

比較ゲノム解析および遺伝子発現プロファイルに基づいたプロモータ情報構造の解析

●野口 英樹

独立行政法人理化学研究所ゲノム科学総合研究センター

〈研究の目的と進め方〉

生命現象を体系的に理解する上で、プロモータの情報構造の解明は構成部品である遺伝子の同定に並び重要な課題である。しかしながら、プロモータやエンハンサーなどの制御領域にはタンパクのコード領域に見られるような強い規則性は観察されず、また実験データも不足していることから、これまでのところ有効な解析手法は確立されていない。本研究では、脊椎動物の比較ゲノム解析から得られる機能領域の情報とDNAマイクロアレイ等を用いた遺伝子発現プロファイルのデータから、組織・時期特異的な遺伝子発現制御に関わるゲノム情報構造を解明し、高精度なプロモータ予測モデルを構築することを目的とする。

〈研究開始時の研究計画〉

まず、ヒトとマウス遺伝子の発現プロファイルを、DNAマイクロアレイを用いて解析する。必要に応じて公共データベースの解析データも利用することで十分なデータを確保する。本年度の研究では発現量に関しては考慮せず、データはユビキタスに発現しているものや特定の組織でのみ有意に発現している（または発現していない）ものなどに分類し、以後の情報解析を行う。

プロモータの比較解析には、完全長cDNAやESTの情報から同定されたプロモータの配列を用いる。比較ゲノムを援用した機能領域の予測では、種間で高度に保存された領域に注目することで予測の擬陽性率を低く抑えることが出来るが、さらに保存パターン（塩基置換やindelの入り方）に注目することでタンパクのコード領域や制御領域を分類することが可能である。本研究では、コアプロモータ領域の種間での保存性について詳しく検討し、保存パターンなどの統計的特長からプロモータ予測を試みる。

〈研究期間の成果〉

プロモータ構造と発現の組織特異性について調べるため、まずはヒトプロモータを、CpGアイランドを持つものと持たないもの（Non-CpG）に分類して、発現プロファイルとの関係を調べた。マイクロアレイのデータを用い、発現の組織特異性をShannonの情報エントロピーとして表現して解析したところ、組織特異性が強い遺伝子ほどNon-CpGプロモータにより発現制御されることが明らかになった。

次に、コアプロモータのヒト・マウス間での保存度を調べたところ、9割以上のプロモータが70%以上の相同性を持つことが分かった。このことは、大部分のプロモータはコード領域と同等かそれ以上（塩基レベルで）に高度に保存されていることを示している。さらに、転写開始点を基準に位置特異的な保存度を解析したところ、Non-CpGプロモータでは転写開始点以上にTATAボックス周辺の保存度がより高く、さらに転写開始点から上流-200bpまでの領域でCpGプロモータに比べて有意に保存されていることが明らかとなった。組織特異的な発現制

御にはこれらの高度に保存された領域が重要な役割を担っているものと考えられる。

次に、プロモータと他の領域における塩基置換の種類（転位と転換）の頻度を調べた。ゲノム全体では転換より転位の方が多く、ヒト・マウスでは転位／転換の比はゲノム全体で1.8であった。コード領域では同義置換のほとんどが転位であることからこの傾向はさらに強いが、逆にプロモータ領域では比は有意に小さく、コード領域とは異なるバイアスがかかっていることが示唆された。これまでに観察されたNon-CpGプロモータの特徴を指標にプロモータ予測を行ったところ、TATAボックスとCAPサイトの重み行列に基づいた予測手法と比較して大幅な精度の改善が見られた。

〈国内外での成果の位置づけ〉

多くの研究者が様々なヒトプロモータ予測のアルゴリズムを開発しているが、いずれの手法もCpGアイランドを持つプロモータの特徴を捉えているに過ぎず、Non-CpGプロモータに対する感度はほとんど無いに等しい。一方で、本研究により発現の組織特異性を定量的に評価した結果、特異性の高い遺伝子はNon-CpGプロモータにより発現が制御されていることが分かった。このことから、発現制御機構を知る上でNon-CpGプロモータは重要であり、Non-CpGプロモータを高感度・特異性で予測する本システムの有用性は高いといえる。

〈達成できなかったこと、予想外の困難、その理由〉

Non-CpGプロモータでは転写開始点のすぐ上流の領域がCpGプロモータと比べて有意に保存されており、この領域に発現に重要なエレメントが存在していると考えられる。しかしながら、本研究期間中には、これらの領域に共通に存在するエレメントを特定することは出来なかった。Non-CpGプロモータは組織特異的な発現に関わることから、プロモータごとに異なるエレメントが存在している可能性も否定できない。

〈今後の課題〉

Non-CpGプロモータの制御に関わるエレメントを同定し、発現プロファイルとの関係モデリングを行うためには、複数種ゲノムの比較を通して少ない（同様の制御を受ける）配列から、高感度かつ特異的にエレメントを抽出する手法が必要だと思われる。今後は、そのための新しいマルチプルアライメント手法の開発を行っていきたいと考えている。

〈研究期間の全成果公表リスト〉

- 2) 比較ゲノムに基づいた遺伝子予測プログラム PHINAL <http://phinal.gsc.riken.jp/>