Genomics 104 (2014) 96–104

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

De novo *LINE-1* retrotransposition in HepG2 cells preferentially targets gene poor regions of chromosome 13

Pasano Bojang Jr. [a], Mark J. Anderton [b], Ruth A. Roberts [b], Kenneth S. Ramos [a,c,*]

[a] Department of Biochemistry and Molecular Biology, University of Louisville School of Medicine, Louisville, KY 40292, USA
[b] Department of Toxicology Sciences, Safety Assessment, AstraZeneca R&D, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK
[c] Center for Environmental Genomics and Integrative Biology, University of Louisville School of Medicine, Louisville, KY 40292, USA

ARTICLE INFO

Article history:
Received 11 February 2014
Accepted 11 July 2014
Available online 17 July 2014

Keywords:
Long interspersed nuclear element-1 (Line-1/L1)
Retrotransposons
HepG2
Fluorescence in situ hybridization
Chromosome 13

ABSTRACT

Long interspersed nuclear elements (Line-1 or L1s) account for ~17% of the human genome. While the majority of human L1s are inactive, ~80–100 elements remain retrotransposition competent and mobilize through RNA intermediates to different locations within the genome. De novo insertions of L1s account for polymorphic variation of the human genome and disruption of target loci at their new location. In the present study, fluorescence in situ hybridization and DNA sequencing were used to characterize retrotransposition profiles of $L1^{RP}$ in cultured human HepG2 cells. While expression of synthetic $L1^{RP}$ was associated with full-length and truncated insertions throughout the entire genome, a strong preference for gene-poor regions, such as those found in chromosome 13 was observed for full-length insertions. These findings shed light into L1 targeting mechanisms within the human genome and question the putative randomness of L1 retrotransposition.

1. Introduction

A full-length human long interspersed nuclear element-1 (Line-1 or *L1*) is approximately 6 kb and contains four major components: a 5′ untranslated region (5′UTR), two open reading frames (ORFs) separated by a 63 bp inter-ORF region and a 3′UTR with a poly A tail and signal [1,2]. *L1* 5′-UTR is 907 bp long and contains an internal promoter that harbors several transcription factor binding sites including, Yin Yang-1 (YY1), Sox11, E2F and RUNX3 transcription factors [3–7]. ORF1 encodes a ~40 kDa protein with RNA binding and nucleic acid chaperone activities [8]. ORF2 encodes a ~150 kDa protein with endonuclease (EN), reverse transcriptase (RT) activities, and a zinc finger domain (ZF) believed to mediate ORF2–DNA interactions [9–11]. *L1s* have been shown to mobilize through target-primed reverse transcription (TPRT), also known as "copy and paste" mechanism; although alternate mechanisms are known to exist [12,13].

Full length *L1* mRNA is transcribed from its internal promoter by RNA polymerase II/III and exported to the cytoplasm [3]. Upon translation, ORF1p and ORF2p exhibit *cis*-preference and bind their encoding mRNA to form a ribonucleoprotein particle (RNP) [14,15]. *L1* RNP translocates into the nucleus, nicks a single-strand of genomic DNA to expose a 3′-OH group which is then used by the RT-domain of ORF2 to prime and synthesize the first strand of *L1* cDNA [16,17]. In the majority of cases, ORF2 falls off the template before reaching the 5′end due to its non-processive nature and the presence of premature cryptic polyadenylation sites in the ORF2 cDNA sequence [18]. This explains the overwhelming presence of truncated *L1s* littered throughout the genome. For example, there are ~516,000 copies of *L1* in the human genome, with 80–100 estimated to be full-length and retrotransposition competent [19,20]. The RNP of *L1* can also act in *trans* to mobilize short interspersed elements (SINES), such as *Alu* sequences, noncoding RNAs such as U6 snRNA, and some cellular mRNAs leading to the formation of processed pseudogenes [21–24]. The integration of *L1* sequences near genes can modulate their expression, induce alternative splicing, reshuffle the genome causing inter-individual genetic variations and/or lead to epigenetic dysregulation at the insertion site [25–27]. Within this context, we recently showed that forced expression of $L1^{RP}$, an active *L1* isolated from exon 1 of the retinitis pigmentosa gene (RP) of a patient with X-linked retinitis pigmentosa [28,29], modulates genetic networks involved in the regulation of inflammation, adhesion and cellular metabolism in HepG2 cells [30]. The HePG2 cell line is frequently used because it retains high functional activity of liver-specific genes [31,32]. Both wildtype and RT-domain mutant (i.e. D702Y) of $L1^{RP}$ induce epithelial to mesenchymal transition (EMT) in HepG2 cells, confirming a biological role for $L1^{RP}$ that does not involve retrotransposition [30].

To further evaluate molecular mechanisms of *L1* mobilization and genetic reprogramming within the HepG2 genome, a follow-up study was conducted to characterize the retrotransposition profiles of $L1^{RP}$. Analysis of *L1* retrotransposition in HepG2 cells by fluorescence in-situ

* Corresponding author at: 1295 North Martin Avenue, P.O. Box 210202, Tucson, AZ 85721-0202.
E-mail address: ksramos@email.arizona.edu (K.S. Ramos).

http://dx.doi.org/10.1016/j.ygeno.2014.07.001
0888-7543/

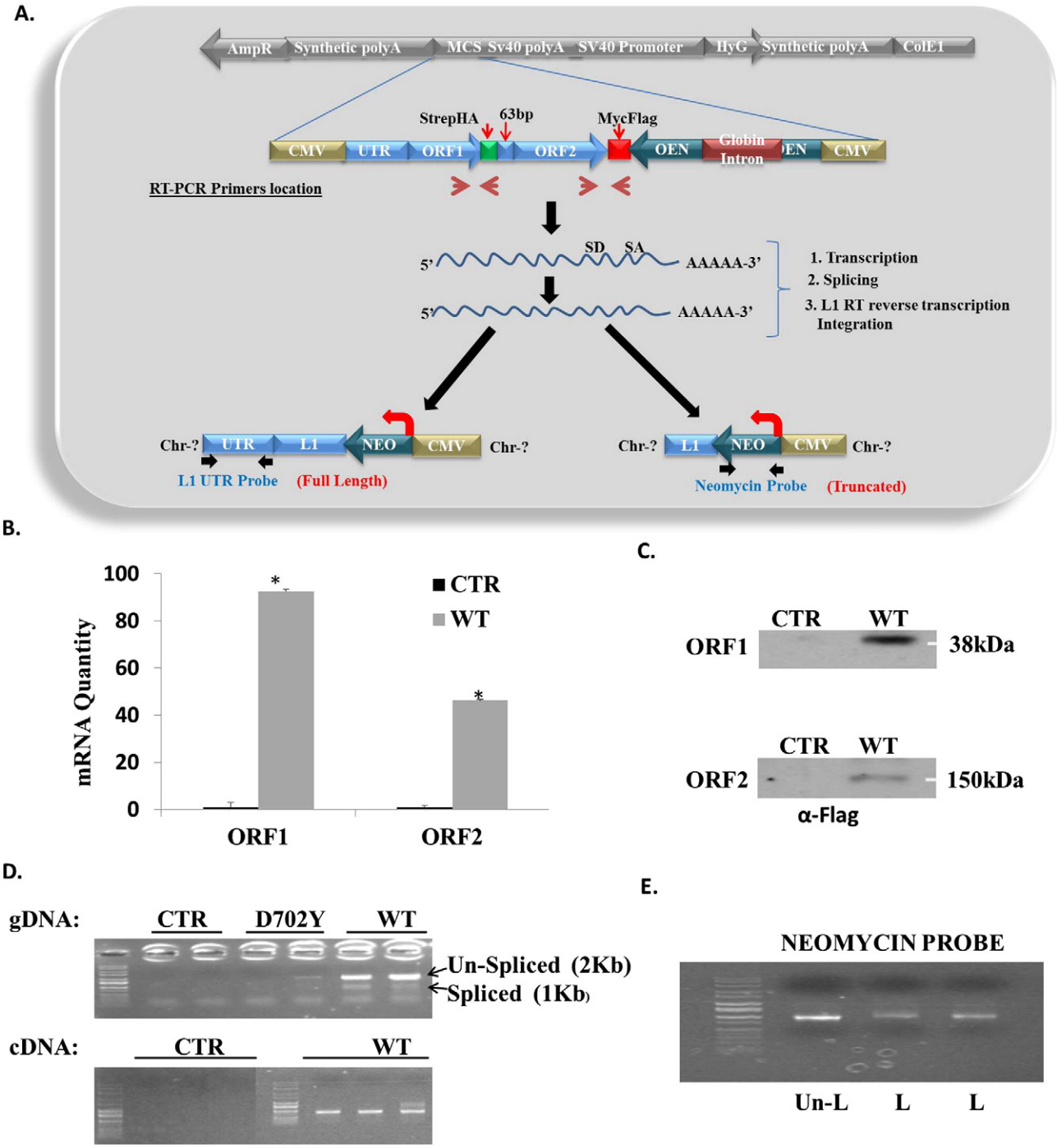


Fig. 1. Expression of ectopic *L1* proteins in HepG2 cells. A. Schematic diagram of *L1* vectors used to examine the expression and retrotransposition of $L1^{RP}$ within the HepG2 genome. $pB001^{CTR}$ (CTR) is the vector backbone while $pB015^{WT}$ (WT) consists of *L1* ORF1 tagged with Strep and HA (green), ORF2 tagged with Myc and Flag (red) and a neomycin cassette placed in opposite orientation. The schematic also shows the retrotransposition of ectopic *L1* leading to full-length or truncated insertions into the HepG2 genome. B. RT-PCR analysis of *L1* ORF1 and ORF2 using primers specific for ectopic *L1*. The location of the primer sets is indicated in Fig. 1A. C. Western blot of ORF1 and ORF2 proteins with antibodies directed against Strep and Flag, respectively, detected ~40 kDA and ~150 kDA bands which are absent in cells transfected with control plasmid. D. Detection of spliced neomycin gene from gDNA and cDNA of HepG2 cells indicates the retrotransposition of ectopic *L1*. The spliced neomycin gene is absent from the cDNA of cells expressing control plasmid. E. Degenerate oligonucleotide primed PCR (DOP-PCR) product of neomycin gene from HepG2 genomic DNA. The results confirmed the increased size of the biotin-dTTP labeled (L) probe compared to unlabeled probe (Un).

hybridization (FISH), an approach previously used to localize genomic consensus sequences of *L1* ORF2 [30,33], coupled with DNA sequencing, established a strong preference for insertion into gene-poor regions of chromosome 13. These findings establish for the first time that *L1* insertions are not entirely random events as originally proposed, and raise questions about the biology and molecular mechanism involved in the regulation of *L1* retrotransposition.

2. Materials and methods

2.1. Cloning, western blotting, RT-PCR and indirect immunofluorescence

Cloning of vectors $pB001^{CTR}$ (vector backbone or CTR), $pB016^{MUT}$ (aspartate (D) to tyrosine (Y) mutant at position 702 (D702Y)) and $pB015^{WT}$ (wildtype or WT) were done as described in Bojang et al.

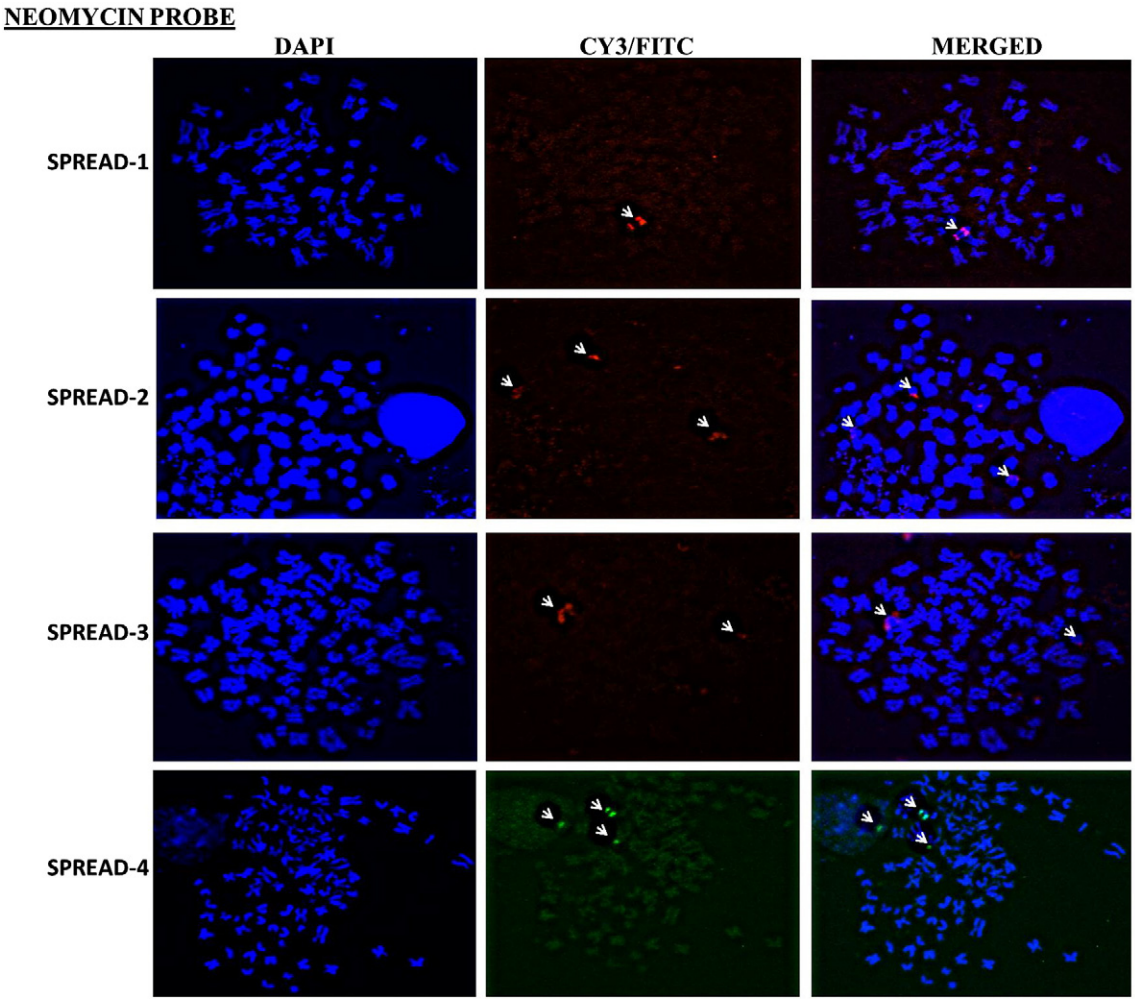


Fig. 2. Analyses of retrotransposition rates in nuclei from different clones of stably transfected HepG2 cells. FISH was completed to evaluate L1 retrotransposition rates in individual nuclei. Column 1 shows chromosome spreads stained with DAPI, column 2 shows the neomycin probe stained with FITC/CY3 and column 3 shows the merged signals. Differences in neomycin staining indicate that L1 retrotransposition rates are specific for individual nuclei.

2013 [30]. These vectors were used to generate stable transfected HepG2 cell lines used to monitor retrotransposition activity of $L1^{RP}$. Total protein was extracted using the m-PER reagent (Thermo Scientific, Rockford, IL) and *L1* ORF1-HASTREP and ORF2-FlAGMYC proteins were detected using antibodies against Strep (Strep (S10D4) sc-52234, Santa Cruz, CA) and Flag (Anti-Flag (F1804): Sigma, St. Louis, MO) tags respectively. RNA was extracted using the RNeasy Mini kit (Qiagen, Maryland, cat# 74104) according to manufacturers' instructions. *L1* ORF1 and ORF2 mRNAs were measured with primers specific for just the transfected *L1* ORF1 (*L1*-ORF1exo-1F: 5′-GAAGGAAGCGCTAAACATGG-3′ and *L1*-ORF1exo-1R: 5′-TGGGACGTCGTATGGGTATT-3′) and ORF2 (*L1*-ORF2exo-1F: *5′*-TGAAATTGGA AACCATCATTCTC-*3′* and *L1*-ORF2exo-1R: 5′-CCTTGTCATCGTCATCCTTGT-3′), and the ΔΔCT method was used to calculate relative levels of message in wildtype and control cells. Indirect immunofluorescence was done as described in Bojang et al., 2013 [30] with antibodies against HA tag (i.e. Anti-HA-tag (6E2) Mouse mAB-Alexa-594 conjugated antibody) of ORF1 and Flag tag (i.e. Anti-flag tag M2-Alexa Fluor-488 conjugated antibody) of ORF2 (Cell Signaling Technology Inc., Danvers, MA). Images were analyzed using the Axiovert Inverted microscope at 63× magnification.

2.2. Degenerate oligonucleotide primed polymerase chain reaction (DOP-PCR) and FISH

Two pairs of degenerate primers were used to amplify the *L1* 5′-UTR (907 bp) (UTR-1f: CAAGATGGCCGAATAGGAAC and UTR-1r: TACTTTTG GTCTTTGATGATGGTG) and the spliced 1 kb neomycin gene (Neo-1f: GGATAGCATTGGGAGATATACCT and Neo-1r: ATTGAACAAGATGGATTG CACGC). The PCR fragment of the spliced neomycin gene was labeled using degenerate oligonucleotide primed PCR (DOP-PCR) as described in Bojang et al., 2013 [30]. After initial PCR of the *L1*-5′UTR, 2 μg of the *L1*-5′UTR was chemically-labeled with CY3 using the MIRUS FISH labeling kit (Cat # MIR 6510) at 37 °C for 1 h. Both PCR products were visualized on a 1% agarose gel, purified and quantitated using the nanodrop. FISH analysis was done as described in Bojang et al., 2013 [30].

2.3. Characterization of L1 insertions by inverse PCR and DNA sequencing

Genomic DNA (gDNA) was isolated from HepG2 cells stably expressing wildtype *L1* and 500 ng was digested with *MluI*. gDNA was then phenol/chloroform extracted, ethanol precipitated, and ligated using T4 DNA ligase. DNA rings containing reverse transcribed and mobilized *L1* sequences were isolated using primers specific for the neomycin gene (Inverse-Neo-1f: AGTGACAACGTCGAGCACAG: Inverse-Neo-1r: ATCAGGACATAGCGTTGGCT). Amplicons were gel purified and cloned into pCR2.1 TOPO TA (Invitrogen). Vectors were digested with *EcoR1* to confirm insertions and each clone was sequenced using M13 forward and reversed primers. DNA sequences were blasted against the NCBI and UCSC Blat genome browser databases to identify *L1* insertion sites.

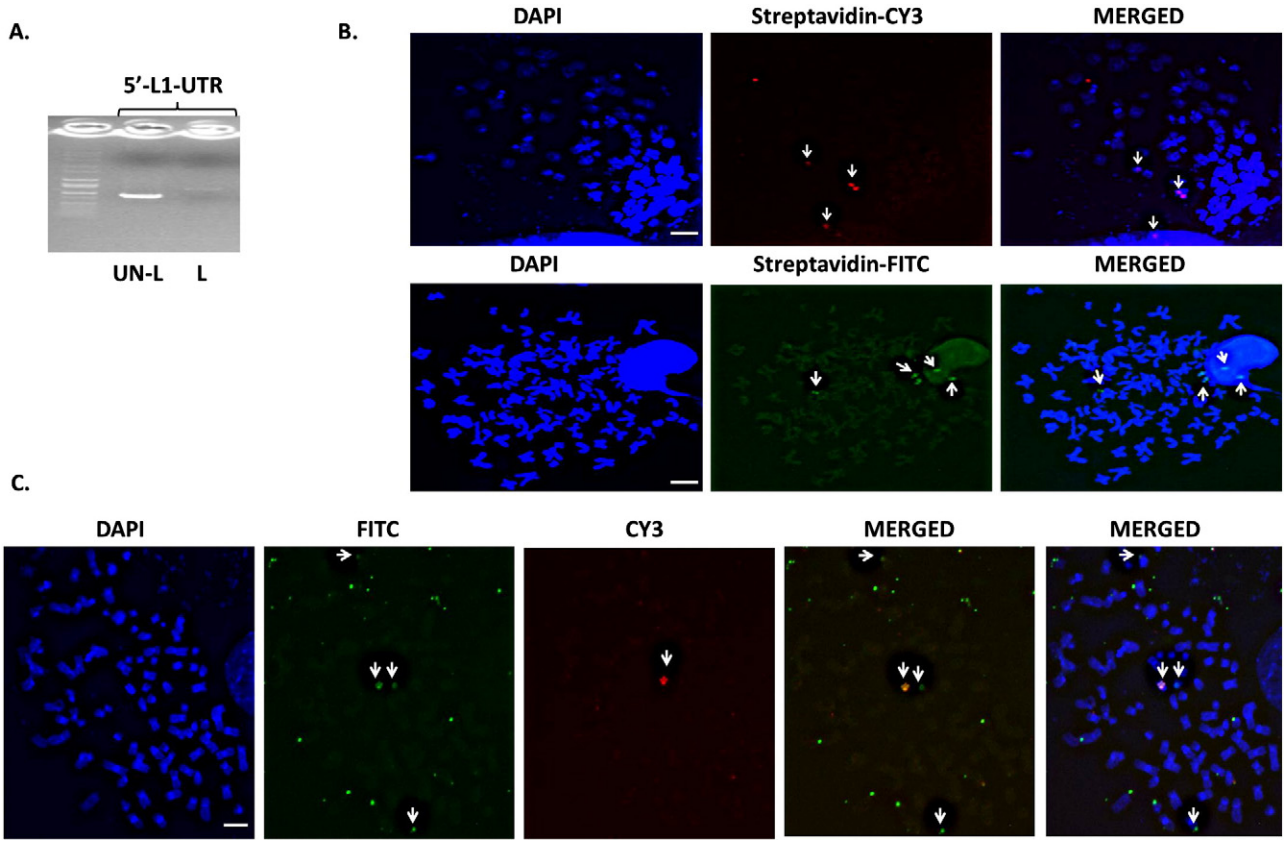


Fig. 3. Detection and quantification of full-length and truncated *L1* insertions. A. CY3 chemically-labeled $L1^{RP}$-5′UTR using the MIRUS FISH labeling kit. Results showed a faint band that is larger in size compared to the unlabeled probe. The increase indicates the incorporation of CY3, while faintness is an indication that UV emission at ~300 nm is not optimal for detection of fluorescent labeled $L1^{RP}$-5′UTR (top). DOP-PCR product of neomycin gene from HepG2 genomic DNA showing increased in size of the biotin-dTTP labeled (L) probe compared to unlabeled probe (Un). B. FISH analysis with $L1^{RP}$-5′UTR (top) and neomycin (bottom) probes. The data indicate that retrotransposition of ectopic *L1* can be readily assayed using FISH analysis. C. Dual Fish analysis of $L1^{RP}$ using the $L1^{RP}$-5′UTR and neomycin probes. Column 1 shows DAPI staining (blue) of chromosomes, column 2 shows FITC staining of the neomycin probe (green), column 3 shows Cy3 staining $L1^{RP}$-5′UTR probe (red), column 4 shows matched CY3 and FITC staining and column 5 shows the merged staining for all dyes. Colocalization of FITC and CY3 (red arrow) indicated full-length $L1^{RP}$ insertions, while single FITC staining indicates truncated insertions. Scale bar is 10 μm.

3. Results

3.1. Ectopic L1 undergoes complete cycles of retrotransposition in cultured HepG2 cells

We have previously shown that $L1^{RP}$ is expressed in HepG2 cells and remains retrotransposition competent after serial passage [30]. Here, we further characterize the expression profiles of *L1* ORF1 and ORF2 proteins in HepG2 cells and evaluated retrotransposition profiles after extended culture. Fig. 1A shows a schematic of the $L1^{RP}$ wildtype vector used in our studies, as described earlier in Bojang et al. 2013 [30,34]. RT-PCR (Fig. 1B) and Western experiments (Fig. 1C) confirmed that *L1* ORF1 and ORF2 proteins are readily detected in HepG2 cells transfected with $L1^{RP}$, but not control plasmid. Measurements of integration and expression of the final spliced neomycin gene product into genomic DNA isolated from control (CTR), D702Y mutant (D702Y) and wildtype (WT) HepG2 cells showed that the un-spliced and spliced forms of the neomycin gene are only detected in wildtype cells (Fig. 1D). In keeping with this observation, a 1 kb neomycin gene product is only amplified from the cDNA of wildtype clones (Fig. 1E). Fig. 2 shows that the rate of retrotransposition is highly variable in different nuclei isolated from different clones of stably transfected HepG2 cells. Together; these data indicate that complete cycles of retrotransposition in HepG2 cells exhibit variable rates of retrotransposition among different nuclei.

Table 1.
Frequency of L1 retrotransposition in HepG2 measured by FISH.

Probe	Criteria	Number of spreads	Spots counted	Average/spread
L1-5′-UTR	# of spots >1	15	40	2.67
Spliced neomycin	# of spots >1	11	63	5.25

Chromosome spreads were counted for full length (L1-5′-UTR probe) or truncated (spliced neomycin probe) insertions, with only spreads showing more than one staining event counted.

3.2. Ectopic full-length and truncated L1 insertions are detected in the genome of HepG2 cells

The non-processive nature of *L1* RT, the presence of cryptic polyadenylation sites in *L1*-ORF2 sequence, and the mode of translation of *L1*-ORF2 often lead to insertion of 5′UTR truncated *L1* sequences [20]. As such, we sought to track the integration of full-length versus truncated *L1* insertions by FISH. Two unique probes, $L1^{RP}$-5′UTR and spliced neomycin gene probes were designed to distinguish full-length and truncated *L1* sequences. Each probe was labeled with either biotin-dTTP or CY3. An increase in the apparent size of labeled probes indicated the incorporation of biotin-dTTP or CY3, respectively (Fig. 3A). These two probes were then used to track full-length and truncated $L1^{RP}$ insertions in cultured HepG2 cells (Fig. 3B), and to quantify the number of insertions. In our studies, only spreads with more than one insertion were counted, as this was judged to represent true, active retrotransposition events as opposed to stable integration. Arrows denote staining of the neomycin probe as an index of *L1* retrotransposition. A total of 15

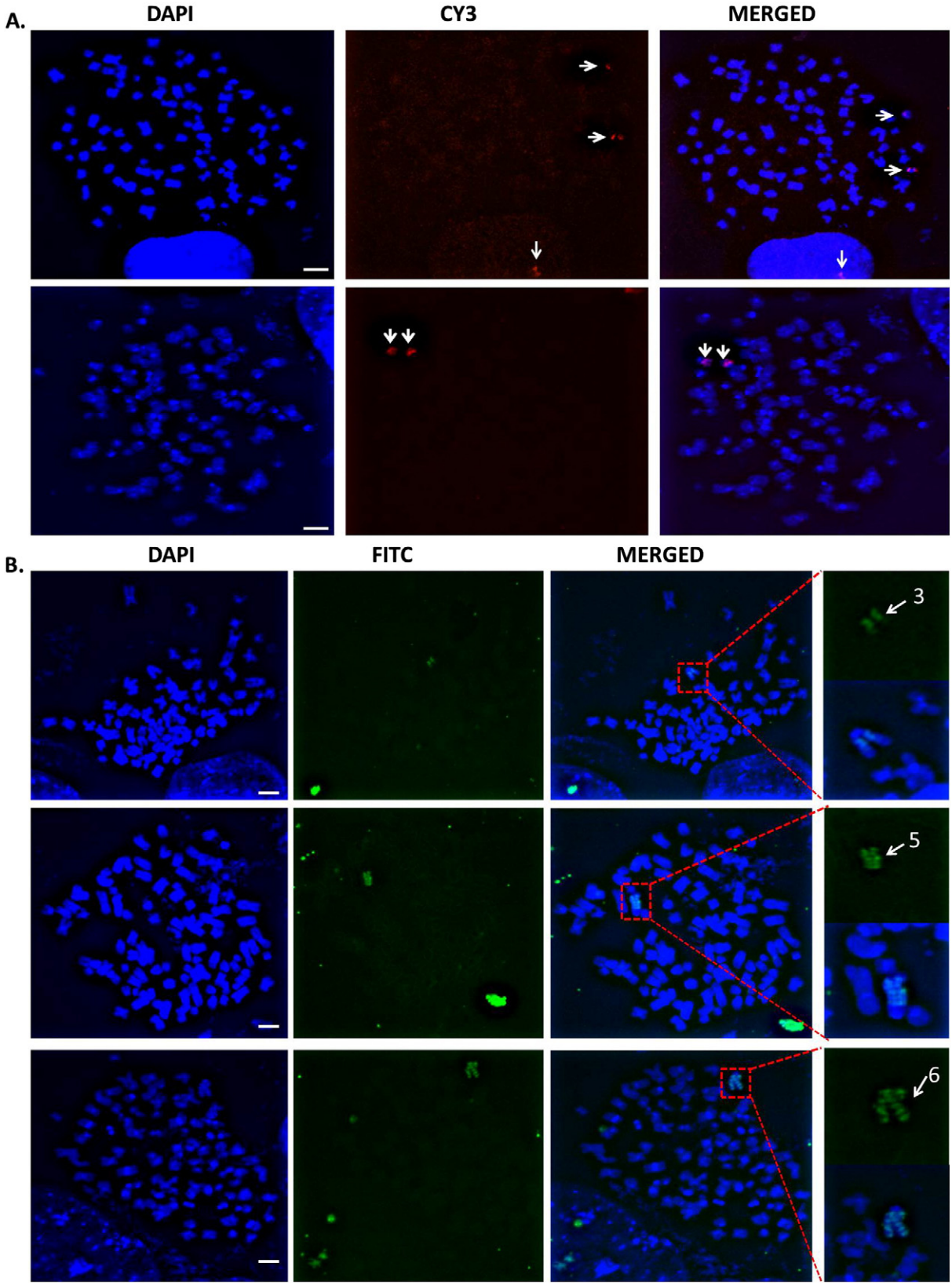


Fig. 4. Patterns of *L1* insertion after prolong culturing. Metaphase chromosome spreads isolated from cells expressing wildtype *L1* and probed biotin-labeled neomycin probes followed by streptavidin-CY3 or streptavidin-FITC secondary antibodies. A. Random insertion of ectopic *L1* into three (top) and two (bottom) different chromosomes. Notice that sometimes not all chromosomes are released from the nucleus. B. Repeated or preferential insertion of ectopic *L1* into the same chromosome. In the first panel (row-1), there are three *L1* insertions, in panel 2 (row-2) there are five *L1* insertions and in panel 3 (row-3) there are six *L1* insertions into the same chromosome. Scale bar is 10 μm.

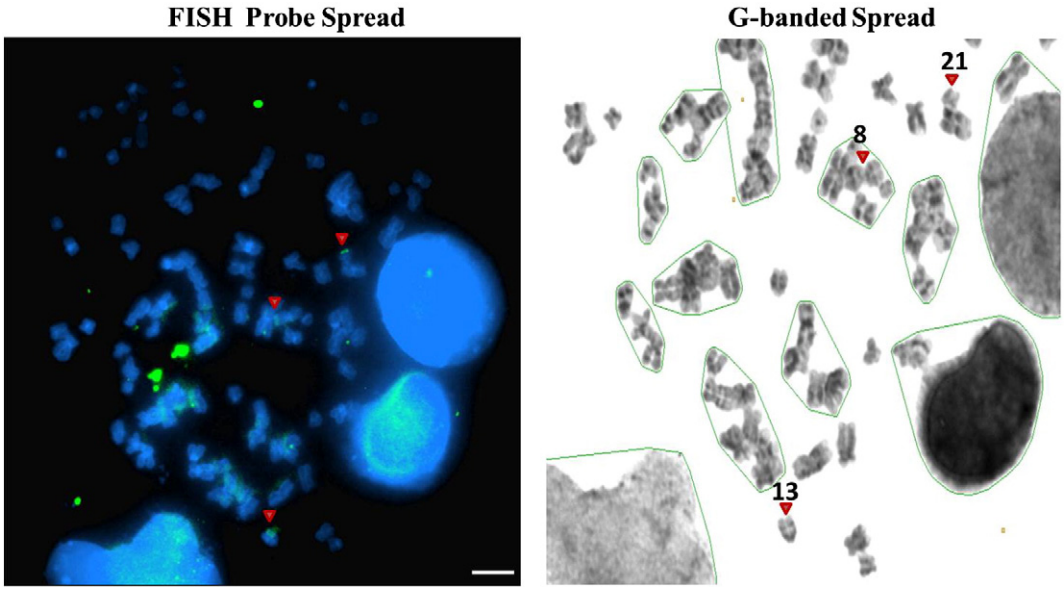


Fig. 5. Identification of the chromosome targeted repeatedly for *L1* insertion. FISH analysis (left) and the corresponding G-banded spread (right) of chromosomes from HepG2 cell stably expressing wildtype *L1*. The neomycin probe stained with FITC indicates the retrotransposition of *L1* and the G-banded spread indicates that *L1* retrotransposed into chromosomes 8, 21 and 13. Data indicate that chromosome 13 is repeatedly targeted by $L1^{RP}$ for insertion. Scale bar is 10 μm.

spreads was examined for the *L1*-5′UTR and 11 spreads for the neomycin gene, with 40 to 67 individual spots counted for each probe (Table 1). The results identified an average of 2.67 full-length insertions compared to 6.10 truncated insertions, confirming the assertion that the majority of *L1* insertions are truncated at the 5′-UTR (Table 1). To further distinguish truncated from full length insertions, we analyzed the expression of *L1* ORF1/2 by indirect immunofluorescence (Supplementary Fig. 1a). We reasoned that cells with truncated 5′*L1* should lose the expression of both proteins and as expected, some populations of wildtype cells lacked expression of both L1 ORF1 and ORF2 (Supplementary Fig. 1a, compare columns 1 and 2 to columns 3 and 4).

Genome analysis has previously demonstrated that the 5′UTR of different *L1*s is remarkably similar in sequence, such that the 5′-UTR probe used may have recognized endogenous *L1s* within the HepG2 genome. To test this hypothesis, combined FISH analysis using $L1^{RP}$-5′UTR and neomycin probes was used as an index of full-length insertions. Fig. 3C shows that CY3 and FITC (column 4) colocalize in HepG2 cells, indicating the presence of full-length ectopic $L1^{RP}$ insertion (red arrow), and the specificity of our probe for ectopic *L1*. It should be noted that the 5′UTR probe used did not stain multiple chromosomal locations in double FISH analysis, however, when the UTR probe was used by itself, several chromosomes were stained including chromosomes with multiple insertions (red arrow) (Supplementary Fig. 1b). The lack of staining might have been caused by the lack of accessibility of the probes to these regions due to compacted heterochromatin of these ancient *L1* sequences. This conclusion is supported by the fact that *L1* sequences are heavily methylated which in turn induces the formation of heterochromatin [7,35]. Fig. 3C also shows that not all neomycin staining colocalized with the 5′UTR probe staining (white arrow), supporting the conclusion that the number of truncated insertions greatly exceeds the number of full-length insertions. Together, these data indicate that the retrotransposition activity of ectopic *L1* can be readily assayed using FISH to differentiate between full-length and truncated *L1* insertions.

3.3. Preferential L1 insertions into gene poor regions of chromosome 13

Next, we sought to examine the randomness of *L1* insertions after extended culturing. HepG2 cells underwent >25 passages and were then processed for FISH using the Cy3-labeled neomycin probe. While the majority of spreads showed a random pattern of retrotransposition (Fig. 4A), multiple insertions were consistently observed into a single chromosome (Fig. 4). The metaphase spread presented in the first row of Fig. 4B displayed three *L1* insertions, which increased to five and then six, as a function of serial passage in culture. To authenticate preferential insertions, chromosome spreads were isolated from two independent clones and FISH analysis was repeated. Again, both of these clones showed preferential insertion into the same chromosome confirming the initial observation (Supplementary Fig. 2). These repeated insertions represent full length *L1* insertions since truncated insertions do not have the ability to retrotranspose.

Since $L1^{RP}$ may be preferentially targeted to this particular chromosome, G-banding experiments were conducted. G-banding analysis identified random insertions into chromosomes 8 and 21, while repeated insertions were identified into a chromosome that could either be 13 or Y based on the G-banding profile (Fig. 5). To authenticate the results, and to more definitively identify the chromosome targeted for preferential insertion, inverse PCR of gDNA was completed. Genomic DNA from HepG2 cells was *MluI* digested and T4 ligated followed by isolation of *L1* rings using primers specific for the neomycin cassette (Fig. 6A). Six unique amplicons (Clones 1–6) of sizes ~2.2, 1.0, 4.0, 0.7, 0.9 and 0.7 kb, respectively, were isolated, gel purified, and cloned into the pCR2.1TA cloning vector (Fig. 6B). Each clone was sequenced using M13 forward and reverse primers specific for the pCR2.1TA vector. The flanking genomic sequences of *L1* insertions were identified using BLAT (http:genome.ucsc.edu) and BLASTN (http://blast.ncbi.nlm.nih.gov/blastn) sequence alignment search engines against the Human Genome Sequence. Four of the clones (Clones 2, 4, 5, 6) matched to chromosome 13 at different sites (Fig. 6C), with Clones 4 and 6 being identical in both size and flanking sequence (Figs. 6B & C). Clone 3 inserted into chromosome 8, while the insertion site of Clone 1 could not be definitively identified given that flanking sequence was not obtained. Overall, these results identify with confidence chromosome 13 as an autosome with preferential targeting or duplication of *L1* insertions.

Lastly, insertion profiles into chromosome 13 were examined given that this autosome has the lowest gene load, CpG island density, and exon coverage of all human autosomes (Dunham et al., 2004). The criteria employed by Dunham and coworkers were applied, where a region is classified as gene rich if it contains five or more genes per megabase (Mb), or gene poor if the density is less than five. Using the

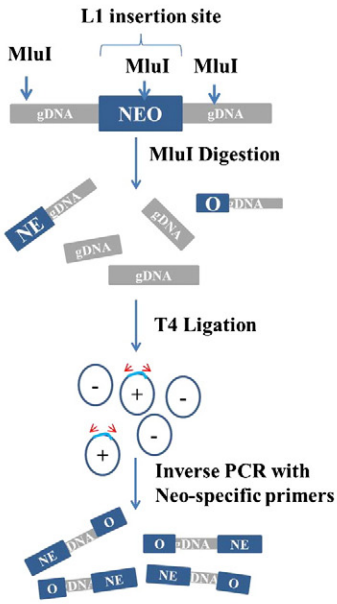
flanking sequence for each clone, Blat analysis showed that all insertions occurred within intronic regions, and more specifically, into gene poor regions denoted as gray shaded segments in Fig. 6D. It should be noted that this analysis was limited to genes validated using the updated GRC38/Hg38 build (http:genome.ucsc.edu).

4. Discussion

Evidence is presented here that the pB015^{WT} plasmid can be reliably used to monitor retrotransposition at the single chromosomal and cellular levels. In the past, detection of *L1* encoded proteins has proven challenging given the stringent restrictions posed by the structural organization of genetic elements and the widespread silencing of *L1* expression. *L1*^{RP} consists of two ORFs in-frame which are separated by an inter-ORF region containing two in-frame stop codons after the stop codon of ORF1. Both ORFs are transcribed from a common promoter, but the modes of translation for the two proteins are different. While *L1* ORF1 is translated using the 40S ribosomal scanning model, *L1* ORF2 relies on the translation/termination model giving rise to lower protein levels [36–38]. Our own findings lend support to these views, with higher levels of ORF1 mRNA and protein than ORF2 mRNA and protein detected in HepG2 cells in all instances examined (Fig. 1).

L1 sequences can be inserted into the genome as full length or 5′-truncated insertions, with only full-length insertions retaining the capacity to remobilize to new locations. Both insertion types can create epigenetic hot spots, alternate splice sites or alternate promoters, which in turn function to fine-tune the expression of nearby genes [39]. Thus, understanding the frequency of full-length and truncated insertions at

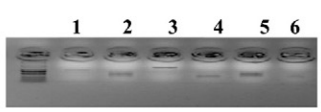
A.



B.

Clone #	Flanking Sequence	Chromosome: Insertion site
1	No flanking region seen(only segment of original vector could be sequenced)	?
2	ATTTTAAGTGTCCAATTCAGAAAGTT	Chromosome 13: 50085006-50085031
3	CCTGGGGTGCCTAATGAGTG	Chromosome 8 : 3382485-3382504
4	ATCATGGTGGAAGGTGAAGGAGGAGCAAAGTCAAGTCTTACATGGCAG CAGGCAAGAGAGtGCGTGCAGGGGAACTTCCCTTTAcAAAACCATCAGA TCTCATGAGACTTATTCACTATCATGAGAACAGTATGGGAAAAACCCAC CCCCATGATTCAGTTACCTCCCAACAGGTCCCTCCCATGACATGTGGGA ATTATGGGAGCTACAATTCAAGGAGATTTGGGTGGGGACAGAGCCAAA CCATATC	Chromosome 13: 95503475-95503724
5	CTATCAGATGATAGCGTTGGCTACCC	Chromosome 13: 43388471-43388496
6	ATCATGGTGGAAGGTGAAGGAGGAGCAAAGTCAAGTCTTACATGGCAG CAGGCAAGAGAGtGCGTGCAGGGGAACTTCCCTTTAcAAAACCATCAGA TCTCATGAGACTTATTCACTATCATGAGAACAGTATGGGAAAAACCCAC CCCCATGATTCAGTTACCTCCCAACAGGTCCCTCCCATGACATGTGGGA ATTATGGGAGCTACAATTCAAGGAGATTTGGGTGGGGACAGAGCCAAA CCATATC	Chromosome 13: 95503475-95503724

C.



D.

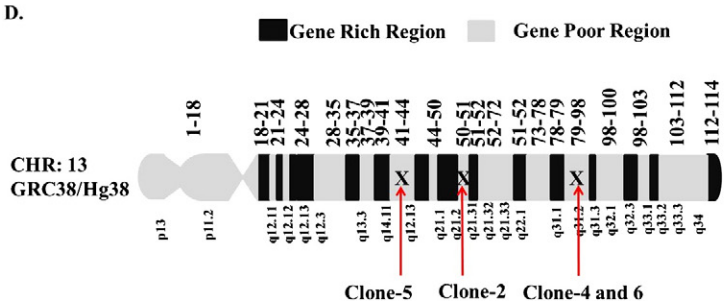


Fig. 6. Authentication of preferential insertions using inverse PCR followed by DNA sequencing to determine chromosome identity and insertion sites of *L1* retrotransposition. A. Schematic of the inverse PCR procedure. Briefly gDNA was digested with *MluL*, then ligated and PCR amplified using neomycin specific primers. B. Agarose gel electrophoresis of independent inverse PCR products. C. Sequencing data after TA-cloning of each gel-purified amplicon. The table indicates clone number, approximate size, flanking sequence adjacent to the neomycin insert, and chromosome location. The results indicate that Clones 2,4,5,6 inserted into chromosome 13, while Clone 3 inserted into chromosome 8. The insertion site of Clone 1 could not be determined as no flanking sequence was resolved for this clone. D. A schematic depiction of *L1* insertion sites into chromosome 13. Dark shaded regions denote gene-rich areas, while gray shaded areas denote gene-poor regions. In all cases, *L1* insertions occur within gene poor regions of the chromosome.

the chromosomal and single cell levels is paramount for advancing our understanding of the genomic basis of human disease. The conventional methodology used for monitoring retrotransposition activity in cultured cells is to count the number of fluorescent signals, or G418-resistant foci, and to factor this as a function of the total number of cells (i.e. number of hygromycin resistant cells). This methodology likely underestimates actual retrotransposition rates because a single cell or chromosome may contain more than one *L1* insertion (see Figs. 1–6). In contrast, FISH methodology allows efficient tracking of *L1* retrotransposition events, and to capture the pattern and number of *L1* insertions at the single chromosomal or cell levels. Of particular note is that identification of cells with more than one insertion, as evidenced by the occurrence of multiple loci with selectable markers, facilitates the differentiation of full length and truncated insertions. As such, the approach can be used reliably to ask questions concerning retrotransposition events in a single G418-positive clone, or chromosome, and also to assess the randomness of this process. The approach can also be readily combined with restriction digestion, inverse PCR and/or sequence alignment to accurately determine both the pattern and mode of *L1* insertions as shown here (Fig. 6).

Although unknown for *L1*, many transposable elements (TEs) have developed highly specific targeting mechanisms that direct their integration to genome safe regions [40–42]. For instance, characterization of the Tf1 fission yeast retrotransposon has shown that 95% of the integrations are clustered upstream of ORFs, with most of the promoters targeted by Tf1s representing genes associated with stress [40,41]. Likewise, Ty5 transposable elements in yeast have been shown to change their target sites in response to stress, with integration into ORFs as opposed to heterochromatin in cells deprived of nitrogen [43]. Interestingly, Ty5 integration into heterochromatin is dependent on phosphorylation of Ser1095, with mutation of Ser1095 redirecting integration to expressed regions of the genome [43,44]. Experiments in maize have shown that integration of DNA transposons lead to variegated corn color phenotypes, while integration of Hatvine1-rrm DNA transposon into the promoter region of VvTFL1A gene influences the branching pattern and the fruit size of grapevines [43,45,46]. A question that remains unanswered is whether similar mechanisms exist in higher organisms to direct the integration of mobile elements into the host genome.

Our analysis of $L1^{RP}$ retrotransposition in HepG2 cells revealed random insertions into all chromosomes, except chromosome 13 where preferential insertions were documented by FISH and DNA sequencing (Figs. 4 & 6). Previous studies have revealed that older *L1s* specifically integrate into gene poor and AT rich regions of the genome, while new *L1* integrations are interspersed, occurring near or within intronic regions at a loosely defined sequence of 5-TTTT/A-3 [27,47]. These findings establish an evolving, but adaptive mechanism for *L1* insertion into the host genome that might not be random, but rather contextual in a manner that affords selective advantage to the host. For example, others have noted the enrichment for recent *L1* insertions into the human Y chromosome, including the unusually high number of full-length *L1*s [48]. Our own G-banding studies would support this notion since we could not readily distinguish between chromosomes 13 and Y, and one of the clones obtained by inverse PCR could be matched in sequence with confidence to either chromosome. The occurrence of preferential insertions into the genome may be linked to a low gene load, with chromosome 13 having the lowest gene density (6.5 genes per Mb) of all human autosomes and containing a central region of 38 Mb where the gene density drops to only 3.1 genes per Mb [49]. We regard these regions as permissive "safe heaven" regions for insertion. Of note is that all of the insertions identified by DNA sequencing targeted intronic regions or repeat regions of high homology among all chromosomes, including the sex chromosomes (Fig. 6). Thus, another likely target for preferential insertion may be the Y chromosome, where faulty selection and a low gene load have also been documented [50,51]. Graves et al. established the inability of the Y chromosome to sort through its genes, and suggested that this may account for its propensity to accumulate junk DNA [50]. A low gene load would make preferential *L1* insertions less harmful to the organism such that insertional mutagenesis would be of modest negative impact on overall survival.

Fish analysis is a routine procedure employed in the clinical genetics laboratory. As such, the approaches described here can be readily adapted in the clinical setting to address questions related to the role of *L1* in human pathogenesis. Such studies are important given increasing recognition that genetic variation between individuals is largely attributed to the polymorphic expression of transposable elements [49, 52]. Further, the activities of TEs can be strongly regulated by environmental cues that define and dictate differences in disease susceptibility [12,13,27]. To date, up to 100 human diseases have been linked to the activity of TEs [53–55]. Given that most repetitive regions of the genome cannot be easily sequenced using current methodologies, FISH analysis of transposable elements can be used in the clinical setting to evaluate polymorphic variations between individuals. These findings shed new light into *L1* targeting within the genome, raise important questions about the cellular mechanisms responsible for *L1* retrotransposition and strongly suggest that *L1* retrotransposition is not entirely random.

Conflict of interest

The authors declare no potential competing interests.

Acknowledgments

This work was supported in part by grants from the National Institute of Environmental Health Sciences (ES014443 and ES017274) and AstraZeneca to KSR. We thank Dr. Alexander Asamoah for his referral and Ms. Margaret Barch for her excellent technical assistance and helpful discussions regarding G-banding measurements. We also gratefully acknowledge the assistance provided by Dr. Bhagavatula Moorthy for his assistance with gDNA sequencing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2014.07.001.

References

[1] A.F. Scott, B.J. Schmeckpeper, M. Abdelrazik, C.T. Comey, B. O'Hara, J.P. Rossiter, T. Cooley, P. Heath, K.D. Smith, L. Margolet, Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence, Genomics 1 (1987) 113–125.
[2] B.A. Dombroski, S.L. Mathias, E. Nanthakumar, A.F. Scott, H.H. Kazazian Jr., Isolation of an active human transposable element, Science 254 (1991) 1805–1808.
[3] G.D. Swergold, Identification, characterization, and cell specificity of a human LINE-1 promoter, Mol. Cell. Biol. 10 (1990) 6718–6729.
[4] K.G. Becker, G.D. Swergold, K. Ozato, R.E. Thayer, Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element, Hum. Mol. Genet. 2 (1993) 1697–1702.
[5] T. Tchenio, J.F. Casella, T. Heidmann, Members of the SRY family regulate the human LINE retrotransposons, Nucleic Acids Res. 28 (2000) 411–415.
[6] N. Yang, L. Zhang, Y. Zhang, H.H. Kazazian Jr., An important role for RUNX3 in human L1 transcription and retrotransposition, Nucleic Acids Res. 31 (2003) 4929–4940.
[7] D.E. Montoya-Durango, Y. Liu, I. Teneng, T. Kalbfleisch, M.E. Lacy, M.C. Steffen, K.S. Ramos, Epigenetic control of mammalian LINE-1 retrotransposon by retinoblastoma proteins, Mutat. Res. 665 (2009) 20–28.
[8] S.E. Holmes, M.F. Singer, G.D. Swergold, Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element, J. Biol. Chem. 267 (1997) 19765–19768.
[9] S.L. Mathias, A.F. Scott, H.H. Kazazian Jr., J.D. Boeke, A. Gabriel, Reverse transcriptase encoded by a human transposable element, Science 254 (1991) 1808–1810.
[10] Q. Feng, J.V. Moran, H.H. Kazazian Jr., J.D. Boeke, Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition, Cell 87 (1996) 905–916.
[11] V.O. Kolosha, S.L. Martin, In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 10155–10160.
[12] J.S. Myers, B.J. Vincent, H. Udall, W.S. Watkins, T.A. Morrish, G.E. Kilroy, G.D. Swergold, J. Henke, L. Henke, J.V. Moran, L.B. Jorde, M.A. Batzer, A comprehensive

analysis of recently integrated human Ta L1 elements, Am. J. Hum. Genet. 71 (2002) 312–326.
[13] D.J. Hedges, P.L. Deininger, Inviting instability. Transposable elements, double-strand breaks, and the maintenance of genome integrity, Mutat. Res. 616 (2006) 46–59.
[14] S.L. Martin, Characterization of a LINE-1 cDNA that originated from RNA present in ribonucleoprotein particles: implications for the structure of an active mouse LINE-1, Gene 153 (1995) 261–266.
[15] W. Wei, N. Gilbert, S.L. Ooi, J.F. Lawler, E.M. Ostertag, H.H. Kazazian, J.D. Boeke, J.V. Moran, Human L1 retrotransposition: cis preference versus trans complementation, Mol. Cell. Biol. 21 (2001) 1429–1439.
[16] J. Jurka, Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 1872–1877.
[17] G.J. Cost, J.D. Boeke, Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure, Biochemistry 37 (1998) 18081–18093.
[18] J.S. Han, J.D. Boeke, A highly active synthetic mammalian retrotransposon, Nature 429 (2004) 314–318.
[19] J.V. Moran, S.E. Holmes, T.P. Naas, R.J. DeBerardinis, J.D. Boeke, H.H. Kazazian Jr., High frequency retrotransposition in cultured mammalian cells, Cell 87 (1996) 917–927.
[20] B. Brouha, J. Schustak, R.M. Badge, S. Lutz-Prigge, A.H. Farley, J.V. Moran, H.H. Kazazian Jr., Hot L1s account for the bulk of retrotransposition in the human population, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 5280–5285.
[21] C. Esnault, J. Maestre, T. Heidmann, Human LINE retrotransposons generate processed pseudogenes, Nat. Genet. 24 (2000) 363–367.
[22] A. Buzdin, S. Ustyugova, E. Gogvadze, T. Vinogradova, Y. Lebedev, E. Sverdlov, A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3′ terminus of l1, Genomics 80 (2002) 402–406.
[23] M. Dewannieux, C. Esnault, T. Heidmann, LINE-mediated retrotransposition of marked Alu sequences, Nat. Genet. 35 (2003) 41–48.
[24] D.C. Hancks, H.H. Kazazian Jr., SVA retrotransposons: evolution and genetic instability, Semin. Cancer Biol. 20 (2010) 234–245.
[25] N. Narita, H. Nishio, Y. Kitoh, Y. Ishikawa, R. Minami, H. Nakamura, M. Matsuo, Insertion of a 5′ truncated L1 element into the 3′ end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy, J. Clin. Invest. 91 (1993) 1862–1867.
[26] T. Takahara, T. Ohsumi, J. Kuromitsu, K. Shibata, N. Sasaki, Y. Okazaki, H. Shibata, S. Sato, A. Yoshiki, M. Kusakabe, M. Muramatsu, M. Ueki, K. Okuda, Y. Hayashizaki, Dysfunction of the Orleans reeler gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon, Hum. Mol. Genet. 5 (1996) 989–993.
[27] J.S. Han, S.T. Szak, J.D. Boeke, Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes, Nature 429 (2004) 268–274.
[28] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.
[29] U. Schwahn, S. Lenzner, J. Dong, S. Feil, B. Hinzmann, G. van Duijnhoven, R. Kirschner, M. Hemberger, A.A. Bergen, T. Rosenberg, A.J. Pinckers, R. Fundele, A. Rosenthal, F.P. Cremers, H.H. Ropers, W. Berger, Positional cloning of the gene for X-linked retinitis pigmentosa 2, Nat. Genet. 19 (1998) 327–332.
[30] P. Bojang Jr., R.A. Roberts, M. Anderton, K.S. Ramos, Reprogramming of HepG2 genome by long interspersed nuclear element-1, Mol. Oncol. 4 (2013) 812–825.
[31] I. Dunham, A. Kundaje, S.F. Aldred, P.J. Collins, C.A. Davis, F. Doyle, C.B. Epstein, S. Frietze, J. Harrow, R. Kaul, J. Khatun, B.R. Lajoie, S.G. Landt, B.K. Lee, F. Pauli, K.R. Rosenbloom, P. Sabo, A. Safi, A. Sanyal, N. Shoresh, J.M. Simon, L. Song, N.D. Trinklein, R.C. Altshuler, E. Birney, J.B. Brown, C. Cheng, S. Djebali, X. Dong, J. Ernst, et al., An integrated encyclopedia of DNA elements in the human genome, Nature 489 (2012) 57–74.
[32] D.G. Jennen, C. Magkoufopoulou, H.B. Ketelslegers, M.H. van Herwijnen, J.C. Kleinjans, J.H. van Delft, Comparison of HepG2 and HepaRG by whole-genome gene expression analysis for the purpose of chemical hazard identification, Toxicol. Sci. 115 (2010) 66–79.
[33] P.D. Waters, G. Dobigny, P.J. Waddell, T.J. Robinson, LINE-1 elements: analysis by fluorescence in-situ hybridization and nucleotide sequences, Methods Mol. Biol. 422 (2008) 227–237.
[34] J.L. Goodier, L. Zhang, M.R. Vetter, H.H. Kazazian Jr., LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex, Mol. Cell. Biol. 27 (2007) 6469–6483.
[35] I. Teneng, D.E. Montoya-Durango, J.L. Quertermous, M.E. Lacy, K.S. Ramos, Reactivation of L1 retrotransposon by benzo(a)pyrene involves complex genetic and epigenetic regulation, Epigenetics 3 (2011) 355–367.
[36] H. Ilves, O. Kahre, M. Speek, Translation of the rat LINE bicistronic RNAs in vitro involves ribosomal reinitiation instead of frameshifting, Mol. Cell. Biol. 12 (1992) 4242–4248.
[37] J.P. McMillan, M.F. Singer, Translation of the human LINE-1 element, L1Hs, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 11533–11537.
[38] R.S. Alisch, J.L. Garcia-Perez, A.R. Muotri, F.H. Gage, J.V. Moran, Unconventional translation of mammalian LINE-1 retrotransposons, Genes Dev. 20 (2006) 210–224.
[39] D.E. Montoya, K.S. Ramos, L1 retrotransposon and retinoblastoma: molecular linkages between epigenetics and cancer, Curr. Mol. Med. 10 (2010) 511–521.
[40] Y.E. Leem, T.L. Ripmaster, F.D. Kelly, H. Ebina, M.E. Heincelman, K. Zhang, S.I. Grewal, C.S. Hoffman, H.L. Levin, Retrotransposon Tf1 is targeted to Pol II promoters by transcription activators, Mol. Cell 30 (2008) 98–107.
[41] S.E. Devine, J.D. Boeke, Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III, Genes Dev. 10 (1996) 620–633.
[42] S. Sandmeyer, Integration by design, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 5586–5588.
[43] Y. Guo, H.L. Levin, High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in *Schizosaccharomyces pombe*, Genome Res. 20 (2010) 239–248.
[44] J. Dai, W. Xie, T.L. Brady, J. Gao, D.F. Voytas, Phosphorylation regulates integration of the yeast Ty5 retrotransposon into heterochromatin, Mol. Cell 27 (2007) 289–299.
[45] B. McClintock, The origin and behavior of mutable loci in maize, Proc. Natl. Acad. Sci. U. S. A. 36 (1950) 344–355.
[46] L. Fernandez, L. Torregrosa, V. Segura, A. Bouquet, J.M. Martinez-Zapater, Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine, Plant J. 61 (2010) 545–557.
[47] C.R. Beck, J.L. Garcia-Perez, R.M. Badge, J.V. Moran, LINE-1 elements in structural variation and disease, Annu. Rev. Genomics Hum. Genet. 12 (2011) 187–215.
[48] S. Boissinot, A. Entezam, A.V. Furano, Selection against deleterious LINE-1-containing loci in the human lineage, Mol. Biol. Evol. 18 (2001) 926–935.
[49] F.M. Sheen, S.T. Sherry, G.M. Risch, M. Robichaux, I. Nasidze, M. Stoneking, M.A. Batzer, G.D. Swergold, Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition, Genome Res. 10 (2000) 1496–1508.
[50] S. Rozen, H. Skaletsky, J.D. Marszalek, P.J. Minx, H.S. Cordum, R.H. Waterston, R.K. Wilson, D.C. Page, Abundant gene conversion between arms of palindromes in human and ape Y chromosomes, Nature 423 (2003) 873–876.
[51] J.A. Graves, E. Koina, N. Sankovic, How the gene content of human sex chromosomes evolved, Curr. Opin. Genet. Dev. 16 (2006) 219–224.
[52] A. Dunham, L.H. Matthews, J. Burton, J.L. Ashurst, K.L. Howe, K.J. Ashcroft, D.M. Beare, D.C. Burford, S.E. Hunt, S. Griffiths-Jones, M.C. Jones, S.J. Keenan, K. Oliver, C. E. Scott, R. Ainscough, J.P. Almeida, K.D. Ambrose, D.T. Andrews, R.I. Ashwell, A.K. Babbage, C.L. Bagguley, J. Bailey, R. Bannerjee, K.F. Barlow, K. Bates, H. Beasley, C.P. Bird, S. Bray-Allen, A.J. Brown, J.Y. Brown, et al., The DNA sequence and analysis of human chromosome 13, Nature 428 (2004) 522–528.
[53] M.C. Seleme, M.R. Vetter, R. Cordaux, L. Bastone, M.A. Batzer, H.H. Kazazian Jr., Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 6611–6616.
[54] V.P. Belancio, D.J. Hedges, P. Deininger, Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health, Genome Res. 18 (2008) 343–358.
[55] J.M. Chen, P.D. Stenson, D.N. Cooper, C. Ferec, A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease, Hum. Genet. 117 (2005) 411–427.