

お隣さんとお引っ越し：L1 の 3' transduction

ヒトの non-LTR retrotransposon、L1 (LINE-1)はしばしば自分の 3'末端よりも後ろ側に隣接した配列を転移させてしまうことが知られている。この現象は 3' transduction と呼ばれている。原理的には逆に 5'側の配列を転移させる現象もあり得る。こちらは 5' transduction と呼ばれているが、5' transduction の報告は、これまでのところゲノムの論文に 2 つ見つけたという記載があるのみ (Lander et al. 2001) だから、ここでは、3' transduction に絞って紹介したい。

3' transduction が実際に起こることが示されたのは 1994 年のことである (Holmes et al. 1994)。Holmes らは家系 JH-1001 の筋ジストロフィーの原因が、ジストロフィン遺伝子のエクソン 48 への L1 の挿入であることを発見した。挿入された配列の長さは 1,968bp であることが他のヒトとの比較から明らかとなった。両側には、転移因子の挿入の特徴である、12 から 15bp の target site duplication (TSD)も認められた。しかし、内部の配列はというと、その 5'側 1,401bp は確かに L1 で、37bp の polyA tail を伴っているが、3'側 530bp は L1 と無関係な配列であり、3'末端に 41bp の polyA tail が見つかった。彼らは、L1 と無関係な配列を Unique Sequence Component (USC)と呼んでいる。USC をプローブとして利用することで転移した元の L1 のコピーが見つかり、LRE2 と名付けられた。USC は LRE2 の TSD の 3'側に存在しているため、LRE2 自身は USC を含んでいない。また、LRE2 の下流の USC は、polyA tail を伴っていない。以上から、LRE2 が 3'側の USC を含んだ形で転写され、polyA シグナル ATTAAG によって新たに polyA 付加され、この polyA tail から逆転写されてジストロフィン遺伝子中に挿入されたことが明らかとなった。

以降、同様の例はいくつか報告されている。McNaughton らはジストロフィン遺伝子第 7 イントロン内に別の USC を伴った L1 が挿入されていることを発見した (McNaughton et al. 1997)。ここでは、15bp の TSD (AGAAATCTGATTGA)が見つかったが、挿入されていない配列との比較はなされていない。L1 の polyA と USC の polyA との間の配列は、975bp で polyA の上流 18-13bp には AATAAG の polyA シグナルも見つかる。また、Rozmahel らは 160bp ほどの L1 と下流側の cystic fibrosis transmembrane conductance regulator (CFTR)遺伝子の配列と一緒に転移していることを明らかにしている (Rozmahel et al. 1997)。L1 の下流には CFTR 遺伝子のエクソン 9 とその周辺のイントロンの配列合わせて 647bp がつながっており、後ろ側に AATAAG の polyA シグナルと 8bp の polyA tail、前後に 9bp の TSD (AAACAGACA) が存在する。これが TSD であることは挿入の無い対立遺伝子座があることから証明された。

Brouha らは chronic granulomatous disease (CGD) の患者の病因を調べ、CYBB 遺伝子の第 4 エクソンに挿入されている L1 を発見した (Brouha et al. 2002)。この挿入は 1,722-1,724bp の L1 と 280bp の 3' transduction された配列とで構成されている。3' transduction された配列の 3' 末端は 101bp の polyA で、少し前に AATAAA の polyA シグナルがある。L1 の 3' 末端の直後には GAAA の繰り返しがあり、ユニークな配列は 153bp しかない。長い TSD は見つからず、最も長く TSD を考えた場合でも 2bp (AA) にしかない。3' transduction された配列と同じ配列は、他にはゲノム中に 1 カ所だけ、2q24.1 座位にのみ存在していた。この患者の 2q24.1 座位には L1 の挿入は無かったが、患者の母親では全長約 6.1kb の L1 が挿入されており、LRE3 と名付けられた。この LRE3 が 3' 側の配列を伴って CYBB 遺伝子座に転移したと考えられる。また、van den Hurk らの見つけた choroideremia (CHM) 遺伝子に挿入された L1 では 414bp の 3' transduction が起きていることが Chen らによって報告されている (van den Hurk et al. 2003; Chen et al. 2005)。

実験的にも 3' transduction が起こることは証明された。Moran らは L1 の転移系において、L1 の polyA ではなく、より下流に人工的に接続した SV40 の polyA tail から逆転写が起きていることを見つけた (Moran et al. 1996)。確認のために、イントロンを挿入した逆向きのネオマイシン耐性遺伝子と SV40 の polyA シグナルを L1 の polyA の下流に付加したコンストラクトを作成し、これが転移するかどうかを調べた (Moran et al. 1999)。この系では、L1 の polyA を通り過ぎて SV40 の polyA シグナルまで転写が起こり、かつ、SV40 の下流の polyA から逆転写が起こらない限りはネオマイシン耐性にならないので、G418 を加えた培地では HeLa 細胞は生育できない。L1 の polyA の前にネオマイシン耐性遺伝子を入れたコンストラクトの場合と比較して、ほとんど違いのない効率で G418 耐性コロニーが観察されたという結果は、L1 が自身の下流の DNA を効率よく転移させることができることを証明している。L1 の転写をサイトメガロウイルスのプロモータからさせた場合と L1 自身の内部プロモータからさせた場合のどちらでも高い効率で 3' transduction が観察できた。SV40 の polyA シグナルを除去すると転移効率は 1% 以下にまで低下したので polyA から逆転写するという制約は存在しているらしい。

このように自然にも実験的にもヒトの L1 が 3' transduction を起こすことは確かめられた。上記の報告はかなり確実に 3' transduction が起きていることを示すものだが、もしかしたら 3' transduction が起きているのかも？というような事例も報告されている。例えば、Holmes らの報告よりも以前の、Miki らの論文 (Miki et al. 1992) では、APC 遺伝子中の L1 の挿入を報告しているが、この L1 では末端の polyA が AATAAAA₁₇TAACAATAAATGAGATAAAATCTA₁₈₇ のよう

な構造を取っており、 A_n に挟まれた（下線を付した）配列はもしかすると 3' transduction によってできた可能性がある。しかし単純に polyA 付加の過程で間違っただけの塩基が付加された可能性もあるし、転移した L1 の polyA tail がそのような配列を既に持っていた可能性も否定できない。Moran らは上記の L1 を 3' transduction の例として挙げている (Moran et al. 1999) が、Chen らは 3' transduction の起こっていない例として数えている (Chen et al. 2005) ように研究者によって解釈の食い違う場合もあるので注意が必要である。この問題は 3' transduction がどの程度の割合で起こるのかを解析する上で無視できない問題である。しかし、実際のところ、3' transduction の頻度の解析で最も問題になるのは以下に示すように、TSD の確からしさのようである。短い TSD を許すと偶然の配列の一致を取ってきやすくなるし、L1 の両側での TSD の検索範囲を広くすればそれだけ偶然のものが増えるのは自明だからである。

筆者の知るところでは、L1 の 3' transduction の頻度について定量的な解析を行ったのは、Goodier らが最初であろう (Goodier et al. 2000)。翌月には Pickeral らも別の手法で L1 の 3' transduction 頻度を報告している (Pickeral et al. 2000)。Goodier らは GenBank の配列から L1 の 3' 末端 100bp を含む配列を 102 個見つけ、シーケンスが不完全なものを除いた後、TSD(target site duplication)を持つものとして既に見つかっていた 15 個に加えて、新しく 66 個の L1 を決定した。以降はこの 66 個の L1 について解析を行っている。66 個の内 15 個 (23%) では、L1 の 3' 末端と TSD との間に余分な配列が見つかった。ただし、この際の TSD の最短は 7 塩基であり、通常の転移をした L1 の TSD の最短は 6 塩基と非常に短い場合も含まれている。5' 側の TSD は L1 の 5' 末端と接していることを条件としている。この余分な配列の長さを、L1 の polyA シグナル(AAUAAA)と余分な配列の後ろの polyA tail との間の塩基数で表現すると、39 から 874 であり、平均は 207 だった。Goodier らは同様の解析を全長のマウスの L1 についても行っており、24 の全長 L1 の内、2 つで 3' transduction が認められた。長さは 2,723bp と 3,327bp とヒト L1 と比べて異常に長い。これはヒトの L1 とマウスの L1 の性質の違いを反映しているのかもしれないが、L1 をとってくるのに 3' 末端 100 塩基を使ったのと全長を使ったのとの違いによる人工的なものかもしれない。彼らは、3' transduction を起こした L1 が更に 3' transduction を起こしているような例も見つけている。また、2 つの 3' transduction が同じ 42bp を共有しており、同じ L1 由来の転移産物である可能性を報告している。この兄弟 L1 は次の Pickeral らの論文でも報告されている。

Pickeral らの解析では、全長の L1 に絞って 6bp 以上の TSD を探索し、5' 末端に TSD が接している場合だけを選んでいく。全長の L1 だけに絞ったのは 5' 末

端の決定の作業が入って不確実性が増すのを避けたためである。こうして、97 個の TSD を持つ L1 を発見し、その内、21 個が 3' transduction を起こしている可能性があった。この 21 個を彼らは 3 グループに分類している。クラス 1 は polyA シグナルを持ち 89 から 975bp の 3' transduction を持っている L1 で、10 配列が該当した。クラス 2 は 4 配列が含まれ、polyA シグナルを持たず、52 から 356bp の 3' transduction 配列を持つ。最後のクラス 3 は 7 から 26bp の 3' transduction 配列を持つ 7 配列が含まれているが、上述の APC 遺伝子中の場合と同様に塩基置換の多い polyA tail を持っているだけとも考えられる。クラス 1 とクラス 2 を含めた L1 の 14 配列のうち、6 つでは、これ以外の位置に良く似た配列が見つかった。彼らは、3' transduction は少なくとも 10% 程度（クラス I のみ）で、14（クラス I と II）-22%（クラス I-III）まで増える可能性がある」と記述している。

彼らは 5' 側の切れている L1 についても試みに 25 配列を選び、同様の解析を行った。内 15 個が TSD を持っており、更にその内 4 個は 3' transduction していると想定されたので、おそらく 5' 側の欠失している L1 でも全長のものと同程度の 3' transduction があるだろうと考えられる。

2001 年にはヒトゲノムのドラフト配列が報告された（Lander et al. 2001）。この論文でも L1 について触れられている。ドラフトゲノムから見つかった 71 個の全長の L1 の内 21% では、3' 側の TSD の前に余分な配列が認められた。長さは 30 から 970bp であった。これ以降の論文は、ドラフトゲノムの情報を利用したものになる。Myers らの L1 の Ta サブファミリーの論文では、26%（何の 26% なのかは論文からは読み取れない）は 3' transduction と考えられるような複雑な配列を 3' 末端に持っている」と記されている（Myers et al. 2002）。

Piskeral らの解析を発展させた論文が Szak らによって 2002 年と 2003 年に報告されている。2002 年の論文では、Piskeral らの解析をプログラム化した perl スクリプト TSDfinder.pl を発表し、これまでにない大量の L1 の挿入についての解析を行っている（Szak et al. 2002）。TSD の最短は 9bp とし、独自のスコアリングをして、不完全な配列一致も考慮されている。L1 の下流 3,000bp から 3' 側の TSD を探索する点は Piskeral らの解析と同じだが、上流側は 5' 側が欠失した場合を含めるために上流 100bp から TSD を探索している。このように条件を緩めた結果か、TSD の長さの分布を調べたところ、ピークが 9bp 未満と 15-16bp の 2 つとなってしまう。Szak ら自身、短い側のピークは false positive が多いためであろうと記述している。TSD の長さの制約を緩めて短いものを許可した場合には、偶然配列が一致する場合が多く含まれることになるのは必然であり、これまでの論文で報告されている 3' transduction にも false positive が含まれている可能性は否定できない。彼らの解析では、72,148 個の L1 の挿入が見つかり、内 10,088 個では 3' transduction の無い挿入が起こっており、6,178 個では 3'

transduction が起こっている。単純計算すると、3' transduction の頻度は 38% となり、これまでのものよりも遥かに大きくなる。しかしながら、L1 の 5' 側と同じ配列が L1 の下流数十塩基中から見つかる場合が通常の挿入で、下流側 3kb から見つかる場合が 3' transduction とされるのだから、短い配列の偶然の一致 (false positive) は 3' transduction とされるものにより多く含まれているはずである。Szak ら自身この割合は segmental duplication によるもので実数は遥かに低いらしいと記しているが、この記述の根拠は記述されていない。

厳密な 3' transduction がどれくらい起こっているのかについて上記の論文の結果を基に詳細に解析を行った結果が翌年に報告されている (Szak et al. 2003)。Szak らはまず、上記の 6,178 個の 3' transduction の可能性がある配列について、transduction の配列と似た配列が他の L1 の挿入の下流 3,000bp にあるかどうかを調べ、2,616 個を得た。更にそこから、以下の 6 種類の場合を取り除いた。(a) 複数の L1 の挿入が近くにあるため下流の配列が一致している、(b) 配列の相同性が 90% 以下、(c) 似た配列の長さが 3' transduction の配列の 3 割以下、(d) L1 に対して向きが逆、(e) 似た配列が L1 の 3' 末端から 20bp 以上離れている、(f) segmental duplication の一部として配列が重複している。以上のように厳密に 3' transduction を絞り込んでいった結果、最終的に 27 個が確実に 3' transduction と言えるものであった。これらを *bona fide* L1-TDs (L1 with transduction-derived sequence) と表現している。

bona fide L1-TDs は配列の類似性により L1-TD1 から L1-TD25 までの 25 のファミリーに分けられた。TSD が見つからない挿入などを含めるとそれぞれのファミリー内のコピー数の平均は 2.5 で、他のコピーの「親」になりうる全長の L1 が見つかるファミリーは 10 個だけだった。3' transduction の長さはほとんどが 500bp 未満で平均は 290bp、最長は 2,883bp であった。明確に 3' transduction と言える L1 の挿入の見つかる量がこれほど少ないことは驚きである。その原因の一つとして、L1-TDs が若いエレメントであり、ヒト内で固定されていないことが考えられる。3' transduction の絞り込みに複数の L1-TDs があることが利用されているため、「親」が失われていたり、転移頻度が低かったりすると見つかってこない。おそらくここで見つかって来た 3' transduction の数は最小値で、実際にはもっと多くの 3' transduction が起こっていると考えられる。だとしても、TSD だけに依存した 3' transduction の割合の算出が false positive という危険を多く含んでいることを認識しておかねばならない。

論文に報告されたごく最近の L1 の転移を調べることで頻度を計算している例もある。Chen らのまとめによると、ヒトの病気の原因として見つかった 13 個の L1 の挿入のうち、3 つ (23%) は 3' 側に余分な配列を伴って転移している (Chen et al. 2005)。この 3 つを示した論文は上述の、Holmes et al. 1994、Brouha et al. 2003、

van den Hurk et al. 2003 である。3' transduction の長さは順に、489bp、153bp、414bp であり、上述のように前 2 者では「親」の L1 (LRE2、LRE3) が見つかったている。

まとめると、L1 の 3' transduction の割合は 20% 台という報告が多いが、false positive の可能性を考慮すると実際にはこれよりも低いと考えるのが妥当であろう。逆に Szak らの解析では確実な 3' transduction を見つけることに焦点が当てられているので割合は過少評価になってしまう。より正確な割合の計測が要求される。

L1 の 3' transduction の割合の算出は、L1 がゲノム進化に果たしている役割を評価する上で重要である。3' transduction された配列中に遺伝子のエクソンが含まれていれば別の遺伝子中にエクソンを挿入するという役割が期待される。しかし今のところ、L1 によって 3' transduction された配列が遺伝子として機能しているという報告は無い。CFTR 遺伝子のエクソン 9 を転移させた例はあるが、挿入された先では機能を持っていない (Rozmahel et al. 1997)。他には、3' transduction か trans mobilization かわからない例として ATM 遺伝子のエクソン 30 を含む領域の転移が報告されている (Ejima et al. 2003)。この例では、エクソン 30 と周辺のイントロン配列が転移しているので、スプライシングされた mRNA が転移する通常の processed (pseudo)gene 形成とは異なっている。15 塩基の TSD や 5' 側の逆位など L1 が転移させた特徴は認められるが、L1 の配列そのものが見られないため、2 通りの解釈が成り立つ。一つは、L1 の 3' transduction が起こったが、L1 まで逆転写されなかった。もう一つは、ATM 遺伝子の antisense RNA が転写されたものが L1 の機構に依存して転移した。いずれにせよ、この場合にも転移後の ATM エクソン配列が機能している証拠は無い。3' transduction によってエクソンが本当に転移するののかも解決しなければならない課題と言えるだろう。

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW,

Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglu S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ; International Human Genome Sequencing Consortium.

Initial sequencing and analysis of the human genome.

Nature. 2001 Feb 15;409(6822):860-921. Erratum in: Nature 2001 Aug 2;412(6846):565. Nature 2001 Jun 7;411(6838):720. Szustakowski, J

Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH Jr.

A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion.

Nat Genet. 1994 Jun;7(2):143-8.

McNaughton JC, Hughes G, Jones WA, Stockwell PA, Klamut HJ, Petersen GB.

The evolution of an intron: analysis of a long, deletion-prone intron in the human dystrophin gene.

Genomics. 1997 Mar 1;40(2):294-304.

Rozmahel R, Heng HH, Duncan AM, Shi XM, Rommens JM, Tsui LC.

Amplification of CFTR exon 9 sequences to multiple locations in the human genome.

Genomics. 1997 Nov 1;45(3):554-61.

Brouha B, Meischl C, Ostertag E, de Boer M, Zhang Y, Neijens H, Roos D, Kazazian HH Jr.

Evidence consistent with human L1 retrotransposition in maternal meiosis I.

Am J Hum Genet. 2002 Aug;71(2):327-36.

van den Hurk JA, van de Pol DJ, Wissinger B, van Driel MA, Hoefsloot LH, de Wijs IJ, van den Born LI, Heckenlively JR, Brunner HG, Zrenner E, Ropers HH, Cremers FP.

Novel types of mutation in the choroideremia (CHM) gene: a full-length L1 insertion and an intronic mutation activating a cryptic exon.

Hum Genet. 2003 Aug;113(3):268-75.

Chen JM, Stenson PD, Cooper DN, Ferec C.

A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease.

Hum Genet. 2005 Sep;117(5):411-27. Review.

Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr.

High frequency retrotransposition in cultured mammalian cells.

Cell. 1996 Nov 29;87(5):917-27.

Moran JV, DeBerardinis RJ, Kazazian HH Jr.

Exon shuffling by L1 retrotransposition.

Science. 1999 Mar 5;283(5407):1530-4.

Goodier JL, Ostertag EM, Kazazian HH Jr.

Transduction of 3'-flanking sequences is common in L1 retrotransposition.

Hum Mol Genet. 2000 Mar 1;9(4):653-7.

Pickeral OK, Makalowski W, Boguski MS, Boeke JD.

Frequent human genomic DNA transduction driven by LINE-1 retrotransposition.

Genome Res. 2000 Apr;10(4):411-5.

Myers JS, Vincent BJ, Udall H, Watkins WS, Morrish TA, Kilroy GE, Swergold GD, Henke J, Henke L, Moran JV, Jorde LB, Batzer MA.

A comprehensive analysis of recently integrated human Ta L1 elements.

Am J Hum Genet. 2002 Aug;71(2):312-26.

Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D, Boeke JD.

Molecular archeology of L1 insertions in the human genome.

Genome Biol. 2002 Sep 19;3(10):research0052.

Szak ST, Pickeral OK, Landsman D, Boeke JD.

Identifying related L1 retrotransposons by analyzing 3' transduced sequences.

Genome Biol. 2003;4(5):R30.

Ejima Y, Yang L.

Trans mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling.

Hum Mol Genet. 2003 Jun 1;12(11):1321-8.

2007/10/16

小島 健司 著

禁 無断複写転載