Retrotransposition in tumors and brains

John L Goodier

Abstract
LINE-1s (L1s), the only currently active autonomous mobile DNA in humans, occupy at least 17% of human DNA. Throughout evolution, the L1 has also been responsible for genomic insertion of thousands of processed pseudogenes and over one million nonautonomous retrotransposons called SINEs (mainly Alus and SVAs). The 6-kb human L1 has a 5′- untranslated region (UTR) that functions as an internal promoter, two open reading frames—ORF1, which encodes an RNA-binding protein, and ORF2, which expresses endonuclease and reverse transcriptase activities—and a 3′-UTR which ends in a poly(A) signal and tail. Most L1s are molecular fossils: truncated, rearranged or mutated. However, 80 to 100 remain potentially active in any human individual, and to date 101 de novo disease-causing germline retrotransposition insertions have been characterized. It is now clear that significant levels of retrotransposition occur not only in the human germline but also in some somatic cell types. Recent publications and new investigations under way suggest that this may especially be the case for cancers and neuronal cells. This commentary offers a few points to consider to aid in avoiding misinterpretation of data as these studies move forward.

Keywords: Alu, Cancer, Mosaicism, Mutation, Neuron, Retrotransposon, Somatic, SVA, Tumor

Retrotransposition of non-long terminal repeat (non-LTR) long interspersed nuclear elements (LINE-1s, or L1s), as well as the mobilization in trans of non-autonomous short interspersed nuclear elements (SINEs) and processed pseudogenes, has built at least 50% of the human genome and remains an ongoing source of gene mutation [1,2]. As a type of “selfish DNA”, L1 activity was formerly thought to occur predominantly in germ cells, where insertions would pass to the next generation. However, in addition to the massive germline expansion of L1s that occurred during mammalian evolution, recent investigations have documented ongoing retrotransposition in selected somatic cell types, including neural progenitor cells, stem cells, early embryos, tumors and induced pluripotent stem cells [3-7]. More than 20 years ago, Miki et al. [8] reported the first instance of somatic retrotransposition, an L1 insertion into the adenomatous polyposis coli tumor suppressor gene of a colorectal cancer. The advent of high-throughput sequencing has made it possible to identify numerous non-germline de novo insertions in various kinds of cancer, as recently described in several high-profile papers [6,9-13] (Table 1).

Cancer-associated hypomethylation and elevated transcription of L1s predicted increased retrotransposition in tumors long before new insertions were detected by next-generation sequencing [16]. It is also reasonable to assume that insertions proliferate preferentially in tumors because cancer cells divide more rapidly than their normal cells of origin. Indeed, cell cycling, though not strictly required for retrotransposition, may increase its frequency [17-19]. Interestingly, investigations to date have detected new insertion events almost exclusively in tumors of epithelial cell types, some of which proliferate and turn over quickly. Carreira et al. [20] speculated that increased retrotransposon insertions in epithelial tumors may relate to a greater “plasticity” of epithelial cells, which are more easily reprogrammed to yield cancer or pluripotent stem cells.

Recent high-throughput sequence studies have reported tumor-specific insertions that vary greatly in number between different tumors of the same type, ranging from 0 in most instances to 106 in a single colorectal tumor identified by Lee et al. [9]. The application of different methodologies clearly accounts for some of this variation. Furthermore, current sequence analysis pipelines lack sensitivity to detect rare insertion events that occur late in tumor development, thus underestimating...
| Study                | Cancer type      | Tumor-normal pairs, n | Tumors with somatic insertions, n | Tumor-only somatic insertions, n | Normal-only somatic insertions | Method                  |
|---------------------|------------------|-----------------------|-----------------------------------|---------------------------------|-------------------------------|--------------------------|
|                     |                  |                       |                                   |                                 |                               |                          |
| Iskow et al. [6]    | Lung             | 20                    | 6                                 | L1: 9                           | 8/9                           | Pyrosequencing            |
|                     | Brain            | 10                    | 0                                 |                                 |                               |                          |
| Lee et al. [9]      | Glioblastoma     | 16                    | 0                                 | L1: 183                         | 38/39                         | Paired-end WGS           |
|                     | Ovarian          | 9                     | 5                                 | Alu: 10                         | 1/3                           |                          |
|                     | Colorectal       | 5                     | 5                                 | ERV1: 1                         | 1/1                           |                          |
|                     | Prostate         | 7                     | 6                                 |                                 |                               |                          |
|                     | Multiple myeloma | 7                     | 1                                 |                                 |                               |                          |
|                     | Normal (Trio)    | 3                     | 0                                 |                                 |                               |                          |
| Solyom et al. [10]  | Colorectal       | 16                    | 13                                | L1: 107                         | 69/107                        | L1-Seq                   |
| Shukla et al. [11]  | Hepatocarcinoma  | 19                    | 5                                 | L1: 17                          | 12/17                         | RC-Seq                   |
|                     |                  |                       |                                   |                                 |                               |                          |
| Ewing et al. [12]b  | Acute myeloid leukemia | 24               | 0                                 | 0                               |                               | Paired-end WGS           |
|                     | Breast           | 12                    | 0                                 | 0                               |                               |                          |
|                     | Colorectal adenocarcinoma | 5                 | 0                                 | 0                               |                               |                          |
|                     | Glioblastoma     | 15                    | 0                                 | 0                               |                               |                          |
|                     | Lung             | 19                    | 2                                 | GRIP: 3                         | 0/0                           |                          |
|                     | Ovarian          | 10                    | 0                                 | 0                               |                               |                          |

*ERV1, Endogenous retrovirus1 (PABL_A type); L1-seq, Hemi-specific PCR coupled to Illumina sequencing [14]; RC-seq, Retrotransposon capture sequencing, involving hybridization of fragmented genomic DNA to custom retrotransposon sequence capture arrays followed by deep sequencing [15]; Trio, mother, father and child; WGS, Whole-genome sequencing. *Ewing et al. examined only gene retrocopy insertion polymorphisms (GRIPs), which are processed gene transcripts present as retrotransposed insertions in one or more individuals but absent from the reference genome.
the total number of tumor-specific insertions. On the other hand, according to the scenario of Figure 1A, claims for the tumor specificity of much de novo retrotransposition and its absence in matched normal cells might be illusory. This scenario assumes that retrotransposition occurs at significant rates in normal somatic cells. However, in non-tumor tissue sampled in bulk, an individual new somatic insertion may be present in only a single cell, or at most several cells, among the large total population of cells sampled and consequently exist in too small a copy number to be detected. An insertion initially present within a normal cell is more easily detected once that cell by chance clonally expands as a tumor, which, upon sampling, high-throughput sequencing and PCR validation, would falsely appear to possess a tumor-only event. Thus, although tumor-specific de novo retrotransposition events may be underestimated, retrotransposition in normal cells is likely grossly underestimated.

Solyom et al. [10] adopted three indirect approaches to conclude that most, if not all, of the insertions they studied occurred post-tumor initiation: (1) by finding an L1 insertion to be absent in a second section of the same tumor, (2) by detecting empty-site X chromosome alleles in males with an X chromosome tumor insertion, and (3) by querying the heterozygosity of single-nucleotide polymorphisms (SNPs) flanking an L1 insertion site (with the presence of both SNP alleles in the empty-site chromosome implying that the insertion occurred after the one-cell stage of the tumor). Because the tumor samples assayed in this study were not microdissected, however, all three pieces of evidence cited may be subject to misinterpretation because of normal tissue present within the tumor sample. Interestingly, Shukla et al. [11] validated by PCR a single de novo L1 insertion in preneoplastic liver tissue that was absent in the corresponding hepatocellular carcinoma, although they also considered the possibility that chromosomal loss within the tumor could have deleted the insertion.

Thus, tumor-specific retrotransposon insertions occur, perhaps frequently in some tumors, but apparently vary greatly in number between different types of tumors and between individual tumors of the same cancer type. Although 80 to 100 L1s are estimated to be potentially active in any given human diploid genome [21], each particular genome can harbor its own unique, active L1s, or L1s shared between different individuals may vary in activity. My “hot” L1 may not be your hot L1; perhaps my mobilome has greater “mutational power” than yours [22,23]. Add to this fact the variability in the epigenetic state of individual active L1s, plus unknown genetic variations in the many cellular factors that associate with the L1 to affect its life cycle [24-26], and the task of assessing rates of cancer retrotransposition becomes complicated indeed.

High-throughput sequencing methods can produce false-positive results, underlining the need for validation of a significant number of the total putative somatic insertions found. As much as possible, it is important to identify 5’ as well as 3’ junctions of an insertion to confirm the presence of a poly(A) tail and a target site duplication (TSD), the hallmarks of a true retrotransposition event. A minor but significant number of L1s insert not by target-primed reverse transcription (TPRT), the standard

![Figure 1](http://www.mobilednajournal.com/content/5/1/11)

**Figure 1** Bulk tissue vs. single-cell detection of somatic retrotransposition. (A) Bulk tissue sampling can underestimate the number of normal tissue retrotransposition events. (B) Single-cell sampling can provide truer estimates of rates of normal somatic cell retrotransposition. The numbers of unique de novo retrotransposon events in each cell are indicated. A minor percentage of normal epithelial cells (light brown) contain new unique insertions. Occasionally, a cancer stem cell (dark brown) gives rise to a tumor (green). Every cell of the tumor will contain the de novo retrotransposon insertions present in the originating stem cell. New tumor-specific events may also subsequently occur. Red circles indicate bulk (A) or single-cell (B) sampling for downstream sequencing analyses.
model for L1 retrotransposition [27], but rather by an endonuclease-independent mechanism. These insertion events likely occur at preexisting DNA lesions and generate integrants lacking TSDs [28-30]. Interestingly, apparent endonuclease-independent insertions have been detected in significantly increased numbers in some tumors (8 of 35 colorectal cancer insertions reported by Solyom et al. [10]).

Obtaining unbiased estimates of de novo retrotransposition in normal as well as tumor cells is critical for understanding somatic mosaicism, cancer induction, tumor heterogeneity, and the etiology of some neurological diseases [31,32]. Evrony et al. [33] recently used multiple displacement amplification of single neurons isolated from the frontoparietal cortex and caudate nucleus of three normal individuals to confirm somatic neuronal retrotransposition, but at a rate much lower (0.04 to fewer than 0.6 unique insertions per neuron) than estimates previously suggested by quantitative PCR (qPCR) analyses (a startling “theoretical” increase of about 80 L1 copies per hippocampal neuron compared with heart and liver samples [34]). Using a high-throughput method called RC-seq to analyze bulk DNA, Baillie et al. [15] found almost 8,000 putative somatic L1 insertions in the hippocampus and caudate nucleus of three individuals, an insertion rate much lower than Coufal et al. [34]'

qPCR techniques are increasingly being used to conclude elevated L1 genomic copy numbers in some cell types or under some cellular conditions [34-39]. Typically, the apparent increase in retrotransposon insertions is not verified by downstream sequencing. Apart from the fact that small changes in protocol can alter the results of such sensitive qPCR analyses, an additional but untested source of bias conceivably exists: that is, the promiscuous reverse transcription (RT) of retrotransposon RNAs “free-floating” in the cell and not engaged in TPRT at a site of chromatin integration. Such ectopic RT reactions might be primed by random complementary nucleic acids or perhaps by fold-back and annealing of the L1 poly(A) tail to one of the ten homopolymeric stretches of four to seven U residues that occur across the length of the L1 RNA molecule. Fold-back self-priming of RNA has been observed for a number of RNA-dependent RNA polymerases, including reverse transcriptases [40-42]. It has also been shown that RT can initiate from internal sites within L1 RNA during endonuclease-independent insertion [28]. One would expect copy numbers of orphan L1 cDNAs generated by promiscuous RT to be higher in cells with elevated expression of L1 ORF2 or perhaps endogenous retroviral reverse transcriptases; indeed, increased RT activity in some types of neuronal and tumor cells has been reported [43-45]. Because these cDNAs are amenable to PCR amplification, qPCR-based estimates of genomic L1 insertion copy numbers in these cells using L1-specific primers would be erroneously high. This source of bias, if true, is not in disagreement with the cis-preference model for L1 retrotransposition. This model states that a retrotransposition-competent integration intermediate consists of L1 RNA bound in cis by its own encoded ORF1 and ORF2 proteins [46,47]. However, the total number of L1 protein [48] and RNA molecules present in the cell likely greatly exceed in number those bound in cis within bona fide insertion intermediates. These molecules can bind to each other in trans and perhaps engage in RT.

Conclusions
The surprising discovery of frequent somatic retrotransposition has important implications for human health. To fully and accurately assess its extent, concerted efforts are needed to sequence many single cells from a large number of cell types (normal and cancerous) from many individuals, with thorough validation by PCR of individual retrotransponson inserts and capillary sequencing of the PCR products to confirm their identity (Figure 1B). Single-cell, whole-genome amplification is a recent technology and is not without biases, including allelic and locus dropouts, chimeric molecules, uneven amplification due to local variations in G-C content, and incorrect nucleotide insertions [33,49,50]. Such artifacts complicate the identification of somatic retrotransposon insertions. However, the ongoing development of new protocols, such as MALBAC (multiple annealing looping-based amplification cycles [51]), promises to reduce bias. Next-generation, single-cell sequencing protocols should help to establish the impact that ongoing L1 retrotransposition manifests in brain biology, cancer, and likely other human diseases where L1 mobilization has not yet been explored. It should also become possible to trace the history of a single somatic retrotransposon back to its origin in a specific cell type or at a specific stage of development.

Endnote
Six RC-Seq libraries, each prepared from 2.5 μg of DNA, yielded 7743 L1 insertions [15]. Since a human diploid cell contains 6.6 pg of DNA, there were on average .003 unique L1 insertions per brain cell. If most insertions were in neurons, the rate was 1 insertion per 30 to 150 neurons (the brain being variously estimated to be 10-50% neurons [52]).

Abbreviations
ERV: Endogenous retrovirus; LINE: Long interspersed nuclear element; LTR: Long-terminal repeat; ORF: Open reading frame; RC-seq: Retrotransposon capture sequencing; SINE: Short interspersed nuclear element; SVA: SINE-R, VNTR and Alu; TSD: Target site duplication; UTR: Untranslated region; TPRT: Target-primed reverse transcription; WGS: Whole-genome sequencing.
Competing interests
The author declares that he has no competing interests.

Acknowledgements
The author thanks Drs Jose Garcia Pérez, Szelvia Solyom, Jens Mayer, Haig Kazazian, Geoffrey Faulkner and John Moran for critical reading of the manuscript.

Received: 10 January 2014 Accepted: 17 March 2014 Published: 7 April 2014

References
1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Hefuda A, Howland J, Kann L, Lehoczyk J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. *Nature* 2001, 409:860–921. A published erratum appears in *Nature* 2001, 412:565.

2. Hancks DC, Kazazian HH Jr: Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* 2012, 22:191–203.

3. Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran JV, Gage FH: Endogenous retrotransposons, cancer stem cells and oncogenesis. *FEBS J* 2014, 281:63–73.

4. Garcia-Perez JL, Marchetto MC, Muotri AR, Coufal NG, Gokcumen O, Haseley P, Lohr JG, Luquette LJ 3rd, Lohr JG, Goodier MW: Endogenous retrotransposons, cancer stem cells and oncogenesis. *FEBS J* 2014, 281:63–73.

5. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Osterag EM, Kazazian HH Jr: L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Gene* 2009, 435:903–910.

6. Garcia-Perez JL, Marchetto MC, Muotri AR, Coufal NG, Gokcumen O, Haseley P, Lohr JG, Goodier MW: Endogenous retrotransposons, cancer stem cells and oncogenesis. *FEBS J* 2014, 281:63–73.

7. Wissing S, Muñoz-Lopez M, Macia A, Yang Z, Montano M, Collins W, Garcia-Perez JL, Moran JV, Greene WC: Reprogramming somatic cells into IPS cells activates L1 retroelement mobility. *Hum Mol Genet* 2012, 21:208–218.

8. Miki Y, Nishishio I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y: Disruption of the APC gene by a retrotransposon insertion of L1 sequence in a colon cancer. *Cancer Res* 1992, 52:643–645.

9. Lee E, Ishikawa Y, Yang L, Kogure K, Haseki F, Ueyama A, Hamaoka T, Ueda M, Kato S, Harada K, Furuta S, Miura H, Uekusa H, Sonoda S, Ueno Y, Komori T, Nothke F, Maruyama Y, Sekiya Y, Miyauchi H, Yoshida A, Ono Y, Nakamura Y: Endogenous retrotransposons, cancer stem cells and oncogenesis. *FEBS J* 2014, 281:63–73.

10. Lee E, Ishikawa Y, Yang L, Kogure K, Haseki F, Ueyama A, Hamaoka T, Ueda M, Kato S, Harada K, Furuta S, Miura H, Uekusa H, Sonoda S, Ueno Y, Komori T, Nothke F, Maruyama Y, Sekiya Y, Miyauchi H, Yoshida A, Ono Y, Nakamura Y: Endogenous retrotransposons, cancer stem cells and oncogenesis. *FEBS J* 2014, 281:63–73.

11. Gross F, Szymanski C, Wurfel JJ, Goodier MW: Endogenous retrotransposons, cancer stem cells and oncogenesis. *FEBS J* 2014, 281:63–73.

12. Hancks DC, Kazazian HH Jr: Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* 2012, 22:191–203.

13. Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran JV, Gage FH: Endogenous retrotransposons, cancer stem cells and oncogenesis. *FEBS J* 2014, 281:63–73.

14. Garcia-Perez JL, Marchetto MC, Muotri AR, Coufal NG, Gokcumen O, Haseley P, Lohr JG, Goodier MW: Endogenous retrotransposons, cancer stem cells and oncogenesis. *FEBS J* 2014, 281:63–73.
39. Bundo M, Toyoshima M, Okada Y, Akamatsu W, Ueda J, Nemoto-Miyauchi T, Sunaga F, Toritsuka M, Ikawa A, Kato M, Kasai K, Kishimoto T, Nawa H, Okano H, Yoshikawa T, Kato T, Iwamoto K: Increased L1 retrotransposition in the neuronal genome in schizophrenia. Neuron 2014, 81:306–313.

40. Levin HL: An unusual mechanism of self-primed reverse transcription requires the RNase H domain of reverse transcriptase to cleave an RNA duplex. Mol Cell Biol 1996, 16:5645–5654.

41. Laurila MR, Salgado PS, Stuart DI, Grimes JM, Barnford DH: Back-priming mode of Φ6 RNA-dependent RNA polymerase. J Gen Virol 2005, 86:521–526.

42. Tuiskunen A, Leparc-Goffart I, Boubis L, Monteil V, Klingström J, Tolou HJ, Lundkvist A, Plumet S: Self-priming of reverse transcriptase impairs strand-specific detection of dengue virus RNA. J Gen Virol 2010, 91:1019–1027.

43. Douville R, Liu J, Rothstein J, Nath A: Identification of active loci of a human endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis. Ann Neurol 2011, 69:141–151.

44. Sciamanna I, Guaitieri A, Cosetti C, Osimo EF, Ferracin M, Macchia G, Anciò E, Prosseda G, Vittulo P, Mistelli T, Spadafora C: A tumor-promoting mechanism mediated by retrotransposon-encoded reverse transcriptase is active in human transformed cell lines. Oncotarget 2013, 4:2271–2287.

45. Pattna R, Lee SH, Dahmstrom JE, Ohms S, Chen L, Dheen ST, Rangasamy D: Inhibition of LINE-1 retrotransposon-encoded reverse transcriptase modulates the expression of cell differentiation genes in breast cancer cells. Breast Cancer Res Treat 2013, 143:239–253.

46. Enault C, Maestre J, Heidmann T: Human LINE retrotransposons generate processed pseudogenes. Nat Genet 2000, 24:363–367.

47. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV: Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol 2001, 21:1429–1439.

48. Mandal PK, Ewing AD, Hancks DC, Kazazian HH Jr: Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles. Hum Mol Genet 2013, 22:3730–3748.

49. Lasken RS: Single-cell sequencing in its prime. Nat Biotechnol 2013, 31:211–212.

50. Macaulay IC, Voet T: Single cell genomics: advances and future perspectives. PLoS Genet 2014, 10:e1004126.

51. Zong C, Lu S, Chapman AR, Xie XS: Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 2012, 338:1622–1626.

52. Hilgetag C, Barbas H: Are there ten times more glia than neurons in the brain? Brain Struct Funct 2009, 213:365–366.

do:10.1186/1759-8753-5-11
Cite this article as: Goodier: Retrotransposition in tumors and brains. Mobile DNA 2014 5:11.