A framework for efficient CRISPRi-mediated silencing of retrotransposons in human pluripotent stem cells

Anita Adami[1,2], Raquel Garza[1,2], Fereshteh Dorazehi[1], Christopher H. Douse[1] and Johan Jakobsson[1,2,*,]**

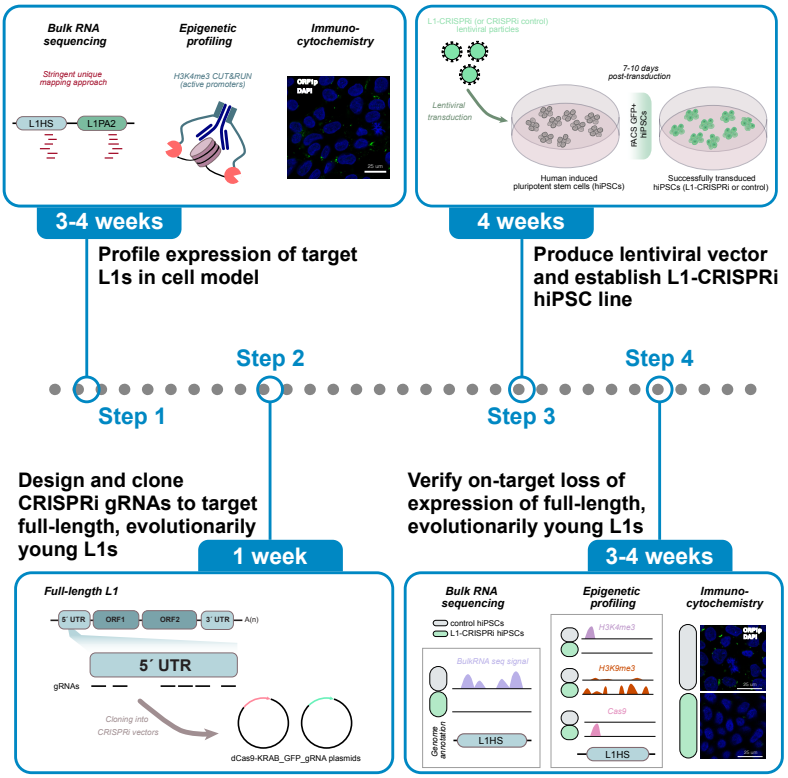
[1] *Department of Experimental Medical Science, Wallenberg Neuroscience Center and Lund Stem Cell Center, BMC A11, Lund University, 221 84 Lund, Sweden.*
[2] *Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, MD 20815, USA.*

* Technical contact: johan.jakobsson@med.lu.se
** Lead Contact: johan.jakobsson@med.lu.se

Summary

This protocol describes the workflow for transcriptional silencing of transposable elements (TEs) in human induced pluripotent stem cells (hiPSCs). The protocol illustrates how to design gRNAs to target families of TEs (or a unique TE locus) and how to validate the efficiency and specificity of large-scale CRISPRi-based silencing using a multiome approach that combines bulk RNA sequencing, CUT&RUN epigenetic profiling, and proteomics. This protocol offers a unique framework to perform and interpret *in vitro* functional studies based on TEs transcriptional manipulation in hiPSC models.
For complete details on the use and execution of this protocol, please refer to Adami *et al.* (2025)[1].

Graphical abstract



Before you begin

Transposable elements (TEs) are repetitive sequences that have colonised the human genome throughout evolution and comprise approximately 50% of the human genome[2,3]. Long considered genomic relics, TEs have been severely understudied. Moreover, their repetitive nature poses considerable challenges in the investigation of these elements. However, advances in sequencing technologies and off-target-minimised CRISPR-based approaches now enable thorough TE investigation at the locus-specific and subfamily levels. Still, downstream on- and off-target analyses of transcriptional manipulations are needed to validate the specificity of these tools and ensure that any observed phenotypic effect in a chosen *in vitro* model is in fact consequence of the TE manipulation.

The protocol described below provides detailed steps for designing gRNAs to target TE families (multiple copies) using CRISPR inhibition (CRISPRi) and to validate the successful targeting of multiple sequences. Specifically, we present a solid framework for thoroughly assessing the specific and efficient silencing of evolutionary young long interspersed nuclear elements-1 (L1s) (Figure 1) in human induced pluripotent stem cells (hiPSCs). This protocol can also be used to design gRNAs for CRISPR-based transcriptional manipulation (CRISPRi or CRISPR activation) of other TE subfamilies, such as human endogenous retroviruses-K (for this, see also Padmanabhan Nair *et al.*, 2022[4]).

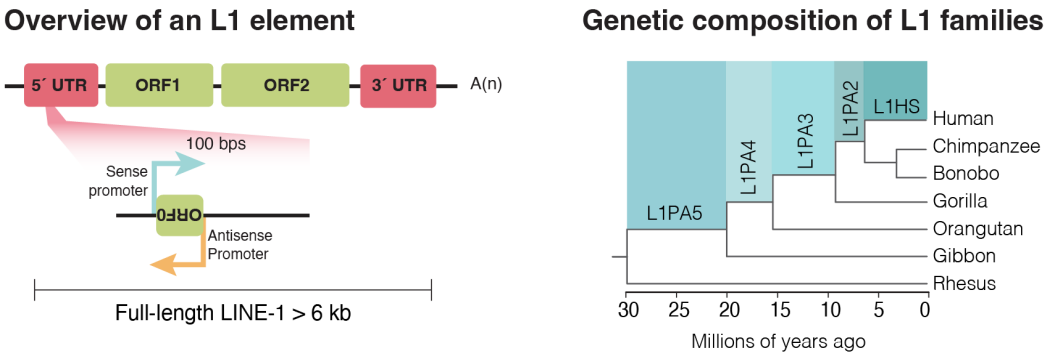


Figure 1. *Left: Overview of the structure of a full-length, evolutionarily young L1 elements which include an intact 5' untranslated region (UTR), three open reading frames (ORFs), and a 3' UTR. Right: Phylogenetic tree showing the evolutionary age of the younger L1 subfamilies.*

We performed the CRISPRi experiments using an all-in-one CRISPRi backbone from Addgene[5] (RRID: Addgene_71237) into which we cloned the different gRNAs. For the successful application of this protocol, techniques such as molecular cloning, lentiviral production, sequencing infrastructure accessibility, hiPSC culture routines, and bioinformatical expertise must be established in the lab. These techniques are not thoroughly covered here.

TE expression profiling *in vitro*

Timing: 3-4 weeks

Before designing and establishing the TE-targeting CRISPRi system, it is essential to profile the expression of the TE family in the chosen *in vitro* model to adequately design the downstream investigations. In this study, we performed an in-depth multiomics analysis that combined bulk RNA-sequencing, CUT&RUN epigenetic profiling and proteomics to quantify the expression of evolutionarily young, full-length (> 6kbp) L1s in human induced pluripotent stem cells (hiPSCs).

CRITICAL: Profiling the expression of L1s (or the TE family of interest) in the chosen model *before* performing any transcriptional manipulation experiment is essential to ensure that only relevant reagents and resources are consumed and only data relevant to achieving the project objectives is collected and processed. Furthermore, this is key to adequately interpret the transcriptional perturbations introduced with the CRISPRi system and perform a locus-specific analysis of the targeted TEs.

1. Culture hiPSCs and harvest pellets for bulk RNA-seq (stranded libraries, paired-end 2 x 150 bp). A minimum of three technical replicates is generally recommended when performing this type of analysis to perform statistics.

CRITICAL: It is important to perform *stranded* RNA sequencing to accurately quantify the transcription of the L1 sequence *in sense* (potentially originating from the element's promoter), or *in antisense*, (potential read-through signal coming e.g., from nearby or overlapping protein-coding genes).

2. Analyse bulk RNA-seq data using a unique mapping approach (See also bulk RNA sequencing analysis of L1-CRISPRi hiPSCs within this publication).

CRITICAL: Apply a unique mapping strategy to detect locus-specific TE transcription. This ensures the detection of L1s on a single-locus level and avoids any mapping ambiguities caused by the repetitive nature of these elements.

3. Perform H3K4me3 CUT&RUN epigenetic profiling[6] to identify active promoter markers of evolutionarily young L1s.

Note: This is a relevant step to assess whether these young elements are expressed from their own promoter, or they undergo passive transcription initiated from another gene promoter[7,8]. H3K4me3 CUT&RUN profiling is particularly valuable for studying the intrinsic expression of evolutionarily young TEs, whose highly repetitive nature poses significant mappability challenges. Pairing short-read bulkRNA sequencing, which does not overcome the mapping ambiguities of the highly repetitive sequences, with epigenetic profiling allows to assess the promoter activity of the youngest, lowly mappable, elements and address the mappability issues.

4. Perform a Western blot and/or an immunostaining (Figure 2) against the L1-derived protein ORF1p to determine whether the expressed L1 elements produce protein.

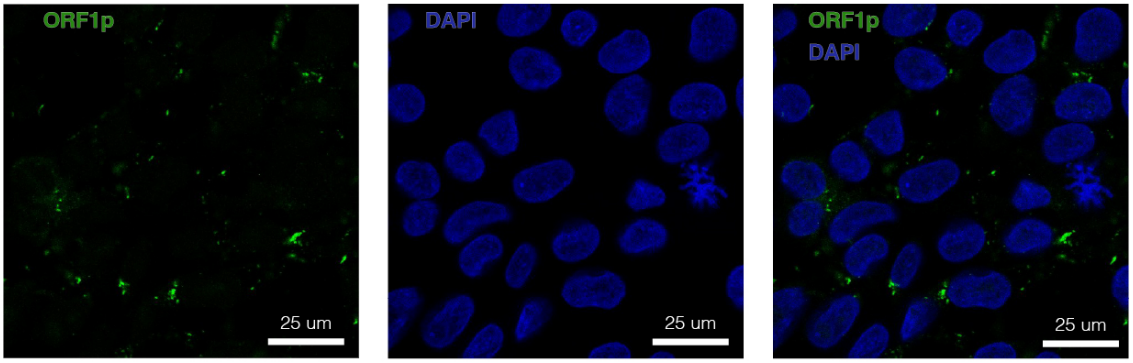


Figure 2. *Immunocytochemistry of the L1-derived protein ORF1p in human induced pluripotent stem cells. ORF1p is shown in green and a blue nuclear dapi staining is included.*

CRITICAL: It is essential that the antibody used is validated and specific to the target. If working with a different subfamily of protein-encoding TEs (e.g. HERV-K elements), this might not be trivial, as many antibodies against TEs can be unspecific.

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
rabbit anti-H3K9me3, polyclonal	Abcam	RRID:AB_306848
rabbit anti-H3K4me3, polyclonal	Active Motif	RRID:AB_2615077
goat anti-rabbit IgG, polyclonal	Abcam	RRID:AB_10681025
mouse anti-ORF1p, clone 4H1	Sigma-Aldrich	RRID:AB_2941775
rabbit anti-Cas9, polyclonal	Takara	RRID:AB_3251495
donkey anti-mouse Cy3	Jackson ImmunoResearch Labs	RRID: AB_2340813
Chemicals, peptides, and recombinant proteins		

ConA-coated magnetic beads (BioMag®Plus Concanavalin A)	Bangs Laboratories	Cat# BP531
Roche cOmplete protease inhibitor	Sigma Aldrich	Cat# 11697498001
Rock-inhibitor, Y27632	Miltenyi Biotec	Cat# 130-106-538
Penicillin-Streptomycin	GIBCO (ThermoFisher)	Cat# 15140122
Draq7	BD Biosciences	Cat# 564904
Knockout replacement serum (KSR)	GIBCO(ThermoFisher)	Cat# 10828010
Laminin-521	Biolamina	Cat# LN521
StemMACS iPS-Brew XF, human	Miltenyi biotec	Cat# 130-104-368
StemPro Accutase	GIBCO(ThermoFisher)	Cat# A1110501
Polyethyleneimine	PEI Polysciences	Cat# PN 23966
SYBR Green I Master	Roche	Cat# 4887352001
DPBS	GIBCO(ThermoFisher)	Cat# 14190086
Critical commercial assays		
NucleoSpin PCR Cleanup and Gel Extraction	Macherey-Nagel	Cat# 740609.50
Rneasy mini kit (miniRNeasy)	QIAGEN	Cat# 74104
TruSeq Stranded mRNA LP (48 Spl)	Illumina	Cat# 20020594
Hyperprep kit (KAPA Biosystems)	Roche	Cat# 7962347001
NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Module	New England Biolabs	Cat#: E6421S for 24 reactions or E6421L for 96 reactions
Bioanalyzer High Sensitivity kit	Agilent	NA
Deposited data		
bulkRNA seq data (raw)	GEO	GSE283600
CUTnRUN data (raw)	GEO	GSE283600
Image resources	This paper	https://doi.org/10.5281/zenodo.17314
Experimental models: Cell lines		
hiPSC line (HS2), male	RIKEN	RRID:CVCL_DQ30
hiPSC line (HS1), female	RIKEN	RRID:CVCL_DQ11
HS1_L1HS-CRISPRi_g1	Adami *et al*[1]	RRID:CVCL_E8I4
HS1_L1HS-CRISPRi_g2	Adami *et al*[1]	RRID:CVCL_E8I5
HS1_L1HS-CRISPRi_LacZ	Adami *et al*[1]	RRID:CVCL_E8I6
HS2_L1HS-CRISPRi_g1	Adami *et al*[1]	RRID:CVCL_E8I7
HS2_L1HS-CRISPRi_LacZ	Adami *et al*[1]	RRID:CVCL_E8I8
Oligonucleotides		
L1-CRISPRi gRNA1 forward: CACCGGTGGAGCCCACCACAGCTCA	Adami *et al*[1]	N/A
L1-CRISPRi gRNA1 reverse: AAACTGAGCTGTGGTGGGCTCCACC	Adami *et al*[1]	N/A
L1-CRISPRi gRNA1 forward: CACCGGCGTGAGCGACGCAGAAGAC	Adami *et al*[1]	N/A
L1-CRISPRi gRNA2 reverse: AAACGTCTTCTGCGTCGCTCACGCC	Adami *et al*[1]	N/A
LacZ gRNA forward: CACCGTGCGAATACGCCCACGCGAT	Eurofins	N/A
LacZ gRNA reverse: AAACATCGCGTGGGCGTATTCGCAC	Eurofins	N/A

Recombinant DNA		
pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP	Addgene/C. Gersbach	RRID: Addgene_71237
FUGW U6 gLacZ dCas9-KRAB-T2a-GFP (LV3599)	Adami *et al*[1]	RRID: Addgene_234883
FUGW U6 gL1HSg1dCas9-KRAB-T2a-GFP (LV3824)	Adami *et al*[1]	RRID: Addgene_234882
FUGW U6 gL1HSg2dCas9-KRAB-T2a-GFP (LV3822)	Adami *et al*[1]	RRID: Addgene_234881
pMDL	Addgene; Zufferey *et al*[9]	RRID: Addgene _12251
psRev	Addgene; Zufferey *et al*[9]	RRID: Addgene _12253
pMD2G	Addgene; Zufferey *et al*[9]	RRID: Addgene_12259
Software and algorithms		
STAR (version 2.7.8a)	Dobin *et al*[10]	RRID:SCR_004463
SAMtools (version 1.16.1 and 1.18)	Li *et al*[11]	RRID:SCR_002105
Deeptools (version 2.5.4)	Ramirez *et al*[12]	RRID:SCR_016366
featureCounts (version 1.6.3)	Liao *et al*[13]	RRID:SCR_012919
Bedtools (version 2.30.0)	Quinlan & Hall[14]	RRID:SCR_006646
Homer (version 5.1)	Heinz *et al*[15]	RRID:SCR_010881

Step-by-step method details

TE-targeting gRNA design (TE subfamily level)

Timing: 1-2 days

1. Retrieve the consensus sequence for the targeted TE-family as a FASTA file from the Dfam database (https://www.dfam.org/home)[16], which includes an open collection of TE consensus sequences. In our case study, we target evolutionary young L1s (Figure 1) and designed gRNAs against a consensus sequence of the 5' regulatory region (5' UTR) obtained from the alignment of the 5' UTR of human-specific L1HS elements and the hominoid-specific L1PA2 subfamily.

Note: At this stage, an alternative option is to perform literature research to find already published gRNA designed to target the TE family or subfamily of interest.

2. Design gRNAs against the desired region within the consensus sequence.

   It is possible to proceed in one of two ways:

***Case 1*:** Manually design the desired gRNAs.

   a. Search for Protospacer Adjacent Motif (PAM) sequences (NGG for the chosen dCas9-based vector in this study) within the TE consensus sequence.
   b. Manually design 20 bp-long gRNAs targeting a sequence that is directly adjacent to the identified PAM.

Note: The PAM sequence is required for the dCas9 machinery to bind the target DNA and needs to be located at the 3' end of the DNA target sequence of choice and will be directly adjacent to the designed gRNA (Figure 3).

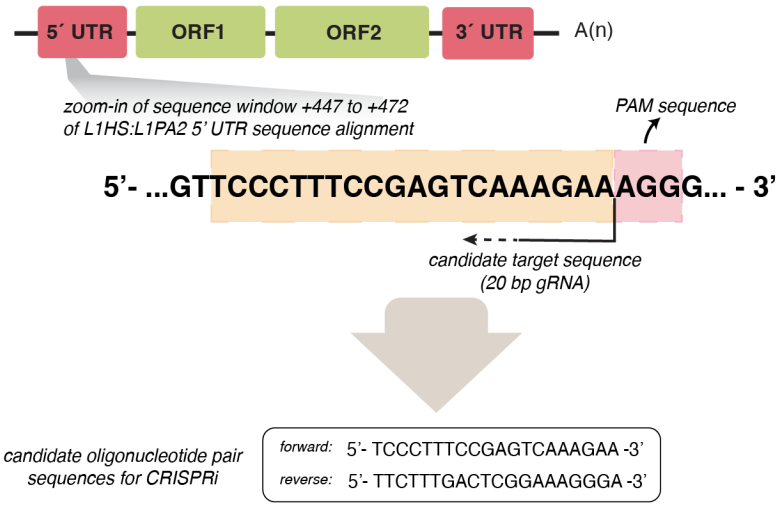


Figure 3. *Example of a manually designed gRNA (oligonucleotide pair) targeting the 5' UTR of full-length, evolutionarily young L1s.*

Case 2: Treat the consensus sequence from Dfam as a single-copy gene.

a. Input the consensus sequence retrieved from Dfam into a gRNA-designing software such as CRISPick (https://portals.broadinstitute.org/gppx/crispick/public).
b. Define the conditions of your experiment:
   i. Reference genome- Human GRCh38 (or the more recent annotation T2T[3] if it is implemented in the chosen gRNA design software/tool)
   ii. Mechanism- CRISPRi
   iii. Enzyme- SpyoCas9
c. Upload the FASTA file retrieved from Dfam with the desired consensus sequence.
d. Inspect the resulting .txt file on Excel and pick the desired candidate gRNAs based on scoring and minimum predicted off-target effect.

For more details on how to use the CRISPick tool, see also: How CRISPick Works and Johansson *et al.* (2022)[17].

Note: Given the repetitive nature of the elements, each gRNA potentially targets different elements within the chosen subfamily. It is therefore preferable to select multiple gRNAs (at least 3-4) before proceeding with the next step to have a better overview of potential targets, off-targets, and downstream functional consequences of the CRISPR-based manipulation.

Alternative: TE-targeting gRNA design (single-locus level)

Timing: 1-2 days

Depending on the research question, it might be of interest to target one single L1 element on a locus-specific level. To do so, gRNAs can be designed as ***Case 2*** above and the single locus-targeting gRNA can be ran through the BLAT tool on the UCSC Genome Browser making sure that the designed gRNA only targets one single locus in the genome.

BLAT your gRNAs (TE subfamily level)

Timing: 1 day

Given that the designed gRNAs will target multiple loci in the genome, it is important to predict *in silico* the targets for each candidate gRNA to proceed with those that present the maximum on-target effects and minimum off-targets.

3. BLAT the sequence either using the BLAT tool on the UCSC Genome Browser, or using the command-line version available for download from the UCSC servers. If using the command-line version, define the following parameters:

    a. Set a small stepSize (1-5 bps).

Note: Step sizes define the frequency at which to query the CRISPR guide into the reference genome (target). For instance, in a target sequence of length 100 bp and a query of 10 bp, a stepSize= 10 will query the sequence to the target every 10 bps, resulting in 10 comparisons. A stepSize= 5 will query the sequence to the target every 5bps, resulting in 20 comparisons.

CRITICAL: The default stepSize in BLAT is optimized for speed, not sensitivity. Thus, lowering the stepSize (e.g. to 1-5 bps) increases the chances of detecting interrupted alignments (non-perfect matches), which is especially relevant for old TE subfamilies (e.g., ERVs) where several mutations have accumulated throughout evolution, or subfamilies prone to carry polymorphisms within their sequence (e.g., SVAs).

    b. Set a large repMatch.

Note: This parameter defines the number of times the query (gRNA) is allowed to appear in the target sequence (reference genome). For TEs, this number should correspond to the minimum the size of the subfamily in question (minimum number of elements found in the chosen TE subfamily).

    c. Set a sensitive minScore.

Note: This threshold defines the minimum alignment quality for a match to be reported, and it will influence the number of hits you get reported by BLAT. The score is calculated based on the number of unique matches in the genome (+1), the number of multimapping matches (+0.5), mismatches (-1), and gaps (-1). Note that the number of multimapping matches are counted as half but are likely to inflate TE scores. For the purposes of TE analyses, where even weaker alignments are relevant, we recommend lowering the default minScore to evaluate partial matches to the target sequence. It is however important to know that setting a lower minScore can potentially lead to false positives. To address this, see Problem 1: Clear false positive hits in the BLAT analysis in the Troubleshooting section.

Taken all these parameters together, here is an example of how a BLAT command could look like:

```
blat -stepSize=1 -repMatch=2000 -minScore=15 ./genome.fa ./guide1.fa ./guide1.blat.out
```

Note: If using the online BLAT tool on the UCSC Genome Browser, stepSize, repMatch, and minScore parameters cannot be manually defined. This can limit the *in silico* predictions as the search might not be sensitive enough to pick up partial matches.

4. To interpret the BLAT results, intersect the coordinates (-a) using bedtools intersect (-wo, only hits with an overlap are reported) to (-b):

    a. **RepeatMasker:** To annotate targeted elements and evaluate the specificity to a particular subfamily or element.

Note: This is especially relevant over composite elements, where CRISPR guides could simultaneously target several subfamilies due to their evolutionary history and sequence composition. For instance, SVA-targeting guides could also target Alus and/or ERVs, since SVAs share portions of their sequence with these other TEs.

    b. **Genes' exons annotation:** To evaluate potential targeting of gene sequences (off-targets).

CRITICAL: This should be performed over exons coordinates rather than genes coordinates, to avoid intersecting with potential intronic on-target TE sequences (covered in the next step).

    c. **Genes' introns annotation:** To annotate intronic targets.

Note: This step complements the previous two, by classifying targeted elements either as intergenic or intragenic.

Taken the parameters, an example of a BEDtools intersect command could look like:

```
bedtools intersect -a guide1.blat.bed -b gencode.gtf -wo > guide1.blat.intersect.gencode.out
```

Clone the gRNAs and produce silencing lentiviral vectors

Timing: 3 weeks

5. Order the oligos containing the designed and chosen 20 bp gRNA and the correct flanking restriction sites (BsmBI in this study) and clone them in the CRISPRi plasmid backbone (we used RRID: Addgene_71237).

Forward oligo: 5' **CACCG**...20bp gRNA... 3'
Reverse oligo: 5' **AAAC**...20 bp gRNA...**C** 3'

Note: Clone one gRNA per plasmid.

CRITICAL: It is crucial to include a non-targeting gRNA (e.g. against the bacterial gene *lacZ*, not present in the human genome, or a scrambled sequence) as a negative control for all the CRISPRi experiments. Plasmids including this gRNA should be produced in parallel with the other CRISPRi plasmids. See the Key Resource Table included in the paper for the exact sequences used in this study.

6. Produce lentiviral vector from the newly cloned plasmids. Lentiviruses for the L1-targeting and non-targeting control vector were produced according to the third generation lentiviral production protocol by Zufferey *et al* (1997)[9]. Virus titration was performed via qPCR of genomic DNA from serial dilution transductions. Titers no lower than 10^8 TU/ml were obtained and used in the experiments.

L1 CRISPRi in hiPSCs via lentiviral transduction

Timing: 2 weeks

Firstly, the L1-CRISPRi hiPSCs line (and the correspondent negative control) needs to be successfully established via lentiviral transduction. For more details on hiPSCs culture, see also Grassi *et al.* (2020)[18], Johansson *et al.* (2022)[19], Adami *et al.* (2025)[1].

7. Establish the L1-targeting and the non-targeting control hiPSCs lines.
    a. Culture hiPSCs on Lam521-coated wells and change media every day using iPS-Brew or an equivalent iPSCs media. Passage cells every 3-4 days using Accutase, seeding 100 000-200 000 cells/well in a 12-wells plate with 900 µl of media.
    b. After expanding the hiPSCs (at least one passage after thawing), transduce cells:
        i. Split cells.
        ii. Seed 250 000 cells/well in a 12-wells plate (at least 1 well/condition, including the negative control).
        iii. Before the cells attach to the well, dilute each virus aliquot 1:10 in iPS-Brew (or other iPSC media) and add the equivalent of MOI 10 of the CRISPRi lentiviral vector to each well. Add **only one virus** per well.
    c. Add double media and supplement with 10 µM Rock inhibitor.
    d. Change media and supplement with Rock inhibitor until the first split, ideally 3 days post-transduction. At this point, the cells should have recovered from the lentiviral transduction.
    e. By day 7-10 after transduction, the CRISPRi system has been constitutively established and the hiPSCs can be selected. Using fluorescence-activated cell sorting (FACS), GFP+ hiPSCs can be sorted. The GFP+ cell population corresponds to the successfully transduced cells and can be collected for bulk RNA sequencing and cryopreservation:
        i. Sort 100 000-200 000 GFP+ cells via FACS.
        ii. Pellet at 500 x *g*.
        iii. Remove supernatant and snap freeze for **bulk RNA sequencing** (stranded libraries, paired end 2 x 150 bp).
        iv. Collect 100 000-200 000 cells/condition, replate them on Lam521-coated wells (12-wells plate) and expand for **cryopreservation and/or downstream experiments**.

If excessive cell death in all the conditions is observed following the lentiviral transduction, see Problem 2: Extensive cell death across all conditions after transduction in the Troubleshooting section.

If not enough GFP+ cells could be recovered, see Problem 3: Not enough GFP+ cells can be recovered after transduction in the Troubleshooting section.

Note: Selection of GFP+ cells at this point is crucial to ensure an efficient silencing of young L1s in this multi-loci-targeting CRISPRi experiment. If the cells are cryopreserved, it is preferrable to resort the GFP+ population prior to any downstream application (e.g. differentiation of CRISPRi hiPSCs into unguided cerebral organoids, as we report in Adami *et al.* (2025)[1]).

Bulk RNA sequencing analysis of L1-CRISPRi hiPSCs

Timing: 1 week

In this multiple-loci transcriptional manipulation, methods like qRT-PCR lack the specificity to analyse transcriptional changes on a locus-specific level. Thus, a bulk RNA sequencing analysis using a stringent unique mapping approach is crucial to assess the consequences of CRISPRi on evolutionarily young L1s. This step is key for any newly designed gRNA, as the *in silico* prediction of the targeted loci might differ to the experimentally observed effects.

8. Map reads using STAR aligner (we used version 2.7.8a) with the latest genome index and GENCODE annotation as guide GTF (--sjdbGTFfile):
    a. Allow a single mapping locus (--outFilterMultimapNmax 1)
    b. Set a ratio of mismatches to the mapped length of 0.03 (--outFilterMismatchNoverLmax)

Optional: It may be informative to determine the sense and antisense transcription originating from evolutionarily young L1s, which harbor a bidirectional promoter in their 5' UTR (Figure 1). One approach is to use alignment files (BAM or SAM) to generate strand-specific bigwig files using the bamCoverage function of deeptools (we used version 2.4.5). These files can then be inputted into computeMatrix and plotHeatmap deeptools functions to produce heatmaps for L1 subfamilies of interest[7]. For a more quantitative approach, alignment files can be split by strand based on SAM flags using samtools view (we used version 1.16.1), followed by generating count matrices for the evolutionary young L1 elements. This allows assessment of in-sense and antisense transcription over the elements. Additionally, one could also define custom GTF entries for regions upstream of the L1 to capture L1-ORF0-derived transcription (we defined these regions as 500bps upstream the L1 TSS in the opposite direction).

Note: This first analysis is sufficient for an initial screening of the CRISPRi gRNAs designed. Testing several multi-loci L1-targeting gRNAs will allow to select the most specific guides or the ones with the most robust effect. Depending on the research question, one can proceed to the next step with 1-3 gRNAs of choice.

H3K4me3, H3K9me3, and dCas9 CUT&RUN of L1-CRISPRi hiPSCs

Timing: 2 days (CUT&RUN protocol) + 1 week (CUT&RUN analysis)

The repetitive nature of evolutionarily young L1s makes it essential to complement bulk RNA sequencing data (as the youngest elements suffer of very low mappability) with techniques such as CUT&RUN epigenetic profiling as described before[6,20] to identify transcription (and loss of transcription upon CRISPRi) from active L1 promoters (H3K4me3 CUT&RUN). To also perform H3K9me3 and dCas9 CUT&RUN on the L1-CRISPRi hiPSCs allows a deeper investigation of both on-target and off-target effects of the targeted L1 elements.

We have shared detailed protocols for both CUT&RUN experimental workflows on protocols.io and they can be found at the following links: dx.doi.org/10.17504/protocols.io.36wgqdb83vk5/v1 (histone markers CUT&RUN) and dx.doi.org/10.17504/protocols.io.bp2l6dr5rvqe/v1 (Cas9 CUT&RUN).

9. Perform H3K4me3 and H3K9me3 CUT&RUN epigenetic profiling of L1-CRISPRi hiPSCs (including the non-targeting control cells).

Note: H3K9me3 is a histone marker found at heterochromatic sites. Performing H3K9me3 epigenetic profiling in L1-CRISPRi hiPSCs (vs control cells) allows for the assessment of H3K9me3 spread around the promoter of the targeted L1s. This analysis is essential to ensure that the CRISPRi system, which deposits H3K9me3 to silence the targeted loci, is specifically targeting the promoters of evolutionarily young L1s and not nearby gene promoters (unless the L1 promoter acts as alternative promoter to an

alternative transcriptional isoform of the gene, in which case they might overlap). By assessing this specificity, one can be sure that any downstream phenotypic effects (e.g. transcriptional changes) are a direct consequence of the L1 silencing and not due to off-targets.

Note: H3K4me3 is an active chromatin mark that can be used for both identifying transcriptionally active L1 elements and measuring on-target specificity of the CRISPRi system by ensuring the loss of this mark over targeted evolutionary young L1 elements.

10. Perform a Cas9 cross-linking CUT&RUN epigenetic profiling of L1-CRISPRi hiPSCs and control cell line.

CRITICAL: Even though most targets such as H3K9me3 and H3K4me3 work well under native condition when profiling by CUT&RUN, there are some transiently binding chromatin factors that benefit from a light cross-linking condition to improve signal recovery. The dCas9 protein is such a target where binding to chromatin must be preserved by cross-linking to the target DNA to ensure a successful outcome. To this end, cells must be cross-linked using a 0.1% formaldehyde solution (for 1 min) prior to chromatin digestion and release. More details on this protocol can also be found in the original research paper Adami *et al.* (2025)[1].

CRITICAL: Always start with a light cross-linking where needed but if that is not sufficient, a moderate cross-linking using 1% formaldehyde solution (for 1 min) could be beneficial. However, it is possible that cross-linking compromises the DNA yield. To minimize this issue, see Problem 4: Cross-linking CUT&RUN compromises DNA yield in the Troubleshooting section.

Note: A cross-linking CUT&RUN requires changes in wash condition such as addition of 0.05% SDS and 1% Triton X-100 to improve cell permeability and increase yield. Cross-linking must be reversed by adding 0.09% SDS and 0.22 mg/ml proteinase K to the supernatant containing CUT&RUN-digested DNA followed by an overnight incubation at 55 °C.

11. Analyse the CUT&RUN data.

Optional: Peak calling using tools such as HOMER (we used version 5.1) in the Cas9, H3K4me3, and H3K9me3 CUT&RUN datasets can be used to create a curated list of targets displaying the expected effects of the CRISPR targeting. Annotation of these sites can be further performed using bedtools intersect to the Repeatmasker, exon, and gene annotations.

Expected outcomes

The expected outcome for this protocol is the generation of human pluripotent stem cell lines where specific and efficient silencing of evolutionarily young, full-length L1s is established via CRISPRi. In these hiPSCs, a dCas9-KRAB-L1-targeting vector (or control dCas9-KRAB vector) is stably expressed, leading to a multi-loci transcriptional silencing of the targeted subfamily of transposable elements (Figure 4A and 4B).

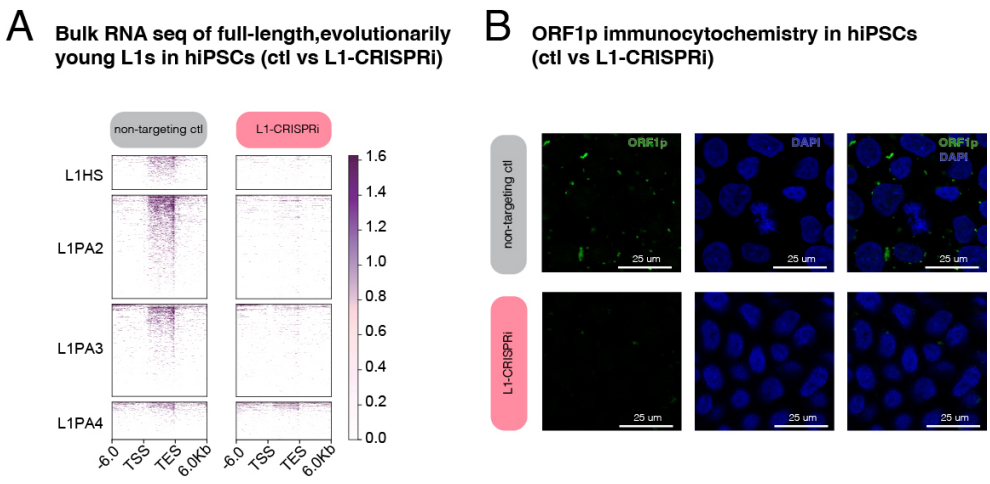


Figure 4. *A) Deeptool plot showing bulkRNA sequencing analysis (stringent unique mapping) of L1 loci targeted with L1-CRISPRi. The non-targeting control (ctl) is shown on the left, and the L1-CRISPRi on the right. B) Immunocytochemistry of ORF1p in control (non-targeting ctl) vs L1-CRISPRi hiPSCs. ORF1p is shown in green, and dapi nuclear staining in blue.*

The generated L1-CRISPRi hiPSC lines can be used for downstream analyses (e.g. differential gene expression analysis) to assess any functional consequences that the manipulation of the targeted L1 elements has on the cells. The hiPSCs can also be differentiation into other hiPSC-derived *in vitro* models to investigate more specific research questions. For instance, we differentiated our newly generated L1-CRISPRi hiPSCs into unguided cerebral organoids to investigate the consequences of evolutionarily young, full-length L1 silencing in human brain development, leveraging this *in vitro* model to describe the *cis*-regulatory role of these TEs on protein-coding genes and long non-coding RNAs in early human brain development[1].

Limitations

This protocol describes how to target evolutionarily young, full-length L1s that have an intact promoter via efficient on-target CRISPRi. One limitation is the use of one single vector per CRISPRi hiPSC line, which makes it quite labor-intensive in terms of cloning, lentiviral vector production, and cell culturing. A CRISPRi vector in which multiple gRNAs can be cloned could be used instead to reduce the production and experimental costs.

The presented framework can also be applied for the silencing of other TEs subfamilies. However, the very limited mappability of some TEs (e.g. SVAs) increases the challenges with gRNA design and downstream transcriptomic analyses, including on- and off-target analyses. Targeting a large portion of other TE subfamilies with high specificity can therefore be complicated, and additional considerations for the gRNA design and the on- and off-target analyses might be required.

Troubleshooting

Problem 1: Clear false positive hits in the BLAT analysis

We have recommended setting a low minScore in the BLAT analysis to be reported even on weak or partial matches of the guide RNA sequences and allow the identification of all potential TE targets. However, if the minScore parameter is set very low, it can lead to BLAT reporting false positives, i.e., sequences which are unlikely to be true, biologically plausible, CRISPR targets because of the very weak match with the target DNA sequence.

Potential solution:

Post filtering of the false positives could be relevant here. Important variables to explore from the BLAT results are the alignment length and percent of identity. Specifically, CRISPR gRNAs are generally 20 bp-long, and if a target sequence matches fewer than 15-16 bases of the gRNA loaded on the Cas9 machinery, it is highly unlikely for the CRISPR system to bind the target DNA. Noteworthy, as the binding of the dCas9 to the target DNA sequence is more tolerant of mismatches compared to the cleaving CRISPR/Cas9 system, a sequence with a mismatch up to 5-6 bases could still be a target for the CRISPRi system. Therefore, BLAT hits with a sequence match of less than 15-16 bases and a mismatch of more than 5-6 bases (out of the 20 bp length of the gRNA) are most likely false positives that can be filtered out. Finally, if a gRNA matches to a genomic location where there is no Protospacer Adjacent Motif (PAM) immediately downstream of the gRNA target site, this is an unequivocable false positive that can be excluded.

Problem 2: Extensive cell death across all conditions after transduction

If cell death is observed across all conditions (including the non-targeting control cell line) after lentiviral transduction, this might be a consequence of a viral load too high for the transduced cells to sustain or an impure viral preparation that leads to cytotoxicity.

Note: Cell death differences across conditions can be a phenotypical consequence of the specific experiment, but if the control cells also die and do not recover after 2-3 days post transduction, that is linked to the toxicity of the viral transduction itself.

Potential solution:

To solve the problem of a too high viral load, one can try to use lower MOIs and test a few (Figure 5), especially if testing a lentiviral vector for the first time. If the transduction with highest MOI presents extensive cell death across conditions and cannot be selected for successfully transduced cells, then the other MOIs can be kept in culture, selected for transduction (e.g. GFP+ cell population), and finally tested for the silencing.

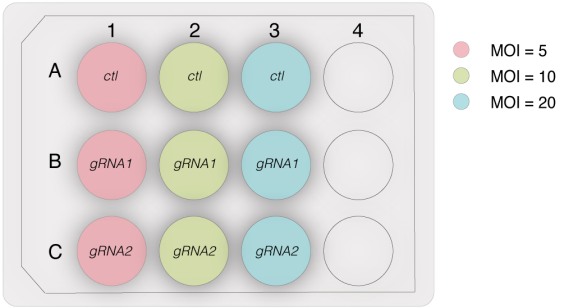


Figure 5. *Schematic of a 12-well culture plate displaying an example of different MOIs used to test the lentiviral vector. Ctl= control.*

If the problem lies in an impure viral preparation, where contaminants such as cell debris and endotoxins are found in the viral suspension, an optimisation of the plasmid purification and virus production workflow is necessary. It is important to work with high-quality and endotoxin-free plasmid DNA as well as adopt proper filtering and centrifugation techniques when collecting the packaged viral particles.

Problem 3: Not enough GFP+ cells can be recovered after transduction

If a low percentage of GFP+ cells is recovered during FACS, this could be due to a low virus titre or low efficiency of the lentiviral transduction.

Potential solution:

A high titre of the lentiviral vector is essential to achieve high transduction efficiency, which can be hindered by the presence of unpackaged/free RNA molecules and non-infectious particles in the viral preparation. This leads to a low effective MOI of the viral preparation, with a drastic reduction in transduction and integration efficiency. It is once again essential to optimize the lentiviral production workflow: specifically, optimized concentration and harvesting methods are essential to obtain high quality and highly concentrated viral particles.

A low transduction efficiency can also occur when the cells are too confluent at the time of transduction. This can lead to uneven transduction across the cells, as well as to a reduction in their metabolic activity, which can prevent optimal transduction. To avoid this, it is crucial to transduce cells which are not overly confluent. An optimal confluency is generally ~70%, but this can vary (and must therefore be optimized) in different cell types. To transduce cells that have been split only once or twice after thawing can also improve the transduction efficiency.

Problem 4: Cross-linking CUT&RUN compromises DNA yield

Adding a cross-linking step could be essential when performing CUT&RUN on reversibly bound proteins such as, in this study, Cas9. However, cross-linking might lead to lower DNA recovery.

Potential solution:

To maximize the DNA recovery, it is crucial to handle the sample with extreme care during the cross-linked steps. Loss of DNA could also occur during the washing steps, and to prevent this, it is possible to slightly increase the centrifugation force (e.g. up to 1200 x g). Finally, a higher number of cells as the initial sample input can improve DNA recovery (e.g. 1 000 000 cells per condition).

Resource availability

Lead contact

Requests for further information, resources, and reagents should be directed to and will be fulfilled by the lead contact, Johan Jakobsson (johan.jakobsson@med.lu.se).

Materials availability statement

This study did not generate new unique reagents. Plasmids and cell lines generated in relation to the original research study (Adami *et al.*, 2025[1]) have been registered on Addgene or Cellosaurus, respectively, are reported in the Key Resource Table included in this publication and are available upon request from the lead contact.

Data and code availability statement

No original data or code was generated in this publication.

All the data or code presented in this work have been published in the original research paper Adami *et al.*, 2025[1]. The sequencing data discussed in the publication have also been previously deposited at the GEO superseries GEO: GSE283600. The code in relation to the original research publication has been deposited on GitHub and is publicly available at https://molecular-neurogenetics.github.io/truster/ (DOI: 10.5281/zenodo.14362613) and git@github.com:Molecular-Neurogenetics/L1CRISPR_Adami_2025.git (DOI: 10.5281/zenodo.15638389).

The image data presented in this publication have been deposited to Zenodo at doi: 10.5281/zenodo.17314543.

Data, code, protocols, and key lab materials used in this study are listed in the Key Resource Table available within this article and deposited to Zenodo at doi: 10.5281/zenodo.17435054.

Acknowledgments

This study was partially funded by the joint efforts of the Michael J. Fox Foundation for Parkinson's Research (MJFF) and the Aligning Science Across Parkinson's (ASAP) initiative. MJFF administers the grants ASAP-000520, ASAP-024296, and ASAP-025170 on behalf of ASAP and itself. The work was also supported by grants from the Swedish Research Council (2022-02673 to J.J. and 2021-03494 to C.H.D.), the Swedish Brain Foundation (FO2023-0232 to J.J.), Cancerfonden (222185Pj to J.J.), Barncancerfonden (PR2023-0099 to J.J.), the Swedish Society for Medical Research (S19-0100 to C.H.D.), and the Swedish Government Initiative for Strategic Research Areas (MultiPark & StemTherapy).

Author contributions

Conceptualization by A.A., R.G., and J.J.; experimental work was performed by A.A. and F.D.; R.G. executed the bioinformatical analyses; A.A., R.G., and F.D. worked on the original draft of the manuscript; all authors contributed to the review and editing of the publication.

Declaration of interests

The authors have no competing interests to declare.

References

1 Adami, A. *et al.* LINE-1 retrotransposons mediate cis-acting transcriptional control in human pluripotent stem cells and regulate early brain development. *Cell Genom*, 100979 (2025). https://doi.org/10.1016/j.xgen.2025.100979

2 de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* **7**, e1002384 (2011). https://doi.org/10.1371/journal.pgen.1002384

3 Hoyt, S. J. *et al.* From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science* **376**, eabk3112 (2022). https://doi.org/10.1126/science.abk3112

4 Nair, V. P., Mayer, J. & Vincendeau, M. A protocol for CRISPR-mediated activation and repression of human endogenous retroviruses in human pluripotent stem cells. *STAR Protoc* **3**, 101281 (2022). https://doi.org/10.1016/j.xpro.2022.101281

5 Thakore, P. I. *et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* **12**, 1143-1149 (2015). https://doi.org/10.1038/nmeth.3630

6 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6** (2017). https://doi.org/10.7554/eLife.21856

7 Garza, R. *et al.* LINE-1 retrotransposons drive human neuronal transcriptome complexity and functional diversification. *Sci Adv* **9**, eadh9543 (2023). https://doi.org/10.1126/sciadv.adh9543

8 Philippe, C. *et al.* Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *Elife* **5** (2016). https://doi.org/10.7554/eLife.13926

9 Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat Biotechnol* **15**, 871-875 (1997). https://doi.org/10.1038/nbt0997-871

10 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013). https://doi.org/10.1093/bioinformatics/bts635

11 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009). https://doi.org/10.1093/bioinformatics/btp352

12 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191 (2014). https://doi.org/10.1093/nar/gku365

13 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014). https://doi.org/10.1093/bioinformatics/btt656

14 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010). https://doi.org/10.1093/bioinformatics/btq033

15 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010). https://doi.org/10.1016/j.molcel.2010.05.004

16 Storer, J., Hubley, R., Rosen, J., Wheeler, T. J. & Smit, A. F. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA* **12**, 2 (2021). https://doi.org/10.1186/s13100-020-00230-y
17 Johansson, P. A., Adami, A. & Jakobsson, J. CRISPRi-mediated transcriptional silencing in iPSCs for the study of human brain development. *STAR Protoc* **3**, 101285 (2022). https://doi.org/10.1016/j.xpro.2022.101285
18 Grassi, D. A. *et al.* Profiling of lincRNAs in human pluripotent stem cell derived forebrain neural progenitor cells. *Heliyon* **6**, e03067 (2020). https://doi.org/10.1016/j.heliyon.2019.e03067
19 Johansson, P. A. *et al.* A cis-acting structural variation at the ZNF558 locus controls a gene regulatory network in human brain development. *Cell Stem Cell* **29**, 52-69 e58 (2022). https://doi.org/10.1016/j.stem.2021.09.008
20 Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006-1019 (2018). https://doi.org/10.1038/nprot.2018.015