

## 生き物の進化における SINE 配列の機能を解明 ～ジャンク DNA ではない、SINE の新たな一面～

国立大学法人東海国立大学機構 名古屋大学大学院生命農学研究科の一柳 朋子 研究員、一柳 健司 教授らの研究グループは、マウス亜種間の比較研究を行い、ゲノム配列の中に何十万コピーとある SINE 配列がゲノムの中を飛び回りながら、周辺の遺伝子のエピジェネティック制御プログラムを書き換え、遺伝子発現に影響を与えていることを発見しました。

生物の進化は遺伝的な変化、表現型変化、そして自然選択を通して起こります。表現型の変化は新しい遺伝子が出来たり、既存の遺伝子が壊れたりする以外に、遺伝子の発現量や発現パターンが変化することでも起こります。後者の場合、どのような遺伝的な変化が重要なのかはあまり分かっていません。一方、ほぼ全ての生物のゲノムには「転移因子」と呼ばれる配列が多数存在します。これらは、ゲノムの中を飛び回りながら自身のコピーを増やすジャンク DNA だと考えられていました。

今回の私たちの研究は、転移因子の一つである SINE が転移することで、ゲノムの変化だけでなく、エピジェネティックな違いが作られて遺伝子発現が変化することを明らかにしました。私たちのゲノムの約半分は転移因子の転移によってできたものです。本研究の結果は、継続的な転移因子の活動によってゲノムが拡大し、その過程で表現型の変化が促されたことを示唆しています。また、本研究では数十万に及ぶ似た配列を個別化して解析することに成功しました。このことは、ポストゲノム時代に残されたダークマターである転移因子の理解を深め、ゲノムの機能や成り立ちの完全解読に寄与すると期待されます。

本研究成果は、2021年2月16日付（日本時間2月17日午前6時）国際分子生物進化学会誌『Molecular Biology and Evolution』オンライン版に掲載されました。

## 【ポイント】

- ・異なる亜種由来のマウス系統 (C67BL6/J [*Mus musculus domesticus* 由来] と MSM/Ms [*Mus musculus molossinus* 由来]) のゲノム配列比較から、約 2000 箇所 SINE 挿入多型 (片方の系統のみに挿入されている SINE コピー) を同定した。
- ・大規模エピゲノム解析により、エピジェネティックな状態 (クロマチンの化学修飾状態) が変化する遷移点、つまりバウンダリーに SINE が多くあることが分かった。
- ・挿入多型になっている SINE コピーは新たにエピジェネティック修飾のバウンダリーとなり、周辺のエピジェネティック状態を変化させ、遺伝子発現量も変化させていた。
- ・転移によってできた SINE コピーも含め、SINE にはクロマチン高次構造の制御因子である CTCF が結合していた。マウス ES 細胞での全 CTCF 結合部位のうち、6 個に 1 個は SINE によって作られた結合部位だった。
- ・これらのことから、SINE は「動くクロマチンバウンダリー」であり、SINE 転移によって新たなバウンダリーがつけられ、この進化プロセスは今もゆっくりと進んでいると考えられる。
- ・SINE の転写 (SINE RNA の合成) は成体組織ではほとんどなく、一方、生殖細胞で盛んであることが分かった。生殖細胞で転移した場合、次世代に受け継がれることになる。
- ・世界で初めて SINE 転写産物の網羅的解析に成功し、1 万を超えるゲノム領域から SINE RNA が合成されていることが分かった。これらの 1 万カ所の中でも発現量の差は 1 万倍以上あり、発現量と発現ロカス数は冪乗則 (べきじょうそく) に従うことを発見した。

## 【研究背景と内容】

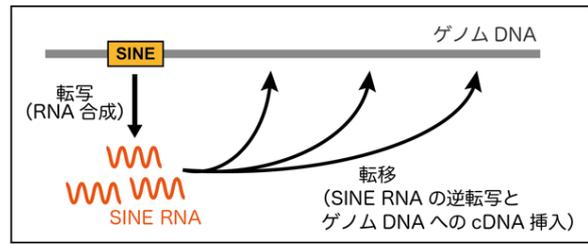
生物は進化の過程で新たな性質 (形質) を獲得することがあります。そのような新規形質獲得はゲノム DNA 配列の変化を伴います。例えば、新しい遺伝子の獲得や既存遺伝子のタンパク質アミノ酸配列の変化などです。しかし、このような変化よりも、既存遺伝子の発現量 (タンパク質の合成量) や発現する組織・細胞種を変化させるようなゲノム配列変化が重要な働きをしていることが示唆されています。

一方、生物のゲノム配列の中には転移因子 (トランスポゾン) と呼ばれる散在性の反復配列がたくさん存在しています。転移因子は 20 世紀の中頃に「遺伝子を制御する因子 (controlling elements)」として発見されたのですが、その後、長い間、「利己的に自分のコピーを増やす DNA で、ガラクタである」と考えられ、ジャンク DNA と呼ばれていました。しかし、21 世紀のポストゲノム時代の到来とともに転移因子が様々な遺伝子の制御に関わる事が分かってきました。しかし、これらの例は転移因子の元の DNA 配列から長い進化の過程で配列変化が蓄積して、新しく遺伝子発現制御機能を獲得したものでした。つまり、「基本的には必要ないので一旦は倉庫の片隅にしまわれる」というガラクタのイメージは変わりません。転移する前からすでに制御機能を内包しており、転移によってすぐさま機能を発揮するものがあるのか、それが現在進行形で転移しながらゲノム機能を拡張しているのかどうかは知られていませんでした。

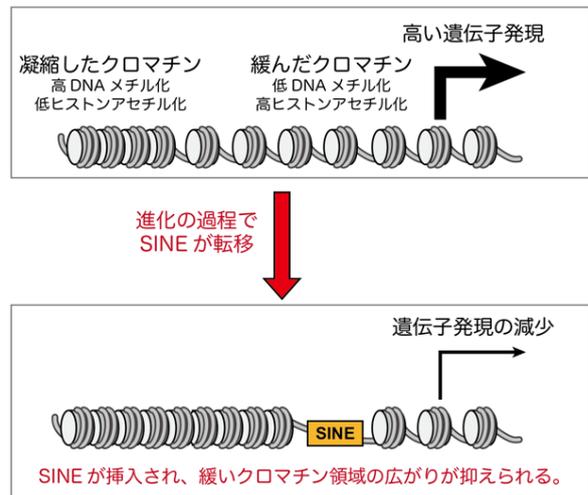
遺伝子をコードする DNA は真核生物の核内ではクロマチンと呼ばれる構造体の中に格納されており、このクロマチンを化学修飾することで、遺伝子の転写活性が制御されています。このような仕組みをエピジェネティクスと呼びます。

SINE(short interspersed elements)は200塩基対程度の短い転移因子で、タンパク質をコードしていない非コードDNAです。SINEはDNAからRNAに転写され、そのRNAから逆転写反応によって相補的なcDNAが作られ、それがゲノムDNAのどこかに挿入されるというメカニズムでコピーを増やします(図1)。ヒトやマウスのゲノム配列の中には150万コピーほどのSINEがあります。ヒトとマウスのSINEは進化の過程でそれぞれ独立に転移して現在の場所に収まっているのですが、不思議なことに、どちらの種でも遺伝子密度の高いゲノム領域にSINEも多く存在しています。これは遺伝子の近くにはあまり存在しない他の転移因子グループとは異なります。SINEの中のいくつかのファミリーは今でも転移活性を持つことが知られており、種内のゲノム配列の多様性の創出に寄与しています。

(図1) SINEの転移機構



(図2) 新たな発見：SINE転移によるエピジェネティック進化



私たちの研究グループでは転移したばかりのマウスSINEに注目し、まずはそのようなSINEコピーをゲノム配列の中から網羅的に探し出すことにしました。これは2つのマウス系統のゲノム配列を比較し、挿入多型のコピーを同定することで行いました。次にクロマチン化学修飾に関する大規模解析を行い(エピゲノム解析と言います)、エピジェネティックな状態(クロマチンの化学修飾状態)が大きく変化する遷移点、つまりバウンダリーにSINEが多くあることが分かりました。しかも、挿入多型になっているコピーに着目すると、SINEがない系統ではクロマチン修飾状態の分断が見られなかったもので、確かにSINEが転移してバウンダリーが形成されたことが分かりました(図2)。さらに、この効果があるせいで、転写開始点近くの緩んだクロマチン領域の範囲が小さくなり、その遺伝子の転写量が大きく減少していることを見つけました。また、SINE配列にはクロマチン高次構造の制御因子であるCTCFが結合しており、全CTCF結合部位のうち、6個に1個はSINEが作り出した結合部位であることが分かり、マウスの進化の過程でSINEによって大きくクロマチン高次構造が変化してきたことが推察されます。

これまでの手法ではSINE RNAの網羅的解析を行うことは不可能で、SINEを含めた転移因子は、特に転写の観点からはポストゲノム時代に残されたダークマターでした。そこで、新たな方法を開発し(meI RNA-seq法と名付けました)、150万のコピーのどこからSINE RNAが作られているのかを解析しました。その結果、約1万のコピーからSINE RNAが作られており、しかもその発現量は均一ではなく、大きな分散があることが分かりました(1万倍もの差)。発現量には冪乗則が観察され、弱く発現する大量のコピーと強く発現する少数のコピーがありました。

### 【成果の意義】

今回の研究によって、遺伝子制御機能(正確に言えばクロマチン制御機能)を配列内に内包したSINEが存在し、それが現在進行形で転移によってコピー数を増やしており、マウスの進

化に寄与していることが分かりました。ゲノムの進化を理解する上で重要な知見です。

現代人では約 20 人に 1 人の割合で転移 SINE コピーが発生していると言われていいますので、ヒトの進化や多様性にとっても SINE は重要な役割をもっていると考えられます。現在、ヒトの多様性、特に疾患感受性に寄与する塩基置換に関する情報が精力的に集められていますが、SINE 挿入多型も考慮に入れた解析が必要になることを示唆する研究結果です。

遺伝子発現においても冪乗則は一般に観測されており、それは遺伝子産物（タンパク質）のいくつかは他の遺伝子の発現を制御する階層的な制御機構があるからと考えられています。一方、SINE はどのコピーもよく似た転写プロモーター配列を持ち、しかも階層的な制御関係はありません。このような配列群で転写量の冪乗則が発見されたことは、転写反応の制御機構に新たな視点をもたらしました。

### 【用語説明】

**転移因子**：自身の配列をコピーして、ゲノムの別の場所に挿入する DNA 配列の総称。ヒトゲノムの中に数百万コピーあり、ゲノムの約 40%を占める。

**SINE**：短い転移因子の一つで、タンパク質はコードしておらず、転移反応においては別の転移因子（LINE と呼ばれる）の逆転写酵素が使われる。網、目、属レベルに固有のファミリーが存在し、霊長目特異的なものとしては Alu、齧歯目特異的なものでは B1、B2 などがある。また、哺乳綱特異的なものでは MIR がある。

**挿入多型**：大部分の転移因子の挿入箇所は種内の個体集団に共通しているが、ごく最近に転移したものは全ての個体にあるわけではなく、一部の個体のみに見られる。このような状態のことを挿入多型という。

**クロマチン**：真核生物の核内で DNA とタンパク質が複合体を形成しており、その複合体のことをいう。タンパク質として最も多いのはヒストンであり、そのほかに転写因子、CTCF のような構造制御因子、ヒストン修飾酵素、DNA メチル化酵素などもクロマチン構成要素である。

**エピジェネティクス**：クロマチンの化学修飾や構造変換により、遺伝子の発現状態を制御するメカニズムのこと。例えば、転写活性化にはヒストンのアセチル化を必要とする。一つの受精卵から同じゲノム配列を持つ細胞を大量に作りながら体を形成していく発生過程では、エピジェネティックな遺伝子発現制御機構は特に重要である。さらに、がん抑制遺伝子の発現抑制のように、成体であってもエピジェネティックな変化が疾患などに関与することが知られている。

**転写プロモーター配列**：転写を開始するために必要な DNA 塩基配列。RNA ポリメラーゼが DNA に結合することを促進する。

**冪乗則（べきじょうそく）**：一つの変数の対数とその頻度の対数が直線的な関係になる分布があらわれること。この分布は冪分布ともいい、遺伝子発現量の分布のほか、戦争の大きさ、地震の大きさ、本の売り上げ、所得などでも見られると言われる。スケール不変性（フラクタル性）があり、どの尺度に拡大、縮小しても同じようになるため、正規分布の平均値のような「集団を代表する値」というのが存在しない。

**【論文情報】**

雑誌名 : Molecular Biology and Evolution

論文タイトル : B2 SINE copies serve as a transposable boundary of DNA methylation and histone modifications in the mouse

著者 : 一柳朋子<sup>1</sup>、加藤大和<sup>1</sup>、毛利嘉伸<sup>1</sup>、平福啓一<sup>2</sup>、Beverly Ann Boyboy<sup>1</sup>、川瀬雅貴<sup>1</sup>、一柳健司<sup>1</sup>

所属 : <sup>1</sup>名古屋大学大学院生命農学研究科、<sup>2</sup>東京慈恵会医科大学附属病院

DOI : 10.1093/molbev/msab033