することで、CTCF/Cohesin の存在密度が高い領域では RNAPII の速度が遅くなり、逆に存在密度が低い領域では速度が速くなるという転写ダイナミクスの制御を明らかにすることができた。通常 DNA は CTCF/Cohesin という蛋白質が作用することで束ねられる。RNAPII は活性化される前からこの CTCF/Cohesin で区切られた染色体上の特定の狭領域に集まっているが、活性化されるとそこから動きだし、先の部分では CTCF/Cohesin の存在密度に応じて速度変化しながら転写していくというダイナミカルな描像を得ることができた。

次に、本論文ではこれら実験データの解析結果に基づいて、RNAPII が遺伝子を転写していく運動をモデル化し、シミュレーションを行うことで転写モデルの実証を行った。具体的には、(1) RNAPII 実体の運動モデルとしてセルオートマトン (CA) を用いて時間発展のシミュレーションを行い、さらに (2) RNAPII が生成した RNA の時空間上での発現量シミュレーションを行うことで、ヒト遺伝子群での RNAPII の挙動から、その産物である RNA を生成するモデリングまで行った。

(1) においては、1 遺伝子の構造を  $n$  個のセルユニットに分割し、その各セルユニットにおける RNAPII の存在量を規定するベクトル、 $\mathbf{r}(t)=(r_0(t), r_1(t), \dots, r_j(t), \dots, r_n(t))$  を導入する。ここで、CA は Rule184 に準拠すると仮定した。この時、 $r_j(t+\tau) = F[r_{j-1}(t)-r_j(t)]$  を満たし、 $j$  番目の RNAPII は 1 つ前の RNAPII から関数  $F$  の影響を受けるものとした。さらに、CTCF/Cohesin 存在領域やエクソン領域のような速度変化を引き起こす領域のゲノム領域に  $\gamma(\mathbf{x},t)$  を導入し、このパラメータは時空間的に変化するものとした。

ChIP-Seq における強度をモデル化するため、TNF $\alpha$  刺激後の細胞の確率密度分布  $P_{cell}(t)$  を導入すると、複数 RNAPII による密度変化  $Q_j(t)$  は  $\Sigma(r_j(t) * P_{cell}(t))$  で表現される。本モデルによるシミュレーション結果は、ヒトの細胞を用いた ChIP-Chip による実験結果を極めてよく再現しており、最適なパラメータ集合を導出することができた。

(2) の RNA の時空間的な発現量解析においては、上記 (1) で得られた RNAPII ダイナミクスの座標情報を引数にとり、各遺伝子について発現されるエクソン・イントロン領域の RNA 量を時空間上でシミュレーションを行った。ここで、時間・空間の分解能は、生物学実験では扱うことの不可能な 1 (秒)・1 (bp) レベルの高密度・高精度な計算を行った。

単一 RNAPII による RNA 減衰モデルは、コピー数を  $\rho$ 、減衰定数を  $\tau$ 、減衰開始時刻を  $T_d$  として、 $\rho * \exp(-(t-T_d)/\tau)$  を適用した。ここに新たに、TSS 蓄積係数  $\lambda$  を用いた Premature termination モデル  $\lambda * \exp(-((x-\xi)/\epsilon + (t-\delta)/\tau))$  を独自に導入し、単一 RNAPII による RNA 算出シミュレーションを行った。さらに、同一遺伝子上には複数の RNAPII が相互作用しながら転写を行っているとは仮定し、上記 (1) のベクトル  $\mathbf{r}(t)$  と結合した協調作用のモデルを導入した。

従来は、衝突・停留など RNAPII の局所的な相関モデルについての実験および数理モデルの研究は行われていたが、今回新たに長距離・非局所的な協調作用が存在することを明らかにした。長距離・非局所的な協調作用を考慮できるモデルとして、複数 RNAPII における  $r_j(t)$  と  $r_{j+1}(t)$  の転写開始時間間隔  $T_{ir}(r_j)$  を用いた  $\Sigma E(\mathbf{x},t,\mathbf{r})$  を導入し、遺伝子転写モデリングを行った。本モデリングによるシミュレーションの結果、イントロン部分の RNA は時

間が経つと消滅し、エクソン部分の RNA は時間とともに蓄積されているのが確認された。また、刺激は常時入ったままであるが、刺激直後から RNA の転写が遺伝子上を波のように伝搬していき、安定的に転写する状態になっていく。この結果は、タイリングアレイにおける RNA 発現量実験の結果を極めてよく再現している。

本モデルによるシミュレーションを、実験結果を再現するためのパラメータバリデーションによって2万回計算することで、複数 RNAPII における  $r_j(t)$  と  $r_{j+1}(t)$  の間隔  $T_{ir}(r_j)$  が、Logistic 関数  $d/dr(T_{ir}(r)) = 0.5 * T_{ir}(r) * (1 - T_{ir}(r)/16)$  上にあることも導出した。現在この関数とゲノム構造との関連を解析しつつある。複数 RNAPII が遺伝子上でこのような関係にあることは、「複数の RNAPII が相互に協調的挙動をとることで、RNA 産出のための転写を効率的に行っている」ということを示唆している。