

論文 / 著書情報
Article / Book Information

| | |
|-------------------|---|
| 題目(和文) | ヒトレトロトランスポゾンL1のプロモーターの生物学的意義とその解析法の構築 |
| Title(English) | |
| 著者(和文) | 東野真之 |
| Author(English) | Saneyuki Higashino |
| 出典(和文) | 学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第9609号, 授与年月日:2014年9月25日, 学位の種別:課程博士, 審査員:相澤 康則,伊藤 武彦,清尾 康志,林 宣宏,梶川 正樹 |
| Citation(English) | Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第9609号, Conferred date:2014/9/25, Degree Type:Course doctor, Examiner:,,,, |
| 学位種別(和文) | 博士論文 |
| Category(English) | Doctoral Thesis |
| 種別(和文) | 要約 |
| Type(English) | Outline |

博士論文 要約ファイル

ヒトレトロトランスポゾン L1 のプロモーターの
生物学的意義とその解析法の構築

東京工業大学

大学院生命理工学研究科 分子生命科学専攻

東野真之

(指導教官 相澤康則 講師)

1-1 本研究の背景

1-1-1 レトロトランスポゾン

レトロトランスポゾン（以下、RE）とは、自身のプロモーターや転写終結配列を有し、そのプロモーターの転写活性化から始まる一連のプロセスによって、そのゲノム内のコピー数を増やすゲノム内転移因子である(Deininger and Batzer, 2002; Kazazian, 2004)。RE 転移では、RE から転写された RNA を鋳型に逆転写反応によって生成される cDNA が、ゲノム内の別の位置に挿入され、新 RE コピーが生じる。そのため、RE 転移が 1 回起こる度に、ゲノム全体での RE のコピー数は 1 つ増える。これに対して、DNA トランスポゾンと呼ばれる転移因子の場合は、ゲノム内の自身の DNA 配列断片を切り出して他の部位に転移させるため、転移前後でゲノム内総コピー数は変動しない。このため、DNA トランスポゾンの転移は、「cut-and-paste」方式と呼ばれるのに対して、RE 転移は「copy-and-paste」方式と表現される（図 1-1）。

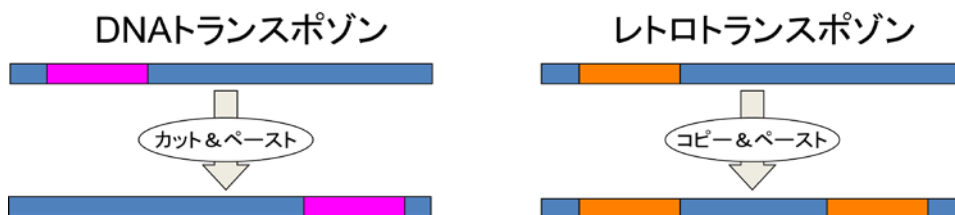


図 1-1 DNA トランスポゾンとレトロトランスポゾンの転移様式の違いを示す模式図。DNA トランスポゾン（ピンク四角）は自らの配列を切り出して他の場所に転移させるのに対して、レトロトランスポゾン（オレンジ四角）は自らの配列をコピーして転移させる。

RE はさらに、以下の大きく 3 つのグループに分類されている：

- LINE (long interspersed nuclear element)
- LTR (long-terminal repeat)
- SINE (short interspersed nuclear element)

これらのうち、前者 2 グループは全長が 5,000–7,000 塩基対からなり、その自身の配列内部に、転移に必要なタンパク質をコードしているのに対して、SINE に属する RE は転移に必須のタンパク質をコードしていない。そのため SINE 型 RE が転移する際には、そのペアーとなる LINE 由来のタンパク質が必要になることが知られている(Kajikawa and Okada, 2002)。一方、全ての RE で共通な特徴として、RE 自身配列の 5' 末端領域に内部プロモーターを有しているが、タンパク質をコードしているかどうかの違いにより、SINE 型 RE のプロモーターは RNA ポリメラーゼ III 用の内部プロモーターを有しているのに対して、LTR 型や LINE 型の RE のプロモーターは主に RNA ポリメラーゼ II によって転写されている。

1-1-2 転移因子 L1

哺乳類ゲノム内の RE は上述のように、そのゲノム進化の過程で「Copy-and-paste」型の転移を繰り返してきたようで、全ての哺乳類ゲノムは多くの RE コピーに占められている。また、種分化ののちに新たに発生した RE ファミリーや、進化の過程で転移能を失った RE ファミリーなど、多種多様な RE ファミリーやサブファミリーが、現存する哺乳類のゲノム内で確認される。その中で、LINE 型に属する LINE1 ファミリー (以下 L1) は、ほぼ全ての哺

乳類ゲノム内で非常に多くのコピー数を誇る RE ファミリーとして知られている。

図 1-2 に全長 L1 内の機能配列構成を示す。全長の L1 は約 6,000 塩基対からなり、大きく以下の 5 つの部分から構成されている。

① 5' 非翻訳領域 (5'UTR) :

RNA ポリメラーゼ II のプロモーター活性をもつ。

② ORF 1 (Open Reading Frame 1) :

約 300 アミノ酸長の ORF1 タンパク質がコードされている。ORF1 タンパク質は後述する ORF2 タンパク質と共に、L1 転移には必須である。ORF1 タンパク質には RNA シャペロン活性があることがインビトロ実験系で示されている。

③ リンカー配列 :

2 つの ORF 領域を分ける約 70 塩基対長の配列である。

④ ORF 2 (Open Reading Frame 2) :

約 1,300 アミノ酸長のタンパク質 (ORF2 タンパク質) がコードされている。後述するように L1 転移に必須な DNA 切断酵素活性と逆転写酵素活性を有している (Mathias et al., 1991; Feng et al., 1996)。

⑤ 3'UTR :

弱いながらもポリ A 付加シグナル配列が存在しており、L1 転写を終結させる (Belancio et al., 2007)。

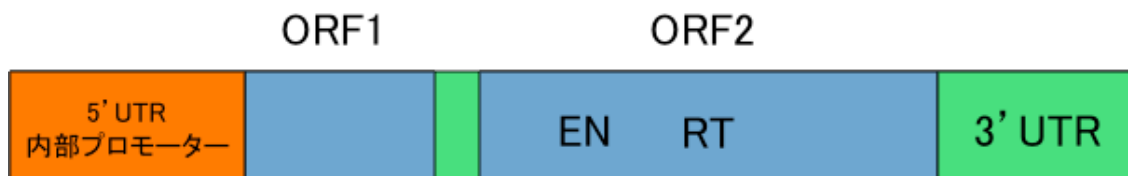


図 1-2 全長 L1 の内部構造。約 6kb。5'UTR はプロモーター活性をもっている。2 つの ORF を持ち、それらから翻訳されるタンパク質は、ともに L1 転移に不可欠である。ORF 2 タンパク質には、エンドヌクレアーゼ (EN) および逆転写酵素 (RT) の両活性をもつドメインがある。

1-1-3 ヒトゲノム内の L1 コピーの数と構造

ヒトゲノムでは、その全配列の少なくとも 60%が RE 由来の配列であり (de Koning et al., 2011)、L1 は、他の RE ファミリーと比べて、ヒトゲノム配列の最も広範囲を占め、その占有率は 17%にのぼる (Lander et al., 2001)。また、そのコピー数は一倍体ゲノム当たり 50 万に達する。これらはお互いに相同性の高い配列を保持しているため L1 由来であると配列解析上認識可能なコピーであり、進化上古い L1 (転移してから長い進化時間を経た L1) の中には多くの変異を蓄積したために L1 と識別不可能になっているコピーもあることが容易に想像できる。ヒトゲノム内部には多くの L1 コピー由来の配列が散在しているわけである。

実際にヒトゲノム内に認められる全 L1 コピーのうち、図 1-2 で示したような全長の構造を保持した L1 コピーはごく一部であり、2002 年の Szak らの報告では約 3,000 コピーしか存在しない。それ以外の L1 コピーは、3'末端のみの配列をもつものであり、5'末端が欠けた不完全な配列となっている (図 1-3) (Szak et al., 2002)。

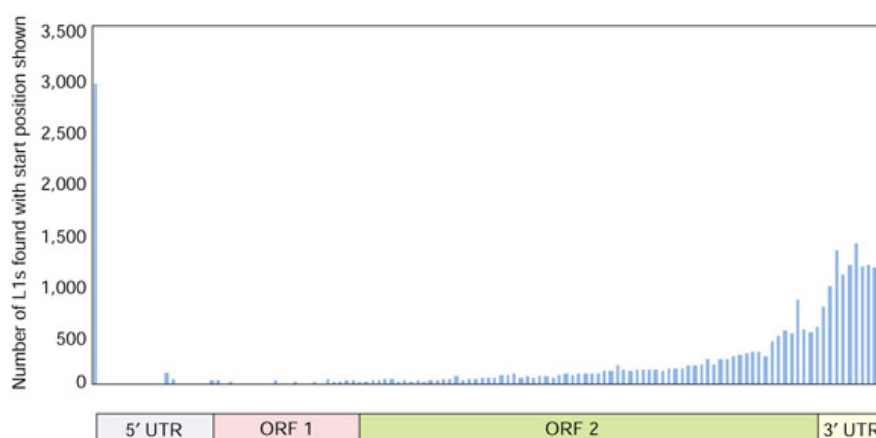


図 1-3 ヒトゲノム内に存在する L1 の全長での開始場所を示すヒストグラム。横軸がそれぞれのコピーの開始点の位置、縦軸が数を表している。Szak et al., 2002 より引用。

しかも、これら全長 L1 コピーの多くは転移後、2つの ORF 内部にフレームシフトやノンセンス変異を蓄積しているおり、転移活性を保持している L1 コピー（2つのタンパク質を全長のままコードしておりかつ、培養細胞を用いた実験により転移活性が確認されているもの）は、1 ヒトゲノムあたり 50 コピーほどしか存在しない(Brouha et al., 2003)。

1-1-4 L1 転移の機構と影響

L1 転移の概要を図 1-4 に示す。L1 の転移は、内部プロモーターから L1 配列が転写されることから始まる。L1 が転写されると、続けて2つの ORF からタンパク質がそれぞれ翻訳され、L1 RNA-タンパク質複合体（以下 L1RNP）を形成する。ORF2 タンパク質が有する DNA 切断ドメインおよび逆転写ドメインは L1 転移を触媒するため、ORF2 タンパク質は転移反応に不可欠である(Mathias et al., 1991; Feng et al., 1996)。一方、ORF1 タンパク質も典型的な L1 転移には不可欠であることが培養細胞を用いた実験で示されているが、ORF1 タンパク質が担う、L1 転移における分子機構はまだ十分には明らかにされていない。これまで生化学的に、ORF1 タンパク質には RNA 結合活性や RNA シャペロン活性が確認されているが、L1 転移との機能的つながりは未解明のままである(Martin and Bushman, 2001; Martin et al., 2005)。

ORF2 タンパク質は、5'-TTTTAA-3'の配列またはこれに類似した配列を特異的に認識し、T と A の間のリン酸ジエステル結合を選択的に切断したのち、その結果生じる 3' 末端から L1 cDNA を伸長することから、L1 転移先配列はこのような配列特異性を示す場合が多い(Symer et al., 2002)。しかし、DNA 損傷

が起こりやすい細胞環境では、この ORF2 タンパク質による DNA 切断を介さずに L1 転移が起こる場合があります、この場合、L1 の転移先はほぼランダムに決まると考えられている(Morrish et al., 2002)。

しかも、L1 転移は、単純な L1 コピーの挿入だけで終わる場合だけでなく、挿入先の上流や下流のゲノム領域を欠損あるいは転座させるケースも高頻度で発生することが知られている(Goodier and Kazazian, 2008)。すなわち、大規模なゲノムの組換えも同時に起こる場合がある。

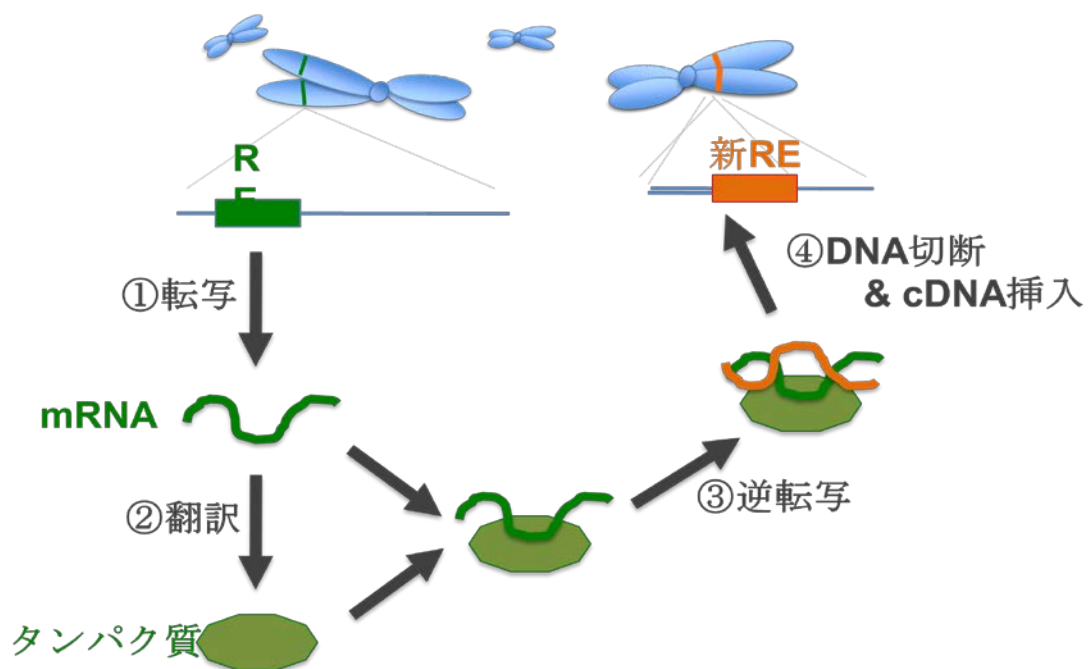


図1-4. RE 転移の仕組み。RE が転写されると、それにコードされているタンパク質が翻訳される。翻訳されたタンパク質は RNA と結合し、逆転写が行われることで、新たな RE コピーがゲノムに転移する。

1-1-5 ゲノム上に存在する L1 の 2 つの機能的意義

以上の研究背景の概要をもとにすると、ヒトゲノムを構成する主要配列であ

る L1 の生物学的インパクトは主に、その転移能に起因するといえるであろう。L1 の転移先は概ね非選択的であり、ゲノム内の様々な部位に転移するため、転移が遺伝子のエキソン内部に起こると、その遺伝子がコードするタンパク質のアミノ酸配列を変化させ表現型に影響することもある。実際、L1 転移による遺伝子破壊が発病の原因である症例が少なくとも 25 件報告されているが(Hancks and Kazazian, 2012)、これらは氷山の一角であると考えられている。これら L1 挿入の発見は、単一遺伝子疾患研究の過程で、注目すべき原因遺伝子領域を詳細に解析した結果であるため、原因遺伝子が判明していない単一遺伝子疾患の場合や、原因遺伝子が単一ではなく複数の多因子遺伝性疾患の場合も、L1 挿入がその原因となっている可能性は高い。このように、ヒトゲノム内で転移活性を保持している約 50 の全長 L1 コピーの生物学的意義については、転移によるゲノムや遺伝子の改変が重要であることは疑いの余地はない。

一方、ヒトゲノム内の残り大多数の、転移活性を失った L1 コピーにも機能的なインパクトがあることが報告されている。例えば、L1 上にはプロモーター、スプライシング部位、ポリ A 付加シグナルといった転写に影響を与える配列が存在する。そのため、L1 がエキソンに挿入せず、イントロン内部や遺伝子の上流や下流等の近傍に挿入した場合、これらの転写制御配列によって、近傍遺伝子の転写に変化が現れる例が知られている。(Perepelitsa-Belancio and Deininger, 2003; Han et al., 2004; Roy-Engel et al., 2005)。特に、L1 の 5'UTR にある内部プロモーターは双方向に転写開始活性をもち、転移の際に L1 自身の mRNA を転写するセンス方向のプロモーター活性だけでなく、アンチセンス方向のプロモーター（以下、ASP）にも転写開始活性があることが知られている

(Speek, 2001; Nigumann et al., 2002; Wheelan et al., 2005; Matlik et al., 2006)。

図 1-5 に示すように、ASP 活性のある L1 と、その近傍遺伝子との位置関係は、大きく次の 3 つのケースが考えられる：

- ① Intronic L1 : L1 が遺伝子のイントロン内部に存在する場合、L1 から転写が始まり、その下流のエキソンとスプライシングによってつながった RNA が生成される。
- ② Upstream L1 : L1 が遺伝子上流にある場合、L1 から転写が始まり、遺伝子の第 2 エキソンとスプライシングによってつながった RNA が生成される。
- ③ Intergenic L1 : L1 が遺伝子間領域にある場合、遺伝子とは無関係に近傍のゲノム配列が転写・スプライシングされる。

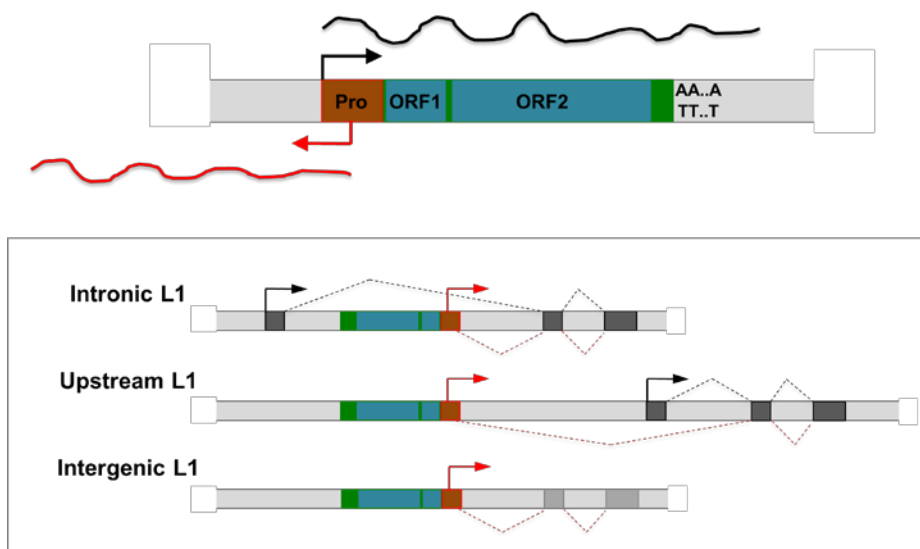


図 1-5 双方向の転写開始活性をもつ L1 内部プロモーターの概略図と、遺伝子内あるいは遺伝子間 L1ASP から生成される RNA のスプライシング構造様式。

前者 2 つのケースでは、転写されてくる RNA は、遺伝子の mRNA の 3' 末端配列とのキメラ体となるため、遺伝子のタンパクコード領域を含む場合がある。

原ガン遺伝子 *MET* のイントロンにある L1ASP から転写されるキメラ RNA については、*MET* タンパク質のドミナントネガティブ変異体がコードされている可能性を提案している過去の報告もある(Weber et al., 2010)。また、遺伝子間領域にある L1ASP からは、新しい遺伝子グループとして注目されている長鎖ノンコーディング RNA が転写される可能性も指摘されている。

しかしながら、L1ASP については、以下のように大切な疑問点が残っている。

(1) どの L1 コピーが ASP からの転写活性をもつのか？

上述のように、内部プロモーター配列をもつ全長 L1 がヒトゲノム内に約 3,000 コピー存在するのに対し、転写活性を有するコピーとしてこれまで同定された L1ASP は約 70 コピーであった。未同定の転写活性能を有する L1ASP は他にいくつくらい、ゲノム内のどこに位置しているのか？

(2) L1ASP の転写制御メカニズムは？

上述のように、L1 転移が宿主遺伝子の破壊につながる高いことから、現在 L1 プロモーターは宿主システム（主にエピジェネティクス機構）によって一義的に抑制されていると考えられているが、この仮説は、L1ASP 由来の多くの転写産物が EST や cDNA として確認されている事実と矛盾している。

1-1-6 L1 によるヒトゲノム構造多型

転移活性能を保持している全長 L1 コピーは、ヒト共通祖先から現代までの間に、さまざまな人種や系統内で転移を繰り返した結果、現代人ゲノム中の一部の L1 は多型として存在している。これをレトロトランスポゾン挿入多型 (Retrotransposons insertion polymorphism, 以下 RIP) という。RIP は、特

定のゲノム座位における L1 コピーの有無の違いであり、一塩基多型(以下 SNP)のようにヌクレオチド塩基の違いや、コピー数多型のように遺伝子コピー数の違いとは別の様式で、個人間のゲノムの差異の原因となっている。

近年のヒトゲノム多型 1000 人ゲノムプロジェクト(Abecasis et al., 2010)によると、ヒトゲノム参照配列上にない L1 コピーが約 1000 カ所に存在することが明らかになっている(Ewing and Kazazian, 2011)。前節で述べたように、L1 は近傍遺伝子の転写に対して、ASP を介して定性的および定量的な影響を与えるため、多型 L1 が個人間での遺伝子発現様式の違いを生み出す可能性も十分ある。しかし現時点では、SNP タイピングのように、多検体に対して並列的にかつ高効率で多型 L1 を評価する方法論は存在しないため、多型 L1 の予防診断医療への新しい可能性を検証することはできない。

1-2 本研究の目的

以上の研究背景をもとに、本博士論文研究では、ヒトゲノムにおける L1 の役割や機能的インパクトをより深く理解するために、L1ASP に関して 2 つの側面から研究を行った。

1 つめとして、L1ASP からの転写活性をゲノムワイドに観察し、L1ASP の転写活性化の機構を明らかにすることを目的とした。具体的にまずは、転写活性化している L1ASP を網羅的に同定するための方法を、次世代シーケンサーを活用して開発することとした。そして新方法の実用性を確認したあとに、その新方法をヒト細胞に対して用い、転写活性がある L1ASP コピーの総数やゲノム座位を特定した。そして、転写活性を確認できた L1ASP と、確認できなかった

た L1ASP の各種特徴を比較することで、L1ASP の転写様式に関する新しい知見を求めた。

L1ASP の多型性が、個人間のゲノム機能の差異につながるかどうかを調べるためには、L1 の多型状態を多くのゲノム検体に対してハイスループット方式で行える必要がある。そこで本博士論文研究の 2 つめの目標として、L1 多型状態を、L1 の近傍 SNP の解析を通して評価できるかどうか、すなわち、多型 L1 と連鎖不均衡にある SNP が存在するかどうかを調べることにした。もし多型 L1 と同じハプロタイプを形成している SNP が高頻度で存在すれば、L1 多型状態はその SNP のヌクレオチドの種類によって間接的に判明するため、L1 多型は現行 SNP タイピング法によってゲノムワイドの評価が可能となり、多型 L1 と疾患との相関性の評価が容易となる。

発表論文

“Polymorphic L1 retrotransposons are frequently in strong linkage disequilibrium with neighboring SNPs.”

Higashino S, Ohno T, Ishiguro K, Aizawa Y.

Gene. 541(1), 55-59 (2014).

“The bonobo genome compared with the chimpanzee and human genomes.”

Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, Koren S, Sutton G, Kodira C, Winer R, Knight JR, Mullikin JC, Meader SJ, Ponting CP, Lunter G,

Higashino S, Hobolth A, Dutheil J, Karakoç E, Alkan C, Sajjadian S, Catacchio

CR, Ventura M, Marques-Bonet T, Eichler EE, André C, Atencia R, Mugisha L,

Junhold J, Patterson N, Siebauer M, Good JM, Fischer A, Ptak SE, Lachmann M,

Symer DE, Mailund T, Schierup MH, Andrés AM, Kelso J, Pääbo S.

Nature. 486, 527-31 (2012).