

論文 / 著書情報
Article / Book Information

題目(和文)	転移因子LINE の転移・増幅における5´連結機構の解明
Title(English)	
著者(和文)	山口勝己
Author(English)	Katsumi Yamaguchi
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9928号, 授与年月日:2015年6月30日, 学位の種別:課程博士, 審査員:梶川 正樹,岩崎 博史,木村 宏,本郷 裕一,相澤 康則,岡田 典弘
Citation(English)	Degree:., Conferring organization: Tokyo Institute of Technology, Report number:甲第9928号, Conferred date:2015/6/30, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

平成 26 年度 博士論文

転移因子 LINE の転移・増幅における 5' 連結機構の解明

東京工業大学大学院

生命理工学研究科

生体システム専攻

山口 勝己

指導教官 梶川 正樹 講師

目次

1. 第一章 序論.....	1
1-1. 転移因子 (Transposable elements)	1
1-2. LINE が宿主に及ぼす影響.....	1
1-3. LINE の構造.....	3
1-4. LINE の転移機構.....	4
1-5. LINE の分類.....	5
1-6. LINE 転移に関与する宿主 DNA 修復タンパク質.....	6
1-7. 標的ゲノム DNA の変化.....	7
1-8. LINE 配列の 5' 末端領域構造① : 5' 欠失配列.....	8
1-9. LINE 配列の 5' 末端領域構造② : 5' 逆位配列.....	9
1-10. LINE の 5' 末端連結部.....	9
1-11. 研究の目的と概要.....	10
2. 第二章 材料と方法.....	12
2-1. LINE 転移検出用プラスミド DNA.....	12
2-2. 培養細胞.....	12
2-3. LINE 転移検出系.....	13
2-4. LINE 転移実験.....	14
2-5. LINE 転移後配列の回収.....	15
2-6. DT40 細胞中で転移させた Zfl2-2 転移後配列の単離.....	16
2-7. ゲノムデータベースを用いた LINE 配列の検索.....	16
3. 第三章 結果.....	18
3-1. 5' EX 生成には宿主因子が関与する.....	18
3-2. 5' EX 生成と NHEJ 経路の遺伝子群.....	18
3-3. 5' EX 生成と LINE 種.....	19

3-4.	5´ EX 生成にはゲノム DNA の 3´ 突出末端が必要である.....	19
3-5.	5´ EX は、その隣接配列に由来する.....	20
3-6.	5´ EX 生成機構の一般性.....	20
3-7.	5´ EX の生成機構.....	21
3-8.	5´ 逆位配列は、ヒト細胞中で転移した.....	22
3-9.	L1 の 5´ 逆位配列は、10~20bp の TSD を伴う.....	23
3-10.	L1 の TSA 分布パターンはヒト培養細胞の違いで異なる.....	24
3-11.	L1 転移におけるゲノム DNA の切断位置.....	24
3-12.	ZfL2-1 転移におけるゲノム DNA の切断位置.....	25
3-13.	LINE の TSA パターンと挿入配列.....	26
3-14.	5´ 末端連結の 2 つの経路 ; アニーリング連結とダイレクト連結.....	26
3-15.	LINE 挿入位置ゲノム DNA の解離しやすさの観察①.....	28
3-16.	LINE 挿入位置ゲノム DNA の解離しやすさの間接的な観察②.....	29
3-17.	ゼブラフィッシュ LINE Nimb-2_DR は 2 つの 5´ 末端連結経路を介して転移する.....	29
4.	第四章 考察.....	32
4-1.	アニーリングを介する 5´ 連結機構.....	32
4-2.	LINE 転移時における 2 つの 5´ 末端連結経路.....	33
4-3.	LINE 転移時にゲノム DNA へ働く宿主因子.....	34
4-4.	5´ 逆位配列生成機構.....	35
4-5.	5´ 末端連結機構の一般性.....	35
	謝辞.....	37
	参考文献.....	38
	図.....	44
	表.....	76
	参考資料.....	85

1. 第一章 序論

1-1. 転移因子 (Transposable elements)

転移因子は、ほぼ全ての真核生物ゲノム中に存在する散在性反復配列である。転移因子は大きく 2 つのグループに分けることができる。1 つは DNA トランスポゾンである (Kleckner, 1990) (図 1)。DNA トランスポゾンは自身の配列を切り出し別のゲノム部位へ挿入する。2 つ目のグループはレトロトランスポゾンである。レトロトランスポゾンは、自身の RNA を逆転写することで新たな cDNA を作り出し、この cDNA コピーをゲノムの別の位置へ挿入する。レトロトランスポゾンは、長い末端反復配列 (Long terminal repeat; LTR) を持つグループ (LTR レトロトランスポゾン) (Garfinkel et al., 1985; Boeke et al., 1985; Eichinger et al., 1988) と LTR を持たないグループ (non-LTR レトロトランスポゾン) (Luan et al., 1993) に分類される。これら 2 つのグループのレトロトランスポゾンの間には、自身のコピー配列をゲノムへ挿入する過程に違いがある。LTR レトロトランスポゾンはウイルス様粒子内で自身の RNA の逆転写を行い、その後、逆転写で生じた自身のコピー配列を標的ゲノム位置へ挿入する (図 1)。一方、non-LTR レトロトランスポゾンは、自身のコピー配列を挿入するゲノム位置で自身の RNA の逆転写を行う (図 1)。non-LTR レトロトランスポゾンには、Long interspersed element (LINE)、Short interspersed element (SINE) (Kajikawa and Okada, 2002; Dewannieux et al., 2003) やレトロ偽遺伝子 (Wei et al., 2001; Esnault et al., 2000) などが含まれる。このように、転移因子には様々な種類が存在する。本研究では、これら転移因子の中でも転移・増幅機構の研究があまり進んでいない LINE に着目して研究を行った。

1-2. LINE が宿主に及ぼす影響

LINE は真核生物ゲノム中に数千から数万コピー存在する転移因子である。ヒトゲノム中では、タンパク質をコードしているエクソン領域は 1~2% であるのに対して、LINE はヒトゲノム全体の約 21% (85 万コピー) を占めることが明らかにされている (図 2; Lander et al.,

2001)。この莫大なコピー数の LINE 配列は、単なるゲノムの構成要素としてのみならず、宿主ゲノムにさまざまな変化を引き起こし、ゲノムの成り立ちやその進化に大きな影響を及ぼしてきたと考えられている (Ostertag and Kazazian, 2001; Symer et al., 2002; Gilbert et al., 2002)。

LINE 転移は、宿主から見ればゲノムへの挿入変異である。LINE は言わば、内在性の変異原であり、宿主生物ゲノムに挿入変異を引き起こすのである。また、単なる挿入変異のみならず、異なる遺伝子座の LINE 配列間に組み換えが起こることで、組み換えを起こす LINE 間の宿主ゲノム配列の、重複、欠失、あるいは逆位などの構造変化が起こることが知られている (Kazazian, 2004)。更に、LINE を含む転移因子は、宿主ゲノムから転写された RNA のスプライシングパターンを変化させたり、ポリ A 付加部位を変化させたりすることで、宿主生物のトランスクリプトームの多様性にも大きな影響を及ぼしていることが示されている (Moran et al., 1999; Goodier et al., 2000)。このように、LINE を含む転移因子の存在は、宿主の遺伝情報に大きな多様性をもたらす、生物進化に大きく寄与していると考えられている。

LINE は内在性の変異原であり、その転移は宿主生物ゲノムの必須機能を破壊する恐れがある。事実、LINE 転移により引き起こされたさまざまな遺伝病が報告されている (Hancks et al., 2012; Kaer et al., 2013)。しかし、一方で、ヒトなどの宿主生物のゲノム中には、多量の LINE 配列が蓄積している。これは、LINE 配列が生殖系列細胞で転移し、その新規転移配列が次世代の生物ゲノムに引き継がれ蓄積したものである。LINE を含む転移因子は、自身のコピー数を増幅させつつ (=宿主生物ゲノムに多様性をもたらす)、宿主生物の生存を脅かすような過剰な転移を避けて、宿主生物とともに存在してきたのであろう。

これまで、LINE の転移・増幅は、宿主生物の生殖系列細胞では‘適度’に起こり、体細胞では高度に抑制されていると考えられてきた。なぜなら、体細胞での LINE 転移は、宿主生物ゲノムの必須情報を破壊する恐れがあるが、その新規転移配列は次世代の生物個体へは引き継がれず、LINE のコピー数増加にはつながらない。すなわち、転移因子 LINE にも宿

主生物にもどちらにも利点がないと考えられるのである。しかし、近年、いくつかの体細胞種で LINE の新規転移が観察されている。そのひとつが、大腸がんなどのある種のがん細胞である (Iskow et al., 2010; Lee et al., 2012)。正常細胞のがん化は遺伝情報の変異により引き起こされるが、LINE 転移が、がん化の主要因なのか、がんの悪性化に寄与するのか、あるいは、がん細胞の成り立ちには寄与せず細胞のがん化の結果 LINE 転移の抑制機構が機能しなくなり起こるのか、現在統一的な見解は得られていない (Iskow et al., 2010; Lee et al., 2012)。体細胞転移のもうひとつの興味深い例は、ラットやマウス、ヒトなどの哺乳類生物の脳神経細胞における LINE 転移である。正常な神経細胞の分化過程で L1 転移の活性化が起きることが示されている (Coufal et al., 2009; Muotri et al., 2005; Muotri et al., 2010)。これらの報告は、神経細胞分化時の LINE 転移が神経系の複雑さや個体内の遺伝的多様性を生み出す可能性を示唆する。この神経細胞での LINE 転移が、宿主生物の発生・生存において機能を持つか否か、非常に興味深い問題である。

1-3. LINE の構造

LINE は 5' 非翻訳領域 (5' UTR)、タンパク質コード領域 (ORF)、3' 非翻訳領域 (5' UTR) から構成されている。5' UTR には RNA ポリメラーゼ II 依存内部プロモーター配列が存在する (Mizrokhi et al., 1998)。この配列は LINE 転移の開始反応である LINE RNA の転写に必要である (Swergold et al., 1990)。3' UTR の 3' 末端にある polyA 配列または AT に富んだ繰り返し配列は自身 RNA の逆転写に必要である (Luan and Eickbush, 1995; Kajikawa et al., 2002)。LINE 配列の内部には、1 つまたは 2 つの ORF が存在する。ORF1 タンパク質 (ORF1p) は核酸結合活性と核酸アニーリング活性を持つことが知られている (REF)。ORF1p は、これらの活性により、自身の mRNA と結合することや、LINE mRNA・LINE タンパク質複合体 (LINE RNP) 形成に関与することが示されている (Hohjoh, 1996; Matsumoto et al., 2006; Kolosha, 1997)。ORF2 タンパク質 (ORF2p) は、エンドヌクレアーゼ (EN) 活性と逆転写酵素 (RT) 活性を持つ (Feng et al., 1996)。これらの活性は、それぞれ、標的ゲノ

ム DNA の切断と自身 RNA の逆転写を行う（次項説明参照）。ORF を 1 つ持つ LINE の ORF は ORF を 2 つ持つ LINE の ORF2 に相当する。

1-4. LINE の転移機構

LINE 転移は、LINE が自身の 5' UTR に存在する RNA ポリメラーゼ II 依存内部プロモーターから RNA ポリメラーゼ II によって LINE mRNA へ転写されることから始まる（図 5）。転写された LINE mRNA は核外へ輸送され、LINE タンパク質が翻訳される。LINE タンパク質は LINE mRNA と LINE RNP を形成し、核内へ輸送される。この LINE RNP は、LINE タンパク質以外に様々な宿主タンパク質を伴っている。これらの宿主タンパク質のいくつかは LINE 転移に関与することが示唆されている (Dai et al., 2012; Goodier et al., 2013; Taylor et al., 2013)。核内へ移行した LINE RNP の EN 活性は、標的となるゲノム DNA の一本鎖を切断する。この切断によって露出したゲノム DNA の 3' 水酸基をプライマーとして、LINE タンパク質の RT 活性により LINE mRNA を鋳型に逆転写が開始され、LINE cDNA が合成される。一本鎖目のゲノム DNA 切断で生じた 3' 水酸基から始まる自身の mRNA の逆転写反応は Target Primed Reverse Transcription (TPRT) と呼ばれる (Luan et al., 1993, Cost et al., 2002)。この TPRT によって標的ゲノム DNA と LINE の 3' 末端が連結される。多くの LINE は自身の RNA の 3' 末端の特異的な配列が TPRT の開始に必要なであるが、哺乳類の L1 クレードの LINE では、この特異的な配列は必要ではないと考えられている (Okada et al., 1997) (クレードは次項で説明)。それぞれ、前者はストリンジェント、後者はリラックスドと呼ばれている (Okada et al., 1997)。TPRT 後、二本鎖目の切断、LINE cDNA 相補鎖の合成、新規に合成された LINE の 5' 末端とゲノム DNA 切断末端の連結を経て、LINE 転移が完了すると考えられている。しかし、TPRT 後のこれらの過程がどのように進行するのか明らかにされていない。本研究では、特に、TPRT 後に起こる、LINE 新規挿入配列の 5' 末端とゲノム DNA 末端との連結機構の解明を目的とした。

1-5. LINE の分類

LINE は、自身のコードする RT ドメインのアミノ酸配列の類似性を基にした系統解析から 5 つのグループに分けることができる (図 3; R2、RandI、L1、RTE、I、Jockey; Malik et al., 1999; Lovsin et al., 2001; Eickbush and Malik, 2002; Kapitonov et al., 2009)。これらのグループはさらにクレードと呼ばれる 28 の小グループに分類される; R2 グループ (CRE クレード、R4 クレード、Hero クレード、NeSL クレード、R2 クレード)、RandI グループ (RandI/Dualen クレード)、L1 グループ (Proto1 クレード、L1 クレード、Tx1 クレード)、RTE グループ (RTETP クレード、Proto2 クレード、RTEX クレード、RTE クレード)、I グループ (Outcast クレード、Ingi クレード、I、Nimb クレード、Tad1 クレード、Loa クレード、R1 クレード)、Jockey グループ (Jockey クレード、Rex1 クレード、CR1 クレード、L2 クレード、L2A クレード、L2B クレード、Daphne クレード、Crack クレード)。R2 グループは系統的に最も古い推定される LINE グループである。このグループの LINE は全て 1 つの ORF を持つ。これらの ORF には、挿入位置の切断に必要な制限酵素様エンドヌクレアーゼ (RLE) ドメインが RT ドメインより C 末端側に位置している。RLE は、特異的な DNA 配列を認識し切断する酵素活性がある。そのため、RLE を持つ LINE はリボソーム RNA、テロメア、マイクロサテライトなど特異的挿入位置に転移する (Kojima and Fujiwara, 2005a)。一方、L1、RTE、I グループは、系統的に比較的新しい LINE グループである。古い LINE グループと新しい LINE グループには 2 つの大きな違いがある。そのひとつは ORF1 の有無であり、もうひとつは挿入位置を切断するヌクレアーゼの違いである (図 3)。新しいグループの LINE には、アプリン/アピリミジンエンドヌクレアーゼ (APE) ドメインが RT ドメインより N 末端側に位置している。APE は、脱塩基 (AP) 部位の DNA を特異的に切断する DNA 修復酵素である。LINE が持つ APE は、EN 活性を持つが AP 部位を認識する特異性がないことが示されている (Feng et al., 1996)。そのため、APE を持つ LINE の多くは、自身配列のゲノム DNA への挿入位置に配列特異性を持たず、ゲノムの様々な位置へ転移・増幅する。APE をもつ LINE のほとんどは、2 つの ORF をコードしているが、RandI グループ、RTETP ク

レード、RTE クレードの LINE 及び、L2 クレードのいくつかの LINE は ORF を 1 つのみコードしている。1 つの ORF が APE ドメインと RLE ドメインに挟まれている RandI グループ (RandI/Dualen クレード) は古いグループと新しいグループの間のように見える。RandI グループは、進化の過程で新しく APE を獲得し RLE を破棄したと示唆されている (Kojima and Fujiwara, 2005b)。

本研究では、新しい LINE グループの LINE 転移における 5´ 連結機構の解明を目指し、ヒトゲノム内に存在し L1 グループ (L1 クレード) に属する L1 (図 4B)、ゼブラフィッシュゲノム内に存在し、Jockey グループ (L2 クレード) に属する ZfL2-2 (図 4A) と ZfL2-1 (図 4C)、ゼブラフィッシュゲノム内に存在し、I グループ (Nimb クレード) に属する Nimb-2_DR (図 4D) を研究対象とした。

1-6. LINE 転移に関与する宿主 DNA 修復タンパク質

序論 1-4. で述べたように、現在提案されている LINE 転移モデルによると (図 5)、LINE 挿入には標的ゲノム部位の 2 本の DNA 鎖が両方共切断される必要がある。そのため、LINE 転移時に DNA 二本鎖切断 (Double strand break; DSB) が起こると推測されている。実際に、哺乳類培養細胞中でヒト L1 を過剰発現させると、ゲノム DNA の DSB が誘導されるという報告がある (Gasior et al., 2006)。一般的にゲノム DNA が二本鎖切断されると、宿主の DNA 修復系が速やかに DSB を修復する (Misteli and Soutoglou, 2009)。LINE 転移時に生じるであろう DSB の連結にも宿主の DNA 修復因子が関与している可能性がある。これまでに、いくつかの DNA 修復に関わる宿主タンパク質が培養細胞中での LINE 転移に関与することが示されている。例えば、Suzuki ら (2009) は、脊椎動物細胞の二本鎖 DNA 切断修復経路の 1 つである非相同末端連結 (Non-homologous end-joining; NHEJ) 経路に関わるタンパク質群が、効率的な LINE 転移に必要であることを報告している (Suzuki et al., 2009)。DSB 応答に関与するタンパク質である ataxia telangiectasia mutated (ATM) タンパク質の場合、LINE 転移に作用する 2 つの相反する報告がされている (Gasior et al., 2006; Coufal et al.,

2011)。一方は、ATM タンパク質が哺乳類培養細胞の L1 転移に必要であることを示唆する報告である (Gasior et al., 2006)。もう一方は、ATM タンパク質がヒト神経前駆細胞中の L1 転移を抑制することを示唆する報告である (Coufal et al., 2011)。これらの報告は異なる細胞種間で ATM タンパク質が LINE 転移に異なる働きをする可能性を示している。以上のように、様々な宿主 DNA 修復タンパク質の LINE 転移への関与は、LINE 転移時に標的ゲノム DNA 末端と LINE の 5' 末端が宿主 DNA 修復タンパク質に DNA 切断部位と認識され、連結されるだろうことを示唆する。

1-7. 標的ゲノム DNA の変化

LINE 挿入配列の配列的特徴は、LINE の 5' 末端連結機構を反映していると考えられる。LINE 挿入配列とゲノム DNA の連結部には、標的ゲノム DNA の変化 (Target site alteration; TSA) が見られる。TSA は、LINE 転移時に LINE 挿入位置のゲノム DNA 配列が重複や欠失をすることで生成され则认为られている。Gilbert ら (2002) は、彼らの LINE の 5' 末端連結モデル (図 6) 中で、LINE 転移時に起こる二本鎖目の切断の位置の違いが様々な TSA を生み出すことを提案している (図 6)。このモデルによれば、ゲノム DNA の二本鎖目の切断が一本鎖目の切断位置より下流で起きると、3' 突出ゲノム DNA 末端が生じる。このゲノム DNA 末端と LINE の 5' 末端が連結すると、転移した LINE の両端に標的ゲノム DNA の重複 (Target site duplication; TSD) が生じる (図 6 下)。一方、ゲノム DNA の二本鎖目の切断が一本鎖目の切断位置より上流で起きると、3' 突出ゲノム DNA 末端が生じる。このゲノム DNA 末端と LINE の 5' 末端が連結すると標的ゲノム DNA の欠失 (Target site truncation; TST) が生じる (図 6 下)。また、ゲノム DNA の二本鎖目の切断が一本鎖目の切断位置と同じ位置で起きると、平滑末端が生じる。このゲノム DNA 末端と LINE の 5' 末端が連結するとゲノム DNA の重複や欠失の生じない LINE 転移 (Blunt-end joining; BEJ) が起きる (図 6 下)。以上のように、生じたゲノム DNA 末端の違いにより、TSD、TST、BEJ のどのタイプの TSA が生じるかを決定していると考えられる。

バイオインフォマティクスを用いた研究によって、様々な生物種に内在する様々な LINE の TSA が解析されている (Ichiyangi et al., 2007; Ichiyangi and Okada, 2008; Kojima, 2010)。これらの解析により、宿主に非依存的で LINE クレードに特異的な TSD 長のピークが観察されている。例えば、L1 クレード LINE の場合、TSD 長のピークは 13-15bp の長さであった (ヒト、ウシ、オポッサム、ゼブラフィッシュ)。L2 クレード LINE の場合、TSD 長のピークは 3-5bp の長さであった (オポッサム、ゼブラフィッシュ)。CR1 クレード LINE の場合、L2 クレード LINE と非常に類似した TSD 長のピークであった (ニワトリ、ゼブラフィッシュ)。RTE クレード LINE の場合、TSD 長のピークは 10-12bp であった (ウシ、オポッサム、ゼブラフィッシュ)。これらの観察から、TSD の長さは LINE の種類によって決定づけられると示唆されている (Ichiyangi and Okada, 2008)。Gilbert ら (2002) の 5' 末端連結モデル (図 6) から考えると、この報告で観察された LINE 転移標的ゲノム DNA 部位の一本鎖目の切断位置からの二本鎖目の切断位置までの塩基数は、LINE によって固有の長さに特定されるように見える。

1-8. LINE 配列の 5' 末端領域構造① : 5' 欠失配列

転移した LINE の 5' 領域の多くは欠失 (5' 欠失) し、全長配列を持つ LINE は稀である (図)。新たに転移した LINE 配列が転移活性を保持するためには、LINE は全長配列を持つ必要がある。なぜなら、LINE の 5' 末端領域には自身配列の転写に必要な内部プロモーター領域が存在するからである (序論 1-3; 図 7A)。従って、5' 欠失 LINE は新たな LINE 転移を起こすことができない。5' 欠失 LINE の生成機構はよく分かっていないが、3' 末端から開始される逆転写反応が何らかの理由で停止してしまうためだと考えられている。以前、私たちは、NHEJ 経路に関わる Ku70 遺伝子の欠損 DT40 細胞中で、ゼブラフィッシュ LINE、ZfL2-2 を転移させた (Suzuki et al., 2009)。この転移後配列の解析から、Ku70 タンパク質欠損細胞で転移した ZfL2-2 は、野生型細胞で転移した ZfL2-2 よりも統計学的有意に多くの全長配列を含んでいた。この結果は、DNA 末端結合タンパク質である Ku70 が LINE の 5' 欠失

配列生成に関与していることを示唆する。Ku70 タンパク質は、LINE RT が逆転写中の LINE cDNA 末端と相互作用し、逆転写反応の進行を妨げることで 5′ 欠失生じさせているのかもしれない。また、Coufal ら (2011) は、DNA 損傷応答に関与する ATM タンパク質の欠損細胞中で転移した L1 は、正常な細胞中で転移した L1 より長い配列長であったことを報告した (Coufal et al., 2011)。これらの結果は、DNA 修復に関わるいくつかのタンパク質が LINE の転移後配列長に関与することを示唆する。以上の宿主タンパク質と LINE の転移後配列長の研究は、LINE 転移中 (逆転写反応中) の転移中間体が宿主 DNA 損傷修復因子に DNA 損傷として認識され LINE の 5′ 末端連結が起こるため、短い LINE 転移後配列、すなわち、5′ 欠失配列が生じることを示しているのかもしれない。

1-9. LINE 配列の 5′ 末端領域構造② : 5′ 逆位配列

LINE 5′ 領域の逆位配列 (5′ 逆位) 構造は、5′ 欠失以外の LINE 転移不可な LINE 構造として知られている (Ostertag and Kazazian, 2001) (図 7A)。この 5′ 逆位の生成機構モデルは、Ostertag ら (2001) のヒトゲノム中に存在するヒト L1 の解析から提案されている (図 8)。5′ 逆位は LINE 転移する標的ゲノム DNA 部位の第一鎖と第二鎖切断によって生じた 2 つのゲノム DNA 末端から LINE RNA の逆転写が開始されるツインプライミングによって生じると考えられている (Ostertag and Kazazian, 2001)。ツインプライミングは、次の (1) から (4) の段階で進行する (図 8)。(1) 第一鎖目の切断位置から TPRT が起こる。(2) 全ての逆転写が完了する前に第二鎖が切断される。(3) 生じたゲノム 3′ 突出末端をプライマーにして LINE mRNA の内部から 2 回目の逆転写反応が起きる。(4) 1 回目と 2 回目の逆転写で生じた 2 つの cDNA 末端が連結し、LINE の 5′ 逆位配列の転移が完了する。このツインプライミングモデルから、L1 転移時に 2 本鎖が切断されて生じた 2 つのゲノム DNA 末端は、2 回の L1 mRNA の逆転写を介して連結されると考えられる。

1-10. LINE の 5′ 末端連結部

LINE 転移後配列とゲノム DNA の 5′ 末端連結部は、次の 3 タイプに分類される (図 9)。

- ①、数塩基の LINE 配列由来かゲノム DNA 由来か区別できない配列(マイクロホモロジー; MH)。
- ②、LINE 配列末端とゲノム DNA 末端の平滑な連結(Direct joining; DJ)。
- ③、由来が不明の数~数十塩基を介した連結 (5′ エクストラヌクレオチド; 5′ EX)。これら 5′ 末端連結部配列の解析は、LINE 転移における 5′ 末端連結機構の解明の重要な研究であると考えられる。これまでに、バイオインフォマティクスを用いたいくつかの研究により、ヒトゲノム中に存在する L1 の MH と DJ が解析されている (Zingler et al., 2005; Kojima, 2010)。これらの解析から、5′ 欠失 L1 配列は頻繁に MH が観察され、一方、全長 L1 配列では DJ が多く観察されることが報告されている。これらの報告は L1 の 5′ 末端連結部に見られる特徴 (MH、DJ) と L1 の転移後配列の構造が相互に関連があることを示している。以上のことから、MH を必要としない全長 L1 配列の 5′ 末端連結と MH を必要とする 5′ 欠失 L1 配列の 5′ 末端連結の少なくとも 2 つの 5′ 末端連結機構が存在することが示唆されている (Zingler et al., 2005)。

ゼブラフィッシュゲノム中の LINE 配列解析により、およそ半分の ZfL2-2 配列が、5′ EX を伴っていることが報告されている (Ichiyangi et al., 2007; Ichiyangi and Okada, 2008)。一方で、ZfL2-2 を HeLa 細胞で転移させた場合、5′ EX は観察されないことが報告されている (Ichiyangi et al., 2007)。ゼブラフィッシュゲノムとヒト HeLa 細胞の異なる宿主間で、ZfL2-2 の 5′ EX の出現頻度が異なることは、宿主因子が 5′ EX の生成に関与している可能性を示唆する。しかしながら、5′ EX 生成にはどのような宿主因子が関与するのか、5′ EX がどのように生成されるのか明らかにされていない。

1-11. 研究の目的と概要

これまで LINE の転移・増幅機構に関する研究から、LINE のコードするタンパク質の機能が主に解明されてきた。その結果、LINE タンパク質による LINE の逆転写の開始反応 (TPRT) が実験的に証明され、この反応により、LINE DNA の 3′ 末端とゲノム DNA の 3′ 末端が連結

されることが示されている。一方で、LINE のコピーDNA の 5′ 末端とゲノム DNA 末端がどのように連結されるのか、その分子機構は明らかにされていない。上述の通り、LINE の 5′ 末端連結機構は、① LINE 転移に伴う標的ゲノム配列の変化 (TSA) や、② LINE 転移配列の 5′ 末端領域構造 (5′ 欠失、5′ 逆位)、あるいは、③ 5′ 末端連結配列の生成 (MH、DJ、5′ EX)、と密接に関係していると考えられる。これまで、LINE の 5′ 末端連結機構に関する研究は主に公開されたリファレンスゲノム配列の解析により行われてきた。しかし、リファレンスゲノムからは、LINE 転移後のゲノム情報のみ得られ、LINE 転移前のゲノム情報を得ることは困難である。一方で、培養細胞内で新規に転移させた LINE 転移後配列の解析の場合、LINE 転移前後のゲノム情報を得ることができるため、上記の記した LINE の 5′ 末端連結機構に関する配列情報を正確に解析することができる。この培養細胞内で新規に転移させた LINE 転移後配列解析方法は、ヒト細胞で転移させたヒト L1 の解析などいくつか報告されているのみである。これらの報告の中でも上述のニワトリ DT40 細胞で転移させたゼブラフィッシュ LINE の解析から、私はこれまで LINE 転移に関与する宿主因子を明らかにしてきた。そこで私は LINE の 5′ 末端連結の分子機構を解明するために、ニワトリ DT40 細胞とヒト HeLa 細胞中で新規に転移させた 300 以上の様々な LINE の転移後配列を決定・解析した。

2. 第二章 材料と方法

2-1. LINE 転移検出用プラスミド DNA

pHLEmH は、ヒト L1 (図 4B) 転移検出用プラスミドであり、米国ミシガン大学 John Moran 博士から譲り受けた pCEP4/L1.3*mneo*₄₀₀/ColE1 (Gilbert et al., 2002) の L1.3 の LINE タンパク質コード領域に存在する 3 つの HindIII サイトに非同義的に点変異を挿入したものである。L1.3 は、ヒトゲノム中のヒト L1 のあるローカスから得た配列である。

pAZ2-2 は、ゼブラフィッシュ LINE ZfL2-2 (図 4A) 転移検出用プラスミドであり、pCEP4 を骨格としゼブラフィッシュ LINE ZfL2-2 (ZL15) 配列とその 3' UTR 領域に *mneo*₄₀₀/ColE1 マーカーを含んでいる (Suzuki et al., 2009)。

TK109-17 は、ゼブラフィッシュ LINE ZfL2-1 (図 4C) 転移検出用プラスミドであり、pCEP4 を骨格とし ZfL2-1 とその 3' UTR 領域に *mneo*₄₀₀/ColE1 マーカーを含んでいる。また、ZfL2-1 ORF1 コード領域の 5' 末端に 3 × FLAG コード DNA が位置している (Sugano et al., 2006)。

Nb2A3-2 は、ゼブラフィッシュ LINE Nimb-2_DR (図 4D) 転移検出用プラスミドであり、pCEP4 を骨格としゼブラフィッシュ LINE Nimb-2_DR コンセンサス配列、その 3' UTR 領域に *mneo*₄₀₀/ColE1 マーカーを含んでいる。Nimb-2 DR コンセンサス配列は、ゼブラフィッシュゲノムデータベース (<http://genome.ucsc.edu>; the July 2010 zebrafish (Danio rerio) Zv9 assembly) より得た。この配列を参照して、ゼブラフィッシュゲノム DNA から PCR 反応によって、~150bp の 3' UTR (TCCGCCCTTTCTCTCCGTTTCATGTAATTTATACACTTCAGAAAGTATTCAGTGTCCATACTCCATTCCAGTCGGTGGCGGTAATGCACCTATAAGTCGGTTGCCAACCGCCAGTAAACCCAAAGAAGAAG AAGAAGAAGAAG) を除いた Nimb-2_DR 配列断片を得た。この断片を NotI と NheI で制限酵素処理をした AZ2-2 にクローニングした。このクローニングで得られたプラスミドを BamHI で制限酵素処理をした後、ゼブラフィッシュゲノム DNA の PCR 反応で除いた~150bp の 3' UTR をクローニングし、Nimb2A3-2 とした。

2-2. 培養細胞

DT40 野生細胞株は理研バイオリソースセンター (number RCB1464) より購入した。また、DT40 欠損細胞株である Ku70^{-/-} (Takata et al., 1998) は京都大学の武田先生より頂いた。全ての DT40 細胞株は、10%ウシ胎仔血清 (FBS : ニチレイ)、1%ニワトリ血清 (CS : インビトロジェン)、10 μ M β -メルカプトエタノール、20U/ml ペニシリン-20 μ g/ml ストレプトマイシン (Invitrogen) を含む RPMI medium 1640 (Invitrogen) を深型 10cm シャーレ (アズノールシャーレ : アズワン) で 5% CO₂ インキュベーター中で 37°Cまたは 33°C条件下で培養した。細胞株ごとに増殖速度が異なるため、1日または2日おきにそれぞれの増殖速度にあわせて5~20倍希釈して継代を行った。G-418耐性を獲得した細胞株は、G-418 sulfate solution (Promega) を添加した RPMI 培地 (1.6mg G-418/ml) を用いて培養した。

当研究室で継代しているヒト HeLa 細胞はオリジナルの HeLa 細胞と比較して性質が変化しているように見える。当研究室の HeLa 細胞を LINE の転移実験に使用した際、オリジナルの HeLa 細胞と比較して約10倍 LINE 転移頻度が高いことが明らかとなっている。オリジナルの HeLa 細胞と当研究室継代の HeLa 細胞を区別するため、当研究室の細胞株を HeLa-RC (retrotransposition-competent) 細胞とした (Kajikawa and Okada, 2002; Kajikawa et al., 2005; Sugano et al., 2006)。

HeLa-RC 細胞 (2×10^5 cells/well) は6ウェルディッシュ中で10%ウシ胎仔血清 (FBS : ニチレイ) を添加した L-グルタミン含有、ピルビン酸ナトリウム非含有の high-glucose Dulbecco's modified Eagle's medium (GIBCO) を用いて 5% CO₂ インキュベーター中、37°Cの条件下で培養した。G-418耐性を獲得した細胞株は、G-418 sulfate solution (Promega) を添加した RPMI 培地 (1.6mg G-418/ml) を用いて培養した。

2-3. LINE 転移検出系

LINE 転移検出に用いるプラスミドには、LINE の 3' UTR に、イントロンで分断された LINE 配列とは逆向きのネオマイシン耐性遺伝子が挿入されている (図 10)。LINE mRNA が逆転写されイントロンがスプライシングにより抜け、ゲノム DNA に挿入されるとネオマイシン耐

性遺伝子を発現する。すなわち、LINE が転移した細胞のみネオマイシン（または、カナマイシン、G418）耐性となる。

2-4. LINE 転移実験

ニワトリ DT40 細胞 (Honda et al., 2007; Suzuki et al., 2009) とヒト HeLa-RC 細胞 (Sugano et al., 2006) 中での LINE 転移実験は、AZ2-2、pLEmH、TK109-17 及び Nb2A3-2 コンストラクトを用いて行った。これらの培養細胞中で転移したそれぞれの LINE 挿入配列の単離方法を次に述べる。

DT40 細胞へのプラスミド DNA 導入はエレクトロポレーション法を用いた。まず、200 μ l 培養液 (3.5×10^4 細胞個/ μ l) に 15 μ l の pEGFPFLAG-1 (1μ g/ μ l) と 15 μ l の pAZ2-2、pLEmH、TK109-17 あるいは Nb2A3-2 いずれかひとつの LINE 転移検出用プラスミド (1μ g/ μ l) を懸濁し、電極間距離 0.4cm キュベット (BIO-RAD) へ移して電気パルスをかけた。pEGFPFLAG-1 は緑色蛍光タンパク質発現プラスミドで、プラスミド DNA の導入効率測定に用いた(後述)。電気パルスは 200V (pLEmH には 250V)、960 μ F、抵抗 ∞ の条件下で Gene Pulser X cell (Bio Rad) を用いて行った。電気パルス直後に 10ml の培地に加えよく懸濁し、そのうちの 0.25ml 懸濁液を新しい 7.5ml の培地入りの 10cm シャーレに加え、5% CO₂、33°Cの条件下でインキュベートした。エレクトロポレーションから 72 \pm 6 時間後、LINE 転移検出のための G-418 薬剤選択は、 7×10^5 個の細胞を新しい 5ml の培地に懸濁し、45°Cにインキュベートした 5ml の G-418 薬剤選択用アガロース (Lonza) 入り RPMI 培地 (0.24%アガロース、30%FBS、3%CS、3.2mg/ml G-418) と懸濁し、5% CO₂、37°Cの条件下で 11 日間インキュベートした (最終濃度 : 0.12%アガロース、20% FBS、2% CS、1.6mg/ml G-418)。遺伝子導入を確認するために、アガロースゲル化と同時に FACSGalibur Flow Cytometry System (Becton-Dickinson) を用いて EGFPFLAG-1 から発現した緑色蛍光タンパク質の蛍光強度測定を行った (Honda et al., 2007)。11 日後に G-418 耐性コロニー (LINE 転移したコロニー) をピックアップし、それぞれのクローンを 1×10^6 細胞個以上に達するまで懸濁培養した。

HeLa-RC 細胞への遺伝子導入はトランスフェクション試薬を用いた。HeLa-RC 細胞 (2×10^5 細胞個/ウェル) は 6 ウェルディッシュ中で 5% CO₂ インキュベーター中、37°C の条件下で培養した。1 日後、これらの細胞に FuGENE6 Transfection Reagent (Roche) を用いて使用説明通りに pAZ2-2、pLEmH、あるいは TK109-17 (それぞれ 500ng/ウェル) のいずれかをトランスフェクションした。1 日後、pAZ2-2、pLEmH、TK109-17 をそれぞれトランスフェクションした細胞に対してハイグロマイシン (200 μ g/mL) を用いて 5 日間薬剤選択を行った。この薬剤選択によって、遺伝子導入された HeLa-RC 細胞のみが選択される。このハイグロマイシン耐性 (Hyg^R) 細胞は、トリプシン処理し、新しい 10cm ディッシュに移し、薬剤 G418 (400 μ g/mL) 入り培地で培養した。この薬剤選択によって、それぞれの LINE が転移したコロニーのみが生き残る (G-418 耐性; G418^R)。この培養から 10 日後、G418^R コロニーをピックアップし、それぞれのクローンを 1×10^6 細胞個以上に達するまで培養した。

2-5. LINE 転移後配列の回収

1×10^6 細胞個以上の G418^R DT40 細胞及び G418^R HeLa-RC 細胞のそれぞれのクローンから、GenElute mammalian genomic DNA miniprep kit (Sigma) を用いて DNA 抽出を行った。抽出したゲノム DNA (20 μ g/クローン) を 75 U の HindIII (TaKaRa) で 37°C、6 時間制限酵素処理を行った。制限酵素処理されたゲノム DNA (~20 μ g) を T4 DNA ligase (350U; TaKaRa) を用いて、500 μ L 反応液中で 16°C、一晩セルフライゲーション反応させた。環状化した DNA を Escherichia coli ElectroMAX DH10B Cells (Invitrogen) もしくは NEB 10-beta Electrocompetent E. coli (New England Biolabs) へ 2500 V, 25 mF and 100 V の条件下で GENE Pulser Xcell (Bio-Rad) を用いてエレクトロポレーション法で導入した。エレクトロポレーションした大腸菌株をすぐに SOC 培地 (600 μ L) 内で 1 時間培養し、その後カナマイシン (70 mg/ml) 培地へプレーティングした。この薬剤選択で、*mneo*/₄₀₀/ColE1 を持つ LINE 転移後配列 (ZfL2-2、L1.3、ZfL2-1 あるいは Nimb-2_DR) とその近傍ゲノム DNA (ニワトリゲノム DNA; DT40 細胞または、ヒトゲノム DNA; HeLa-RC 細胞) の環状 DNA を含む大腸菌の

みが生き残る (図 10)。これらのクローンから単離した LINE 転移後配列はそれぞれ LINE 特異的なプライマーを用いて LINE 挿入配列の 5' 末端連結部と 3' 末端連結部の DNA 配列を決定した。LINE 配列及びゲノム DNA との連結部配列は、それぞれの LINE のコンセンサス配列と、ニワトリゲノムデータベース (<http://genome.ucsc.edu>; the May 2006 chicken (*Gallus gallus*) WUGSC 2.1/galGal3 assembly) もしくはヒトゲノムデータベース (<http://genome.ucsc.edu>; the Human (*Homo sapiens*) Feb. 2009 (GRCh37/hg19) assembly) を用いて決定した。

本研究で新規転移させた LINE 配列に加え、以前報告された論文から、HeLa 細胞 (Gilbert et al., 2005) 及び HCT116 細胞 (Symer et al., 2002) で新規転移した L1.3 配列のデータを抽出し本研究の解析に使用した。L1.3 は、ヒトゲノム中のヒト L1 のあるローカスから得た配列である。今後、L1.3 を L1 として表記する。

2-6. DT40 細胞中で転移させた ZfL2-2 転移後配列の単離

DT40 細胞中で転移させた ZfL2-2 転移後配列は以前の研究 (Suzuki et al., 2009) で得られたデータを使用して解析を行った。

2-7. ゲノムデータベースを用いた LINE 配列の検索

ゼブラフィッシュ LINE RTE-1_DR 全長コンセンサス配列は Rebase Update (<http://www.girinst.org/rebase/update/index.html>) から取得した (Kapitonov and Jurka, 2005)。この配列をゼブラフィッシュゲノムデータベース (<http://genome.ucsc.edu>; The Jul. 2010 zebrafish (*Danio rerio*) Zv9 assembly) で検索した。この検索で 6 つのゼブラフィッシュ LINE RTE_DR 全長配列を得た。

ゼブラフィッシュ LINE Nimb-2_DR 全長コンセンサス配列は Rebase Update (<http://www.girinst.org/rebase/update/index.html>) から取得した (Kojima and Jurka, 2010a)。この配列をゼブラフィッシュゲノムデータベース (<http://genome.ucsc.edu>; The

Jul. 2010 zebrafish (Danio rerio) Zv9 assembly) で検索した。この検索で合計 16 の全長配列及び 5´ 欠失配列のゼブラフィッシュ LINE Nimb-2_DR を得た。

トカゲ LINE RTEX_ACar 全長コンセンサス配列は Rebase Update (<http://www.girinst.org/rebase/update/index.html>) から取得した (Kojima and Jurka, 2010b)。この配列をトカゲゲノムデータベース (<http://genome.ucsc.edu>; The May 2010 Anolis carolinensis draft assembly (Broad version AnoCar2.0)) で検索した。この検索で 9 つのトカゲ LINE RTEX_ACar 全長配列を得た。

3. 第三章 結果

3-1. 5' EX 生成には宿主因子が関与する

以前私たちは、DT40 細胞でゼブラフィッシュ LINE ZfL2-2 の転移実験を行った (Suzuki et al., 2009)。この実験で、私は、野生型 DT40 細胞と NHEJ 経路に関わる遺伝子欠損 DT40 細胞のゲノム DNA へ転移した 122 の ZfL2-2 転移後配列を決定した。まず、私は、5' EX の生成機構に関する知見を得るため野生型 DT40 細胞で転移した ZfL2-2 の 5' 末端配列を解析した。野生型 DT40 細胞で転移した ZfL2-2 は、26 配列中 12 配列 (~45%) が 5' EX を伴っていた (図 11A)。この 5' EX 出現頻度 (~45%) は、私の所属する研究室の先行研究で示されたゼブラフィッシュゲノム中に存在する内在 ZfL2-2 の 5' EX 出現頻度 (~50%) と類似していた (Ichiyanagi and Okada, 2008)。この結果は、ゲノム DNA への ZfL2-2 挿入機構が、ゼブラフィッシュ細胞 (ホモ) とニワトリ DT40 細胞 (ヘテロ) で類似していることを示すのかもしれない。次に、ZfL2-2 を HeLa-RC 細胞中で新規転移させ、その 5' 末端配列を解析した。HeLa-RC 細胞で転移した ZfL2-2 は 20 配列中 2 配列 (10%) のみが 5' EX を伴っており、DT40 細胞での場合 (~45%) と比較して優位に少なかった (図 11A; $P=0.019$, マン・ホイットニーの U 検定)。HeLa-RC 細胞と野生型 DT40 細胞間の 5' EX の出現頻度の違いは、これらの細胞間で異なる発現量 (あるいは機能) を持つ何らかの宿主因子が 5' EX 生成に関与することを示していると考えられる。

3-2. 5' EX 生成と NHEJ 経路の遺伝子群

次に私は、NHEJ 経路の遺伝子群が 5' EX の生成に関与するか否か検証するため、NHEJ 経路遺伝子の欠損細胞株で新規転移した ZfL2-2 の 5' EX の出現頻度を解析した。NHEJ 経路の遺伝子群は ZfL2-2 の DT40 細胞での効率的転移に必要であることが示されている (Suzuki et al., 2009)。しかし今回の解析から、遺伝子欠損細胞株において 5' EX の出現頻度に違いは見られず (図 11A)、NHEJ 経路の遺伝子群は 5' EX の生成に関与していないと考えられた。

3-3. 5' EX 生成と LINE 種

次に私は、LINE 種の違いが 5' EX の出現頻度に影響を及ぼすか否か検証するため、2 つの LINE 種、ヒト L1 とゼブラフィッシュ ZfL2-1 を HeLa-RC 細胞と DT40 細胞で転移させ 5' EX の出現頻度を解析した (図 11B)。ゼブラフィッシュ ZfL2-1 の場合、HeLa-RC 細胞では 30 配列中 2 配列 (~5%) しか 5' EX を伴っていなかったが、DT40 細胞中では 32 配列中 15 配列 (~45%) が 5' EX を伴っていた (図 11B)。これはゼブラフィッシュ ZfL2-2 と同様の出現頻度である。一方、ヒト L1 の場合、HeLa-RC 細胞では 38 配列中 4 配列 (~10%) が 5' EX を伴っていたが、DT40 細胞では全く観察されなかった (図 11B)。これらの結果は、HeLa-RC 細胞中では LINE 種が異なっても 5' EX の出現頻度はおよそ一定 (~10%) であるが、DT40 細胞では、LINE 種が異なる (ゼブラ LINE とヒト LINE) と 5' EX の出現頻度が大きく異なることを示している。これらの結果から示唆されることとして、第一に、5' EX の生成には、宿主因子のみならず LINE の違いも関与することが挙げられる。さらに第二に、LINE の違いはいかなる場合でも 5' EX の生成に影響するわけではなく、HeLa-RC 細胞でのように 5' EX の生成に変化をもたらさない場合も存在する。

3-4. 5' EX 生成にはゲノム DNA の 3' 突出末端が必要である

次に、5' EX の生成が LINE 挿入位置のゲノム DNA の切断のされ方と関係があるか否か検証するため、5' EX と TSA (図 6) との関係性を検証した。5' EX の出現頻度の高い DT40 細胞で転移した ZfL2-2 配列の 5' EX の有無と TSA の種類を解析したところ、5' EX を伴う配列はすべて TSD (12 配列) であったが、5' EX を伴わない配列には TST (13 配列中 5 配列) が含まれている結果を得た (図 11C; $P=0.039$ 、フィッシャーの正確確率検定)。TSD は、LINE が挿入されるゲノム切断位置に 3' 突出末端を伴うと考えられることから (図 6)、5' EX が常に TSD を伴う本解析結果は、5' EX の生成にゲノム DNA の 3' 突出末端が必要であることを示唆する。

3-5. 5' EX は、その隣接配列に由来する

5' EX 配列がどのように合成されるのか、その合成に鋳型配列は存在するのかもしれないのか、5' EX 配列の由来はこれまで明らかにされていなかった。本研究では、5' EX 配列が、LINE 5' 連結部近傍の配列に由来するのではないかと仮説を立て DT40 細胞で転移した ZfL2-2 が伴う 5' EX 配列をホモロジー検索で解析した。短い 5' EX 配列は相同性の判定が困難なため、5bp 以上の 5' EX 配列について、その配列が近傍のゲノム DNA 配列あるいは ZfL2-2 配列 (5' EX との連結部からそれぞれ 25bp の領域) と相同性を持つか否か解析した。このホモロジー検索で、5' EX 配列の少なくとも 70% がその近傍配列と相同性を持つことが明らかとなった (表 1; 図 12A-D)。この結果は、DT40 細胞中で転移した ZfL2-2 の 5' EX はその隣接配列を鋳型に生成されることを示唆する。これらの相同性は 4 つのタイプに分けることができた。1 つ目は、隣接するゲノム DNA と一致する配列であった (タイプ 1)。2 つ目は隣接する ZfL2-2 と一致する配列であった (タイプ 2)。3 つ目は、隣接するゲノム DNA の相補鎖と一致する配列であった (タイプ 3)。4 つ目は隣接する ZfL2-2 の相補鎖と一致する配列であった (タイプ 4)。60% 以上の 5' EX 配列は、タイプ 1 とタイプ 2 であった (表 1; それぞれ 39% と 22%)。相補鎖と一致するタイプ 3 とタイプ 4 は稀に観察されるタイプであった (表 1; それぞれ 7% と 1%)。

次に、細胞種の異なる HeLa-RC 細胞で転移した ZfL2-2 の 5' EX 配列についても同様の解析を行った。解析可能な配列 (5bp 以上) は 1 サンプルであったが、この 5' EX 配列も隣接するゲノム DNA 配列と完全に一致した (図 12E)。この結果は、ZfL2-2 の 5' EX 配列は細胞種が異なってもその形成過程は類似しており、その近傍配列を鋳型に合成されることを示唆する。

3-6. 5' EX 生成機構の一般性

5' EX の生成機構は LINE 種が異なっても同様であるか否かを明らかにするために、私は、

ZfL2-2 と異なる LINE の転移で生じた 5' EX についても同様の解析を行った。以前 Gilbert ら (2005) によって、HeLa 細胞内で転移させたヒト L1 が 100 配列報告されている。そのうち 15 配列が 5' EX を伴っている (Gilbert et al., 2005)。私はこの 15 配列のうち 5bp 以上の 5' EX を持つ 7 配列について同様のホモロジー検索を行った。この解析により、7 配列中 5 配列の 5' EX がその隣接するゲノム DNA 配列または挿入された L1 の 5' 末端配列と同一性のある配列を含んでいた (図 13A)。

次に私は、ZfL2-2 や L1 以外の LINE についても解析するため、ゼブラフィッシュゲノムデータベース (<http://genome.ucsc.edu>; The Jul. 2010 zebrafish (Danio rerio) Zv9 assembly) やトカゲゲノムデータベース (<http://genome.ucsc.edu>; The May 2010 Anolis carolinensis draft assembly (Broad version AnoCar2.0)) を検索した。ゼブラフィッシュゲノム中からは 6 つのゼブラフィッシュ LINE RTE_DR 全長配列、トカゲゲノム中からは 9 つのトカゲ LINE RTEX_ACar 全長配列を得た。これらの LINE 配列は、自身のコンセンサス配列とほぼ一致したため、これらの LINE 配列は挿入された時の配列を高度に維持していることが示唆された。これらの全長配列の中でも、TSD を伴う全長配列は LINE の 5' 末端とゲノム DNA の境界が区別できる。従って、これらの TSD を伴う全長配列の解析により、私は、2 配列の RTE_DR (図 13B; RTE_DR #1 は、タイプ 2、RTE_DR #2 はタイプ 3)、1 配列の RTEX_ACar (図 13C; タイプ 1) がタイプ 1 からタイプ 4 までのいずれかの 5' EX を伴っていることを発見した。以上の結果を総合すると、様々な生物種に存在する異なった種類の LINE の 5' EX 配列は共通の機構で生成されることが示唆される。

3-7. 5' EX の生成機構

上述で得られた 5' EX 配列データから、私は、どのように 5' EX が生じるか、考えられる可能性を探った。5' EX は隣接する配列由来であること (図 12、図 13) から考えられる最も簡潔な解釈は、5' EX は同一性のある隣接配列を鋳型として合成されるということである。加えて、5' EX 生成には、標的となるゲノム部位の 3' 突出末端が必要であることから

(図 11C)、このゲノム DNA の 3′ 末端と TPRT 反応で生じた LINE cDNA の 3′ 末端がアニーリングを介して相互作用する可能性が考えられる。このアニーリングは、隣接するゲノム DNA を鋳型として 5′ EX を合成する開始反応に重要であるのかもしれない。タイプ 1 の 4 サンプル (図 14A、C、E、G、I) を例に、5′ EX の生成機構モデルを図 14B、D、F、H、J に示す。まず、ゲノム DNA と LINE DNA の 2 つの 3′ 末端がアニーリングする (図 14B、D、F、H、J)。このアニーリング部から DNA が数塩基合成される。その後、このアニーリングは一度解離し、新たな末端で再度アニーリングが起こる。その後、ゲノム DNA と LINE DNA の 2 つの末端は連結される。タイプ 2 の 5′ EX 生成も同様の機構で説明できる (図 15)。タイプ 3 およびタイプ 4 の場合、それぞれ、ゲノム DNA の 3′ 末端あるいは LINE cDNA の 3′ 末端でアニーリングを介したヘアピン構造が形成され、隣接する DNA 配列を鋳型に 5′ EX が合成される (図 16)。このヘアピン構造は一度解離し、その後、ゲノム DNA と LINE DNA の末端結合反応が進行するのだろう。

これら 5′ EX の合成を行う DNA ポリメラーゼは同定されていないが、これら DNA ポリメラーゼの役割は、LINE 転移の 5′ 末端連結の過程で、ゲノム DNA 末端と LINE DNA 末端が安定に相互作用し解離しないようアニーリング領域の長さを伸長する役目があるのかもしれない。それぞれの末端が安定的に相互作用し続けることで、その後の 5′ 末端連結過程が進行するのかもしれない。私は、この時、初めのアニーリングが一度も解離しなければ MH を伴った 5′ 末端連結、解離が起これば、5′ EX を伴った 5′ 末端連結が生じるのではないかと推定している。

3-8. 5′ 逆位配列は、ヒト細胞中で転移した L1 でのみ起こる

Ostertag ら (2001) によって提案されている 5′ 逆位配列生成機構、ツインプライミングは、ゲノム DNA の 3′ 突出末端と逆転写途中の LINE RNA がアニーリングし、第二の逆転写反応が起こるモデルである (Ostertag and Kazazian, 2001; 序論 1-9; 図 7C、図 8)。この 5′ 逆位配列生成にも宿主のなんらかの因子が関与している可能性がある。そのため私は、

ヒト細胞 (HeLa-RC, HeLa, HCT116) とニワトリ細胞 (DT40) 中で転移した L1、ZfL2-1、Nimb-2_DR の 3 種類の LINE について 5' 逆位配列の有無を調べた (図 17A)。まず、ヒト培養細胞中で転移した L1 と ZfL2-1 の 5' 欠失配列について調べた。その結果、L1 では、細胞種によりわずかな差はあるが、~20%の転移後配列において 5' 逆位配列が観察された。しかし、ZfL2-1 では解析した 30 配列すべてが 5' 逆位配列を持たなかった (図 17A)。ヒト L1 でのみ 5' 逆位配列が検出されたことから、5' 逆位配列生成は、LINE の性質に依存することが示唆される。ニワトリ DT40 細胞での転移の場合、ZfL2-1 のみならず L1 でも 5' 逆位配列は観察されなかった。唯一、Nimb-2_DR で 31 配列中 1 例のみ観察されたが、その割合は非常に低いものであった (図 17A)。この L1 で観察された 5' 逆位配列生成の細胞種特異性は、細胞種間で異なって発現または機能している宿主因子が 5' 逆位に関与することを示唆する。

3-9. L1 の 5' 逆位配列は、10~20bp の TSD を伴う

次に、5' 逆位配列の生成が LINE 挿入位置のゲノム DNA の切断のされ方と関係があるか否か検証するため、5' 逆位配列と TSA (図 17B-D) との関係性を検証した。LINE 転移によって様々なタイプの TSA がどのように生じるかを説明するモデルは以前に示されている (図 6; Gilbert et al., 2002)。このモデルでは、主に見られる短い TSA と稀に観察される長い TSA は異なったメカニズムで生じることが提案されている。そこで、私は、TSA の種類を次の 5 つに分類した: long TST (L-TST, >20 bp)、short TST (S-TST, ≤20 bp)、BEJ (0 bp)、short TSD (S-TSD, ≤20 bp)、long TSD (L-TSD, >20 bp)。ヒト細胞で転移した 5' 逆位を伴っていない L1 配列では、TST から TSD まで様々な TSA が観察された (図 18)。一方で、5' 逆位配列を持つ L1 のほとんどは、~15bp にピークのある 10 から 20bp の TSD であった。これらの結果と Gilbert ら (2002) の TSA 生成モデルから、L1 が 5' 逆位配列を伴って転移するためには、挿入位置のゲノム DNA の二本鎖目の切断が一本鎖目の切断位置より 10~20bp 下流の位置で起こる必要があると考えられる (図 6、図 17B-D)。(詳細は考察 4-4 で述べる。)

5′ 逆位配列は、2回の逆転写開始反応を必要とする。すなわち、5′ 逆位配列を伴う 5′ 末端連結は逆転写反応が関与するのである。一方、5′ 逆位配列を伴わない 5′ 末端連結に逆転写反応が関与する積極的証拠は得られていない。このことから、5′ 逆位配列を伴う LINE 転移と伴わない LINE 転移では、その 5′ 末端連結連結機構に大きな違いが存在すると考えられる。そこで、以後の解析では、5′ 逆位配列を伴う LINE 配列のデータを除外し、5′ 逆位配列を伴わない LINE 転移の 5′ 末端連結機構について解析した。

3-10. L1 の TSA 分布パターンはヒト培養細胞の違いで異なる

まず、ヒト L1 配列を異なるヒト培養細胞種で転移させた場合、TSA 分布パターンには変化が生ずるか否か解析した (図 18A)。HeLa-RC 細胞中で転移した L1 の TSA 分布は、HCT116 細胞中で転移した L1 の TSA 分布 (Symer et al., 2002) と類似していた (図 18A、図 18B)。一方、HeLa 細胞中で転移した L1 の TSA 分布 (Gilbert et al., 2005) は HeLa-RC 細胞および HCT116 細胞のものと統計学的に有意に異なっていた (図 18A-C)。特に、HeLa 細胞の L1 TSA では、HeLa-RC 細胞と HCT116 細胞のものでは観察されなかった L-TSD がその 39% を占め、HeLa-RC 細胞と HCT116 細胞では観察される 1 から 9bp の TSD が存在しなかった (図 18A-C)。これらのデータは同じヒト L1 配列であっても、転移するヒト培養細胞の種類が異なれば、TSA の分布が変化することを示している。これは、ヒト細胞種間で異なる発現を示す宿主因子が L1 の TSA 生成機構に関与していることを示唆する。更に、細胞種を変えて LINE を転移させ解析することで、LINE 転移における宿主因子の機能を解明することができる可能性を示している。

3-11. L1 転移におけるゲノム DNA の切断位置

次に、私はニワトリ DT40 細胞中でヒト L1 を転移させ、37 個の L1 転移後配列を解析した。驚くべきことに、様々な TSA が観察されたヒト細胞での L1 転移とは異なり (図 18)、DT40 細胞での L1 転移配列は、15bp にはっきりとしたピークを持つ 10 から 20bp の S-TSD から

なる TSA 分布を示した (図 19A)。Gilbert らによって提案された TSA 生成モデルによると、LINE 挿入位置のゲノム DNA の二本鎖目切断位置の違いが様々な TSA を生み出す (図 6)。DT40 細胞で観察された片寄った L1 の TSA 分布 (10 から 20bp の S-TSD) は、DT40 細胞中でのゲノム DNA の二本鎖目切断が、一本鎖目の切断位置より下流 15bp の位置を中心に約 10bp の範囲で起こる可能性を示唆する。この仮説が正しければ、HeLa-RC 細胞での L1 転移では、DT40 細胞と同様の位置で切断されたゲノム DNA の末端配列がプロセッシング (二次的切断) を受け、様々な TSA が生じているのかもしれない。

3-12. ZfL2-1 転移におけるゲノム DNA の切断位置

次に私は、L1 以外の LINE を転移させた時に、どのような分布の TSA が観察されるかを調べるため、ゼブラフィッシュ ZfL2-1 を DT40 細胞及び HeLa-RC 細胞内で転移させその転移後配列を解析した (図 19B、図 19C)。DT40 細胞中で転移した ZfL2-1 配列のほとんど (93%) が 5bp にはっきりとしたピークのある 1 から 8bp の S-TSD であり、TST はわずかであった (7%) (図 19B)。Gilbert ら (2002) のモデルを考慮すると、DT40 細胞での ZfL2-1 TSA 分布は、ゲノム DNA の二本鎖目の切断が一本鎖目の切断位置より下流 5bp を中心に約 10bp の範囲で起こると考えられる。一方で、HeLa-RC 細胞内で転移した 30 個の ZfL2-1 配列は、21 配列 (70%) が S-TST であった (図 19C)。残りの 30% は、1 つが BEJ、6 つが S-TST、2 つが L-TST であった (図 19C)。HeLa-RC 細胞中で転移した ZfL2-1 の TSA は、DT40 細胞中の TSA と比較して (図 19B、図 19C)、明らかな 4bp のピークがあるという点では同様であった。しかし、HeLa-RC 細胞中で転移した ZfL2-1 の TSA 分布は、DT40 細胞中での ZfL2-1 の TSA 分布と比較して統計学的に有意に異なっていた (図 19B、図 19C; $P < 0.05$ 、フィッシャーの正確確率検定)。この結果は、L1 の TSA 分布から推定された、二本鎖目の切断は、LINE が転移する細胞の種類によらず起こり、HeLa-RC 細胞で観察された TST は、生じた 3' 突出末端のプロセッシングによって生じるという仮説と矛盾しない。また、L1 の TSD のピークが ~15bp であったのに対して、ZfL2-1 の TSD のピークは ~5bp であったことから、二本鎖目の切断位

置は LINE の種類によって異なることが示唆される。このことは、私の研究室で以前に報告されたクレード特異的な TSD のピークが観察された報告と一致する (Ichianagi and Okada, 2008)。

3-13. LINE の TSA パターンと挿入配列長

LINE 転移を経てゲノムへ挿入された LINE 配列は頻繁にその 5' 領域が欠失しているため、様々な長さの LINE 挿入配列が存在する。上述の通り、私は、様々な TSA の生成は細胞種によって異なる働きを持つ宿主プロセッシング因子に依存する可能性があることを示した。様々な長さに欠失している LINE の 5' 末端すなわち、LINE の挿入配列長も HeLa-RC 細胞と DT40 細胞の間で異なる可能性がある。私は、この 2 つの培養細胞種間での L1 挿入配列長の比較を箱ひげ図を用いて行った (図 19D、19E)。この比較によって、DT40 細胞より短い L1 挿入配列が HeLa 細胞で観察された (図 19D; $P < 0.05$, Mann-Whitney U test)。ZfL2-1 の場合も、挿入配列長を HeLa-RC 細胞と DT40 細胞間で比較した。L1 の場合と同様に、HeLa-RC 細胞では DT40 細胞より短い ZfL2-1 配列長が観察された (図 19E; $P < 0.05$, Mann-Whitney U test)。TSA の解析結果と LINE 転移後配列長の解析結果 (図 18A、図 19A-C) を合わせると、HeLa-RC 細胞では、様々な種類の TSA と短い挿入配列、DT40 細胞では、LINE の種類によって特定の範囲に現れる TSD と長い挿入配列長がヒト L1 とゼブラフィッシュ ZfL2-1 両方で観察されることが示された。これらの結果から、TSA のパターンと挿入配列長の間に関係があるということが示された。これらの結果は、DT40 細胞より HeLa-RC 細胞中の方がより活性のある宿主のプロセッシング因子が LINE の 5' 末端連結反応に関与するということを示唆する。

3-14. 5' 末端連結の 2 つの経路 ; アニール連結とダイレクト連結

ゲノム DNA と LINE 転移後配列の 5' 末端連結部は、MH、5' EX、DJ の 3 つの場合のいずれかが存在する (序論 1-10; 図 9)。私は 5' EX の解析 (結果 3-3) から、5' EX の生成頻

度は宿主の違いや LINE の違いによって変化することを示した。次に私は、5′ 末端部を MH、5′ EX、DJ の 3 つで分類した場合、宿主の違いや LINE の違いでこれらの割合が変化するか否か解析した。まず、DT40 細胞内で転移した L1 と ZfL2-1 を解析した (図 20A)。DT40 細胞中で転移した ZfL2-1 挿入配列の場合、すべての転移後配列が 5′ EX もしくは MH で 5′ 連結されていた (図 20A)。5′ EX の生成機構 (結果 3-6) の項で述べたように、5′ EX あるいは MH を伴う 5′ 連結機構は、どちらもその途中段階でゲノム末端と LINE 末端のアニーリングを介する連結機構であると考えられる (図 27)。一方、DT40 細胞中で転移した L1 挿入配列は、その 5′ 末端連結部の多くがアニーリングを必要としない DJ で占められていた (図 20A; $P < 0.05$, Fisher's exact test)。これらの結果は、DT40 細胞では少なくとも 2 つの異なる 5′ 末端連結経路; アニーリング連結とダイレクト連結、が存在すると考えられる。また、その経路は LINE 種によって異なり、ZfL2-1 がアニーリング連結を、ヒト L1 がダイレクト連結を介して転移していると考えられる。一方、HeLa-RC 細胞で転移した L1 の場合、L1 挿入配列の大多数がアニーリングを介する連結である MH (55%) もしくは 5′ EX (13%) が観察され、残りがアニーリングを介さない DJ (32%) であった (図 20A)。同様に、HeLa-RC 細胞で転移した ZfL2-1 も、アニーリングを介する連結である MH (77%) と 5′ EX (7%) が大多数であり、DJ (16%) は少数であった (図 20A)。これらのデータは、ヒト細胞では、L1 と ZfL2-1 の両方がアニーリング連結を介して転移していることを示唆する。すなわち、2 つの末端連結経路のうちのどちらを介して転移するのかは、LINE 種だけではなく宿主因子も関与していると考えられる。

ここで私は、二本鎖目の切断位置の違いがこれら 2 つの連結経路のどちらを介して転移するか決定するというモデルを提案する。ヒト L1 では、一本鎖目の切断位置から平均 15bp 下流で二本鎖目が切断される (図 19A; 図 21A)。この切断により、ゲノム DNA は ~15bp の塩基対を介して相互作用している。一方、ZfL2-1 では、一本鎖目の切断位置から平均 5bp 下流で二本鎖目が切断される (図 19B; 図 21B)。この切断により、ゲノム DNA は ~5bp の塩基対を介して相互作用している。言い換えれば、L1 挿入位置のゲノム DNA の相互作用は解離

しにくく、ZfL2-1 挿入位置のゲノム DNA の相互作用は解離しやすいのだろう。アニーリング連結をするためにはゲノム DNA の解離が必要だと考えられる。そのため、解離しやすい ZfL2-1 の挿入位置ゲノム DNA は、アニーリング連結を介して LINE 配列と連結され、解離しにくい L1 の挿入位置ゲノム DNA は、ダイレクト連結を介して LINE 配列と連結されるのだろうと考えられる。加えて、LINE 挿入位置のゲノム DNA の解離のしやすさは、塩基対部位の長さだけでなく宿主因子にも依存するため、細胞種の違いで連結経路の選択に違いが生じるのだろうと考えられる。

3-15. LINE 挿入位置ゲノム DNA の解離しやすさの観察①

Ku70 タンパク質は、DNA の DSB 修復に関わる主要な宿主因子である。このタンパク質が、切断された二本鎖 DNA 末端に結合することで DSB 修復が開始される。私の研究室での以前の研究により、Ku70 遺伝子を欠損させた DT40 細胞中でのゼブラフィッシュ LINE ZfL2-2 挿入配列に高頻度で L-TST が観察された (Suzuki et al., 2009)。この結果は、LINE 転移時に生じたゲノム DNA 末端と Ku70 タンパク質の相互作用がゲノム DNA 末端を分解から保護することを示していると考えられる。この保護がなくなることで、ゲノム DNA が大きく欠失した L-TST が生成されるのだろう。そこで私は、L1 と ZfL2-1 の両方を DT40 細胞中で転移させその転移後配列を解析した (図 20B)。LINE 転移中のゲノム DNA の解離のしやすさが異なれば、L-TST の出現率の差として検出できると考えられる。野生型 DT40 細胞では、両方の LINE 共にほとんどすべての挿入配列 (~97%) が short-TSA (≤ 20 bp の TSD と TST 及び BEJ) であった (図 20B)。一方、Ku70 欠損 DT40 細胞中で転移させた ZfL2-1 挿入配列では野生型 DT40 細胞と比較して多くの L-TST が観察され (図 20B; $P < 0.05$ 、フィッシャーの正確確率検定)、Ku70 欠損 DT40 細胞中で転移させた L1 挿入配列では野生型 DT40 細胞中と同様の TSA が観察された (図 20B)。これらの観察より、末端 DNA を保護する Ku70 タンパク質の存在しない状態では、ZfL2-1 転移時の塩基対領域 (図 21B) の方が、L1 転移時の塩基対領域 (図 21A) よりもゲノム DNA 末端へ解離しやすく、そのため、ZfL2-1 において多

くの L-TST が形成されたと考えられる。また、L1 転移では Ku70 タンパク質の非存在下でもほとんど L-TST が観察されないことから、L1 転移時の塩基対領域は強固に結合していると考えられる (図 20B)。これらの観察結果は、2つの 5' 末端連結経路 (アニーリングとダイレクト) が塩基対領域の解離の度合いにより決定されるという仮説と矛盾しない。

3-16. LINE 挿入位置ゲノム DNA の解離しやすさの間接的な観察②

LINE 転移時に ZfL2-1 と L1 の間で塩基対領域の解離の度合いが異なるならば、5' 欠失配列と全長配列の割合も LINE 間で変化する可能性がある。なぜなら、塩基対領域が速やかにゲノム DNA 末端へ解離すれば、その末端が修復されるべき DNA 末端として認識され、逆転写の完了前に 5' 末端連結反応が開始されるかもしれないからである。一方、塩基対領域の解離が起こらなければ、逆転写を行うのに十分な時間が得られ、より長い LINE 配列の挿入が起こるかもしれない。そこで、私は、DT40 細胞中での全長 LINE 配列転移と 5' 欠失 LINE 配列転移の数を比較した (図 20C)。L1 の挿入配列の半分以上は全長配列転移 (62%; 図 20C) であったが、ZfL2-1 挿入配列では全長配列転移は稀にしか観察されなかった (6%; 図 20C)。この観察結果は、2つの 5' 末端連結経路 (アニーリングとダイレクト) が塩基対領域解離の度合いにより決定されるという仮説と矛盾しない。また、この観察結果は、塩基対領域の解離の度合いが LINE の長さの決定に影響を及ぼす可能性を示唆する。

3-17. ゼブラフィッシュ LINE Nimb-2_DR は 2つの 5' 末端連結経路を介して転移する

2つの 5' 末端連結経路 (アニーリング連結とダイレクト連結) と塩基対領域の解離の度合いとの関係性をさらに解析するため、ゼブラフィッシュゲノム中に存在する LINE の Nimb-2_DR (Kojima and Jurka, 2010) を用いて解析を行った。Nimb-2_DR は I グループの Nimb クレードに属する (図 3)。この LINE の全長コンセンサス配列は ORF1 と ORF2 の構造を持つ (Kojima and Jurka, 2010a)。私は初めに、内在性 Nimb-2_DR の TSA 情報を得るため Nimb-2_DR のコンセンサス全長配列を用いてゼブラフィッシュゲノムデータベース

(<http://genome.ucsc.edu>; the Jul. 2010 zebrafish (*Danio rerio*) Zv9 assembly) から 16 個の内在性 Nimb-2_DR 配列を得た。序論 1-11. の説明にあるように、内在性 LINE 配列の正確な TSA 情報を得ることは困難である。しかし、TSD の場合、ゲノム中の Nimb-2_DR 配列の 5' 末端及び 3' 末端に隣接するゲノム DNA の重複を検索することで TSD の長さを推定できる。一方で、BEJ や TST の場合、そのどちらであるのか判定できず、また、TST の長さを求めることはできない。そこで、本解析では内在性 Nimb-2_DR 配列の TSA を 0 (TST または、BEJ)、S-TSD、L-TSD に分類して解析した (図 22A)。この解析の結果、およそ 70% (16 配列中 11) の内在性 Nimb-2_DR 配列は 12bp 付近にピークが見られる 10-15bp の TSD を持つことが明らかとなった (図 22A)。次に、DT40 細胞中で Nimb-2_DR 配列を転移させ、30 個の Nimb-2_DR 転移後配列を得た (図 22B)。ゼブラフィッシュゲノム中の Nimb-2_DR 配列のデータと同様に、DT40 細胞中で転移した Nimb-2_DR 配列の 70% は 12-13bp 付近にピークが見られる 10-15bp の TSD を持っていた (図 22C)。次に 5' 連結部の特徴を調べたところ、30 個の Nimb-2_DR 転移後配列中、6 配列 (20%) が 5' EX、14 配列 (47%) が MH、10 配列が DJ (33%) であった (図 22C)。5' 連結部に 3 つすべてのタイプが同程度出現したことは、DT40 細胞中では Nimb-2_DR はアニーリング連結とダイレクト連結の 2 つの 5' 連結経路の両方を用いて転移する可能性を示唆している。

次に、Nimb-2_DR 転移後配列の全長配列と 5' 欠失の割合を調べた (図 22D)。その結果、30 配列中 11 配列 (37%) が全長配列であった (図 22D)。DT40 細胞中で転移させた LINE の全長配列の割合が L1 では 62% (34 配列中 21 配列)、ZfL2-1 では 6% (33 配列中 2 配列) であった (図 20C) ことから考えると、Nimb-2_DR の全長配列の割合は、L1 と ZfL2-1 の全長配列の割合の間である (図 20C、図 22C)。

L1 の全長配列と 5' 欠失配列では 5' 末端連結の経路が異なるといういくつかの報告 (Zingler et al., 2005; kojima, 2010) がされたため、次に、私は DT40 細胞内で転移させた L1、ZfL2-1 及び Nimb-2_DR を全長配列と 5' 欠失配列に分けてその TSA の分布を解析した (図 22E、図 22F)。L1 の場合、全長配列と 5' 欠失配列共に DJ が多く観察された (それ

ぞれ、76%、54%)。対照的に、ZfL2-1 の場合では全長配列と 5´ 欠失配列の両方ともに、全て 5´ EX または MH であった。興味深いことに、全長 Nimb-2_DR 配列の 5´ 末端は、主に DJ (82%) を伴っていた。一方、5´ 欠失 Nimb-2_DR 配列はその 5´ 末端に、5´ EX もしくは MH (95%) のどちらかを伴っていた。全長 Nimb-2_DR 配列の 5´ 末端は、ヒト L1 配列の 5´ 末端パターンと類似しており (図 22E)、5´ 欠失 Nimb-2_DR 配列の 5´ 末端は、ZfL2-1 配列の 5´ 末端パターンと類似していた (図 22E)。これらの観察から、Nimb-2_DR は DT40 細胞中で異なった 2 つの 5´ 連結経路の両方を用いて転移することが示された。1 つは、ダイレクト連結により全長配列を転移させる。もう一方は、アニーリング連結を介して 5´ 欠失配列を生み出す。本研究で提案した仮説、5´ 末端連結経路は LINE 挿入位置のゲノム DNA 塩基対領域の解離のしやすさに支配されている、から考えると、DT40 細胞で転移した Nimb-2_DR 配列の TSD のピークが、L1 (~15bp) と ZfL2-1 (~5bp) の間の値 (~12bp) であることが、Nimb-2_DR の 2 つの経路の使い分けに関係しているのかもしれない (図 19A、図 19B、(図 22A)。

4. 第四章 考察

本研究では、4種の異なるLINE、ヒトL1、ゼブラフィッシュ ZfL2-2、ゼブラフィッシュ ZfL2-1、ゼブラフィッシュ Nimb-2_DR をヒト HeLa-RC 細胞とニワトリ DT40 細胞中で新規に転移させ 300 以上の LINE 転移後配列を決定し、解析した。これらの配列解析から、私は、LINE 転移時における 5' 末端連結の分子機構に関して新たな知見を得ることができた。

4-1. アニーリングを介する 5' 連結機構

ZfL2-2 の 5' EX のデータ (図 11-図 16) から、5' EX の生成は次の一連の機構によって説明が可能である。TPRT 後、まず、LINE cDNA の 3' 末端がゲノム DNA の 3' 末端と相互作用する。DNA 合成は、これらの 2 つの DNA 末端の片方または両方から起こる (図 14-図 16、図 27)。この DNA 合成の後に LINE cDNA の 3' 末端とゲノム DNA の 3' 末端と相互作用部が解離し、より長くなった一本鎖 DNA 領域が露出する (図 14-図 16、図 27)。本研究では、「アニーリング、DNA 合成、および、解離」のサイクルが複数回起こったことを示唆する例を発見している (図 15G-H、図 23-図 26)。この 5' EX の生成過程は、より長いアニーリング領域の形成、すなわち、ゲノム DNA と LINE DNA の両末端の安定な相互作用に寄与しているのかもしれない。この安定な相互作用は、宿主の DNA 修復系が LINE の 5' 末端連結を完了させるために必要なのだろう。このことが正しいならば、DNA 合成によって生じた 5' 連結部の安定性の違いが、5' 連結部の様々な構造を決定しているのであろう (図 27)。最初に形成されたアニーリングがそれに続く連結反応のために十分に安定である時、その DNA 末端は、アニーリング部位の解離なしで連結される。その結果、マイクロホモロジー形成された 5' 連結部となる (図 27A-C)。一方で、形成されたアニーリングが安定でない時、より安定となるような次の DNA 合成のためのアニーリングの解離と再アニーリングが起こる。その結果、5' EX が生成される (図 27)。宿主が異なると 5' EX の出現頻度が異なるという結果から、この連結部の安定性を調節する宿主因子が LINE の 5' 末端連結時に働いているのかもしれない。

現在、5′ EX の生成に関与するポリメラーゼや連結部の安定性に作用する宿主因子は明らかにされていない。これらの宿主因子の同定は、LINE 転移機構の全容解明に重要である。

4-2. LINE 転移時における 2 つの 5′ 末端連結経路

今回の研究で得られたデータから、ゲノム DNA と LINE の 5′ 末端が連結するためのアニーリング連結とダイレクト連結の 2 つの異なった経路の存在が明らかとなった。この 2 つの経路は、異なった 2 種類の培養細胞中でそれぞれの LINE によって区別して使われていた。図 28 で、この 2 つの経路が異なった状況下でどのように区別し使用されているのかの総合的なモデルを示す。詳細な二本鎖目切断の分子機構は今回の研究では明らかにできなかったが、限定された位置で二本鎖目が切断されることが示唆された (図 19A、19B、図 28)。また、一本鎖目の切断位置からの二本鎖目の切断位置は、それぞれの LINE によって異なることが示唆された。ヒト L1 の場合、二本鎖目の切断は、一本鎖目の切断位置から ~15bp 下流の位置で起こる。その結果、ターゲットとなる部位に 2 つの切断で挟まれた ~15bp の塩基対領域が生じる (図 19A、図 21A、図 28)。一方で、ゼブラフィッシュ ZfL2-1 の場合、二本鎖目の切断は、一本鎖目の切断位置から ~5bp 下流の位置で起こる。その結果、ターゲットとなる部位に 2 つの切断で挟まれた ~5bp の塩基対領域が生じる (19B、図 21B、図 28)。DT40 細胞中では、L1 によって生じた ~15bp の塩基対領域は安定であり、その塩基対部位は解離しにくい。その結果、主に全長配列が産み出され、ダイレクト連結経路をとる (図 28A)。安定な塩基対形成は、おそらく、L1 mRNA が L1 RT に全て逆転写される時間を与えているのだろう。その結果、平滑末端の LINE DNA-RNA ハイブリットが形成され、ゲノム DNA とダイレクトに連結されることが可能となるのであろう。対照的に、DT40 細胞中では、ZL2-1 によって生じた ~5bp の塩基対領域は不安定であり、その塩基対は解離しやすい。その結果、主に 5′ 欠失配列が産み出され、アニーリング連結経路をとる。塩基対領域の解離によって生じた 3′ 突出ゲノム DNA 末端は、合成途中の LINE DNA とアニーリングすることで、5′ 末端連結反応へ進む。アニーリング連結経路にはこの合成途中の LINE DNA と 3′ 突出末端が

必要であると考えられる。言い換えれば、DT40 細胞では、ターゲットとなるゲノム位置で生じる塩基対領域の長さが、その塩基対の解離のしやすさを決定しているということである。そして、この解離のしやすさの違いがどちらの 5' 末端連結経路を選択するかを決定づける。一方で、HeLa-RC 細胞では、5' 末端連結部の特徴が L1 と ZfL2-1 とともに類似していた (図 28C-D)。L1 と ZfL2-1 によって生じたそれぞれ、15bp と 5bp の塩基対領域は両方ともに同程度で解離しているのだろう。HeLa-RC 細胞では、5' 末端連結部で MH と 5' 欠失配列が支配的である。このことは、HeLa-RC 細胞では塩基対領域の解離が主に逆転写の完結前に起こり、アニーリング経路による 5' 末端連結反応へ進むことを示唆する。

4-3. LINE 転移時にゲノム DNA へ働く宿主因子

私が提案した 5' 末端連結のモデル (図 28) は、LINE 挿入位置ゲノム DNA の塩基対領域の解離の程度が DT40 細胞では塩基対領域の長さに依存するが、HeLa-RC 細胞では依存しないということになる。これは、少なくとも HeLa-RC 細胞では、この解離の程度に宿主因子が関与していることを想定している。HeLa-RC 細胞では、おそらく LINE TPRT 転移中間体は、迅速に宿主 DNA 修復因子に認識されるのであろう。その結果、塩基対領域の解離が速やかに進行し、次のアニーリング経路へと続くのだろう。HeLa-RC 細胞中での LINE 転移後配列は、DT40 細胞中の LINE 転移後配列より短いものが多く観察された (図 19D-E)。これは、HeLa-RC 細胞で見られたなだらかで広い範囲の TSA (図 18A、図 19C) および DT40 細胞で見られた狭い範囲の TSA (図 19A、図 19B) と相互に関連する。このことは、HeLa-RC 細胞で標的となるゲノム DNA の塩基対領域の解離が進行して露出した DNA 末端は、頻繁に宿主プロセッシング因子の作用を受けているが、DT40 細胞では露出した DNA 末端は、宿主プロセッシング因子の作用を受けていないことを示唆する。TPRT 中間体に働く宿主プロセッシング因子の活性は、DT40 細胞より HeLa-RC 細胞の方がより強い可能性がある。反対に、5' EX が HeLa-RC 細胞より DT40 細胞の方に高頻度で生じていたことから、TPRT 中間体に働く (ある種の) ポリメラーゼ活性は、HeLa-RC 細胞より DT40 細胞の方が強い可能性がある。

4-4. 5´ 逆位配列生成機構

今回の研究で、Nimb-2_DR の 1 サンプルを除いて、5´ 逆位配列はヒト細胞中で転移したヒト L1 配列でのみ観察された。5´ 逆位は第一鎖と第二鎖切断によって生じた 2 つのゲノム DNA 末端から LINE RNA の逆転写が開始されるツインプライミングによって生じると考えられている (Ostertag et al., 2001)。第二鎖切断位置から逆転写が始まる (二回目の逆転写反応) ためには、第一鎖切断位置から始まる逆転写 (一回目の逆転写反応) が完了する前に LINE 挿入位置のゲノム DNA の塩基対領域が解離し、3´ 突出末端とならなくてはならない (図 28B)。興味深いことに、5´ 逆位を伴わない L1 の転移配列は様々な TSA が観察されたが (図 18A-C)、5´ 逆位を伴う L1 の転移配列は、主に 10-20bp の TSD が観察された (図 17B-D)。この結果は、塩基対領域の解離により生じた 3´ 突出末端が、宿主プロセッシング因子より先に L1 の RT に認識されなくてはならないことを示唆する (図 28B-C)。一方で、プロセッシングされて生じた短い L1 の 3´ 突出末端は、二回目の逆転写反応開始に利用できないのであろう (図 28C)。また、DT40 細胞では、5´ 逆位を伴う L1 配列は観察されなかった (図 17A)。L1 の ~15bp の塩基対領域が解離されず、ツインプライミングが起らなかったことが考えられる。ZfL2-1 の場合、一回目の逆転写反応で全ての LINE RNA の逆転写が完了する前に ZfL2-1 の ~5bp の塩基対領域が解離され、3´ 突出末端を生じたように見えるが、HeLa-RC 細胞と DT40 細胞両方ともに、5´ 逆位配列は観察されなかった (図 17A)。解離によって生じた ~5 塩基の 3´ 突出末端は、二回目の逆転写の開始に必要な LINE RNA と安定にアニーリングするには短すぎる可能性がある。このことは、ヒト細胞中での L1 の 5´ 逆位配列が 10-20bp の長い TSD のみであったことと関係しているのかもしれない。

4-5. 5´ 末端連結機構の一般性

本研究のデータ (図 20A) は、HeLa-RC 細胞では主にアニーリング連結が優先的であることを示しており、ダイレクト連結の存在は明らかにできていない。しかしながら、Zingler

ら (2005) は以前に、ヒトゲノムに存在する L1 の解析により、5´ 欠失した L1 は、5´ 末端連結に MH を必要とし、全長 L1 は、5´ 末端連結に MH を必要としない (すなわち、DJ) ことを報告した。また、Kojima (2010) は、ヒトゲノム中の 5´ 欠失した L1 と比較して全長 L1 でははっきりしたピークのある TSD を持つことを報告している。加えて、Beck ら (2010) によって解析された個々のヒトゲノムに異なって存在する 66 の全長 L1 中で、私は、64 の全長 L1 配列が 10~20bp の TSD を伴っていることを発見した。これらの報告はヒト細胞でもアニーリング連結とダイレクト連結の 2 つ 5´ 末端連結経路が存在することを示唆する。これらの報告及び本研究は、アニーリング連結とダイレクト連結の 2 つの 5´ 末端連結経路はどのような細胞種においても普遍的に存在し LINE 転移に使用されていることを示唆する。しかし一方で、LINE 挿入部位で起こる 2 本鎖目の切断位置の違いや (様々な) 宿主因子の関与により、この 2 つの経路の LINE 転移における関与の度合いは様々に異なるようである。今後、これらの連結経路に関与する宿主因子を同定しその機能を解明していくことで、LINE の転移機構においてその機構がほとんど明らかにされていない 5´ 末端連結の全容が解明されるだろう。

謝辞

本研究を行うにあたり、研究の場や多くの研究発表の場を与え、ご指導くださった岡田典弘名誉教授に心から感謝いたします。研究を進めるにあたり、深い専門知識から実験技術、論文作成までの的確な助言および研究者としての姿勢や考え方をご指導して頂いた梶川正樹講師に深く感謝いたします。また、研究のサポートや研究者としての姿勢をご指導いただいた一柳健司博士、小島健司博士に深く感謝いたします。博士後期課程の研究生活や研究活動・発表において、多大なる支援をしていただいた東京工業大学グローバル COE に深く感謝いたします。そして、研究に関する様々なご指導をしてくださった先輩方や私の研究生活を支えて頂いた岡田・梶川研究室の皆様にも深く感謝いたします。最後に、私の長い研究生活を支援および応援してくれた家族に心から感謝いたします。

参考文献

Arkhipova, I., Meselson, M. (2000) Transposable elements in sexual and ancient asexual taxa. *Proc. Natl. Acad. Sci. USA*, **97**, 14473-14477.

Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., Moran, J.V. (2010) LINE-1 retrotransposition activity in human genomes. *Cell*, **141**, 1159-70.

Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. (1985) Ty elements transpose through an RNA intermediate. *Cell*, **40**, 491-500.

Burma, S., Chen, B.P.C., Chen, D.J., (2006) Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA Repair*, **5**, 1042-1048.

Babushok, D.V., Ostertag, E.M., Courtney, C.E., Choi, J.M., Kazazian, H.H. (2006) L1 integration in a transgenic mouse model. *Genome Res.*, **16**, 240-250.

Cost, G.J., Feng, Q., Jacquier, A., Boeke, J.D. (2002) Human L1 element target-primed reverse transcription in vitro. *EMBO J.*, **21**, 5899-5910.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Marchetto, M.C., Muotri, A.R., Mu, Y., Carson, C.T., Macia, A., Moran, J.V., Gage, F.H. (2011) Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc. Natl. Acad. Sci. U S A.*, **108**, 20382-20387.

Dewannieux, M., Esnault, C., Heidmann, T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet.*, **35**, 41-48.

Eichinger, D. J. & Boeke, J. D. (1988) The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. *Cell*, **54**, 955-966.

Esnault, C., Maestre, J., Heidmann, T. (2000) Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.*, **24**, 363-367.

Feng, Q., Moran, J.V., Kazazian, H.H. Jr, Boeke, J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905-916.

Garfinkel, D. J., Boeke, J. D. & Fink, G. R. (1985) Ty element transposition: reverse transcriptase and virus-like particles. *Cell*, **42**, 507-517.

Gasior, S.L., Wakeman, T.P., Xu, B., Deininger, P.L. (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.*, **375**, 1383-1393.

Gasior, S.L., Roy-Engel, A.M., Deininger, P.L. (2008) ERCC1/XPF limits L1 retrotransposition. *DNA Repair*, **7**, 983-989.

Gilbert, N., Lutz-Prigge, S., Moran, J.V. (2002) Genomic deletions created upon LINE-1 retrotransposition. *Cell*, **110**, 315-325.

Gilbert, N., Lutz, S., Morrish, T.A., Moran, J.V. (2005) Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol. Cell. Biol.*, **25**, 7780-7795.

Goodier, J.L., Ostertag, E.M., Kazazian, H.H. Jr. (2000) Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum Mol Genet.*, **9**, 653-7.

Goodier, J.L., Kazazian, H.H. Jr (2008) Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell*, **135**, 23-35.

Hancks, D.C., Kazazian, H.H. Jr. (2012) Active human retrotransposons: variation and disease. *Curr Opin Genet Dev.*, **22**, 191-203.

Honda, H., Ichiyanagi, K., Suzuki, J., Ono, T., Koyama, H., Kajikawa, M., Okada, N. (2007) A new system for analyzing LINE retrotransposition in the chicken DT40 cell line widely used for reverse genetics. *Gene*, **395**, 116-124.

Ichiyanagi, K., Nakajima, R., Kajikawa, M., Okada, N. (2007) Novel retrotransposon analysis reveals multiple mobility pathways dictated by hosts. *Genome Res.*, **17**, 33-41.

- Ichiyonagi, K., Okada, N. (2008) Mobility pathways for vertebrate L1, L2, CR1, and RTE clade retrotransposons. *Mol. Biol. Evol.*, **25**, 1148-1157.
- Kaer, K., Speek, M. (2013) Retroelements in human disease. *Gene*, **518**, 231-41.
- Kajikawa, M., Okada, N., (2002) LINEs Mobilize SINEs in the eel through a shared 3' sequence. *Cell*, **111**, 433-444.
- Kajikawa, M., Yamaguchi, Y., Okada, N. (2012) A new mechanism to ensure integration during LINE retrotransposition : A suggestion from analysis of the 5' extra nucleotides. *Gene*, **505**, 345-351.
- Kapitonov, V.V., Jurka, J. (2003) The esterase and PHD domains in CR1-like non-LTR retrotransposons. *Mol. Biol. Evol.*, **20**, 38-46.
- Kapitonov, V.V., Jurka, J. (2005) RTE-1_DR non-LTR retrotransposon from zebrafish genome. *Rebase Reports*, **5**, 94.
- Kapitonov, V.V., Tempel, S., Jurka, J. (2009) Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. *Gene*, **448**, 207-13.
- Kazazian, H.H. Jr. (2004) Mobile elements: drivers of genome evolution. *Science*, **303**, 1626-32.
- Kleckner, N. (1990) Regulation of transposition in bacteria. *Annu Rev Cell Biol.*, **6**, 297-327.
- Kojima, K.K., Fujiwara, H. (2005a). Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol. Biol. Evol.*, **22**, 2157-2165.
- Kojima, K.K., and Fujiwara, H. (2005b) An extraordinary retrotransposon family encoding dual endonucleases. *Genome Res.*, **15**, 1106-1117.
- Kojima, K.K. (2010) Different integration site structures between L1 protein-mediated retrotransposition in cis and retrotransposition in trans. *Mobile*

DNA, **1**, 17.

Kojima, K., Jurka, J. (2010a) Non-LTR retrotransposons from vertebrates. *Rebase Reports*, **10**, 487.

Kojima, K., Jurka, J. (2010b) Non-LTR retrotransposons in tetrapods. *Rebase Reports* **10**, 488.

Kulpa, D.A., Moran, J.V. (2006) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.*, **13**, 655-660.

Luan, D.D., Korman, M.H., Jakubczak, J.L., Eickbush, T.H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell*, **72**, 595-605.

Martin, S.L., Cruceanu, M., Branciforte, D., Li P.W., Kwok, S.C., Hodges, R.S., Williams MC, (2005) LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *J. Mol. Biol.*, **348**, 549-561.

Matsumoto, T., Hamada, M., Osanai, M., Fujiwara, H., (2006) Essential domains for ribonucleoprotein complex formation required for retrotransposition of telomere-specific non-long terminal repeat retrotransposon SART1. *Mol. Cell. Biol.*, **26**, 5168-5179.

Moran, J.V., DeBerardinis, R.J., Kazazian, H.H. Jr., Exon Shuffling by L1 Retrotransposition. (1999) *Science*, **283**, 1530-4.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., Kazazian, H.H. Jr (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917-927.

Martin, S.L. (2010) Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. *RNA Biol.*, **7**, 706-711.

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E.,

Batzer, M. A. and Moran, J. V. (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition.

Nat. Genet., **31**, 159-165.

Morrish, T. A., Garcia-Perez, J. L., Stamato, T. D., Taccioli, G. E., Sekiguchi, J., Moran, J. V. (2007) Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature*, **446**, 208-212.

Nakamura, M., Okada, N., Kajikawa, M. (2012) Self-Interaction, nucleic acid binding, and nucleic acid chaperone activities are unexpectedly retained in the unique ORF1p of zebrafish LINE. *Mol. Cell. Biol.*, **32**, 458-469.

Okada, N., Hamada, M., Ogiwara, I., Ohshima, K. (1997) SINEs and LINEs share common 3' sequences: a review. *Gene*, **205**, 229-243.

Ostertag, E. M., Kazazian, H. H. Jr (2001) Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.*, **11**, 2059-65.

Seleme, M. C., Busseau, I., Malinsky, S., Bucheton, A., Teninges, D. (1999) High-frequency retrotransposition of a marked I factor in *Drosophila melanogaster* correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes. *Genetics*, **151**, 761-771.

Sugano, T., Kajikawa, M., Okada N. (2006) Isolation and characterization of retrotransposition-competent LINEs from zebrafish. *Gene*, **365**, 74-82.

Suzuki, J., Yamaguchi, K., Kajikawa, M., Ichiyangi, K., Adachi, N., Koyama, H., Takeda, S., Okada, N. (2009) Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet.*, **5**, e1000461.

Symer, D. E., Connelly, C., Szak, S. T., Caputo, E. M., Cost, G. J., Parmigiani, G., Boeke, J. D. (2002) Human L1 retrotransposition is associated with genetic instability in vivo. *Cell*, **11**, 327-338.

Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., Takeda, S. (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.*, **17**, 5497-508.

Takahashi, H., Fujiwara, H. (2002) Transplantation of target site specificity by swapping the endonuclease domains of two LINEs. *EMBO J.*, **21**, 408-417.

Zingler, N., Willhoeft, U., Brose, H.P., Schoder, V., Jahns, T., Hanschmann, K.M., Morrish, T.A., Lower, J., Schumann, G.G. (2005) Analysis of 5' junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5-end attachment requiring microhomology-mediated end-joining. *Genome Res.*, **15**, 780-789.

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., Moran, J.V. (2001) Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol.*, **21**, 1429-39.

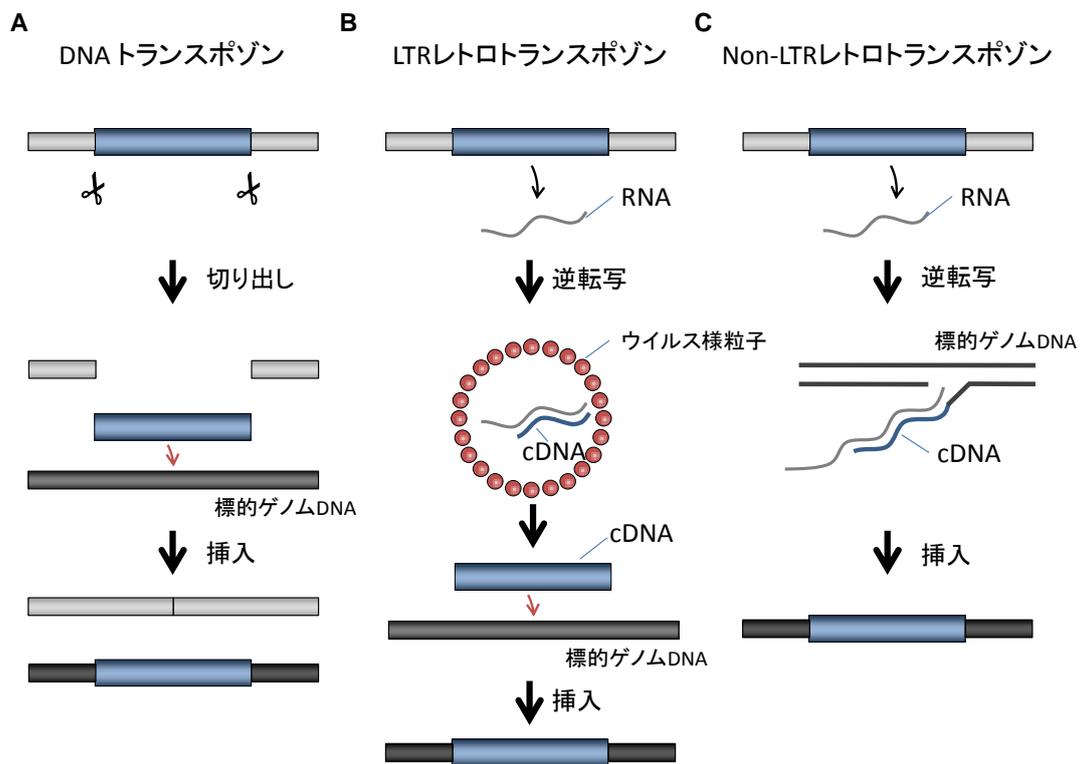


図 1. 転移因子種類と転移様式

- A. DNA トランスポゾンの転移様式
- B. LTR レトロトランスポゾンの転移様式
- C. non-LTR レトロトランスポゾンの転移様式

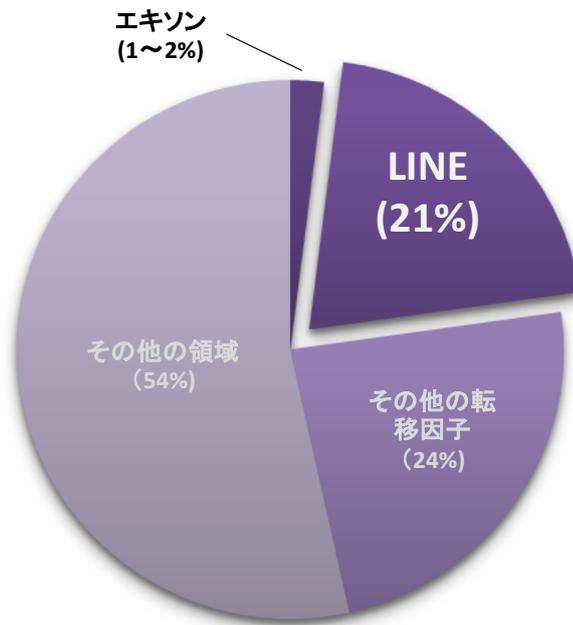


図 2. ヒトゲノム中の LINE の内訳

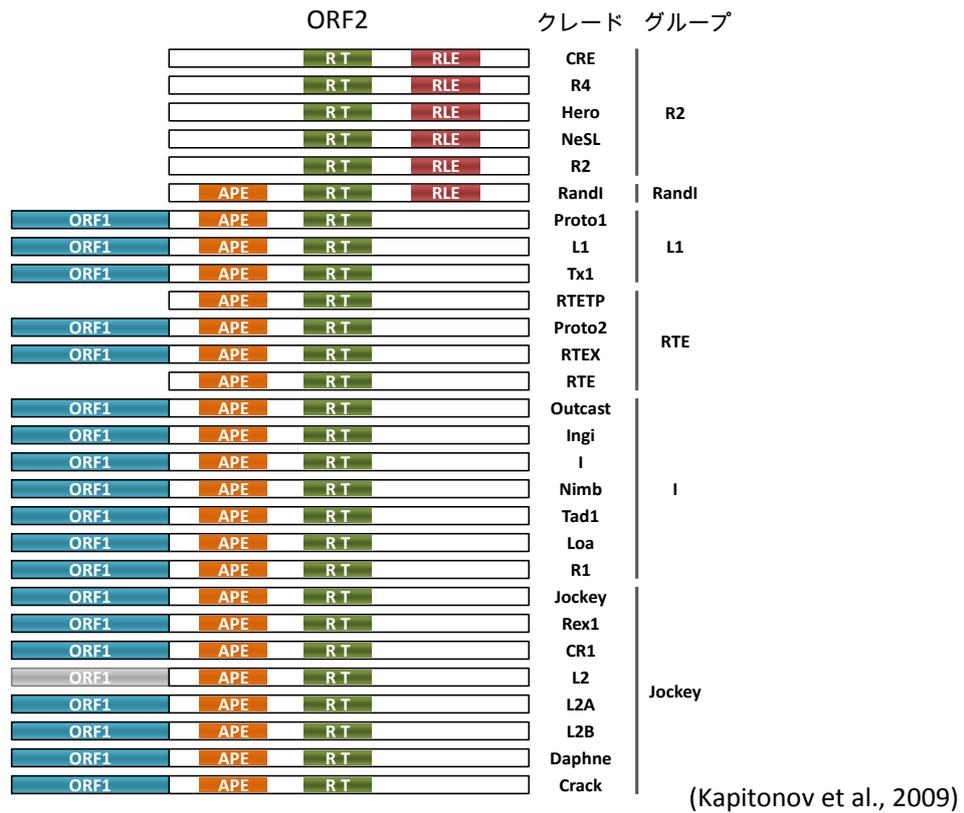


図 3. LINE の分類

LINE は、自身のコードする逆転写酵素 (RT) ドメインのアミノ酸配列系統解析を基に5つのグループに分けることができる。これらのグループはさらにクレードと呼ばれる28の小グループに分類される (Malik et al., 1999; Lovsin et al., 2001; Eickbush and Malik, 2002; Kapitonov et al., 2009)。制限酵素様エンドヌクレアーゼ (RLE) ドメイン; アプリン/アピリミジンエンドヌクレアーゼ (APE) ドメイン; L2 クレード中のいくつかのLINEファミリーにのみ存在する ORF1 はグレーで示されている。

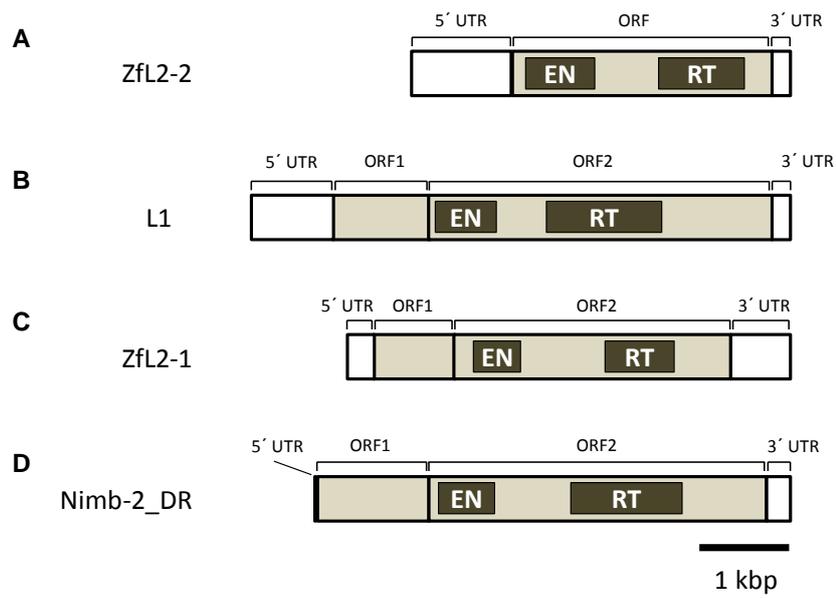


図 4. 本研究で使用した LINE の構造

- A. ゼブラフィッシュゲノム内に存在する L2 クレードに属する ZfL2-2 の構造。ORF を 2 つもつ LINE の ORF2 に相当する ORF を 1 つだけ持つ。
- B. ヒトゲノム内に存在する L1 クレードに属する L1 の構造。
- C. ゼブラフィッシュゲノム内に存在する L2 クレードに属する ZfL2-1 の構造。
- D. ゼブラフィッシュゲノム内に存在する Nimb クレードに属する Nimb-2_DR の構造。

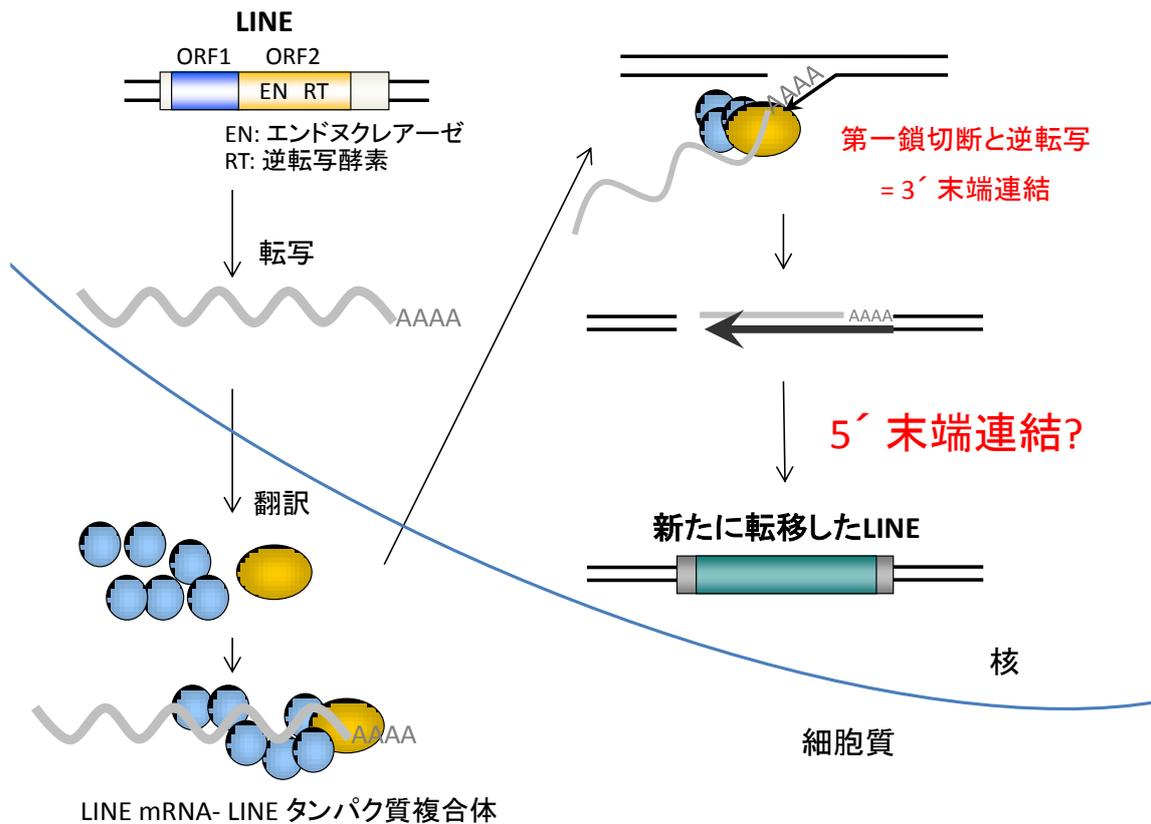


図 5. LINE 転移機構のモデル

LINE は、ORF1 と ORF2 の 2 つのタンパク質をコードしている。この ORF2 タンパク質は自身の転移に必要なエンドヌクレアーゼと逆転写酵素活性を持つ。LINE 転移は、LINE mRNA の転写で開始される。この LINE mRNA は核外へ輸送され、LINE タンパク質が翻訳される。この LINE タンパク質は LINE RNA と複合体を形成する。この複合体は核内へ輸送され、ゲノム DNA の LINE 挿入位置の一本鎖を切断する。生じたゲノム DNA 末端の 3' 水酸基から LINE RNA を逆転写する。この反応により、LINE の 3' 末端と宿主ゲノム DNA が連結される。続いて、ゲノム DNA の LINE 挿入位置の二本鎖目の切断、二本鎖目の LINE DNA の合成がされると考えられている。これらの反応の詳細は未解明である。その後、LINE 5' 末端と宿主ゲノム DNA が連結されるが、連結させる活性のない LINE タンパク質だけでは連結できないため、様々な宿主因子が 5' 末端連結に関与していると考えられている。しかし、LINE の 5' 末端連結機構は明らかにされていない。

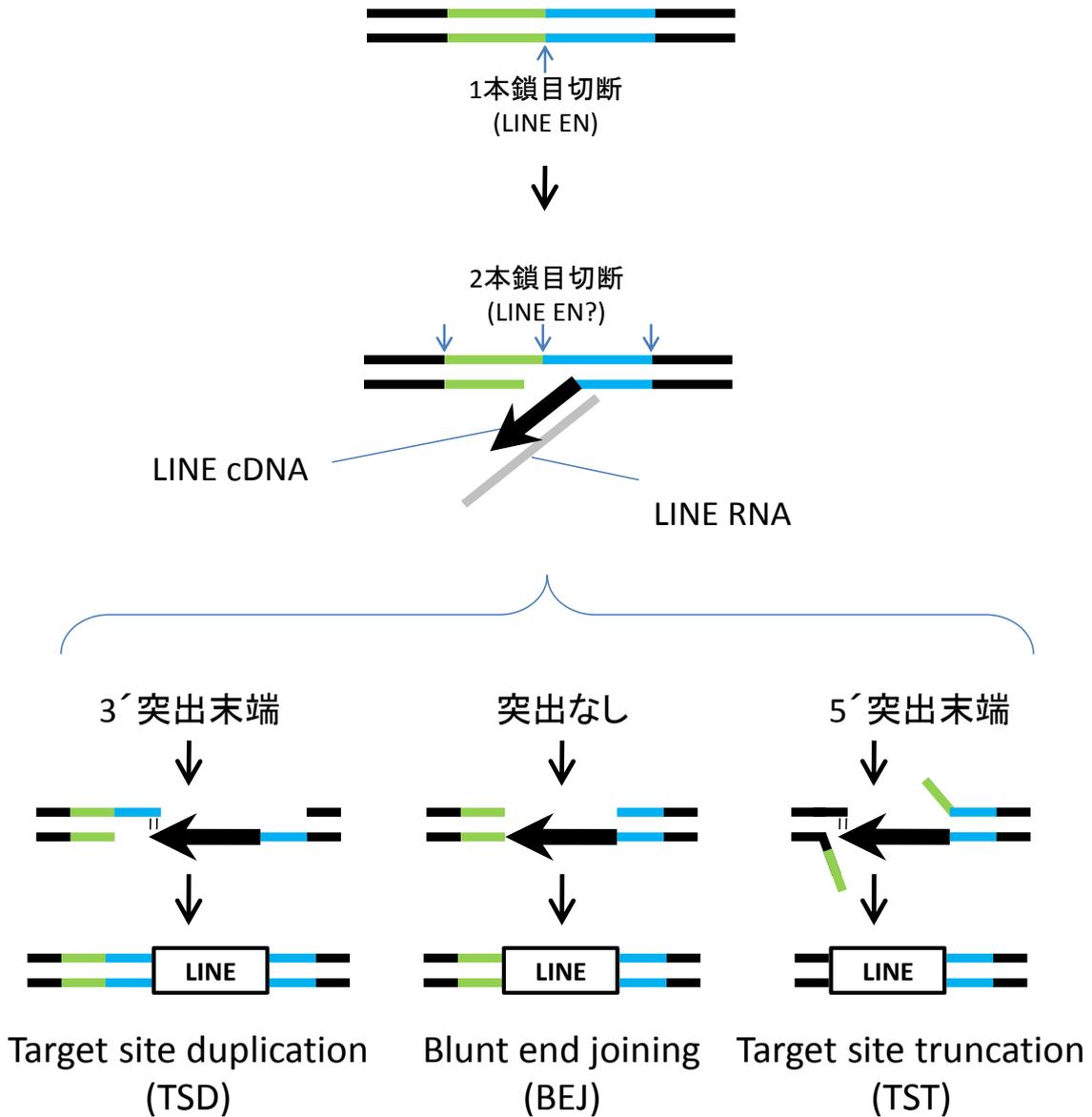


図 6. LINE 転移したゲノム DNA 部位の重複や欠失のモデル

LINE の 5' 連結部には、Target site alteration (TSA) と呼ばれる LINE 転移によって生じるゲノム DNA の欠失 (Target site truncation; TST) や重複 (Target site duplication; TSD)、が見られる。この TSA の生成は LINE の 5' 連結反応と密接に関係すると考えられている。この TSA 生成機構モデルは Gilbert ら (2002) によって提案された。このモデルによれば、ゲノム DNA の 2 本鎖目切断位置の違いが様々な TSA を生み出すと考えられている。2 本鎖目切断が 1 本鎖目の切断位置より下流の位置で起こると 3' 突出末端が生じ、この 3' 突出末端は LINE cDNA と連結され、その結果 LINE 配列の両端には、ゲノム DNA の重複 TSD が生じる。一方上流の位置で起こると 5' 突出末端が生じ、LINE との連結時に、この 5' 末端の欠失 TST が生じる。また、突出なしで 2 本鎖目の切断がされると、ゲノム DNA の変化

なく連結される Blunt-end joining (BEJ) となる。しかし、どのように 2 本鎖目の切断が起こるのか、TSA 生成に宿主因子は関与するのかなど、TSA 生成メカニズムの実験的証拠は明らかにされていない。

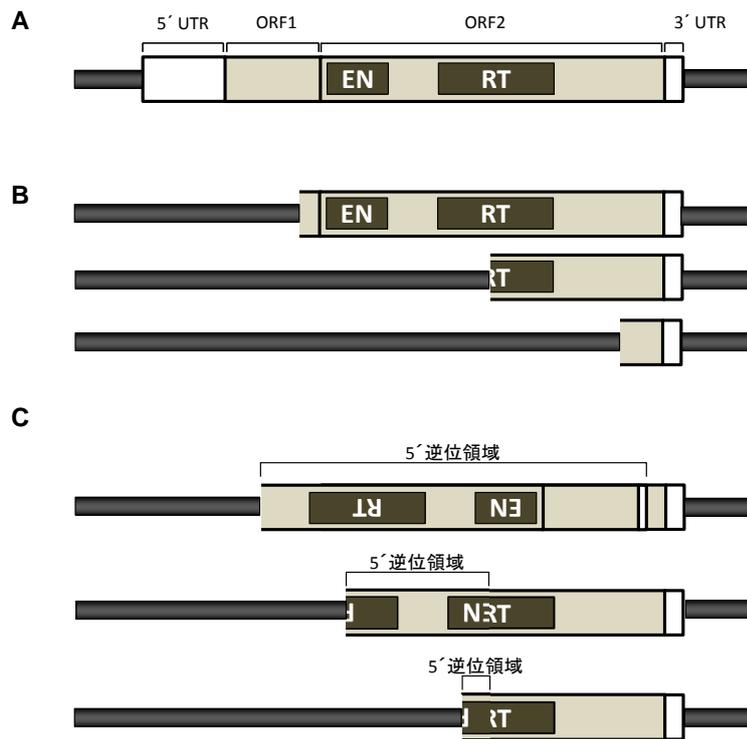


図 7. LINE 配列の 5' 末端領域構造

- A. LINE の全長配列
- B. LINE の 5' 欠失配列
- C. LINE の 5' 逆位配列

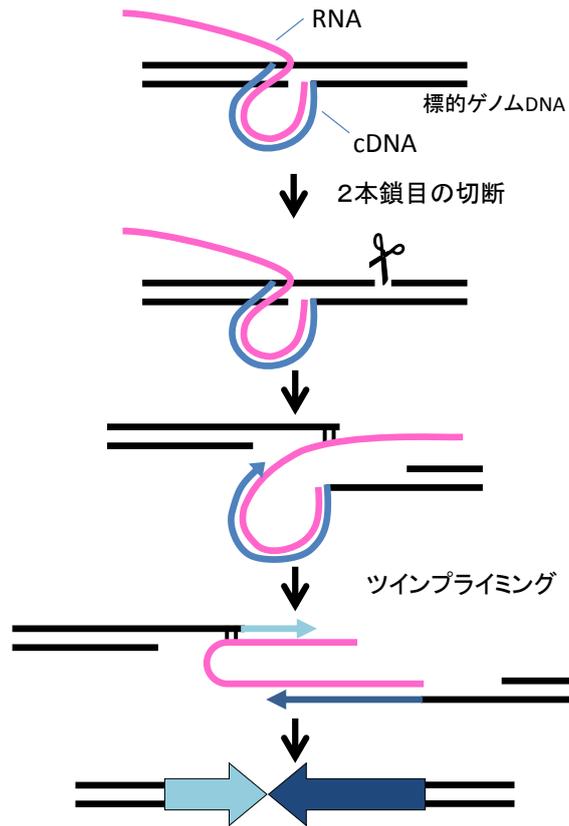


図 8. ツインプライミングモデル

5′ 逆位は LINE 転移する標的ゲノム DNA 部位の第一鎖と第二鎖切断によって生じた 2 つのゲノム DNA 末端から LINE RNA の逆転写が開始されるツインプライミングによって生じると考えられている (Ostertag and Kazazian, 2001)。ツインプライミングは、次の段階で進行する。第一鎖目の切断位置から TPRT が起こる。全ての逆転写が完了する前に第二鎖が切断される。生じたゲノム 3′ 突出末端をプライマーにして LINE mRNA の内部から 2 回目の逆転写反応が起きる。1 回目と 2 回目の逆転写で生じた 2 つの cDNA 末端が連結し、LINE の 5′ 逆位配列の転移が完了する。

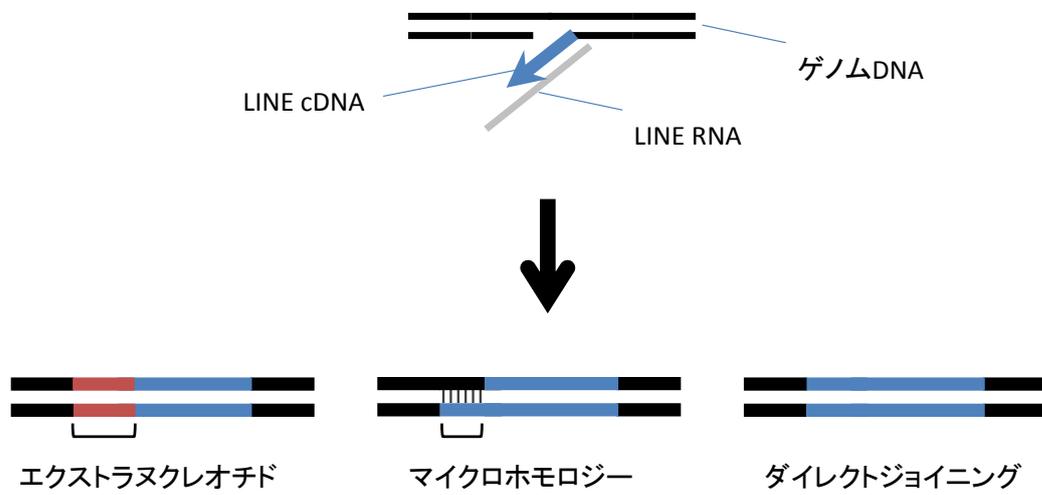


図9. LINEの5′末端とゲノムDNA末端の連結部に見られるエクストラヌクレオチド(5′EX)、マイクロホモロジー (MH) 及び、ダイレクトジョイニング (DJ)。

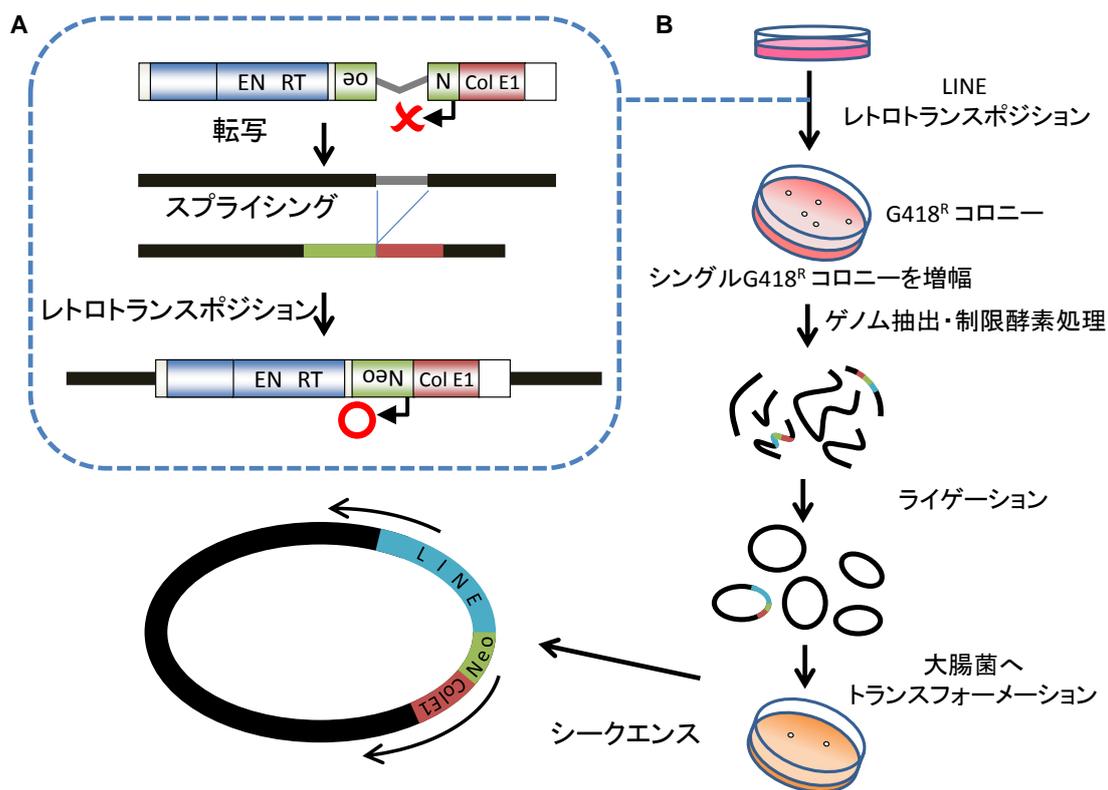


図 10. 培養細胞中で転移した LINE 配列の決定

A. 培養細胞内での LINE 転移検出に用いるプラスミドには、LINE の 3' UTR に、イントロンで分断された LINE 配列とは逆向きにネオマイシン耐性遺伝子が挿入されている。LINE mRNA が転写されイントロンがスプライシングにより抜け、LINE RNP の逆転写反応によってゲノム DNA に挿入されるとネオマイシン耐性 (Neo) 遺伝子を発現する。すなわち、LINE 転移した細胞のみネオマイシン (または、G418、カナマイシン) 薬剤耐性を獲得する。

B. プラスミド中の LINE は細胞内で LINE 転移によって宿主ゲノム中に組み込まれ、細胞は Neo 遺伝子を獲得し、G-418 薬剤耐性となる。この G-418 耐性細胞のシングルコロニーを培養後、細胞を回収し、ゲノム DNA を抽出する。抽出したゲノム DNA を LINE 転移によってゲノム中へ挿入された配列 (LINE、ColE1 及び、イントロンが抜け落ちた Neo 遺伝子) にはない切断部位をもつ *Hind*III で制限酵素消化する。次に、制限酵素消化産物を T4DNA ライゲースでセルフライゲーションさせる。このセルフライゲーション反応によって環状化した DNA 産物を大腸菌にエレクトロポレーション法で導入する。LINE インサートが含まれた環状 DNA のみ、大腸菌内で ColE1 複製起点を介して複製され、カナマイシン耐性 (Neo 遺伝子による) となるため、プレート上で生き残る。この薬剤選択により 1 から 100 個の大腸菌コロニーが得られる。このシングルコロニーから LINE インサートが含まれたプラスミド DNA を抽出する。このようにして得られたプラスミドは、LINE 挿入配列とゲノム DNA を含む。LINE 特異的なプライマーを用いて、LINE5' 及び 3' 連結部の配列を決定する。

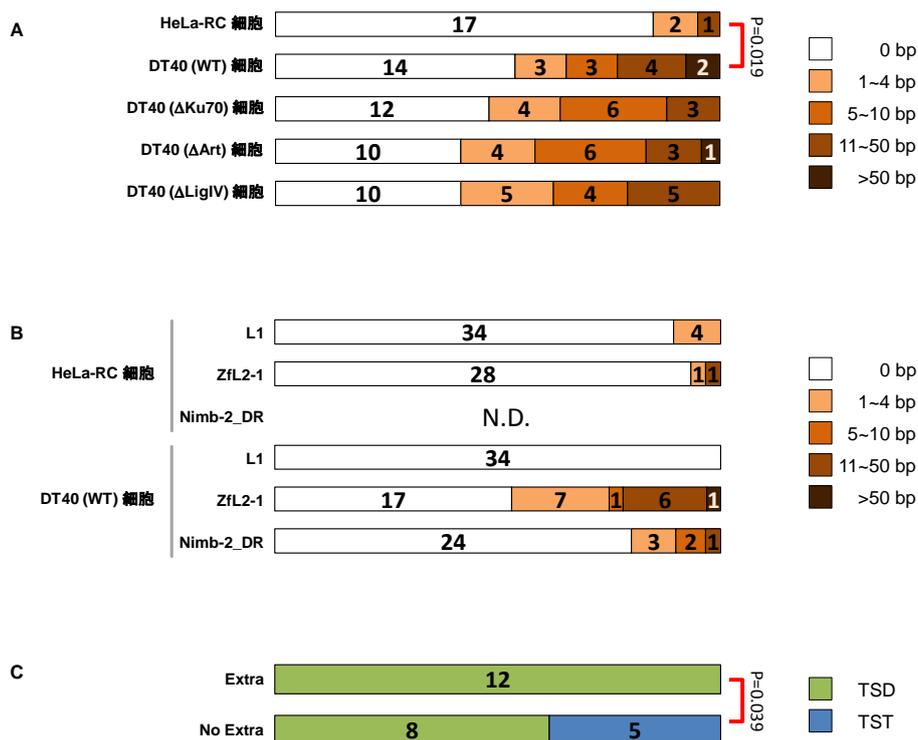


図 11. 培養細胞内で転移した ZfL2-2 の 5' EX

A. 異なる細胞種間での ZfL2-2 の 5' EX 出現頻度の比較。5' EX は長さに応じて、0bp、1~4bp、5~10bp、11~50bp、51bp 以上の 5 つに分類した。WT は野生型 DT40 細胞、 Δ Ku70 は Ku70 遺伝子欠損 DT40 細胞、 Δ Art は Artemis 遺伝子欠損 DT40 細胞、 Δ LigIV は DNA ligase IV 遺伝子欠損 DT40 細胞。統計解析には、マン-ホイットニーの U 検定を使用した。

B.

C. 野生型 DT40 細胞中で転移した ZfL2-2 の TSA。5' EX を伴っている ZfL2-2(Extra)と 5' EX

を伴っていない ZfL2-2 (No Extra) の間で比較した。統計解析には、フィッシャーの正確

確率検定を使用した。

- B. Ku70 遺伝子欠損 DT40 細胞で転移した ZfL2-2 の 5' EX 配列
- C. Artemis 遺伝子欠損 DT40 細胞で転移した ZfL2-2 の 5' EX 配列
- D. DNA ligase IV 遺伝子欠損 DT40 細胞で転移した ZfL2-2 の 5' EX 配列
- E. HeLa-RC 細胞で転移した ZfL2-2 の 5' EX 配列

A

Clone #13 CATAATAATAAAAAAATAATAAAATAA-----TGGTGGGGTCGGGGAGGGGGGAGG
Clone #14 AGAATTAGGAGCATCTTGCAGAAGTTTATTAT-----TTATGCAGCCATAAAAAATGATGAG
Clone #15 AGGACAGAGGCAGAGATTGAAGAGATTGAAGATTC-----AACAGGAAGAAGTTGAATCTCTGAA
Clone #16 CTGCTTTACCTCAATATGTCATTTGAAGACATATTG-----AACACATGAAGAAATGCTCATGATC
Clone #17 TTCTGAAATAACCTTCGCAGGGGTTAATGGTTTTTAAGGGTTTT--AATGGCAATCATAAAAAGTCAGGA
Clone #18 CAACAGAGTGAGACCCTGTCTGTTTATGTTTATGTTTTAAATAAAGACAATTATGCAGCCAAAAACA
Clone #86 TTTCACAAGTTATCGTCTTTAATTCACAAGTTAT-----CGTTTTATTCTGTCTTTTTATTGCC

B

RTE_DR #1 TTATGTGGACCGTACTCAGTACTTTCCAAGTTTTG-----GGTCCAAGGTTGTCATAGCCGGGG
RTE_DR #2 AGATAACAGTAAACTATAAAGCTATGGTTTATAGTTATAT-----GGTCCAAGGTTGTCATAGCCGGGG

C

RTEX_ACar CTGTTTGCAAGAAGTATTCTAGTGAAGAAGTATTC-----GGGATTGCAAGATGGCGCCGGAGTA

図 13. ZfL2-2 以外の LINE の 5' EX 配列。表示方法は、図 12 と同様である。
A. HeLa 細胞で転移したヒト L1 の 5' EX 配列 (Gilbert et al., 2005).
B. ゼブラフィッシュゲノムに内在する RTE_DR の 5' EX 配列
C. トカゲゲノムに内在する RTEX_ACar の 5' EX 配列

A

p70A4. 2a
 5' CCAAACATTATATATACAAGAATTC**AAGAA**TTCTTATCTGGAACATGCAGCTCAAC 3'
 |||
 3' GGGTTGTAATATATATATGTTCTTAAG**TCTT**AAGAATAGACTTGTACGTCGAGTTG 5'

B

5' CCAAACATTATATATACAAGAATTC
 AAGAATAGACTTGTACGTCGAGTTG 5'

↓ 2つの3'末端の相互作用

5' CCAAACATTATATATACAAGAATTC
 |||
 AAGAATAGACTTGTACGTCGAGTTG 5'

↓ DNA合成

5' CCAAACATTATATATACAAGAATTC
 |||
GTTCTTAAGAATAGACTTGTACGTCGAGTTG 5'

↓ 解離と再アニーリング

5' CCAAACATTATATATACAAGAATTC
 |
GTTCTTAAGAATAGACTTGTACGTCGAGTTG 5'

↓ 5'末端連結の完了

5' CCAAACATTATATATACAAGAATTC**AAGAA**TTCTTATCTGGAACATGCAGCTCAAC 3'
 |||
 3' GGGTTGTAATATATATATGTTCTTAAG**TCTT**AAGAATAGACTTGTACGTCGAGTTG 5'

C

pWA1. 17a
 5' AAGAGCAAGGGCTCCCATTGGAAGA**TATTNNAG**ACCATACGCTACTTGAAACTACA 3'
 |||
 3' TTCTCGTCCCGAGGGTAAACTTCT**ATAAANN**TCTGGTATGCGATGAACCTTTTGATGT 5'

D

5' AAGAGCAAGGGCTCCCATTGGAAGA
 CTGGTATGCGATGAACCTTTTGATGT 5'

↓ 2つの3'末端の相互作用

5' AAGAGCAAGGGCTCCCATTGGAAGA
 |||
 CTGGTATGCGATGAACCTTTTGATGT 5'

↓ DNA合成

5' AAGAGCAAGGGCTCCCATTGGAAGA
 |||
TAAANNTCTGGTATGCGATGAACCTTTTGATGT 5'

↓ 解離と再アニーリング

5' AAGAGCAAGGGCTCCCATTGGAAGA
 |
TAAANNTCTGGTATGCGATGAACCTTTTGATGT 5'

↓ 由来未知なT付加を伴う5'末端連結の完了

5' AAGAGCAAGGGCTCCCATTGGAAGA**TATTNNAG**ACCATACGCTACTTGAAACTACA 3'
 |||
 3' TTCTCGTCCCGAGGGTAAACTTCT**ATAAANN**TCTGGTATGCGATGAACCTTTTGATGT 5'

E

pAA6. 1a
 5' CTCTACATAATCCATCTTTCTCTAA**TCTT**GGGAAAAATCACTTTGCGGAACGC 3'
 |||
 3' GAGATGATTAGGTAGAAAGAGATT**AGAAA**CCCTTTTTTAGTGAAAGCGCTTGCG 5'

F

5' CTCTACATAATCCATCTTTCTCTAA
 CCCTTTTTTAGTGAAAGCGCTTGCG 5'

↓ 2つの3'末端の相互作用

5' CTCTACATAATCCATCTTTCTCTAA
 |||
 CCCTTTTTTAGTGAAAGCGCTTGCG 5'

↓ DNA合成

5' CTCTACATAATCCATCTTTCTCTAA
 |||
TAGAAACCCTTTTTTAGTGAAAGCGCTTGCG 5'

↓ 解離と再アニーリング

5' CTCTACATAATCCATCTTTCTCTAA
 |
TAGAAACCCTTTTTTAGTGAAAGCGCTTGCG 5'

↓ 5'末端連結の完了

5' CTCTACATAATCCATCTTTCTCTAA**TCTT**GGGAAAAATCACTTTGCGGAACGC 3'
 |||
 3' GAGATGATTAGGTAGAAAGAGATT**AGAAA**CCCTTTTTTAGTGAAAGCGCTTGCG 5'

G

p4A1. 2a
 5' ATGCTGAACAGATTTTTTTTAAAT**TTTTA**AAGCCTCTCAACTGCTCCAGAATG 3'
 |||
 3' TACGACTTGCTAAAAAAATTTA**AAAA**TTCGGAGAAGTTGACGAGGTCTTAC 5'

H

5' ATGCTGAACAGATTTTTTTTAAAT
 TTCGGAGAAGTTGACGAGGTCTTAC 5'

↓ 2つの3'末端の相互作用

5' ATGCTGAACAGATTTTTTTTAAAT
 |||
 TTCGGAGAAGTTGACGAGGTCTTAC 5'

↓ DNA合成

5' ATGCTGAACAGATTTTTTTTAAAT
 |||
AAAAATTTCGGAGAAGTTGACGAGGTCTTAC 5'

↓ 解離と再アニーリング

5' ATGCTGAACAGATTTTTTTTAAAT
 |
AAAAATTTCGGAGAAGTTGACGAGGTCTTAC 5'

↓ 5'末端連結の完了

5' ATGCTGAACAGATTTTTTTTAAAT**TTTTA**AAGCCTCTCAACTGCTCCAGAATG 3'
 |||
 3' TACGACTTGCTAAAAAAATTTA**AAAA**TTCGGAGAAGTTGACGAGGTCTTAC 5'

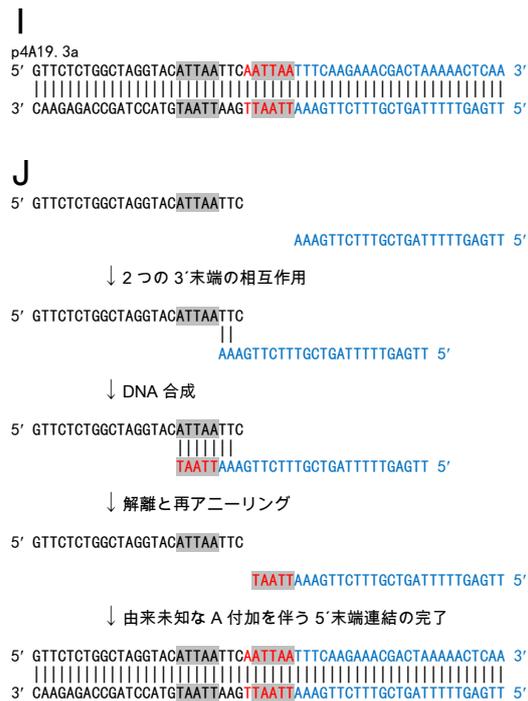


図 14. 想定されるタイプ1の5' EX生成機構。表示方法は、図12と同様である。

- A. p70A4. 2a 配列 5' 末端連結部
- B. p70A4. 2a の 5' EX 生成機構
- C. pWA1. 17a 配列 5' 末端連結部
- D. pWA1. 17a の 5' EX 生成機構
- E. pAA6. 1a 配列 5' 末端連結部
- F. pAA6. 1a の 5' EX 生成機構
- G. p4A1. 2a 配列 5' 末端連結部
- H. p4A1. 2a の 5' EX 生成機構
- I. p4A19. 3a 配列 5' 末端連結部
- J. p4A19. 3a の 5' EX 生成機構

A

pAA3. 5a

5' AAAATACCTCTCTGTACCTGTAGATCTGAACATCTGAACATGCAGCTCAACTCCTT 3'
 |||
 3' TTTTATGGAGAGACATGGACATCTAGACTTGTAGACTTGTACGTCGAGTTGAGGAA 5'

B

5' AAAATACCTCTCTGTACCTGTAGAT

TAGACTTGTACGTCGAGTTGAGGAA 5'

↓ 2つの3'末端の相互作用

5' AAAATACCTCTCTGTACCTGTAGAT
 |||
 TAGACTTGTACGTCGAGTTGAGGAA 5'

↓ DNA 合成

5' AAAATACCTCTCTGTACCTGTAGATCTGAACAT
 |||
 TAGACTTGTACGTCGAGTTGAGGAA 5'

↓ 解離と再アニーリング

5' AAAATACCTCTCTGTACCTGTAGATCTGAACAT
 |||
 TAGACTTGTACGTCGAGTTGAGGAA 5'

↓ 5'末端連結の完了

5' AAAATACCTCTCTGTACCTGTAGATCTGAACATCTGAACATGCAGCTCAACTCCTT 3'
 |||
 3' TTTTATGGAGAGACATGGACATCTAGACTTGTAGACTTGTACGTCGAGTTGAGGAA 5'

E

p4A1. 4a

5' ACTACCAGAGCATTGTATTGAGTATCTCTAGAACTGGGTAGATCTCTAGAACTGGGTACCAGC 3'
 |||
 3' TGATGGTCTCGTAACATAACTCATAAGATCTTAGACCCA TCTAGAGATCTTCGACCCA TGGTCG 5'

F

5' ACTACCAGAGCATTGTATTGAGTAT

TCTAGAGATCTTCGACCCA TGGTCG 5'

↓ 2つの3'末端の相互作用

5' ACTACCAGAGCATTGTATTGAGTAT
 |||
 TCTAGAGATCTTCGACCCA TGGTCG 5'

↓ DNA 合成

5' ACTACCAGAGCATTGTATTGAGTATCTCTAGAACTGGGTA
 |||
 TCTAGAGATCTTCGACCCA TGGTCG 5'

↓ 解離と再アニーリング

5' ACTACCAGAGCATTGTATTGAGTATCTCTAGAACTGGGTA
 |||
 TCTAGAGATCTTCGACCCA TGGTCG 5'

↓ 5'末端連結の完了

5' ACTACCAGAGCATTGTATTGAGTATCTCTAGAACTGGGTAGATCTCTAGAACTGGGTACCAGC 3'
 |||
 3' TGATGGTCTCGTAACATAACTCATAAGATCTTAGACCCA TCTAGAGATCTTCGACCCA TGGTCG 5'

C

pAA16. 1a

5' GAGAAAATTGTATAAGTGGATATTTCTAGATCTCTAGAACTGGGTACCAGCTG 3'
 |||
 3' CTCTTTTAAACATATTCACCTATAAAGATCTTAGAGATCTTCGACCCATGGTCGAC 5'

D

5' GAGAAAATTGTATAAGTGGATATTT

TAGAGATCTTCGACCCATGGTCGAC 5'

↓ 2つの3'末端の相互作用

5' GAGAAAATTGTATAAGTGGATATTT
 |||
 TAGAGATCTTCGACCCATGGTCGAC 5'

↓ DNA 合成

5' GAGAAAATTGTATAAGTGGATATTTCTAGAA
 |||
 TAGAGATCTTCGACCCATGGTCGAC 5'

↓ 解離と再アニーリング

5' GAGAAAATTGTATAAGTGGATATTTCTAGAA
 |||
 TAGAGATCTTCGACCCATGGTCGAC 5'

↓ 5'末端連結の完了

5' GAGAAAATTGTATAAGTGGATATTTCTAGAACTCTAGAACTGGGTACCAGCTG 3'
 |||
 3' CTCTTTTAAACATATTCACCTATAAAGATCTTAGAGATCTTCGACCCATGGTCGAC 5'

G



H

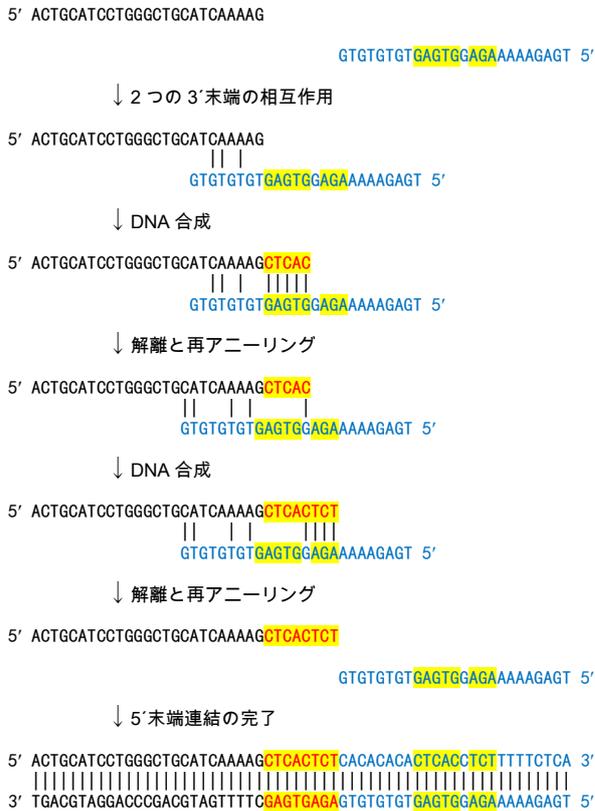


図 15. 想定されるタイプ 2 の 5' EX 生成機構。表示方法は、図 12 と同様である。

- A. pAA3. 5a 配列 5' 末端連結部
- B. pAA3. 5a の 5' EX 生成機構
- C. pAA16. 1a 配列 5' 末端連結部
- D. pAA16. 1a の 5' EX 生成機構
- E. pAA6. 1a 配列 5' 末端連結部
- F. pAA6. 1a の 5' EX 生成機構
- G. p4A1. 4a 配列 5' 末端連結部
- H. p4A1. 4a の 5' EX 生成機構
- I. ZfL2-2 配列 5' 末端連結部
- J. pWA11. 5a の 5' EX 生成機構

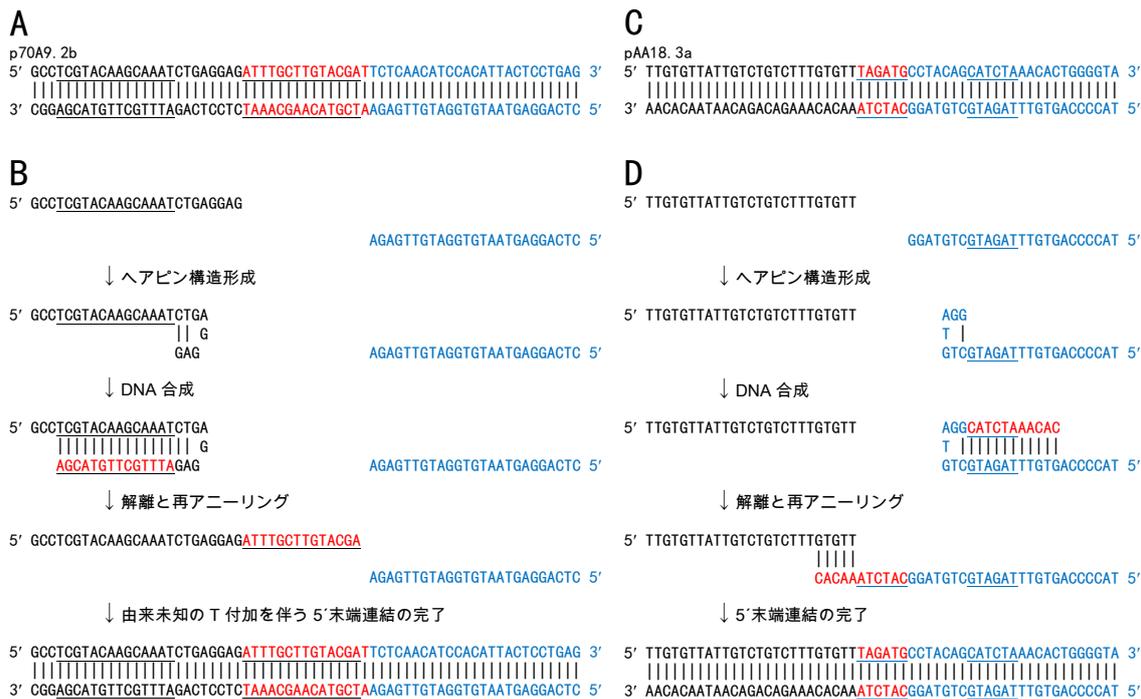


図 16. 想定されるタイプ 3 (p70A9. 2a) とタイプ 4 (pAA18. 3a) の 5' EX 生成機構。表示方法は、図 12 と同様である。

- A. p70A9. 2a 配列 5' 末端連結部
- B. p70A9. 2a の 5' EX 生成機構
- C. pAA18. 3a 配列 5' 末端連結部
- D. p AA18. 3a の 5' EX 生成機構

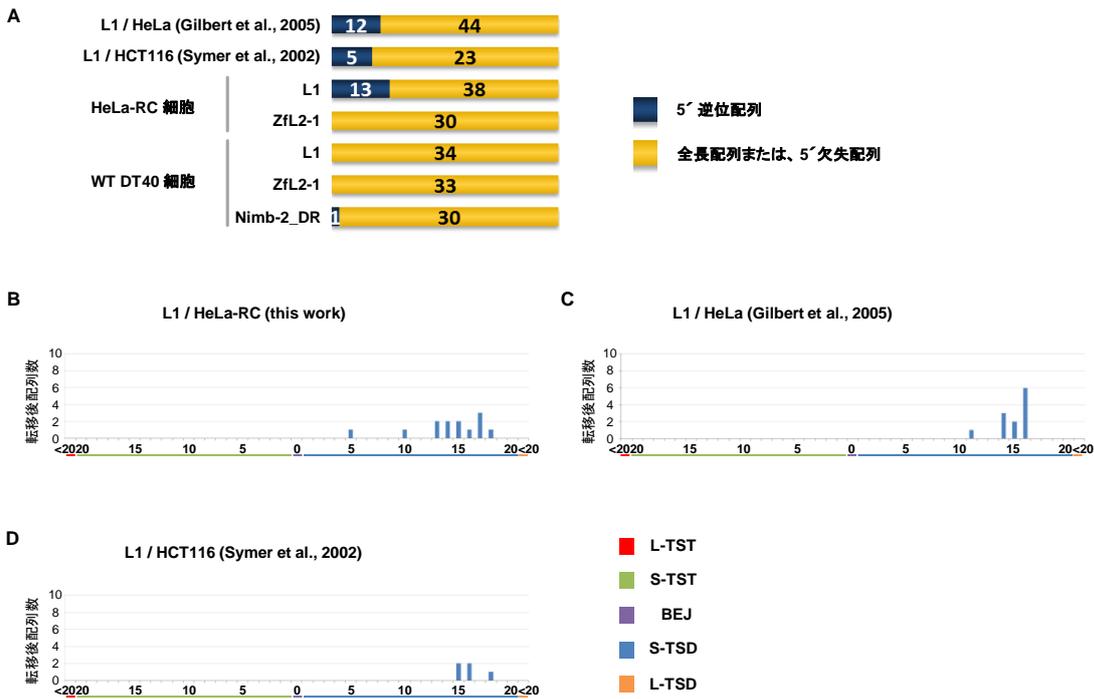


図 17. 5′ 逆位転移後配列は、ヒト細胞で転移した L1 に見られ、~15bp にピークのある TSD を持つ。

A. 5′ 逆位配列と、全長配列または 5′ 欠失配列に分類した LINE 配列の構造。ヒト細胞を上部、DT40 細胞を下部に示す。

(B-D) 培養細胞内で転移した 5′ 逆位配列の L1 TSA 分布。TSA は 5 つのタイプに分類される: long TST (L-TST, >20 bp)、short TST (S-TST, ≤20 bp)、BEJ (0 bp)、short TSD (S-TSD, ≤20 bp)、long TSD (L-TSD, >20 bp)。

B. HeLa-RC 細胞内で転移した 5′ 逆位配列の L1 TSA 分布。

C. Gilbert ら (2005) によって報告された HeLa-RC 細胞内で転移した 5′ 逆位配列の L1 TSA 分布。

D. Symer ら (2002) によって報告された HCT116 細胞内で転移した 5′ 逆位配列の L1 TSA 分布。

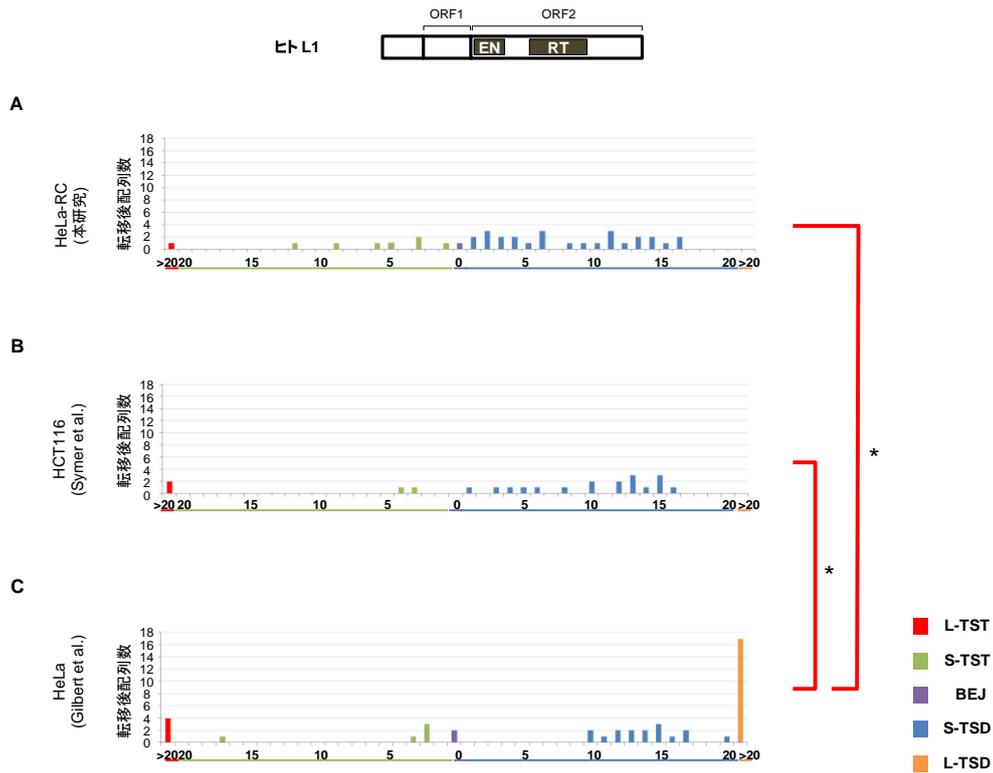


図 18. ヒト培養細胞の種類によって L1 の TSA パターンは変化する。

A. HeLa-RC 細胞中での L1TSA の長さの分布

B. Gilbert ら (2005) が報告した HeLa 細胞中での L1 TSA の長さの分布。Gilbert ら (2005) が報告したデータの中から L1.3 のデータのみを抽出した。

C. Symer ら (2002) が報告した HCT116 細胞中での L1 TSA の長さの分布。Symer ら (2002) が報告したデータの中から L1.3 のデータのみを抽出した。

統計解析は TSA それぞれのタイプに分けた分布を細胞種ごとに対してフィッシャーの正確確率検定を Bonferroni 補正して行った。この検定の結果、 $P < 0.05$ の組み合わせに対してアスタリスクで示した。図の最上部にヒト L1 構造の模式図を示す。

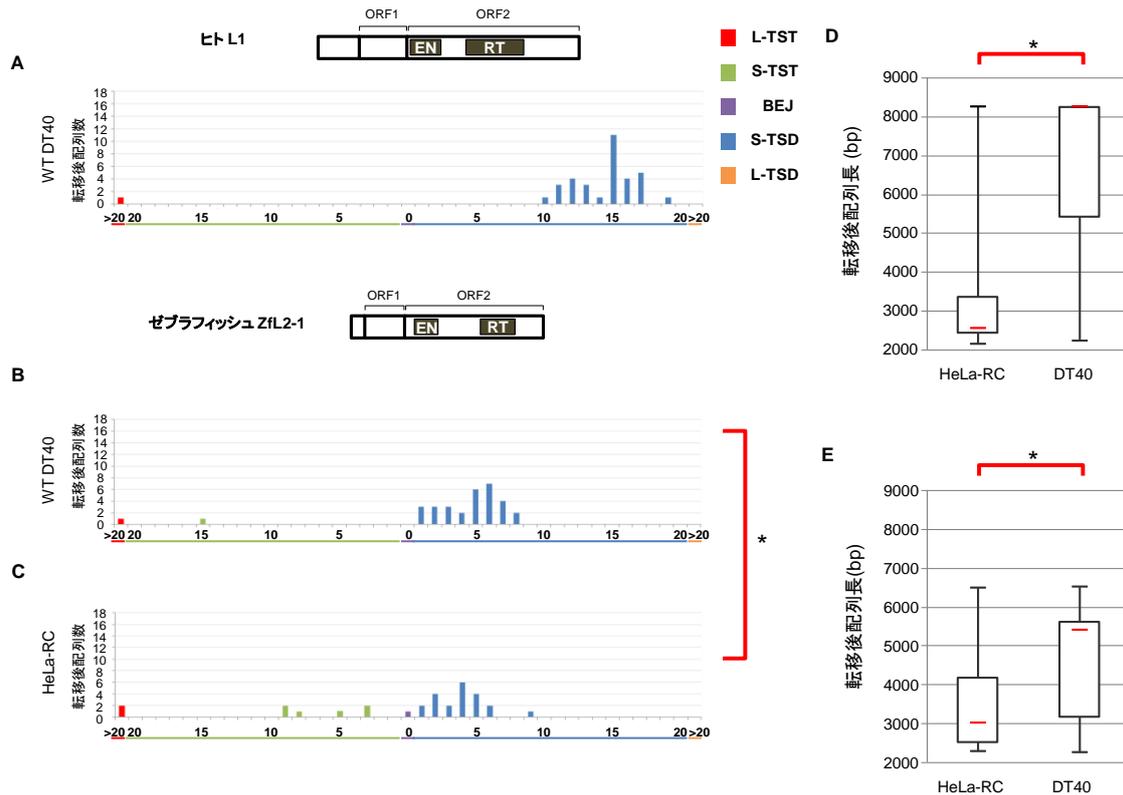


図 19. L1 と ZfL2-1 の転移後配列は、細胞種によって異なる。

A. DT40 細胞で転移したヒト L1TSA の分布

B. DT40 細胞で転移したゼブラフィッシュ ZfL2-1TSA の分布

C. HeLa-RC 細胞で転移したゼブラフィッシュ ZfL2-1TSA の分布

統計解析は TSA それぞれのタイプに分けた分布を細胞種ごとに対してフィッシャーの正確確率検定を行った。この検定の結果、 $P < 0.05$ の場合をアスタリスクで示した。

D. HeLa-RC 細胞と DT40 細胞で転移したヒト L1 の転移後配列長

E. HeLa-RC 細胞と DT40 細胞で転移したゼブラフィッシュ ZfL2-1 の転移後配列長

箱ひげ図は中央値 (赤線)、最大値 (バーの上端)、最小値 (バーの下端)、第一四分線 (箱の下端)、第三四分線 (箱の上端) を示す。細胞種ごとマン-ホイットニーの U 検定を行った。この検定の結果、 $P < 0.05$ の場合にアスタリスクで示した。転移後配列長の分布のグラフの上部にヒト L1 構造とゼブラフィッシュ ZfL2-1 構造の模式図を示す。

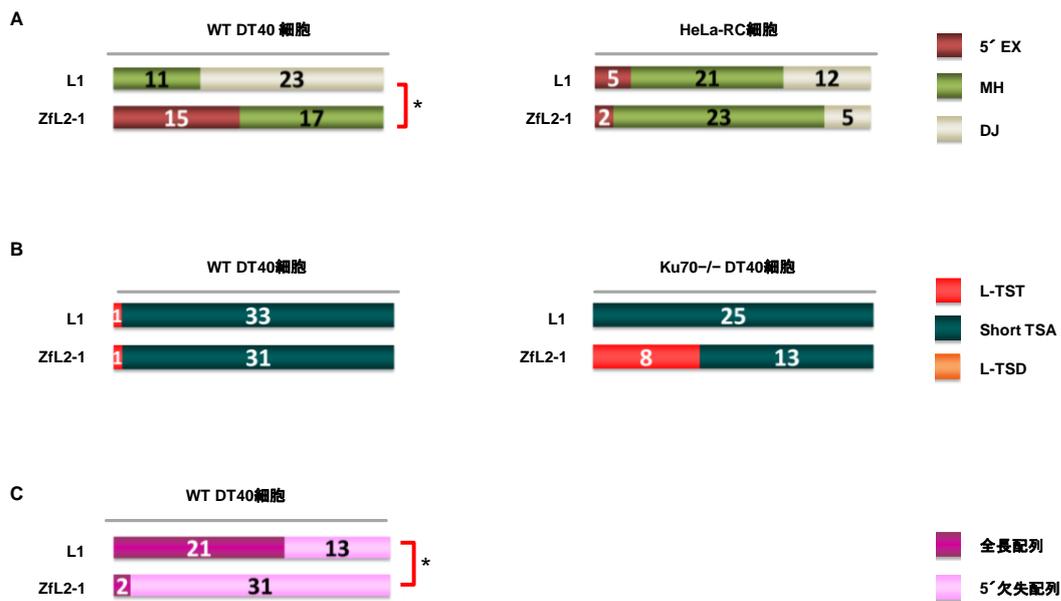


図 20. LINE 転移時の 5' 末端連結はアニーリングとダイレクトの 2 つの経路で行われる。
 A. DT40 細胞（左）または、HeLa-RC 細胞（右）で転移した L1 と ZfL2-1 の 5' 末端連結部（5' EX、MH、DJ）
 B. DT40 細胞（左）または、DT40 Ku70^{-/-}細胞（右）で転移した L1 と ZfL2-1 を Long (>20 nt) TSA と short (≤20 nt) TSA に分類した。
 C. DT40 細胞で転移した LINE（L1 と ZfL2-1）の全長配列と 5' 欠失配列の比較。
 統計解析は、フィッシャーの正確確率検定を行った。この検定の結果、P<0.05 の場合をアスタリスクで示した。

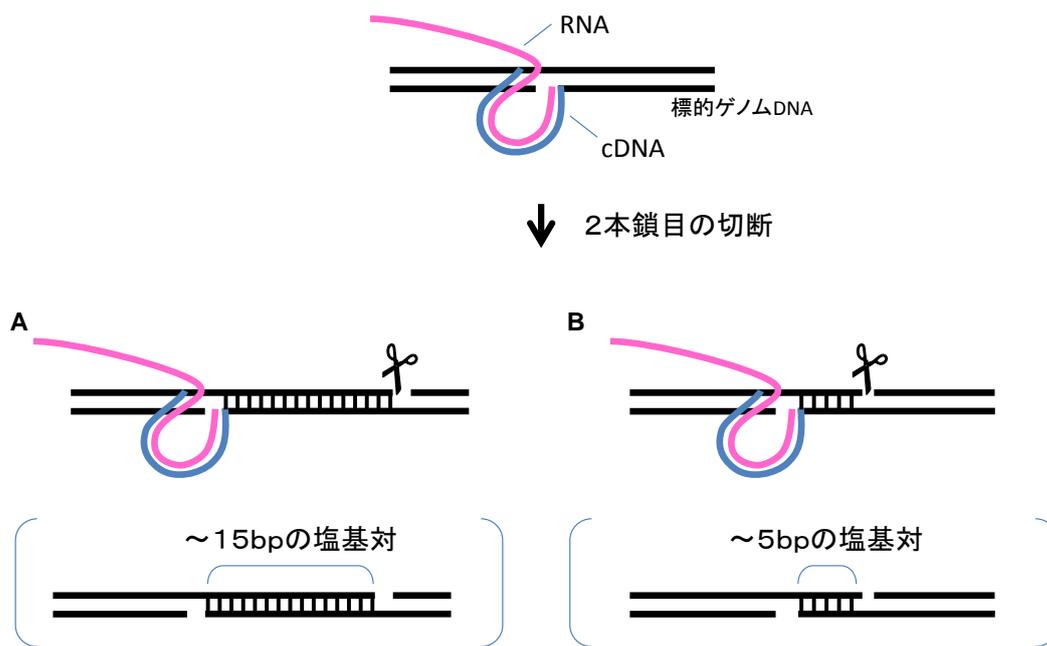


図 21. 二本鎖目の切断で生じる塩基対領域。

A. ~15bp の塩基対領域の生成。

B. ~5bp の塩基対領域の生成。

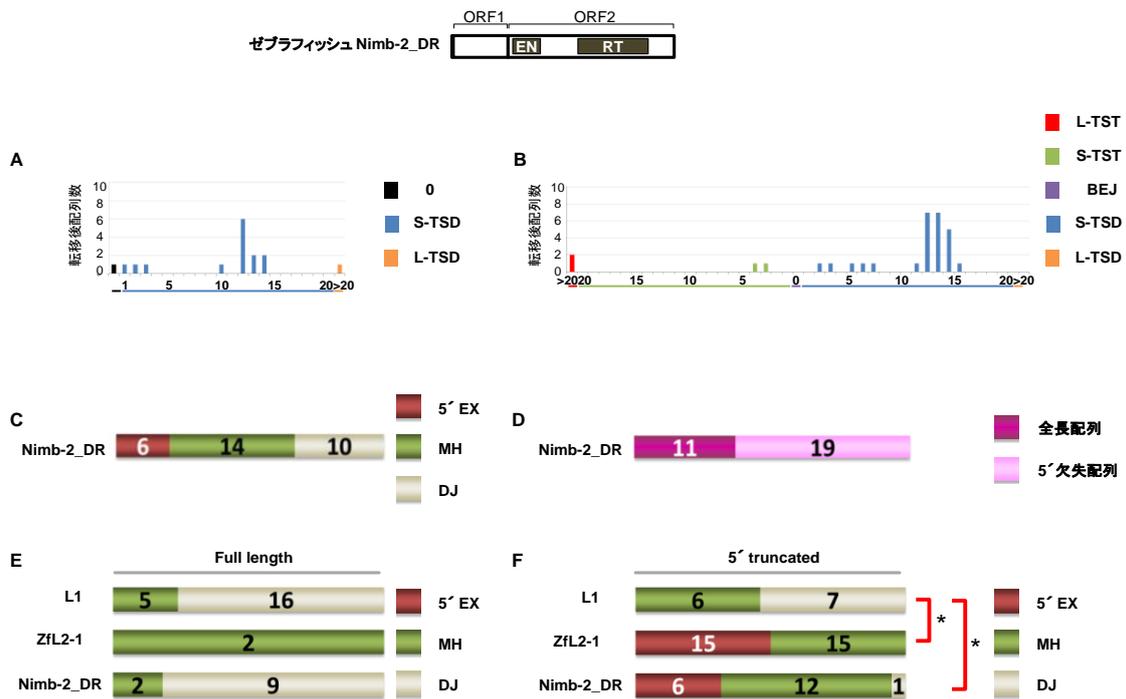


図 22. DT40 細胞中で転移したゼブラフィッシュ Nimb-2_DR は、アニーリングとダイレクト 2 つのタイプの 5' 末端連結の特徴を示す。

- A. ゼブラフィッシュゲノム中の Nimb-2_DR TSA の分布
- B. DT40 細胞中で転移した Nimb-2_DR TSA の分布
- C. DT40 細胞で転移した Nimb-2 DR の 5' 末端連結部 (5' EX、MH、DJ)
- D. DT40 細胞で転移した Nimb-2 DR の全長配列と 5' 欠失配列の比較
- E. DT40 細胞で全長配列が転移した L1、ZfL2-1Nimb-2_DR の 5' 末端連結部 (5' EX、MH、DJ)
- F. DT40 細胞で 5' 欠失配列が転移した L1、ZfL2-1Nimb-2_DR の 5' 末端連結部 (5' EX、MH、DJ)

統計解析は、フィッシャーの正確確率検定を行った。この検定の結果、 $P < 0.05$ の場合をアスタリスクで示した。図の最上部にゼブラフィッシュ Nimb-2 DR 構造の模式図を示す。

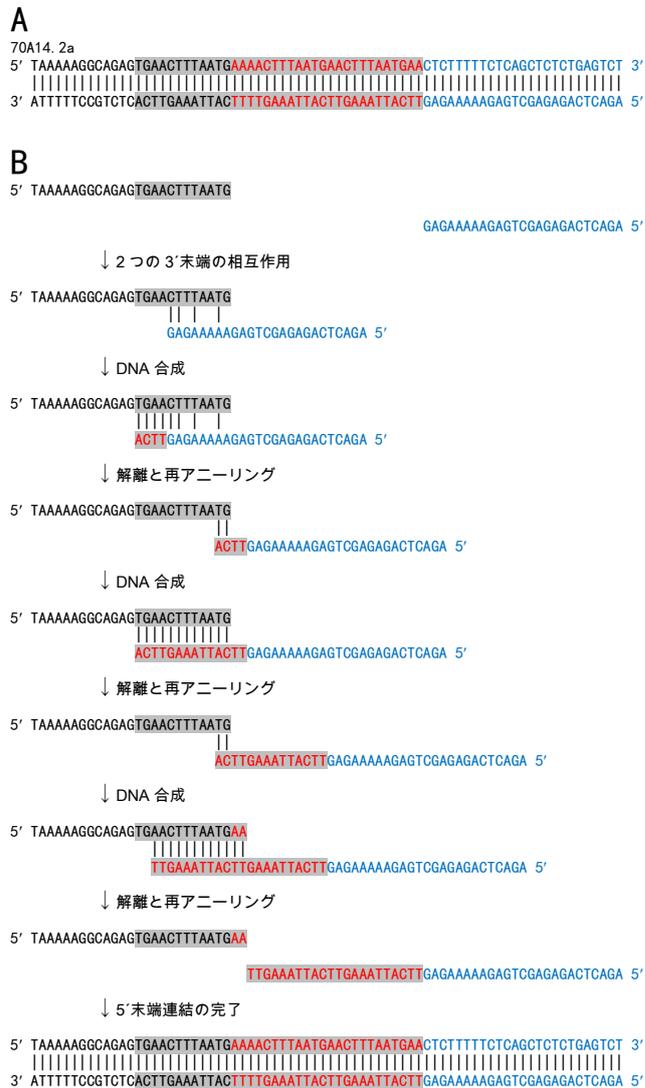


図 23. p70A14. 2a 配列の 5' EX 生成機構。表示方法は、図 12 と同様である。

A. p70A14. 2a 配列 5' 末端連結部

B. p70A14. 2a の 5' EX 生成機構

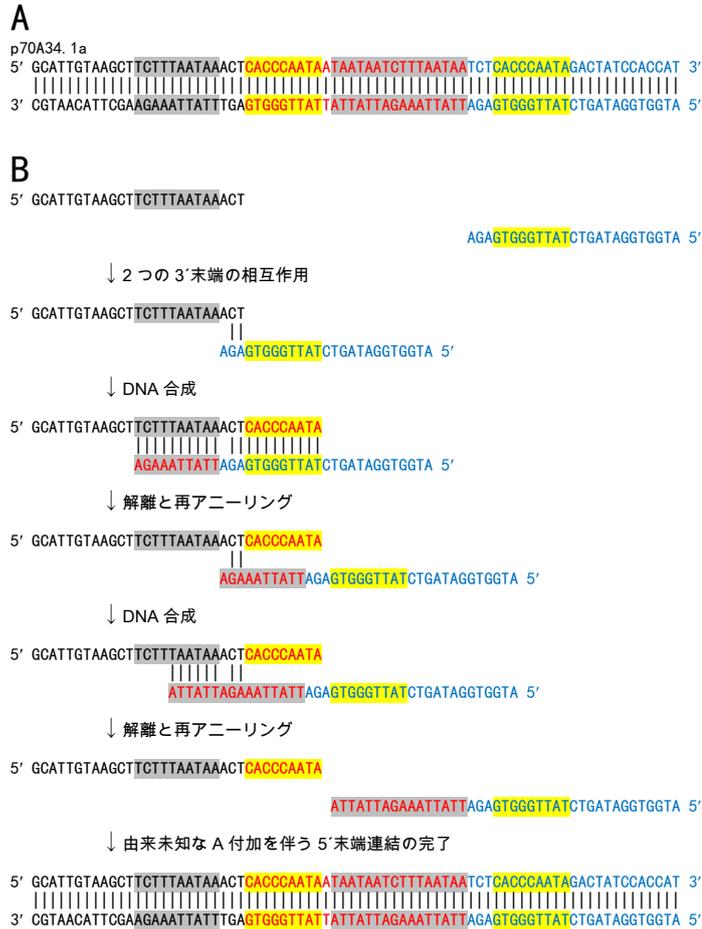


図 24. p70A34. 1a 配列の 5' EX 生成機構。表示方法は、図 12 と同様である。

A. p70A34. 1a 配列 5' 末端連結部

B. p70A34. 1a の 5' EX 生成機構

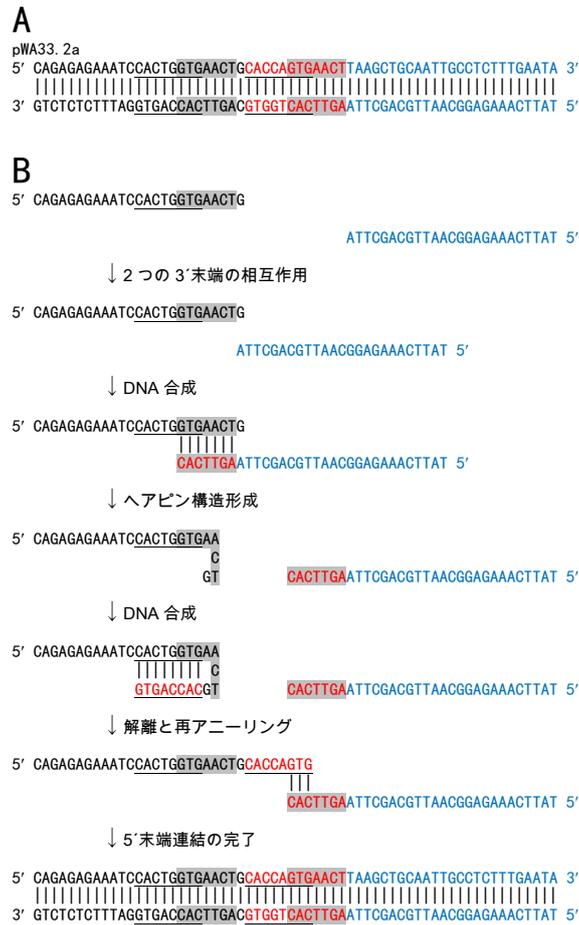


図 25. pWA33. 2a 配列の 5' EX 生成機構。表示方法は、図 12 と同様である。

A. pWA33. 2a 配列 5' 末端連結部

B. pWA33. 2a の 5' EX 生成機構

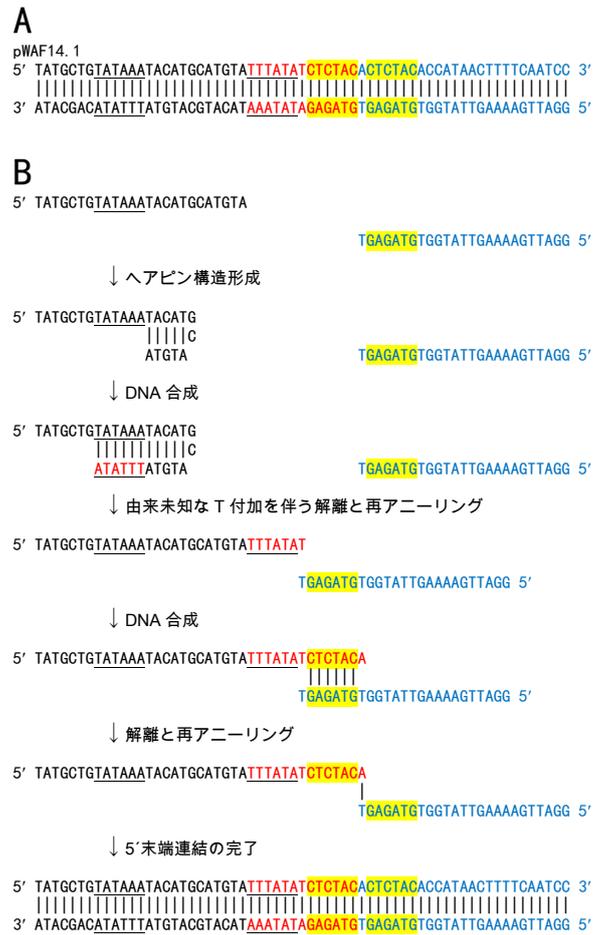


図 26. pWAF14.1 配列の 5' EX 生成機構。表示方法は、図 12 と同様である。

- A. pWAF14.1 配列 5' 末端連結部
- B. pWAF14.1 の 5' EX 生成機構

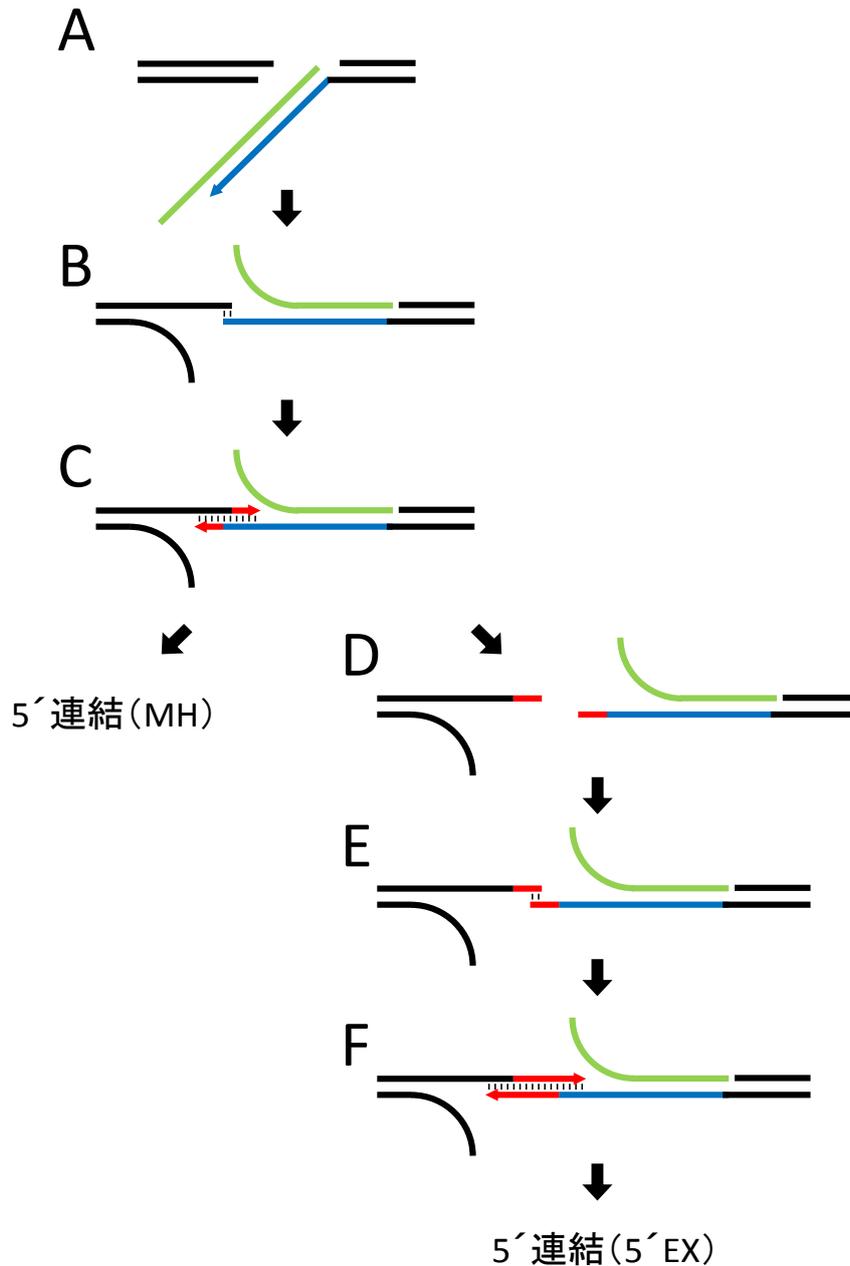


図 27. LINE の 5' 末端連結時の MH と 5' EX 生成機構モデル。
 図は宿主ゲノム DNA (黒)、LINE RNA (緑)、LINE cDNA (青)、新規に合成された DNA (赤) を示す。TPRT 後、二本鎖目のゲノム DNA が切断される。(A) 生じたゲノム DNA の 3' 突出末端と LINE cDNA の 3' 末端が数塩基のアニーリングを介して相互作用する。(B) その 3' 末端で DNA 合成が起きる。(C) この合成された相互作用部分が安定な場合、この連結部は宿主 DNA 修復系によって修復される。その結果、MH を伴った LINE の 5' 末端連結部となる。一方、この合成された相互作用部分が不安定な場合、この 3' 末端の相互作用部は解離し再度相互作用する (D, E)。その後、次の DNA 合成が起きる。その結果さらに安定な相互作用部となった時、連結部は宿主 DNA 修復系によって修復される。その結果、5' EX を伴った LINE の 5' 末端連結部となる。

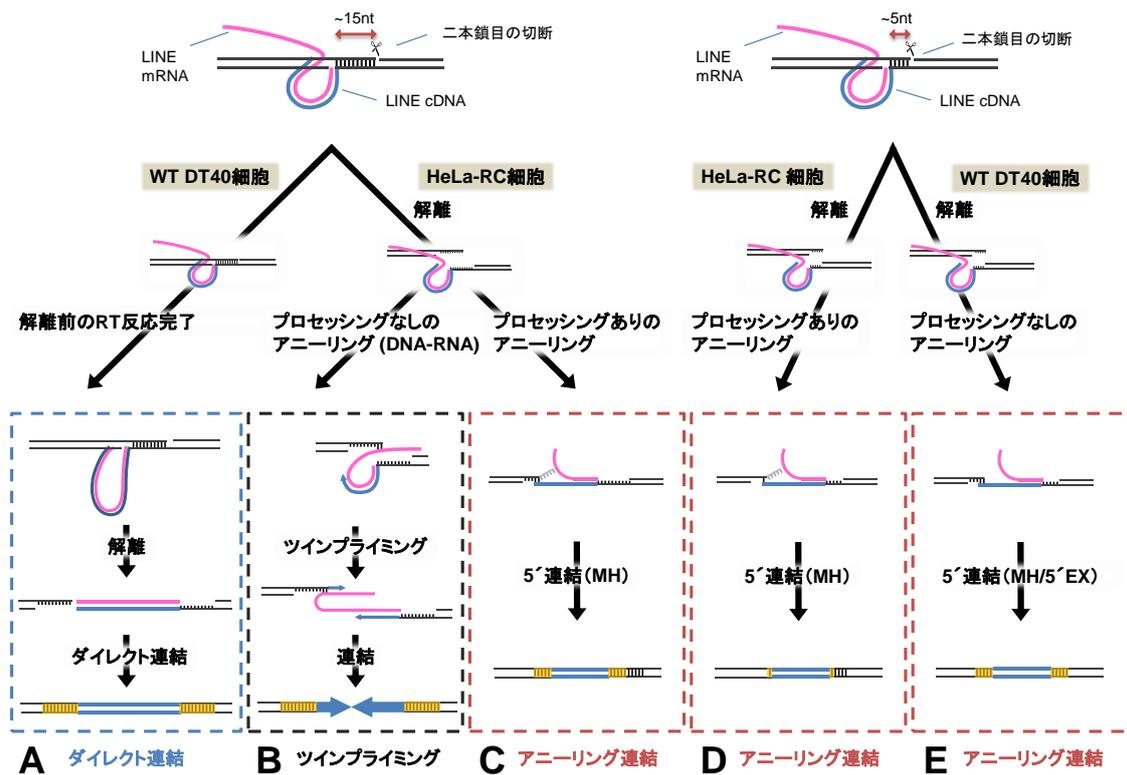


図 28. LINE の 5' 末端連結機構モデル。

LINE は 2 つの内 1 つの 5' 連結経路を通り連結される。2 回の標的ゲノム部位切断部位に挟まれている dsDNA 領域の解離しやすさや宿主因子によって、この 2 つの経路は支配されている。詳細な説明は、考察に記述した。

表 1. DT40 細胞中で転移した Zfl2-2 の 5' EX 配列の分類。

細胞種	全 5'EX (bp)	タイプ 1 (bp)	タイプ 2 (bp)	タイプ 3 (bp)	タイプ 4 (bp)	由来不明 (bp)
野生型 DT40 細胞	244 (100%)	120 (49%)	48 (19%)	22 (9%)	0 (0%)	54 (22%)
Ku70 欠損 DT40 細胞	114 (100%)	63 (55%)	11 (10%)	21 (18%)	0 (0%)	19 (17%)
Artemis 欠損 DT40 細胞	172 (100%)	41 (24%)	53 (31%)	0 (0%)	6 (3%)	72 (42%)
LigaseIV 欠損 DT40 細胞	112 (100%)	26 (23%)	32 (29%)	0 (0%)	0 (0%)	54 (48%)
合計	642 (100%)	250 (39%)	144 (22%)	43 (7%)	6 (1%)	199 (31%)

タイプ 1、タイプ 2、タイプ 3、タイプ 4 については結果 3-5. を参照。括弧内の数はそれぞれの 5' 配列の割合を示している。

表 2. 培養細胞中で転移した LINE 配列

Clones	LINE	Cell	LR	IL	5'MH	5'EX	3'MH	3'EX	TSA	Chr.	Direction	Position	Comment
					(bp)	(bp)	(bp)	(bp)	(bp)				
pWLE23.1a	L1.3	DT40 WT	An	8236	3	0	2	0	14	24	+	3225377	
pWLE25.1a	L1.3	DT40 WT	An	8238	0	0	2	0	11	2	+	72624351	
pWLE25.2a	L1.3	DT40 WT	An	8235	0	0	1	0	16	20	-	5095026	
pWLE25.3a	L1.3	DT40 WT	An	8235	0	0	2	0	15	1	+	75750183	
pWLE25.4a	L1.3	DT40 WT	An	8236	0	0	2	0	15	1	-	61623126	
pWLE27.7a	L1.3	DT40 WT	An	8235	*	*	*	*	*	4	-	44294761	
pWLE28.3a	L1.3	DT40 WT	An	8235	0	0	5	0	-32737	4	+	5016533	
pWLE28.5a	L1.3	DT40 WT	An	2250	0	0	1	0	15	20	-	5794983	
pWLE28.8a	L1.3	DT40 WT	An	4736	0	0	4	0	15	1	+	143556465	
pWLE30.5a	L1.3	DT40 WT	An	8235	0	0	0	0	15	2	-	134013415	
pWLE34.2a	L1.3	DT40 WT	An	4374	0	0	0	0	15	1	+	139303632	
pWLE34.3a	L1.3	DT40 WT	An	4132	3	0	1	0	17	3	-	72061595	
pWLE34.4a	L1.3	DT40 WT	An	2454	0	0	0	0	15	1	+	60201654	
pWLE34.5a	L1.3	DT40 WT	An	8235	0	0	0	0	10	5	+	58834606	
pWLE34.7a	L1.3	DT40 WT	An	8236	3	0	5	0	15	1	-	99439776	
pWLE35.1a	L1.3	DT40 WT	An	6734	0	0	3	0	13	1	+	45978730	
pWLE35.2a	L1.3	DT40 WT	An	2222	0	0	0	0	17	1	-	159097841	
pWLE35.3a	L1.3	DT40 WT	An	8235	0	0	2	0	12	4	-	36760632	
pWLE35.5a	L1.3	DT40 WT	An	8235	0	0	0	0	11	Z	+	10602247	
pWLE35.7a	L1.3	DT40 WT	An	3219	0	0	1	0	16	4	-	1449196	
pWLE35.8a	L1.3	DT40 WT	An	6042	1	0	6	0	17	2	-	76972240	
pWLE35.9a	L1.3	DT40 WT	An	8235	0	0	2	0	11	1	-	62334731	
pWLE36.1a	L1.3	DT40 WT	An	8235	0	0	3	0	12	1	-	11371195	
pWLE36.2a	L1.3	DT40 WT	An	8236	3	0	2	0	13	12	-	1910914	
pWLE36.5a	L1.3	DT40 WT	An	8235	1	0	2	0	15	5	-	38756270	
pWLE36.6a	L1.3	DT40 WT	An	6576	1	0	2	0	15	2	+	44445952	
pWLE36.7a	L1.3	DT40 WT	An	2611	1	0	0	0	16	11	+	3425508	
pWLE36.8a	L1.3	DT40 WT	An	4554	0	0	6	0	17	7	+	17905368	
pWLE36.9a	L1.3	DT40 WT	An	7243	3	0	2	0	12	2	+	33389906	
pWLE36.12a	L1.3	DT40 WT	An	8235	0	0	5	0	15	1	-	92147807	
pWLE36.10a	L1.3	DT40 WT	An	8235	2	0	2	0	17	1	-	46255029	
pWLE36.13a	L1.3	DT40 WT	An	8235	0	0	1	0	16	1	+	183124051	
pWLE36.14a	L1.3	DT40 WT	An	8235	0	0	0	0	13	1	+	52483201	

pWLE36.16a	L1.3	DT40 WT	An	8235	0	0	2	0	19	2	+	48979032	
pWLE36.20a	L1.3	DT40 WT	An	8235	0	0	2	0	12	2	-	125547059	
p7LE13.1a	L1.3	DT40 Ku70 ^{-/-}	An	2684	1	0	5	0	15	4	-	18206346	
p7LE13.2a	L1.3	DT40 Ku70 ^{-/-}	An	8234	0	0	1	0	16	2	+	137959525	
p7LE13.3a	L1.3	DT40 Ku70 ^{-/-}	An	8235	0	0	4	0	16	18	-	7704300	
p7LE13.5a	L1.3	DT40 Ku70 ^{-/-}	An	8236	2	0	7	0	18	2	-	95658631	
p7LE13.6a	L1.3	DT40 Ku70 ^{-/-}	An	2242	0	2	5	0	13	9	-	17241776	
p7LE14.1a	L1.3	DT40 Ku70 ^{-/-}	An	3557	0	0	1	0	16	8	+	3519554	
p7LE14.2a	L1.3	DT40 Ku70 ^{-/-}	An	4019	4	0	4	0	17	10	-	13239916	
p7LE14.3a	L1.3	DT40 Ku70 ^{-/-}	An	4576	1	0	2	0	16	12	-	14623953	
p7LE14.4a	L1.3	DT40 Ku70 ^{-/-}	An	8235	1	0	0	0	14	11	-	10437145	
p7LE14.5a	L1.3	DT40 Ku70 ^{-/-}	An	3650	0	0	5	0	16	3	+	9428477	
p7LE14.6a	L1.3	DT40 Ku70 ^{-/-}	An	2701	0	0	4	0	16	9	-	25387660	
p7LE15.1a	L1.3	DT40 Ku70 ^{-/-}	An	3830	0	0	0	0	14	2	-	109072699	
p7LE15.2a	L1.3	DT40 Ku70 ^{-/-}	An	2166	0	0	0	0	15	12	+	3983982	
p7LE15.3a	L1.3	DT40 Ku70 ^{-/-}	An	6386	0	0	5	0	14	2	-	137960349	
p7LE15.4a	L1.3	DT40 Ku70 ^{-/-}	An	2341	2	0	5	0	13	1	+	70209170	
p7LE15.5a	L1.3	DT40 Ku70 ^{-/-}	An	8236	2	0	2	0	11	4	-	44758241	
p7LE15.6a	L1.3	DT40 Ku70 ^{-/-}	An	2192	2	0	2	0	19	7	-	33829319	
p7LE16.1a	L1.3	DT40 Ku70 ^{-/-}	An	3820	2	0	4	0	15	2	-	16190585	
p7LE16.2a	L1.3	DT40 Ku70 ^{-/-}	An	8235	0	0	5	0	12	5	+	20628621	
p7LE16.4a	L1.3	DT40 Ku70 ^{-/-}	An	2311	0	0	5	0	13	6	+	20010622	
p7LE18.2a	L1.3	DT40 Ku70 ^{-/-}	An	8235	0	0	2	0	10	2	-	48941915	
p7LE18.6a	L1.3	DT40 Ku70 ^{-/-}	An	8235	0	1	0	0	15	2	-	130125877	
p7LE19.1a	L1.3	DT40 Ku70 ^{-/-}	An	8235	0	0	2	0	11	1	+	28353431	
p7LE19.3a	L1.3	DT40 Ku70 ^{-/-}	An	2147	1	0	0	0	20	3	-	99341940	
p7LE19.6a	L1.3	DT40 Ku70 ^{-/-}	An	7512	0	0	0	0	14	3	+	1502528	
pHLE1.2a	L1.3	HeLa-RC	An	8235	0	0	2	0	12	4	-	174264167	
pHLE1.4a	L1.3	HeLa-RC	An	3306	0	0	12	0	15	2	+	50237288	
pHLE1.5a	L1.3	HeLa-RC	An	2408	0	0	3	0	-1	3	+	140095592	
pHLE1.6a	L1.3	HeLa-RC	An	2272	6	0	4	0	-9	17	+	70254859	
pHLE2.3a	L1.3	HeLa-RC	An	2529	0	4	7	0	Translocation	3	+	116533961	Translocation chr3 + → chr2 +
pHLE2.5a	L1.3	HeLa-RC	An	2305	1	0	0	0	13	2	-	74067970	
pHLE2.9a	L1.3	HeLa-RC	An	2405	1	0	2	0	-3	4	-	27765199	
pHLE2.10a	L1.3	HeLa-RC	An	2552	0	0	2	0	2	6	-	158574145	

pHLE2.11a	L1.3	HeLa-RC	An	3379	1	0	1	0	6	12	+	60539559	
pHLE3.5a	L1.3	HeLa-RC	An	2638	2	0	0	0	0	5	+	12751859	
pHLE3.9a	L1.3	HeLa-RC	An	3039	3	0	0	0	14	7	-	125209937	
pHLE4.4a	L1.3	HeLa-RC	An	2299	2	0	3	0	5	15	+	54588828	
pHLE4.6a	L1.3	HeLa-RC	An	7737	0	0	0	0	13	5	+	33371381	
pHLE4.8a	L1.3	HeLa-RC	An	2642	2	0	3	0	Translocation	2	+	228280194	Translocation chr2 + -> chrX +
pHLE4.11a	L1.3	HeLa-RC	An	2427	1	0	5	0	1	20	-	17275693	
pHLE4.12a	L1.3	HeLa-RC	An	4515	0	0	1	0	16	21	+	10863481	
pHLE5.8a	L1.3	HeLa-RC	An	3402	0	0	2	0	-1111620	6	-	86521790	
pHLE5.9a	L1.3	HeLa-RC	An	4515	0	0	2	0	8	7	-	14282589	
pHLE5.10a	L1.3	HeLa-RC	An	2138	1	0	3	0	-5	1	-	50237288	
pHLE5.14a	L1.3	HeLa-RC	An	2474	2	0	3	0	11	13	-	99661284	
pHLE5.15a	L1.3	HeLa-RC	An	2307	1	0	2	0	3	6	-	100225182	
pHLE5.20b	L1.3	HeLa-RC	An	2510	0	2	4	0	-3	5	+	20358830	
pHLE5.21a	L1.3	HeLa-RC	An	2455	2	0	4	0	16	6	+	50874477	
pHLE6.2a	L1.3	HeLa-RC	An	2797	1	0	7	0	-12	8	-	78615108	
pHLE6.7a	L1.3	HeLa-RC	An	2889	3	0	4	0	6	18	+	62361366	
pHLE6.11a	L1.3	HeLa-RC	An	2324	3	0	5	0	11	7	+	51233595	
pHLE6.18a	L1.3	HeLa-RC	An	2276	0	0	4	0	6	5	-	118850203	
pHLE7.3a	L1.3	HeLa-RC	An	3362	0	0	2	0	14	9	-	76039525	
pHLE7.10a	L1.3	HeLa-RC	An	2393	0	4	0	0	4	X	-	23967100	
pHLE7.28a	L1.3	HeLa-RC	An	8235	0	0	0	0	11	6	-	110222597	
pHLE7.30a	L1.3	HeLa-RC	An	2616	0	U6snRNA	0	0	9	6	-	110222597	
pHLE8.2c	L1.3	HeLa-RC	An	2533	1	0	4	0	-6	2	+	232295820	
pHLE8.3a	L1.3	HeLa-RC	An	5135	0	0	2	0	2	13	-	57285681	
pHLE8.15a	L1.3	HeLa-RC	An	2228	0	1	2	0	3	5	-	72250784	
pHLE8.18a	L1.3	HeLa-RC	An	2282	1	0	4	0	4	4	+	43485686	
pHLE8.22a	L1.3	HeLa-RC	An	4931	4	0	2	0	2	7	-	114495454	
pHLE8.24b	L1.3	HeLa-RC	An	2521	1	0	0	0	1	11	-	41434114	
pHLE8.30a	L1.3	HeLa-RC	An	2630	2	0	0	0	10	12	-	47717375	
pHA1.1a	ZfL2-2	HeLa-RC	19	2439	5	0	0	0	2	8	+	128420122	
pHA1.5a	ZfL2-2	HeLa-RC	24	2421	2	0	1	0	Translocation	3	+	65883590	
pHA1.6a	ZfL2-2	HeLa-RC	24	2401	1	0	2	0	-4	9	-	139333224	
pHA1.7a	ZfL2-2	HeLa-RC	25	3154	5	0	4	0	10	1	+	65495199	
pHA1.9a	ZfL2-2	HeLa-RC	22	5472	2	0	3	0	6	11	+	41928450	
pHA2.1a	ZfL2-2	HeLa-RC	24	5497	1	0	2	0	-5	8	-	115469670	

pHA2.2a	ZfL2-2	HeLa-RC	31	2297	2	0	6	0	1	4	-	90791096
pHA2.3a	ZfL2-2	HeLa-RC	23	2705	0	4	2	0	1	8	+	127010612
pHA2.5a	ZfL2-2	HeLa-RC	23	2450	5	0	2	0	1646	22	+	40029338
pHA2.8a	ZfL2-2	HeLa-RC	23	2419	3	0	2	0	5	13	-	51112057
pHA2.11a	ZfL2-2	HeLa-RC	24	2750	0	0	4	0	-8	4	-	33871495
pHA3.4a	ZfL2-2	HeLa-RC	20	2662	2	0	2	0	-9	21	+	45552803
pHA3.5a	ZfL2-2	HeLa-RC	18	2487	4	0	1	0	6	22	-	29796462
pHA3.9a	ZfL2-2	HeLa-RC	25	2683	2	0	3	0	-1	9	+	135549475
pHA3.11a	ZfL2-2	HeLa-RC	26	3144	0	0	2	0	0	1	+	173187869
pHA4.2a	ZfL2-2	HeLa-RC	20	2748	*	*	*	*	*	17	+	22046402
pHA4.3a	ZfL2-2	HeLa-RC	26	2483	0	3	2	0	4	16	-	3144811
pHA4.6a	ZfL2-2	HeLa-RC	18	2735	2	0	1	0	6	17	-	49029297
pHA4.8a	ZfL2-2	HeLa-RC	23	2439	0	15	5	0	6	18	-	61132070
pHA4.10a	ZfL2-2	HeLa-RC	28	4174	0	0	4	0	6	6	-	64704132
pHA4.11a	ZfL2-2	HeLa-RC	22	2547	5	0	4	0	6	8	-	130802612
pWC1.1a	ZfL2-1	DT40 WT	39	5072	0	1	3	0	7	4	+	14885032
pWC1.2a	ZfL2-1	DT40 WT	29	2318	1	0	6	0	5	3	-	25282407
pWC4.1a	ZfL2-1	DT40 WT	23*	3661	0	2	1*	0*	6	19	-	6265350
pWC7.2a	ZfL2-1	DT40 WT	25	5606	2	0	3	0	6	8	-	2345122
pWC9.1a	ZfL2-1	DT40 WT	22	3281	2	0	4	0	1	1	-	144812675
pWC9.2a	ZfL2-1	DT40 WT	25	2557	0	11	4	0	7	4	-	23242359
pWC10.1a	ZfL2-1	DT40 WT	25	5509	5*	0*	1*	0*	*	*2	+	111993962
pWC10.2a	ZfL2-1	DT40 WT	15	4820	2	0	2	0	2	4	-	55056605
pWC10.4a	ZfL2-1	DT40 WT	23	5597	0	4	2	0	-2643	4	+	41775045
pWC10.6a	ZfL2-1	DT40 WT	18	6365	0	3	3	0	5	3	-	27147736
pWC11.2a	ZfL2-1	DT40 WT	26	5606	0	2	2	0	1	17	+	7803614
pWC11.6a	ZfL2-1	DT40 WT	23	5598	1	0	2	0	5	17	-	2451977
pWC12.1a	ZfL2-1	DT40 WT	22	6084	3	0	2	0	6	4	+	59499361
pWC13.1a	ZfL2-1	DT40 WT	22	5405	0	1	2	0	7	7	-	10415731
pWC14.1a	ZfL2-1	DT40 WT	23	3130	0	17	3	0	6	1	-	9179165
pWC12.4a	ZfL2-1	DT40 WT	20	5613	1	0	2	0	3	3	+	73285080
pWC12.5a	ZfL2-1	DT40 WT	21*	2333	0	8	1*	0*	8	3	+	71146657
pWC13.2a	ZfL2-1	DT40 WT	20	5594	2	0	1	0	1	14	+	8190351
pWC13.4a	ZfL2-1	DT40 WT	24	3792	1	0	1	0	4	4	+	3111857
pWC13.5a	ZfL2-1	DT40 WT	21	5676	3	0	2	0	2	2	-	101087287
pWC13.6a	ZfL2-1	DT40 WT	24	2253	0	20	4	0	5	1	+	186772972

pWC14.3a	ZfL2-1	DT40 WT	19	5592	0	4	1	0	5	5	-	10370532	
pWC14.5a	ZfL2-1	DT40 WT	20	5452	2	0	1	0	7	2	+	120175640	
pWC14.6a	ZfL2-1	DT40 WT	21	4736	0	43	2	0	6	1	+	101462742	
pWC15.1a	ZfL2-1	DT40 WT	19	2677	0	158	0	0	6	2	+	13951003	
pWC15.2a	ZfL2-1	DT40 WT	24	2925	0	18	2	0	8	1	-	165780575	
pWC15.3a	ZfL2-1	DT40 WT	18	6493	1	0	2	0	5	24	-	60761	
pWC16.1a	ZfL2-1	DT40 WT	19	2342	1	0	2	0	4	2	-	26064282	
pWC16.2a	ZfL2-1	DT40 WT	17	2516	2	0	2	0	-15	Z	+	21810624	
pWC16.3a	ZfL2-1	DT40 WT	24	4732	1	0	4	0	6	3	+	61311327	
pWC21.1a	ZfL2-1	DT40 WT	16	6365	0	11	2	0	2	5	+	46583457	内部配列の欠失
pWC21.3a	ZfL2-1	DT40 WT	24	6376	1	0	2	0	3	10	-	7394242	
pWC22.1a	ZfL2-1	DT40 WT	23	6494	1	0	3	0	3	15	-	3636795	
p7C2.1a	ZfL2-1	DT40 Ku70-/-	25	3271	0	2	1	0	-438	14	-	1093646	
p7C3.1a	ZfL2-1	DT40 Ku70-/-	19	2216	0	2	0	0	-206	Z	-	13282674	
p7C7.2a	ZfL2-1	DT40 Ku70-/-	20	2614	1	0	2	0	10	14	-	1093646	
p7C24.1a	ZfL2-1	DT40 Ku70-/-	20	2329	0	2	3	0	10	1	-	3986932	
p7C34.2a	ZfL2-1	DT40 Ku70-/-	29	5046	0	6	1	0	-1153	1	-	160093519	
p7C38.1a	ZfL2-1	DT40 Ku70-/-	21	4210	5	0	2	0	-1093	2	-	14679331	
p7C39.3a	ZfL2-1	DT40 Ku70-/-	11	6364	0	55	1	0	10	5	-	1394378	
p7C42.1a	ZfL2-1	DT40 Ku70-/-	21	4571	0	25	4	0	10	4	-	58092284	
p7C43.1a	ZfL2-1	DT40 Ku70-/-	20	6207	3	0	3	0	-157	15	-	8446157	
p7C44.1a	ZfL2-1	DT40 Ku70-/-	30	4226	0	30	1	0	7	6	+	17618477	
p7C45.1a	ZfL2-1	DT40 Ku70-/-	43	4605	0	0	2	0	-1424	3	+	66185500	
p7C47.1a	ZfL2-1	DT40 Ku70-/-	22	2787	0	8	4	0	3	1	-	100042704	
p7C48.1a	ZfL2-1	DT40 Ku70-/-	19	3468	2	0	0	0	0	3	-	8330631	
p7C49.1a	ZfL2-1	DT40 Ku70-/-	22	4322	0	31	2	0	0	8	-	17439933	
p7C50.1a	ZfL2-1	DT40 Ku70-/-	19	6494	2	0	1	0	1	4	+	64482117	
p7C51.1a	ZfL2-1	DT40 Ku70-/-	24	2210	0	3	3	0	2	3	-	79436066	
p7C52.1a	ZfL2-1	DT40 Ku70-/-	20	5602	3	0	1	0	-460	2	+	47868106	
p7C55.1a	ZfL2-1	DT40 Ku70-/-	25	5428	0	39	1	0	2	2	+	116023820	
p7C61.1a	ZfL2-1	DT40 Ku70-/-	25	5453	0	6	2	0	2	2	+	4607884	
p7C62.1a	ZfL2-1	DT40 Ku70-/-	26	6495	0	2	2	0	7	13	-	981739	
p7C63.1a	ZfL2-1	DT40 Ku70-/-	24	3084	0	0	1	0	-4422	12	-	4041036	
Ra102	ZfL2-1	HeLa-RC	26	6489	1	0	2	0	-8	13	+	90383371	
Ra103	ZfL2-1	HeLa-RC	35	3465	4	0	0	3	0	11	-	61758135	

Ra105	ZfL2-1	HeLa-RC	19	2663	2	0	3	0	3	9	-	122199727	
Ra112	ZfL2-1	HeLa-RC	16	2613	0	0	2	0	2	13	-	110221224	
Ra115	ZfL2-1	HeLa-RC	16	2563	2	0	2	0	-3	1	-	54784857	
Ra121	ZfL2-1	HeLa-RC	18	3779	3	0	2	0	6	10	-	23753319	
Ra123	ZfL2-1	HeLa-RC	21	2474	0	1	2	0	-3	6	+	1547787	
Ra124	ZfL2-1	HeLa-RC	20	2336	2	0	2	0	3	7	+	149475024	
Ra125	ZfL2-1	HeLa-RC	26	5507	0	0	3	0	4	1	+	68098583	
Ra126	ZfL2-1	HeLa-RC	19	2337	1	0	1	0	2	12	-	21303956	
Ra201	ZfL2-1	HeLa-RC	19	5577	2	0	1	0	-5	8	+	75006605	
Ra203	ZfL2-1	HeLa-RC	26	4063	5	0	3	0	1	*1	-	194598935*	
Ra205	ZfL2-1	HeLa-RC	22*	2957	2	0	2*	*0	5	12	+	6643780	*異常な3'末端リピート
Ra207	ZfL2-1	HeLa-RC	18	5449	1	0	2	0	2	2	+	135866793	
Ra227	ZfL2-1	HeLa-RC	25	2276	2	0	1	0	4	18	-	1538347	
Ra230	ZfL2-1	HeLa-RC	21	2464	3	0	2	0	-42	7	+	138849623	
Ra232	ZfL2-1	HeLa-RC	24	2888	0	0	2	0	-9	15	+	75924940	
Ra501	ZfL2-1	HeLa-RC	20	4194	4	0	1	0	4	8	+	2989176	
Ra506	ZfL2-1	HeLa-RC	24	5801	0	0	0	0	4	6	+	85414340	
Ra507	ZfL2-1	HeLa-RC	25	2335	1	0	1	0	4	5	-	124785563	
Ra510	ZfL2-1	HeLa-RC	25	3481	1	0	2	0	5	5	+		
Ra511	ZfL2-1	HeLa-RC	20	3042	3	0	1	0	4	19	-	49955315	
Ra512	ZfL2-1	HeLa-RC	19	6485	0	0	3	0	5	2	-	112252488	
Ra513	ZfL2-1	HeLa-RC	14	2918	2	0	0	0	9	1	+	1265095	
Ra514	ZfL2-1	HeLa-RC	19	2350	3	0	1	0	2	21	+	34821087	
Ra515	ZfL2-1	HeLa-RC	29	5500	2	0	3	0	-9	3	+	44446472	
Ra701	ZfL2-1	HeLa-RC	26	2385	2	0	3	0	5	4	-	174629124	
Ra705	ZfL2-1	HeLa-RC	21*	3128	0	40	1*	0*	-24	3	-	1539076	*異常な3'末端リピート
Ra707	ZfL2-1	HeLa-RC	26*	3086	2	0	1*	0*	1	20	-	49947545	*異常な3'末端リピート
Ra708	ZfL2-1	HeLa-RC	21*	2986	2	0	1*	0*	6	1	+	110228418	*異常な3'末端リピート
pWG2.3a	Nimb2	DT40 WT	21	7378	*0	0	2	0	14	1	-	14886468	
pWG2.4a	Nimb2	DT40 WT	23	2725	1	0	4	0	7	1	+	51373420	
pWG3.1a	Nimb2	DT40 WT	15	2041	0	15	2	0	11	12	+	11297602	
pWG3.2a	Nimb2	DT40 WT	18	6543	4	0	1	0	-21	5	+	49034845	
pWG3.3a	Nimb2	DT40 WT	18	2361	1	0	2	0	15	2	-	100409803	
pWG4.5a	Nimb2	DT40 WT	17	5958	0	3	1	0	12	2	-	27211870	
pWG5.1b	Nimb2	DT40 WT	26	7378	0	0	2	0	12	4	+	55631541	
pWG5.2a	Nimb2	DT40 WT	22	7378	0	0	2	0	-25	2	-	106046497	

pWG5.3a	Nimb2	DT40 WT	22	2903	0	4	3	0	13	1	+	17154081	
pWG5.4a	Nimb2	DT40 WT	18	3575	6	0	2	0	13	5	-	3874075	
pWG5.5a	Nimb2	DT40 WT	24	7378	0	0	2	0	12	20	-	7231360	
pWG6.3a	Nimb2	DT40 WT	21	7379	1	0	1	0	13	9	+	17176134	
pWG7.5a	Nimb2	DT40 WT	18	7378	0	0	2	0	13	2	-	151147006	
pWG7.2a	Nimb2	DT40 WT	23	6099	2	0	2	0	14	3	+	66396868	
pWG7.6a	Nimb2	DT40 WT	21	2563	0	3	1	0	-3	7	-	32379370	
pWG8.1a	Nimb2	DT40 WT	25	7378	*0	0	2	0	12	2	-	107240246	
pWG8.2a	Nimb2	DT40 WT	20	3725	2	0	2	0	-4	2	-	7962818	
pWG8.3a	Nimb2	DT40 WT	18	2600	0	10	4	0	14	4	+	58952821	
pWG9.2a	Nimb2	DT40 WT	23	3986	3	0	5	0	6	3	-	24931310	
pWG9.3a	Nimb2	DT40 WT	21	2351	2	0	4	0	12	3	+	7962818	
pWG10.1a	Nimb2	DT40 WT	13	7378	2	0	2	0	13	4	-	59955414	
pWG10.3a	Nimb2	DT40 WT	19	7378	*0	0	2	0	12	10	-	7795479	
pWG11.2a	Nimb2	DT40 WT	21	4450	1	0	4	0	14	1	-	79972434	
pWG12.2a	Nimb2	DT40 WT	19	2440	0	6	2	0	14	12	+	19784079	
pWG13.1a	Nimb2	DT40 WT	17	6407	3	0	2	0	2	13	+	3406951	
pWG13.2a	Nimb2	DT40 WT	19	7379	0	0	0	0	13	18	-	3093921	
pWG14.1a	Nimb2	DT40 WT	22	7378	*0	0	2	0	12	1	-	172048449	
pWG14.2a	Nimb2	DT40 WT	21	3566	3	0	6	0	13	2	-	15111955	
pWG15.1a	Nimb2	DT40 WT	14	2453	3	0	*0	*4	3	15	+	2079463	
pWG16.1a	Nimb2	DT40 WT	21	2730	0	0	3	0	5	8	+	7931493	
pHLE2.1a	L1.3	HeLa-RC	An	2947	2	0	1	0	13	5	-	28052433	5' 逆位, 5'逆位 LINE 長 (1080 bp)
pHLE3.8a	L1.3	HeLa-RC	An	3009	2	0	0	0	16	9	-	3034971	5' 逆位, 5'逆位 LINE 長 (1288 bp)
pHLE3.11a	L1.3	HeLa-RC	An	6046	3	0	1	0	17	7	+	76875115	5' 逆位, 5'逆位 LINE 長 (2043 bp)
pHLE4.2a	L1.3	HeLa-RC	An	7212	1	0	3	0	15	8	+	109437894	5' 逆位, 5'逆位 LINE 長 (524 bp)
pHLE4.5a	L1.3	HeLa-RC	An	219	0	0	0	0	14	5	-	152282810	5' 逆位, 5'逆位 LINE 長 (1959 bp)
pHLE2.6a	L1.3	HeLa-RC	An	4901	2	0	3	0	17	16	+	66873766	5' 逆位, 5'逆位 LINE 長 (1332 bp)
pHLE4.9a	L1.3	HeLa-RC	An	2230	3	0	0	0	10	2	-	154790258	5' 逆位, 5'逆位 LINE 長(4017 bp)
pHLE5.16a	L1.3	HeLa-RC	An	843	3	0	2	0	15	X	+	29334865	5' 逆位, 5'逆位 LINE 長 (1395 bp)
pHLE6.19a	L1.3	HeLa-RC	An	1220	6	0	0	0	14	11	-	54910101	5' 逆位, 5'逆位 LINE 長 (953 bp)
pHLE7.8a	L1.3	HeLa-RC	An	828	5	0	1	0	18	5	-	107821707	5' 逆位, 5'逆位 LINE 長 (2017 bp)
pHLE7.20b	L1.3	HeLa-RC	An	-204	0	0	0	0	5	5	+	110222597	5' 逆位, 5'逆位 LINE 長 (2870 bp)
pHLE8.12a	L1.3	HeLa-RC	An	3554	2	0	3	0	13	11	+	101662258	5' 逆位, 5'逆位 LINE 長 (177 bp)
pHLE8.25a	L1.3	HeLa-RC	An	43	4	0	0	0	17	5	-	36306098	5' 逆位, 5'逆位 LINE 長 (2458 bp)

LR は、3' 末端リピートの長さ (bp) を示す。LR 中で表示されている An は長さ未知のポリ

Aを示す。ILはLINE配列長 (bp) を示す。5´ MHは5´ MHの長さ (bp) を示す。length of the 5´ microhomology. 5´ EXは5´ EXの長さ (bp) を示す。3´ MHは3´ MHの長さ (bp) を示す。3´ EXは3´ EXの長さ (bp) を示す。TSAのプラスの数字はTSD、ゼロは、BEJ、マイナスの数字はTSTの長さ (bp) を示す。Chrは、LINE転移が起きた染色体番号、Direction (+/-)は、LINE配列の挿入向き、PositionはLINE転移が起きた染色体中の位置を示す。NDは決定できなかったものを示す。

参考資料

培養細胞中で転移した LINE の 3' および 5' 末端配列

上部の配列は、LINE の 5' 末端配列を示す。中央の配列は、LINE 転移前のゲノム DNA 配列を示す。ヒトゲノムは、'hs'、ニワトリゲノムは、'gg' で表されている。染色体番号と染色体の向き (+/-) をその右に示す。下部の配列は、LINE の 3' 末端配列を示す。配列はそれぞれ、LINE (黒)、5' 逆位領域 (紫)、5' EX (赤)、TSD (緑)、TST (橙)、3' 末端繰り返し配列 (下線)、異常な 3' 末端繰り返し配列 (黄)、ホモロジー (アスタリスク: *)、MH (灰色塗りつぶし)、HindIII 認識配列 (イタリック) を示す。

L1.3 / DT40 WT cells

pWLE23. 1a
5' end GGGCTGGCCCTTTGTAGTAAGACGAATTTGAGATCTCTAGAAGCTGGGTACCAGCTGCT
gg_chr24+ GGGCTGGCCCTTTGTAGTAAGACGAATTTGAGATCTCTAGAAGCTGGGTACCAGCTGCT
3' end AAAAAAAAAAAAAAAAAAGACGAATTTGAGCAAGATTTGCGAATAACACCT

pWLE25. 1a
5' end ACATGCTTTTAAGACGATAGACGTGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCTAGCTTGC
gg_chr2+ ACATGCTTTTAAGACGATAGAACTGTAACCTGTAGTTTGTGCTGGCCAAAGAAGTTTCTCCTAAAAG
3' end AAAAAAAAAAGAGCATAGAACTGTAACCTGTAGTTTGTGCTGGCCAAAGAAGTTTCTCCTAAAAG

pWLE25. 2a
5' end TTTGTTTGACCAACAGAACACAATTTGTTAGATCTCTAGAAGCTGGGTACCAGCTGCTAG
gg_chr20- TTTGTTTGACCAACAGAACACAATTTGTTAGATCTCTAGAAGCTGGGTACCAGCTGCTAG
3' end AAAAAAAAAAGAACACAATTTGTTAGCAAGAAACACCTGGGTTTGTACATG

pWLE25. 3a
5' end GAATATCCACAATTTCTACAGCTCCCAATGAAATGCATTTGGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCTAGC
gg_chr1+ GAATATCCACAATTTCTACAGCTCCCAATGAAATGCATTTGTTTCAGGGTGTGTTTGGATCTGAGAGACAGGATGTCTCAAG
3' end AAAAAAAAAAAAAAAAAAGTAAATGCATTTGTTTCAGGGTGTGTTTGGATCTGAGAGACAGGATGTCTCAAG

pWLE25. 4a
5' end CCAAGAAAAGACATGGCGCAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCTAGCTTGTAGC

gg_chr1- CCAAGAAAAGACATGGCAGCTTTCTTTTTATGTCAAAAGCTGACTGAAATCAATATTAGTCTA

3' end AAAAGAAAAGACATGGCAGCTTTCTTTTTATGTCAAAACCTGACTGAAATCAATATTAGTCTA

pWLE27. 7a
5' end AGTGTAAAAAGAAAAAGAAAAAAGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAA

gg_chr4- AGTGTAAAAAGAAAAAACAACAACCCCAAAAACAGTAAACAACAAATCTAACTC

3' end AAAAAAGAAAAAAGAAAAAACAACAACCCCAAAAACAGTAAACAACAAATCTAACTC

pWLE28. 3a
5' end ACTTTTTGTGAGC7TTCAGCACCCGATCTGCAGCCAGGACCCCTGATGTCAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCTA

gg_chr 4 + GAAAAGGGAGAGC7TTCAGCACCCGATCTGCAGCCAGGACCCCTGATGTCAGAAATAGATTTTGCCCAAGGGCTCATGGAGGGTTGCTA
4983796
5016533
gg_chr4 + GGGGATGCATCAAAATTTGCAAGATAGTAAAAACAATGGAGAGAAGCGATGCTTATGAGTGGCTGTTTCAGGCTATTTTTCCACAAGGG

3' end AAAAAAGAAAAAAGAAAAAACAACAATGGAGAGAAGCGATGCTTATGAGTGGCTGTTTCAGGCTATTTTTCCACAAGGG

pWLE28. 5a
5' end TTACACTGCATCTGACATCTTAGACTCAGCACTAATCAGAACTTCATGTGCTTAGTGGGTGCAGCCAGCATGGCACATGTATAC

gg_chr20 - TTACACTGCATCTGACATCTTAGACTCAGCACTAATCAGAACTTCATGTGCTTAGTGGGTGCAGCCAGCATGGCACATGTATAC

3' end AAAAAAGAAAAAAGAAAAAACAACAATGGAGAGAAGCGATGCTTATGAGTGGCTGTTTCAGGCTATTTTTCCACAAGGG

pWLE28. 8a
5' end AGAAGAGAGAGCGACGAGAAAAATCCCTATAAAAATTCCTCTGGAAGGAAGTGTACCATTCTCTGAAACTATCCAAATCAATA

gg_chr1 + AGAAGAGAGAGCGACGAGAAAAATCCCTATAAAAATTCCTCTGGAAGGAAGTGTACCATTCTCTGAAACTATCCAAATCAATA

3' end AAAAAAGAAAAAAGAAAAAACAACAATGGAGAGAAGCGATGCTTATGAGTGGCTGTTTCAGGCTATTTTTCCACAAGGG

pWLE30. 5a
5' end CTTGTGAAAGGTTAATACTTCTCTTTAAAAACAATAATGCGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCTAGCTTGC

gg_chr2 - CTTGTGAAAGGTTAATACTTCTCTTTAAAAACAATAATGCGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCTAGCTTGC

3' end AAAAAAGAAAAAAGAAAAAACAACAATGGAGAGAAGCGATGCTTATGAGTGGCTGTTTCAGGCTATTTTTCCACAAGGG

pWLE34. 2a
5' end CTGGACAGGCTGGAGAAAGTGGGCCATGAAAGCCTAATGAGGAAATCAACAACCCCTCATGCTAAAACTCTCAATAAATTAGT

gg_chr1 + CTGGACAGGCTGGAGAAAGTGGGCCATGAAAGCCTAATGAGGAAATCAACAACCCCTCATGCTAAAACTCTCAATAAATTAGT

3' end AAAAAAGAAAAAAGAAAAAACAACAATGGAGAGAAGCGATGCTTATGAGTGGCTGTTTCAGGCTATTTTTCCACAAGGG

pWLE34. 3a
5' end TGTGGAGTCTGGCAGGGCAATCAGGCAGGAGAAGAAATAAGGGTATTCAATTAGAAAAGAGGAAGTCAAATTGTCC

gg_chr3 - TGTGGAGTCTGGCAGGGCAATCAGGCAGGAGAAGAAATAAGGGTATTCAATTAGAAAAGAGGAAGTCAAATTGTCC

3' end AAAAAAGAAAAAAGAAAAAACAACAATGGAGAGAAGCGATGCTTATGAGTGGCTGTTTCAGGCTATTTTTCCACAAGGG

pWLE34. 4a
5' end AAGCTGTTATTGATCTTCAAGTTGCTTTGAAAAGCAATGTTAGTGGATGAAATGGAAACCATCTTCTCAGTAA

gg_chr1+ AAGCTGTTATTGATCTTCAAGTTGCTTTGAAAAGCAATGTTAGTGGATGAAATGGAAACCATCTTCTCAGTAA

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAGCAATGTTAGATGGCCAAATTCACAAGCACTACTTTCTC

pWLE34. 5a
5' end TCAATATCAATATAACAGTAGGTTGAAAGAACAAAGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGC

gg_chr5+ TCAATATCAATATAACAGTAGGTTGAAAGAACAAAGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGC

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAGAACAAAGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGC

pWLE34. 7a
5' end AAAACTATCTTAATTCCTTAAAAACACAACACAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAG

gg_chr1 - AAAACTATCTTAATTCCTTAAAAACACAACACAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAACAACAACACAGCTCTTGAAGCTAGAAGAACAATTTGGTAAGAAA

pWLE35. 1a
5' end GCATTTACAATGTTTATTAAATCTATACACATATTCCAGGAGAAGCTCCCAATATAGCAA

gg_chr1 + GCATTTACAATGTTTATTAAATCTATACACATATTCCAGGAGAAGCTCCCAATATAGCAA

3' end AAAAAAAAAAAAAAAAAAAAAAAAAATCTATACACAGAAAGTTGAAGTAGATGTTATTTTCATCGTG

pWLE35. 2a
5' end AAGTTGCTACTGAGCATCTGTGCTAAAAATGGTGAAGTATCATGTATACATATGTAACCTAACCCGATCCGAAACAA

gg_chr1 - AAGTTGCTACTGAGCATCTGTGCTAAAAATGGTGAAGTATCATGTATACATATGTAACCTAACCCGATCCGAAACAA

3' end AAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAATGGTGAAGTATGAGAGACTGTATCCATGAAGTCACACTATATTG

pWLE35. 3a
5' end AAGCCTCAGTCTGCTAAAAATCAACAGCATTGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCTAGCTGCTAGCGGGCGGGGG

gg_chr4 - AAGCCTCAGTCTGCTAAAAATCAACAGCATTGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCTAGCTGCTAGCGGGCGGGGG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAACAGCATTGATCCATTTTAGAGCATAGAAGCTATTAAAGACTACATTAGAACACAATAAAAAACATTTA

pWLE35. 5a
5' end AAAACATAGGGAAGATTATTGAATTATGAAAATAAATCGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAG

gg_chrZ + AAAACATAGGGAAGATTATTGAATTATGAAAATAAATCGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAGAAAATAAATCGAGGACACAGGCAGGACTTTATTGCCAGCCTGGAA

pWLE35. 7a
5' end AAGAGAGCATCTGCTCAGGGAGTAGAAAGCCAGTGCCTGAGGACTTCTATGCTCCAAAACACCAAAAGCAATG

gg_chr4 - AAGAGAGCATCTGCTCAGGGAGTAGAAAGCCAGTGCCTGAGGACTTCTATGCTCCAAAACACCAAAAGCAATG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGCCAGTGCCTGAGGACTTCTATGCTCCAAAACACCAAAAGCAATG

pWLE35. 8a
5' end AACTTTGGTTGTTTCATCTCCCTTAAAAACATATATTGACACATAGGCTCAAAATAAAGGATGGAGGAAGATCTACCAAG

gg_chr2 - AACTTTGGTTGTTTCATCTCCCTTAAAAACATATATTGAGCTTATTAATATTTTAGATTATTCACCTGCTTCATTTCGAAC

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATATTTGAGTTATTAATATTTTAGATTATTCACCTGCTTCATTTCGAAC

pWLE35. 9a
5' end GCTAGCCATGTATTTGGTCTTGGTTGAAAGAGAAGAAATGCAGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCTAGCTTGC

gg_chr1 - GCTAGCCATGTATTTGGTCTTGGTTGAAAGAGAAGAAATGCAGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCTAGCTTGC

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAATGCAGTATCTTTTCACTCTGGTGCAGTCTCAAAATTTCTAGCTA

pWLE36. 1a
5' end GTCTAATTAAAAGGCTGAACAGTTTCAGAAAGAAATGGTTAAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCT

g_chr1 - GTCTAATTAAAAGGCTGAACAGTTTCAGAAAGAAATGGTTAAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCT

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGATGGTTAGCTTCTGAGAAAGTGATATTCTAAATCAGAAAAGCAG

pWLE36. 2a
5' end GCATGTGCTCAGGTATACAGCTGACAACCTTCTTAAAGAGAAGCCAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAG

gg_chr12 - GCATGTGCTCAGGTATACAGCTGACAACCTTCTTAAAGAGAAGCCAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAAAGCCAGCACTGGAAAAGGGGAACTCAACTATGGATTTT

pWLE36. 5a
5' end TATTTTCCTTTCAGAAAGTGTGAGGTCAAGAAATATATAGTAGATCTCTAGAAGCTG

gg_chr5 - TATTTTCCTTTCAGAAAGTGTGAGGTCAAGAAATATATAGTAGATGCCCCAACCTTAT

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAATATATAGTAGTCCCAACCTTAT

pWLE36. 6a
5' end TTGCTACCCGTCAGAGTCAATATGCCAAACGCAGCCAGAGAGAAAGTCCGGTTACCCCTCAAAGGAAAGCCCA

gg_chr2 + TTGCTACCCGTCAGAGTCAATATGCCAAACGCAGCCAGAGAGAAAGTCCGGTTACCCCTCAAAGGAAAGCCCA

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAATATGCCAAACGCAGCCAGAGAGAAAGTCCGGTTACCCCTCAAAGGAAAGCCCA

pWLE36. 7a
5' end CTGTAGTCTGTGACAATAGTTGTGAAAGTACTCTGTGGAGCTATGTTATTGCGGCACTATTACAAATAGCAA

gg_chr11 + CTGTAGTCTGTGACAATAGTTGTGAAAGTACTCTGTGGAGCTATGTTATTGCGGCACTATTACAAATAGCAA

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGTACTCTGTGGAAGTGGTCTTAAACACCGCAACACTTTGAGTCA

pWLE36. 8a
5' end GCTCTATCAGAGTTGAAGCAATTAATAAACCAGTATAACTGGCAACCGAATCCAGCAGCACATCAAAAACTTATCCA

gg_chr7 + GCTCTATCAGAGTTGAAGCAATTAATAAACCAGTATAACTGGCAACCGAATCCAGCAGCACATCAAAAACTTATCCA

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAACCAGTATAACATAACCGATTGAATATATGAGCATACATTGCTCTATTCTT

pWLE36. 9a
5' end GACCCAGATTTTCAGTTCCTGTTAATAAGAAGTTCAAACTCTAAAAGCGAGCGGCTCTCCTCCTC

gg_chr2 + GACCCAGATTTTCAGTTCCTGTTAATAAGAAGTTCAAAAGGATAGATTATAAAGCTTTACCTACCA

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGTTCAAAGCAGTAGATTATAAAGCTTTACCTACCA

pWLE36. 10a
5' end GTAGTTTGCCTTGTATAAGGCATGATCAAGAACAAGGCTCARGATCTCTAGAAGCTGGGTACCAGTGCTAGCAAGTAGCT

gg_chr1 - GTAGTTTGCCTTGTATAAGGCATGATCAAGAACAAGGCTCARGCTTGGTGTGGTTAGGGTTACTTTCTTAGATAAAAATC

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAGAACTAAGGCTC.AAGCTTAAAGACCTTTGCTGAGCAAGAAGTGAATGGTCACA

pWLE36. 12a
5' end GGGATCAGTAACCAAAACACAACCTCCTCACCACAAACAATATGATGAGATCTCTAGAAGCTGGGTACCAGTGCTAGCAAGCTAGCTTGCTA

gg_chr1 - GGGATCAGTAACCAAAACACAACCTCCTCACCACAAACAATATGATGAGATCTCTAGAAGCTGGGTACCAGTGCTAGCAAGCTAGCTTGCTA

3' end AAAAAAAAAAAAAAAAAAAAAAAAAACAATATGATCTGTTGTAAGACACTAGGGTGTGCACCTCATTGTGTAAGACTGAG

pWLE36. 13a
5' end CCAAAGTAGCCAGAGTCTGCTAGTTTCCTTGAACACAGCAGGGGAGAGATCTCTAGAAGCTGGGTACCAGTGCTA

gg_chr1 + CCAAAGTAGCCAGAGTCTGCTAGTTTCCTTGAACACAGCAGGGGAAACAGTCAGATTAAACAGTATTAACGTCAT

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAGAAACAGCAGGGGAAACAGTCAGATTAAACAGTATTAACGTCAT

pWLE36. 14a
5' end AAAGTGAAATCTGGTATAAATCAACTTGAACACCAGTCTGAGATCTCTAGAAGCTGGGTACCAGTGCTAGCAAGTAGCT

gg_chr1 + AAAGTGAAATCTGGTATAAATCAACTTGAACACCAGTCTGTTTCATTTGAAAGGAAAAAATGCAAGTCTATGTGTCAG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAGAAACACCAGTCTGTTTCATTTGAAAGGAAAAAATGCAAGTCTATGTGTCAG

pWLE36. 16a
5' end AAATGATTTTATGATCTATGACAAATTAAGAAATATAATAATCCGAGATCTCTAGAAGCTGGGTACCAGTGCTAGC

gg_chr2 + AAATGATTTTATGATCTATGACAAATTAAGAAATATAATAATCCGAGATCTCTAGAAGCTGGGTACCAGTGCTAGC

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAGAAATAATAATAATCCATAAATGCCACTCATCAAGCAAAGGATGCTTA

pWLE36. 20a
5' end CAATTCAGTTGAAGATTTCTAGTCAATTAAGATCTTTAGAGATCTCTAGAAGCTGGGTACCAGTGCTAGCAAGC

gg_chr2 - CAATTCAGTTGAAGATTTCTAGTCAATTAAGATCTTTAGAGATCTTTAGCATTATTTTACCTTTACAATATAATCTGTTCAC

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAGATCTTTAGCATTATTTTACCTTTACAATATAATCTGTTCAC

p7LE14. 4a
5' end ATCAATTAATAACATAAAGGGTTGATGAAAGTAAGTGAAGATCTCTAGAAAGCTGGGTACCAGCTGTAGCAAG

gg_chr11- ATCAATTAATAACATAAAGGGTTGATGAAAGTAACTGAGATGCTTATGTTGTTAATTTTTATAAAAT

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGTAACTGAGATGCTTATGTTGTTAATTTTTATAAAAT

p7LE14. 5a
5' end CACATAAACACAACAATTTTCATAGCTGCCAAAAATCAATGAACCTATGGAACCAAAAA

gg_chr 3 + CACATAAACACAACAATTTTCATAGCGCTCCAAAAATCAGATGACCAAGAGAGATGTACTGC

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAATCAATGAACCAAGAGATGTACTGC

p7LE14. 6a
5' end TTAGGAGAGAATTTGTATTCTGGAATCAAAGACTAGAATCCGAACTAGAAATACCATTTGACC

gg_chr 9 - TTAGGAGAGAATTTGTATTCTGGAATCAAAGACTAGAAATCCATTTAAGAGCTTGGGATGAAGAC

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAGACTAGAATCCATTTAAGAGCTTGGGATGAAGAC

p7LE15. 1a
5' end CCACCAACAAATACAGAGAACATATCCATTTTGAAGAGCTGSGATGCTCAAGGAATAAAAGAGGACACAAACAATGGAAG

gg_chr2- CCACCAACAAATACAGAGAACATATCCATTTTGAAGAGCTGSGAAATTTGAAAAATCTTCTCAAAAACAGTAGCATTGTAT

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGCTGTGAAAAATTTGAAAAATCTTCTCAAAAACAGTAGCATTGTAT

p7LE15. 2a
5' end TCCAGAGGTGCTCTGTTGATATGGCCTTGAAGATGACAGATAGTCTTTTATGCGGATCCCTCAGAAGAAGCTGTCAGAAGG

gg_chr12+ TCCAGAGGTGCTCTGTTGATATGGCCTTGAAGATGACAGATAAATTTTAGTGTTCAGGAATTGCTGCTGGCTGTTACTCAGTCT

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAATGACAGATAAATTTTAGTGTTCAGGAATTGCTGCTGGCTGTTACTCAGTCT

p7LE15. 3a
5' end TGTGTTGATGATCAAAATCTCAGCTACAAAAATCTATTTTTACTTTATAGACAAGCAATGTTGAGAGATTTGTACCACCAGGCC

gg_chr2- TGTGTTGATGATCAAAATCTCAGCTACAAAAATCTATTTTTAACATAGTTACCACCACAGCTATGCATTTTGCAGTGATGAACAA

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAATCTATTTTTAACATAGTTACCACCACAGCTATGCATTTTGCAGTGATGAACAA

p7LE15. 4a
5' end GGTAAAGTCCATTTCCAGAAAAGTGTCTAGACTTGATTTAAAACTCCAGAGGGGAATATCACACTGTGGGACTGTGGGGTGGGGGAGGGGG

gg_chr1+ GGTAAAGTCCATTTCCAGAAAAGTGTCTAGACTTGATTTAAAACTCCAGAGAGAAAATTTGCTGTTTCTGTATAAGTCTATCCAAATCATAATG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAATCTCCAGAAAAGTGTCTAGACTTGATTTAAAACTCCAGAGAGAAAATTTGCTGTTTCTGTATAAGTCTATCCAAATCATAATG

p7LE15. 5a
5' end GCAGAAAATCCCTGTACATTCAGACAAGTAACAAGAAATCACAGATCTCTAGAAAGCTGGTACCAGC

gg_chr4- GCAGAAAATCCCTGTACATTCAGACAAGTAACAAGAAATCACAGATGCACTGTAATGACAGTAAATTC

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAGAAATCACATGCACTGTAATGACAGTAAATTC

p7LE15. 6a
5' end GCATGTAATTTGGGT TAAGAAAAAAGCGAGA CAAAACGACCCAAACCCGCGTTTTATTCTG

gg_chr7- GCATGTAATTTGGGTT AAGAAAAAAGCGAGAATAATATTTTGATGCATCTCCGTGCAAGTGTGA

3' end AAAAAAAAAAAAAAAAAAAAGAAAAATAAGCGAGAAATAATATTTTGATGCATCTCCGTGCAAGTGTGA

p7LE16. 1a
5' end GGGTTT ACCATTAAGTGGAGAAGTAACAGCAAAAGATGATTTCA TAAAAGAGGACACAAACAATGGAAGACATTCATG

gg_chr 2 - GGGTTT ACCATTAAGTGGAGAAGTAACAGCA AAAGATGATTCAAGACCAATTTCTAATAGAACTCTCCTTTTCCCAGAGG

3' end AAAAAAAAAAAAAAAAAAAAAAAGATGATTTCAAGACCAATTTCTAATAGAACTCTCCTTTTCCCAGAGG

p7LE16. 2a
5' end TTTAGACAACTTGGTCAATAAAAAGTTTTTAAAAATTTCC AGATCTCTAGAAGCTGGGTACCAGCTGCTAGC

gg_chr 5 + TTTAGACAACTTGGTCAATAAAAAGTTTTT AAAAATTTCCAAAGGATTAGGATTTACCTGCTGAAAACATGCTT

3' end AAAAAAAAAAAAAAAAAAAAAAATTTCCAAAGGATTAGGATTTACCTGCTGAAAACATGCTT

p7LE16. 4a
5' end GACAAGAGCTCTGCATTATTATTGGCAGTAAAAAGTATTTCC TGGGTCGGGGGAGGGGAGGGATAGCATTGGGAGATAT

gg_chr6+ GACAAGAGCTCTGCATTATTATTGGCAGT AAAAAGTATTTCCATAGCTGAAAAGGCTTACTTGGTGGTCTGTGTTCTCTATT

3' end AAAAAAAAAAAAAAAAAAAAAGTATTTCCATAGCTGAAAAGGCTTACTTGGTGGTCTGTGTTCTCTATT

p7LE18. 2a
5' end TTTCCACCGATATCCCTAAGTAACAGCTAGTT TAAGAACA AGATCTCTAGAAGCTGGGTACCAG

gg_chr 2 - TTTCCACCGATATCCCTAAGTAACAGCTAGTTT AAGAACACAACAACAAAAGCAAGAGAAAATCCC

3' end AAAAAAAAAAAAAAAAAAAAGAACAACAACAACAAGCAAGAGAAAATCCC

p7LE18. 6a
5' end CCCTTAATTTGCAATGCTAAAATTAATTAGAT AGATCTCTAGAAGCTGGGTACCAGCTGCTAG

gg_chr2 - CCCTTAATTTGCAATGCT AAAATTAATTAGATGAAACAGTAAAATCCCAGGAAATTTTCATATG

3' end AAAAAAAAAAAAAAAAAAAATAAAATTAATTAGATGAAACAGTAAAATCCCAGGAAATTTTCATATG

p7LE19. 1a
5' end TAATTTTTAGCAAATGGTTTCATTGTGTAAAGACAGGT AGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGTAGC

gg_chr1 + TAATTTTTAGCAAATGGTTTCATTGTGT AAGACAGGTAATCTCAGCGACACATGTCAGTGTCTCCTGCTCCGCTGA

3' end AAAAAAAAAAAAAAAAAAAAAGACAGGTAACTCAGCGACACATGTCAGTGTCTCCTGCTCCGCTGA

p7LE19. 3a
5' end CTTCTGATTTGCAGTGTTTAAAATAAAGGGCAACAG AGATCCCTCAGAAACTCGTCAAG

gg_chr3 - CTTCTGATTTGCAGTGTTTAAAATAAAGGGCAACAGACACT TGGATTGAGATAACCCAGCACAG

3' end AAAAAAAAAAAAAAAAAAAATAAAATAAAGGGCAACAGACACTTGGATTGAGATAACCCAGCACAG

p7LE19. 6a
 5' end AATAGTGGGTTTTCTAGTGACATGAAAAGCATCTCTCAAGGGCAGACAGACTGCCTCTCA

 gg_chr3 + AATAGTGGGTTTTCTAGTGACATGAAAAGCATCTCTCACTCTAGCTGAAAAACAGGTAC

 3' end AAAAAAAAAAAAAAAAAAAAAAAAGAAAAGCATCTCTCACTCTAGCTGAAAAACAGGTAC

L1.3 / HeLa-RC cells

pHLE1. 2a
 5' end TTGTAATATGCCACCTTTATATTTATAAGAGTATTTAGAGATCTCTAGAAGCTGGGTACCAGCTGTAGCAAGCTAGCTTGCTAG

 hs_chr4 - TTGTAATATGCCACCTTTATATTTATAAGAGTATTTAGAGATCTCTAGAAGCTGGGTACCAGCTGTAGCAAGCTAGCTTGCTAG

 3' end AAAAAAAAAAAAAAAAAAAAAAAAAGAGTATTTACACAAGAATTTGCAGTTTAAAGTCGTATCAATCATTGAGTTGATATAA

pHLE1. 4a
 5' end TCCATTTTGTCCAAAAGCAAAAACAAAAGCACTCAAGTTAAAAAAGTCTGGATTAAAGATTTAAACGTTAAACCTAAAACCATAA

 hs_chr2+ TCCATTTTGTCCAAAAGCAAAAACAAAAGCACTCAAGTTAAAAAAGTCTGGATTAAAGATTTAAACGTTAAACCTAAAACCATAA

 3' end GCTGCAATAACAAGTTAACAAAAAAAAAAAAAAAAAAAAAAAAAAAGTCAAAATTTATAAAGGAGAACACCAAAATATTATTTCATG

pHLE1. 5a
 5' end TTTGGCCACAGGAGAGCAGTATCTACATTTAATTGAAAAAAAACCAACACCCGCATATCTCACTCATAGGTGGGAATTGAAC

 hs_chr3 + TTTGGCCACAGGAGAGCAGTATCTACATTTAATTGAAAAAAAACCAACACCCGCATATCTCACTCATAGGTGGGAATTGAAC

 3' end AAAAAAAAAAAAAAAAAAAAAAAACTACCATTGTTATTGTAGTATTCTCATAACAATCTTGGAAG

pHLE1. 6a
 5' end GCCACTTCAGTTGTTTTGTTGAAATACCTAATGCTAGTAGACACATTAGTGGGTGCAGCGCACCCAGCATGGCACATG

 hs_chr17 + GCTGTGCACCTCAGTTGTTTTGTTGAAACAACAACACAAAAACAAAACAGATGCTCTGGACACTGATTTTC

 3' end AAAAAAAAAAAAAAAAAAAAAAAACAAAAACAAAACAGATGCTCTGGACACTGATTTTC

pHLE2. 3a
5' end GGGGGGGTGGCTACGCCTGTGGTCAAAATGTGGGCACATATACACCATGGAATAGTATGCAGCCATAAAAAAT

hs_chr2+ GGGGGGGTGGCTACGCCTGTAAATCCCAACACTTTGGGAGGCCAAGGTGGTGGATCACAAAGTCAGGAGTTC

116533960
hs_chr3 + TTCTAATTATGATGTAATCCATGTCTCAATAAAAAACACTCACATGCTCCAGAGGTTAAAAATGGGAGGAA

3' end [AA](#)CACTCACATGCTCCAGAGGTTAAAAATGGGAGGAA

(comment)
pLEmH TCCAACAATGATAGACTGGATTAAAGAAATGTGGCAGATATACACCATGGAATAGTATGCAGCCATAAAAAAT

5' end GGGGGGGTGGCTACGCCTGTGGTCAAAATGTGGGCACATATACACCATGGAATAGTATGCAGCCATAAAAAAT

hs_chr2+ GGGGGGGTGGCTACGCCTGTAAATCCCAACACTTTGGGAGGCCAAGGTGGTGGATCACAAAGTCAGGAGTTC

hs_chr3+ GGATGGAGGTTACAAGGAAATGGTCAAACTTTGAATGAAACTAGCCAAATATCCTTTTGT
116534108

pHLE2. 5a
5' end GTAAATTAATAACCTTTTCATCTTAAGGAGGGAAAAATGGAAGCTCGGGGGAGGGGGAGGGATAGCATTGGGAGATATACCTAATGCTAGATGACA

hs_chr2 - GTAAATTAATAACCTTTTCATCTTAAGGAGGGGAAAATGGAAGCTAATTATCAAAATTTAAACGCTGATATCCTTTTAAATTTGACTATT

3' end [AA](#)GAAAATGGAAGCTAATTATCAAAATTTAAACGCTGATATCCTTTTAAATTTGACTATT

pHLE2. 9a
5' end ATAAAAATGTTAGTGTAAAACACATTCAATTTTCCAATGGCGAAACCAACACCGCATATTCTCACTCATAGGTGGGAATGAACAATGAGAT

hs_chr4 - ATAAAAATGTTAGTGTAAAACACATTCAATTTTCCAATGGCGAATCAAGATTCAGTAAAGATAAGGAAATATGCTTGTAAATCATTAGTAA

3' end [AA](#)GATTCAAGTAAAGATAAGGAAATATGCTTGTAAATCATTAGTAA

pHLE2. 10a
5' end GCATATATAGTGTAAAACCTGTAAGAAAAGCACATTGTTATAACAACAATGATAGACTGGATTAAGAAAATGTGGCAGATATACACCATGGA

hs_chr6 - GCATATATAGTGTAAAACCTGTAAGAAAAGCACATTGTTATAAAAAATCAGAAATCATATATAGAGATGAGAACAGATCATTATCGGGGAG

3' end [AA](#)TAAAAATCAGAAATCATATATAGAGATGAGAACAGATCATTATCGGGGAG

pHLE2. 11a
5' end TATAATAATAATAAATCATGAAACTGGCACTTTAGAAATAACTGGCTAGCCATATGTAGAAAGCTGAAACTGGACCCCTTCCCTACACC

hs_chr12 + TATAATAATAATAAATCATGAACTTTGGCACTTTAGAAATAGCAATAGCCTAAGACTTGAACTTATAGTAATGAGTTGATGTTAGTCAAGT

3' end [AA](#)GAAATAGCCTAAGACTTGAACTTATAGTAATGAGTTGATGTTAGTCAAGT

pHLE3. 5a
5' end TTGAGCTATTGTTAGTACCCTTTATCTTTTGAAAAGACATTATCATGCTGCTATAAAGACACATGCACAGTATGTTATTGGGGCACTATT

hs_chr5 + TTGAGCTATTGTTAGTACCCTTTATCTTTTGAAAAGACATTATTTTTATGTACATGATATCTAATTATTGCACAAATCCTTACCCTCT

3' end [AA](#)TATTTTATGTACATGATATCTAATTATTGCACAAATCCTTACCCTCT

pHLE3. 9a
5' end CGTGTGTTTCTTAAATTAGTCAATCTATAACCCGCAAGGAAAAATCAAAGATCTACAATGAACCTAAACAAATTTAGAAGAAAAACA

hs_chr7 - CGTGTGTTTCTTAAATTAGTCAATCTATAACCCGCAAGGAAAAATCAAAGATAATGCTTGGACTTAAATGAAAGCTAGTTCGGGCTG

3' end [AA](#)GAATACTAAAGATAATGCTTGGACTTAAATGAAAGCTAGTTCGGGCTG

pHLE4. 4a
5' end CCTGGCACTTAGAAGGGAAGAACTTTGCAACACAATGACGATTAAGAGGGGGAGGGATAGCATTGGGAGATATACCTAATGCTAGATGACACATT

hs_chr15 + CCTGGCACTTAGAAGGGAAGAACTTTGCAACACAATGACGATTAAGAGATAACAGTGAAAGTCAAGAGCAGATTCACAAAGGGAAGTCACTTTGA

3' end [AA](#)GAATAACAGTGAAAGTCAAGAGCAGATTCACAAAGGGAAGTCACTTTCA

pHLE7. 10a
5' end ACAAATCCCGACTTTCCTGTGTTTTAAATTTAGTTTTAAATTTAGCATATTCCTACTATAGTGGGAATGAACAATGAGAT

hs_chrx - ACAAATCCCGACTTTCCTGTGTTTTAAATTTAGTTTTAAATTCAAATGCATCTGCAGTTTTAAATCAAATGATCTCCA

3' end AATAAATCAAATGCATCTGCAGTTTTAAATCAAATGATCTCCA

pHLE7. 28a
5' end TGCTGTGATGGAGCTGGGCAAAGTGAATGAAAATCTTGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCAAGCTAGCTT

hs_chr6 - TGCTGTGATGGAGCTGGGCAAAGTGAATGAAATCTCTGAAAAGGATCCCAATTTAGATGATGCCATTAATAATGTT

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAATCTCTGAAAAGGATCCCAATTTAGATGATGCCATTAATAATGTT

pHLE7. 30a U6 snRNA insertion at 5'-end
5' end AGCATGGCCCTGGGCAAGATGACACGCAAAATCGTGAAGCGTCCATATTTTGCACAGTATGTTTTGCGGCAGCTATTACAAATAGCAAAGACT
HLE7. 30aL1. 3Ct-R1 CATAATAAAATACTAAAACATCCATAATCGTCAATTGCTTCTGTTTCTAAAAATTTAGTCTGCGCTCGGCGAGCACAATATAAAATGGAACGATAC

hs_chr6 - CATAATAAAATACTAAAACATCCATAATCGTCAATTGCTTCTGTTTCTAAAAATTTAGTCTGCGCTCGGCGAGCACAATATAAAATGGAACGATAC

3' end AATAAATTTATAGGCTATTTCTTAATAGGTGCATCAACCCAGTCATCA

pHLE8. 2c
5' end TAGAGGGCAGCAT AAGAATGGGAATTGAGGT TAAAAGT AAGAAAATGTGGCACATATACACCATGGAATACATATGACAGCC

hs_chr2 + TAGAGGGCAGCAT AAGAATGGGAATTGAGGT TAAAAGT GAGTATAAAGTAAAGATAGGATTTTCATCAGGGTGAAAGTA

3' end AAAGTAAAGATAGGATTTTCATCAGGGTGAAAGTA

pHLE8. 3a
5' end AGATCAGAGCTGCTGCTGACTGCTAGTCAAATGTGACATGATCAA AAAAAACCCTCAAAAAATCAATGAATCCAGGAGCTGTTTTTGAA

hs_chr13 - AGATCAGAGCTGCTGCTGACTGCTAGTCAAATGTGACATGATCA AAGAATAAACCTTTGTTGATgTCAGATGTCAAAGCCATTGAGTTTTT

3' end AAAGAAATAAACCTTTGTTGATGTCAGATGTCAAAGCCATTGAGTTTT

pHLE8. 15a
5' end AAGGAATAATGGATAATGGTATTGTTGCTACTTTTTAAGTATGGCACATGTACATATGTAACAAACCCGATCCGAAACAAAGCACC

hs_chr5 - AAGGAATAATGGATAATGGTATTGTTGCTACTTTTTAAGAAATATCTGTTAGAAGATATAGGATGCAGAACATCTAGGATTTGCTGAAAG

3' end AAAGAAATATCTGTTAGAAGATATAGGATGCAGAACATCTAGGATTTGCTGAAAG

pHLE8. 18a
5' end GAAAACCCCAACAAGGCCATTATAGTTAAATCACTGGAACCAAAATTTGGGAGATATACCTAATGCTAGTGCACACATTAGTGGGTGCAGCG

hs_chr4 + GAAAACCCCAACAAGGCCATTATAGTTAAATCACTGGAACCAAAAGAAACATGTTAATGGCATCCAGGAGAAAAATACATAGTGAAGAG

3' end AAAGAACATGTTAATGGCATCCAGGAGAAAAATACATAGTGAAGAG

pHLE8. 22a
5' end GGACTGATTTTAAAAAGTGTGACTCTCCATCTACTTTGTCATGTGATAAAGTCTAGAAAATCTAGAAGAAATGGATACATCTCTCGA

hs_chr7 - GGACTGATTTTAAAAAGTGTGAAATCTCCATCTACTTTGTCATGTGATAAAGAAATTTGTTTTCAATCTCTTGATGAATCATTTGAT

3' end AAAGAAATTTGTTTTCAATCTCTTGATGAATCATTTGAT

pHA4.11a
HA4.11aR6.9 AAAACATGAGCAAAGTGAATTCAATAACACCTTAAAAATCACTTTGGGGAACGCTCAGGCTCAATCTGCC

hs_chr8 - AAAACATGAGCAAAGTGAATTCAATAACACCTTAAAAAGATCATTGATCAAGACCAAGTGGGATTTATCC

HA4.11aNeoPrif9 GTCTGCTAAATGACTAAATGTAATGTAATGTAAAAGATCATTGATCAAGACCAAGTGGGATTTATCC

pWC10. 4a
 5' end ATGTTACAAGGCAACAGCACACCCATACCACACATTATTC²³²⁴⁵⁰⁰²AAATCTTTCTGGGAGCGTCTAGGCTCTCGTCTGACGGCCTGCACCCC

 gg_chr4 - ATGTTACAAGGCAACAGCACACCCATACCACACATTACACAGATAAATCTATTCTTATCATCGGGTACAATAAATCAAGTACAAAACTC
 23245002
 gg_chr4 - GAAAACCAATCTGATCTTTATTTTACAGTACTCTAAG²³²⁴²³⁵⁹GACACCACTCATACGCGAGCATATTTGAAACTGAATCTATTTACATTTATGA

 3' end AAGCGCTATACAAATAAAGCTGAATTGAATTGAATTGA²³²⁴²³⁵⁹ACCACCTCATACGCGAGCATATTTGAAACTGAATCTATTTACATTTATGA

pWC10. 6a
 5' end TGTGTCGTATCCTCCCAAGGATG⁴⁴⁶²⁷⁷AGCTGTCATTTCTGGGAGAAGCATCGATGGAGCGTTGGAGCTGGAGCTGGAAGAAGTGGAGTCCCAGAT

 gg_chr1 - TTACTTACCAGTCAATGACAAGAA⁴⁴⁶²⁷⁷AGCTGTCATTTGTTCAAGCTCATTAACTTTTTTCTTTTGGATCAGGATTTCACTCCGAAATCTGCAA

 3' end CATTGTAAGGCGCTATACAAATAAAGCTGAATTGAATTGAATTGAATTGA⁴⁴⁶²⁷⁷TAACCTTTTTTCTTTTGGATCAGGATTTCACTCCGAAATCTGCAA

pWC11. 2a
 5' end CAAAAGTAAAGTGGGAAATGAGTTTCTGGCTTGTACTGTTAATAT¹TACAATACTGGAATCTTTCTGGGAGCGTCTAGGCTCTTCGG

 gg_chr3 - CAAAAGTAAAGTGGGAAATGAGTTTCTGGCTTGTACTGTTAATATTTTTCTCTATTTGAAGATCTGTTTCAAACCTCGCTG

 3' end AAAGCGCTATACAAATAAAGCTGAATTGAATTGAATTGAATTGA¹TTTCTCTATTTGAAGATCTGTTTCAAACCTCGCTG

pWC11. 6a
 5' end AGCCACTGAGTGGCTGAGATGGTGCAGAGGGTCAAGAAGCGAGCG¹AACTCTTTCTGGGAGCGTCTAGGCTCTTCGGTC

 gg_chr17 + AGCCACTGAGTGGCTGAGATGGTGCAGAGGGTCAAGAAGCGAGCGAGCGCTGCTGCCCTTCCCATGGGGCAGCGGTA

 3' end AAAGCGCTATACAAATAAAGCTGAATTGAATTGAATTGA¹AGGGCTGCTGCCCTTCCCATGGGGCAGCGGTA

pWC12. 1a
 5' end GAGCAGCTGAACCCATCCCAGACCCCTGTGTTAAAGCTGCCAAG¹GTCTTCCCAGATCCGGGGCAGAACGTCTCCGCTGTG

 gg_chr17- GAGCAGCTGAACCCATCCCAGACCCCTGTGTTAAAGCTGCCAAGAGCGGAGAGCTAGCTTTATGCTCCAGAGCCCTCCAAAC

 3' end GCGCTATACAAATAAAGCTGAATTGAATTGAATTGA¹CCAAAGGAGAGCTAGCTTTATGCTCCAGAGCCCTCCAAAC

pWC12. 4a
 5' end TCCCACAGTGCAAAAATTTCCAAAGTTAAATCTGAT¹GAATGCTCTTTGCTAACTGGAATCTTTCTGGGAGCGTCTAGG

 gg_chr4 + TCCCACAGTGCAAAAATTTCCAAAGTTAAATCTGATGTTGTTTTAAAGAGAAGGCTAGTTAAAAATCAAGAATATAATTATA

 3' end AAGCGCTATACAAATAAAGCTGAATTGAATTGAATTGA¹TTGTTTTAAAGAGAAGGCTAGTTAAAAATCAAGAATATAATTATA

pWC12. 5a
 5' end ATGTAGTTTTGGATTGTGGCCTGATGAAGCAAGTAAATAAGAGTTATG¹TAAAGTTATTCTCTATTTTCACTGGGGATA

 gg_chr3 + ATGTAGTTTTGGATTGTGGCCTGATGAAGCAAGTAAATAAGAGTTATTGACAAGATGACGAACCTGAAAGCTTGGTTCTGCTT

 3' end AAGCGCTATACAAATAAAGCTGAATTGAATTGA¹AAAAAAGTTATTGACCAGATGACGAACCTGAAAGCTTCATATAGCAC

pWC13. 1a
 5' end TTTGGCTTTGTTTTGGAATTGATGCTTCTCCACAGCTGCCTGTAGT¹ACCATCTGACTAGCAGTAAAAATTCACAAAATTCACACTATAG

 gg_chr4 + TTTGGCTTTGTTTTGGAATTGATGCTTCTCCACAGCTGCCTGTAGTGGCACTGGGGATAGAGGATGTTGTACTATTCTCTTTTAACT

 3' end AAAGCGCTATACAAATAAAGCTGAATTGAATTGAATTGA¹TAGTGGCACTGGGGATAGAGGATGTTGTACTATTCTCTTTTAACT

pWC22. 1a
 5' end AGCTTGTCCAATGTGAACATAAAGGCATCACCAGATTAGAGGGATTCAGCTGTGATCTCTAGAAGCTGGTACTA

 gg_chr 24 - AGCTTGTCCAATGTGAACATAAAGGCATCACCAGATTAGAGGGATTCAGCTGTGAGTTGCTAACCCAATGCATCT

 3' end TCTACATTTGAAAAGCGCTATACAAATAAAGCTGAATTGAATTGAATTGAATTGAGTTGCTAACCCAATGCATCT

ZfL2-1 / DT40Ku70-/- cells

p7C2. 1a
 5' end AAATTATTTGCTTTCCTTTATTTACTGAATGCAGCTTTGGGCAGTACAAGGGTCAGCTTCTACTGCTATGCAGATGATACTCAATTATATTTCCACTA

 gg_chr14 - AAATTATTTGCTTTCCTTTATTTACTGAATGCAGCTTTGGGCAGTACAAGAAATCCTCACAAAGGGCACAAAGCACTTTAATGAGGCCAAGAAGTGTG
 1098029
 gg_chr14 - AGTCATTTCCAACGTAAAGTTTGTGGCATCAAACACAGGTAACACCCCGAGGAACAGCTAGTGATGCTTCCTGACCACTTCCATCCTGTGCTCCCGGCCCTT
 1093646

 3' end AATCTACATTTGAAAAGCGCTATACAAATAAAGCTGAATTGAAAAAATGAATTGAATTGAGTTGATGCTTCCTGACCACTTCCATCCTGTGCTCCCGGCCCTT

p7C3. 1a
 5' end GTAACTCTTTTTACACCTGCTCAAACCTGAACCATGGATTGGTTTTCTGAGGATCCCTCAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATG

 gg_chrZ - GTAACTCTTTTTACACCTGCTCAAACCTGAACCATTTCTGGGAAGTTCTACTCTTACATTTATCTAGCTTTTACTACAGCATAGTCAGGTCTAGATT
 13282880
 gg_chrZ - AGGCTGTAACCTCTGAAATGTACAAGCCAGATAGCTAAGTGAGGGTGGGAAGTGCAAGGAAAGCAGAAAGACGTCCTTTAGTCTGCAGAGAAAAACA
 13282674

 3' end CTACATTTGAAAAGCGCTATACAAATAAAGCTGAATTGAATTGAATTGAAAAGTGCAAGGAAAGCAGAAAGACGTCCTTTAGTCTGCAGAGAAAAACA

p7C7. 2a
 5' end CTACAGTACAGAGCACTGGGCATCATATGGCTGAACTCGCTCCTTAAGATCTCAAAATCAGGGCTTCTGGTAGTACCTAGAATA

 gg_chr3 - CTACAGTACAGAGCACTGGGCATCATATGGCTGAACTCGCTCCTTAAGATCTCAAAATCAGGGCTTCTGGTAGTACCTAGAATA

 3' end ATAAAGCTGAATTGAATTGAATTGAATTGAGTTGATGCTTCCTGACCACTTCCATCCTGTGCTCCCGGCCCTT

p7C24. 1a
 5' end AAGCTTAAGCTATAGGAGATAATGAGAAGTTCATTCTCTATTTTCACCTGGGGTACTCATCCGA

 gg_chr1 - AAGCTTAAGCTATAGGAGATAATGAGAAGTTCATTCTCTATTTTCACCTGGGGTACTCATCCGA

 3' end AAGCTTAAGCTATAGGAGATAATGAGAAGTTCATTCTCTATTTTCACCTGGGGTACTCATCCGA

p7C48. 1a
5' end ATTGAACAGTTTCTCTCCTTTTAAACAACCTAGAATAAAGACAGTCTACAGGTGTCCAGGACAGGTCTACAATGGTTAAGTCATACT

gg_chr3 - ATTTGAACAGTTTCTCTCCTTTTAAACAACCTAGAATAAAGACAGAGTGTGCTACCCGTTTTGTAAACATACGGAGGCTCTCAAGTGG

3' end CATTGTAAGGCGTATACAATAAAGCTGAATTGAATTGAATTGAAGTGTGCTACCCGTTTTGTAAACATACGGAGGCTCTCAAGTGG

p7C49. 1a
5' end CCAGGTGCTCAATCAGGATTCAATCAGGAATCAATCAGGATCAATCATGGAAAAAATAATTTAGA

gg_chr8 - CCAGGTGCTCAATCAGTGGTTCAGGCGGTGACTCAGCAGTTACCATACAGATGCTAACACCTCTGGCT

3' end TTGAATTGAATTGAATTGGTTTCAGGCCGTGACTCAACAGTTACCATACAGATGCTAACACCTCTGGCT

p50. 1a
5' end ACTCTTTTAGCTGGGGACAACAAGTTTACTACCATGATCTCTAGAAAGCTGGGTACTAGCGCCGCCACCA

gg_chr4 + ACTCTTTTAGCTGGGGACAACAAGTTTACTACCATGAAATGGCATTAAAGGCCACAGACTCCCAGCTAA

3' end AAAGCGCTATACAATAAAGCTGAATTGAATTGAATTGAATGGCATTAAAGGCCACAGACTCCCAGCTAA

p7C51. 1a
5' end CTAGGTGTTTCATAGTACTAATCAAACAGACTCATGCTAACCATAGAGTTACCTCAGAAGAAGCTCGTCAAGAAGGCGATAGAAGGCGATGC

gg_chr3 - CTAGGTGTTTCATAGTACTAATCAAACAGACTCATGCTAACCATAGAGAAATTAACATATCATGGACTTCTGATTAACAGATCTCTTA

3' end TTGTAAGGCGCTATACAATAAAGCTGAATTGAATTGAATTGAATTGAATAAATTAACATATCATGGACTTCTGATTAACAGATCTCTTA

p7C52. 1a
5' end CCAAAAAACAATCCAGCTAGTTTCTAACCAATAGAGAATTCACCAAGGACCAAACTGGAAATCTTTCTGGGAGCGTCTAGGCTC

gg_chr2 + CCAAAAAACAATCCAGCTAGTTTCTAACCAATAGAGAATTCACCAAGGACCAAACTGGTTGTGTAGTCTATACACTACTCAGAAT
47868106
47867647
gg_chr2 + AGAGCCAGCTCTTTTGAAGGTGATTTTAACTAAATAGTTAATAACATCATGTGCAATTTCTTAGTCCCAAGGTGGGCAACATT

3' end AAAGCGCTATACAATAAAGCTGAATTGAATTGAATTGAATTGAATAAATTAACATCATGTGCAATTTCTTAGTCCCAAGGTGGGCAACATT

p7C55. 1a
5' end TACAGCATAAAGACTAAAAAAAGACTAAAAAAGACATAAAGACACTTAAACACTTAAACAT

gg_chr2 + TACAGCATAAAGACTACTAAGTGAATCTAAAAGTCTGTGCACATCCAGCAAAAGTATTTCAC

3' end TTGAATTGAATTGAATTACTAAGTGAATCTAAAAGTCTGTGCACATCCAGCAAAAGTATTTCAC

p7C61. 1a
5' end TACATTTACCTCTGCAGTTTTATTGTAGCCAGATATAATGCAGATATAATGTAGTCCAGCTTGTTCACCCACTTAAACATCAGTAAAC

gg_chr2 + TACGTTTACCTCTGCAGTTTTATCGTAGCCAGATATAATGTAGTCCAGCTTAAATCCACCAAGACAGCATTTAAAAAGCAGCAGTCTTTTAAAG

3' end AAGCGCTATACAATAAAGCTGAATTGAATTGAATTGAATTGAATTGACGCTTTATACCACCAAGACAGCATTTAAAAAGCAACAGTCTTTTAAAG

p7C62. 1a
5' end CGCTGTGCCCCGAGACTTGCCCGAGTGAAGCTGTAGCTTTGTGATACCCCTGAGATCTCTAGAAGCTGGGTACTAGCGCCGCCA

gg_chr13 - CGCTGTGCCCCGAGACTTGCCCGAGTGAAGCTGTAGCTTTGTGATACCCCTCCAATGTTGCTGAATTTATCTTTTCAGCACACCT

3' end AAGCGCTATACAATAAAGCTGAATTGAATTGAATTGAATTGAATTGACCTCCAATGTTGCTGAATTTATCTTTTCAGCACACCT

Ra115
5' end TTCTTTCTAATCCTCTAATAAOCCTAGAGAAAGTGAGGAAGTAGCAAAATCAAGTAAAGGAGTGGAGCCTTCTCTTTCATGGCTCCTACA

hs_chr1 + TTCTTTCTAATCCTCTAATAAOCCTAGAGAAAGTGAGGAAGTATCCTTATCCGAGCCTACAGATGAGAAAACCTGAGGCTCAGCGAAGAGC

3' end CTACATTGTAAGGCGCTATACAAATAAAGCTGAATTGAATTGAATTATCCGAGCCTACAGATGAGAAAACCTGAGGCTCAGCGAAGAGC

Ra121
5' end AAAAAAAAAAACACACACAAAGAAAAATAATCTTACCATTATTTAAGGAGATTATTAAAGAACCGCACTAGACCCAACTTAGCTAACT

hs_chr10 - AAAAAAAAAAACACACACAAAGAAAAATAATCTTACCATTATTTAAGGAGATTATTTAAAAAATAGTTACTGTGTAAGGGAAGTCTCAAAA

3' end ATCTACATTGTAAGGCGCTATACAAATAAAGCTGAATTGAATTGAATTGAGATTATTTAAAAAATAGTTACTGTGTAAGGGAAGTCTCAAAA

Ra123
5' end CTTTTGGGCAAGATATGGTGGAAATAATGCTGTAGATACACTCTCCGAGTCAAACCTAGATTAAGACCTATCTGTTAGTAAAGCA

hs_chr6 + CTTTTGGGCAAGATATGGTGGAAATAATGCTGTAGAGCCTTATAATGATGATGACTTTTATCACTTGGCTAAGATAGTGCCGCCC

3' end AAAGCGCTATACAAATAAAGCTGAATTGAATTGAATTGATATAATGATGATGACTTTTATCACTTGGCTAAGATAGTGCCGCCC

Ra124
5' end GGACAGGGAGGAGAGAGAGCCCGAGCCCTTTCCACTCTGCATGTTATTCTCTATTTCACTGGGATACTCATCCCGAGGTC

hs_chr7 + GGACAGGGAGGAGAGAGAGCCCGAGCCCTTTCCACTCTGCATCTCCCAAGTCCAGTGACTATGGACCCGAGGAGGTGCTGA

3' end TGTAAGCGCTATACAAATAAAGCTGAATTGAATTGAATTGACTCTCCCAAGTCCAGTGACTATGGACCCGAGGAGGTGCTGA

Ra125
5' end CTCAGGCCGCCCTCAGGGCTGATTTAGACTCTCGCACCCTGCTAGCAGTAAAAATTC

hs_chr1 + GGTGATGATCTGGGAAACTCAGGCCGCCCTCAGGGCTGATTTGATCTTTGCTATTGGACTCTGTGACACTGTCAAGTCTG

3' end GCTATACAAATAAAGCTGAATTGAATTGAATTGAATTGATCTTTGCTATTGGACTCTGTGACACTGTCAAGTCTG

Ra126
5' end GGTGTGTAGTAGGCTATAATCTAGGTTTATGTATGTACACGCTAAGATGTTGCATCTTTATTCTCTATTTCACTGGGATACTCATCCCGAGG

hs_chr12 - GGTGTGTAGTAGGCTATAATCTAGGTTTATGTATGTACACGCTAAGATGTTGCATCTTTATTCTCTATTTCACTGGGATACTCATCCCGAGG

3' end ACACAATCTACATTGTAAGGCGCTATACAAATAAAGCTGAATTGAATTGAATTGACCAATGACAAAATATCTAACAATGCAATTTCTCATAATGCAAT

Ra201
5' end TTTGAAAAAGGTGGGACATAATGAACCTCATGAGTAGATACCTATCATAGGCTCTCCGCTCAGCGCCTGCACCCAGTGGAGCCGAGCTGAAC

hs_chr8 + TTTGAAAAAGGTGGGACATAATGAACCTCATGAGTAGATACCTATCATATAACACATGCATCCTGTCTTTCTGTGCTAGTATTGTTGAAGATAGCT

3' end ACAATCTACATTGTAAGGCGCTATACAAATAAAGCTGAATTGAATTGAATTGACCAATGCAATCCTGTCTTTCTGTGCTAGTATTGTTGAAGATAGCT

Ra203
5' end TTGAGTTCCTGGTACATGCCTAAGATGTGCAGATTTGTTACATAGGTAAATAGGAGATGCCAACTTTCTGCACCGCTTATACCTCCAAATCC

hs_chr1 - TTGAGTTCCTGGTACATGCCTAAGATGTGCAGATTTGTTACATAGGTAAATATTGCCATGGTGGTTGCTGCACCTATCAACCTATCACTTA

3' end ACATTGTAAGGCGCTATACAAATAAAGCTGAATTGAATTGAATTGACCAATGCAATCCTGTCTTTCTGTGCTAGTATTGTTGAAGATAGCT

Ra205
 5' end AGACGGGCGGAGAGAAACCCGGGAGGCTAGGGACGGCCTGAAGG**CCCTCTTCATCTGAGAAATATCGCTAAATACGAAATATGCTA**

 hs_chr12 + AGACGGGCGGAGAGAAACCCGGGAGGCTAGGGACGGCCTCA**AGGCGGCAGGGGCGGGCCAGGCCGATGTGTTCGGCCGCTGCGGG**

 3' end AAGCGCTATACAAATAAAGCTGAATTGAATTGAATGA**A**AAGCGGCAGGGGCGGGCCAGGCCGATGTGTTCGGCCGCTGCGGG

Ra207
 5' end TTAACCTCAGTTTGTGGATATCCTCACGATAACTTGGTGAG**CCCTAGACTCTTGTTCACCCCACTAAACATCAGTAACGCAT**

 hs_chr2 + TTAACCTCAGTTTGTGGATATCCTCACGATAACTTGGTGAGTTTGA**TTGAGATTGCATTGAATCTGTAGATGAAGTTGGAA**

 3' end GTAAAGCGCTATACAAATAAAGCTGAATTGAATTGA**TTG**TTTTGATTGAGATTGCATTGAATCTGTAGATGAAGTTGGAA

Ra227
 5' end CATGTACAACAATACCTTCACAGAGCTAAGAAACCGGTAAGAGATTAC**ACTGATTGGATCTAGAAACGCCAACCCCGTGCGTTTTA**

 hs_chr18 - CATGTACAACAATACCTTCACAGAGCTAAGAAACCGGTAAGAGATT**ACAGCACCTGGGTATAGCACAGAAATAAGAAAAGGCACACTGA**

 3' end AAAAGCGCTATACAAATAAAGCTGAATTGAATTGA**TTG**AAATACAGCACCTGGGTATAGCACAGAAATAAGAAAAGGCACACTGA

Ra230
 5' end GGATAGAGAG**AGTTC**AAAACCTAGATTAAAGACCTATCTGTTTAGTAAAGCATACACTCAG

 hs_chr7 + GGATAGAGAGAGTGA**TAATGGAGTAGGTTTACATAGGGTTAAAAATAAATG**TTAATTGACAAGGCTGTTAAGAAGGATCAAAGCAGAGAGGAGAGA

 3' end CACAATCTACATTGTAAAAGCGCTATACAAATAAAGCTGA**ATTGAATTGAATTGAATTGA**TTAATTGACAAGGCTGTTAAGAAGGATCAAAGCAGAGAGGAGAGA

Ra232
 5' end CAGTAGATGATCTCAGCTCACTGCAACCTCTG**GTCCATGCTTTTATGACTTCGAGACTGGATTACTGTAATGCTCTATTT**

 hs_chr15 + CAGTAGATGATCTCAGCTCACTGCAACCTCTG**CCTCCAGGGTTCAAGTATTCTTTTGCCTAGCCTCCTGAGTAGCTGT**

 3' end AAGCGCTATACAAATAAAGCTGAATTGAAAAAAA**AAATGAATT**CAAGTATTCTTTTGCCTAGCCTCCTGAGTAGCTGT

Ra501
 5' end GGTCATGCCCAAGCCAGCTAAATTTTGTATTTTAGT**AGATTTTATTTAACACCATCTCTAAATTAGCAAAATATCGGTCATCGTTG**

 ucsc_chr8 + GGTCATGCCCAAGCCAGCTAAATTTTGTATTTTAGT**AGAGACGGGGTTTTGCCATGTTGTCAGGCTGAACCTCCTGACCTCAGGTGA**

 3' end AAAAGCGCTATACAAATAAAGCTGA**ATTGAATTGAATTGAATTGA**AAATTTACAAGGAAAAAATAAACACCCCATCAAAGTGGGCAAG

Ra506
 5' end GCAATCTACTCGTCTGACAAGGACTAATATCCAGAATCTACAATGAAC**TCAAATCTCCGGCAGTCGGAGATCCTGAAGAAGGACTTCAGGAGCCTGA**

 UCSC_hs_chr6 + GCAATCTACTCGTCTGACAAGGACTAATATCCAGAATCTACAATGAAC**TCAAATTTACAAGGAAAAAATAAACACCCCATCAAAGTGGGCAAG**

 3' end ATTGTAAGCGCTATACAAATAAAGCTGA**ATTGAATTGAATTGAATTGA**AAATTTACAAGGAAAAAATAAACACCCCATCAAAGTGGGCAAG

Ra507
 5' end GGGAAAGCTTTCATTCAAGTTTCTCTCTTAGTACTCTGCTAG**CTTTATCTCTATTTTCACCTGGGATACTCAT**

 hs_chr5 - GAAGAAGCTTTCATTCAAGTTTCTCTCTTAGTACTCTGCT**AGATGTACAGCAAGAAAAAGAACTCTGCCTTCAGGT**

 3' end GCTATACAAATAAAGCTGA**ATTGAATTGAATTGAATTGA**TTAGATGTACAGCAAGAAAAAGAACTCTGCCTTCAGGT

Ra707
 5' end TGTCCTGACCCACCCAGCCTCTTTGCTGATCCCTTGACACAGATCTGCGCAACTCAATTTACAATTAGAGGGATACAAAAGTTAG

 hs_chr20 - TGTCCTGACCCACCCAGCCTCTTTGCTGATCCCTTGACACACTCAAGCTCCTGCCGCCAGCCTCTGATCTGCTCATCCCTCT

 3' end CTATACAAATAAAGCTGAATTGAAAAAAAAAAAAAAAAAAAAGCTCAAGCTCCTGCCGCCAGCCTCTGATCTGCTCATCCCTCT

Ra708
 5' end TGGTTTGAACCTTTTTGATATACTGGGTTTCATCCTGTGGAGTCTTAAACCATATATCCCATGTACAAAAAAGCTGCTTCTTTC

 hs_chr1 + TGGTTTGAACCTTTTTGATATACTGGGTTTCATCCTGTGGAGTCTTAAACAATGTTTTCAGTAATTTACAAAAAACAACAAAAAC

 3' end AAAAGCGCTATACAAATAAAGCTGAAGTGAATTGAAAAAAAAGCTTAAACAATGTTTTCAGTAATTTACAAAAAACAACAAAAAC

Nimb2 / DT40 WT cells

pWG2. 3a
 5' end TAAAGAAATGGTTCTCTGCTTTCATAAACAGTCAATGTGCAGAGTTGGCCAGGCGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGA

 gg_chr1 - TAAAGAAATGGTTCTCTGCTTTCATAAACAGTCAATGTGCAGAGTTGGCCAGGCGCTGATGCCCCCGAGCTCGTTCCCTCCAGCCGATTGAA

 3' end GTTGCCAAACCGCCAGTAAAACCCAAAGAAAGAAAGAAAGAAAGTTGGCCAGGCGCTGATGCCCCCGAGCTCGTTCCCTCCAGCCGATTGAA

pWG2. 4a
 5' end GACCACCACTGACACCCACATCCGGGCCTACAAGAACGGAGTCAAGGTTAGAACCAAGTACACTGGGATAACATTGATACAGGTAGACATTTATATAATATAC

 gg_chr1 + GACCACCACTGACACCCACATCCGGGCCTACAAGAACGGAGTCAAGGTTAGAACCAAGTACACTGGGATAACATTGATACAGGTAGACATTTATATAATATAC

 3' end TAAGTCTGGTTGCCAACCGCCAGTAAAACCCAAAGAAAGAAAGAAAGAAAGAACCCACAGAACCCATTGAGGCTGGAAAAGACCTCAGATATACCCGGTCCAA

pWG3. 1a
 5' end TGGCAAGTTCAGTGAGACTTTGTGGAGAGGATGATTATCGTAGGGAATCGTATGAAATCGTAGGGAAGGAGTATTATAACTCG

 UCSG_chr12 + TGGCAAGTTCAGTGAGACTTTGTGGAGAGGATGATTATCGTAGGGAATCGTATGAAATCGTAGGGAAGGAGTATTATAACTCG

 3' end ACCGCCAGTAAAACCCAAAGAAAGAAAGAAAGTATGATTATGGCAGGACAGGAGGATTAGCATTGCACCTCTGCATCTAAATAAT

pWG3. 2a
 5' end AAGTTTGCATATGTGCTCTAAGAGAGGGTGCAAAAGTTGAAATGTTGCAACTGTGGAGGGATCATAGCTCAGCGT

 gg_chr5 + AAGTTTGCATATGTGCTCTAAGAGAGTTCCTTAAGTGTGATTGAAATGTTTCACTCTGGAGATCGGAGCAGCTTCC

 3' end GTCTGTTGCCAACCCGCGCTAAGAACCCAAAGAAAGAAAGAAAGTTTCACTCTGGAGATCGGAGCAGCTTCC

pWG7. 2a
5' end AGCACTCAACAATGCCATACAGGATCTTGTAAATGTAATGACAAAAATGTAATCATGTTAAACGTTCTACAATGGAATGCAAGGAGCTTAATTGCGAAT

gg_chr3 + AGCACTCAACAATGCCATACAGGATCTTGTAAATGTAATGACAAAAATGTAATCATTTTTGTCTAGGAAGACAGCCTATCTACGATTGATAATTTCTTC

3' end TTGCCAACCCGAGTAAACCCAAAGAAGAAGAAGAAGAAATGTAATCATTTTTGTCTAGGAAGACAGCCTATCTACGATTGATAATTTCTTC

pWG7. 5a
5' end GCAATAATGGTAGAAAGTCTTGTACTTCTGCATTTTTTCTCAGGCCAGTTCTTCCAGGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGATGGCAA

gg_chr2 - GCAATAATGGTAGAAAGTCTTGTACTTCTGCATTTTTTCTCAGGCCAGTTCTTCCAGGAGCTCTTTATGCCAAGCATCCTTTAAACTTGAAGGCTATTGT

3' end AAGTCTGGTTGCCAACCCGAGTAAACCCAAAGAAGAAGAAGAAGTTTCTTCCAGGCCAGTTTATGCCAAGCATCCTTTAAACTTGAAGGCTATTGT

pWG7. 6a
5' end GCGGGGCTGCCCGGGGGGGGGGGGCGGGGACTATAAAGAGCCCGTCCCGATTACCCCAACTGGGCAATGTATACATTGTATCAACAAGAAACAATAGAACAATT

gg_chr7 - GCGGGGCTGCCCGGGGGGGGGGGGCGGGGACTATAAAGAGCCCGTCCCGGGGAGCGGGGCTGTGCGGTGGGTGCGCTGCGCTGCGGCGCTCGGACGCGCCGG

3' end CTATAAGTCTGGTTGCCAACCCGAGTAAACCCAAAGAAGAAGAAGAAGAGGGAGCGGGGCTGTGCGGTGGGTGCGCTGCGCTGCGGCTCGGACGCGCCGG

pWG8. 1a
5' end GCAATATTGACAGAAAACTGACATACCAATTTGCAGTGAGTTGAGAACTGTATAGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGATGG

gg_chr2 - GCAATATTGACAGAAAACTGACATAGCAATTTGCAGTGAGTTGAGAACTGTATAGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGATGG

3' end TTGCCAACCCGAGTAAACCCAAAGAAGAAGAAGAAGAAAGAACTGTATACAAATAGACTGTTACAGACTGAAAGGTGCGAACAGCTTTGT

pWG8. 2a
5' end TTCTCCAGCTTATGAAATCCATGCAATATGACCTTCTCTCCATTAATCATCAGTACAAAAATATACTGGGCCCTCATGAGATCTGTTTTGATTATGGAT

gg_chr2 - TTCTCCAGCTTATGAAATCCATGCAATATGACCTTCTCTCCATTAATCCAACAACTGAATGTCTACTTTAGGAGACATTCTGACATTTATCTCTTAA

3' end TGCACTATAAGTCTGGTTGCCAACCCGAGTAAACCCAAAGAAGAAGAAGAAGAACTGAATGTCTACTTTAGGAGACATTCTGACATTTATCTCTTAA

pWG8. 3a
5' end AGGCTTCTGAAGTCTGCCAACAACAACTGGGAAGTCTATTGTGTCACACTGGACTGACTCAATCATACATTACATAAAATAGG

gg_chr4 + AGGCTTCTGAAGTCTGCCAACAACAACTGGGAAGTCTATTGTGTCACACTGGACTGACTCAATCATACATTACATAAAATAGG

3' end CAACCCGAGTAAACCCAAAGAAGAAGAAGAAGGTTCTATTGTGACTTCTGATAAGAATTTATTTACCTCAATTAATTTGAGGC

pWG9. 2a
5' end TATCAAGCATGACACAGTAGTCATTTATCCCAAGTGGATTTATCTTTTGGCAAGATGGCAAAATAATGGGGATCAAACATCTGTGCAAAAACAGGTCATCT

gg_chr3 - TATCAAGCATGACACAGTAGTCATTTATCCCAAGTGGATTTATCTTTTGGCAAGATTCCTCTGGATAGAAGCATAGAAGTTTGAGTTTCTTTCCCTAAAAACAA

3' end ACCTATAGTCTGGTTGCCAACCCGAGTAAACCCAAAGAAGAAGAAGAAGAAATTCCTCTGGATAGAAGCATAGAAGTTTGAGTTTCTTTCCCTAAAAACAA

pWG9. 3a
5' end CTGGTGGATGAAAAGTGGACATGACCCAGCAGTATGTCCTATAGCCTGGAAGTCAACAGAAAGAATATAGAGAGCCCGCCGACAAACGCCAACCCCGTGC

gg_cbr3 + CTGGTGGATGAAAAGTGGACATGACCCAGCAGTATGTCCTATAGCCTGGAAGTCAACAGTATTTGGGAGCATCAAAGAGGGATGCGGGAGGGTGAGGGAGGT

3' end ATAGTCTGGTTGCCAACCCGAGTAAACCCAAAGAAGAAGAAGAAGGTTCAACAGTATTTGGGAGCATCAAAGAGGGATGCGGGAGGGTGAGGGAGGT

pWG10. 1a
5' end GGATGATCCCGGGGCTTTTCCAACTTAGCGATTCTATGATTCTATGATAAGCCATTAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGATGGCAAGGC

gg_chr4 - GGATGATCCCGGGGCTTTTCCAACTTAGCGATTCTATGATTCTATGATAAGCCATTAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGATGGCAAGGC

3' end ACCTATAAGCTGGTTGCCAACCGCCAGTAAACCCAAAGAAAGAAAGATTAAGCCATTAGTAGCTAGAAAAGTTCAATGTTACAAGCCATGCTGATTTTGACT

pWG10. 3a
5' end AATAGGAGCGGGGAGGGTATTAAAAGCAACCGCGGCGAGGCAAGCAGTTGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGATGGCAAG

gg_chr10 - AATAGGAGCGGGGAGGGTATTAAAAGCAACCGCGGCGAGGCAAGCAGTTGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGATGGCAAG

3' end GGTTGCCAACCGCCAGTAAACCCAAAGAAAGAAAGAAAGATTAAGCCATTAGTAGCTAGAAAAGTTCAATGTTACAAGCCATGCTGATTTTGACT

pWG11. 2a
5' end TGCTAAATGGCCAAAGAGTAAAACACGTTGGCTTTCGTATAGATGAAGTCAATCAGTATTCAGAGTGGTTCGCGAGGCGAATACTATGGACTCAG

gg_chr1 - TGCTAAATGGCCAAAGAGTAAAACACGTTGGCTTTCGTATAGATGAAGTCAATCAGTATTCAGAGTGGTTCGCGAGGCGAATACTATGGACTCAG

3' end TCTGGTTGCCAACCGCCAGTAAACCCAAAGAAAGAAAGAAAGATTAAGCCATTAGTAGCTAGAAAAGTTCAATGTTACAAGCCATGCTGATTTTGACT

pWG12. 2a
5' end CTATTATTGCTAGCCTTTATGTTATTCAGTCCGATGGAGAAAAGCACTGAGAGATCAGCATTTTACATTGGCAGGTTTATTAGGAACA

gg_chr12 + CTATTATTGCTAGCCTTTATGTTATTCAGTCCGATGGAGAAAAGCACTGAGAGATCAGCATTTTACATTGGCAGGTTTATTAGGAACA

3' end GCCAACCGCCAGTAAACCCAAAGAAAGAAAGAAAGATTAAGCCATTAGTAGCTAGAAAAGTTCAATGTTACAAGCCATGCTGATTTTGACT

pWG13. 1a
5' end AGACTTTGAAAAATCCTTGAATTCGTGTGACACTGACGAAGTAACTAATAAAAAGTTCCAGAAATATGACTGTGTCTAAACAAAATGAAACAGGAA

gg_chr13 + AGACTTTGAAAAATCCTTGAATTCGTGTGACACTGACGAAGTAACTAATAAAAAGTTCCAGAAATATGACTGTGTCTAAACAAAATGAAACAGGAA

3' end ACCTATAAGCTGGTTGCCAACCGCCAGTAAACCCAAAGAAAGAAAGAAAGATTAAGCCATTAGTAGCTAGAAAAGTTCAATGTTACAAGCCATGCTGATTTTGACT

pWG13. 2a
5' end GCCCGGGGCGCTCGGTCGTTTTAAAGCCTCCCGCGTGCCTGCCGCGGCTGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGATGGCAA

gg_chr18 - GCCCGGGGCGCTCGGTCGTTTTAAAGCCTCCCGCGTGCCTGCCGCGGCTGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGATGGCAA

3' end GCCAACCGCCAGTAAACCCAAAGAAAGAAAGAAAGATTAAGCCATTAGTAGCTAGAAAAGTTCAATGTTACAAGCCATGCTGATTTTGACT

pWG14. 1a
5' end GCGCGGGGCGCCCGCTATATGAGCTCCGCGGCGCTGACTCGGCCCTTGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGATGGCAAGG

UCSC_gg_chr1 - GCGCGGGGCGCCCGCTATATGAGCTCCGCGGCGCTGACTCGGCCCTTGAGATCTCTAGAAGCTGGGTACCAGCTGCTAGCGGAGGATGGCAAGG

3' end CAACCGCAGTAAACCCAAAGAAAGAAAGAAAGATTAAGCCATTAGTAGCTAGAAAAGTTCAATGTTACAAGCCATGCTGATTTTGACT

pWG14. 2a
5' end ATGCTACCAGTCCAAAGCTTCTACCAGAACTTGGATGTTTACAAGAGAAACATCTATGCAAGTTGAAATGGGAAATGCCCTAAGCATTAGGAG

gg_chr2 - GCAACTTTTGTAAATAAGCTTCTACCAGAACTTGGATGTTTACAAGAGAAACATCTATGCAAGTTGAAATGGGAAATGCCCTAAGCATTAGGAG

3' end AAGTCTGGTTGCCAACCGCCAGTAAACCCAAAGAAAGAAAGAAAGATTAAGCCATTAGTAGCTAGAAAAGTTCAATGTTACAAGCCATGCTGATTTTGACT

pWG15. 1a
 5' end CTGAAGGGTGGCAGGGAGGGATGGATA TGAAGCTGAGGGTTTTTTTGGCAGACAAAAGAGCAATCTTCAGCATTTCACATTGGCAGGTTTATAGGAAACAAATCTA

 gg_chr15 + CTGAAGGGTGGCAGGGAGGGATGGATA TGAAGCTGAGGGTTTTTTTGGCAGACAAA TAGGTATTGTACAAAATTCGAAAGCTCCCACTGCTGATAGGCCTGGC

 3' end TGCACCTATAAGTCTGGTTGCCAACCCGAGTAAACCCCAAGAAGAAGAAGAAATAAATAGGTATTGTTACAAAATTCGAAAGCTTCAAAGTAATTCAGTGTTTAT

pWG16. 1a
 5' end ACCTGCACTGCTCCAGTTTTGAGAGGTCGTGAGG-TTTTTTAATCCAGAATAAGTTTGGCAAGTACACTGGGATAACATTGATACAGGTAGACATTTATAT

 gg_chr8 + ACCTGCACTGCTCCAGTTTTGAGAGGTCGTGAGGTTTTTTAATCCAGAATAAGTTTCCCTTTTAAAAATAAATGGGATGTAGTAATGGGTTGTACGGGA

 3' end TATAAGTCTGGTTGCCAACCCGAGTAAACCCCAAGAAGAAGAAGAAGAGTTCCCTTTTAAAAATAAATGGGATGTAGTAATGGGTTGTACGGGA

L1.3 inversion/ HeLa-RG cells

pHLE2. 1a
 5' end GCATTTATCTGGTACTATTTGTTTGTAGGTTTAGTTTAGAAAGTCAGACTGCATAAATGCTCTCTTTGAGAAGTGTCTGTCATGTCCT

 hs_chr5 - GCATTTATCTGGTACTATTTGTTTGTAGGTTTAGTTTAGAAAGTCAGACTGAAAGTACTAAAAGTTGTCTCATGCGTCCGTGGAAGAGAC

 3' end AAAGAAAGTCAAGTAAAGTACTAAAAGTTGTCTCATGCGTCCGTGGAAGAGAC

pHLE3. 8a
 5' end CAACATAAGCCATAATGCAAGCCGAGAAGTACATGTTAAAAAACCAAAAAGAAAGTGTCTGTCATGCTCCGCCCACCTTTTGTAGGGTTGT

 hs_chr9 - CAACATAAGCCATAATGCAAGCCGAGAAGTACATGTTAAAAAACCAAAAAGAAAGTGTCTGTCATGCTCCGCCCACCTTTTGTAGGGTTGT

 3' end AATAAAATAACCAAAAAGAAAGTGTCTGTCATGCTCCGCCCACCTTTTGTAGGGTTGT

pHLE3. 11a
 5' end TTAGCAACACATCTGGAATTCAGAAAAATTAGAATGTGACAGAAGAGGCATGTAGTTTCCTGAATACAGCACACTGATGGTCTTGACTCTTT

 hs_chr7 + TTAGCAACACATCTGGAATTCAGAAAAATTAGAATGTGACAGAAGAGGCATGTAGTTTCCTGAATACAGCACACTGATGGTCTTGACTCTTT

 3' end AAAGAAAGGCATGTAGTTTCCTGAATACAGCACACTGATGGTCTTGACTCTTT

pHLE4. 2a
 5' end CAGAATTTGGGAGACTCAAAAAATCAGCATAAATTGTAAGGGATTGTACAGATGGGTTTTCGGTGTAGATGTCCTTTTGTGGTTGTTAGTTT

 hs_chr8 + CAGAATTTGGGAGACTCAAAAAATCAGCATAAATTGTAAGGGATTGTACAGATGGGTTTTCGGTGTAGATGTCCTTTTGTGGTTGTTAGTTT

 3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTGTAAGGGATTGTACAGATGGGTTTTCGGTGTAGATGTCCTTTTGTGGTTGTTAGTTT

pHLE4. 5a
5' end CTTTGAGATTCAGTGACAATAGAGTCCAATT TAAAACTACTGGAAACAAAAAACCCACCGCTACCGAGGGTGGTTGTTTGCCGGATCAA

hs_chr5 - CTTTGAGATTCAGTGACAATAGAGTCCAATT TAAAACTACTGGAAACAAAAATTTGAAATTTCTCACTGAAAACATAACTCTTCTTTA

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA TAAAACTACTGGAAACAAAAATTTGAAATTTCTCACTGAAAACATAACTCTTCTTTA

pHLE2. 6a
5' end GCCACTACGGAACAGTGTGGAGATTCCTT AAAGAACTAAAAGTAGAGGAATGTATCCATTTCTCTAGATTTCTAGTTTATTG

hs_chr16 + GCCACTACGGAACAGTGTGGAGATTCCTT AAAGAACTAAAAGTAGAGGAATGTATCCATTTCTCTAGATTTCTAGTTTATTG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAGAACTAAAAGTAGAGGAATGTATCCATTTCTCTAGATTTCTAGTTTATTG

pHLE4. 9a
5' end ACTTTACAAGAGGCTTTAAGGGAGTGGTAAACATGAAATGAAAGACCATATGTATACATGTGCCATGCTGGTGGGCTGCACCCACTAATGT

hs_chr2 - ACTTTACAAGAGGCTTTAAGGGAGTGGTAAACATGAAATGAAAGACCATATGTATACATGTGCCATGCTGGTGGGCTGCACCCACTAATGT

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA GAAAGACCATATGTATACATGTGCCATGCTGGTGGGCTGCACCCACTAATGT

pHLE5. 16a
5' end GGTTTGTCTAAAAATTTGGACTAAGCAGGAAGGAATGTTTAAAGTTAAGATAAGGATTAACCGTATTACCGCCATGCATCTGGAATTCGGCTTACGG

hs_chrX + GGTTTGTCTAAAAATTTGGACTAAGCAGGAAGGAATGTTTAAAGTTAAGATAAGGATTAACCGTATTACCGCCATGCATCTGGAATTCGGCTTACGG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA GGTTAAGATAAGGATTAACCGTATTACCGCCATGCATCTGGAATTCGGCTTACGG

pHLE6. 19a
5' end TCTCACACTGTTAAACCTTCCTTTTGATGAGAAGTTTGAAGATCTGTTTTGGAGGCCTAGGCTTTGCAACGGCTCTGCAGCAG

hs_chr11 - TCTCACACTGTTAAACCTTCCTTTTGATGAGAAGTTTGAAGATCTGTTTTGGAGGCCTAGGCTTTGCAACGGCTCTGCAGCAG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA GAAATCTGTTTTATAGAATCTCCCAATGGATATTTTGAGGGCTTTGA

pHLE7. 8a
5' end ATTTGCGAGCATAGAACTGGTCTAAGTGAACATATAAAATAGAAATGCTCACATGTTCTTCTCGGTTATCCCTGATTC

hs_chr5 - ATTTGCGAGCATAGAACTGGTCTAAGTGAACATATAAAATAGAAATGCTCACATGTTCTTCTCGGTTATCCCTGATTC

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA TAAAAATAGAAATGCTCACATGTTCTTCTCGGTTATCCCTGATTC

pHLE7. 20b
5' end CATTGAGCCATACCTAGAATGTCAAATGGTCTCTCTTAACTGGAATGAAAAAGAGGTTTACTGCTTTAAAAACCTCCACACCTCCCCCTG

hs_chr5 + CATTGAGCCATACCTAGAATGTCAAATGGTCTCTCTTAACTGGAATGAAAAAGAGGTTTACTGCTTTAAAAACCTCCACACCTCCCCCTG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA GAAAGAGATGAAAGAACTGTGTGACCAGGAGGCGACTGTAACCTATGA

pHLE8. 12a
5' end TCAATCGAGTAAAAATAAAAAGAAATAATTAACAATAAACAATATATTTGAAGTCAGTAGTGTGATGCCCTCCAGC

hs_chr11 + TCAATCGAGTAAAAATAAAAAGAAATAATTAACAATAAACAATATATTTGAAGTCAGTAGTGTGATGCCCTCCAGC

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA ACAATATATTGATAGATTAATAAGAAAGATATACCT

pHLE8. 25a
5' end ATACAGGACCCATTCTTGTGGCAGCTGGATGAAAAGACATGGGTTTAAACTTCATTTTTAATTTAAAGGATCTAGG

hs_chr5 - ATACAGGACCCATTCTTGTGGCAGCTGGATGAAAAGACATGGGTTTAAACTTCATTTTTAATTTAAAGGATCTAGG

3' end AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA GAAAAGACATGGGTTTATCTTGTGCAAGGCCAAGGCTATTTAATATT