**Direct insertional mutagenesis by L1 resulted in diseases including muscular dystrophy, hemophilia, and breast cancer**

**Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man.**
Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE.
We now report insertions of L1 elements into exon 14 of the factor VIII gene in two of 240 unrelated patients with haemophilia A. Both of these insertions (3.8 and 2.3 kilobases respectively) contain 3' portions of the L1 sequence, including the poly (A) tract, and create target site duplications of at least 12 and 13 nucleotides of the factor VIII gene.

- **Characterization of a nondeleterious L1 insertion in an intron of the human factor VIII gene** Woods-Samuels P, Wong C, Mathias SL, Scott AF, Kazazian HH Jr, Antonarakis SE.

    - **A 20.7 kb deletion within the factor VIII gene associated with LINE-1 element insertion.**

      Van de Water N, Williams R, Ockelford P, Browett P.

Table 2 L1 EN-mediated retrotranspositions associated with human genetic diseas

| Disrupted gene[a] | Chromosomal location | Disorder[b] | Inserted element | Insertion size (bp) | |
|---|---|---|---|---|---|
| *Simple insertions* | | | | | |
| *APC* | 5q | Colon cancer | L1 Ta | 520 | |
| *CHM* | Xq | Choroideremia | L1 Ta | 6,017 | |
| *CYBB* | Xp | CGD | L1 Ta | 836 | |
| *CYBB* | Xp | CGD | L1 Ta | 1,722 | |
| *DMD* | Xp | DMD | L1 Ta | 1,400 | |
| *DMD* | Xp | XLDCM | L1 Ta | 530 | |
| *F8* | Xq | Haemophilia A | L1 Ta | 3,800 | |
| *F8* | Xq | Haemophilia A | L1 preTa | 2,300 | |
| *F9* | Xq | Haemophilia B | L1 Ta | 463 | |
| *F9* | Xq | Haemophilia B | L1 Ta | 163 | |
| *HBB* | 11p | β-Thalassemia | L1 Ta | 6,000 | |
| *RP2* | Xp | XLRP | L1 Ta | 6,000 | |
| *RPS6KA3* | Xp | CLS | L1 HS | 2,800 | |
| *APC* | 5q | Desmoid tumor | *Alu*Yb8 | 278 | |
| *BCHE* | 3q | Acholinesterasemia | *Alu*Yb9 | 289 | |
| *BRCA2* | 13q | Breast cancer | *Alu*Yc1 | 281 | |
| BTK | Xq | XLA | *Alu*Y | –[e] | |

# Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes

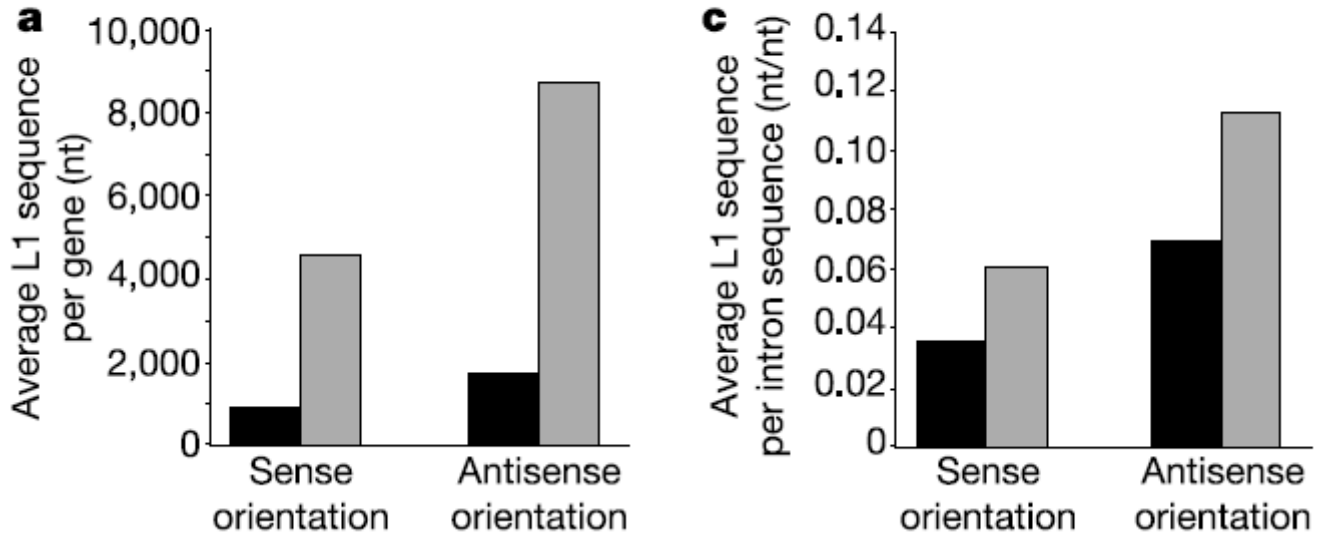
**Jeffrey S. Han[1], Suzanne T. Szak[2] & Jef D. Boeke[1]**

[1]*Department of Molecular Biology and Genetics and High Throughput Biology Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA*
[2]*Biogen, Inc., Cambridge, Massachusetts 02142, USA*

**LINE-1 (L1) elements are the most abundant autonomous retrotransposons in the human genome, accounting for about 17% of human DNA. The L1 retrotransposon encodes two proteins, open reading frame (ORF)1 and the ORF2 endonuclease/reverse transcriptase. L1 RNA and ORF2 protein are difficult to detect in mammalian cells, even in the context of overexpression systems. Here we show that inserting L1 sequences on a transcript significantly decreases RNA expression and therefore protein expression. This decreased RNA concentration does not result from major effects on the transcription initiation rate or RNA stability. Rather, the poor L1 expression is primarily due to inadequate transcriptional elongation. Because L1 is an abundant and broadly distributed mobile element, the inhibition of transcriptional elongation by L1 might profoundly affect expression of endogenous human genes. We propose a model in which L1 affects gene expression genome-wide by acting as a 'molecular rheostat' of target genes. Bioinformatic data are consistent with the hypothesis that L1 can serve as an evolutionary fine-tuner of the human transcriptome.**

# L1 intragenica ed espressione genica

High Low



# L1 intragenica ed espressione genica

High Low

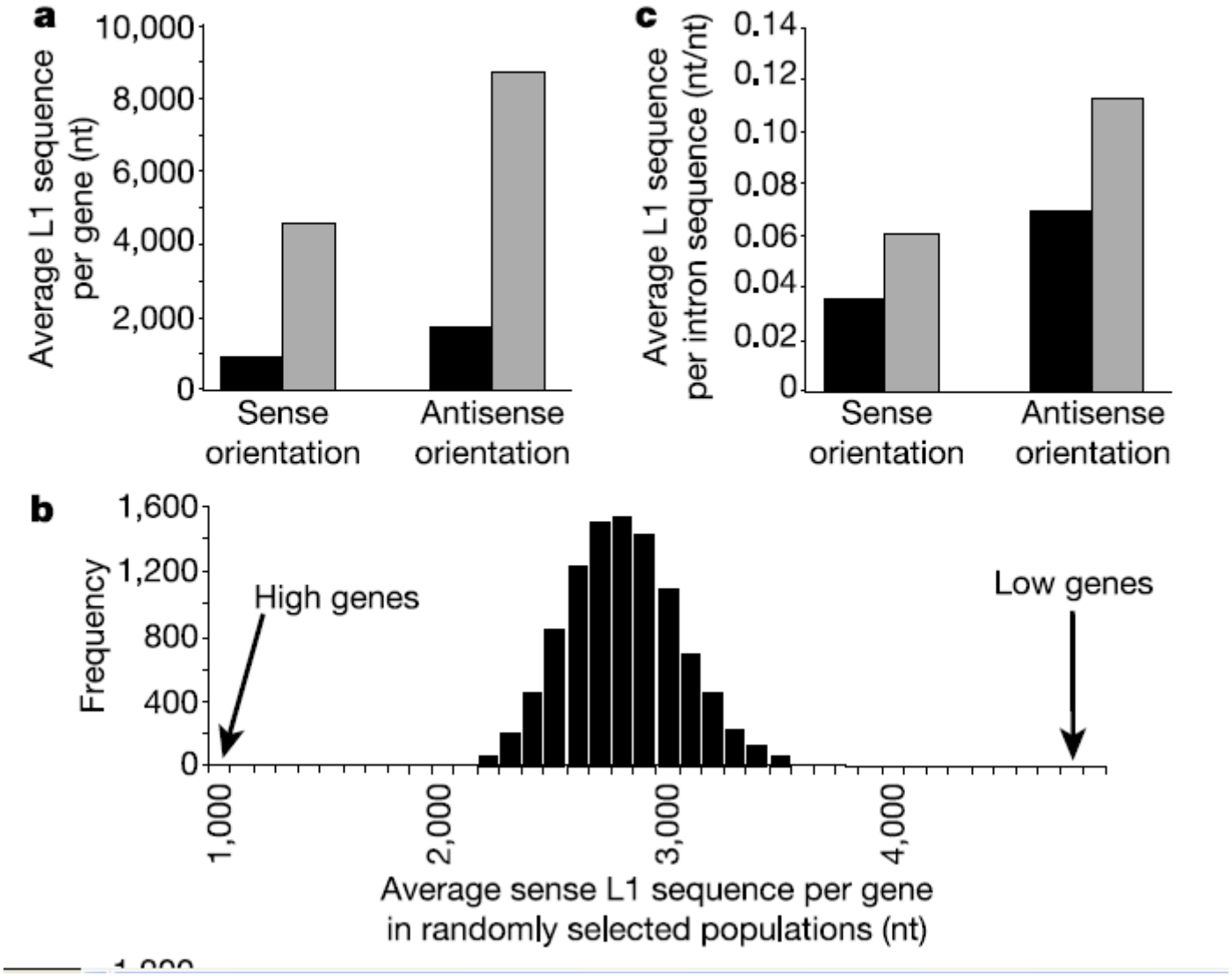
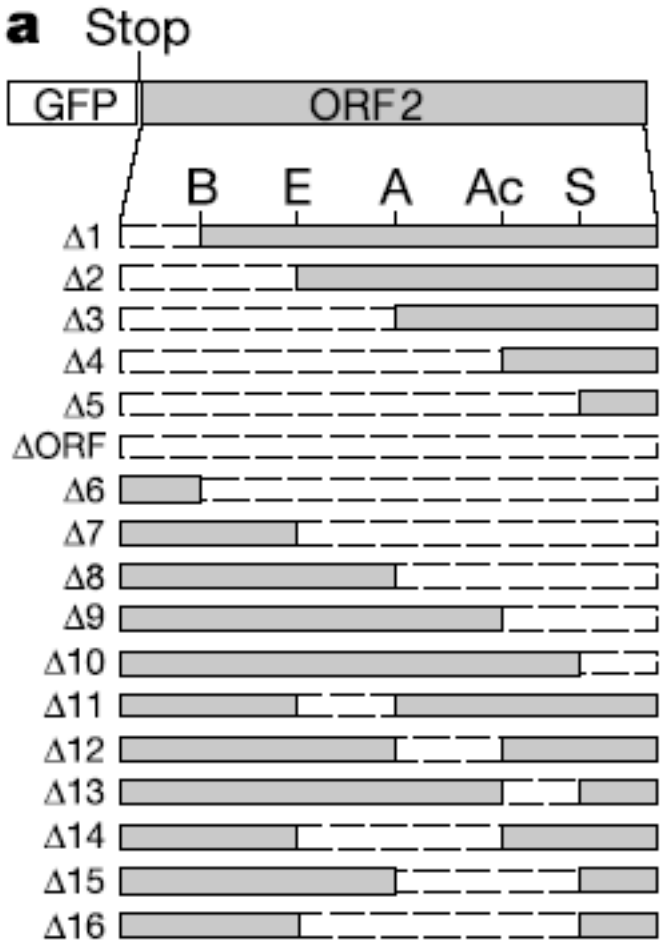
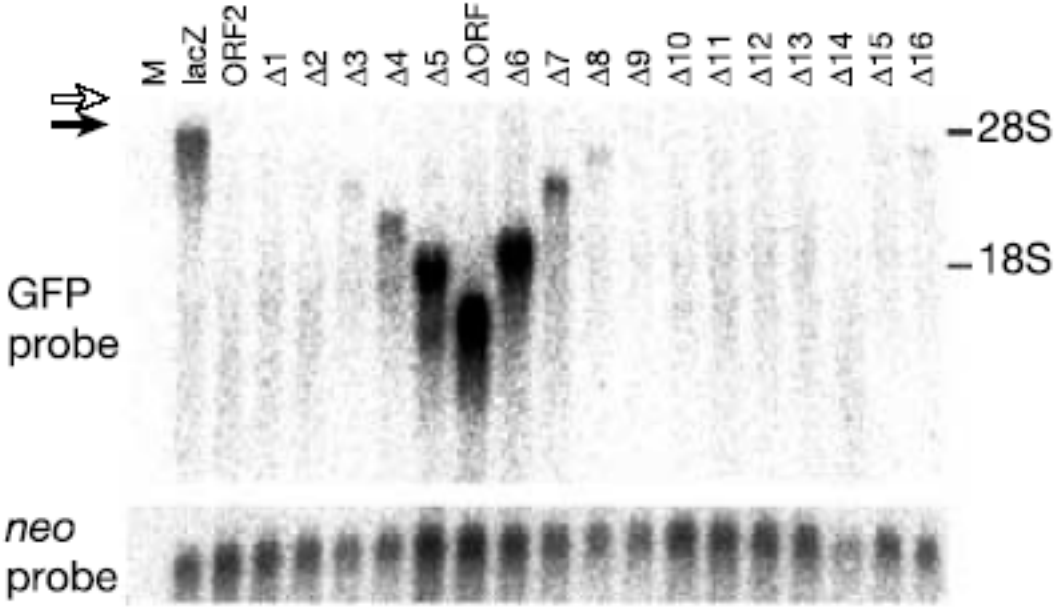


**Figure 5** Bioinformatic analysis of L1 content in genes. **a**, Average L1 content of genomic loci of sets of highly (black bars) and poorly (grey bars) expressed genes (see Methods). **b**, Average L1 content in sets of randomly selected populations of genes (see Methods). Positions where the highly and poorly expressed genes would be (data superimposed from **a**) are indicated and are outside the random distribution ($P < 0.01$). **c**, Data from **a**, normalized to total intron content. **d**, Highly and poorly expressed genes were sorted into high GC, low L1 isochore or low GC, high L1 isochore[50] classes. The percentage of each expression class falling into each isochore is indicated. Subpopulations were analysed as described in **a** and **c**.

L1 mutants

RNA analysis




Decrease in expression depends on L1 length

**Figure 3** Decrease in L1 expression is dependent on length. **a**, The left panel depicts the structures of deletion constructs. Hollow regions represent deleted sequences. B, *Bbv*CI; E, *Eco*RI; A, *Afl*II; Ac, *Acl*I; S, *Spe*I. The right panel shows a total RNA analysis of HeLa transfections. Lanes: M, mock; lacZ, pGFPstoplacZ; ORF2, pGFPstopORF2. Open and black arrows show the expected positions of GFPstopORF2 and GFPstoplacZ, respectively. **b**, The adenosine base composition of the sense strand, in 50-nucleotide windows, was plotted for each position in L1.2 with MacVector 6.5.3 (Oxford Molecular). **c**, The top panel shows the structures of GFPstopORF1, GFPstop4ORF1 and GFPstop5UTR. The 4ORF1 repeat is about 4,500 nucleotides long and the 5′ UTR repeat is about 4,000 nucleotides long. The bottom panel shows a total RNA analysis of HeLa transfections. Open, black and grey arrows show the expected positions of GFPstop4ORF1, GFPstop5UTR and GFPstopORF1, respectively.

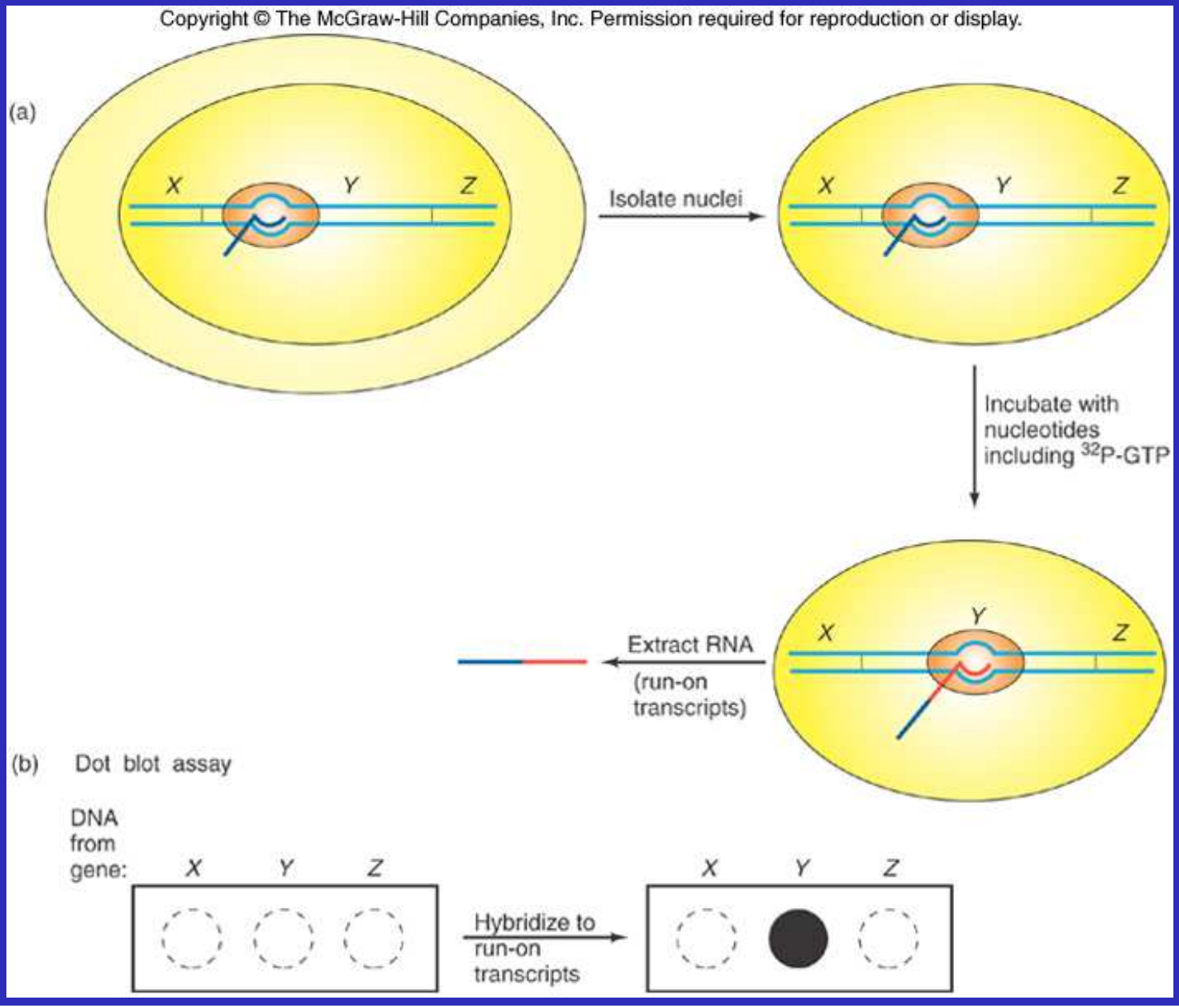
# Nuclear Run-On Transcription

- Isolate nuclei from cells, allow them to extend in vitro the transcripts already started in vivo in a technique called run-on transcription
- RNA polymerase that has already initiated transcription will "run-on" or continue to elongate same RNA chains
- Effective as initiation of new RNA chains in isolated nuclei does not generally occur

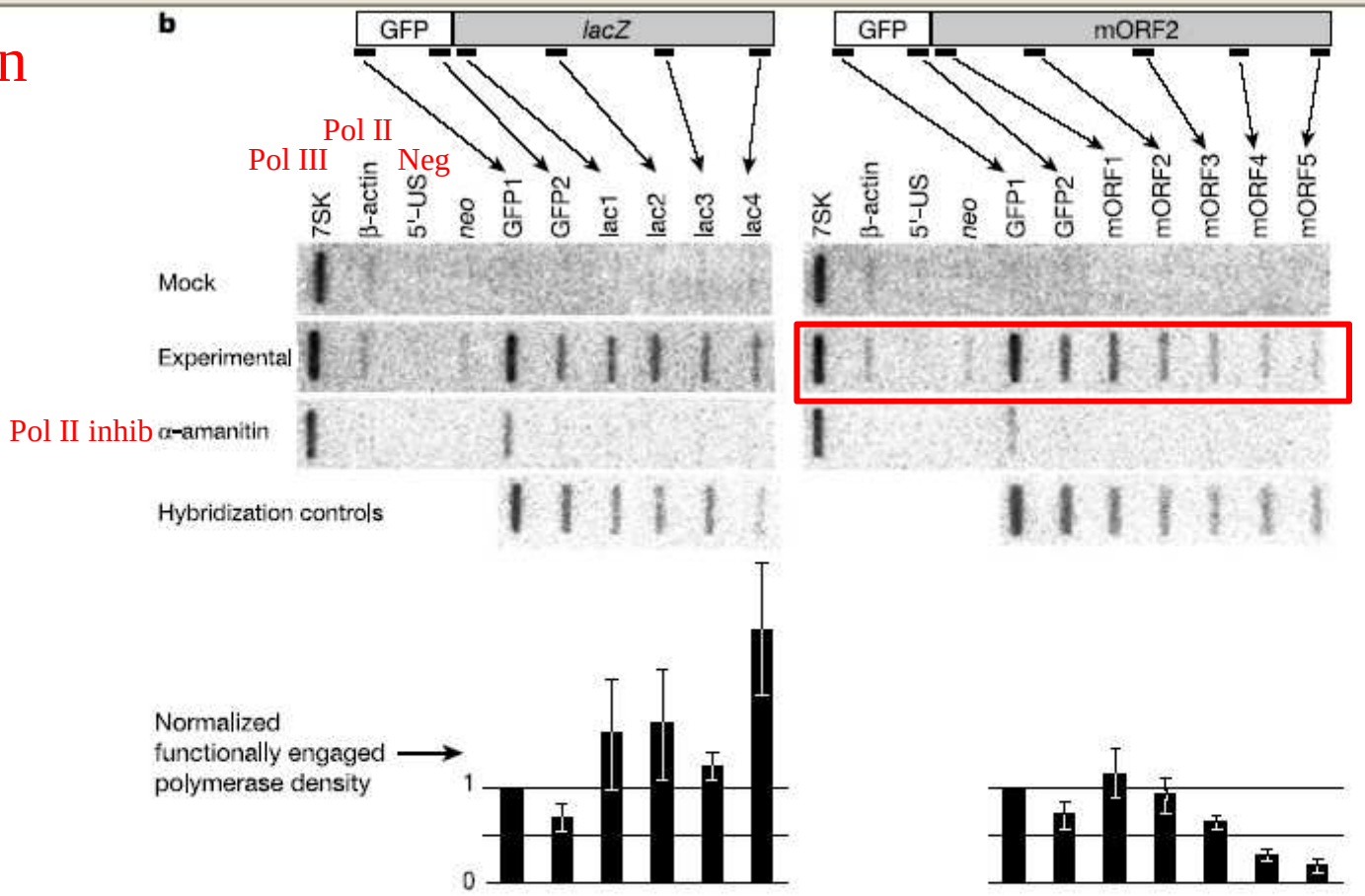
# Run-On Analysis

- Results will show transcription rates and an idea of which genes are transcribed
- Identification of labeled run-on transcripts is best done by dot blotting
  - Spot denatured DNAs on a filter
  - Hybridize to labeled run-on RNA
  - Identify the RNA by DNA to which it hybridizes
- Conditions of run-on reaction can be manipulated with effects of product can be measured

# Nuclear Run-On Transcription Diagram



# Run-on




positions. 7SK controls for RNA polymerase III transcription. β-actin is a control for RNA polymerase II transcription. 5′-US is a negative control that hybridizes to a region upstream of the cytomegalovirus (CMV) promoter. *neo* controls for transfection. Hybridization controls are described in Methods. Normalized functionally engaged polymerase density is the signal ($N = 3$) corrected for α-amanitin-resistant transcription and hybridization efficiency, with GFP1 set to 1. Error bars show the standard deviation.

**b**, Nuclear run-on analysis (NRO). Nuclei were isolated from HeLa cells 36 h after transfection and used for NRO. Bold lines under GFP, *lacZ* and mORF2 indicate probe

α-Amanitin is an inhibitor of  RNA pol II
This mechanism makes it a deadly toxin.
α-Amanitin can also be used to determine which types of RNA polymerase are present. This is done by testing the sensitivity of the polymerase in the presence of α-amanitin. RNA polymerase I is insensitive, RNA pol II is highly sensitive, and RNA pol III is slightly sensitive.

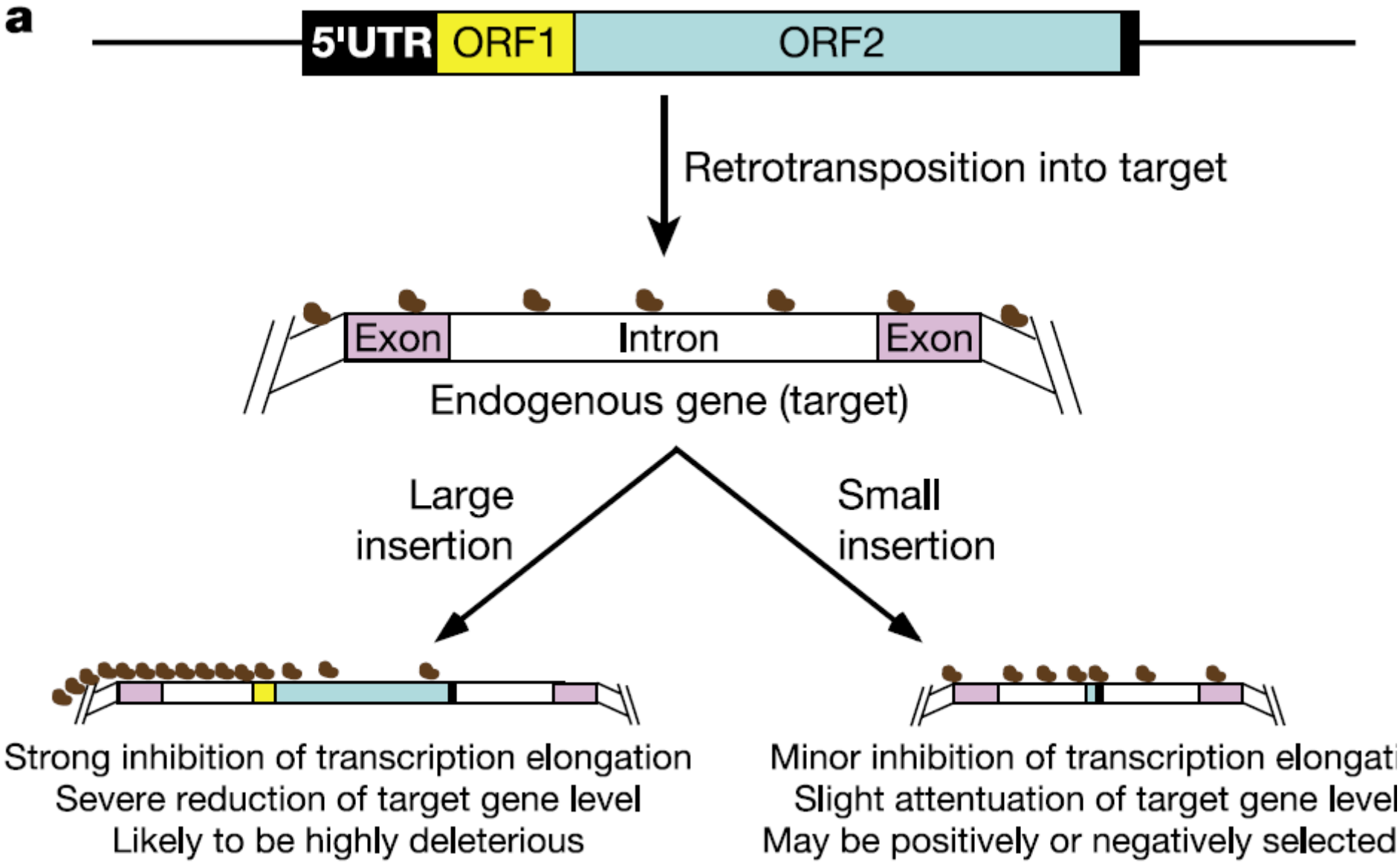

a
5'UTR
ORF1
ORF2
Retrotransposition into target
Exon
Intron
Exon
Endogenous gene (target)
Large insertion
Small insertion
Strong inhibition of transcription elongation
Severe reduction of target gene level
Likely to be highly deleterious
Minor inhibition of transcription elongati
Slight attentuation of target gene level
May be positively or negatively selected

**Figure 6** Models for L1-mediated modulation of gene expression/structure. **a**, Effects on transcription. Brown dots represent transcriptional complexes, which could be slowed, paused or dissociated from the templates on encountering significant amounts of L1 sequence. **b**, Effects on mRNA and protein structure. Left, hypothetical gene with three exons. Middle, intronic sense L1 insertions can produce a minor amount of prematurely polyadenylated mRNA, potentially giving rise to a truncated protein with additional, previously untranslated amino acids at the C terminus (white segment). Right, intronic antisense L1 insertions can produce a major amount of prematurely polyadenylated mRNA.

# A highly active synthetic mammalian retrotransposon

**Jeffrey S. Han & Jef D. Boeke**

*Department of Molecular Biology and Genetics and High Throughput Biology Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA*

LINE-1 (L1) elements are retrotransposons that comprise large fractions of mammalian genomes[1]. Transcription through L1 open reading frames is inefficient owing to an elongation defect[2], inhibiting the robust expression of L1 RNA and proteins, the substrate and enzyme(s) for retrotransposition[3–5]. This elongation defect probably controls L1 transposition frequency in mammalian cells. Here we report bypassing this transcriptional defect by synthesizing the open reading frames of L1 from synthetic oligonucleotides, altering 24% of the nucleic acid sequence without changing the amino acid sequence. Such resynthesis led to greatly enhanced steady-state L1 RNA and protein levels. Remarkably, when the synthetic open reading

Italian Ministry for University and Research and Consiglio Nazionale delle Ricerche to P.P.D.F. We thank E. Frittoli for technical assistance. L.L. was supported in part by a fellowship from the Associazione Sviluppo Piemonte.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to P.P.D.F. (difiore@ifom-firc.it).

# A highly active synthetic mammalian retrotransposon

**Jeffrey S. Han & Jef D. Boeke**

*Department of Molecular Biology and Genetics and High Throughput Biology Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA*

LINE-1 (L1) elements are retrotransposons that comprise large fractions of mammalian genomes[1]. Transcription through L1 open reading frames is inefficient owing to an elongation defect[2], inhibiting the robust expression of L1 RNA and proteins, the substrate and enzyme(s) for retrotransposition[3–5]. This elongation defect probably controls L1 transposition frequency in mammalian cells. Here we report bypassing this transcriptional defect by synthesizing the open reading frames of L1 from synthetic oligonucleotides, altering 24% of the nucleic acid sequence without changing the amino acid sequence. Such resynthesis led to greatly enhanced steady-state L1 RNA and

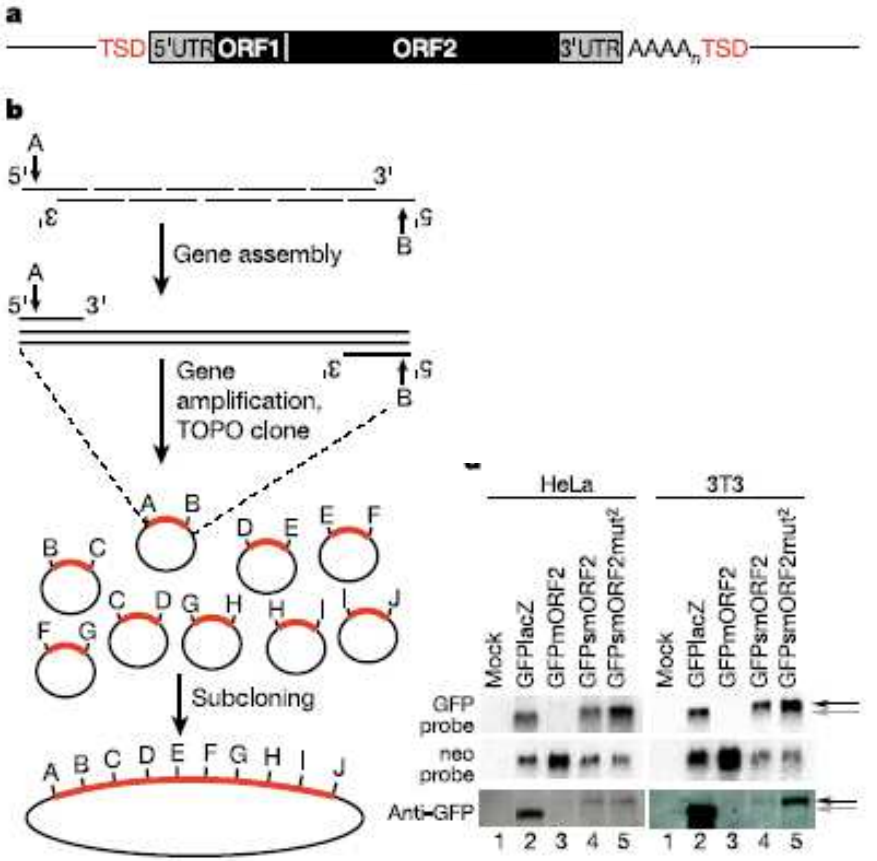
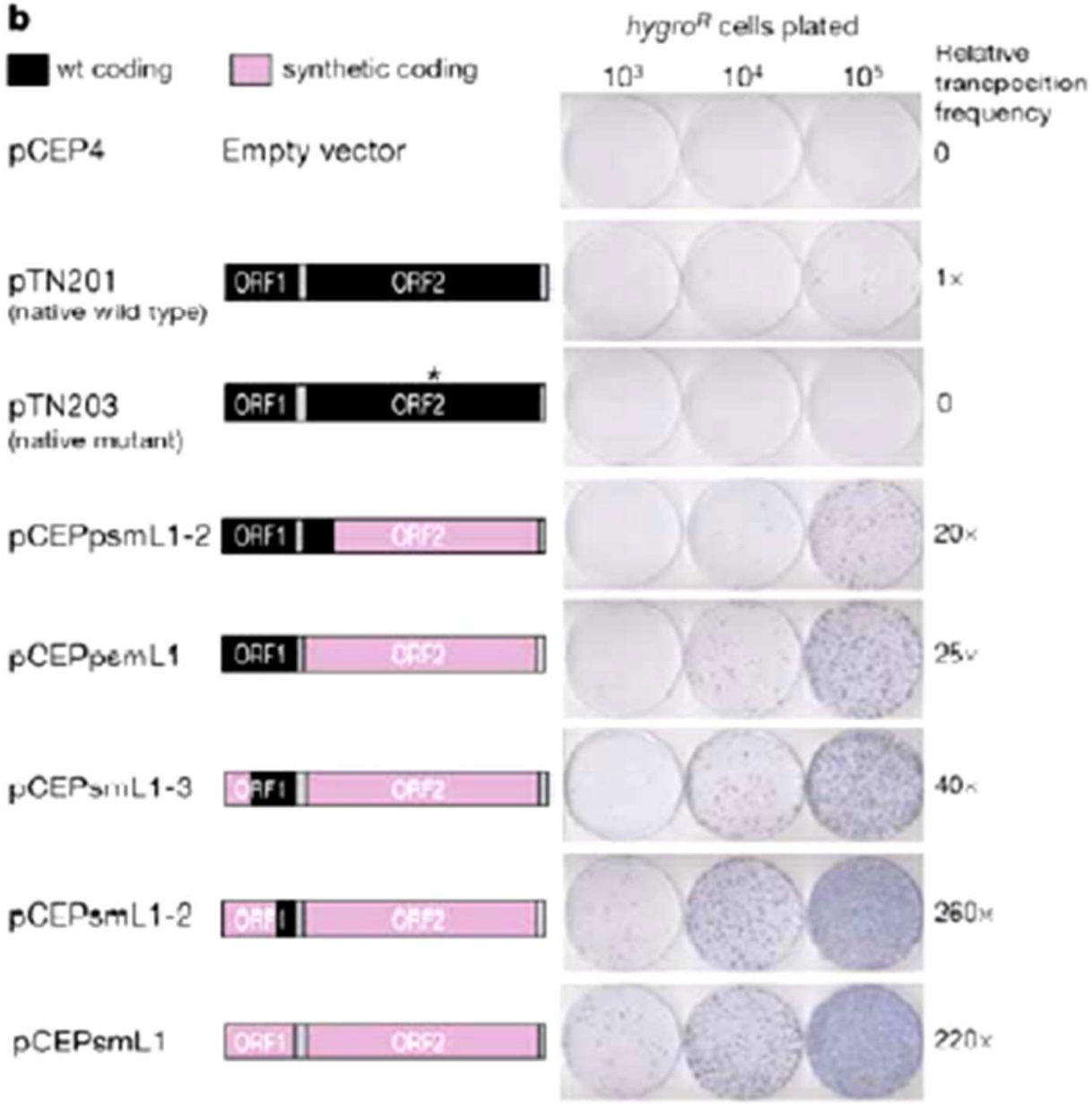



**Figure 1** Synthesis and expression of synthetic mouse ORF2. **a**, L1 structure. TSD, target site duplication; UTR, untranslated region. **b**, Overview of gene synthesis. Oligonucleotides encoding each fragment were mixed in a PCR assembly reaction and subsequently used as template amplification. Amplification products were cloned and ligated together with unique restriction sites (labelled A to J). **c**, Plasmid structures. The test sequences (*lacZ*, mORF2 or smORF2) are fused, in frame, downstream of the GFP ORF. An independent *neo* transcript is used to monitor transfection efficiency and loading. Blue lines represent probes used in **d**. **d**, Analysis of smORF2 expression. Top, RNA


b
wt coding
synthetic coding
hygroR cells plated
10³
10⁴
10⁵
Relative transposition frequency
pCEP4
Empty vector
0
pTN201
(native wild type)
ORF1
ORF2
1×
pTN203
(native mutant)
ORF1
ORF2
0
pCEPpsmL1-2
ORF1
ORF2
20×
pCEPpsmL1
ORF1
ORF2
25×
pCEPsmL1-3
ORF1
ORF2
40×
pCEPsmL1-2
ORF1
ORF2
260×
pCEPsmL1
ORF1
ORF2
220×

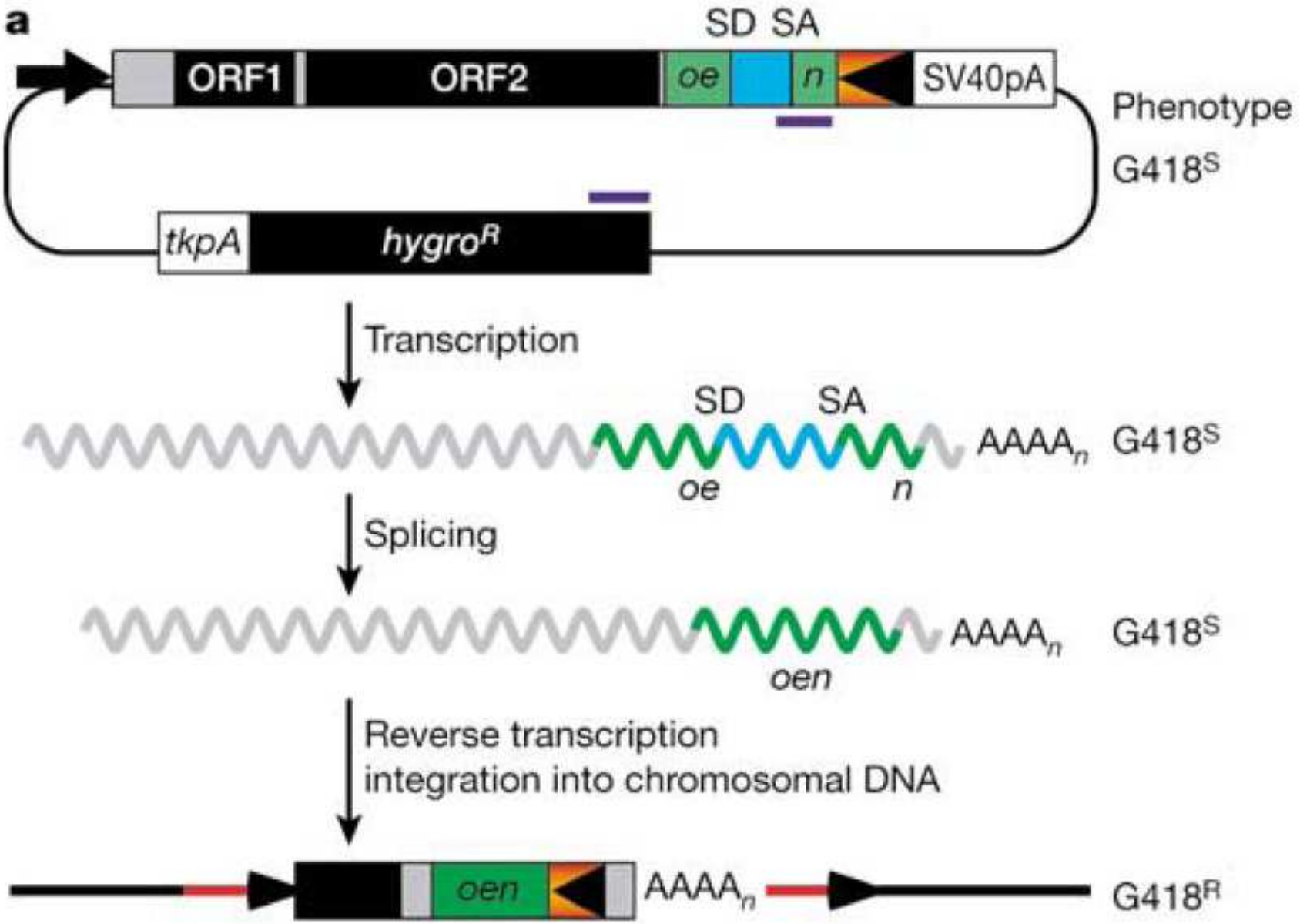

a
SD SA
ORF1
ORF2
oe
n
SV40pA
Phenotype
G418S
tkpA
hygroR
Transcription
SD
SA
AAAAn
G418S
oe
n
Splicing
AAAAn
G418S
oen
Reverse transcription
integration into chromosomal DNA
oen
AAAAn
G418R

**Figure 2** Retrotransposition of synthetic mL1. **a**, The retrotransposition assay. The L1 element contains an intron-interrupted *neo* reporter in the 3′ untranslated region with its own promoter and polyadenylation signal. Only when *neo* is transcribed from the L1 promoter, spliced, reverse transcribed and integrated into the genome does a cell become G418-resistant[4]. Blue lines represent probes for RNA analysis (Fig. 4). SD, splice donor; SA, splice acceptor. **b**, Retrotransposition was assayed in HeLa cells ($N = 3$). pTN201 contains only wild-type native mouse L1 sequence, and pTN203 contains wild-type native mouse L1 sequence with a D709Y reverse transcriptase point mutation[22]. The average absolute number of colonies for pTN201 was 440 events per $10^6$ transfected cells.

# SVA

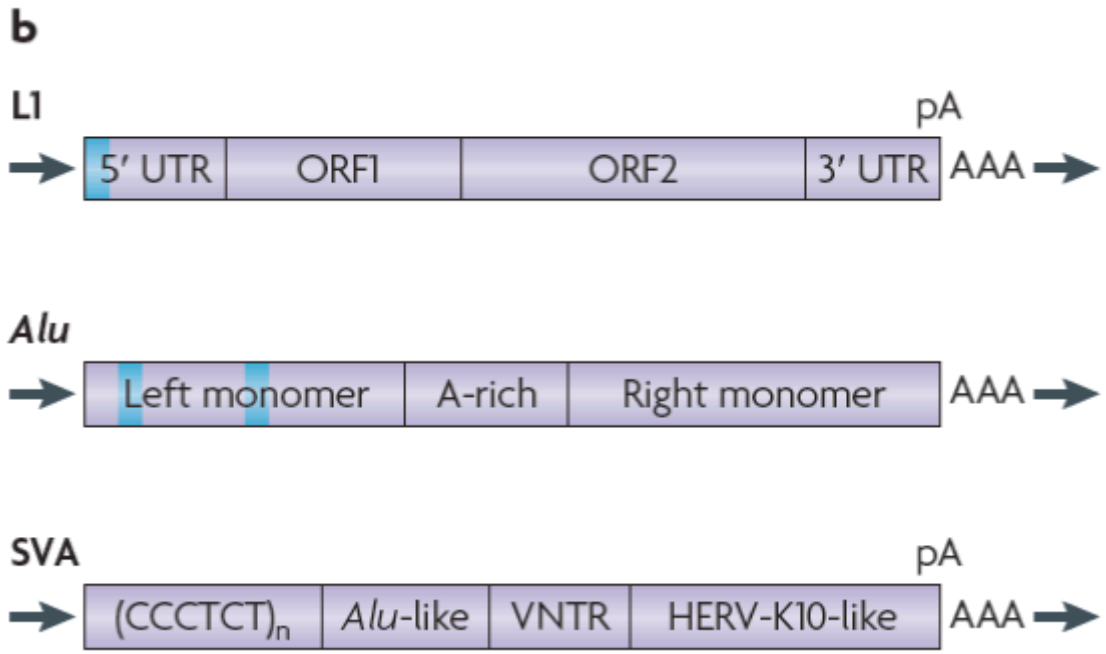
SINE-VNTR-Alu (SVA) elements are nonautonomous, hominid-specific non-LTR retrotransposons and distinguished by their organization as composite mobile elements.

They represent the evolutionarily youngest, currently active family of human non-LTR retrotransposons

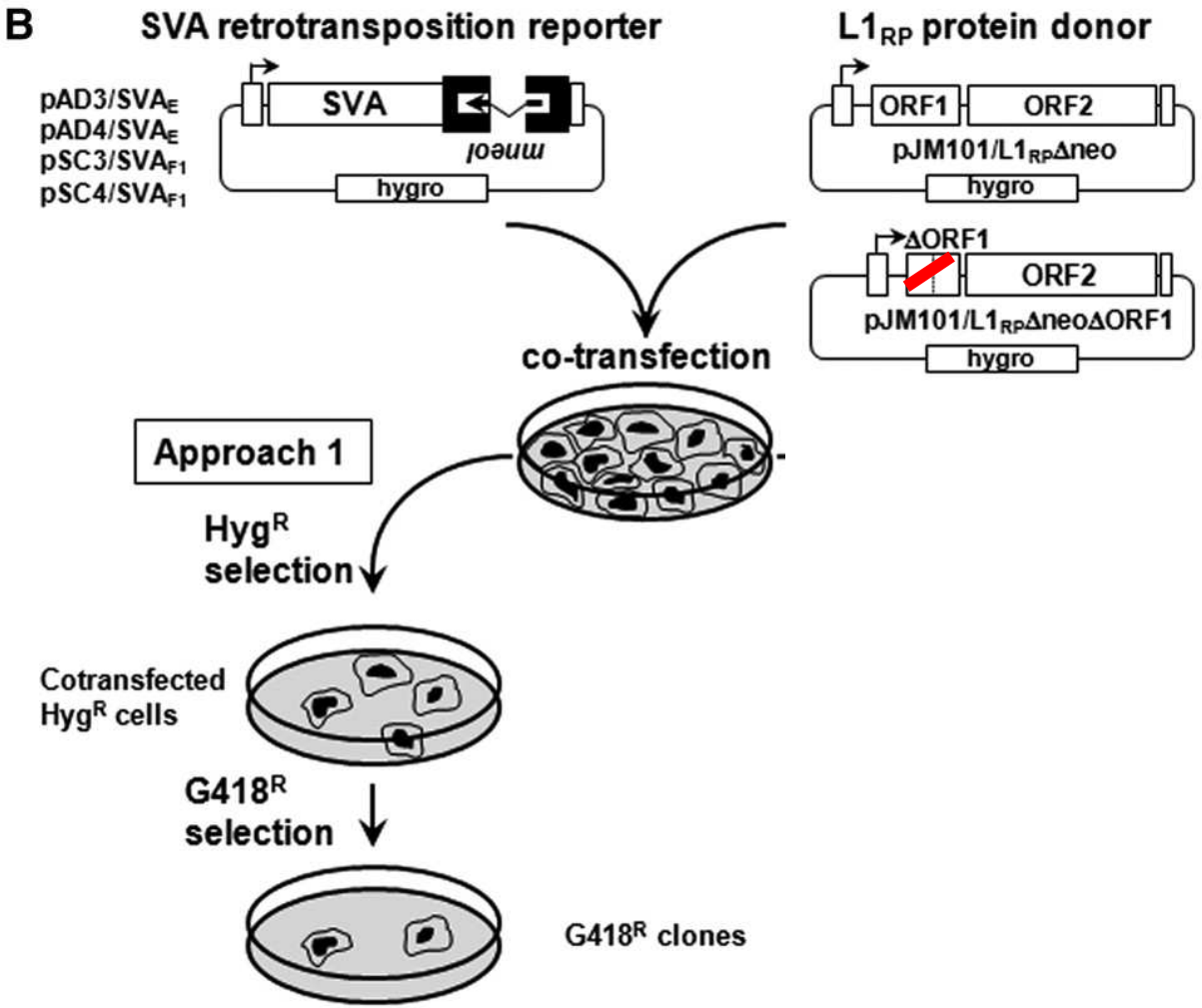
# *Alu* elements as regulators of gene expression

Julien Häsler and Katharina Strub*

*Alu* elements are the most abundant repetitive elements in the human genome; they emerged 65 million years ago from a 5′ to 3′ fusion of the 7SL RNA gene and amplified throughout the human genome by retrotransposition to reach the present number of more than one million copies. Over the last years, several lines of evidence demonstrated that these elements modulate gene expression at the post-transcriptional level in at least three independent manners. They have been shown to be involved in alternative splicing, RNA editing and translation regulation. These findings highlight how the genome adapted to these repetitive elements by assigning them important functions in regulation of gene expression. *Alu* elements should therefore be considered as a large reservoir of potential regulatory functions that have been actively participating in primate evolution.


b
L1
pA
5′ UTR
ORF1
ORF2
3′ UTR
AAA
Alu
Left monomer
A-rich
Right monomer
AAA
SVA
pA
(CCCTCT)n
Alu-like
VNTR
HERV-K10-like
AAA

## Rationale of the SVA trans-mobilization assay.



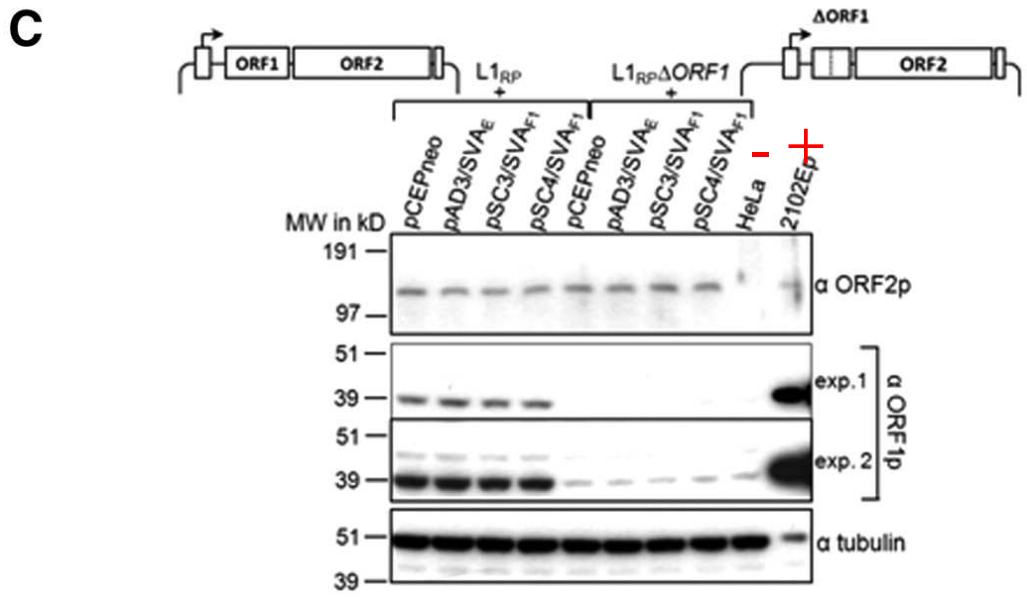

Raiz J et al. Nucl. Acids Res. 2011;nar.gkr863

© The Author(s) 2011. Published by Oxford University Press.

Nucleic Acids Research

**L1 ORF1p is required for trans-mobilization of SVA reporter elements.**

Immunoblot analysis of L1 protein expression after cotransfection of L1 protein donors with SVA retrotransposition reporter plasmids or pCEPneo.



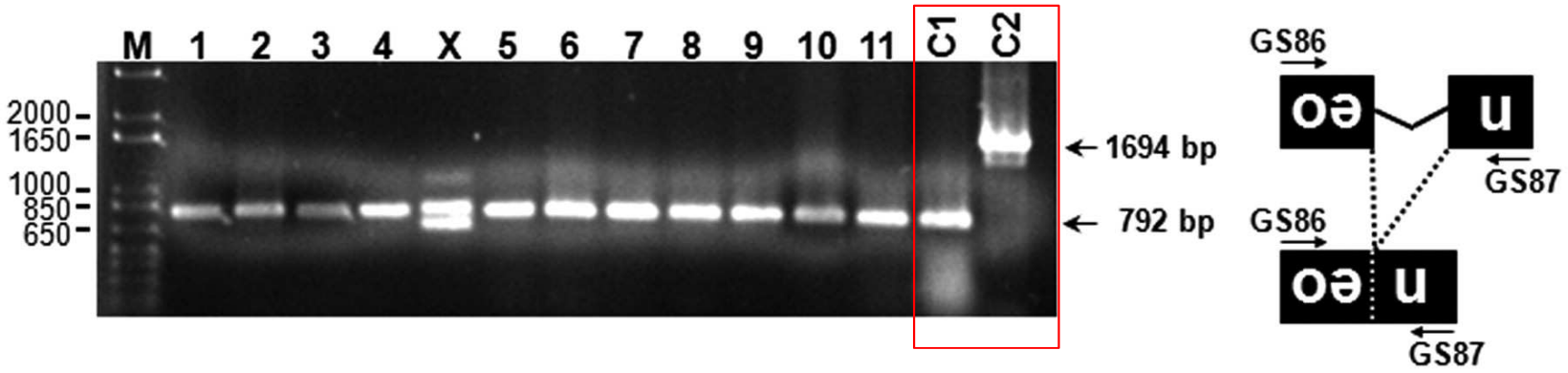

Raiz J et al. Nucl. Acids Res. 2011;nar.gkr863

© The Author(s) 2011. Published by Oxford University Press.

Nucleic Acids Research

**Diagnostic PCR to test for correct splicing of the intron from the mneoI indicator cassette.**

PCR to test for correct splicing of the intron from the mneoI cassette




integration into the genome via authentic “retro”transposition

**Raiz J et al. Nucl. Acids Res. 2011;nar.gkr863**

© The Author(s) 2011. Published by Oxford University Press.

Nucleic Acids Research

## Intact L1 trans-mobilization.


pJM101/L1RP =L1 cis activity Controllo positivo 100%

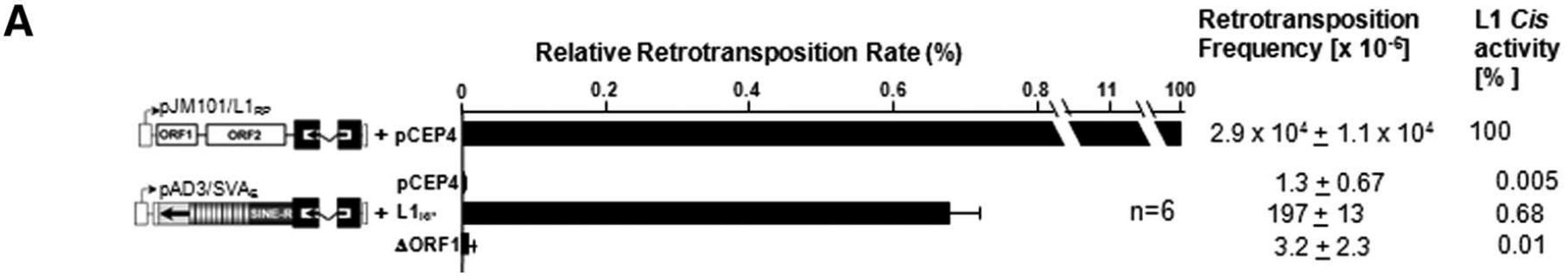
pCEP4 empty vector

**Raiz J et al. Nucl. Acids Res. 2011;nar.gkr863**

© The Author(s) 2011. Published by Oxford University Press.

Nucleic Acids Research

**L1 ORF1p is required for trans-mobilization of SVA reporter elements.**




pJM101/L1RP =L1 cis activity Controllo positivo 100%

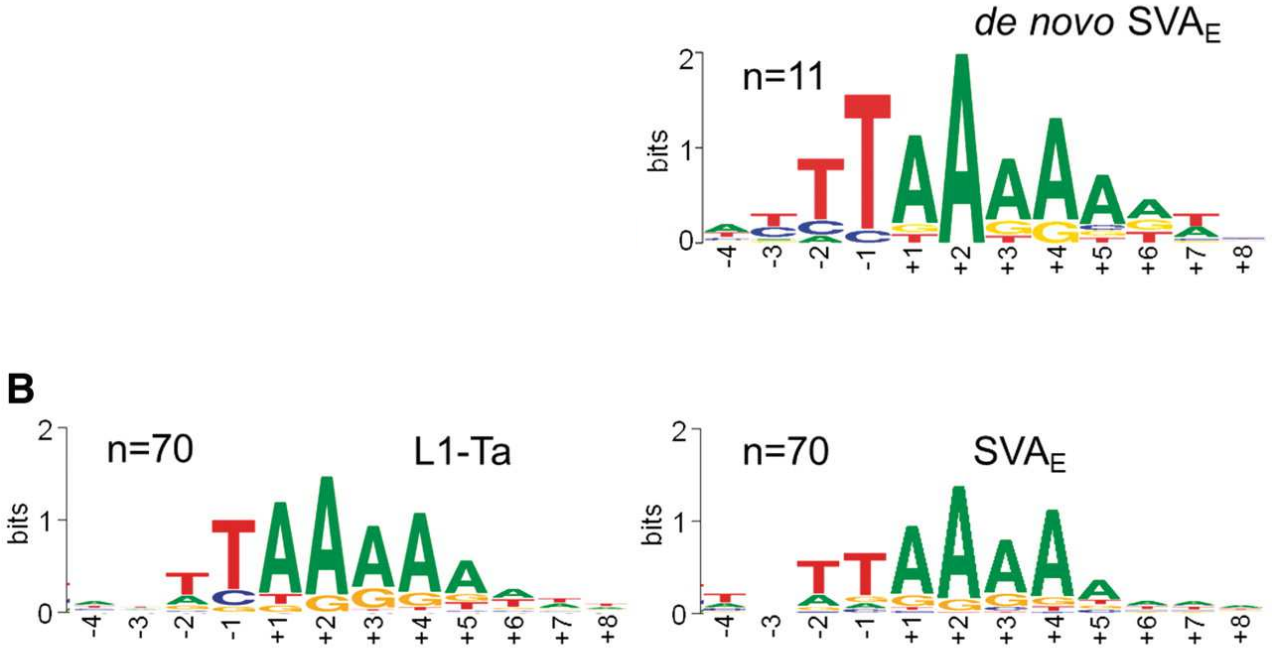
intact (L1RP) and mutant (ΔORF1) L1 protein donor plasmid
pCEP4 empty vector

**Raiz J et al. Nucl. Acids Res. 2011;nar.gkr863**

© The Author(s) 2011. Published by Oxford University Press.

Nucleic Acids Research

**The nucleotide profile of SVAE de novo insertion sites resembles the consensus target sequence of pre-existing human-non-LTR retrotransposons.**




**Raiz J et al. Nucl. Acids Res. 2011;nar.gkr863**

© The Author(s) 2011. Published by Oxford University Press.

Nucleic Acids Research