Aperito Journal of Recent Trends in Genomics and Proteomics

Received: Feb 28, 2015
Accepted: Mar 16, 2015
Published: Mar 19, 2015

http://dx.doi.org/10.14437/AJRTGP-1-104 Review Chi Zhang, Aperito J Recent Trends Genome Proteom 2015: 1:1

Recent Advances in LINE-1 Retrotransposon Biology in the Brain

Chi Zhang*

Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA, 02115, USA

Abstract

It is now widely recognized that retrotransposons such as Long Interspersed Element-1 (LINE1 or L1s) have played an instrumental role in shaping the structure and function of our genomes. L1s contribute to ~17% of total human genome, and there is increasing evidence of its active role in neuronal development. L1s retrotransposition was proposed as a potential mechanism of generating neuronal genome diversity, and multiple groups have attempted to characterize L1s retrotransposition events and their impacts on brain function and physiology. L1s are also linked to cognitive and psychiatric diseases such as Rett syndrome and schizophrenia. Here, I discuss the recent progress in characterizing the L1s retrotransposition in neuronal development, and how the knowledge has contributed to our understanding of brain function and diseases.

***Corresponding Author:** Chi Zhang, Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA, 02115, USA; E-mail: chizhang@genetics.med.harvard.edu

Introduction

Long Interspersed Element-1 (LINE1 or L1s) are important components of the genome, constituting ~17% of the human genome [1]. L1s are the only type of retrotransposons that are known to mobilize autonomously in humans, and therefore have played an important role in shaping the human genome and contributing to our genome evolution.

A more comprehensive review of the different aspects of L1s biology can be found elsewhere (Beck et al. 2011), so only a very brief overview of the general L1 biology will be presented here. A full length L1 element has a length of ~6kb, and contains two Open Reading Frames (ORFs) and 5' and 3' UnTranslated Regions (UTRs) (Figure 1). The 5' UTR harbors an internal RNA polymerase II promoter that directs synthesis of full-length, poly adenylated L1 transcripts [2]. ORF1 encodes a high-affinity RNA binding protein that forms the L1 Ribo Nucleo Protein particles (L1 RNP) with the L1 RNA. ORF1 protein (ORF1p) is also thought to function as an RNA chaperon that contributes to L1 integration into the genomic loci [3]. Compared to ORF1, ORF2 is better known, and encodes a 146kd protein that functions as the endonuclease and reverse transcriptase, both of which are critical for L1 retrotransposition [4, 5]. The ORF2p also contains a cysteine-rich region near the C-terminus that is also important for retrotransposition, although its exact function remains unknown [6]. The 3'UTR of L1s has a relatively weak polyA signal, and also is suggested to bind the nuclear export factor NXF1 (TAP) to aid in nuclear export of the L1s RNA [7]. It's unclear whether the 3'UTR is critical for retrotransposition, since deleting a 145bp region of the 3'UTR containing the polyA signal had little effect on exogenous L1 retrotransposition in Hela cells [6].

Although the human genome contains roughly half a million copies of L1s, majority of these (over 99.9%) are

inactive due to 5' truncations, inversion, and/or point mutations within the two ORFs [8]. In fact, the average length of all L1 copies in the human genome is ~900bp, and the median size is ~1kb [1]. It is estimated that the average human genome contains 80-100 copies of retrotransposition-competent L1s [9].

Evidence of L1s retrotransposition during Brain development

Although originally considered a mutational force limited in germ cells or in early embryogenesis before lineage commitment [10], L1s retrotransposition is increasingly being recognized as a means of generating somatic mutations in contexts such as brain development and tumorigenesis. Importantly, somatic mutations generated by L1s in the brain have been proposed to play a role in diversifying neuronal functions through generating mosaicism. I will discuss in detail the recent findings regarding L1s retrotransposition in the brain.

The first piece of evidence regarding L1s retrotransposition during brain development came when Muotri et al. examined RNA transcripts enriched in rat adult hippocampus neural stem cells treated with glycosylated form of Cystatin C (CCg) [11]. CCg stimulates proliferation of multi potent Neural Progenitor Cells (NPCs) from a heterogeneous population. L1s transcripts were enriched by 1.5-2 folds in CCg-responsive cells, suggesting an upregulation of L1 retrotransposition. Importantly, using an artificial reporter of L1s retrotransposition (called L1 cassettes) where EGFP expression is turned on only after retrotransposition into a genomic locus, the authors showed that rat NPCs can support retrotransposition of human L1s, whereas no evidence of L1 cassette retrotransposition was detected in rat primary neurons and astrocytes, nor mesenchymal stem cells or fibroblasts. Functionally, L1 cassette insertions are shown to affect expression of target genes and result in phenotypic changes in cell behavior. The authors further examined in vivo retrotransposition in the mouse brain using the same L1 cassette strategy, and showed that retrotransposition-positive cells colocalized with neuronal but not glial markers. Therefore, this study demonstrated for the first time that L1s retrotransposition is active in neural progenitors during the neurogenesis stage.

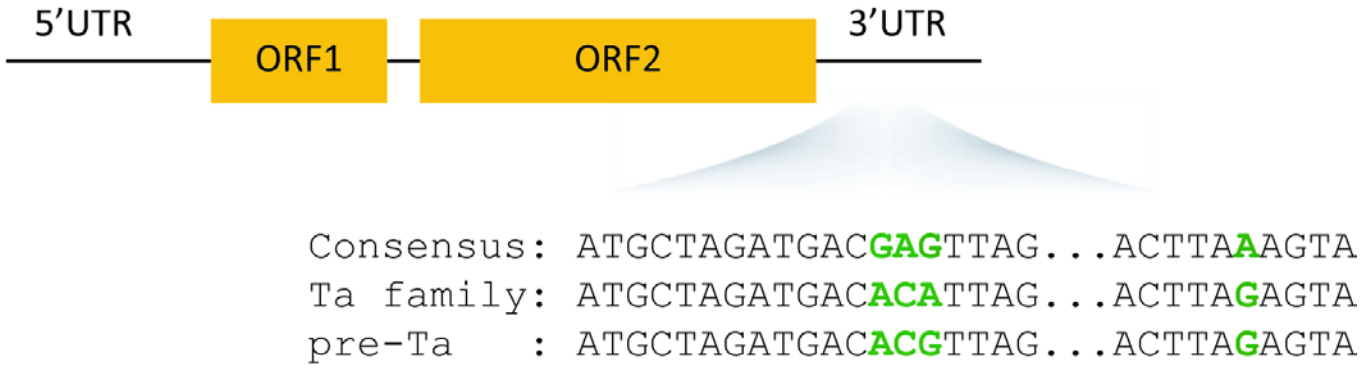
The finding of L1s retrotransposition during neuronal development spurred a significant interest in understanding the mechanistic details of how L1s activity is controlled during neurogenesis, and what the functional impacts are. Kuwabara et al. identified that wnt/ -catenin pathway promotes L1s transcription and activity [12], and Muotri et al. further show that MECP2 normally suppresses L1s retrotransposition in neurons, likely through its role in mediating DNA methylation of L1 locus [13].

Similar to NPCs derived from rat adult hippocampal progenitor cells, NPCs isolated from human fetal brains or derived from human embryonic stem cells support retrotransposition of L1 cassettes in vitro [14]. Using a quantitative PCR method on samples derived from adult human tissue, the authors found that copy number of L1s in the hippocampus is significantly higher compared to samples derived from heart and liver from the same individual. This evidence suggests that the brain supports L1s retrotransposition to a significantly greater extent than other tissues, and the extent of somatic mosaicism generated by L1s might have important functional implications to brain function.

Recent advances of Next-Generation sequencing (Next-Gen sequencing) has enabled accurate mapping of genomic locations of individual L1 elements within the genome. The first study by Ewing et al. examined locations of L1s elements in individual human genomes [15]. Their approach took advantage of the fact that a subfamily of L1s, termed the Ta subfamily (also called L1H family), is responsible for majority of retrotransposition events in humans, and carry nucleotides at the 3'UTR that are distinct from other subfamilies (Figure 1). Taking advantage of these nucleotides (called "diagnostic nucleotides") and a clever strategy of using degenerate nucleotide sequences that anneal to the flanking genome, the authors were able to PCR up the 3' end of L1s, together with a piece of the flanking genomic DNA 3' to the L1s sequence. The information of the 3' genomic sequences allows mapping the specific L1s element to its genomic

location. Using this strategy, the authors discover that on average, there are 285 distinct L1 insertion presences or absences when comparing any two individual genomes, and estimate the rate of germline L1 retrotransposition between 1/95 to 1/270 births. This study demonstrated the power of using Next-Gen sequencing approaches to systematically study L1 genomic insertions.

Figure 1: Schematic of the genomic structure of L1s. Signature nucleotides at the 3'UTR specific for Ta (also called L1H) and pre-Ta subfamilies [29] are shown in green and bold.



A number of studies followed Ewing et al. to examine the L1s landscape in the brain genome. Baillie et al. was the first group that applied Next-Gen sequencing to examine L1s in the brain [16]. The authors developed Retrotransposon Capture sequencing (RC-seq) that hybridizes fragmented genomic DNA onto sequence capture arrays targeting the 5' and 3' termini of full-length retrotransposons including L1s and other retrotransposons. The captured DNA is then subjected to deep sequencing and afterwards mapped to the genome to identify known and novel retrotransposon insertions. Remarkably, out of all RC-seq-identified L1 insertion sites in the brain, 98.1% were annotated as somatic insertions by the definition that brain somatic insertions occur only in one region of a single individual, do not overlap with insertions identified in the blood samples, and do not match known L1 polymorphisms. The authors conclude that on average, each human individual contains ~2,500 potential L1 somatic insertions in the brain, and somatic insertions disproportionately affect protein-coding loci, with an enrichment for genes relevant to neurogenesis and synaptic function. As an example, L1 somatic insertion into the HDAC1 locus, an epigenetic regulator implicated in multiple neurodevelopmental and neuroplastic processes, is identified in one individual. Therefore, somatic L1 retrotranspositions are likely to fundamentally alter the genetic landscape of the brain, with important functional implications on brain functions.

One important caveat of the above mentioned studies is that pools of bulk tissue are used for genomic DNA extraction, and therefore, L1 insertions with a higher abundance in the tissue are more likely to be detected compared to low-abundance L1 insertions, which might introduce systemic biases. To address this problem, Evrony et al. developed a method to perform single-cell genome amplification to study somatic mutations. Single purified pyramidal neuronal nuclei are subjected to Multiple Displacement Amplification (MDA) for whole-genome amplification [17]. This amplification method was efficient and is estimated to capture up to 90% of retrotransposon insertions per single cell. To enrich for L1 insertions, the authors adopted a strategy similar to that of Ewing et al. where primers specific to L1H subfamily are used to amplify the 3' end of L1Hs together with a piece of the genomic DNA. Indeed, the authors are able to identify specific L1H insertions in single neuron genome samples that were

undetected in pooled tissue DNA, or amplified DNA from 100 neuron samples, suggesting that this method has a higher sensitivity for low-abundance somatic insertions.

Using the single-genome amplification strategy, Evrony et al. studied somatic L1H insertion rate in the cerebral cortex and caudate nucleus of normal individuals, and estimate that 68% of neurons sampled in the study do not have detectable somatic L1H insertions, and on average, each single neuron harbors 1.1 somatic insertions. When examining unique somatic insertions that are present in one 1-neuron sample but not other 1-neuron samples from the same individual, it is estimated that 82% of 1-neuron samples do not have detectable unique somatic insertions, with the average being 0.6 unique insertions per neuron. The authors further speculate that real insertion rates are likely even lower, taking into account errors introduced by sequencing and amplification etc.

The strategy of tracing transpositions has been used to successfully perform lineage tracing along developmental and pathological time points [18], and similarly, Evrony et al. leveraged the single neuronal genome amplification strategy to trace cell lineages with L1 locations, and showed that by whole-genome sequencing of single neurons and distinguishing between two independent L1H retrotranspositions, they were able to distinguish two clones of cells and their clonal distribution in the human brain [19].

Consensus and discrepancies from multiple studies

The power of massive parallel genome sequencing is just starting to be applied to study L1 retrotransposition during brain development, and discrepancies exist regarding rates and frequencies of L1s retrotransposition from multiple studies employing different strategies. Table 1 summarizes the current data on the rate and frequency of L1s retrotransposition and the strategies used to perform these studies.

Table 1: Summary of L1s retrotransposition frequencies reported from different studies and their experimental approaches.

Organ or region (Species)	Experimental approach	Frequency of L1s retrotransposition	Reference
Hippocampus (Human)	**Quantitative PCR** (compare copy number of L1 ORF2 in the hippocampus vs liver)	80 per cell	Coufal et al. 2009
Adult hippocampal progenitor cells (Rat)	**EGFP-L1 cassettes** (EGFP expression dependent on retrotransposition of cassette into the genome)	0.75%-1.5% of all cells	Muotri et al. 2005
iPS-induced NPCs (Human)	**EGFP-L1 cassettes**	0.4% of all cells	Muotri et al. 2010
Cultured Hela cells (Human)	**EGFP-L1 cassettes**	0.5% of all cells per day	Ostertag et al. 2000
Cerebral cortex and caudate nuclei (Human)	**Next-Gen sequencing**	<0.6 unique somatic insertions per cell, and 82% neurons with no detectable unique somatic insertions	Evrony et al. 2012

In general, the consensus is that L1s retrotransposition is a rare event, especially in the brain. The discrepancies between the results obtained might be due to a number of factors including:

1) The use of exogenous reporter plasmids might result in an over-estimation of retrotransposition rates due to its high level of expression and high-copy number.
2) Quantitative PCR-based approach used by Coufal et al. which measures the number of L1 copies in the genome, uses spiked-in L1 plasmids to generate a standard curve of L1 copy number. This approach might be subject to systematic biases of amplification from genomic DNA vs plasmid DNA, which might explain the hundred fold differences between its and other estimates.
3) Recent approaches using Next-Gen sequencing with either bulk tissue sample or samples amplified from single cell genomes offer a potentially more accurate view of the genomic landscape of L1 retrotransposition, yet are restrained by sequencing depth limitations, amplification bias during library preparation (especially for single cell samples), amplification/sequencing errors, and informatics difficulties in successfully aligning short and rare reads.

Functional consequences of L1 retrotransposition

Given the low frequency of L1 retrotransposition events, it has been difficult to study the functional impacts of L1 retrotransposition to brain function. The field has been speculating that the somatic mosaicism generated by L1 insertions might be a mechanism of diversifying neuronal function during neuronal development. On a gene expression level, retrotransposition has been associated with the following functions:

1) Mutagenesis of target genes

Transposable elements have long been known as a mutagenic force in the genome, and can lead to disruptive mutations of the genome such as exon disruption, altered splicing, genomic deletion etc. Consequently, such disruptive mutations, if occurring on essential genes, can lead to diseases. At least 11 human diseases have been described that are caused by L1 insertion into their corresponding loci, such as X-linked dilated cardiomyopathy [20, 21].

2) Regulation of target gene expression

L1 insertion can impact target gene expression via multiple manners. Insertion of L1 sequence onto a transcript reduces target RNA and protein expression through impaired transcriptional elongation [22]. L1s located immediately 5' of protein-coding genes frequently function as alternative promoters, whereas genes with L1s inserted at the 3' UTR frequently show reduced expression than those that do not have L1s insertion [23]. Last but not least, extensive splicing of L1 transcript might also lead to mis-incorporation of L1 splice sites into the splice products of targets genes that the L1 inserts into, which very likely will significantly interfere with normal gene expression [24].

Recent evidence is showing that L1 retrotransposition might contribute to the pathogenesis of several neurological diseases, potentially through functionally impacting the structural integrity and expression output of the genome. A few of the examples are illustrated below:

1) Rett syndrome

Rett syndrome is a neurodevelopmental disorder characterized by intellectual disability, seizure and loss of functional capabilities during normal early development. Almost all cases of Rett syndrome are caused by mutations of MECP2, a gene critical for epigenetic silencing and transcriptional repression. Muotri et al. demonstrated that MECP2 silences L1 expression in neural stem cells, and in both animal models and Rett patient iPS-induced NPCs,

there is a significant increase of L1 retrotransposition [13]. Importantly, quantification of L1 copy number in patient brains also revealed an increased L1 copy number in Rett patients. These data suggest that L1 retrotransposition is a characteristic of Rett pathology, and might lead to the genetic and epigenetic changes of mature neurons.

2) Schizophrenia

Genetic factors have long been speculated to play a causative role in schizophrenia due to its high heritability. Bundo et al. quantified copy numbers of L1s in normal and schizophrenia patient brain and iPS-induced neurons, and report that samples from schizophrenia patients show higher L1 copy numbers [25]. This is also true in mouse and monkey models of schizophrenia. Interestingly, whole genome sequencing and gene ontology analysis revealed that in schizophrenia patients, L1 preferentially inserts into genes related to synaptic functions and psychiatric disorders, suggesting a potential causative correlation between L1 insertion and schizophrenia pathogenesis.

3) Age related neurological diseases

The endonuclease activity encoded by L1 ORF2 is known to nick chromosomal DNA non-specifically. This has led to the thinking that even if L1s retrotransposition failed to happen efficiently, the ORF2 nicking activity might result in cumulative DNA damage and negatively impact genomic integrity [26]. In addition, cellular stress, such as DNA damage [27] and oxidative stress [28], is known to induce L1 expression, which might further exacerbate mutagenic forces in the genome. All these have led to the hypothesis that expression of retrotranspositionally active L1s or L1 ORF2 alone might compromise normal cell function and contribute to cellular aging [26]. This might have relevance to age-related neurodegenerative diseases such as Alzheimer's disease, given that adult neural stem cells continue to generate functional neurons that populate the hippocampus. Further studies are needed to test this possibility.

Conclusion

L1s have been a significant player of human genome evolution, and recent studies established its activity during neuronal development. Although further studies are needed to better understand the frequency and general nature of L1s in the brain, it likely plays a role in shaping the neuronal genome and the pathophysiology of neurological diseases. Further improvements of the sequencing technologies will also aid in further discoveries of L1 biology in the brain.

References

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).
2. Swergold, G. D. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Molecular and cellular biology* **10**, 6718-6729 (1990).
3. Martin, S. L. The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. *Journal of biomedicine & biotechnology* **2006**, 45621, doi:10.1155/jbb/2006/45621 (2006).
4. Feng, Q., Moran, J. V., Kazazian, H. H., Jr. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905-916 (1996).
5. Mathias, S. L., Scott, A. F., Kazazian, H. H., Jr., Boeke, J. D. & Gabriel, A. Reverse transcriptase encoded by a human transposable element. *Science (New York, N.Y.)* **254**, 1808-1810 (1991).
6. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917-927 (1996).
7. Lindtner, S., Felber, B. K. & Kjems, J. An element in the 3' untranslated region of human LINE-1 retrotransposon mRNA binds NXF1(TAP) and can

function as a nuclear export element. *RNA (New York, N.Y.)* **8**, 345-356 (2002).

8. Doucet, A. J. *et al.* Characterization of LINE-1 ribonucleoprotein particles. *PLoS genetics* **6**, doi:10.1371/journal.pgen.1001150 (2010).
9. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5280-5285, doi:10.1073/pnas.0831042100 (2003).
10. Ostertag, E. M. *et al.* A mouse model of human L1 retrotransposition. *Nature genetics* **32**, 655-660, doi:10.1038/ng1022 (2002).
11. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903-910, doi:10.1038/nature03663 (2005).
12. Kuwabara, T. *et al.* Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nature neuroscience* **12**, 1097-1105, doi:10.1038/nn.2360 (2009).
13. Muotri, A. R. *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468**, 443-446, doi:10.1038/nature09544 (2010).
14. Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127-1131, doi:10.1038/nature08248 (2009).
15. Ewing, A. D. & Kazazian, H. H., Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome research* **20**, 1262-1270, doi:10.1101/gr.106419.110 (2010).
16. Baillie, J. K. *et al.* Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**, 534-537, doi:10.1038/nature10531 (2011).
17. Evrony, G. D. *et al.* Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* **151**, 483-496, doi:10.1016/j.cell.2012.09.035 (2012).
18. Sun, J. *et al.* Clonal dynamics of native haematopoiesis. *Nature* **514**, 322-327, doi:10.1038/nature13824 (2014).
19. Evrony, G. D. *et al.* Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49-59, doi:10.1016/j.neuron.2014.12.028 (2015).
20. Belancio, V. P., Hedges, D. J. & Deininger, P. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome research* **18**, 343-358, doi:10.1101/gr.5558208 (2008).
21. Yoshida, K., Nakamura, A., Yazaki, M., Ikeda, S. & Takeda, S. Insertional mutation by transposable element, L1, in the DMD gene results in X-linked dilated cardiomyopathy. *Human molecular genetics* **7**, 1129-1132 (1998).
22. Han, J. S., Szak, S. T. & Boeke, J. D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**, 268-274, doi:10.1038/nature02536 (2004).
23. Faulkner, G. J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nature genetics* **41**, 563-571, doi:10.1038/ng.368 (2009).
24. Belancio, V. P., Hedges, D. J. & Deininger, P. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic acids research* **34**, 1512-1521, doi:10.1093/nar/gkl027 (2006).
25. Bundo, M. *et al.* Increased l1 retrotransposition in the neuronal genome in schizophrenia. *Neuron* **81**, 306-313, doi:10.1016/j.neuron.2013.10.053 (2014).
26. St Laurent, G., 3rd, Hammell, N. & McCaffrey, T. A. A LINE-1 component to human aging: do LINE elements exact a longevity cost for evolutionary advantage? *Mechanisms of ageing and development* **131**, 299-305, doi:10.1016/j.mad.2010.03.008 (2010).
27. Hagan, C. R. & Rudin, C. M. Mobile genetic element activation and genotoxic cancer therapy: potential clinical implications. *American journal of pharmacogenomics : genomics-related research in*

drug development and clinical practice **2**, 25-35 (2002).

28. Rockwood, L. D., Felix, K. & Janz, S. Elevated presence of retrotransposons at sites of DNA double strand break repair in mouse models of metabolic oxidative stress and MYC-induced lymphoma. *Mutation research* **548**, 117-125, doi:10.1016/j.mrfmmm.2004.01.005 (2004).
29. Beck, C. R., Garcia-Perez, J. L., Badge, R. M. & Moran, J. V. LINE-1 elements in structural variation and disease. *Annual review of genomics and human genetics* **12**, 187-215, doi:10.1146/annurev-genom-082509-141802 (2011).