Published in final edited form as:

Methods Mol Biol. 2018 ; 1769: 169–181. doi:10.1007/978-1-4939-7780-2_11.

A Role for Retrotransposons in Chromothripsis

Hancks DC^{1,2}

¹Eccles Institute of Human Genetics, School of Medicine, University of Utah, Salt Lake City, UT, USA.

²Department of Immunology, UT Southwestern Medical Center, Dallas, TX, USA.

Abstract

Chromothripsis is a mutational event driven by tens to hundreds of double-stranded DNA breaks which occur in a single event between a limited number of chromosomes. Following chromosomal shattering, DNA fragments are stitched together in a seemingly random manner resulting in complex genomic rearrangements including sequence shuffling, deletions, and inversions of varying size. This genomic catastrophe has been observed in cancer genomes and the genomes of patients harboring developmental and congenital defects. The mechanisms catalyzing DNA breakage and coordinating the “random” assembly of genomic fragments are actively being investigated. Recently, retrotransposons—a type of “jumping gene”—have been implicated as one means to generate double-stranded DNA breaks during chromothripsis and as sequences which can contribute to the final configuration of the derived chromosomes. In this methods chapter, I discuss how to apply available bioinformatic tools and the hallmarks of retrotransposon mobilization to breakpoint junctions to assess the role for active and inactive retrotransposon sequences in chromothriptic events.

Keywords

Chromothripsis; Retrotransposon; *Alu*; L1; LINE-1; SVA; Nonallelic homologous recombination; Deletion; Inversion

1 Introduction

It is well established that genetic mutations result in human disease including the initiation and progression of cancer. Thus, significant effort has been invested in characterizing the mechanisms and nature of genetic mutations in patients. Mutational analysis has been transformed over the past decade with the implementation of high-throughput next-generation DNA sequencing (NGS). Together with advances in computational biology, whole-genome sequencing (WGS) has become feasible and commonplace. Perhaps naturally, whole-genome sequencing analysis has been rapidly applied to cancers of various origins and stages to understand genetic factors shaping the biology of these malignancies.

An important finding came when characterizing paired-end sequencing data from the genome of a patient with chronic lymphocytic leukemia [1] with 42 somatically acquired rearrangements involving four chromosomes. Additional analysis of other cancer genomes using NGS identified similar mutational profiles. Specifically, novel complex rearrangements, which included sequence shuffling, inversions, and oscillation between two copy number states involving only a few chromosomes, indicated pulverization of DNA in a single event. Subsequently, this phenomenon was termed chromothripsis—meaning “chromosome shattering” [1].

To date, chromothripsis has now been observed in many types of cancers with some types of bone cancers being the most prevalent [1-3]. This complex class of rearrangements in cancer cells often involves tens to hundreds of chromosomal fragments dispersed across only a limited number of chromosomes (e.g., 2–4). Further investigations have identified similar genomic rearrangements in patients with congenital abnormalities [4, 5]. Notably, at least one instance has been reported where a stably segregating chromothripsis event was identified in healthy individuals [6, 7].

The cellular mechanisms driving chromothripsis and mediating DNA repair of damaged chromosomes are actively being investigated. Of particular interest are defects in chromosomal segregation which can result in the formation of structures termed micronuclei [8, 9]. Chromosomes partitioned into micronuclei often become highly fragmented and reassembled in a random manner due to defects in DNA replication. Likewise, hallmarks of chromothripsis have been observed in cells where telomeres have fused—a phenomenon known as telomere crisis—to generate dicentric chromosomes [10]. Along with these studies, one report identified a role for mobile DNA such as retrotransposons, also known as “jumping genes,” as factors contributing to a germline chromothripsis event [7].

Retrotransposons are genetic elements widespread in nature and major components of most genomes (Fig. 1) [11]. In the human genome, retrotransposons are responsible for ~35% of the genomic real estate [12]. Long Interspersed Element-1 (LINE-1 or L1) is the only retrotransposon in humans able to duplicate itself (autonomous) (Fig. 1a). Mobilization of L1 sequences occurs via an RNA intermediate—a process termed retrotransposition. A full-length active L1 encodes two proteins, both of which are required for retrotransposition *in cis* [13]. ORF1p displays RNA-binding activity, while ORF2p encodes reverse transcriptase [14] and DNA endonuclease (EN) [15] activities. Along with its own RNA, L1 ORF2p is able to copy other RNAs, such as *Alu* elements (Fig. 1b), hominid-specific SVA elements (Fig. 1c), and even protein-coding RNAs (e.g., processed pseudogenes) (Fig. 1d), to new genomic locations by a coupled integration reverse-transcription mechanism [16] (Fig. 2a). Briefly, DNA is cleaved at an L1 EN target sequence (5'-TTTT/AA-3') to produce a free 3'-OH which is subsequently used by ORF2p to prime first-strand cDNA synthesis. Along with contributing to genome expansion through RNA-based duplication, it is well established that due to their high copy numbers, retrotransposons can result in nonallelic homologous recombination (NAHR) to generate indels or genetic inversions [17, 18].

Analysis of a stably segregating germline chromothripsis event [6, 7] identified the insertion of a 5'-truncated SVA element at one of the breakpoint junctions (BPJs) into a sequence

resembling the L1 EN consensus cleavage site (5'-TTTT/AA-3'). This insertion was associated with a 110 kb deletion in the derived chromosome 5' - of the SVA and attached to a distal DNA fragment located ~4.7 megabases away relative to the reference genome. Notably, the 5'-end of the SVA insertion displayed microhomology with sequence at the BPJ of the distal fragment. Previous studies have indicated that reverse transcription by ORF2p of RNAs like SVA may function as “molecular Band-Aids” at double-stranded DNA (dsDNA) breaks [19] and that L1-mediated insertions are often associated with large genomic deletions (Fig. 2b) [20, 21]. These data suggest a model where ORF2p EN activity can be a source of DNA breaks and may bridge distal DNA fragments brought in close proximity presumably through chromatin looping during chromothripsis [7].

Further characterization of this mutational event identified two pairs of *Alu* elements flanking four breakpoint junctions, one of which was associated with the SVA insertion. Specifically, in the reference genome, there were two *Alu* elements in the same orientation flanking the 110 kb sequence deleted on the derived chromosome. In the second instance, there were two *Alu* elements on the same chromosome in opposite orientation relative to each other downstream of the 110 kb deletion flanking a 620 kb sequence inverted during chromothripsis. Both *Alu*-mediated deletions (Fig. 3a) and inversions (Fig. 3b) have been characterized as mechanisms contributing to primate genome evolution [22, 23], and in some cases, *Alu*-mediated rearrangements have resulted in single-gene disease [17]. These data highlight how *Alu* elements in the genome can contribute to rearrangements in germline chromothripsis events [7].

Perhaps the most striking about this case of chromothripsis was that at some of the BPJs, including the distal BPJ which the 5'-end of the SVA element was ultimately attached to, sequences resembling the L1 EN cleavage site were identified. Together these data indicate a model by which active and inactive retrotransposons can contribute to chromothriptic rearrangements (Fig. 3c): (1) generation of dsDNA breaks, (2) orientation of DNA fragments to impact derived chromosome structure, (3) function as molecular Band-Aids, and (4) bridge distal DNA fragments. Interestingly, several studies indicate that retrotransposons, such as L1, are active where chromothripsis is known to occur, namely, cancer and cells responsible for producing germ cells (e.g., embryonic stem cells) [24, 25]. Based on these observations, methods are described below to aid in the characterization of retrotransposon sequences in chromothriptic events.

2 Materials

The methods here begin making the assumption that the breakpoint junctions from the DNA sequence analysis have been identified. Briefly, the UCSC genome browser (genome.ucsc.edu) [26] is a powerful tool to analyze breakpoint junctions. In addition, the analysis outlined below involves manual inspection of breakpoint sequences complemented by the UCSC genome browser for visualization of genomic features such as repetitive elements like *Alu* sequences.

2.1 Breakpoint Junction Sequences

These data should be obtained by aligning sequencing data to the reference genome sequence (e.g., hg19). Several computational tools have been developed to aid in the characterization of retrotransposon insertions (reviewed in [27]).

2.2 RepeatMasker (www.repeatmasker.org)

RepeatMasker [28] is an analysis program freely available online. It contains libraries and annotations for known repetitive elements including retrotransposon sequences.

RepeatMasker is frequently used by transposon researchers and in the annotation of transposable elements during genome analysis. RepeatMasker is able to identify different types of retrotransposons along with the subfamily which they belong to. Notably, only certain retrotransposon sequences, such as *AluYa5* and *L1-Hs*, are active (e.g., retrotransposition competent) in the human genome; however, the ability to mitigate nonallelic homologous recombination by a retrotransposon is not known to be restricted to specific subfamilies.

2.3 UCSC Genome Browser (genome.ucsc.edu)

The UCSC genome browser [26] is a powerful tool to easily visualize a variety of genomic features including repetitive sequences like retrotransposons.

1. To display annotations for repetitive DNA, from the browser page, scroll below the browser window to the section labeled “Repeats” and select “full” from the drop-down heading below the RepeatMasker heading. This selection will add a “track” on the browser displaying different repeat families (e.g., LINE, SINE). The color of each annotated repeat (gray-scaled box) corresponds to the similarity of this repeat to the known RepeatMasker consensus.
2. To obtain information pertaining to specific repetitive elements, select the repeat (gray box). This will display a subsequent page and element specific information including subfamily, sequence length of element, and element orientation (e.g., positive strand). The orientation of the element is useful in determining whether sequences such as a pair of *Alu* elements may have contributed to the rearrangement resulting in either an indel or genetic inversion.

3 Methods

3.1 Characterization of Retrotransposon Element Insertions

1. If insertions of notable size are identified at breakpoint junctions in a chromothriptic genome, these sequences should be screened using RepeatMasker to determine their identity including type and subfamily. Indeed, not all insertions identified at breakpoint junctions may be derived from retrotransposon sequences. While insertion of a small number of nucleotides is likely due to cellular DNA repair machinery, some retrotransposon sequences can be very short in length (e.g., 46 bp) or may only consist of a homopolymer of adenines derived from the polyA tail from the mRNA (see Note 1).

2. Manually characterize insertions for hallmarks of L1-mediated retrotransposition (Fig. 2). Retrotransposon insertions can be broadly grouped into two classes [11, 29]: (1) L1 endonuclease-dependent (Fig. 2a) and (2) L1 endonuclease-independent (ENi) (Fig. 2c) [19, 30] (*see Note 2*). The most common hallmarks of EN-dependent insertions include:
 - a. Insertion at a sequence resembling the L1 EN site, polyA tail of variable length at the 3'-end of the insertion, most insertions being 5'-truncated (e.g., lacking sequence at the 5'-end relative to the full-length retrotransposon consensus sequence), target-site duplication of varying length (4–20 bp) directly flanking the insertion, 3'-transductions (e.g., duplication of downstream sequence due to transcriptional readthrough) (*see Note 3*), and occasionally untemplated nucleotides at the 5'-end of the insertion
3. If a retrotransposon sequence is present at a BPJ but does not display hallmarks of L1 EN-dependent insertions, the insertion may have occurred in an L1 EN-independent manner (Fig. 2c) (*see Notes 4 and 5*). In the context of chromothripsis, this class of insertions has the potential to serve as a bridge through microhomology-mediated interactions to link fragmented DNA. Common hallmarks of L1 ENi insertions include:
 - a. Truncation at both the 5'- and 3'-ends, lack of polyA tail, and insertion of untemplated nucleotides 5'- and 3'- of the insertion

3.2 Characterization of L1 Endonuclease Cleavage Sites

1. Assuming no additional mutations have occurred, the easiest means to identify the L1 EN cleavage site is to inspect the 5'-target-site duplication (Fig. 2a). For example, if the following bold sequence 5'-**TTAAAACTG**-3' is the 5'-TSD in the same orientation as the L1-mediated insertion, then the L1 EN cleavage site was (relative to the bottom strand, underlined) 5'-TTTT/AA-3' where the "/" indicates the insertion site (*see Notes 6 and 7*).

¹The short insertion size of retrotransposons is due to their inherent nature to become truncated—primarily at their 5'-end—during integration. For example, an *Alu* insertion was identified in the *SERPINC1* gene associated with type 1 antithrombin deficiency only six base pairs long with a 40 bp polyA tail [29, 32]. Similarly, insertions containing only a homopolymer stretch of adenines (e.g., mRNA polyA tail) and lacking any known retrotransposon sequence have also been characterized [33].

²Majority of insertions are L1 EN dependent (Fig. 2a) and integrate by a mechanism termed target-primed reverse transcription (TPRT) [16]. Briefly, ORF2p cleaves the bottom strand of DNA at a sequence resembling its target site (5'-YYYY/RR-3', where Y = pyrimidine and R = purine with 5'-TTTT/AA-3' being most frequent) [15, 34, 35]. Cleavage liberates a 3'-OH which is used to prime first-strand cDNA synthesis by ORF2p. Currently, the mechanism of top-strand DNA cleavage and 5'-attachment of the cDNA to upstream DNA are incomplete. The resolution of the insertion presumably involves cellular DNA repair machinery.

³Weak polyA signals encoded by L1 and SVA often promote transcriptional readthrough by RNA polymerase in turn generating "chimeric" RNAs consisting of the retrotransposon sequence, a polyA tail, 3'-transduced downstream sequence, and another polyA tail (Fig. 1) [36, 37].

⁴L1 ENi insertions (Fig. 2c) are less common and were first identified in cell lines lacking key DNA repair proteins using engineered L1 elements containing an inactivating mutation in the L1 endonuclease domain of ORF2p [19]. L1 ENi insertions have since been identified and characterized in detail in the human genome [30].

⁵Another sequence feature occasionally observed for both L1 EN and ENi insertions is deletion of genomic sequence at the insertion site also known as target-site deletions (Fig. 2b). For example, as noted above, a large deletion of 110 kb occurred 5'- of the inserted SVA element in the case of Nazaryan et al. [7]. For L1 EN insertions associated with deletions, a sequence resembling the L1 EN consensus cleavage site (5'-TTTT/AA-3') should be identifiable at the pre-insertion locus, also commonly referred to as the empty site.

3.3 Characterization of the Contribution of Inactive Retrotransposons in Chromothripsis

Chromothripsis is characterized by rearrangements, e.g., inversions and deletions involving numerous DNA fragments [31]. It has been established that direct repeats such as two *Alu* elements on different chromosomes or the same one may mediate nonallelic homologous recombination resulting in sequence deletion, duplication, or inversion (Fig. 3) [17, 18].

1. To identify similar events at breakpoint junctions, go to the BPJ on UCSC genome browser with the RepeatMasker track on.
2. Look for two of the same type of repetitive elements (e.g., *Alu* elements) immediately adjacent to the BPJ (*see* Note 8). In the case of a deletion or inversion in the derived chromosome, one expectation if repetitive sequences were involved in NAHR is that you are able to identify two repeats of the same family flanking the BPJ either in the same or opposite orientation (*see* Note 9 and 10).
3. If repetitive (retrotransposon) elements are identified in the reference genome flanking a deletion, duplication, or inversion, inspect the sequence element in the derived chromosome for chimeric nature—a hallmark of recombination—using nucleotide sequence alignment (*see* Note 8).

Acknowledgments

D.C.H. is funded by a K99/R00 Pathway to Independence Award from the National Institutes of Health (USA, NIGMS) and the Cancer Prevention & Research Institute of Texas (CPRIT).

References

1. Stephens PJ, Greenman CD, Fu B et al. (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144:27–40. 10.1016/j.cell.2010.11.055 [PubMed: 21215367]

⁶It has long been hypothesized that if left unrestrained, ORF2p EN can be deleterious to the cell [38]. Although the L1 EN is thought to cleave only its target sequence on the bottom strand during classical TPRT, evidence from insertions resulting in single-gene disease lacking TSDs but inserting at L1 EN sites and Nazaryan et al. [7] suggest that ORF2p can cleave the top-strand following bottom-strand cleavage at sequences resembling the L1 EN cleavage consensus sequence.

⁷The canonical L1 cleavage site is 5'-TTTT/AA-3' with almost all variations displaying similarity to 5'-YYYY/RR-3'. To date, several datasets including those characterizing L1-mediated insertions using next-generation sequencing have been published (reviewed in [27]). Gilbert et al. (2005) [34] contains a list of 100 characterized insertions produced by engineered L1s in cell culture. Hancks and Kazazian [29] contains an up-to-date list of over 120 published insertions resulting in single-gene disease in humans and their hallmarks of retrotransposition.

⁸Retrotransposon sequences including *Alu* elements are very common in the human genome. Repeat elements involved in a NAHR event must be directly at the breakpoint and not “nearby.” Due to their abundance, many repeats by chance will likely be close to a breakpoint. Further indication that a repetitive (retrotransposon) element was involved in a NAHR event is the presence of either a solo chimeric element in the case of a deletion or duplication or two chimeric elements in the case of an inversion in the derived chromosomes.

⁹Along with L1s, *Alus*, and SVA elements, endogenous retroviruses (ERVs), which comprise ~8% of the human genome, are also frequent mediators of NAHR [39]. For instance, most ERVs in the genome are present as solo long terminal repeats (LTRs). Considering that the majority of ERV insertions initially are full length (LTR-gag-pol-env-LTR), these solo LTRs are evidence that ERVs frequently undergo NAHR presumably to reduce the burden of these selfish genetic elements.

¹⁰While high-sequence identity is important to facilitate NAHR between two repetitive sequences, it is not absolutely required. For example, in Nazaryan-Petersen et al. [7], the inversion was mediated by two *Alu* elements from two distantly related subfamilies *AluSq* and *AluJb*, which share ~77% nucleotide identity by BLASTn analysis.

2. Forment JV, Kaidi A, Jackson SP (2012) Chromothripsis and cancer: causes and consequences of chromosome shattering. *Nat Rev Cancer* 12:663–670. 10.1038/nrc3352 [PubMed: 22972457]
3. Storchová Z, Kloosterman WP (2016) The genomic characteristics and cellular origin of chromothripsis. *Curr Opin Cell Biol* 40:106–113. 10.1016/j.ceb.2016.03.003 [PubMed: 27023493]
4. Kloosterman WP, Guryev V, van Roosmalen M et al. (2011) Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. *Hum Mol Genet* 20:1916–1924 [PubMed: 21349919]
5. Kloosterman WP, Cuppen E (2013) Chromothripsis in congenital disorders and cancer: similarities and differences. *Curr Opin Cell Biol* 25:341–348. 10.1016/j.ceb.2013.02.008 [PubMed: 23478216]
6. Bertelsen B, Nazaryan-Petersen L, Sun W et al. (2016) A germline chromothripsis event stably segregating in 11 individuals through three generations. *Genet Med* 18:494–500. 10.1038/gim.2015.112 [PubMed: 26312826]
7. Nazaryan-Petersen L, Bertelsen B, Bak M et al. (2016) Germline chromothripsis driven by L1-mediated retrotransposition and Alu/Alu homologous recombination. *Hum Mutat* 37:385–395. 10.1002/humu.22953 [PubMed: 26929209]
8. Crasta K, Ganem NJ, Dagher R et al. (2012) DNA breaks and chromosome pulverization from errors in mitosis. *Nature* 482:53–58. 10.1038/nature10802 [PubMed: 22258507]
9. Zhang C-Z, Spektor A, Cornils H et al. (2015) Chromothripsis from DNA damage in micronuclei. *Nature* 522:179–184. 10.1038/nature14493 [PubMed: 26017310]
10. Maciejowski J, Li Y, Bosco N et al. (2015) Chromothripsis and kataegis induced by telomere crisis. *Cell* 163:1641–1654. 10.1016/j.cell.2015.11.054 [PubMed: 26687355]
11. Richardson SR, Doucet AJ, Kopera HC et al. (2015) The influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Microbiol Spectr* 3:MDNA3-0061–2014. 10.1128/microbiolspec.MDNA3-0061-2014
12. Kazazian HH Jr, Moran JV (2017) Mobile DNA in health and disease. *N Engl J Med* 377:361–370. 10.1056/NEJMra1510092 [PubMed: 28745987]
13. Moran JV, Holmes SE, Naas TP et al. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell* 87:917–927. 10.1016/S0092-8674(00)81998-4 [PubMed: 8945518]
14. Mathias SL, Scott AF, Kazazian HH et al. (1991) Reverse transcriptase encoded by a human transposable element. *Science* 254:1808–1810 [PubMed: 1722352]
15. Feng Q, Moran JV, Kazazian HH Jr et al. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87:905–916. 10.1016/S0092-8674(00)81997-2 [PubMed: 8945517]
16. Luan DD, Korman MH, Jakubczak JL et al. (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. 10.1016/0092-8674(93)90078-5 [PubMed: 7679954]
17. Deininger PL, Batzer MA (1999) Alu repeats and human disease. *Mol Genet Metab* 67:183–193 [PubMed: 10381326]
18. Cordaux R, Batzer MA (2009) The impact of retrotransposons on human genome evolution. *Nat Rev Genet* 10:691–703. 10.1038/nrg2640 [PubMed: 19763152]
19. Morrish TA, Gilbert N, Myers JS et al. (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* 31:159–165. 10.1038/ng898 [PubMed: 12006980]
20. Symer DE, Connelly C, Szak ST et al. (2002) Human L1 retrotransposition is associated with genetic instability in vivo. *Cell* 110:327–338. 10.1016/S0092-8674(02)00839-5 [PubMed: 12176320]
21. Gilbert N, Lutz-Prigge S, Moran JV (2002) Genomic deletions created upon LINE-1 retrotransposition. *Cell* 110:315–325. 10.1016/S0092-8674(02)00828-0 [PubMed: 12176319]
22. Han K, Lee J, Meyer TJ et al. (2007) Alu recombination-mediated structural deletions in the chimpanzee genome. *PLoS Genet* 3:1939–1949. 10.1371/journal.pgen.0030184 [PubMed: 17953488]
23. Lee J, Han K, Meyer TJ et al. (2008) Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. *PLoS ONE* 3:e4047 10.1371/journal.pone.0004047 [PubMed: 19112500]

24. Burns KH (2017) Transposable elements in cancer. *Nat Rev Cancer* 17:415–424. 10.1038/nrc.2017.35 [PubMed: 28642606]
25. Faulkner GJ, Garcia-Perez JL (2017) L1 Mosaicism in mammals: extent, effects, and evolution. *Trends Genet.* 10.1016/j.tig.2017.07.004
26. Kent WJ, Sugnet CW, Furey TS et al. (2002) The human genome browser at UCSC. *Genome Res* 12:996–1006 [PubMed: 12045153]
27. Ewing AD (2015) Transposable element detection from whole genome sequence data. *BMC Mobile DNA* 6:24 10.1186/s13100-015-0055-3
28. Smit AFA, Hubley R & Green P RepeatMasker Open-4.0.2013-2015 <http://www.repeatmasker.org>
29. Hancks DC, Kazazian HH (2016) Roles for retrotransposon insertions in human disease. *Mob DNA* 7:9 10.1186/s13100-016-0065-9 [PubMed: 27158268]
30. Sen SK, Huang CT, Han K et al. (2007) Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome. *Nucleic Acids Res* 35:3741–3751. 10.1093/nar/gkm317 [PubMed: 17517773]
31. Zhang CZ, Leibowitz ML, Pellman D (2013) Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes Dev* 27:2513–2530. 10.1101/gad.229559.113 [PubMed: 24298051]
32. Beauchamp NJ, Makris M, Preston FE et al. (2000) Major structural defects in the antithrombin gene in four families with type I antithrombin deficiency—partial/complete deletions and rearrangement of the antithrombin gene. *Thromb Haemost* 83:715–721 [PubMed: 10823268]
33. Chen J-M, Stenson PD, Cooper DN et al. (2005) A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. *Hum Genet* 117:411–427. 10.1007/s00439-005-1321-0 [PubMed: 15983781]
34. Gilbert N, Lutz S, Morrish TA et al. (2005) Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol Cell Biol* 25:7780–7795. 10.1128/MCB.25.17.7780-7795.2005 [PubMed: 16107723]
35. Hancks DC, Kazazian HH Jr (2012) Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* 22:191–203. 10.1016/j.gde.2012.02.006 [PubMed: 22406018]
36. Moran JV, DeBerardinis RJ, Kazazian HH (1999) Exon shuffling by L1 retrotransposition. *Science* 283:1530–1534 [PubMed: 10066175]
37. Xing J, Wang H, Belancio VP et al. (2006) Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc Natl Acad Sci U S A* 103:17608–17613. 10.1073/pnas.0603224103 [PubMed: 17101974]
38. Gasior SL, Wakeman TP, Xu B et al. (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* 357:1383–1393. 10.1016/j.jmb.2006.01.089 [PubMed: 16490214]
39. Mager DL, Stoye JP (2015) Mammalian endogenous retroviruses. *Microbiol Spectr* 3 :MDNA3–0009-2014. 10.1128/microbiolspec.MDNA3-0009-2014
40. Batzer MA, Deininger PL (2002) Alu repeats and human genomic diversity. *Nat Rev Genet* 3:370–379. 10.1038/nrg798 [PubMed: 11988762]
41. Hancks DC, Kazazian HH (2010) SVA retrotransposons: evolution and genetic instability. *Semin Cancer Biol* 20:234–245. 10.1016/j.semcancer.2010.04.001 [PubMed: 20416380]
42. Hancks DC, Ewing AD, Chen JE et al. (2009) Exon-trapping mediated by the human retrotransposon SVA. *Genome Res* 19:1983–1991. 10.1101/gr.093153.109 [PubMed: 19635844]
43. Damert A, Raiz J, Horn AV et al. (2009) 5'-transducing SVA retrotransposon groups spread efficiently throughout the human genome. *Genome Res* 19:1992–2008. 10.1101/gr.093435.109 [PubMed: 19652014]
44. Wei W, Gilbert N, Ooi SL et al. (2001) Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol* 21:1429–1439. 10.1128/MCB.21.4.1429-1439.2001 [PubMed: 11158327]
45. Esnault C, Maestre J, Heidmann T (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 24:363–367. 10.1038/74184 [PubMed: 10742098]

46. Zhang Z, Harrison PM, Liu Y, Gerstein M (2003) Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res* 13:2541–2558 [PubMed: 14656962]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1. Retrotransposons active in the human genome [11]. (a) Long Interspersed Element-1 (LINE-1 or L1) is the only autonomous retrotransposon in the human genome (> 500,000 copies). Proteins encoded by L1 (ORF1p and ORF2p) are required for mobilization of its own RNA and other retrotransposon RNAs. L1 encodes its own promoter in the 5'-UTR (*black bent arrow*), a protein with RNA-binding activity (ORF1p), and a protein with reverse transcriptase and endonuclease activities (ORF2p). A full-length L1 terminates at its 3'-end in a polyA tail of variable length (AAA_n). Majority of L1-mediated retrotransposition events are flanked by a target-site duplication of varying length (4–20 bp, *black horizontal arrows*). Occasionally, transcriptional readthrough will occur by bypassing the weak polyA signal (AATAA) and terminating transcription at a downstream sequence in the flank generating a chimeric transcript. If mobilized, these chimeric transcripts generate 3'-transductions [36]. (b) *Alu* elements are ~300 bp active primate-specific Short Interspersed Elements (SINEs), > 1.2 million in the human genome [40]. *Alus* are RNA pol III transcripts which encode an (a) and (b) box. Efficient *Alu* transcription requires an upstream enhancer sequence and a downstream pol III terminator (TTTT). (c) SINE-VNTR-Alu (SVA) elements are active composite hominid-specific retrotransposons [41], >2700 copies in the human genome. Starting at the 5'-end, SVA elements consist of a hexameric repeat (CCCTCT)_n, an *Alu*-like

domain derived from two antisense *Alu* elements, a GC-rich VNTR, and a SINE-R domain derived from an extinct HERV-K retrovirus. SVA transcription can occur upstream or within the element. Upstream transcriptional initiation [42, 43] or transcriptional readthrough can generate chimeric transcripts which if mobilized will produce 5' - and 3' -transductions [37], respectively. **(d)** In addition to mobilizing retrotransposon RNAs, L1 can copy and paste cellular RNAs to new genomic locations to produce processed pseudogenes [44, 45] also known as retrogenes (~8000 in human genome [46]). Typically, processed pseudogene insertions resemble cDNA sequences (e.g., lack introns), terminate in a 3' -polyA tail, and are flanked by a target-site duplication. *UTR* untranslated region, *ORF* open-reading frame, *AA* amino acid, *kB* kilobase, *VNTR* variable number of tandem repeats, *env* envelope, *LTR* long terminal repeat. Not drawn to scale

Fig. 2.

Modes of L1-mediated retrotransposition. **(a)** Majority of insertions occur by a coupled integration and reverse-transcription mechanism termed target-primed reverse transcription (TPRT) [16]. Specifically, L1 ORF2p cleaves the bottom strand of its target sequence (5'-TTT/AA-3') followed by reverse transcription of L1 RNA (*gray line*) using the liberated 3'-OH. The mechanism of top-strand cleavage and resolution of the insertion are incomplete but likely involve cellular factors including DNA repair proteins. Briefly, the insertion is characterized by a target-site duplication (*horizontal black arrows*), 3'-polyA tail, and in this instance 5'-truncation. Most genomic L1 insertions (99%) are highly truncated at their 5'-end. **(b)** L1 EN-dependent target-site deletions. Although this pathway also requires L1 EN activity, the major difference is that a double-strand break occurs upstream of L1 EN bottom-strand cleavage. This DSB ultimately results in the loss of DNA sequence 5'- of the insertion. It is hypothesized that microhomology interactions either by the RNA or nascent first-strand cDNA stabilize this intermediate. These insertions are characterized by deletions 5'- of the target-site (*gray box*), L1 EN cleavage site, frequent 5'-truncations, and a 3'-polyA tail. **(c)** L1 EN-independent insertions are an alternative means by which retrotransposon sequences can be deposited into the genome. Specifically, it has been reported that when double-strand breaks occur in DNA, retrotransposon RNAs can function

as molecular Band-Aids by bridging DNA fragments through microhomology interactions. As in the other pathways, L1 ORF2p reverse transcribes the RNA into cDNA at the DNA break. These insertions are characterized by integration at sites sharing no resemblance to L1 EN cleavage site, target-site deletions, 5' - and 3' -truncations of the retrotransposon sequence including lack of a 3' -polyA tail, and untemplated nucleotides (ins) at the junctions of the insertion. cDNA—complementary DNA. Not drawn to scale

C) Roles for active and inactive retrotransposons in chromothriptic rearrangements

Fig. 3. Genomic rearrangements mediated by genomic retrotransposons. The rearrangements described here can be produced, in principle, by any genomic repeat. **(a)** Retrotransposon recombination-mediated deletions (RRMDs) can transpire due to nonallelic homologous recombination. Mispairing of nonhomologous repetitive elements in the same orientation, such as *Alus* (black and gray arrows), between homologous chromosomes (interchromosomal, *left*), can generate a duplication on one chromosome and a deletion on the other, while mispairing between repetitive elements on the same chromosome (intrachromosomal, *right*) can result in a genomic deletion and a ring fragment. RRMDs have resulted in human genetic disease [17] and contributed significantly to genome evolution [22]. A hallmark of recombination between repetitive elements is the production of a chimeric element (*half-black, half-gray arrow*) [40]. **(b)** Retrotransposon recombination-mediated inversions can occur when repetitive elements such as *Alus* in opposite orientation mispair followed by recombination [23]. Recombination results in the generation of

chimeric elements (*half-black, half-gray arrows*) and an inversion of the intervening sequencing (reversal of *black arrowheads* on *gray line*). (c) Model for how retrotransposons may contribute to chromothripsis. NAHR between retrotransposons (*black and gray arrows*) or other repetitive elements can produce genomic deletions or inversions. Retrotransposon RNAs can serve as a means to bridge distal fragments via microhomology interactions. L1 ORF2p is a source of DNA endonuclease activity. Not drawn to scale