doi:10.1038/nature08248

LETTERS

L1 retrotransposition in human neural progenitor cells

Nicole G. Coufal[1], José L. Garcia-Perez[2,3], Grace E. Peng[1], Gene W. Yeo[1]†, Yangling Mu[1], Michael T. Lovci[1]†, Maria Morell[4], K. Sue O'Shea[4], John V. Moran[2,5] & Fred H. Gage[1]

Long interspersed element 1 (LINE-1 or L1) retrotransposons have markedly affected the human genome. L1s must retrotranspose in the germ line or during early development to ensure their evolutionary success, yet the extent to which this process affects somatic cells is poorly understood. We previously demonstrated that engineered human L1s can retrotranspose in adult rat hippocampus progenitor cells *in vitro* and in the mouse brain *in vivo*[1]. Here we demonstrate that neural progenitor cells isolated from human fetal brain and derived from human embryonic stem cells support the retrotransposition of engineered human L1s *in vitro*. Furthermore, we developed a quantitative multiplex polymerase chain reaction that detected an increase in the copy number of endogenous L1s in the hippocampus, and in several regions of adult human brains, when compared to the copy number of endogenous L1s in heart or liver genomic DNAs from the same donor. These data suggest that *de novo* L1 retrotransposition events may occur in the human brain and, in principle, have the potential to contribute to individual somatic mosaicism.

The human nervous system is complex, containing approximately 10^{15} synapses with a vast diversity of neuronal cell types and connections that are influenced by complex and incompletely understood environmental and genetic factors[2]. Neural progenitor cells (NPCs) give rise to the three main lineages of the nervous system: neurons, astrocytes and oligodendrocytes. To determine whether human NPCs can support L1 retrotransposition, we transfected human fetal brain stem cells (hCNS-SCns) (Fig. 1a)[3] with an expression construct containing a retrotransposition-competent human L1 (RC-L1) driven from its native promoter ($L1_{RP}$). The RC-L1 also contains a retrotransposition indicator cassette in its 3′ untranslated region (UTR), consisting of a reversed copy of the enhanced green fluorescent protein (EGFP) expression cassette, which is interrupted by an intron in the same transcriptional orientation as the RC-L1 (refs 4–7). The orientation of the cassette ensures that EGFP-positive cells will only arise if the RC-L1 undergoes retrotransposition (Supplementary Fig. 1a).

A low level of $L1_{RP}$ retrotransposition, averaging 8–12 events per 100,000 cells, was observed in three different human fetal brain stem cell lines (BR1, BR3 and BR4; Fig. 1d). By comparison, an L1 containing two missense mutations in the open reading frame 1 (ORF1)-encoded protein ($JM111/L1_{RP}$)[5,7] did not retrotranspose (Fig. 1b, d). Controls demonstrated precise splicing of the intron from the retrotransposed EGFP gene (Fig. 1b and Supplementary Figs 1 and 4), and indicated that L1 retrotransposition events were detectable by both PCR and Southern blotting 3 months after transfection (Fig. 1c). Moreover, reverse transcriptase PCR (RT–PCR) revealed that hCNS-SCns express endogenous L1 transcripts and that some transcripts are derived from the human-specific (L1Hs) subfamily[4,9,10] (Supplementary Fig. 6a, b and Supplementary Tables 4 and 5).

To determine whether L1 retrotransposition occurred in undifferentiated cells, we conducted immunocytochemical localization of cell-type-restricted markers in EGFP-positive hCNS-SCns. These cells expressed neural stem cell markers, including SOX2, Nestin, Musashi-1 and SOX1 (Fig. 1e and Supplementary Fig. 2a, b), and some co-labelled with Ki-67, indicating that they continued to proliferate (Supplementary Fig. 2c). EGFP-positive hCNS-SCns could also be differentiated to cells of both the neuronal and the glial lineages (Fig. 1f, g). Notably, $L1_{RP}$ did not retrotranspose using our experimental conditions in primary human astrocytes or fibroblasts, although a low level of endogenous L1 expression was detected in both cell types (Fig. 1d and Supplementary Figs 2d, e and 6a, b).

We next used two different protocols to derive NPCs from five human embryonic stem cell lines (hESCs; Fig. 2a). As in our previous study[1], NPC differentiation led to a ~25-fold increase in L1 promoter activity over a 2-day period, and then a decline (Fig. 2c); there was also a ~250-fold increase in synapsin promoter activity during differentiation (Supplementary Fig. 4b). H13B-derived NPCs expressed both endogenous L1 RNA and ORF1 protein[8], although the level of ORF1 protein expression was less than in the parental H13B hESC lines (Fig. 2d). HUES6-derived NPCs also expressed endogenous L1 RNA (Supplementary Fig. 6a, b), and sequencing indicated that some transcripts are derived from the L1Hs subfamily (Supplementary Tables 4 and 5). Similar studies performed with fetal brain, liver, and skin samples showed evidence of endogenous L1 transcription (Supplementary Fig. 6c, d and Supplementary Tables 4 and 5).

RC-L1 retrotransposition was readily detected at varying efficiencies in hESC-derived NPC lines (Supplementary Table 1 and Supplementary Figs 1 and 4f, g). Again, we determined that $JM111/L1_{RP}$ could not retrotranspose (Supplementary Table 1), that EGFP-positive NPCs expressed canonical neural stem cell markers (Fig. 2b, e and Supplementary Fig. 3c, d), and that EGFP-positive HUES6-derived NPCs could be differentiated to cells of both the neuronal and glial lineages (Fig. 2f and Supplementary Fig. 3e, f). The variability in retrotransposition efficiencies in hESC-derived NPCs probably depended on several factors (see Supplementary Table 1 for specific details).

Characterization of EGFP-positive neurons showed that some expressed subtype-specific markers (tyrosine hydroxylase (Fig. 2g) and GABA (γ-aminobutyric acid; data not shown)) and whole-cell perforated patch-clamp recording demonstrated that some HUES6-derived NPCs are functional (Fig. 2h–k; $n = 4$ cells). Furthermore, we demonstrated that an RC-L1 tagged with neomycin or blasticidin retrotransposition indicator cassettes could retrotranspose in NPCs (Supplementary Figs 1 and 4c–e)[5,11]. Some G418-resistant foci also expressed SOX3 and could be differentiated to a neuronal lineage (Fig. 2b).

[1]Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [2]Departments of Human Genetics and Internal Medicine, 1241 East Catherine Street, University of Michigan Medical School, Ann Arbor, Michigan 48109-5618, USA. [3]Andalusian Stem Cell Bank, Center for Biomedical Research, Avda Conocimiento s/n, University of Granada, 18100, Spain. [4]Department of Cell and Developmental Biology, 109 Zina Pitcher, University of Michigan Medical School, Ann Arbor, Michigan 48109-2200, USA. [5]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815-6789, USA. †Present address: Stem Cell Program, Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-5004, USA.

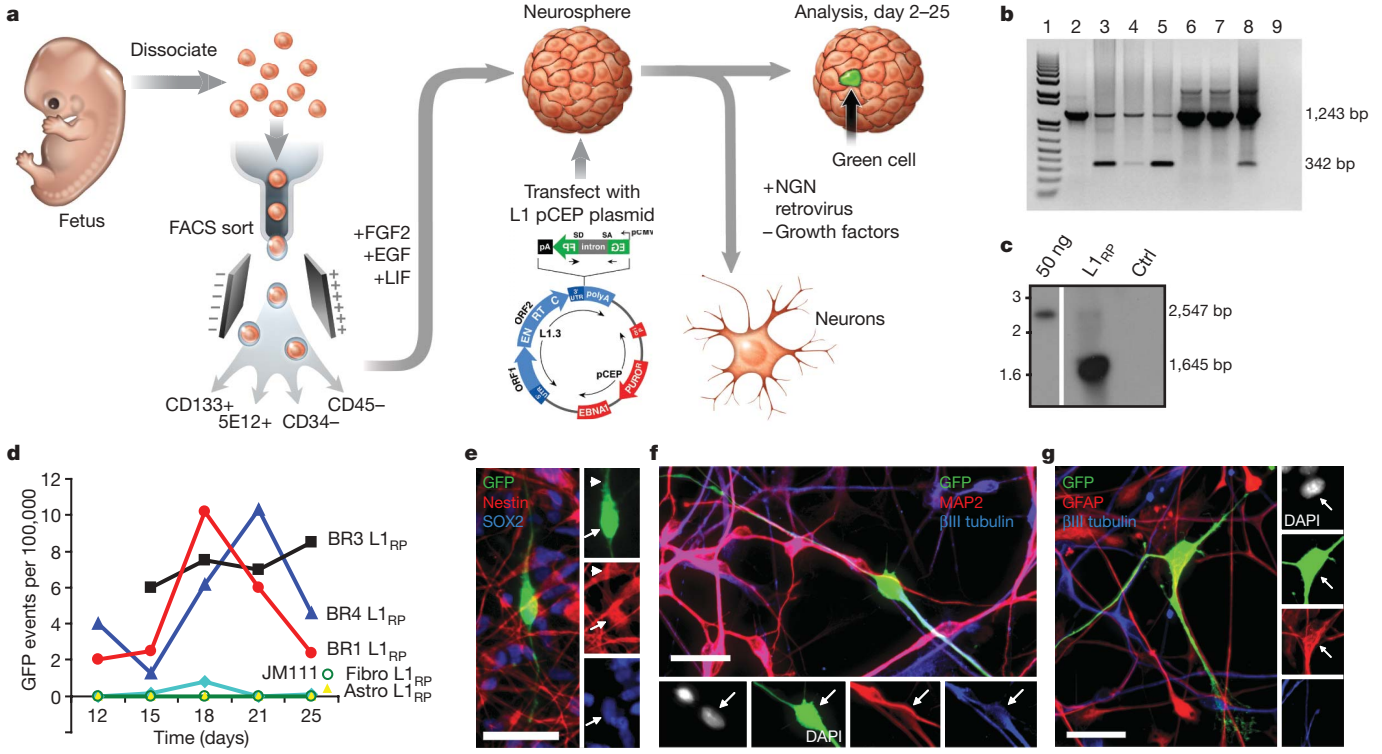


Figure 1 | L1 retrotransposition in hCNS-SCns. **a**, Experimental rationale. **b**, PCR of genomic DNA. The 1,243-bp product contains the intron, the 342-bp product indicates intron loss and retrotransposition. Lane 1, standards; lane 2, hCNS-SCns transfected with JM111/$L1_{RP}$; lanes 3–5, three human fetal brain stem cell lines transfected with $L1_{RP}$; lanes 6–7, primary astrocytes and fibroblasts transfected with $L1_{RP}$; lane 8, positive control; lane 9, water. **c**, Southern blot of hCNS-SCns (line FBR-BR3). The 2,547-bp band represents plasmid, the 1,645-bp band is diagnostic for genomic insertion. Ctrl, control. **d**, Time course of L1 retrotransposition. Astro, astrocytes; fibro, fibroblasts. **e**, EGFP-positive cells express Nestin and SOX2. Arrows indicate co-labelled cell body, arrowheads indicate co-labelled processes. **f**, EGFP-positive cells can differentiate to neurons (βIII tubulin and MAP2ab positive). **g**, EGFP-positive cells can differentiate into glia (GFAP-positive, βIII-tubulin-negative). Scale bars, 25 μm.

We next characterized 19 retrotransposition events from EGFP-positive NPCs (Supplementary Fig. 7b and Supplementary Table 2). Comparison of the pre- and post-integration sites demonstrated that retrotransposition occurred into an actual or inferred L1 endonuclease consensus cleavage site (5′-TTTT/A and derivatives). Five of eight fully characterized events were flanked by target site duplications, and no large deletions were detected at the insertion site[5,9,12] (Supplementary Fig. 7b and Supplementary Table 2). Interestingly, 16 out of 19 retrotransposition events were fewer than 100 kilobases (kb) from a gene and some occurred in the vicinity of a neuronally expressed gene[1,12,13].

Notably, we consistently observed higher L1 retrotransposition efficiencies in hESC-derived NPCs when compared to fetal NPCs. A Euclidian distance map on the basis of exon-array expression analysis[14] indicated that hCNS-SCns cluster closer to HUES6 cells, whereas HUES6-derived NPCs cluster closer to fetal brain (Supplementary Fig. 11a). Thus, hESC-derived NPCs and hCNS-SCns may represent different developmental stages in progenitor differentiation. That being stated, we conclude that engineered human L1s can retrotranspose in human NPCs.

Several studies have reported an inverse correlation between L1 expression and the methylation status of the CpG island in their 5′ UTRs[15,16]. Thus, we performed bisulphite conversion analyses on genomic DNAs derived from matched brain and skin tissue samples from two 80–82-day-old fetuses (Fig. 3a, one male/one female sample). We then amplified a portion of the L1 5′ UTR containing 20 CpG sites and sequenced the resultant amplicons. Notably, the L1 5′ UTR exhibited significantly less methylation in both brain samples when compared to the matched skin sample (two-sample Kolmogorov–Smirnov test, $P \leq 0.0079$ day 80 female, $P \leq 0.0034$ day 82 male; Fig. 3b). The analysis of individual L1 5′ UTR sequences demonstrated the greatest variation between the brain and skin at CpG residues located near the 3′ end of the amplicon, and six amplicons from the brain samples were unmethylated (Fig. 3e and Supplementary Fig. 8a, b). Restricting this analysis to ten L1s from brain and skin with the highest sequence homology to an RC-L1 showed that 19 out of 20 sequences were derived from the L1Hs subfamily (data not shown), and that one L1Hs element from the brain was completely unmethylated (Fig. 3c). In all cases, control experiments showed that the bisulphite conversion efficiency was >90% (Supplementary Fig. 8c).

Previous data indicated that SOX2 and MECP2 could associate with the L1 promoter and repress L1 transcription under some experimental conditions[1,17]. Two putative SRY-binding sites are located in the L1 5′ UTR immediately 3′ to the CpG island (Fig. 3a and Supplementary Fig. 11b)[18]. Thus, we performed chromatin immunoprecipitation (ChIP) for SOX2 and MECP2 in hCNS-SCns, HUES6-derived NPCs, and HUES6-derived neurons. SOX2 associated with the L1 5′ UTR in a pattern that correlates with the decrease in SOX2 expression observed during neural differentiation (Fig. 3d and Supplementary Fig. 4h). MECP2 expression was lower in both hCNS-SCns and HUES6-derived NPCs than in neurons (Supplementary Fig. 4h), and both hCNS-SCns and HUES6-derived NPCs expressed similar levels and types of L1 transcripts (Supplementary Fig. 6a, b). However, higher levels of MECP2 were detected in association with the L1 promoter in hCNS-SCns than in HUES6-derived NPCs (Fig. 3d). We propose that less L1 promoter methylation in the developing brain may correlate with increased L1 transcription and perhaps L1 retrotransposition, and the differential interaction of SOX2 and MECP2 with L1 regulatory sequences may modulate L1 activity in different neuronal cell types.

Although NPCs are useful to monitor L1 activity, they only allow monitoring of a single L1 expressed from a privileged context. By comparison, the average human genome contains ~80–100 active L1s, the expression of which may be affected by chromatin structure[4].

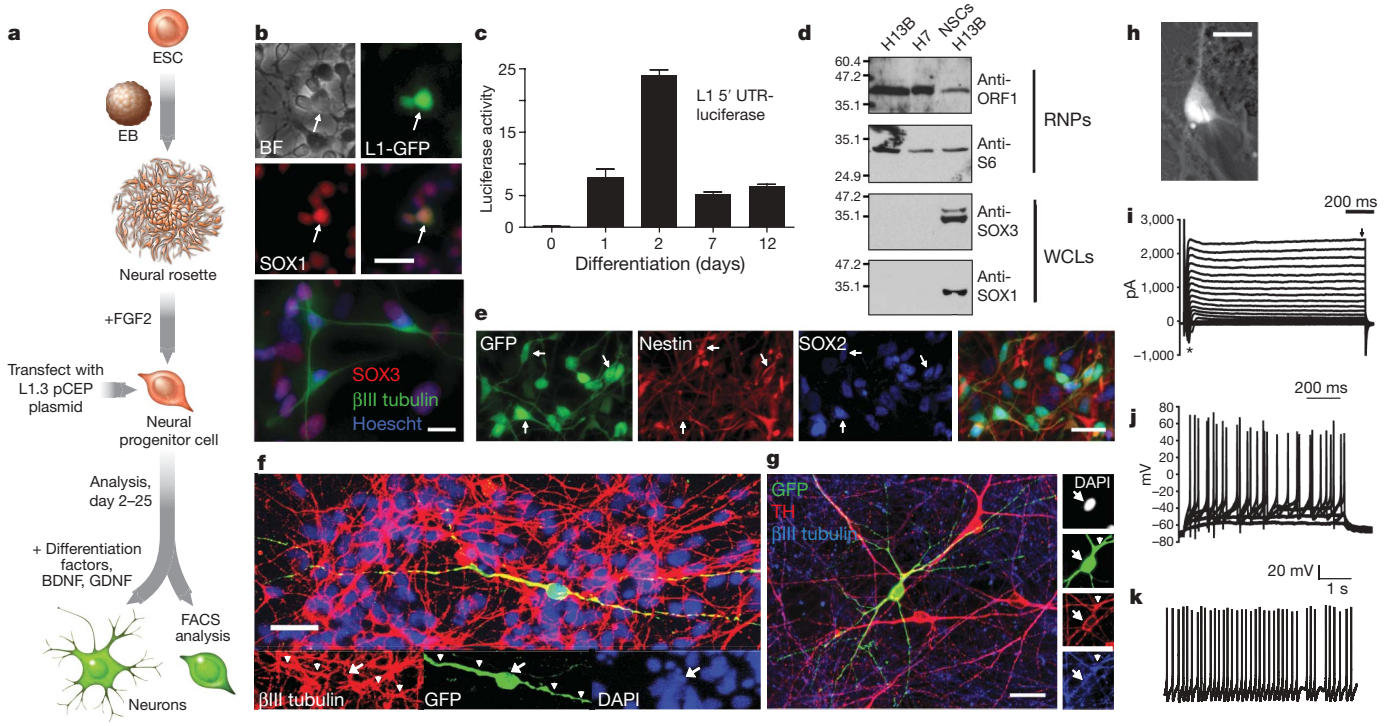
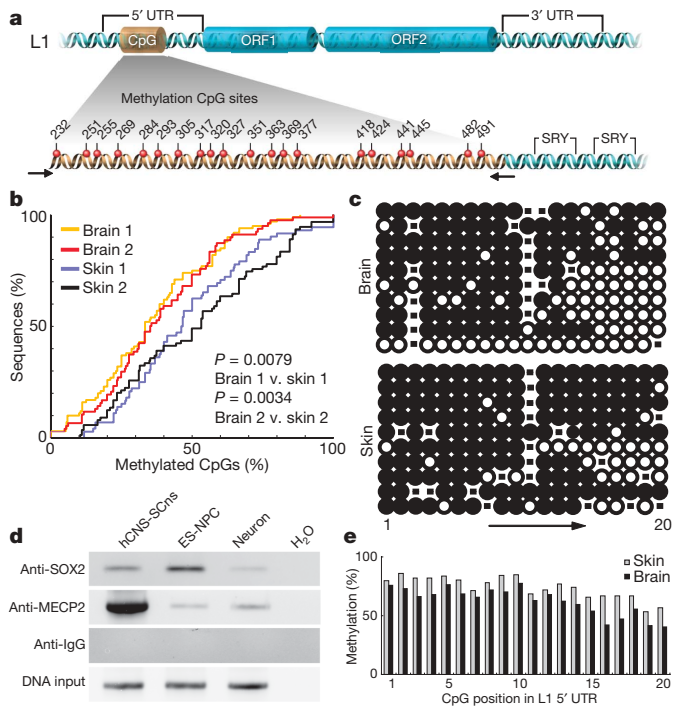


Figure 2 | L1 retrotransposition in hESC-derived NPCs. **a**, Experimental rationale. BDNF, brain-derived neurotrophic factor; EB, embryoid body; GDNF, glial-cell-derived neurotrophic factor. **b**, L1 retrotransposition in H13B (top, LRE3 with EGFP reporter) and H7 (bottom, LRE3 with neomycin reporter)-derived NPCs (BF, bright field). G418-resistant foci can express progenitor (SOX3) and neuronal (βIII tubulin) markers. **c**, L1 5′ UTR is induced upon differentiation. Error bars denote s.d. **d**, H13B-derived NPCs express endogenous ORF1 protein. RNP, ribonucleoprotein particle samples; WCL, whole cell lysates. Lane markers denote kDa. **e–g**, EGFP-positive, HUES6-derived NPCs express SOX2 and Nestin and can differentiate to be tyrosine hydroxylase (TH) positive. Arrows indicate cell soma co-localization, arrowheads indicate co-labelled processes. Scale bars, 25 μm. **h**, An LRE3-EGFP positive neuron. Scale bar, 10 μm. **i–k**, Data are derived from the neuron in **h**. **i**, Transient Na^+ (asterisk) and sustained K^+ (arrow) currents in response to voltage step depolarizations. **j**, Suprathreshold responses to somatic current injections. **k**, Spontaneous action potentials ($V_m = -50$ mV).

Therefore, we developed a quantitative multiplexing PCR strategy to investigate endogenous L1 activity in the human brain, proposing that active retrotransposition would result in increased L1 content in brain genomic DNA as compared to other tissues (Fig. 4a).



In brief, we designed Taqman probes against a conserved 3′ region of ORF2 (conjugated with the VIC fluorophore), as well as several control probes (conjugated with the 6FAM fluorophore). Controls were designed against the L1 5′ UTR and other non-mobile DNA sequences in the genome that have copy numbers that are higher (such as α satellite[19]) or lower (such as HERVH and the 5S ribosomal DNA (rDNA) gene) than ORF2. Furthermore, because most L1 retrotransposition events are 5′ truncated[9,20,21], we reasoned that the L1 5′ UTR probes should detect a smaller copy number increase than the L1 ORF2 probes. Each probe set amplified a single product of the predicted size (Supplementary Fig. 10b). Moreover, sequencing PCR products derived from both ORF2 probe sets showed enrichment for members of the L1Hs subfamily (Supplementary Table 3).

We next isolated genomic DNA from the hippocampus, cerebellum, liver and heart from three adult humans. We consistently observed a statistically significant increase in L1 ORF2 content in the hippocampus when compared to heart and liver samples from the same individual (Fig. 4b, c and Supplementary Figs 9a and 10a). Notably, two individuals (1079 and 1846) showed more marked copy number differences

Figure 3 | Methylation analysis and ChIP for the endogenous human L1 5′ UTR. **a**, Schematic illustrating the L1 CpG island, and SRY-binding sites. **b**, Cumulative distribution function plot, comparing overall methylation and collapsing CpG sites into a single data point (two-sample Kolmogorov–Smirnov test). **c**, Individual methylation of sequences showing highest sequence similarity to consensus RC-L1s. Open and closed circles denote unmethylated and methylated CpG dinucleotides, respectively. Dash indicates mutated CpG site. **d**, ChIP identifying MECP2 and SOX2 occupying the endogenous human L1 promoter, extracts were analysed by PCR towards the L1 5′ UTR SRY-binding region (SOX2 immunoprecipitation) or CpG island region (MECP2 immunoprecipitation). **e**, CpG dinucleotides exhibited higher methylation at the 5′-end of the CpG island; higher methylation overall was observed in skin samples.

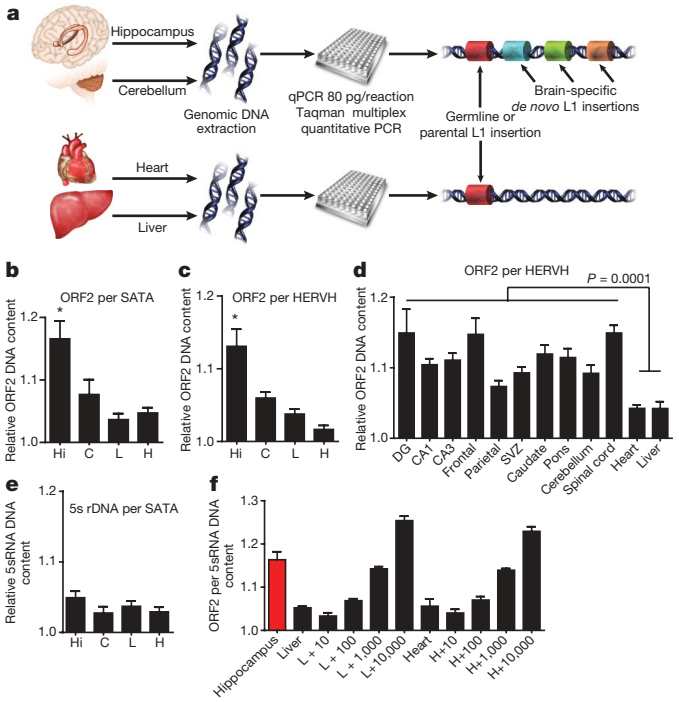


Figure 4 | Multiplex quantitative PCR analyses of L1 copy number in human tissues. **a**, Experimental schematic. **b**, **c**, Relative quantity of L1, standardized such that the lowest liver value was normalized to 1.0. C, cerebellum; H, heart; Hi, hippocampus; L, liver. Further L1 ORF2 assays with other internals controls are shown in Supplementary Figs 9 and 10. All error bars are s.e.m. $*P < 0.05$ (repeated measures one-way analysis of variance (ANOVA) with Bonferroni correction, $n = 3$ individuals, with three repeat samples from each tissue). SATA, α satellite. **d**, Ten samples from various brain regions ($n = 3$ individuals) compared to somatic liver and heart. DG, dentate gyrus; SVZ, subventricular zone. One-way *t*-test, $P \leq 0.0001$ with 34 degrees of freedom. **e**, Multiplexing of 5S rDNA with α-satellite indicated no significant change, $P \leq 0.5054$. **f**, Hippocampal tissue compared to liver (L) and heart (H) spiked with estimated plasmid copy numbers of L1 (10, 100, 1,000 and 10,000 copies).

than a third individual (4590) (Supplementary Fig. 10a). Controls demonstrated that the ratio of the 5S rDNA gene to α satellite DNA between each tissue remained relatively constant (Fig. 4e).

We extended this analysis to ten brain regions from three other individuals (Fig. 4d and Supplementary Fig. 9b). The samples were derived from the frontal and parietal cortex, spinal cord, caudate, CA1 and CA3 areas of the hippocampus and pons, as well as from the hippocampal dentate gyrus and the subventricular zone[22]. As described earlier, there was marked variation between different brain areas and between individuals (Supplementary Fig. 9c). However, an unpaired *t*-test comparing all the grouped brain samples to the heart and liver DNA again showed a small, but statistically significant increase in ORF2 content in the brain (Fig. 4d).

To independently corroborate the observed increase in L1 copy number in the hippocampus and cerebellum samples, we spiked 80 pg of liver and heart genomic DNA (approximately 12 genomes) from individual 1846 with a calculated quantity of L1 plasmid, then we repeated the multiplexing approach to assay ORF2 quantity relative to the 5S rDNA internal control (Fig. 4f). Three replications of this experiment indicated that the hippocampus samples contained approximately 1,000 more L1 copies than the heart or liver genomic DNAs, suggesting a theoretical increase in ORF2 of approximately 80 copies per cell. The spiked L1 copies were in the form of a plasmid, which probably affects the copy number estimates, providing an estimate of relative change and not precise quantification of the absolute number of L1s per cell. Therefore, ultimate proof that endogenous L1s are retrotransposing in the brain requires identification of new retrotransposition events in individual somatic cells.

The large degree of variability in L1-ORF2 copy numbers between brain regions and individuals may represent unsystematic rates of L1 retrotransposition, or another level of regulation that requires further determination. That being stated, our *in vitro* findings in NPCs coupled with the observed L1-ORF2 copy number changes in the brain make it tempting to speculate that somatic retrotransposition events occur during the early stages of human nervous system development. This study contributes to a growing body of evidence indicating that engineered L1s can retrotranspose during early development, and in select somatic cells[1,6,23–25]. Future experiments will determine whether endogenous L1s truly retrotranspose in the brain and whether these events are simply 'genomic noise' or have the potential to affect neurogenesis and/or neuronal function.

METHODS SUMMARY

Cell culture, transfection and analysis. Fetal hCNS-SCns lines[3] and hESCs[24,26] were cultured as previously described. Neural progenitors were derived from hESCs as previously described[14,27]. NPCs were transfected by nucleofection (Amaxa Biosystems), and either maintained as progenitors in the presence of FGF2 or differentiated as previously described[14]. Cells were transfected with L1s containing an EGFP retrotransposition cassette in pCEP4 (Invitrogen) that lacks the CMV promoter and contains a puromycin-resistance gene[7].

Cell lysates. Ribonucleoprotein particles were isolated and analysed as previously described[8]. Luciferase assays were performed as previously described[1]. ChIP was performed using primers towards the L1 5′ UTR and a ChIP assay kit (Upstate/Millipore) as per the manufacturer's protocol.

Bisulphite analysis. Fetal tissues were obtained from the Birth Defects Research Laboratory at the University of Washington. Bisulphite conversions were performed by manufacturer's instructions using the Epitect kit (Qiagen). BLASTN (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to align sequences to a database of full-length L1s.

PCR. Adult human tissues were obtained from the NICDH Brain and Tissue Bank for Developmental Disorders (University of Maryland). Taqman probes and primers were designed using L1 Base (http://l1base.molgen.mpg.de/) and copy number estimates were based on the UCSC genome browser (http://genome.ucsc.edu). Experiments were performed on an ABI Prism 7000 sequence detection system (Applied Biosystems). For each tissue, three separate tissue samples were extracted and considered as repeated measures. Whole genome size was estimated based on the equation: cell genomic DNA content $= 3 \times 10^9$(bp) $\times$ 2(diploid) $\times$ 660(molecular mass of 1 bp) $\times 1.67 \times 10^{12}$ (weight of 1 Da), (in which bp denote base pairs), resulting in the approximation that one cell contains 6.6 pg genomic DNA[28]. Therefore, the 80 pg of genomic DNA used per reaction is derived from approximately 12 cells. Inverse PCR was performed as previously described[1,24].

Received 31 March; accepted 1 July 2009.
Published online 5 August 2009.

1. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435,** 903–910 (2005).
2. Tang, Y., Nyengaard, J. R., De Groot, D. M. & Gundersen, H. J. Total regional and global number of synapses in the human brain neocortex. *Synapse* **41,** 258–273 (2001).
3. Uchida, N. *et al.* Direct isolation of human central nervous system stem cells. *Proc. Natl Acad. Sci. USA* **97,** 14720–14725 (2000).
4. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl Acad. Sci. USA* **100,** 5280–5285 (2003).
5. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87,** 917–927 (1996).
6. Ostertag, E. M. *et al.* A mouse model of human L1 retrotransposition. *Nature Genet.* **32,** 655–660 (2002).
7. Ostertag, E. M., Prak, E. T., DeBerardinis, R. J., Moran, J. V. & Kazazian, H. H. Jr. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **28,** 1418–1423 (2000).
8. Kulpa, D. A. & Moran, J. V. Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum. Mol. Genet.* **14,** 3237–3248 (2005).
9. Moran, J. & Gilbert, N. *Mammalian LINE-1 Retrotransposons and Related Elements* (ASM Press, 2002).
10. Myers, J. S. *et al.* A comprehensive analysis of recently integrated human Ta L1 elements. *Am. J. Hum. Genet.* **71,** 312–326 (2002).
11. Morrish, T. A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature Genet.* **31,** 159–165 (2002).
12. Gilbert, N., Lutz, S., Morrish, T. A. & Moran, J. V. Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol. Cell. Biol.* **25,** 7780–7795 (2005).
13. Symer, D. E. *et al.* Human L1 retrotransposition is associated with genetic instability *in vivo*. *Cell* **110,** 327–338 (2002).

14. Yeo, G. W. *et al.* Alternative splicing events identified in human embryonic stem cells and neural progenitors. *PLoS Comput. Biol.* **3,** e196 (2007).
15. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431,** 96–99 (2004).
16. Takai, D. & Jones, P. A. The CpG island searcher: a new WWW resource. *In Silico Biol.* **3,** 235–240 (2003).
17. Yu, F., Zingler, N., Schumann, G. & Stratling, W. H. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res.* **29,** 4493–4501 (2001).
18. Tchenio, T., Casella, J. F. & Heidmann, T. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* **28,** 411–415 (2000).
19. Lee, C., Wevrick, R., Fisher, R. B., Ferguson-Smith, M. A. & Lin, C. C. Human centromeric DNAs. *Hum. Genet.* **100,** 291–304 (1997).
20. Pavlicek, A., Paces, J., Zika, R. & Hejnar, J. Length distribution of long interspersed nucleotide elements (LINEs) and processed pseudogenes of human endogenous retroviruses: implications for retrotransposition and pseudogene detection. *Gene* **300,** 189–194 (2002).
21. Grimaldi, G., Skowronski, J. & Singer, M. F. Defining the beginning and end of KpnI family segments. *EMBO J.* **3,** 1753–1759 (1984).
22. Gage, F. H. Mammalian neural stem cells. *Science* **287,** 1433–1438 (2000).
23. Prak, E. T., Dodson, A. W., Farkash, E. A. & Kazazian, H. H. Jr. Tracking an embryonic L1 retrotransposition event. *Proc. Natl Acad. Sci. USA* **100,** 1832–1837 (2003).
24. Garcia-Perez, J. L. *et al.* LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.* **16,** 1569–1577 (2007).
25. van den Hurk, J. A. *et al.* L1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.* **16,** 1587–1592 (2007).
26. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282,** 1145–1147 (1998).
27. Zhang, S. C., Wernig, M., Duncan, I. D., Brustle, O. & Thomson, J. A. *In vitro* differentiation of transplantable neural precursors from human embryonic stem cells. *Nature Biotechnol.* **19,** 1129–1133 (2001).
28. Forslund, O. *et al.* Nucleotide sequence and phylogenetic classification of candidate human papilloma virus type 92. *Virology* **312,** 255–260 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Simon for excellent schematic drawings, M. L. Gage, J. Kim and H. Kopera for editorial comments, B. Miller and R. Keithley for cell culture assistance, C. T. Carson for hESC advice, D. Chambers and J. Barrie for flow cytometry assistance, L. Randolph-Moore for molecular advice, B. Aimone for statistics advice, T. Liang for microarray assistance, and Y. Lineu and J. Mosher for helpful comments. We also thank T. Fanning and M. Klymkowsky for the ORF1 protein and SOX3 antibodies, respectively. F.H.G. and N.G.C. are supported by the Picower Foundation, G. Harold and Leila Y. Mathers Charitable Foundation, Lookout Fund (MH082070), and the California Institute for Regenerative Medicine (CIRM). J.L.G.-P. is supported by Plan Estabilizacion Grupos SNS ENCYT 2015 (EMER07/56, Instituto de Salud Carlos III, Spain) and through the IRG-FP7-PEOPLE-2007 Marie Curie program. K.S.O. was supported by grants GM069985 and NS048187 from the National Institutes of Health (NIH). J.V.M. was supported by grants GM082970 and GM069985 from the NIH and by the Howard Hughes Medical Institute. Work in the laboratories of K.S.O. and J.V.M. only used NIH-approved stem cell lines.

Author Contributions N.G.C. and F.H.G. directed the project. J.V.M. and J.L.G.-P. directed aspects of the project conducted at Michigan. N.G.C., J.L.G.-P., J.V.M. and F.H.G. designed experiments and drafted the manuscript. N.G.C., F.H.G., J.L.G-P. and G.E.P. performed the experiments. G.W.Y. and M.T.L. carried out bioinformatics data analysis. Y.M. performed electrophysiology experiments. M.M. and K.S.O. provided hESC culture and NPC differentiation assistance. All authors commented on or contributed to the current manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to F.H.G. (gage@salk.edu).