

論文内容の要旨

論文題目: マウス・トランスクリプトームにおける
アンチセンス転写と mRNA-like noncoding RNA の関係

氏名: 片山 慎太郎

ヒトゲノムの解読プロジェクトは 1991 年から始まり、2000 年 6 月にドラフト配列を発表した後、2003 年 4 月には全作業を終了した。国際ヒトゲノムコンソーシアムの解析によれば、生命を形作るために必要な約 22,000 個の遺伝子に含まれるタンパク質の情報は、ゲノム中僅か 1.5%程度の領域を占めるに過ぎなかった。しかし、この 1.5%およびその周辺にある僅かな遺伝子発現制御領域だけで高度に複雑化されたヒトの生命活動を実現しているとは、にわかには信じがたい話であった。

一方、ゲノムからの転写産物の総体であるトランスクリプトームに目を移すと、1.5%のゲノム領域を遥かに超える多種多様なトランスクリプトの存在が示唆されており、生物が非常に非効率的であると思える程であった。例えば Kampa らは、ヒト 21, 22 番染色体のゲノムタイリングチップを用いて 11 の細胞株のポリ(A)テールを持つトランスクリプトームを解析したところ、トランスクリプトが検出された領域の半分は、公共データベースに登録されている messenger-RNA (mRNA) のいずれとも一致しない事を示した。また我々も完全長 complementary DNA (cDNA) ライブラリ の技術を用いて、約 300 のマウス由来組

織および細胞株のトランスクリプトームを解析し、ゲノムの 62%が転写されていることを明らかにした。

そしてゲノムおよびトランスクリプトーム研究の進展により、2 つの新しいトランスクリプトに関心が寄せられるようになった。1 つは mRNA-like noncoding RNA (ncRNA)であり、これは mRNA 同様ポリメラーゼ II で転写されキャップ構造とポリ(A)テールを持つものの、有効な open reading frame (ORF) が存在しないトランスクリプトである。沼田らは、当時発表されていたマウスのトランスクリプトームから ncRNA として信頼度の高い 4,280 個に絞り、それらが CpG 含有量の高い領域から転写されている傾向がある事を示した。Wang らは、これら ncRNA の配列がその他ゲノム中の遺伝子間領域と同程度の進化速度である事を示し、一方我々は最新のマウス・トランスクリプトームに含まれる ncRNA の transcription start site (TSS)について解析を行い、TSS 近傍の保存性が高い事を明らかにした。また今西らは当時発表されていたヒトのトランスクリプトームから 1,377 の ncRNA を特定し、また前述の Kampa らは、得られたヒト新規トランスクリプトのうち 75bp 以上の ORF はせいぜい 24%程度である事を示した。

ゲノムおよびトランスクリプトーム研究の進展により関心が寄せられたもう 1 つのトランスクリプトは、アンチセンス RNA である。あるトランスクリプトをコードする DNA 鎖(センス鎖側)の逆鎖から転写され、センス鎖側のトランスクリプトと部分的もしくは全体的に 2 本鎖 RNA を作る可能性がある。清澤らは、当時発表されていたマウスのトランスクリプトームから、sense-antisense (S/AS) のペアを 2,481 個特定し、imprinting を受ける遺伝子と S/AS ペアに関連がある事を示した。Chen らは、当時発表されていたヒトのトランスクリプトームから、約 20%の遺伝子にアンチセンス RNA が存在している事を明らかにした。また Cheng らは、ヒトの 10 本の染色体に対して設計したゲノムタイリングチップを用いて特定した新規トランスクリプトから 768 個をランダムに選びさらに実験的検証を行った結果、61%に両鎖のトランスクリプトが存在している事を明らかにした。

ゲノムワイドなトランスクリプトーム解析が行われるまでは、ncRNA とアンチセンス RNA いずれも数十個程度しか発見・機能解析がなされておらず、遺伝子発現制御に関わる要素ではあったが、その数からごく特別な事例であると考えられていた。またゲノムワイドなトランスクリプトーム解析の後大量の

ncRNA およびアンチセンス RNA が特定されてからも、その大部分は機能未知のままであった。しかしゲノムの 1.5%を遥かに超える領域にコードされ、そして実際に転写されている ncRNA およびアンチセンス RNA は、果たして生命活動に無関係なジャンクなのであろうか?

本研究では、前述した新しいトランスクリプト、ncRNA とアンチセンス RNA について解析を行う。まずアンチセンス RNA についてであるが、アンチセンス RNA が多く存在している事はこれまでの解析でも示唆されているが、これを最新のマウス・トランスクリプトームを元に再度確認する。特に、マイクロアレイベースだけでなく、cDNA のシーケンスベースでも検証を行う事は、得られた事象の信頼度を高める点において価値がある。またこれまでのアンチセンス RNA の解析が、様々なサンプルから得られたトランスクリプトを同列に並べたものであったり、ごく少数の細胞・組織サンプルだけを用いた物であったりしたので、同一組織や細胞群の中での一般性が不明瞭である。そこで 90 サンプルから得られた cap analysis gene expression (CAGE) 法による転写開始点レベルでの発現プロファイリングデータを用いて、アンチセンス転写の共発現性を考察する。

次に ncRNA についてであるが、ゲノムタイリングチップの解析によると新規トランスクリプトの多くが ncRNA であり、また多くの新規トランスクリプトで両鎖からの転写が重なるように発現していた事から、cDNA によるシーケンスベースでも ncRNA とアンチセンス転写の関係性について考察する。

またその関係性に基づき、アンチセンス RNA と ncRNA の発現動態を把握して機能解析につなげる為に必要な、新しい発現プロファイリングツールを提案する。

そしてこれらの解析の結果、まず ncRNA とアンチセンス転写の関係性については、マウスゲノムには遺伝子数とほぼ同数の noncoding TU が存在し、それら noncoding TU はゲノム DNA の混入による人工産物ではなく、その 23% (端読み配列を含めると 59%) が S/AS ペアを形成して、また coding TU と divergent もしくは full-overlap する傾向にある。そして divergent S/AS ペアは共発現する傾向にある。また S/AS ペアはパートナーの発現量を活性化もしくは抑制している事が分かった。そして発現プロファイリングツールとしては、mRNA のキャップ構造からラベルしたトランスクリプトを、TSS 直下流にデザインしたプローブへ hybridization させる CAGE-TSSchip というマイクロアレイを開発した。プロト

コール自体は、dye-swap 間で良い相関を示す程度に完成しており、優位差を示す TSS については qRT-PCR および CAGE と一貫した結果を示し、また必要感度に応じて増える CAGE のシーケンスコストに比べ良いコストパフォーマンスを持ち、さらには活性に大きな差のある TSS 近傍のプロンプが有意差を示す事を示した。また CAGE-TSSchip を用いてマウスの肝臓と肝臓癌由来細胞株 Hepa1-6 で活性に有意差のある TSS を調べた所、細胞周期や癌マーカーとして知られている遺伝子の 5'UTR に存在する TSS がスクリーニングされ、また Hepa1-6 で有意に活性の高い TSS の近傍には E2F など癌関連転写因子の結合サイトが頻出している事が分かり、CAGE-TSSchip により妥当な発現制御エレメントをスクリーニングすることができた。さらに CAGE-TSSchip を用いてマウスの肝臓と小脳で発現量が有意に変動している divergent S/AS ペアをスクリーニングした所、Rsh12a の long isoform TSS とそのアンチセンスが逆相関するように発現変動している事が分かり、divergent S/AS ペアの発現プロファイリング用途としての可能性も示した。

本研究の1つの成果は、翻訳されてタンパク質を作る事ではなく、転写される事が重要なトランスクリプトが、S/AS ペアという形でゲノム上に大量にコードされている事を示した点である。トランスクリプトームプロジェクトの進展により既に多くの機能未知 ncRNA の存在が指摘され、またゲノムプロジェクトの進展もあって既に多くの S/AS ペアの存在が指摘されていたが、実はそれらは密接にリンクし、細胞群や組織内で ncRNA を含む S/AS ペアが共発現して、静的に見えた細胞の中で互いを制御しあっていることを指摘した点が新しい。つまり細胞の状態は転写因子やクロマチンの構造だけではなく、転写される事が重要なトランスクリプトによるアンチセンス転写においても規定され得る事を示唆している。転写される事が重要なトランスクリプトとして ncRNA を想定する事は、多くの ncRNA が配列レベルでは保存性が低く、しかし転写開始点レベルでは保存性が高かった事実と矛盾しない。

本研究の2つ目の成果は、TSS レベルでアンチセンス転写や ncRNA の発現量をコストパフォーマンス良く測定する技術として CAGE-TSSchip を開発した事である。いずれ来るであろう次世代シーケンス技術と完全に競合する技術であるが、今後、これら新規トランスクリプトを含めた遺伝子発現制御ネットワークの解明に資するものと期待される。