

2006年4月29日
独立行政法人 理化学研究所

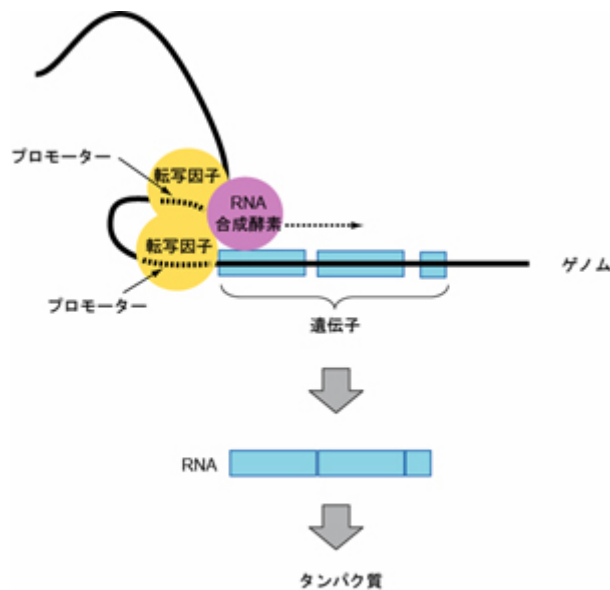
新しい遺伝子調節メカニズム「ブロード型」の調節領域を発見

- 高等動物の遺伝子の「スイッチ」、多くは進化速度が速いニュータイプ -

生命の設計図であるゲノム。ヒトの場合、このゲノムには約30億個の塩基対がつくるDNAという物質が遺伝暗号として並んでいます。生体内では、DNAの並びからRNAという物質の並びが作りだされ、必要な部分が選択され、最終的にはタンパク質がつくられます。このタンパク質がつくられる部分のDNAの並びのことを遺伝子と呼んでいたのですが、最近、この仕組みについて、多くのことがわかってきて、これまで常識と思われていた概念を考え直さなければならない状況になってきました。「遺伝子」という言葉の定義すら考えなおさなければならないようになってきています。

文部科学省ゲノムネットワークプロジェクトと理研が主催する国際FANTOMコンソーシアムは、RNA生成の量やタイミングを調節するスイッチが、ヒトとマウスにおいて、これまでわかっていたよりも5～10倍もあることを発見しました。さらにこれが、「ブロード型」と名付けた新しいメカニズムで働いていること、進化速度が早いものであったことなどを突きとめました。

これらの発見は、単純に考えられてきた遺伝子にまつわる仕組みが、実はとても複雑な仕組みになっているという証拠をさらに提示したことになります。



(図) ゲノムからタンパク質がつけられる

2006年4月29日
独立行政法人 理化学研究所

新しい遺伝子調節メカニズム「ブロード型」の調節領域を発見

- 高等動物の遺伝子の「スイッチ」、多くは進化速度が速いニュータイプ -

◇ポイント◇

- ・ヒトとマウスで従来の5～10倍の遺伝子調節領域を見つける
- ・ゲノム上のさまざまな場所からRNA合成:「ブロード型遺伝子調節領域」が大半を占める
- ・ブロード型は進化が速く、ヒトなどの高等動物の複雑さをもたらす原動力だと示唆

独立行政法人理化学研究所(野依良治理事長)を中核機関におく文部科学省ゲノムネットワークプロジェクト(笹月健彦推進委員会委員長、榊佳之実施会議議長、林崎良英中核機関代表)と、理研が主催する「国際FANTOMコンソーシアム^{*1}」(統括:林崎良英、科学コーディネーター:カルニンチ・ピエロ)は、遺伝子のオン・オフ(RNA生成のタイミングと量)を決めるスイッチの役割をする遺伝子調節領域(プロモーター^{*2})を、ヒトとマウスにおいて、従来の約5～10倍の数に相当する数(マウス:236,498個、ヒト:190,513個)同定しました。そして、その詳細解析から、新しいタイプの調整メカニズムを発見し、これを「ブロード型」と命名しました。

従来、教科書には「一般的なプロモーターによるRNAの生成は、ゲノム上の一点(遺伝子の最上流)から開始される」と記されてきました。しかし今回の研究で命名した「ブロード型プロモーター」は、高等生物においてプロモーターの主要な型であり、RNA生成の開始点がバリエーションに富み、ゲノム上のある一定の広い範囲内でRNA生成が始まるという特徴を持っていました。また、高等動物においていろいろな組織で発現している遺伝子のRNA生成開始点はブロード型である傾向が強く、逆に特定の組織で特徴的に発現している遺伝子の開始点は厳密に一点に規定される傾向があることがわかりました。

さらに、ヒト・マウス・ラット・イヌ・チンパンジーのプロモーターの比較から、霊長類の進化の過程において、ブロード型のプロモーターの進化が速い(突然変異の発生率が高い)ことがわかり、ブロード型プロモーターとそれによって調節される遺伝子が、高等動物の遺伝子の機能や進化、ひいては生物の複雑さを考える上で重要であることが示唆されました。

また、RNA生成の開始点は、これまで遺伝子の第1エクソン^{*3}の先頭に限定されると考えられていましたが、第2エクソンなどの他のエクソンにおいてもしばしばあることを見つけ、特に最終エクソンからRNA生成が開始されることが多いことがわかりました。

一方、生物が進化していく過程で化石化し、機能していないと考えられていた「偽遺伝子^{*4}」の9,583個が、実際はRNAを生成していることなども突き止めました。

これらのプロモーターやRNAのデータベース(<http://gerg01.gsc.riken.jp/cage/>)は、バイオ医学研究分野に主要な研究材料を提供することになります。

また同時に、関連する基礎情報は、国立遺伝学研究所の日本DNAデータベース(DNA Data Bank of Japan :DDBJ, <http://www.ddbj.nig.ac.jp/>)のデータベース上

で公開されます。

この成果は、米国の科学雑誌「Nature Genetics」(4月28日号)と「PLoS Genetics」ゲノムネットワーク・FANTOM特集号(4月28日号)に発表されます。「PLoS Genetics」特集号に掲載される12報の論文は、異なる視点からトランスクリプトーム^{*5}を解析し、その結果、予想を上回る複雑さを明らかにしています。

1. 背景

近年、哺乳動物のDNA遺伝暗号がつぎつぎと解読されていますが、ゲノムの塩基配列の中にいったい何が書かれているのかという情報の解読は不十分で、まだまだ未知の領域です。2005年、米国の科学雑誌『Science』のRNA特集号(9月2日号)に、文部科学省ゲノムネットワークプロジェクトと理研を中心とする国際FANTOMコンソーシアムにより報告した2報の成果は、「RNA新大陸^{*6}の発見」として大きな反響を与えました(プレスリリース:哺乳動物のトランスクリプトームの総合的解析による「RNA新大陸」の発見)。それは、細胞が生産するRNAに関し、今までにない大規模なスケールで調べたところ、従来100個ぐらいいしか知られていなかった「非タンパクコードRNA(Non-coding RNA; ncRNA^{*7})」が、実は23,000個以上、つまり、全遺伝子の半分以上(53%)を占めているという新しい事実を判明したことでした。同時に、二重鎖DNAの双方の鎖が読まれるセンス/アンチセンス^{*8}のRNAペア^{*9}が多数存在することも発見しました。このことは、タンパク質がゲノムにコードされている最終生理活性物質であるというこれまでの常識を覆し、予想を凌ぐトランスクリプトームの複雑さを認識させるもので、哺乳動物ゲノムの情報内容に対するこれまでの理解(「遺伝子」という領域が散在しているゲノムのイメージ)を根幹から変えてしまうものでした。

2. 研究手法と成果

今回の研究では、理研が独自に開発した種々の技術を用いて、ヒトおよびマウスのトランスクリプトーム解析を行い、これまでに知られていたプロモーターの10倍に及ぶ数のプロモーターを同定しました。すなわち、マウスで236,498個、ヒトで190,513個(ヒトで報告のあるプロモーターデータベースでは30,964個、本年4月10日)のプロモーターを同定するとともに、それらの解析を行い遺伝子発現に関する数多くの興味深い知見を得ました。

研究グループは、主にCAGE法^{*10}を用いて得られた約1400万個のRNA5'末端^{*11}の塩基配列を決定し、これをゲノム上にマッピングすることにより、マウスで約236,498個、ヒトで190,513個のプロモーターを同定しました。

これまでゲノムDNAからRNA生成が始まる点(転写^{*12}開始点)はゲノム上の一点に厳密に決まっている(ナロー型)と考えられていましたが、今回の研究で、実際には、高等動物においては、多くの場合、転写開始点は曖昧で、ある程度の長さの領域のさまざまな場所から始まっていること(ブロード型)を突き止めました。また、このような転写開始点の分布の形とプロモーターの構造について解析した結果、ナロー型にはTATA配列^{*13}をもったプロモーターが多く、ブロード型ではCpG配列^{*13}が豊富なプロモーターが大部分を占めていることがわかりました。さらに、

多くの組織で普遍的に発現している遺伝子の多くはブロード型プロモーターを持ち、他方、組織特異的、あるいは時期特異的に発現する遺伝子はナロー型プロモーターを持つ傾向があることも突き止めました。

さらに、プロモーター構造について、ヒト、マウス、ラット、イヌ、チンパンジーの比較解析を行った結果、全ゲノムのうち、ブロード型のプロモーターの突然変異発生率が、ナロー型よりも高いことがわかりました。つまり、霊長類に近い種から霊長類への進化の過程において、プロモーターの進化速度が速い、ということが明らかとなったこととなります。この結果は、高等動物の遺伝子の機能と進化を考える際には、遺伝子産物そのものの構造的な違いのみならず、プロモーターの機能が生物の複雑さを解明する上で重要であることを示唆する発見であると考えています。

また、従来、転写開始点は主に遺伝子の最前部（第1エクソンの上流）にあると考えられていましたが、今回の大規模な解析により、遺伝子の中ほど（第2エクソンやその下流）や後部（最終エクソン）にも数多くの転写開始点が存在することを見つめました。これらの転写開始点から生成されたRNAは、異なる種類のタンパク質を生成したり、あるいはタンパク質をコードしないRNA（ncRNA）として、発現制御などの生理活性を担っていると予測しています。

PLoS Geneticsに掲載される12報の論文には、プロモーターだけでなくRNA大陸の詳細な解析によるいくつもの興味深い新たな発見を報告しています。具体的には、「マクロRNA」と名づけた非常に長いRNAを作る一群の新しい種類の遺伝子（66個）の発見があげられます。従来、マクロRNAは、少数で特殊なRNAであると考えられてきました。しかし今回の研究で、遺伝子の発現を調節する機能を持つマクロRNAも見つかっており、ゲノム上のひとかたまりの領域（メガベース単位）に含まれる複数の遺伝子のオン・オフをまとめてコントロールする「領域スイッチ」として働いていると予想しています。このことは、プロモーターが単一の遺伝子のオン・オフをコントロールする働きと対照的です。

また、遺伝子の発現をコントロールする仕組みとして“chains（鎖）”と名づけた新しい概念を提案しました。“chains（鎖）”は遺伝子が密集している領域であり、その中では二つの遺伝子の発現が単一のプロモーター（双方向性プロモーター）で制御されていたり、センス/アンチセンスRNAによる遺伝子発現制御が行われていたりなど、一連の遺伝子の転写が領域全体の活動として制御されているという非常に複雑な構造を有するものです。今回の研究において、1,000以上の“chains（鎖）”を見つめました。また、通常は進化上の古い遺伝子が化石化したものと考えられていた偽遺伝子が、実際にはRNAを生成している例を9,583個発見し、その中にはタンパク質をコードしているRNAもありました。これらのタンパク質は未知の機能を持っていることが推測され、またタンパク質をコードしていないncRNAも遺伝子発現制御などの機能があると考えています。

同時にRNAから翻訳されるタンパク質として新たに約1,100個の非常に短いタンパク質（100アミノ酸以下）の存在を明らかにしました。これらは、非常に短いため、タンパク質をコードしないRNAであると考えられていましたが、今回計算機による詳しい解析によりタンパク質として発現していることが予測され、実験により生体内で機能しうることを示しています。

3. 今後の展開

ヒト・マウスをはじめとする哺乳類ゲノムに書き記されている遺伝子には、正しい組織・細胞で正しいタイミングで作動する遺伝子調節領域（プロモーター）が備わっており、今回それらが予想を超える数を発見しました。一つの遺伝子には複数のプロモーターがあり、さまざまな目的で生体は使い分けています。さらに、プロモーターとして、従来の教科書に載っている典型的構造をしたナロー型のプロモーターはあまり多くなく、主としてブロード型のプロモーターが多いことがわかりました。これらのプロモーターは、あらゆる生命現象を分子レベルで説明するために必須の研究対象となります。たとえば、疾患の原因遺伝子そのものの異常により、疾患が発症することもあります。疾患の原因遺伝子のプロモーター異常による多数の例も知られています。今回、解析構築したプロモーターのデータベースは、研究者が解析するゲノム配列のどの領域がプロモーターであるかという情報を提供することから、バイオ医学研究分野に不可欠の研究基盤材料になると期待されます。

(問い合わせ先)

独立行政法人理化学研究所 横浜研究所
ゲノム科学総合研究センター
遺伝子構造・機能研究グループ

プロジェクトディレクター 林崎 良英
Tel : 045-503-9222 / Fax : 045-503-9216

横浜研究推進部

溝部 鈴

Tel : 045-503-9117 / Fax : 045-503-9113

(報道担当)

独立行政法人理化学研究所 広報室

Tel : 048-467-9272 / Fax : 048-462-4715

Mail : koho@riken.jp

<補足説明>

※1 FANTOM

理研が中心となって結成された哺乳動物（マウス）の遺伝子を網羅的に機能注釈することを主眼とする国際的研究コンソーシアム共同集団（Functional ANnotation Of Mouse cDNA）の略称。オーストラリア、シンガポール、スウェーデン、南アフリカ、イタリア、ドイツ、ギリシャ、スイス、英国、米国などを含む全世界の11ヶ国／45ヶ所の研究機関等が参加している。

※2 プロモーター

転写開始を促す活性を持つDNA上の特定の領域・塩基配列のこと。

※3 エクソン

最終的にメッセンジャーRNAとして、うつしとられて残る（コードされる）DNAの配列部分（領域）。従来、教科書では、遺伝子とはタンパク質をつくる（コードする）もので、タンパク質をつくるための情報を持つ遺伝子領域（DNA配列）をエクソンと説明していた。しかし、最近になって、タンパク質をコードしない mRNA がたくさん見つかってきており、遺伝子という言葉ですら定義の変換の必要に迫られている。

※4 偽遺伝子

DNAの配列のうち、かつては遺伝子産物（特にタンパク質）をコードしていたと思われるが、現在はその機能を失っているものをいう。活性な遺伝子と配列が類似しているが、転写されなかったり、終始コドンや変異を持つなどして不活性化している。

※5 トランスクリプトーム

RNA合成酵素によってゲノム情報から写し取られた転写物集団。狭義な旧来のセントラルドグマの定義では、mRNAを主要なものとして考え、それ以外をジャンク（不要物）としていた。

※6 RNA 新大陸

多様な細胞内RNA集団の莫大な可能性を示す比喩的表現。2005年9月の記者発表で、遺伝子の定義を新たに提案した際に用いた新用語。

※7 ncRNA

非タンパクコードRNA（Non-coding RNA）のことで、このRNAからはタンパク質は翻訳されない。

※8 センスとアンチセンス

遺伝情報としてタンパク質に合成されうる配列の方向性のこと。アンチセンスはセンス配列に対して相補的で逆の方向性。

※9 RNA ペア

同一ゲノム上のセンスとアンチセンスの両方向の転写RNAが、相補的に結合した複合物の状態。

※10 CAGE 法（Cap Analysis of Gene Expression）

耐熱性逆転写酵素や cap-trapper 法を組み合わせ、5'末端から20塩基のタグ配列を切り出し、塩基配列を決定する実験技法のこと。

※11 5'端(と3'端)

核酸合成は、構成単位のヌクレオチド分子内の五単糖の炭素の位置で考えると5'から3'方向へ進むので、鎖の5'端が先頭になり、3'端が末尾となる。

※12 転写

遺伝子 DNA から RNA が読み取られること。

※13 CpG と TATA 配列

ゲノム DNA 中でシトシン (C) とグアニン (G) の二種の塩基が連続した配列で、遺伝子上流に多く存在し、プロモーター活性を持つ場合が多い。遺伝子上流部分の指標として利用されている。TATA は、チミン (T) とアデニン (A) の繰り返しで転写開始点の上流 (25bp 以上) にしばしば見出される。

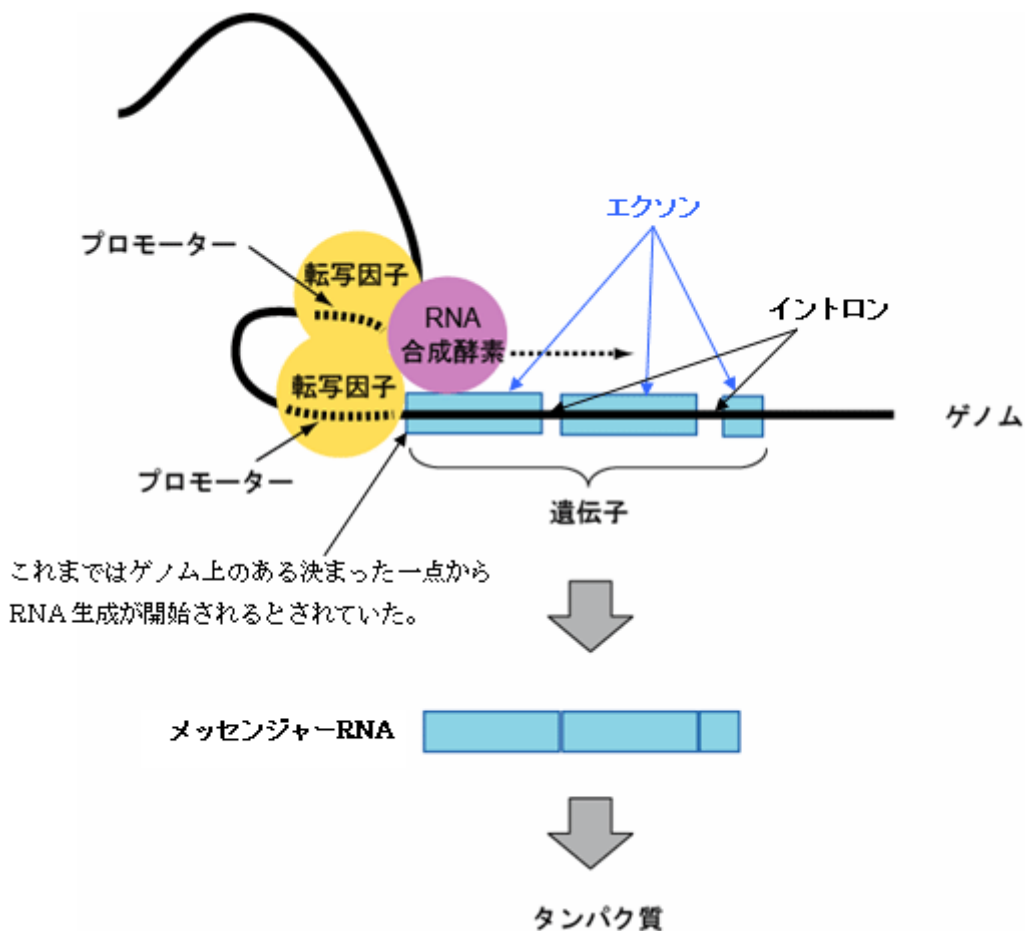


図1 プロモーターとは？

ゲノムは、ある生物を作り、維持し、増殖させるために必要な遺伝子という概念の全体、または、ある生物の遺伝的特性全体を規定する遺伝情報の最小単位である。プロモーターは、転写開始を促すゲノム上の配列のことをさす。RNA 合成酵素により DNA 配列から RNA 配列が合成される。さらにそこからイントロンと呼ばれる部分が抜け落ちたメッセンジャーRNA ができ、それからタンパク質がつくられる。今回、エクソン部分の端から読まれていると考えられていた RNA 生成の開始点が、ある一点ではなく、ある一定の範囲内から開始していたことがわかった。このある一定の範囲内から RNA 生成開始を指定するプロモーターを「ブロード型」と命名した。

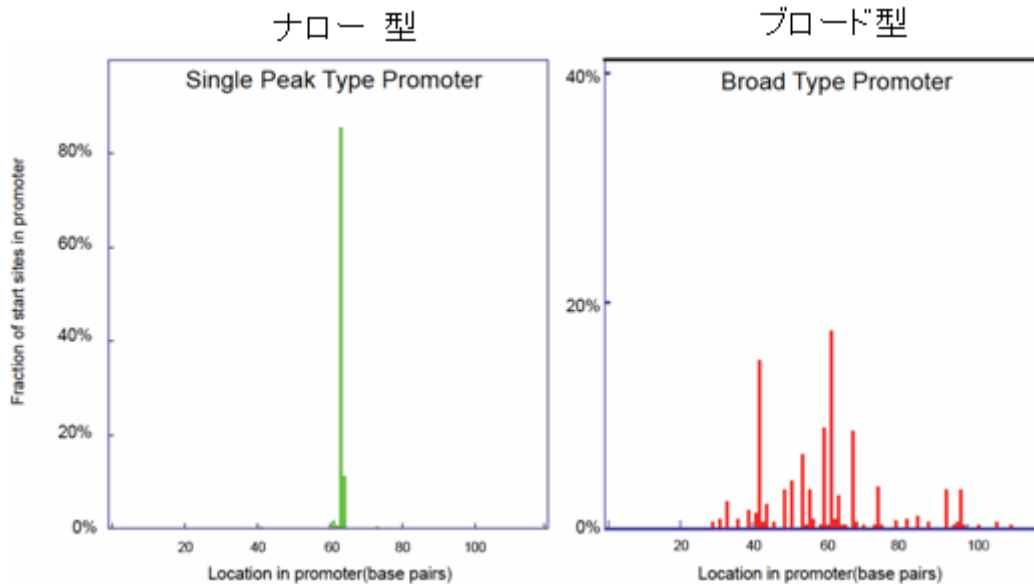


図2 プロモータータイプの比較

ナロー型、ナロー型のプロモーターの比較。ナロー型プロモーターでは転写開始点が厳密に決まっているのに対し、ブロード型プロモーターでは転写開始点は幅を持つ。

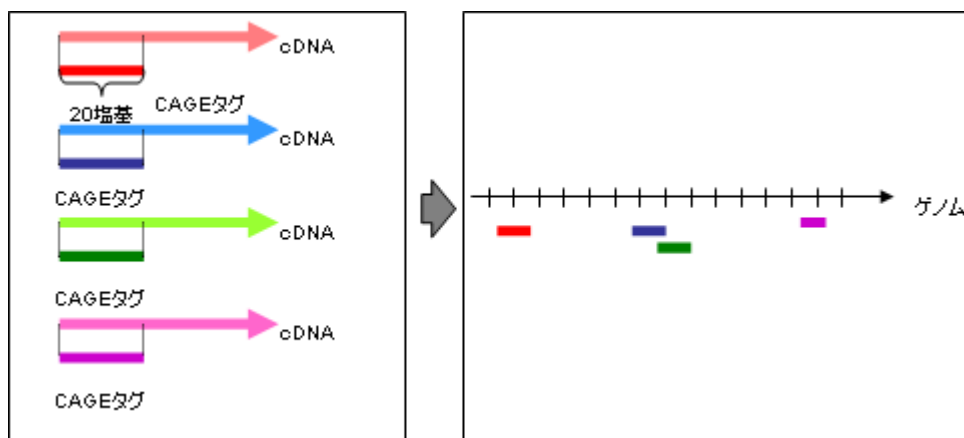


図3 CAGE法

CAGE法は、cDNA (RNA から DNA に変換したもの) の5'末端から20塩基のタグ配列を切り出し、塩基配列を決定する方法。RNAの転写開始点の同定ができる。