Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells

Sabine Klawitter1,*, Nina V. Fuchs1,2,*, Kyle R. Upton3,*, Martin Muñoz-Lopez4,*, Ruchi Shukla3, Jichang Wang2, Marta García-Cañadas4, Cesar Lopez-Ruiz4, Daniel J. Gerhardt3, Attila Sebe1, Ivana Grabundzija2, Sylvia Merkert5, Patricia Gerdes3, J. Andres Pulgarin4, Anja Bock1, Ulrike Held1, Anett Wittuhhn5, Alexandra Haase5, Balázs Sarkadi6, Johannes Löwer1, Ernst J. Woldteng7, Ulrich Martin5, Zoltán Ivics1, Zsuzsanna Izsvák2, Jose L. Garcia-Perez4, Geoffrey J. Faulkner3,8 & Gerald G. Schumann1

Human induced pluripotent stem cells (hiPSCs) are capable of unlimited proliferation and can differentiate in vitro to generate derivatives of the three primary germ layers. Genetic and epigenetic abnormalities have been reported by Wissing and colleagues to occur during hiPSC derivation, including mobilization of engineered LINE-1 (L1) retrotransposons. However, incidence and functional impact of endogenous retrotransposition in hiPSCs are yet to be established. Here we apply retrotransposon capture sequencing to eight hiPSC lines and three human embryonic stem cell (hESC) lines, revealing endogenous L1, Alu and SINE-VNTR-Alu (SVA) mobilization during reprogramming and pluripotent stem cell cultivation. Surprisingly, 4/7 de novo L1 insertions are full length and 6/11 retrotransposition events occurred in protein-coding genes expressed in pluripotent stem cells. We further demonstrate that an intronic L1 insertion in the CADPS2 gene is acquired during hiPSC cultivation and disrupts CADPS2 expression. These experiments elucidate endogenous retrotransposition, and its potential consequences, in hiPSCs and hESCs.
Transcription and translation of functional L1 elements are prerequisites for L1-mediated retrotransposition. To confirm that reports of pronounced L1 expression in hiPSCs by Wissing et al. could be extended to the hiPSC lines used in our study, we measured L1 mRNA abundance, L1 promoter methylation status and L1 ORF1 protein (ORF1p) expression in fibroblast (HFF-1-) and hCBE-derived hiPSCs (hiFF-iPS4, hiPS-SB4, hiPS-SB5, hCBiPS1 and hCBiPS2) and their parental cells (Table 1; Fig. 1a–d). TaqMan qRT–PCR (quantitative PCR with reverse transcription) targeting the L1 5′UTR (Fig. 1a; Supplementary Table 1) revealed significantly elevated L1 mRNA levels in each hiPSC line relative to their parental cells ($P < 0.05–0.0001$, analysis of variance (ANOVA)), that peaked in earlier passages of cell lines hiPS-SB4 and hiPS-SB5 (Fig. 1b,c). Northern blot analyses with an L1 5′UTR-specific probe (Fig. 1a) confirmed elevated expression of full-length L1 transcripts in hiPSCs (Fig. 1c). Notably, extended hiPSC culture led to reduced L1 mRNA abundance (Fig. 1b, left panel; hiPS-SB4, hiPS-SB5; $P < 0.05–0.001$, ANOVA) and resembled levels observed in hESCs (HES-3, Fig. 1b). Bisulfite DNA sequencing of the CpG island present in the canonical L1 promoter revealed strong hypomethylation in all tested hiPSC lines compared to parental cells ($P_{12} < 2 \times 10^{-6}$, Fig. 1d; $P_1 = 2.6 \times 10^{-12}$, $P_2 = 1.8 \times 10^{-3}$, Supplementary Fig. 3; $\chi^2$ test). Consistently, L1 ORF1p was abundant in hiPSCs, based on immunoblot (Fig. 1e) and immunofluorescence assays (Fig. 2; Supplementary Fig. 4). In agreement with previous reports of cytoplasmic L1 ORF1p expression in human tumours and cancer cell lines, in hiPSCs, we found L1 ORF1p predominantly expressed in cytoplasmic foci (Fig. 2b). However, unlike recent studies focused on other cell types, we did not resolve whether L1 ORF1p was directed to stress granules in hiPSCs. Finally, quantitative immunoblot analyses (Supplementary Methods) revealed a tenfold increase in L1 ORF1p expression in hiPSCs when compared with parental cells (Supplementary Fig. 5).

Taken together, our results revealed a spike in L1 expression during or immediately after reprogramming, confirming previous findings, followed by attenuation in later hiPSC passages (Fig. 1b,c). To extend these results, we measured L1 mRNA levels upon differentiation of late passage hiPSCs (hiPS-SB4 (p98) and hiFF-iPS4 (p50)) into embryoid bodies. We observed 49% and 58% reductions in L1 mRNA levels after 1 and 10 days of embryoid body differentiation, respectively (Fig. 1b, middle panel). A parallel assay conducted with early passage hiPSCs indicated a gradual and significant decrease of L1 mRNA abundance by up to 65% after 8 days of embryoid body differentiation and a concomitant increase in differentiation markers (Fig. 1b, right panel; Supplementary Fig. 6). Hence, elevated L1 expression in hiPSCs was triggered by reprogramming and attenuated by short-term cultivation, while, in turn, subsequent differentiation gradually reduced L1 expression.

Endogenous retrotransposition in pluripotent stem cells. To unambiguously determine whether activation of the L1 mobilization machinery produced L1-mediated retrotransposition, we used retrotransposon capture sequencing (RC-seq) to map the genomic integration sites of de novo retrotransposition insertions. Briefly, RC-seq involved liquid phase sequence capture to enrich DNA for the 5′ and 3′ junctions of recent L1, Alu and SVA insertions and the surrounding genome. Putatively immobile long terminal repeat (LTR) retrotransposons were also probed as negative controls. Multiplexed, paired-end 150mer Illumina sequencing of RC-seq libraries, followed by contig assembly,
Figure 1 | Reprogramming-induced expression of the L1 retrotransposition machinery is abrogated during embryoid body formation. (a) Schematic of organization and expression of a functional human L1 element. Binding sites of TaqMan primer/probe combinations (small convergent arrows) on L1 cDNA used for qRT-PCR analyses and of the 1,299-bp [α-32P]dCTP-labelled PCR product in the 5′UTR region (black bar) used for northern analysis are shown. Methylation status of the CpG island (position number 232–491 of the L1.3 reference sequence) was analysed. Open circles, CpG residues. (b) Relative full-length L1 (FL-L1) mRNA transcript levels were assessed by qRT-PCR from early passage (until p24) HFF-1-derived (hFF-iPS4, hiPS-SB4 and hiPS-SB5) and hCBEC-derived (hCBiPS1 and hCBiPS2) hiPSC lines (left panel), and after differentiation of hFF-iPS4 (p50) and hiPS-SB4 (p98) lines into embryoid bodies (EBs) (middle panel) (*P<0.05, **P<0.01, ***P<0.001). hiPS-SB5.1 cells (p10) were differentiated into EBs. L1 transcript levels were quantified on day 0 before initiation of differentiation, and after 2, 4, 6 and 8 days of differentiation by qRT–PCR (right panel; ***P<0.001, linear regression t-test). Bars represent arithmetic means ± s.d. from experiments performed as technical duplicates of biological triplicates, or, in the case of hCBEC, hCBiPS1 and hCBiPS2 (green bars), arithmetic means of technical duplicates of one biological sample. (c) Northern analysis of cytoplasmic poly-A+ mRNA with a 1,299-bp L1 5′UTR-specific probe confirmed exceeding activation of FL-L1 transcription during hiPSC cultivation. β-Actin mRNA (1.8 kb, lower panel) served as loading control. (d) Endogenous L1 promoter sequences are significantly hypomethylated in hiPSC lines relative to their parental HFF-1 and hCBEC cells. Overall percentage methylation of 5′UTR CpG islands in HFF-1 and hCBEC cells (n = 29 CpG islands; blue bar) and in five derived hiPSC lines (n = 95 CpG islands; red bar), respectively, is presented. Error bars indicate s.e.m. ***P<0.001; χ² test. (e) Immunoblot analysis of cell lysates from HFF-1 and hCBEC cells and their respective derived hiPSC lines measures L1 ORF1p (40 kDa) and Oct-3/4 expression (A isoform, 45 kDa; B isoform 33 kDa). Shorter (exp.1) and longer exposures (exp.2) of the α-Actin immunoblot are provided. Lysates from hESC lines HES-3 (left panel) and H1 (right panel) served as positive control for L1 ORF1p and Oct-3/4 expression. β-Actin (42 kDa) served as loading control.
provided high-fidelity, single nucleotide resolution of insertions absent from the reference genome, even at low read depth. We analysed all eight hiPSC lines and their matched parental cells by RC-seq. For five fibroblast-derived hiPSC lines (Table 1), we included two separate passages each to detect mobilization events that may have accumulated during cell culture. Similarly, we analysed two passages each of three hESC lines to evaluate endogenous retrotransposition during hESC cultivation (Table 1).

Table 1 | Analysed pluripotent stem cell lines and their characteristics.

| Stem cell line | Parental cells | Reprogramming factors | Factor delivery by | Reference | Passages (p) assayed by RC-seq |
|----------------|----------------|-----------------------|--------------------|-----------|-------------------------------|
| hiPS-SB4       | HFF-1          | OCT-4, SOX2, KLF4 and c-MYC | Sleeping Beauty transposon | 26        | 43, 53                        |
| hiPS-SB5       |               | OCT-4, SOX2, KLF4, c-MYC and LIN28 |               |           | 40, 59                        |
| hiPS-CRL1502   | CRL1502        | OCT-4, SOX2, NANOG, LIN28, KLF4 and c-MYC | oriP/EBNA1-based pCEP4 episomal vector | 25        | 15, 40                        |
| hiPS-CRL2429   | CRL2429        | Dermal fibroblasts (male) |               |           | 11, 40                        |
| hiPS-FB        | FB             | OCT-4, SOX2, KLF4 and c-MYC | Lentiviral vector | 68        | 7, 23                         |
| hFF-iPS4       | HFF-1          | OCT-4, SOX2, NANOG and LIN28 | Unpublished | 58        |                               |
| hCBIPS1        | hCBEC          | OCT-4, SOX2, NANOG and LIN28 |               | 27        | 30, 23                        |
| hCBIPS2        | hESC lines     | NA                     | NA                 |           | 73, 74, 92, 102, 30, 60, Unpublished 10, 23 |

Figure 2 | Immunofluorescence staining of hiPSC colonies and their parental cells for endogenous L1 ORF1p expression in HFF-1-derived hiPSCs. (a) ORF1p staining indicates activation of endogenous L1 expression after reprogramming of HFF-1 cells into lines hiPS-SB4, hiPS-SB5 and hFF-iPS4. Cells were analysed at passages (p) 23, 16 and 60, respectively. Oct-3/4 staining confirmed the pluripotent status of the analysed stem cell colonies. Mesenchymal stem cell marker CD105 (endoglin) is reported to be expressed in HFF-1 cells but not expressed in pluripotent stem cells. (b) Enlarged areas indicated by boxed dashes in a demonstrate cytoplasmic localization of endogenous L1 ORF1p and its accumulation in foci. Scale bars, 20 μm.
RC-seq detected a total number of 40,608 non-reference retrotransposon insertions including on average 214 L1, 1,411 Alu, 53 SVA and 14 LTR non-reference genome insertions per hiPSC and hESC sample (Supplementary Fig. 7; Supplementary Data 1). Insertions were annotated as de novo in pluripotent cells if they were not (i) reported previously in non-reference retrotransposon insertion databases9,12,33–38, (ii) found in parental cells, (iii) found in an earlier hESC passage or (iv) found in multiple hiPSC or hESC lines. In total, we detected eight L1, seven Alu and two SVA putative de novo insertions (Supplementary Data 1). We found no de novo LTR retrotransposon insertions, despite observing profound upregulation of HERV-K group HML-2 transcription in hiPSCs and hESCs (Supplementary Methods; Supplementary Fig. 8).

Five retrotransposon subfamilies (L1-Ta, L1-pre-Ta, AluYa5 and SVA3) known to be active in humans contributed putative de novo insertions10,11,39. These were first validated by genotyping PCR, with seven L1, two Alu and one SVA insertion confirmed as de novo in hiPSCs and a single Alu insertion (Alu-2) in hESCs (Fig. 3a; Supplementary Figs 9 and 10; Supplementary Table 3). The remaining six putative de novo insertions (one L1, four Alu and one SVA) were detected by PCR in parental cells or an earlier hESC passage, suggesting that these variants were present but were not de novo. Next, we determined the entire nucleotide sequence of 10/11 confirmed de novo retrotransposon insertions (Supplementary Figs 9 and 10). For one event, SVa2, a member of the SVA subfamily, we could sequence only the 3' junction, which included a poly-A tail characteristic of L1-mediated trans mobilization (Fig. 3a; Supplementary Fig. 9). Our efforts to PCR amplify the matching 5' junction of SVa2 with multiple primer combinations, intended to detect a possible 5' SVa truncation or a small proximal genomic deletion, were unsuccessful (see Methods). One reasonable explanation for this outcome was the occurrence of a large 5' genomic deletion at the SVa2 integration site, as reported previously13,40–42. Additional sequence analyses revealed that 9/10 of the remaining insertions exhibited the canonical hallmarks of L1-mediated target-primed reverse transcription43 including: (i) a target site duplication (TSD), (ii) a variable length L1 poly-A tail and (iii) an integration site resembling the L1 endonuclease target motif 5'-TTT/am-3' (refs 44,45; Fig. 3a; Supplementary Fig. 9). The one exception, insertion L1-dn4, was 3' truncated within its poly-A signal and devoid of an L1 endonuclease motif, but nevertheless incorporated an 8-bp TSD. These features were consistent with L1 endonuclease-independent retrotransposition46. Insertions L1-dn6 and L1-dn14 presented one and two untemplated G nucleotides at their 5' ends, respectively, as seen elsewhere40,42. Insertions L1-dn3, L1-dn4, L1-dn13 and L1-dn15 exhibited microcomplementarities of one to five nucleotides at their 5' end, a structural feature reported previously for L1 integration sites47. L1-dn13 and L1-dn15 were also 5' truncated and inverted, consistent with 'twin-priming'48, and in one instance a 5' inversion was displaced from the remaining L1 sequence by a 25-bp DNA fragment of unknown origin (L1-dn15). Thus, L1-mediated retrotransposition in hiPSCs and hESCs occurs via mechanisms described previously in mammalian cells.

The rate of L1-mediated retrotransposition occurring in pluripotent stem cells was difficult to accurately assess given the unknown genomic heterogeneity of each population. However, by estimating the sensitivity of RC-seq, we were able to determine the approximate L1 mobilization rate in hiPSCs. First, we identified that the overall RC-seq false positive rate was 1.5%, based on our recent PCR validation rate of 98.5% for insertions found by RC-seq in a cohort of hepatocellular carcinoma patients33 using the same detection thresholds as used here. Next, we determined that 88.5, 92.8, 88.3 and 89.8% of germline L1, Alu, SVA and LTR insertions, respectively, found in a parental cell line or early hESC passage were also detected in the matched hiPSC or later hESC passage, indicating an overall RC-seq false negative rate of 7.9%. To then model the sensitivity of RC-seq for de novo insertions, we randomly sampled each library and determined the fraction of the total germline events detected in that library as a function of sampling depth (Supplementary Fig. 11). At 50% library sampling depth (that is, modelling 50% variant allele fraction) 71.4%, 76.2%, 68.4% and 87.3%, respectively, of the germline L1, Alu, SVA and LTR insertions found in hiPSC lines were detected, dropping to 5.9%, 5.8%, 7.4% and 27.1% at 5% sampling depth. The estimated overall false negative rates at 50% and 5% variant allele fraction for de novo insertions detected in hiPSC lines were therefore 30.5% and 94.4%, respectively. These figures were similar for hESC lines (31.5 and 94.1%). Thus, we concluded that although RC-seq reliably detected high variant allele fraction retrotransposon insertions, a large pool of low variant allele fraction events may have been overlooked at the RC-seq thresholds used here. This would be particularly acute in the chosen hESC lines where, unlike iPSCs, cells had not undergone a recent population bottleneck in vitro. Using these parameters and the observed de novo L1 insertion counts, we estimated that hiPSC lines carried 3.7 de novo L1 insertions with allele frequencies ≥5%, on average, extrapolating to ~1 de novo L1 insertion per cell (see Methods). However, the low number of insertions identified precluded similar estimates for hESC lines.

Figure 3 | RC-seq reveals endogenous de novo L1, Alu and SVA retrotransposition in pluripotent stem cells. (a) Structures of validated de novo L1, Alu and SVA retrotransposition events (red box, untranslated region; white box, L1 ORF; green diamonds, TSDs). Names of insertions (for example, L1-dn10), and gene (for example, RPP25) was used. (b) Relative L1-dn13 content at passages 43, 56, 58 and 60 of culture 1 (red lettering) and at passages 28, 34, 43 and 49 of culture 2 (blue lettering) were quantified by qPCR. L1-dn13 is present in passages 43 to 60 of culture 1, but absent from culture 2.
RC-seq were absent from the earlier hiPSC passage surveyed and therefore may have arisen after reprogramming. To better establish the temporal dynamics of L1 retrotransposition in hiPSCs, we performed multiplex TaqMan qPCR incorporating a 5′ junction-spanning probe (Fig. 3b) to quantify L1-dn13 and L1-dn14 copy-number variation in hiPS-SB4. We observed an eightfold increase in L1-dn13 copy number upon extended cultivation (Fig. 3b, left panel) and a ~two-fold decrease in L1-dn14 copy number (Fig. 3b, right panel), indicating the presence of two different hiPS-SB4 subpopulations carrying
insertions L1-dn13 or L1-dn14, respectively, with opposite growth dynamics. L1-dn13 and L1-dn14 were not detected in hESCs (HES-3) or the parental fibroblast (HFF-1) population, again showing that L1-dn13 and L1-dn14 were de novo insertions. As the hiPS-SB4 line was cultivated from a single-cell-derived hiPSC clone, these data showed that either one or both of these insertions occurred during or after reprogramming, confirming our RC-seq and genotyping PCR data. To discriminate whether L1-dn13 arose during hiPSC reprogramming or cultivation, we thawed and extensively cultured a passage (p19) of hiPS-SB4 isolated well before the later passages analysed by RC-seq (p43 and p53) (Fig. 3c). L1-dn13 was not detected by qPCR in this second hiPS-SB4 cultivar (Fig. 3d). Hence, L1-dn13 likely arose in the original hiPS-SB4 cultivar between p19 and p43. We concluded that cultivation of hiPSCs, and hESCs, as described above for the Alu-2 insertion (Fig. 3a), can lead to endogenous retrotransposition.

De novo L1 insertions retain retrotransposition competency. Intriguingly, 4/7 de novo L1 insertions were full length, a surprising result given that most preexisting genomic L1 retrotransposition events are 5′ truncated. Indeed, only ~15% of L1 copies in the reference genome and <1% of somatic L1 insertions found thus far in tumours are full length.33,50,51. PCR amplification and sequencing of three full-length de novo L1s (L1-dn4, L1-dn6 and L1-dn14) revealed no deleterious nonsense mutations in their ORFs (Supplementary Fig. 10), suggesting each insertion likely retained retrotransposition competency. As a proof-of-principle, we used an established cell culture-based L1 retrotransposition reporter assay52 to evaluate the mobility of L1-dn6 in HeLa cells. L1-dn6 subclones retrotransposed at a relative efficiency of 20–30% of that obtained for the benchmark L1.3 (accession no. L19088.1)53 element (Fig. 4) and were therefore classified as highly active or ‘hot’53,11. These data indicated that new, full-length L1 insertions in hiPSCs could retain substantial competence in initiating further rounds of mobilization.

L1 insertional mutagenesis disrupts CADPS2 expression. Six de novo retrotransposition events mapped to introns of protein-coding genes. These included key factors in neuron (CADPS2 and NREP) and nephron (SLC12A1) biology, as well as genes with established and predicted roles in cell cycle regulation and oncogenesis (PTPN9, RNF38 and PLXDC2). Insertions showed a marked bias for the 5′ end of genes, with insertions falling on average in the 20th percentile of gene length measured from the annotated RefSeq transcription start site (TSS), a significant deviation from random expectation (P<0.006, permutation test). Albeit based on a small sample of insertions, this outcome could be explained by L1 endonuclease preference for open chromatin54 and increased chromatin accessibility around transcription start sites55.

Given that intrinsic L1 insertions can disrupt host gene transcription,15,56, we noted with interest that all six genes were expressed in hiPSCs and hESCs57. For example, L1-dn13 occurred in an intron of CADPS2 and, as noted above, exhibited copy-number variation during hiPSC cultivation (Fig. 3b,d). This afforded us an opportunity to analyse differential CADPS2 expression with reference to L1-dn13 copy number. First, we measured and compared CADPS2 mRNA expression in early versus late hiPS-SB4 passages via TaqMan qRT–PCR (Fig. 5a) and observed a fivefold reduction in CADPS2 expression in the latter cells (Fig. 5b). Importantly, this assay tested CADPS2 expression at an exon junction located downstream of the L1-dn13 integration site (Fig. 5a) and indicated opposing changes in L1-dn13 copy number (Fig. 3b, left panel) and CADPS2 expression for hiPS-SB4 cells in culture, suggesting that L1-dn13 interfered with CADPS2 expression.

To further test this possibility, we employed a human triose phosphate isomerase/Renilla luciferase reporter assay developed to monitor the effects of different introns on mammalian gene expression.58 We generated three constructs (Supplementary Methods; Supplementary Fig. 12a,b) respectively containing: (i) 825 bp spanning the empty L1-dn13 target intron of L1.3, (ii) 423 bp spanning the same region but in this case containing the 389 bp L1-dn13 insertion and its TSDs to produce a 825 bp sequence (pSTM06_01), (ii) 423 bp spanning the same region but in this case containing the 389 bp L1-dn13 insertion and its TSDs to produce a 825 bp sequence (pSTM06_01), (ii) 423 bp spanning the same region but in this case containing the 389 bp L1-dn13 insertion and its TSDs to produce a 825 bp sequence (pSTM06_01), (ii) 423 bp spanning the same region but in this case containing the 389 bp L1-dn13 insertion and its TSDs to produce a 825 bp sequence (pSTM06_01). We cloned each of these fragments into the triose phosphate isomerase/Renilla reporter cassette and quantified their effect on luciferase activity (Supplementary Fig. 12c). Interestingly, the CADPS2 intron sequence harbouring L1-dn13 (pSTM06_03) had the strongest inhibitory effect and reduced mRNA expression in early vs late cultures (Fig. 3b, left panel), and CADPS2 expression for hiPS-SB4 cells in culture, suggesting that L1-dn13 interfered with CADPS2 expression.
**Figure 5 | L1-dn13 affects CADPS2 expression.** (a) Schematic of the human CADPS2 allele of the hiPS-SB4 line harbouring insertion L1-dn13. A CADPS2 transcript including exons 7, 8, 27 and 28 is presented. Binding sites of the TaqMan primer/probe combination spanning the exon27/exon28 junction on CADPS2 cDNA used for qRT-PCR analysis are shown (red arrows and line). (b) Relative CADPS2 mRNA levels in early (p16) and late passage (p50) hiPS-SB4 cells were assessed by qRT–PCR. HES-3 and hFF-iPS4 cells served as positive controls. qRT–PCR results were normalized to 18S rRNA using CADPS2 expression in parental HFF-1 cells as control. Bars, arithmetic means ± s.e.m. of technical triplicates. (c) Structure of the L1-dn13 integration site in the CADPS2 gene in hiPS-SB4 subclones. hiPS-SB4_D differs from hiPS-SB4_B by the presence of the L1-dn13 de novo insertion in CADPS2 intron 7. Binding sites of L1-dn13-specific validation PCR primers OP1, ISP1 and ISP2 and expected lengths of the resulting PCR products are indicated. Black diamonds, TSDs. (d) Genotyping PCR validating the L1-dn13 presence in subclone hiPS-SB4_D and its absence from hiPS-SB4_B in gDNAs isolated from HFF-1, hiPS_SB4_B, and hiPS_SB4_D cells and from the original mixed population of the hiPS-SB4 culture (hiPS-SB4(Mix)). Primer combinations used are indicated in blue; H2O.ISP1/ISP2 and H2O.ISP1/OP1, negative control PCRs using H2O instead of gDNA; 100-bp ladder, size marker. (e) qPCR analyses confirming absence of L1-dn13 from hiPS-SB4_B and HFF-1 cells, and its presence in hiPS-SB4_D cells and the hiPS-SB4 culture. gDNAs from HFF-1 cells and from hiPS-SB4(Mix) cells served as negative and positive controls, respectively. For normalization, a primer/probe combination specific for the human RPP25 gene was used. ΔΔCt values measured the relative quantity of L1-dn13. Bars, arithmetic means ± s.e.m. of technical triplicates. (f) Relative CADPS2 mRNA levels in hiPS-SB4_B, hiPS-SB4_D and hiPS-SB4(Mix) cells were determined by qRT–PCR using cytoplasmic RNA and primer/probe combinations spanning exon 27/exon 28 junction of CADPS2. Bars, arithmetic means ± s.e.m. of technical triplicates.
luciferase activity by 62%, a significant decrease beyond the constructs lacking L1-dn13 (P = 0.022).

As further corollary, we isolated two clones from the original hiPS-SB4 culture by single cell cloning (see Methods) where L1-dn13 was identified by RC-seq, one carrying L1-dn13 (hiPS-SB4_D) and the other not carrying L1-dn13 (hiPS-SB4_B) (Fig. 5c). The identity of each clone was verified by genotyping PCR (Fig. 5d) and qPCR (Fig. 5e). qRT–PCR applied to cytoplasmic RNA extracted from each clone indicated that CADPS2 expression was ~95% lower in hiPS-SB4_D than in hiPS-SB4_B (Fig. 5f). Consistently, CADPS2 expression in the original hiPS-SB4 culture, which was heterogeneous for the L1-dn13 allele, was in between expression levels observed for the hiPS-SB4_D and hiPS-SB4_B clones. We then employed end point quantitative RT-PCR with subsequent capillary electrophoresis to compare the relative expression of each CADPS2 allele in hiPS-SB4_D (Supplementary Methods; Supplementary Fig. 13), as distinguished by a single nucleotide polymorphism located in the 3′ UTR of CADPS2. Notably, L1-dn13 was associated with complete inactivating of the CADPS2 mutant allele in hiPS-SB4_D while, interestingly, the CADPS2 wild-type allele was also downregulated by >90% relative to hiPS-SB4_B. Again, expression of each CADPS2 allele in the hiPS-SB4 culture heterogeneous for L1-dn13 lay between levels observed for the hiPS-SB4_B and hiPS-SB4_D subclones. Altogether, these results conclusively indicate that L1-dn13 interfered with CADPS2 expression.

Discussion

Here we have demonstrated that endogenous L1-mediated retrotransposition can occur in hiPSCs and hESCs, building upon earlier reports of engineered L1 retrotransposition in stem cells. By contrast, two previous studies reported an absence of endogenous retrotransposition events in mouse or human iPSCs. A more recent study reported low-level L1 mobilization in hiPSCs, though in this case no insertions could be confirmed by PCR, leaving the validity of the reported putative L1 insertions unclear. We unequivocally demonstrated here by RC-seq, gold-standard PCR validation and capillary sequencing, including L1 integration site structural characterization, that fibroblast-derived hiPSCs clearly can support the mobilization of endogenous non-LTR retrotransposons. We speculate that our use of clonally derived hiPSCs, and the robustness of RC-seq in detecting somatic L1 insertions, enabled us to discover retrotransposition events that may have otherwise remained undetected.

We estimated that hiPSCs each carried ~1 de novo L1 insertion, with the notable caveat that this calculation was based on a small number of observed events. Nonetheless, this is a much lower rate than recently found for human hippocampal neurons and glia (13.7% and 6.5% somatic L1 insertions per cell, respectively). Our sensitivity calculations suggested that most de novo insertions with a variant allele fraction of <5% in hiPSC and hESC populations were overlooked by RC-seq at the detection thresholds used here, and these were not included in the above rate estimates. This is a major consideration in concluding whether parental cell type or choice of reprogramming vector affects endogenous retrotransposition activity in hiPSCs. Low frequency or subclonal retrotransposition may indeed occur in our hCBECDerived hiPSC lines, that were reprogrammed via lentiviral systems, and escaped detection by RC-seq here. Therefore, we would propose that additional experiments are required to better define how these and other considerations affect L1 activity. Indeed, one explanation for the low number of insertions characterized in hESCs is that these cell populations were not clonally derived and were therefore likely to present more extensive genomic heterogeneity than hiPSCs. The lone Alu insertion found here in H9 cells is nonetheless the first endogenous retrotransposition event reported in hESCs, reinforcing evidence that L1-mediated mobilization can occur in early human development.

L1 activity was highly dynamic during reprogramming and hiPSC cultivation. Parental cells, early hiPSC passages, later hiPSC passages and re-differentiated cells presented grossly different levels of L1 expression. As corroborated by RC-seq, genotyping PCR and qPCR, the majority of retrotransposition in hiPSCs likely took place during or immediately after reprogramming, where we observed a peak in expression of the L1 mobilization machinery. As a result, each detected variant could affect substantial hiPSC subpopulations. Interestingly, major induction of L1 mRNA and protein expression, far in excess of that seen in hESCs and neural stem cells, was accompanied by a comparatively modest increase in L1 mobilization rate. Due to drastic epigenetic changes occurring upon reprogramming, it is possible that reprogramming per se may activate the expression of cellular L1 restriction factors such as APOBEC proteins and PIWIL2 (ref. 64). Consistently, APOBEC3B and PIWIL2 have been demonstrated to control engineered L1 retrotransposition in hiPSCs. Thus, it is tempting to speculate that the cellular milieu of hiPSCs and hESCs may permit L1 upregulation but also limit L1-mediated mutagenesis.

That 4/7 of the de novo L1 insertions reported here were full-length was consistent with 2/3 of the engineered L1 de novo insertions characterized by Wissing et al. and being full-length. This >50% incidence of full-length L1 de novo insertions in hiPSCs is unexpected as only ~15% of L1 copies in the human reference genome and <1% of somatic L1 insertions identified in tumours are full-length. However, 7/7 engineered L1 retrotransposition events found in hESCs were recently reported to be significantly 5′ truncated, suggesting that pluriptosity factors common to hiPSCs and hESCs might not play any role in the observed overlaprepresentation of full-length de novo L1 insertions found in hiPSCs. The mechanism of L1 5′ truncation is not fully understood. On one hand, the preponderance of 5′ truncated L1 copies in the genome has long been explained by an inability of the L1 reverse transcriptase encoded by L1 ORF2p to copy the entire template L1 RNA, either due to premature dissociation of the L1 reverse transcriptase from its RNA or competition from an unknown cellular RNase that digests the L1 RNA before completion of reverse transcription. Therefore, it is possible that hiPSCs provide a nuclear environment allowing a more stable association of the L1 reverse transcriptase with L1 RNA, or the L1 reverse transcriptase does not have to compete with a cellular RNase which might be differentially expressed in hiPSCs. On the other hand, a recent study demonstrated that the DNA-damage-signalling protein ATM may control the length or number of de novo L1 insertions in human neural stem cells. Thus, it is possible that subtle differences in the DNA repair mechanisms operating in hiPSCs and hESCs could be related to the high frequency of full-length L1 insertions characterized in hiPSCs.

Each de novo L1 insertion reported here integrated in a protein-coding gene expressed in pluripotent cells. In one case, we identified an L1 insertion (L1-dn13) that arose during hiPSC cultivation and integrated into an intron of the gene CADPS2. It remains to be determined whether acquisition of L1-dn13, and a concurrent reduction in CADPS2 expression, imputed carrier hiPSCs with a selective advantage in vitro. Furthermore, it remains unclear why transcription of the CADPS2 allele lacking L1-dn13 was reduced by >90%. To speculate, it is possible that...
CADPS2 expression involves a direct or indirect positive feedback loop where, for example, transcription from CADPS2 reinforces open chromatin\(^7\). A reduction in CADPS2 expression caused by L1-dn13 could hence have a strongly negative effect on transcription from the wild-type CADPS2 allele.

In closing, it is notable that intrinsic L1, Alu and SVA insertions can alter cellular phenotype and are associated with numerous instances of human disease\(^8\). Future in-depth experiments are however required to definitively establish whether endogenous retrotransposition alters the phenotype of hiPSC derivatives sufficiently to impact their use in medical or research applications. We can nevertheless conclude that retrotransposition, in addition to other sources of genetic and epigenetic variation\(^9\), can change the functional landscape of the hiPSC genome.

**Methods**

**Cell lines and culture conditions.** hiPSC lines hiP-SB4 (hiFF-T2-Oskml) and hiPS-SB5 (hiFF-T2-Oskml/hiPS-SB5.1) were generated by reprogramming HFF-1 cells (ATCC-Number: SCRC-1041) using Sleeping Beauty (SB) transposon-based plasmids pT2-OSKM or pT2-OSKML which contain polycistronic OSKM (OCT4, SOX2, KLF4 and c-myc) or OSKML (OSKM + LIN28) expression cassettes\(^26\). Briefly, HFF-1 cells (4 × 10^5 cells per well) were transfected by nucleofection according to the manufacturer’s instructions. In each transfection, 2 μg of transposon plasmid (pT2-OSKM or pT2-OSKML) and 0.2 μg of CMV- SB100X vector (harbouring the enhanced Sleeping Beauty transposase gene under control of a CMV promoter\(^28\)) were used. After transfection cells were plated onto Matrigel-coated six-well plates (hESC-qualified Matrix, BD Biosciences) and were grown in MEF-conditioned ESC medium used for the cultivation of hESCs and hiPSCs. ESC medium consisted of Knockout DMEM (Life Technologies) supplemented with 4 ng ml\(^{-1}\) basic fibroblast growth factor 2 (FGF2, Invitrogen), 20% Knockout Serum Replacement ( Gibco, 1 mM l-glutamine, (Biochrom AG), 50 μM B-mercaptoethanol and 0.1 mM nonessential amino acids. The medium was changed every 2 days. Embryoid bodies were cultured for 8 days in total. Embryoid bodies were collected by sedimentation under gravity from three dishes on day 0 (untransfected hiPSCs), 2, 4, 6 and 8, respectively (Supplementary Fig. 6). Total RNA was extracted from each well using Trizol (Invitrogen) following the instructions of the manufacturer.

**Analysis of expression in embryoid bodies by qRT-PCR.** To analyse the expression of both pluripotency markers and L1, real-time quantitative RT-PCR was performed on total RNA per well used for cell sorting by the High Capacity RNA-to-CDNA kit (Applied Biosystems). For each time point and transcript to be quantified, qRT-PCR analyses were done in triplicate. qRT–PCR for pluripotency/differentiation markers was carried out using Power SYBR Green PCR Master Mix (Applied Biosystems) on the ABI7900HT Sequence Detection System (Applied Biosystems). Data was normalized to GAPDH expression. qRT–PCR for L1 was performed with Absolute QPCR Mix (Agene), and data was normalized to 18S rDNA expression.

**qRT-PCR using TaqMan fluorogenic probes.** Cytoplasmic RNA was extracted from 5 × 10^5 to 3 × 10^6 somatic cells, hiPSCs or embryoid body cells using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. Cytoplasmic RNA (0.5–1 μg) was incubated with 2 μl Rnase-Free DNasel (Life Technologies, Darmstadt, Germany) for 30 min at room temperature. DNasel digestion was stopped by adding 2 μl of 25 mM EDTA and incubation for 10 min at 65 °C. DNasel-digested cytoplasmic RNA (0.1–0.5 μg) was used for cDNA synthesis using the SuperScript III First-Strand Synthesis Kit (Invitrogen) in combination with a Random 6–8 and PrimeIt (0.5 μg ml\(^{-1}\); Invitrogen) according to the manufacturer’s instructions. Quantitative real-time PCR was carried out in AbGene plates using an Applied Biosystems 7900HT Fast Real-Time PCR System. The primer and probe combination L1 5’UTR (ref. 60) was used to quantify transcripts expressed from each L1-T2 allele. Four oligonucleotides and probes used for qRT–PCR are listed in Supplementary Table 1. The probe specific for the L1 5’UTR was labelled with the fluorophore 6-carboxy-fluorescein (FAM) and a non-fluorescent quencher. 18S rRNA expression was quantified using Eukaryotic 18S rRNA endogenous control (VIC/TAMRA Probe, Primer Limited, Part number 4310893E, Applied Biosystems). Transcript levels of the human CADPS2 gene were monitored using a gene specific assay (Life Technologies, Hs00645283_m1) spanning exon sequences (Fig. 5a). Cycling conditions were the following: 95 °C for 15 min (one cycle), 95 °C for 15 s and 60 °C for 1 min (40 cycles). A total of 1–5 μl of cDNA per sample were used for the quantification of endogenous L1 and CADPS2 mRNA levels. Analysis of real-time and end point fluorescence was performed using the software SDS version 2.3 as well as RQ manager 1.2 (Applied Biosystems).

**Northern blot analysis.** Total RNA was isolated from the cell lines HFF-1, 2102Ep (ref. 60), HEK-3 and hiPS-SB4 using TRIzol (Invitrogen) according to the manufacturer’s instructions. Poly(A) + RNA was isolated applying the Dynabeads mRNA Purification Kit (Life Technologies) according to the manufacturer’s instructions. Denatured mRNA (2.8 μg) from each cell line was subjected to denaturing electrophoresis in a horizontal 1% agarose gel containing morpholinopropanesulfonic acid buffer and 6% formaldehyde, and transferred onto a Hybond-N + Nylon membrane (Amersham) by overnight capillary transfer using 10 × SSC as transfer buffer. A total of 4 μl RiboBlot High Range RNA ladder (MBI Fermentas, St.Leon-Rot, Germany) were loaded as size marker. After crosslinking the RNA onto the membrane by ‘baking’ at 80 °C for 2 h, the membrane was prehydrized overnight in 50% Formamide/4×SSC/1%SDS at 65 °C. The full-length L1 mRNA-specific probe was generated by PCR amplification of a 1299 bp fragment ranging from position numbers (pos.) 58–1356 of a full-length L1 element by using primers L1 FW1 and L1 RV1 (Supplementary Table 1) and pM101/L1PMACMV\(^7\) as template. Pos. refer to the
for 1 h at room temperature in 5% BSA/1% PBS (pH 7.4). Subsequently, cells were washed three times with 1 × PBS for 5 min each at room temperature. Cells were incubated with 5% BSA/1% PBS/0.1% Triton X-100 (Sigma) for 1 h at room temperature. Membrane preparations were washed thrice for 5 min each at room temperature using 1 × PBS. Subsequently, membranes were blocked for 1 h with 5% milk powder in PBS-T and incubated overnight with the respective primary antibody. The analysis was performed using an Axio Observer A1 microscope (Carl Zeiss MicroImaging, Goettingen, Germany).

**RC-seq library preparation, sequencing and analysis.** GemRNA was isolated from 1 × 10^6 cells from each hESC and hiPSC line and its respective parental cells using DNAzol Genomic DNA Isolation Reagent (MRC Inc, Cincinnati, OH, USA) according to the manufacturer’s instructions. 

**Immunofluorescence staining.** hiPSC colonies were detached from their tissue culture dish by incubation with 250 μl of a 1 mg/ml collagenase type IV/DMEM and washed subsequently with PBS (pH 7.4). 1 × 10^5 cells in 1 ml hESC media were plated in lysis buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 10% Glycerin, 1% Triton X-100, 2 mM EDTA, 2 mM EGTA, 40 mM β-Glycerolphosphate disodium salt hydrate, 50 mM NaF, 10 mM Na3PO4, 200 μM Na3VO4, 2 mM DTT; 1 × complete protease inhibitor cocktail (Roche Applied Science), homogenized by passing the lysate ten times through a 26 G needle, and lysates were cleared by centrifugation. A total of 50 μg of each protein lysate were boiled in 3 × SDS sample buffer (NEB), loaded on 4–12% BisTris gels (Invitrogen), subjected to SDS–polyacrylamide gel electrophoresis, and electroblotted onto nitrocellulose membranes. After protein transfer, membranes were blocked for 2 h at room temperature in a 10% solution of non-fat dry milk in PBS-T (1:17 mM NaCl, 3 mM KCl, 16.5 mM NaHPO4, 1.5 mM KH2PO4, 0.05% Tween 20) followed by 30 s at 58°C. Amplified products were gel purified (QaAquick gel extraction kit, Qiagen), cloned in pGEM-T Easy (Promega) and at least 30 individual clones were sequenced for each sample. The unique sequence in each clone was analysed using RepeatedMasker at http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker. Next, the fraction of unmethylated CpG sites was calculated by comparison to a consensus L1-Ta sequence. In addition, each individual sequence was compared to the existing insertion site-specific primers if required. In cases where an insertion was detected by RC-seq at one terminus only, PCR and capillary sequencing were applied to the remaining end to resolve integration site structure.

PCR reactions contained 0.125 μl of Primer Taq (New England Biolabs), 5 μl PCR-buffer, 10 μl of each Primer, 10 mM dNTPs and 10–20 ng genomic template DNA of 25 μl. The following cycle parameters were used: 95°C for 2 min, then 35 cycles of 95°C for 30 s, 58°C for 30 s, 68°C for 40 s, followed by a single extension step at 68°C for 5 min. Optimization in some cases required adjusted annealing temperatures and cycle number. PCR products of the correct size (Fig. 3a) that were obtained with the retrotransposon primer in combination with the gemRNA primer were TA-cloned and sequenced. The same method was applied to both the 5′ and the 3′ ends of all de novo insertions to fully characterize each, apart from SVA-2. To PCR amplify the 5′ junction of the SVA-2 insertion from genmic DNA, we designed three SVA-specific primers and three oligonucleotides binding 50–300 bp upstream of the SVA-2 integration site. To facilitate the detection of a potentially 5′-truncated SVA, the SVA-specific primers were placed within the sequenced 123 bp of the SVA-2 3′ end (Supplementary Fig. 9), at the junctions of the SVA-2-specific Alu-like and VNTR region, and the (CCCTCT)n repeat and Alu-like region, respectively. Combinatorial use of these genome-SVA primer pairs did either not result in a PCR product or generated too nonspecific products. For a complete list of used primers see Supplementary Table 3. Eleven de novo insertions (seven L1, three Alu and one SVA) were confirmed by PCR as de novo. Six additional insertions were determined as gemline insertions, already present in the parental cell line or an early hESC passage. Control genotyping PCR of the single-copy gene GAPDH in genomic DNA preparations of parental and hiPSC lines used for RC-seq insertions were applied to identify the de novo insertions is presented in Supplementary Fig. 15. PCR amplification was performed using primers GAPDH1 (5′-CAAAGCTTGGGCGACAGCTTG3′) and GAPDH-β (5′-GAGAGCTGACCTGGAATGTTG3′) which bind in exon 8 (chr12:6646861-66468680) and intron 7 (chr12:6647005-6647026), respectively, resulting in a 466-bp DNA fragment. Cycling conditions were identical to those described above.

**Quantification of L1-dn13 and L1-dn14 copy numbers by qPCR.** To determine relative copy numbers of de novo insertions L1-dn13 and L1-dn14 within the
hiPSC-SB4 culture, we applied real-time qPCR using TaqMan fluorogenic probes. To this end, genomic DNA was isolated using 1 mL DNAzol Genomic DNA Isolation Reagent (Gibco-Invitrogen, Cincinnatti, OH, USA) from 1 × 10⁶ cells, according to the manufacturer’s instructions. A total of 100 ng of genomic DNA was used for quantitative real-time PCR (qPCR). Primer and probe combinations specific to the genomic 5′ junctions of the de novo insertions L1-dn13 and L1-dn14 (Fig. 3b) were used to quantify the copy number of the respective insertion in hiPSC cultivars. Each probe was labeled with fluorescein-carboxyfluorescein and a non-fluorescent quencher. For normalization the single-copy gene RPP25 (Ribonuclease P/MRP 25kDa subunit; FAM/non-fluorescent quencher, primer limited, HS00706565_S1; Applied Biosystems) was used. Cycling conditions were: 95 °C for 15 min (one cycle), 95 °C for 15 s and 60 °C for 1 min (40 cycles). For analysis of real-time and end point fluorescence, the software SDS version 2.3 as well as RQ manager 1.2 (Applied Biosystems) were used.

Isolation of hiPSC-SB4 single-cell subclones. To isolate single cell subclones from the hiPSC-SB4 culture by limiting dilution, hiPSC-SB4 cells of passage 64 representing a mixed population of cells with and without the L1-dn13 de novo retrotransposition event, were magnetically separated from feeder cells by applying a Feeder Removal Kit (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) according to manufacturer’s instructions. A total of 100 ng of genomic DNA was used for relative L1 mRNA levels determined by qRT–PCR was performed by ANOVA, as well as RQ manager 1.2 (Applied Biosystems) were used. A total of 100 ng of genomic DNA was used for relative L1 mRNA levels determined by qRT–PCR was performed by ANOVA, as well as RQ manager 1.2 (Applied Biosystems) were used.

L1 retrotransposition reporter assays. De novo full-length L1 insertions were amplified from genomic DNA using an Expand Long Template PCR system (Roche) using primers L1-prim6 and L1-prim9. Upon upstream extension the insert becomes full-length (available upon request). For each PCR we used: 0.3 μL Expand Long Template Taq (Roche), 1 × buffer#1, 400 μM dNTPs, 1 μM each Primer and 30 ng genomic DNA in 50 μL per tube. Cycling conditions were: 95 °C for 5 min, then 30 cycles of 95 °C for 1 min, 56 °C for 30 s, 60 °C for 1 min, followed by a single extension step at 68 °C for 10 min. To avoid the generation of mutants that may lead to retrotransposition defective elements, we conducted at least four independent PCRs per L1. PCR products were resolved on 0.9% agarose gels, and fragments of the expected length of ~6 kb representing potential full-length L1 elements were excised and purified using a Qiaquick kit (Qiagen) and cloned in the Topo-XL plasmid vector (Invitrogen). The clones were then transferred to fecoder-coated 12-well plates (Thermo Fisher/Nunc) and further expanded. To isolate genomic DNA from each clone, cells were harvested after collagenase IV treatment, centrifuged, washed and pelleted. Genomic DNA was isolated as described in the previous paragraph. Genotyping PCR conditions applied to screen for the presence of the L1-dn13 insertion and to determine its presence/absence (Fig. 5d) are identical to those described above for insertion PCR validation. Primers used to demonstrate presence/absence of L1-dn13 are provided in Supplementary Data 1 and Supplementary Table 3. PCR products were visualized on a 1.5% agarose gel after ethidium bromide staining.

Reduction in full-length transcript levels in the embryoid body time kinetics experiment was evaluated by means of Linear Regression for data from day 0 to day 8 ($R^2 = 0.79$). Analyses were performed with SAS/STAT software, version 9.2 SAS system for Windows.

References
1. Yamanaka, S. Induced pluripotent stem cells: past, present, and future. Cell Stem Cell 10, 678–684 (2012).
2. Gore, A. et al. Somatic coding mutations in human induced pluripotent stem cells. Nature 471, 63–67 (2011).
3. Hussein, S. M. et al. Copy number variation and selection during reprogramming to pluripotency. Nature 471, 58–62 (2011).
4. Laurent, L. C. et al. Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. Cell Stem Cell 8, 106–118 (2011).
5. Lister, R. et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 471, 68–73 (2011).
6. Mayhara, Y. et al. Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell 7, 521–531 (2010).
7. Ben-David, U. & Benvenisty, N. The tumorigenicity of human embryonic and induced pluripotent stem cells. Nat. Rev. Cancer 11, 268–277 (2011).
8. Levin, H. L. & Moran, J. V. Dynamic interactions between transposable elements and their hosts. Nat. Rev. Genet. 12, 615–627 (2011).
9. Beck, C. R. et al. LINE-1 retrotransposition activity in human genomes. Cell 141, 1149–1150 (2010).
10. Mills, R. E., Bennett, E. A., Iskow, R. C. & Devine, S. E. Which transposable elements are active in the human genome? Trends Genet. 23, 183–191 (2007).
11. 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).
12. Ewing, A. D. & Kazazian, Jr. H. H. Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. Genome Res. 21, 985–990 (2011).
13. Garcia-Perez, J. L. et al. LINE-1 retrotransposition in human embryonic stem cells. Hum. Mol. Genet. 16, 1569–1577 (2007).
14. Kano, H. et al. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Gene Dev. 23, 1303–1312 (2009).
15. Han, J. S., Szak, S. T. & Boeke, J. D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcripomes. Nature 429, 268–274 (2004).
16. Hancks, D. C. & Kazazian, Jr. H. H. Active human retrotransposons: variation and disease. Curr. Opin. Genet. Dev. 22, 191–203 (2012).
17. Beck, C. R., Garcia-Perez, J. L., Badge, R. M. & Moran, J. V. LINE-1 elements in structural variation and disease. Annu. Rev. Genomics Hum. Genet. 12, 187–215 (2011).
18. Bourc’his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon inactivation in male germ cells lacking Dnmt3L. Nature 431, 96–99 (2004).
19. Maherl, N. et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 1, 55–70 (2007).
20. Wissing, S. et al. Reprogramming somatic cells into iP cells activates LINE-1 retroelement mobility. Hum. Mol. Genet. 21, 2411–2423 (2012).
21. Friedli, M. et al. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. Genome Res. 24, 1251–1259 (2014).
22. Wissing, S., Montano, M., Garcia-Perez, J. L., Moran, J. V. & Greene, W. C. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. J. Biol. Chem. 286, 36427–36437 (2011).
23. Cheng, L. et al. Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. Cell Stem Cell 10, 337–344 (2012).
24. Quinlan, A. R. et al. Genome sequencing of mouse induced pluripotent stem cells reveals retroelement stability and infrequent DNA rearrangement during reprogramming. Cell Stem Cell 9, 366–373 (2011).
25. Briggs, J. A. et al. Integration-free induced pluripotent stem cells model genetic and neural developmental features of down syndrome etiology. Stem Cells 31, 467–478 (2013).
26. Grabundzija, I. et al. Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells. Nucleic Acids Res. 41, 1822–1837 (2013).
27. Haase, A. et al. Generation of induced pluripotent stem cells from human cord blood. Cell Stem Cell 5, 434–441 (2009).
28. Abyzov, A. et al. Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. Nature 492, 438–442 (2012).
29. Groddler, J. L., Zhang, L., Vetter, M. R. & Kazazian, Jr. H. H. LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. Mol. Cell. Biol. 27, 6469–6483 (2007).
30. Doucet, A. J. et al. Characterization of LINE-1 ribonucleoprotein particles. PLoS Genet. 6, e1001150 (2010).
31. Rountsirn, N. et al. Transpositional element-1 protein expression is a hallmark of many human cancers. Am. J. Pathol. 184, 1280–1286 (2014).
32. Moldovan, J. B. & Moran, J. V. The zinc-finger antiviral protein ZAP inhibits LINE and Alu retrotransposition. PLoS Genet. 11, e1005121 (2015).
33. Shukla, R. et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell 153, 101–111 (2013).
34. Ewing, A. D. & Kazazian, Jr. H. H. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res. 20, 1262–1270 (2010).
35. Iskow, R. C. et al. Natural mutagenesis of human genomes by endogenous retrotransposons. Cell 141, 1253–1261 (2010).
36. Wang, J. et al. dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. Hum. Mutat. 27, 323–329 (2006).
37. Mir, A. A., Philippe, C. & Cristofari, G. euL1db: the European database of L1 retrotransposition insertions in humans. Nucleic Acids Res. 43, D43–D47 (2015).
38. Wang, H. et al. SVA elements: a hominid-specific retroposon family. J. Mol. Biol. 354, 994–1007 (2005).
39. Symer, D. E. et al. Human L1 retrotransposition is associated with genetic instability in vivo. Cell 110, 327–338 (2002).
40. Raiz, J. et al. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res. 40, 1666–1683 (2012).
41. 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).
42. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eckbush, T. H. Reverse transcription of R2BN RNA is primed by a nicks at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72, 395–605 (1993).
43. Cost, G. L., Feng, Q., Jacquier, A. & Boeke, J. D. Human L1 element target-primed reverse transcription in vitro. EMBO J. 21, 5899–5910 (2002).
44. Jurka, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroelements. Proc. Natl Acad. Sci USA 94, 1872–1877 (1997).
45. Morrish, T. A. et al. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat. Genet. 31, 139–165 (2002).
46. Zingler, N. et al. Analysis of 5’ junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5’-end attachment requiring microhomoology-mediated end-joining. Genome Res. 15, 780–789 (2005).
47. Ostertag, E. M. & Kazazian, Jr. H. H. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. Genome Res. 11, 2059–2065 (2001).
48. Grimaldi, G., Skowronski, J. & Singer, M. F. Defining the beginning and end of KpnI family segments. EMBO J. 3, 1753–1759 (1984).
49. Lee, E. et al. Landscape of somatic retrotransposition in human cancers. Science 337, 967–971 (2012).
50. Solymos, S. et al. Extensive somatic L1 retrotransposition in colorectal tumors. Genome Res. 22, 2328–2338 (2012).
51. Moran, J. V. et al. High frequency retrotransposition in cultured mammalian cells. Cell 87, 917–927 (1996).
52. Sassaman, D. M. et al. Many human L1 elements are capable of retrotransposition. Nat. Genet. 16, 37–43 (1997).
53. Cost, G. L., Feng, Q., Schilissel, M. S. & Boeke, J. D. Target DNA chromatization modules nicking by L1 endonuclease. Nucleic Acids Res. 29, 573–577 (2001).
54. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. Nature 489, 75–82 (2012).
55. Kær, K. & Speek, M. Retroelements in human disease. Genie 518, 231–241 (2013).
56. Guenther, M. G. et al. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell 7, 249–257 (2010).
57. Nott, A., Meislin, S. H. & Moore, M. J. A quantitative assessment of intron effects on mammalian gene expression. RNA 9, 607–617 (2003).
58. Schütte, T. D. et al. Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. Anal. Biochem. 285, 194–204 (2000).
59. Coufal, N. G. et al. L1 retrotransposition in human neural progenitor cells. Nature 460, 1127–1131 (2009).
60. Arokium, H. et al. Deep sequencing reveals low incidence of endogenous LINE-1 retrotransposition in human induced pluripotent stem cells. PLoS ONE 9, e108682 (2014).
61. Upton, K. R. et al. Ubiquitous L1 mosaicism in hippocampal neurons. Cell 161, 228–235 (2015).
Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Klawitter, S. et al. Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. *Nat. Commun.* 7:10286 doi: 10.1038/ncomms10286 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/
Author Correction: Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells

Sabine Klawitter1,10, Nina V. Fuchs1,2, Kyle R. Upton3, Martin Muñoz-Lopez4, Ruchi Shukla5, Jichang Wang2, Marta García-Cañadas4, Cesar Lopez-Ruiz4, Daniel J. Gerhardt3, Attila Sebe1, Ivana Grabundzija2, Sylvia Merkert6, Patricia Gerdes3, J. Andrés Pulgarín4, Anja Bock1, Ulrike Held1, Anett Witthuhn6, Alexandra Haase6, Balázs Sarkadi7, Johannes Löwer1, Ernst J. Wvetang8, Ulrich Martin6, Zoltán Ivics1, Zsuzsanna Izsvák2, Jose L. García-Perez4, Geoffrey J. Faulkner3,9 & Gerald G. Schumann1

Correction to: Nature Communications; https://doi.org/10.1038/ncomms10286; published online 8 January 2016

This Article contains an error in the author affiliations. The correct affiliation for author Ruchi Shukla is 'MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh, EH4 2XU, UK', and is not 'Mater Research Institute – University of Queensland, TRI Building, Woolloongabba QLD 4102, Australia'.

Published online: 19 December 2018

OPEN

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018

1Division of Medical Biotechnology, Paul-Ehrlich-Institute, D-63225 Langen, Germany. 2Max-Delbrück-Center for Molecular Medicine, D-13125 Berlin, Germany. 3Mater Research Institute, University of Queensland, TRI Building, Woolloongabba, Brisbane, Queensland 4102, Australia. 4Department of Human DNA Variability, Pfizer/University of Granada and Andalusian Regional Government Center for Genomics and Oncology (GENYO), PTS Granada, 18016 Granada, Spain. 5MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK. 6Leibniz Research Laboratories for Biotechnology and Artificial Organs (LEBAO), Department of Cardiac, Thoracic, Transplantation, and Vascular Surgery; REBIRTH, Cluster of Excellence, Hannover Medical School, D-30625 Hannover, Germany. 7Department of Biophysics and Radiation Biology, Semmelweis University, H-1094 Budapest, Hungary. 8Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St Lucia, Queensland 4072, Australia. 9Queensland Brain Institute, University of Queensland, Brisbane, Queensland 4072, Australia. 10Present address: Division of Inborn Metabolic Diseases, University Children’s Hospital, D-69120 Heidelberg, Germany. These authors contributed equally: Sabine Klawitter, Nina V. Fuchs, Kyle R. Upton, Martin Muñoz-Lopez. The original article can be found online at https://doi.org/10.1038/ncomms10286. Correspondence and requests for materials should be addressed to J.L.G.-P. (email: jlgp@genyo.es) or to G.J.F. (email: faulknergj@gmail.com) or to G.G.S. (email: Gerald.Schumann@pei.de)