Supplemental Figure and Table Legends

Figure S1: Rationale for the LINE-1 retrotransposition assay. **a,** A cartoon of an RC-L1 is shown at the top of the figure. The dark blue rectangle represents the 5´UTR. The yellow and blue arrows represent ORF1 and ORF2, respectively. The relative positions of the endonuclease (EN), reverse transcriptase (RT) and C-domain (C) in L1 ORF2 are indicated. The 3' UTR of the L1 was tagged with a retrotransposition indicator cassette, which consists of a reporter gene in the reverse orientation (REP, gray arrow) containing its own promoter and polyadenylation signal. The reporter gene is also interrupted by an intron in the same transcriptional orientation as the RC-L1 (IVS2, black rectangle). This arrangement ensures that the reporter cassette will only be activated and expressed (gray oval) if the spliced RC-L1 mRNA undergoes a successful round of retrotransposition. **b,** Several retrotransposition markers (EGFP, top left; NEO, top right; BLAST, below) are useful for studying retrotransposition. Each scheme also indicates the relative position of the primers used to confirm splicing of the intron from the retrotransposition indicator cassette (red arrows). A schematic showing the anticipated results of the PCR-intron removal assay is shown at the right of each figure.

Figure S2: Characterization of L1 retrotransposition events in hCNS-SCns. **a-b,** L1-EGFP-positive hCNS-SCns express the neural stem cell markers Musashi1 (nuclear) and SOX1 (nuclear) as well as nestin (cytoplasmic). DAPI, nuclear stain. **c,** L1-EGFP-positive hCNS-SCns are still capable of cell division (Ki-67-positive; white, nuclear); Ki-67-positive cells also express the cytoplasmic progenitor marker Nestin. In all images, arrows indicate co-labeled cells. **d,** Bright-field images of primary astrocytes and fibroblasts. **e,** FACS analyses of hCNS-SCns cells (FBR4, see methods) transfected with $L1_{RP}$ results in a low, but reproducible rate of L1 retrotransposition. Retrotransposition events were not observed in hCNS-SCns transfected with JM111/$L1_{RP}$ or in primary human astrocytes or fibroblasts transfected with $L1_{RP}$.

Figure S3: Retrotransposition in hESC-derived NPCs. **a,** The schematic at the top of the figures outlines the NPC differentiation protocol used for the H7, H9, H13B and BG01 hESCs. The number of days for each step of differentiation is indicated above the arrows. The medium used in the derivation is indicated below the arrow. Bright-field images of representative cells at each stage of the derivation are shown below the graphic. **b,** Dissociated neurospheres express the nuclear neural stem cell markers SOX1 and SOX3. **c,** L1-EGFP-positive, HUES6-derived NPCs express SOX1 (nuclear) and Nestin (cytoplasmic). **d,** H13B-derived NPCs support L1-EGFP retrotransposition and express SOX3 (nuclear). **e,** L1-EGFP-positive, HUES6-derived neurons co-label for the neuronal markers βIII tubulin and Map2a+2b (both cytoplasmic). **f,** L1-EGFP-positive, HUES6-derived NPCs can differentiate to a glial lineage that is positive for the cytoplasmic marker GFAP but negative for the neuronal marker βIII tubulin. Arrows indicate co-labeled cells; arrowheads indicate cellular processes that are co-labeled.

Figure S4: Characterization of L1 retrotransposition in hESC-derived NPCs. **a,** FACS analysis of HUES6-derived NPCs transfected with either $L1_{RP}$ or JM111/$L1_{RP}$. **b,** The synapsin promoter is strongly induced upon NPC differentiation. X axis, differentiation time course (days post-differentiation); Y axis, luciferase activity (fold activity). **c-d,** Transfection of both HUES6- (C) and H7-(D) derived NPCs with the engineered LRE3-*mneoI* construct indicates that G418-resistant colonies contain a retrotransposition event, lacking the intron from the indicator cassette. **e,** The L1.3 *mblastI* construct also retrotransposes in HUES6-derived NPCs. **f-g,** LRE3 *mEGFPI* also retrotransposes in both H13B-(F) HUES6- (G) derived NPCs. Molecular size standards are shown at the right of the gel images in panels C-G. **h,** Western blot for SOX2 and MeCP2 indicates expression of SOX2 decreases with neural differentiation, whereas MeCP2 expression is upregulated.

Figure S5: NPCs exhibit a grossly normal karyotype. **a-c,** The three hCNS-SCns cell lines have a normal karyotype. **d,** FISH (fluorescence *in situ* hybridization) using a probe cocktail specifically designed to identify small populations of cells with changes in chromosome 12 and 17 copy number, a common karyotypic abnormality observed in the culturing of hESC[29], revealed that HUES-6 cells demonstrated a normal signal pattern for the ETV6 BAP (TEL) gene located on chromosome 12. All cells also demonstrated a normal signal pattern for the chromosome 17 centromere. Two hundred interphase nuclei were examined using this procedure. In sum, we did not detect any evidence of trisomy 12 and/or trisomy 17. **e-f,** The HUES6 (E) and H9 ES (F) hESCs exhibit a grossly normal karyotype.

Figure S6: Quantification of L1 RNA transcripts **a,** Quantitative RT-PCR analysis of L1 ORF2 transcripts in *in vitro* cell types, standardized to actin. **b,** RT-PCR analysis of L1 ORF1 transcripts, with GAPDH as a loading control. **c,** Quantitative RT-PCR analysis of L1 ORF2 transcripts from fetal brain, skin, and liver, n=3 individuals. **d,** RT-PCR analysis of L1 ORF1 transcripts from the same tissues.

Figure S7: Analyses of the EGFP-positive and EGFP-negative FACS-sorted NPC populations. **a,** PCR on genomic DNA from EGFP-positive and EGFP-negative cell populations revealed L1 retrotransposition events (342 bp product) in the EGFP-positive cells and little of the original LRE3 expression construct (1,243 bp product) in either sample. **b,** Characterization of retrotransposition events in hESC-derived NPCs revealed structural hallmarks of LINE-1 retrotransposition. The caricature represents a fully characterized engineered L1 retrotransposition LRE3 event in HUES6-derived NPCs (see Supplemental Table 2). The schematic shows the sequence of the pre-integration (bottom) and post-integration (top) sites. Also shown are the nucleotide position of the truncation site within L1 (in this example, truncation occurred in the EGFP cassette), the approximate length of the poly (A) tail, target-sited duplications that flank the

retrotransposed L1, and the endonuclease recognition site. **c,** Both EGFP-positive and EGFP-negative sorted HUES6-derived NPC populations expressed the neural stem cell markers Nestin and SOX2. **d,** Both EGFP-positive and EGFP-negative HUES6 derived NPC populations could differentiate to cells of both the neuronal (βIII tubulin) and glial (GFAP) lineages.

Figure S8: Methylation analysis of the human L1 5' UTR. **a-b,** The X axis shows the sequence identity (percent) of each L1 5' UTR analyzed from the brain and skin samples as compared to the database of RC-L1 5' UTR sequences. The cutoff for analysis was the mean sequence identity to the RC-L1 database minus one standard deviation. The Y-axis shows the percentage of unmethylated CpG dinucleotides in each sample. **c,** The conversion of isolated cytosine residues that were not part of a CpG dinucleotide was used to measure the efficiency of the bisulfite conversion reaction. We obtained a conversion efficiency of >90% for all analyzed sequences, with no statistically significant difference between samples (left, D80 female, right, D82 male). **d,** The dinucleotide sequences of the L1 5' UTRs from the brain and skin samples were compared to one another. The only statistically significant difference between brain and skin samples was in the conversion of CpG to TpG dinucleotides, indicating a lesser degree of L1 methylation in the brain samples when compared to the skin samples. Statistically significant changes were not observed in the first base of any other dinucleotide sequence between the two samples, indicating there was no statistically significant sampling bias of different L1 subtypes between data sets.

Figure S9: Multiplex qPCR data from human brain areas and somatic tissues**. a,** The ratio of ORF2/internal control represents the amount of L1 ORF2 DNA sequence in each sample relative to the amount of L1 5' UTR, standardized such that the lowest liver value is normalized to 1.0 and all other samples are reported relative to the lowest liver value. Hi = Hippocampus, C = Cerebellum, H = Heart, and L = Liver. Under these conditions, the copy numbers of L1 ORF2 sequences were higher in the hippocampus and, to a lesser extent, the cerebellum when

compared to the heart and liver samples. Graphs were obtained by grouping data from different individuals in Fig. S10A.*$p<0.05$ as a result of a repeated measures one-way ANOVA with a Bonferroni correction (n=3 individuals, with 3 repeat samples from each tissue). **b,** Ten additional samples from various brain regions (n=3 individuals) were compared to somatic liver and heart samples (ORF2/5S rDNA). An unpaired t-test comparing grouped brain samples with the somatic tissues, $p \leq 0.002$ with degrees of freedom = 34. **c,** Data from the three individuals were combined for each tissue into a single data point to generate the data in Fig. 4D (Error bars = SEM from 3 different tissue samples).

Figure S10: Multiplex qPCR data from hippocampus, cerebellum, liver and heart DNAs isolated from three individuals. **a,** Data from the three individuals were combined for each tissue into a single data point to generate the data in Fig. 4B-C and Fig. S9A (n=3 for each tissue in A, n=9 collapsed across the three individuals in Fig. 4). Notably, in two individuals (1079 and 1846), the difference in L1 copy number was evident in multiple multiplex PCR reactions; however, the increase in L1 copy number was more modest in a third individual (4590). **b,** Each primer set amplified only a single PCR product, tested on a single hippocampal tissue, 60 cycles of PCR.

Figure S11: Euclidian distance map based on exon-splicing array data[14]. **a,** Each cell type was assayed in triplicate and compared to duplicates of human fetal brain standardized RNA (Ambion/Applied Biosystems). In all cases the replicates clustered well together. Fetally derived NPCs clustered closer to HUES6 cells, whereas the HUES6-derived NPCs cluster closer to fetal brain. **b,** Promoter analysis of the L1.3 5' untranslated region indicated two SRY/SOX2 binding sites that were assayed in ChIP experiments. Analysis of a scrambled L1 5' UTR is included on the right.

Table S1: Result of L1 retrotransposition assays in hESC-derived NPCs. From left to right, column 1 indicates the hESCs cell line from which NPCs were

derived, column 2 indicates the lab where the experiments were performed, column 3 indicates if selection (puromycin 0.2 μg/mL) was used in the assay, and column 4 indicates the percentage of EGFP-expressing cells with s.d. We observed that L1 retrotransposition events could be efficiently silenced in some hESC-derived NPCs (column 1, marked H13B*). This silencing could be overcome by treating the cells with histone deacetylase inhibitors (see Supplemental methods for experimental details; Garcia-Perez and Moran, unpublished data), and it may reflect idiosyncrasies that arise during the differentiation protocol.

Table S2: Analysis of L1 insertions in hESC-derived NPCs. From left to right: column 1: if the insertion was characterized from a clone or from FACS-sorted cells (derivations 1 and 2 are from separate NPC derivations, and separate transfections of L1); column 2: if the insertion characterization was full or partial; column 3: the truncation site of the retrotransposed tagged L1; column 4: the estimated length of the poly (A) tail; column 5: the sequence of the actual or inferred LINE-1 endonuclease bottom strand cleavage site; column 6: the chromosomal locus of the insertion; column 7: the insertion target site of the tagged L1. Note that eight insertions were characterized completely; however, only the 3' end was characterized of the remaining insertions because the restriction enzyme utilized in the ligation step of the inverse PCR protocol was also present in the retrotransposed L1 sequence.

Table S3: Sequencing analysis of QPCR genomic DNA products. PCR products from both ORF2#1 and ORF2#2 primer sets (column b) were cloned and sequenced from PCR reactions run with both hippocampus and liver genomic DNA (column c). Percentage sequence identity to an RC-L1 consensus sequence (column d) was determined. Mismatches to a known RC-L1 are indicated in red (column e). Sequence analysis using the UCSC genome browser and Repeatmasker indicates that the majority of amplified sequences belong to the L1Hs subfamily of elements (column f and g). Notably, due to their short

length, some amplicons could not be definitively assigned to a single L1 subfamily.

Table S4: Sequencing analysis of QPCR products from L1 RT-PCR. Quantitative RT-PCR products from ORF2 #1 primer sets were cloned from three sample types: fetal brain, hCNS-SCns, and HUES6-derived NPCs (column b). Percentage sequence identity to an RC-L1 consensus (column c) was determined, and sequence analysis using UCSC genome browser and Repeatmasker indicated that most sequences belonged to the L1Hs subfamily of elements (column e and f, respectively). Complete sequence of the QPCR product is indicated in column D.

Table S5: Sequencing of actively transcribed ORF1 fragments from RT-PCR. RT-PCR fragments from Fig. S6 were cloned and sequenced from three samples: fetal brain, hCNS-SCns, and HUES6-derived NPCs (column b). Percentage sequence identity to an active RC-L1 (L1.3) (column c) was determined, as well as sequence analysis using the UCSC genome browser and Repeatmasker (columns e and f). In addition, since these are larger fragments than those resulting from QPCR, most mapped to a unique genomic location (column g). Complete sequence of the RT-PCR product is indicated in column D.

Methods References

[29] Draper, J.S. *et al.*, Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat Biotechnol* 22 (1), 53-54 (2004).

[30] Zhao, C., Teng, E.M., Summers, R.G., Jr., Ming, G.L., & Gage, F.H., Distinct morphological stages of dentate granule neuron maturation in the adult mouse hippocampus. *J Neurosci* 26 (1), 3-11 (2006).

[31] Kimberland, M.L. *et al.*, Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Hum Mol Genet* 8 (8), 1557-1560 (1999).

[32] Brouha, B. *et al.*, Evidence consistent with human L1 retrotransposition in maternal meiosis I. *Am J Hum Genet* 71 (2), 327-336 (2002).

[33] Gage, F.H. *et al.*, Survival and differentiation of adult neuronal progenitor cells transplanted to the adult brain. *Proc Natl Acad Sci U S A* 92 (25), 11879-11883 (1995).

[34] Athanikar, J.N., Badge, R.M., & Moran, J.V., A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res* 32 (13), 3846-3855 (2004).

[35] Sambrook, J., Fritsch, E., & Maniatis, A., *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Press, Cold Spring Harbor, 1989).

[36] Kent, W.J. *et al.*, The human genome browser at UCSC. *Genome Res* 12 (6), 996-1006 (2002).