Table 1. Mobile element dynamics in model organisms. Tns, DNA transposons; Rtns, retrotransposons. Organisms are budding yeast, *S. cerevesiae*; mustard weed, *A. thaliana*; roundworm, *C. elegans*; fruit fly, *D. melanogaster*; mouse, *M. musculis*; human, *H. sapiens*.

Organism	Mobile element type (% of genome)			Active element(s)	Estimate of insertion freq. per generation	Estimate of removal freq.
	Tns	LTR Rtns	Non-LTR Rtns			
Budding yeast	0	3	0	LTR Rtn	10^{-5}–10^{-7}*	High (LTR recombination)
Mustard weed	5	5	0.5	Tn, LTR Rtn	?	?
Roundworm	12	0	<0.4	Tn	Very low	?(Low)
Fruit fly	0.3	2.7	0.9	Tn, LTR Rtn, non-LTR Rtn	10^{-1}–10^{-2}†	High (deletion and selection)‡
Mouse	0.9	10	27	LTR Rtn, non-LTR Rtn	$>10^{-1}$	Low
Human	3	8.5	35	Non-LTR Rtn	10^{-1}§	Low

*See (*63*). †Mobile element insertion rates for P and I element hybrid dysgenesis crosses are ~10°. In natural crosses, transposition and retrotransposition rates are 10^{-1} to 10^{-2} [for copia and Doc, see (*65*); for mariner, see (*66*)]. ‡See (*67*). §See (*31*).

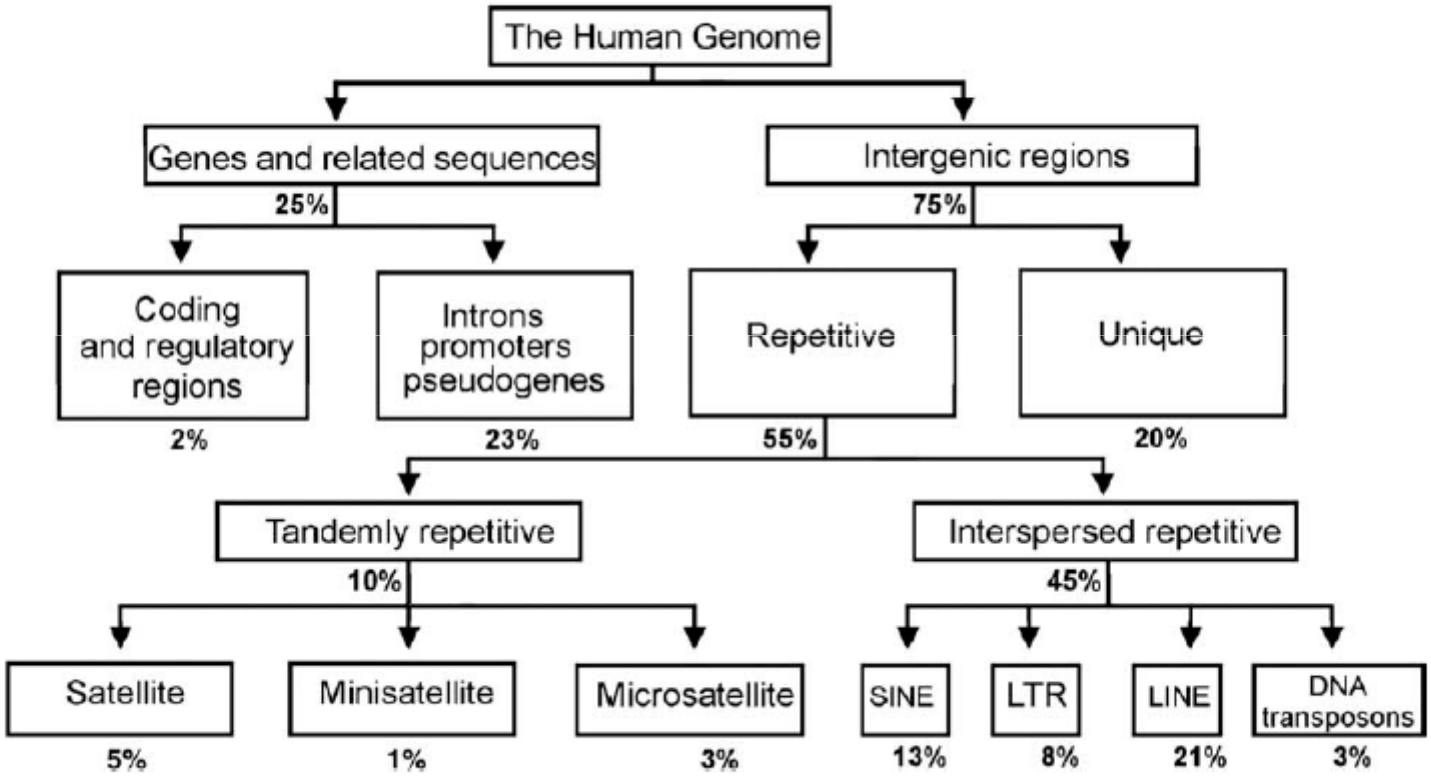
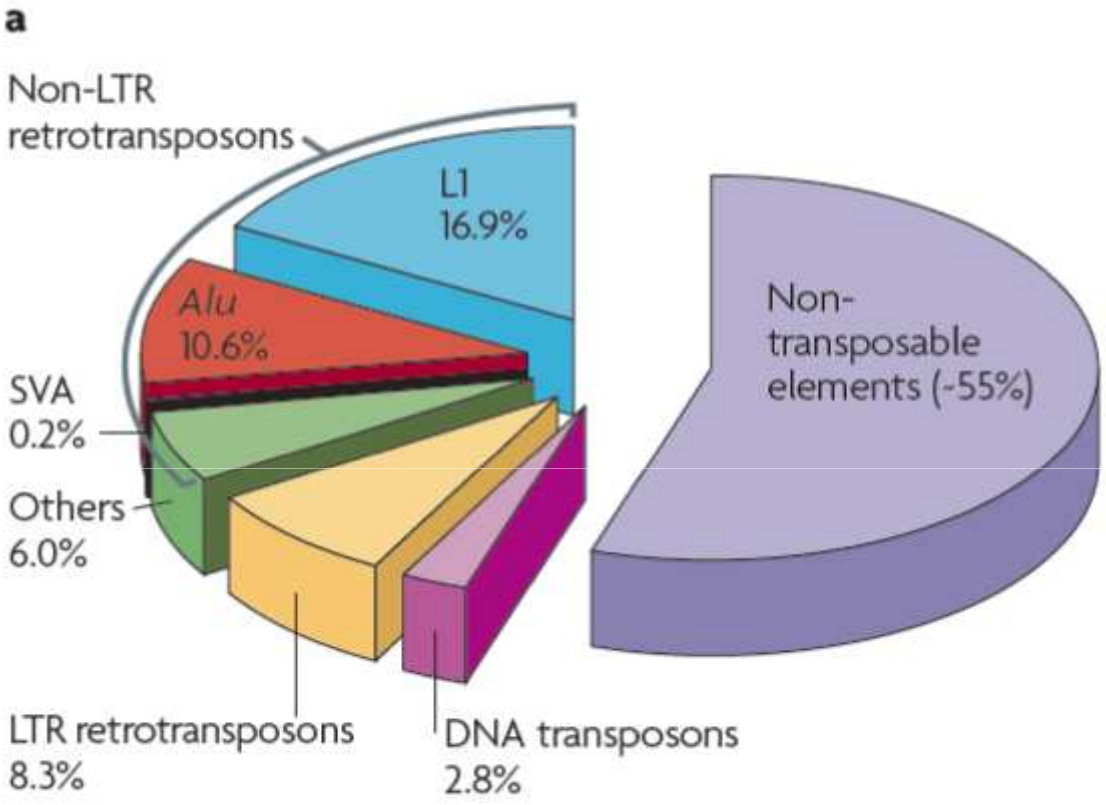
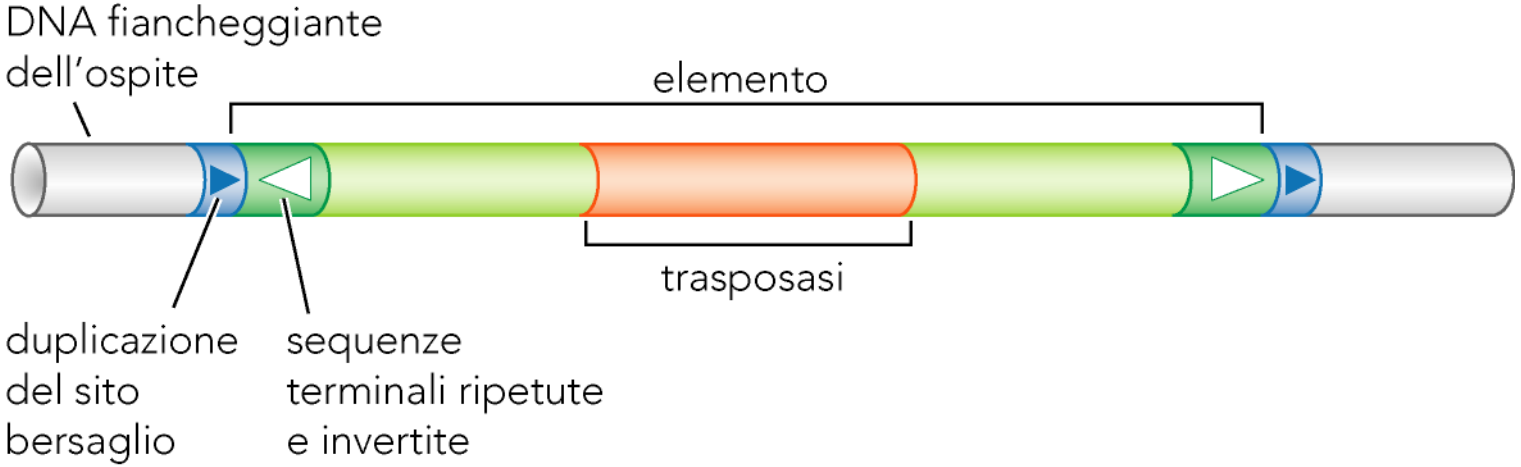


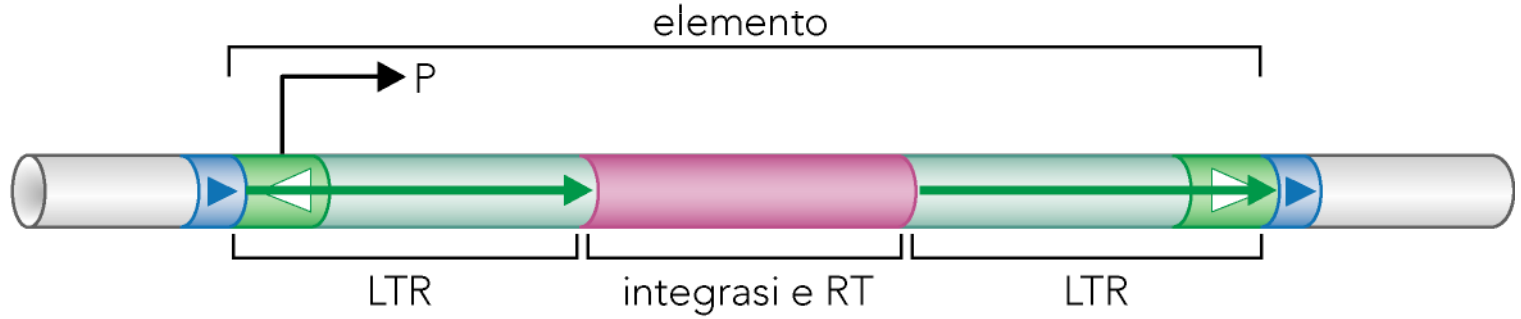
Fig. 1. Composition of the human genome. The percentage shares of various functional and non-functional sequences are shown.


a
Non-LTR retrotransposons
L1 16.9%
Alu 10.6%
SVA 0.2%
Others 6.0%
LTR retrotransposons 8.3%
DNA transposons 2.8%
Non-transposable elements (~55%)

a trasposoni a DNA



b retrotrasposoni tipo virus/retrovirus



c retrotrasposoni poli-A

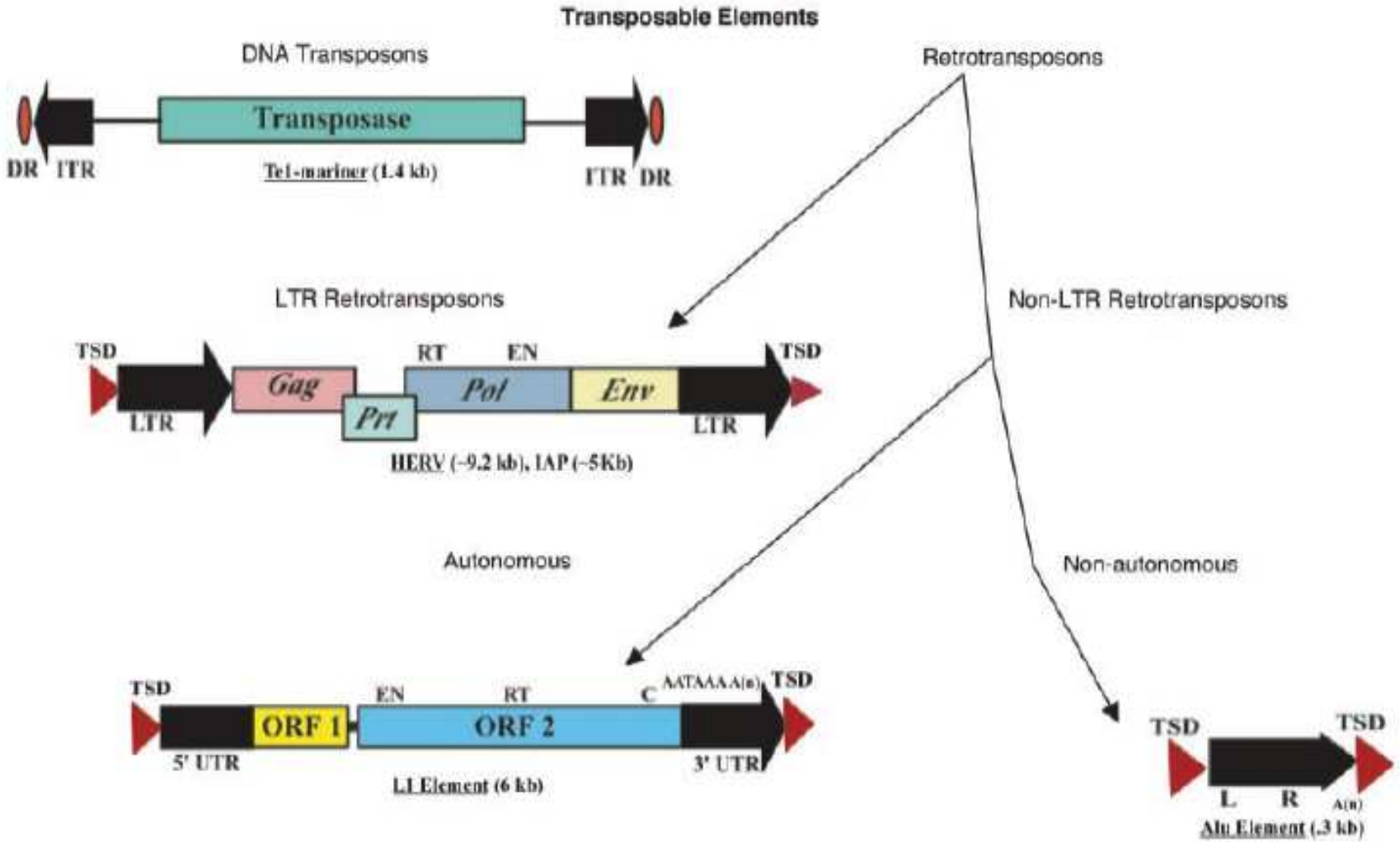
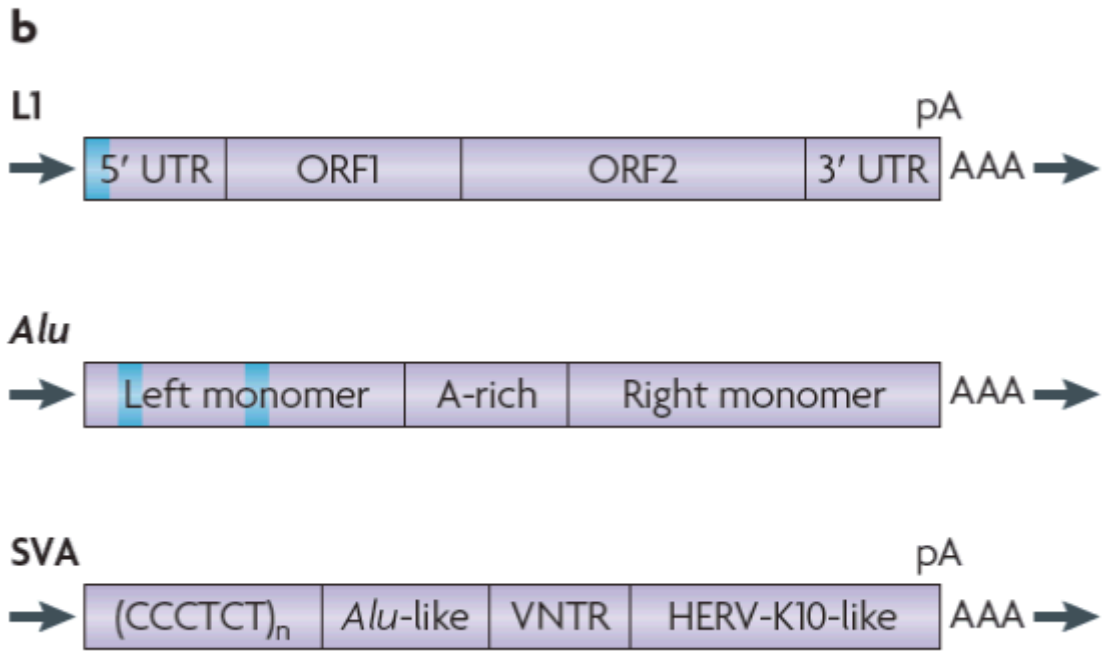


Fig. 1. Classes of mobile elements. DNA transposons, e.g., Tc-1/mariner, have inverted terminal inverted repeats (ITRs) and a single open reading frame (ORF) that encodes a transposase. They are flanked by short direct repeats (DRs). Retrotransposons are divided into autonomous and nonautonomous classes depending on whether they have ORFs that encode proteins required for retrotransposition. Common autonomous retrotransposons are (i) LTRs or (ii) non-LTRs (see text for


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

REVIEW

Mobile Elements: Drivers of Genome Evolution

Haig H. Kazazian Jr.*

Mobile elements within genomes have driven genome evolution in diverse ways. Particularly in plants and mammals, retrotransposons have accumulated to constitute a large fraction of the genome and have shaped both genes and the entire genome. Although the host can often control their numbers, massive expansions of retrotransposons have been tolerated during evolution. Now mobile elements are becoming useful tools for learning more about genome evolution and gene function.

Mobile, or transposable, elements are prevalent in the genomes of all plants and animals. Indeed, in mammals they and their recognizable remnants account for nearly half of the genome (*1*, *2*), and in some plants they constitute up to 90% of the genome (*3*). If, as many believe, the origins of life are in an "RNA world" followed by reverse transcription into DNA, then mobile elements could

Because sequence specificity of integration is limited to a small number of nucleotides—e.g., TA dinucleotides for Tc1 of *Caenorhabditis elegans*—insertions can occur at a large number of genomic sites. However, daughter insertions for most, but not all, DNA transposons occur in proximity to the parental insertion. This is called "local hopping." Active transposons encode a transposase enyme

residues, then a glutamate) and a handlike three-dimensional structure (*6*, *8*).

Although these elements generally transpose to genomic sites less than 100 kb from their original site (e.g., the *Drosophila* P element), some are able to make distant "hops" (e.g., the fish Tc1/mariner element; see below).

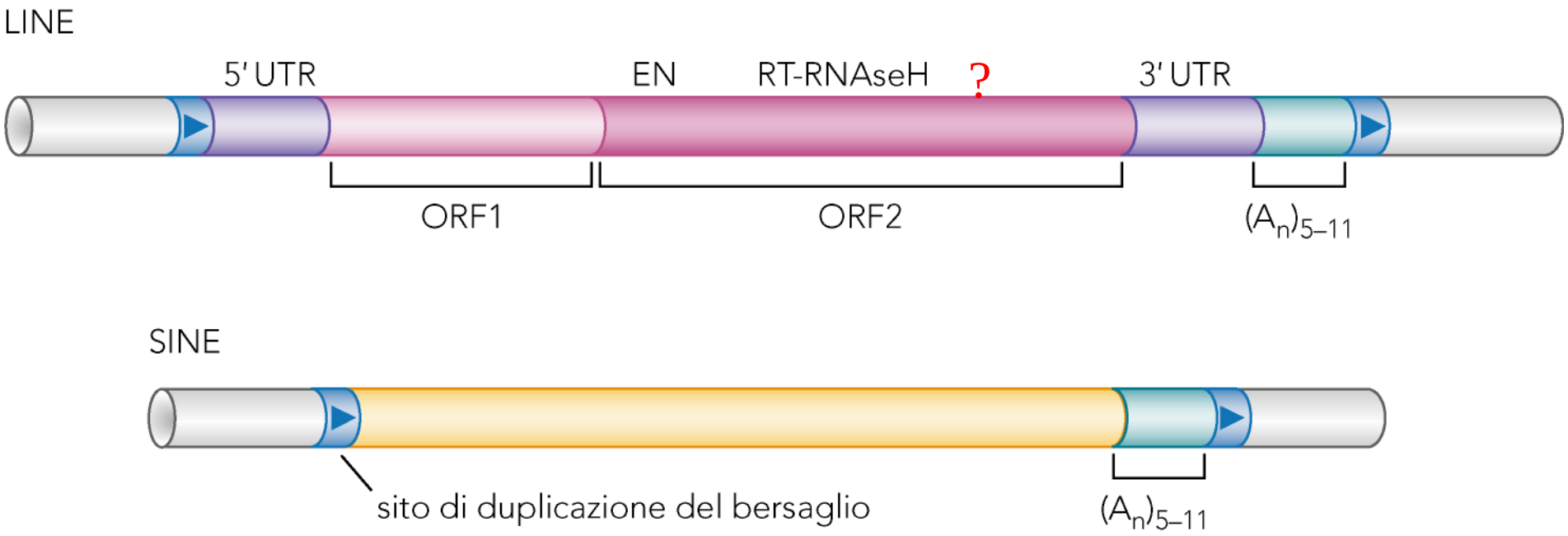
LTR Retrotransposons

Retrotransposons are transcribed into RNA, and then reverse transcribed and reintegrated into the genome, thereby duplicating the element. The major classes of retrotransposons either contain long terminal repeats at both ends (LTR retrotransposons) or lack LTRs and possess a polyadenylate sequence at their 3′ termini (non-LTR retrotransposons).

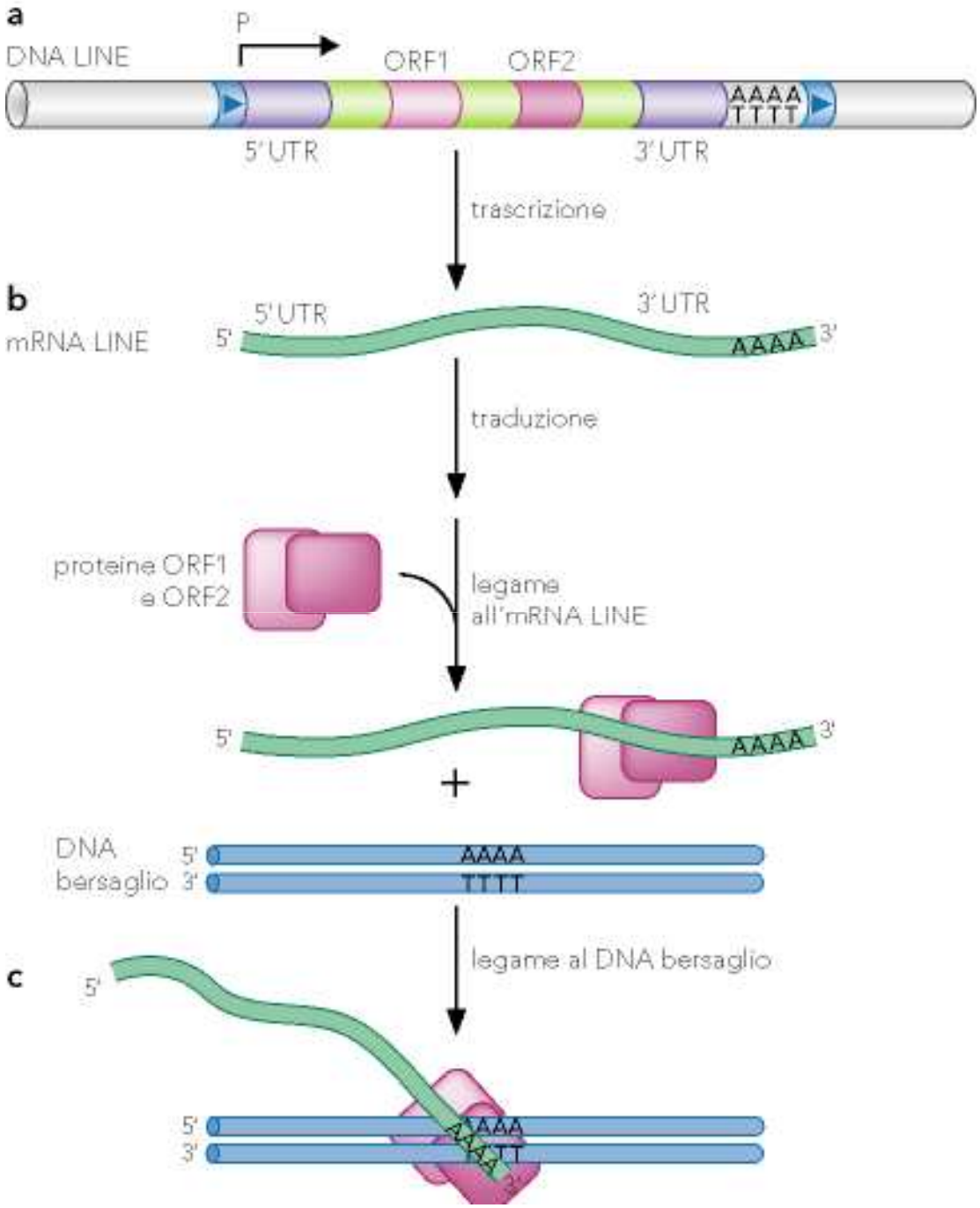
LTR retrotransposons and retroviruses are

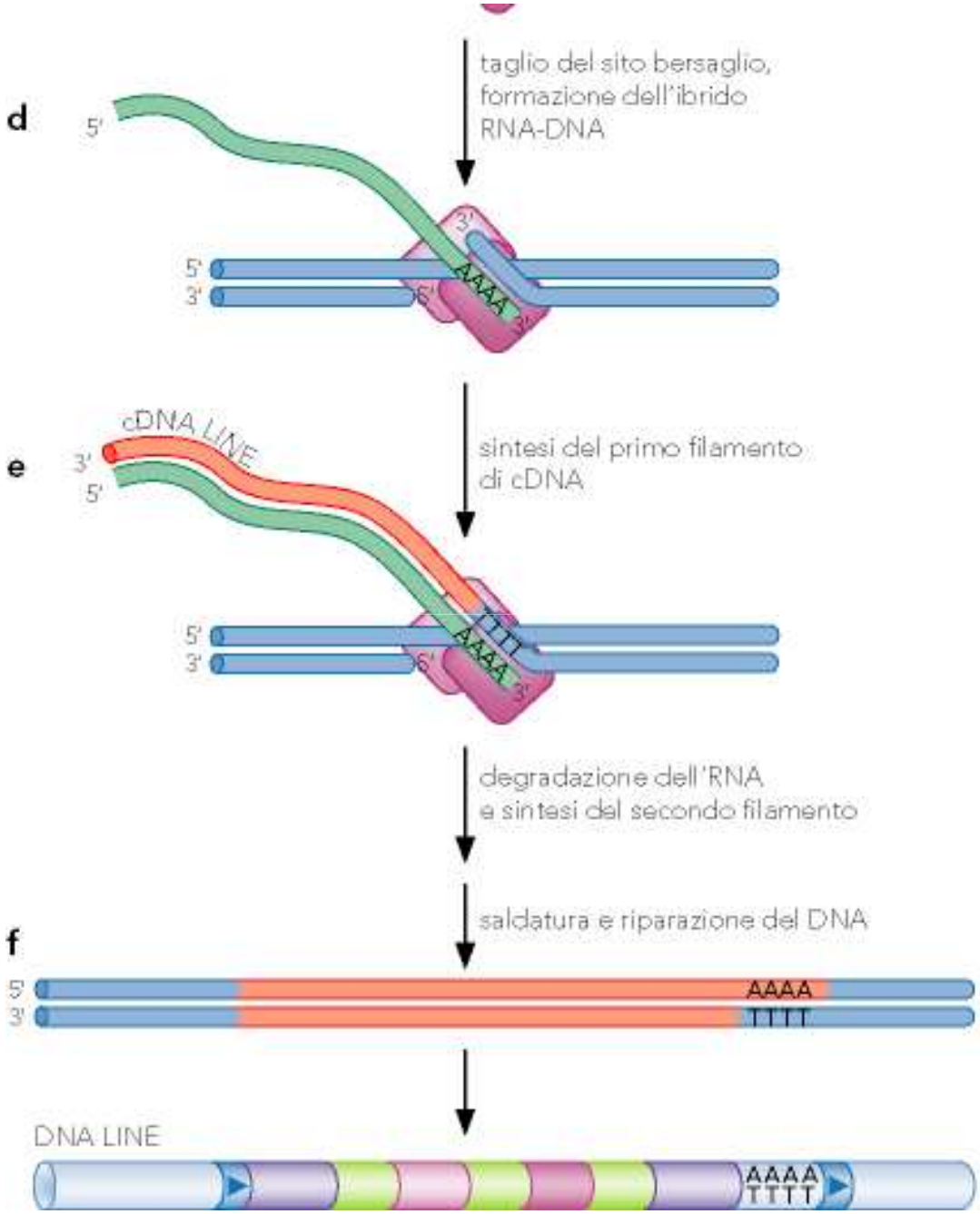
Watson et al., BIOLOGIA MOLECOLARE DEL GENE, Zanichelli Editore S.p.A.

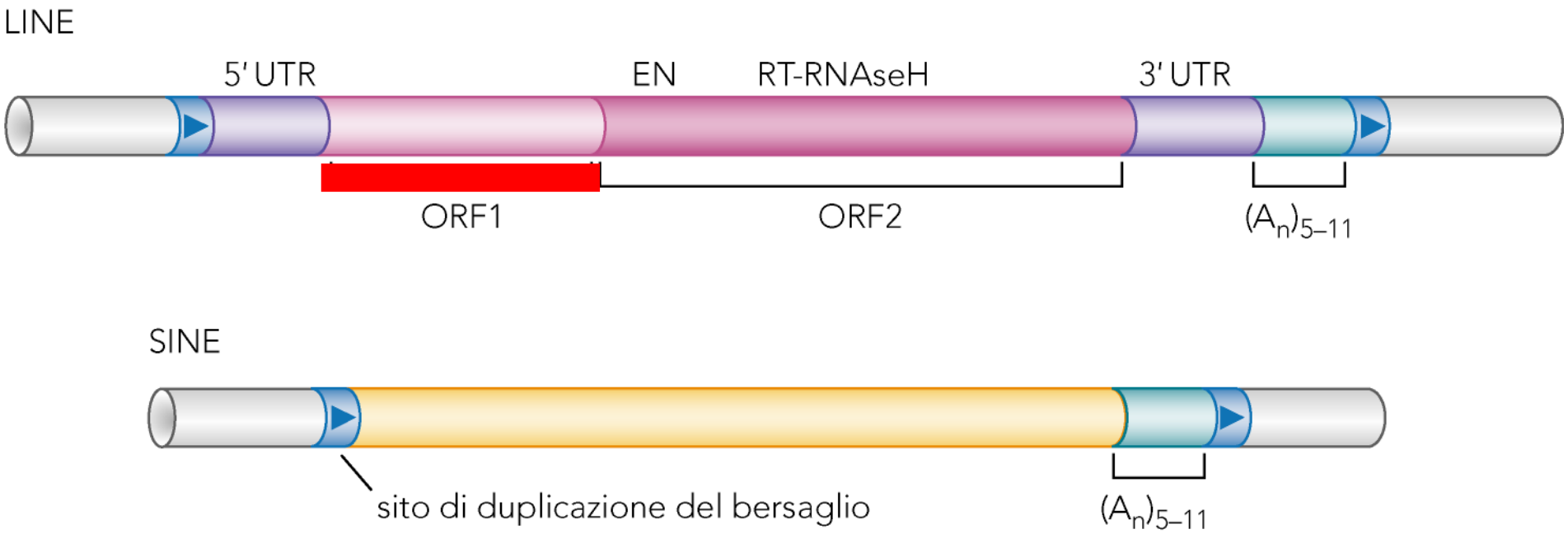
Watson BM Gene Capitolo 11


LINE
5′ UTR
EN
RT-RNAseH
?
3′ UTR
ORF1
ORF2
$(A_n)_{5–11}$
SINE
sito di duplicazione del bersaglio
$(A_n)_{5–11}$

Le proteine della trasposizione


a
DNA LINE
P
ORF1
ORF2
AAAA
TTTT
5' UTR
3' UTR
trascrizione
b
mRNA LINE
5' UTR
3' UTR
5'
AAAA
3'
traduzione
proteine ORF1
e ORF2
legame
all'mRNA LINE
5'
AAAA
3'
+
DNA
bersaglio
5'
3'
AAAA
TTTT
legame al DNA bersaglio
c
5'
5'
3'
AAA
TT


d
taglio del sito bersaglio,
formazione dell'ibrido
RNA-DNA
5'
3'
5'
3'
AAAA
e
cDNA LINE
sintesi del primo filamento
di cDNA
TTT
degradazione dell'RNA
e sintesi del secondo filamento
saldatura e riparazione del DNA
f
AAAA
TTTT
DNA LINE



Watson et al., BIOLOGIA MOLECOLARE DEL GENE, Zanichelli editore S.p.A.

Trimeric structure for an essential protein in L1 retrotransposition

Sandra L. Martin*†, Dan Branciforte*, David Keller‡, and David L. Bain§

Watson et al., BIOLOGIA MOLECOLARE DEL GENE, Zanichelli editore S.p.A.

The function of the protein encoded by the 5-most ORF, ORF1p, is incompletely understood,

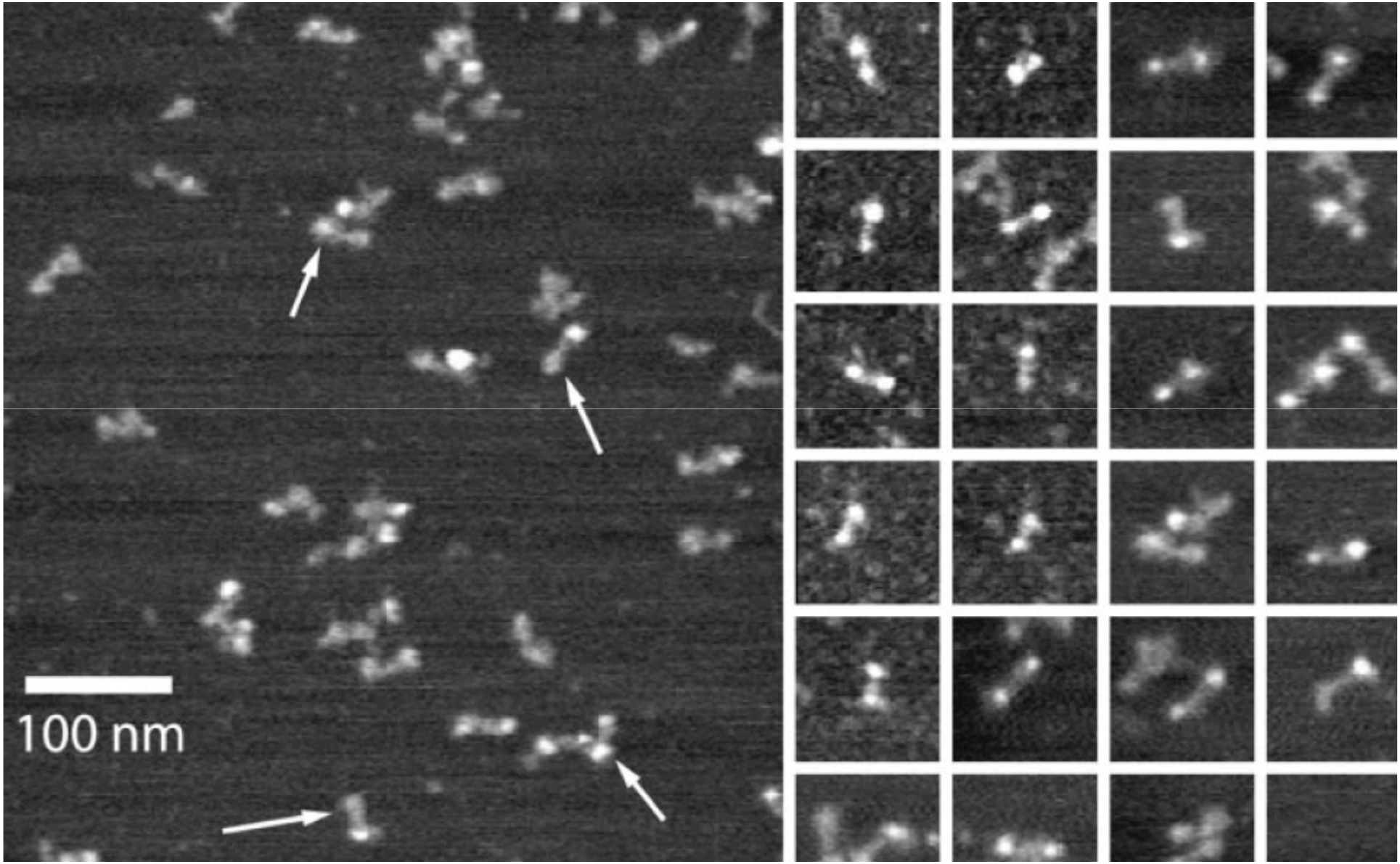
the ORF1p from mouse L1 is known to bind single-stranded nucleic

acids (L1 RNA and DNA) and function as a nucleic acid chaperone.

Structural features are compatible with the nucleic acid binding

Watson et al., BIOLOGIA MOLECOLARE DEL GENE, Zanichelli editore S.p.A.

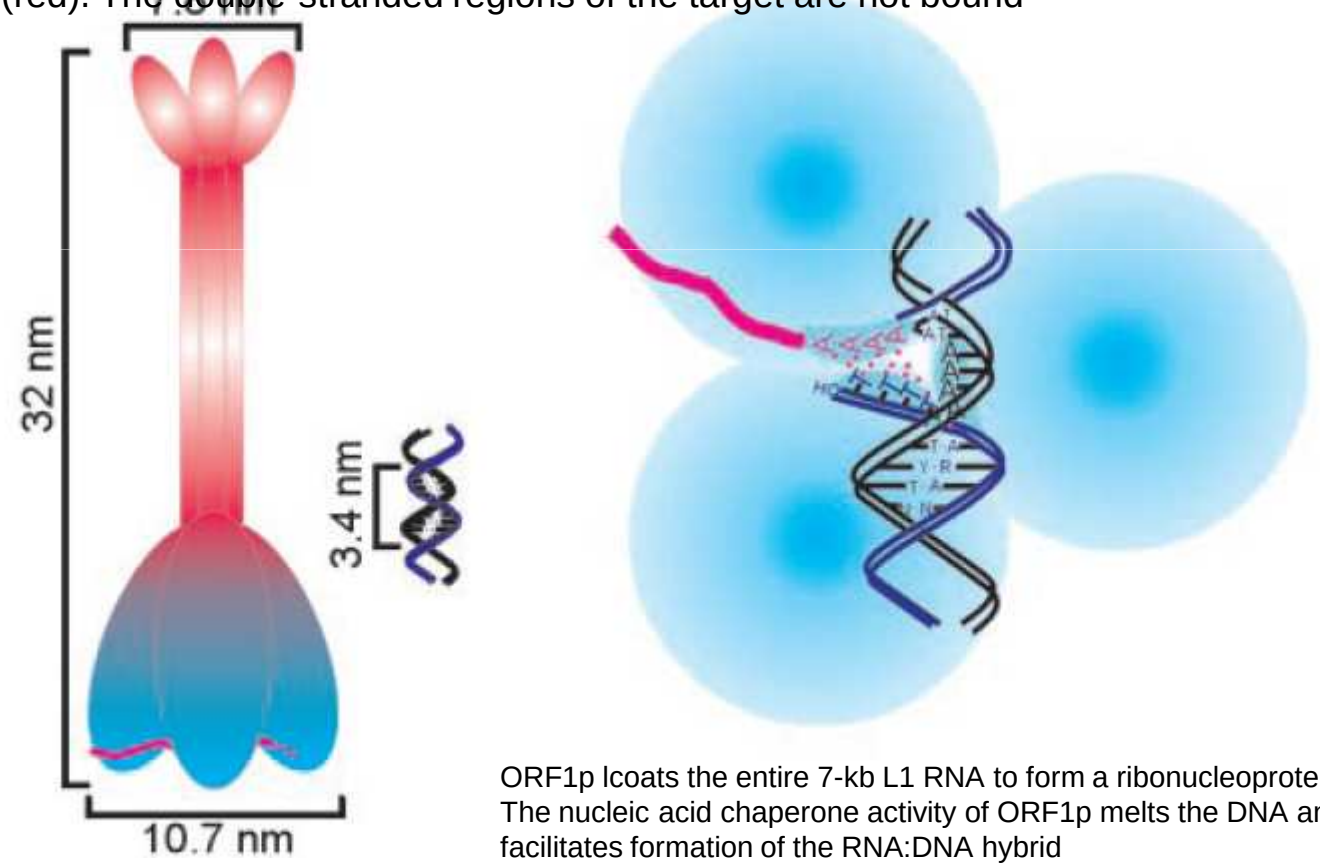
LINE-1 Orf1



acidic(red)

basic (blue)

Each subunit of the trimer contains one single-stranded nucleic acid binding interface which is bound with one of the DNA target strands or the polyA tail of the L1 RNA (red). The double-stranded regions of the target are not bound

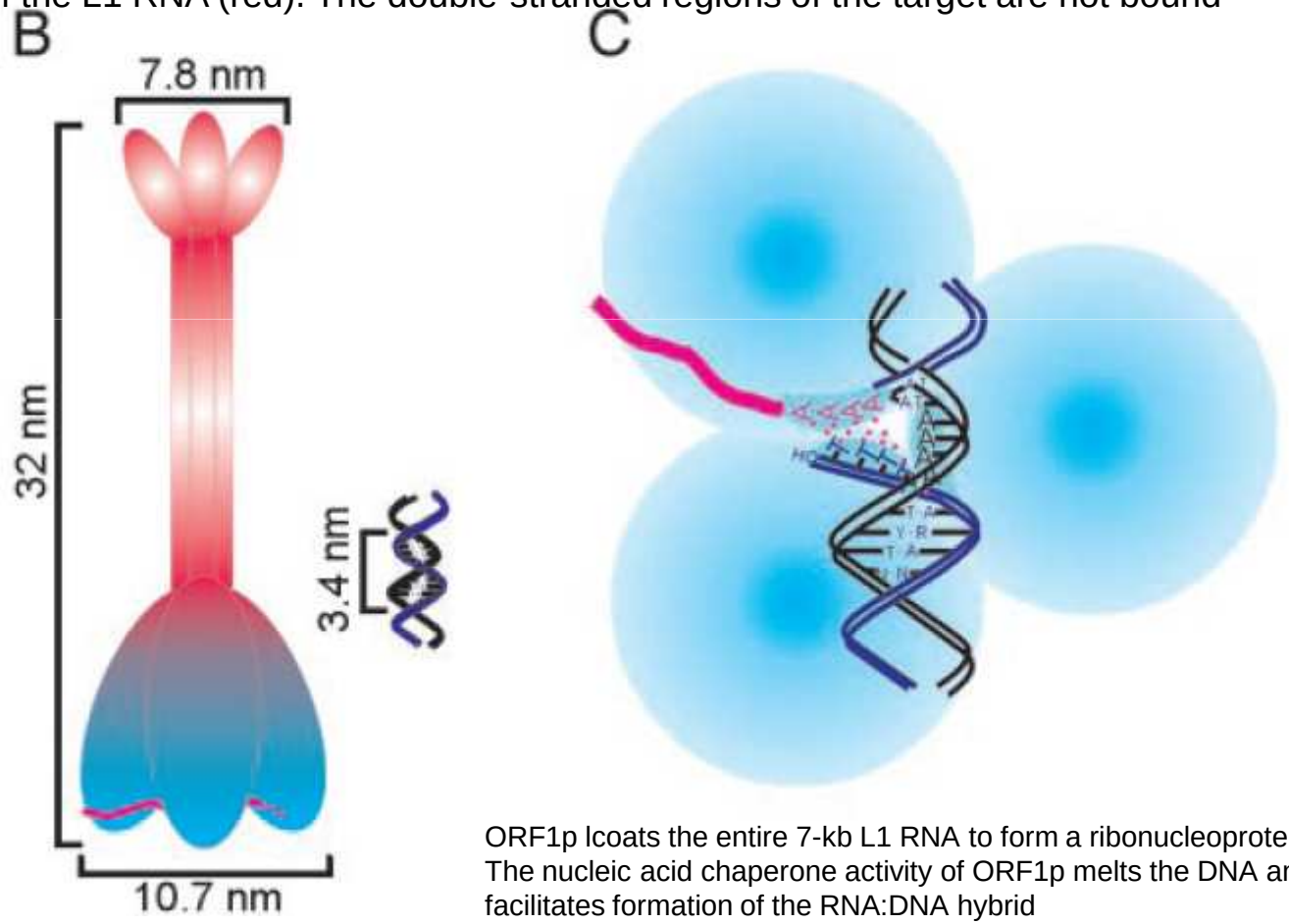


ORF1p lcoats the entire 7-kb L1 RNA to form a ribonucleoprotein particle
The nucleic acid chaperone activity of ORF1p melts the DNA and then facilitates formation of the RNA:DNA hybrid

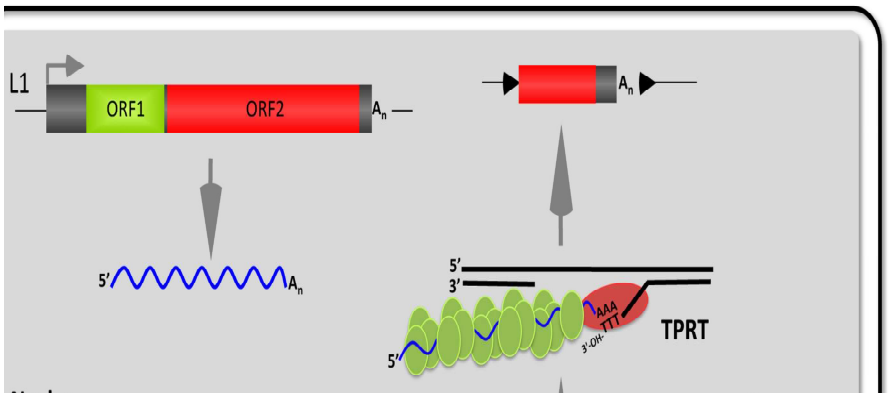
acidic(red)

basic (blue)

Each subunit of the trimer contains one single-stranded nucleic acid binding interface (shaded crescent), which is shown bound with one of the DNA target strands (the nicked strand with HO-TTTT and the intact strand with AAAA; ref. 19) or the polyA tail of the L1 RNA (red). The double-stranded regions of the target are not bound

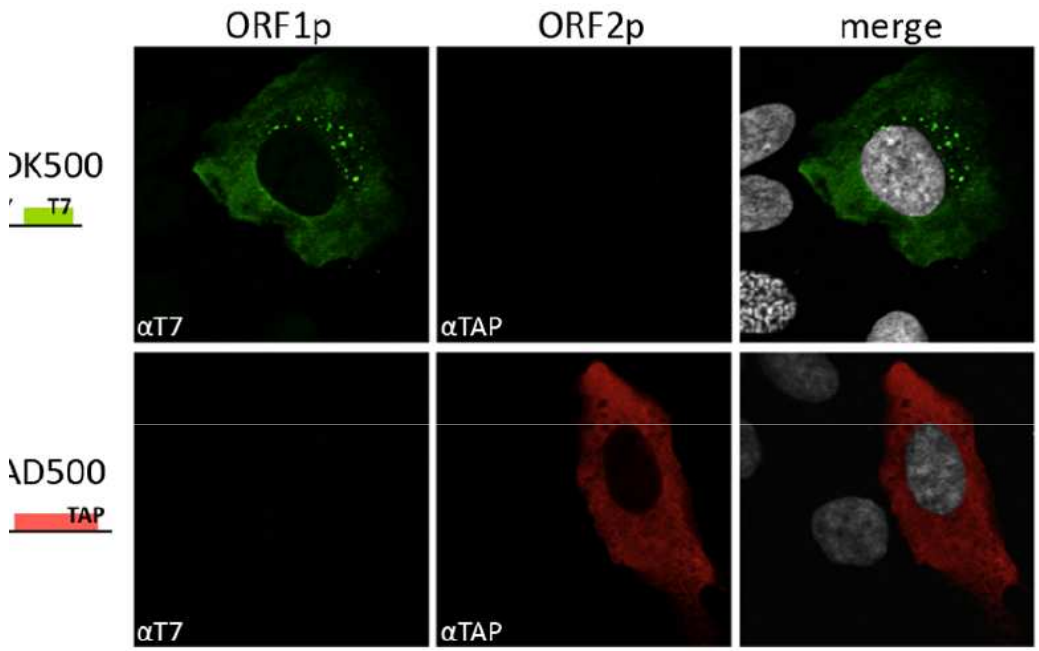


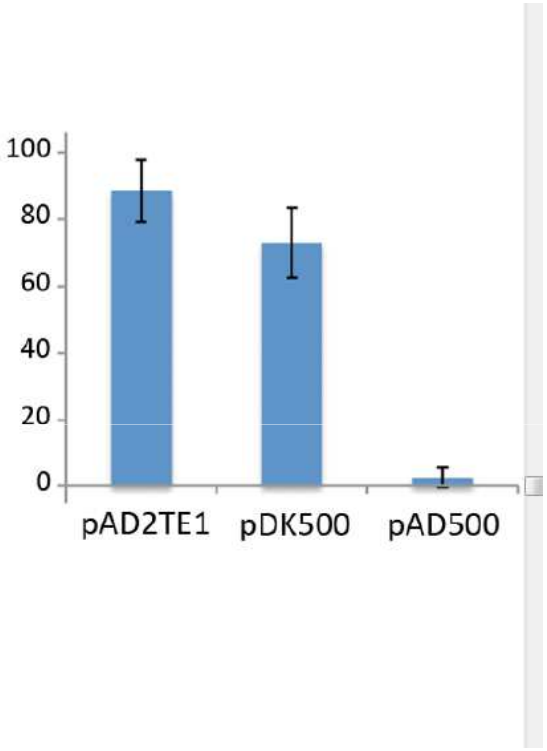
ORF1p lcoats the entire 7-kb L1 RNA to form a ribonucleoprotein particle The nucleic acid chaperone activity of ORF1p melts the DNA and then facilitates formation of the RNA:DNA hybrid



target-site primed reverse transcription

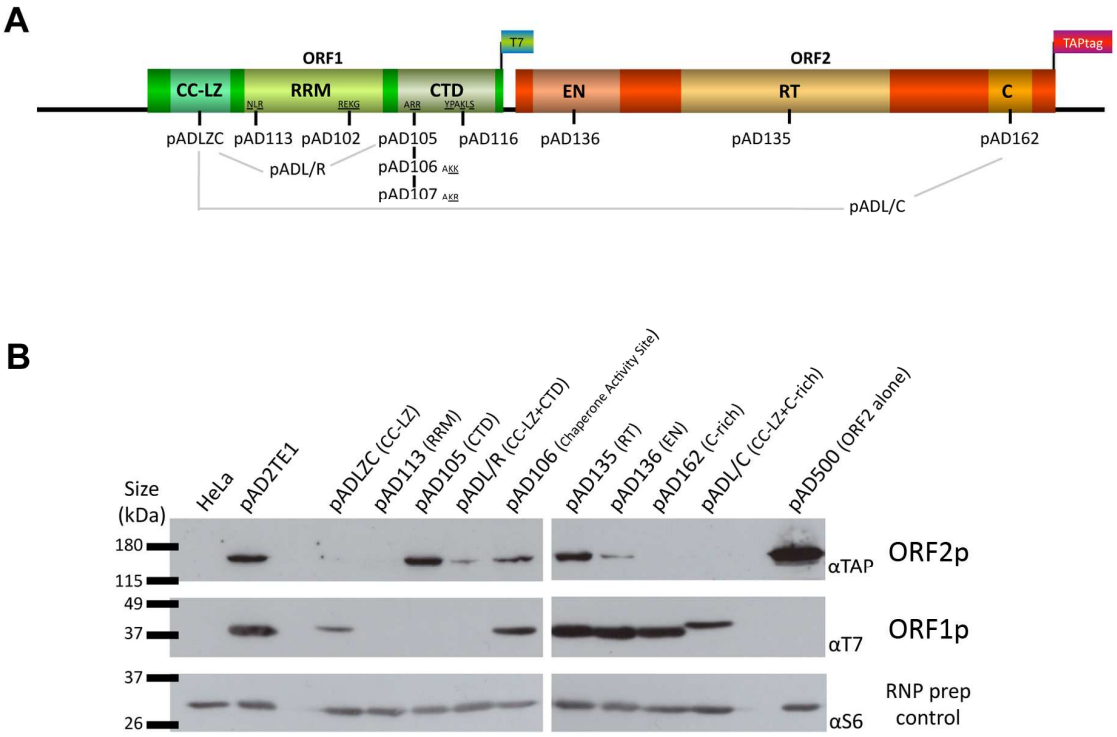
ORF1p is necessary and sufficient for L1 cytoplasmic foci formation





T7-tagged ORF1p green

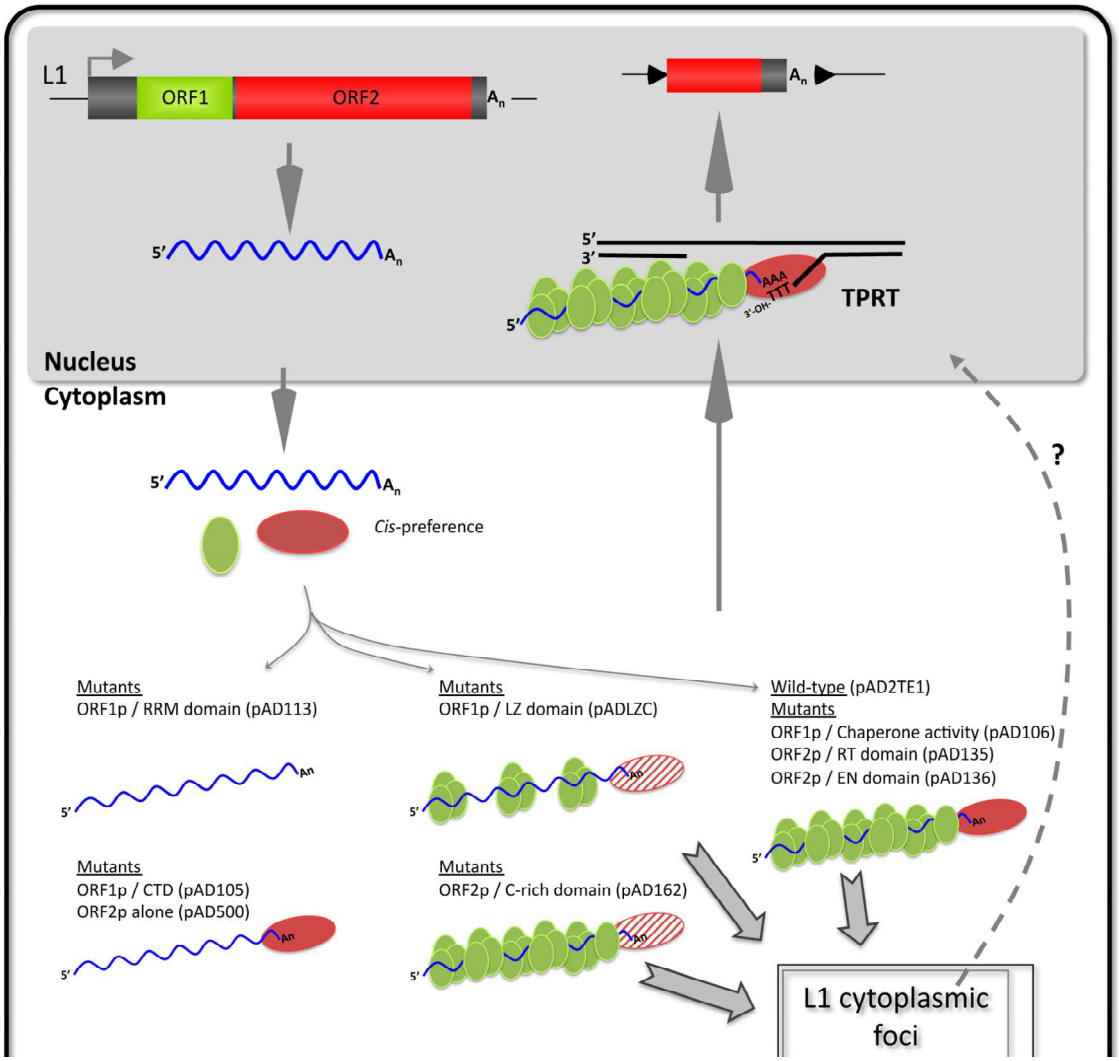
TAP-tagged ORF2p red;



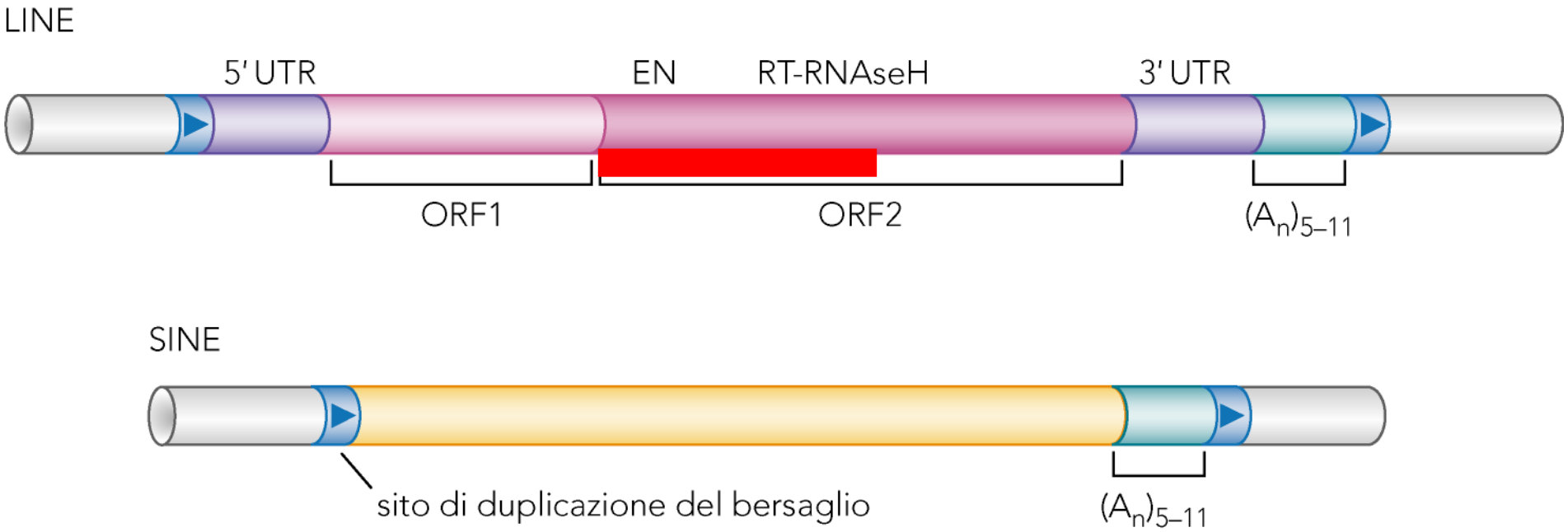
T7 gene 10 epitope tag

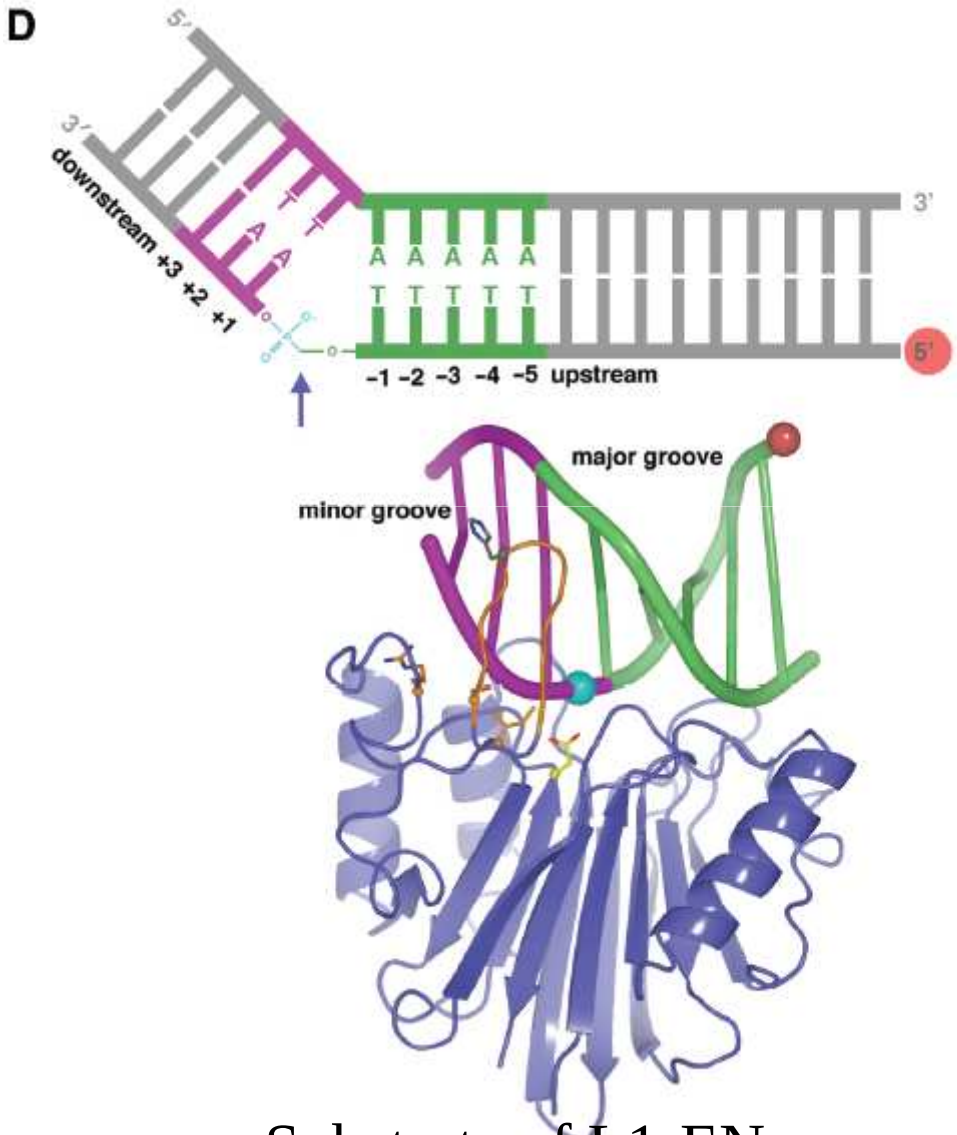
Tandem Affinity Purification epitope tag (TAP tag)

Mutations in the ORF1p RNA recognition motif (pAD113) disrupt L1 cytoplasmic foci formation and lead to a severe reduction of ORF1p and ORF2p in cytoplasmic RNP complexes



target-site primed reverse transcription (TPRT)

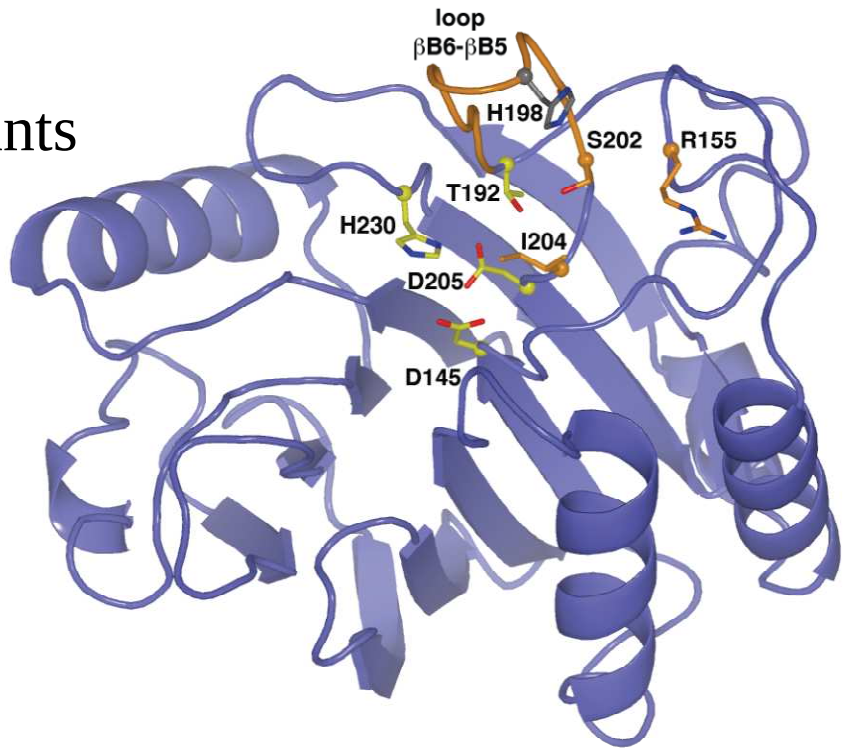

LINE
5′ UTR
EN
RT-RNAseH
3′ UTR
ORF1
ORF2
$(A_n)_{5–11}$
SINE
sito di duplicazione del bersaglio
$(A_n)_{5–11}$



Substrate of L1-EN

L1-EN point mutants and bB6–bB5 hairpin loop variants

A



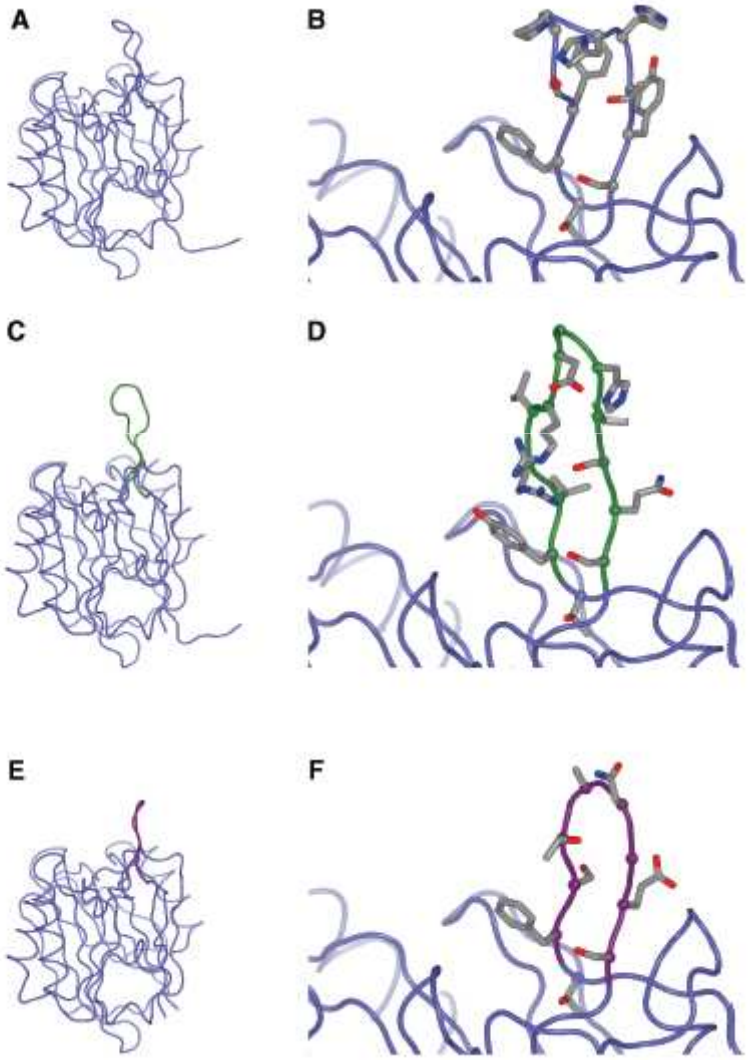
orange: residues
potentially contacting DNA

B

	191	192	193	194	195	196	197	.	.	198	199	200	201	202	203	204	205
L1-EN	Y	T	F	F	S	A	P	.	.	H	H	T	Y	S	K	I	D
Tx1L-EN	F	T	Y	V	R	V	R	D	G	H	V	S	Q	S	R	I	D
R1Bm-EN	P	T	F	S	T	A	.	.	.	.	N	G	E	S	Y	V	D
Tras1-EN	P	T	F	D	T	I	R	G	G	K	R	Y	Q	S	R	V	D
	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211

Crystal structures of L1-EN betaB5–beta B6 hairpin loop variants.

WT



Substrate requirements of L1-EN

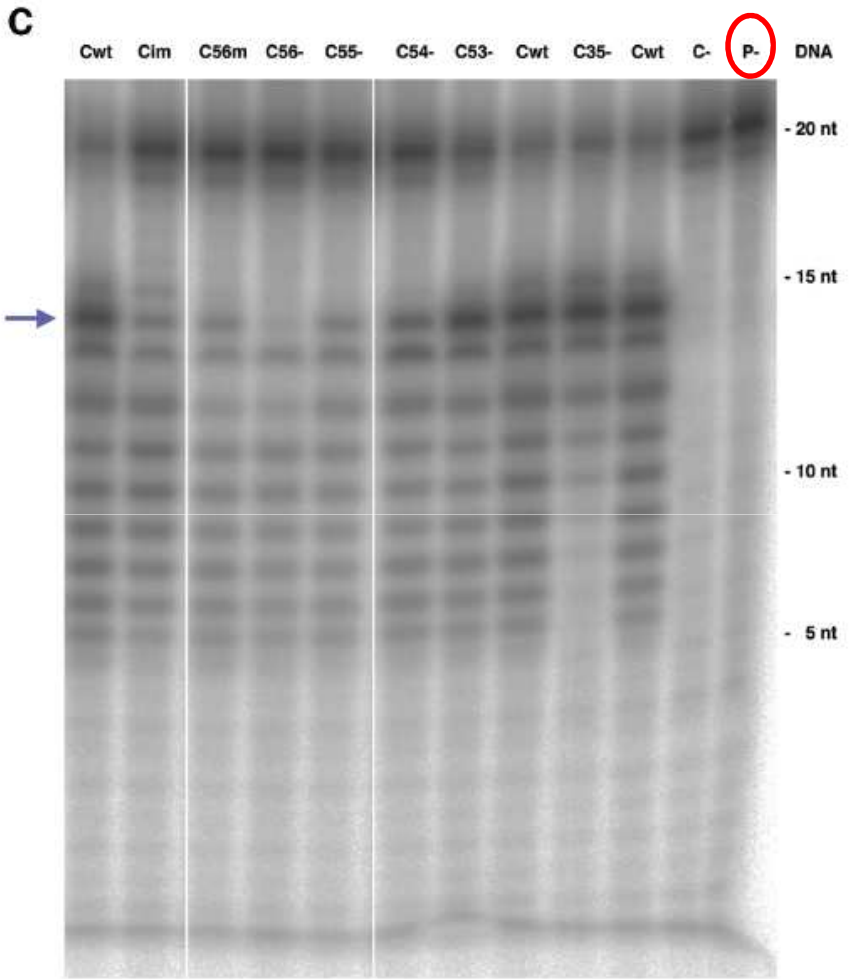
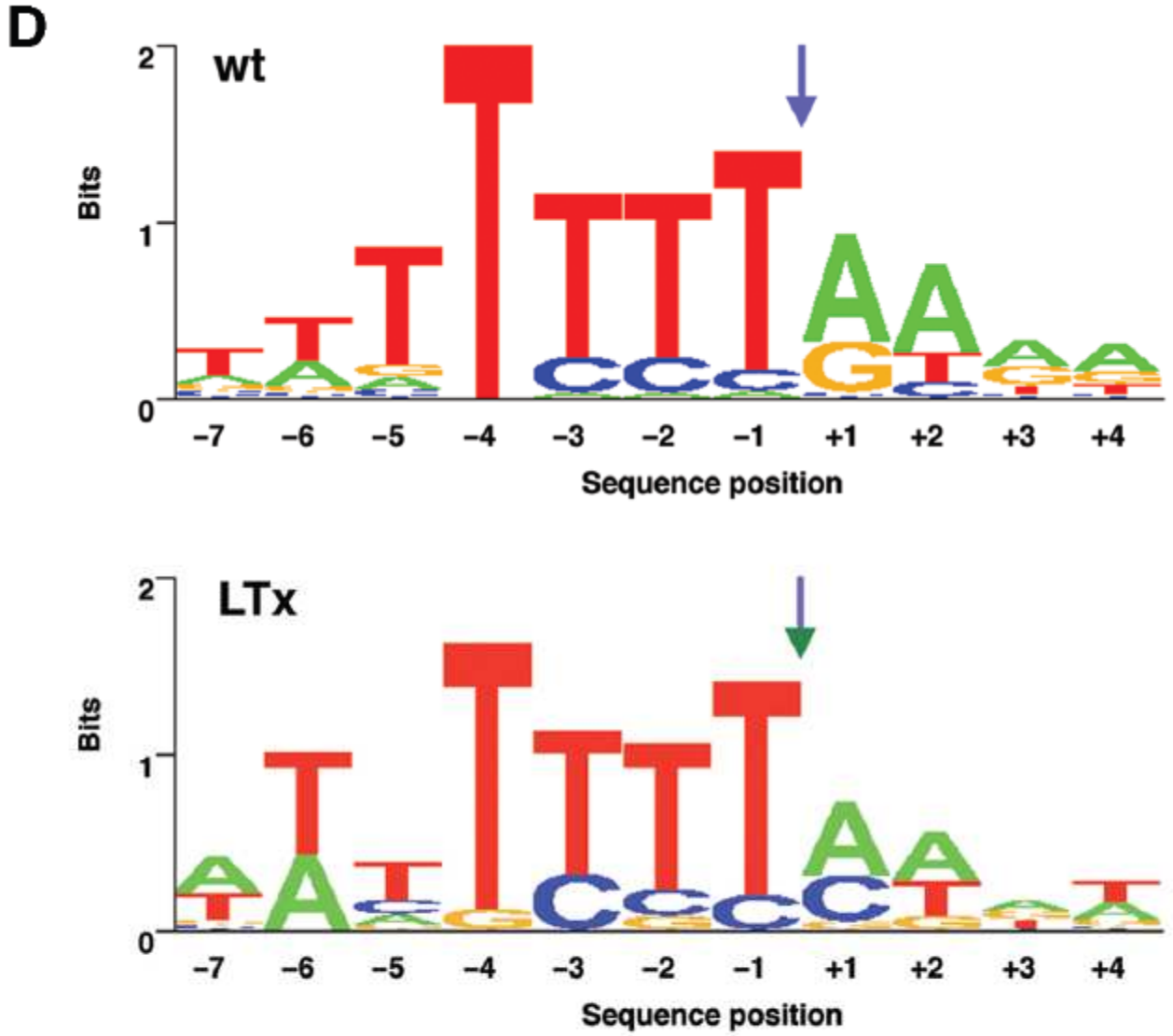


Table 1. Comparison of retrotransposition frequencies *in vivo* and plasmid nicking activities *in vitro*

L1-EN variant	Retrotransposition frequency[a], %	Plasmid nicking activity[b], %
wt	100 ± 17.1	100 ± 0.8
LTx	21 ± 2.4	29 ± 2.6
LR1	2 ± 2.3	6 ± 0.8
L3G	0 ± 2.2	10 ± 1.8
D145A	0[c]	3 ± 1.0
R155A	12 ± 3.3	19 ± 3.4
T192V	5 ± 3.0	–
S202A	32 ± 7.8	28 ± 2.2
I204Y	1 ± 1.1	4 ± 1.2
H230A	0	–

[a]Corrected for background activity (≤5%); for details see Supplementary Data.
[b]Normalized to L1-EN (wt) activity, (–) not analyzed.
[c]As a D145A/N147A double mutant.

Specificity of L1-EN



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

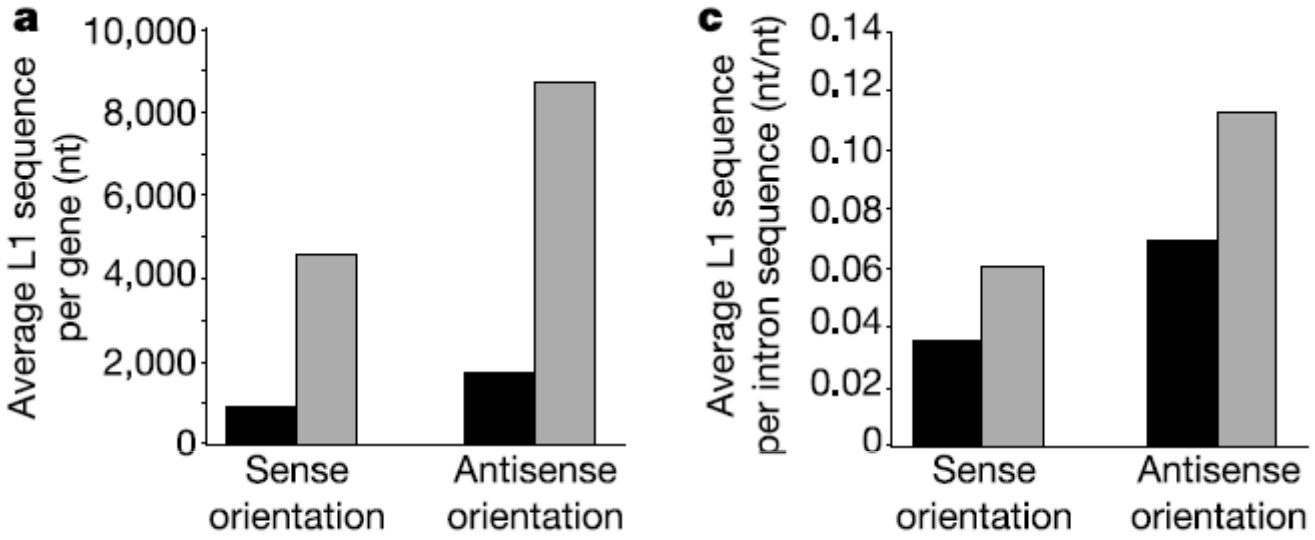
Jeffrey S. Han[1], Suzanne T. Szak[2] & Jef D. Boeke[1]

[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

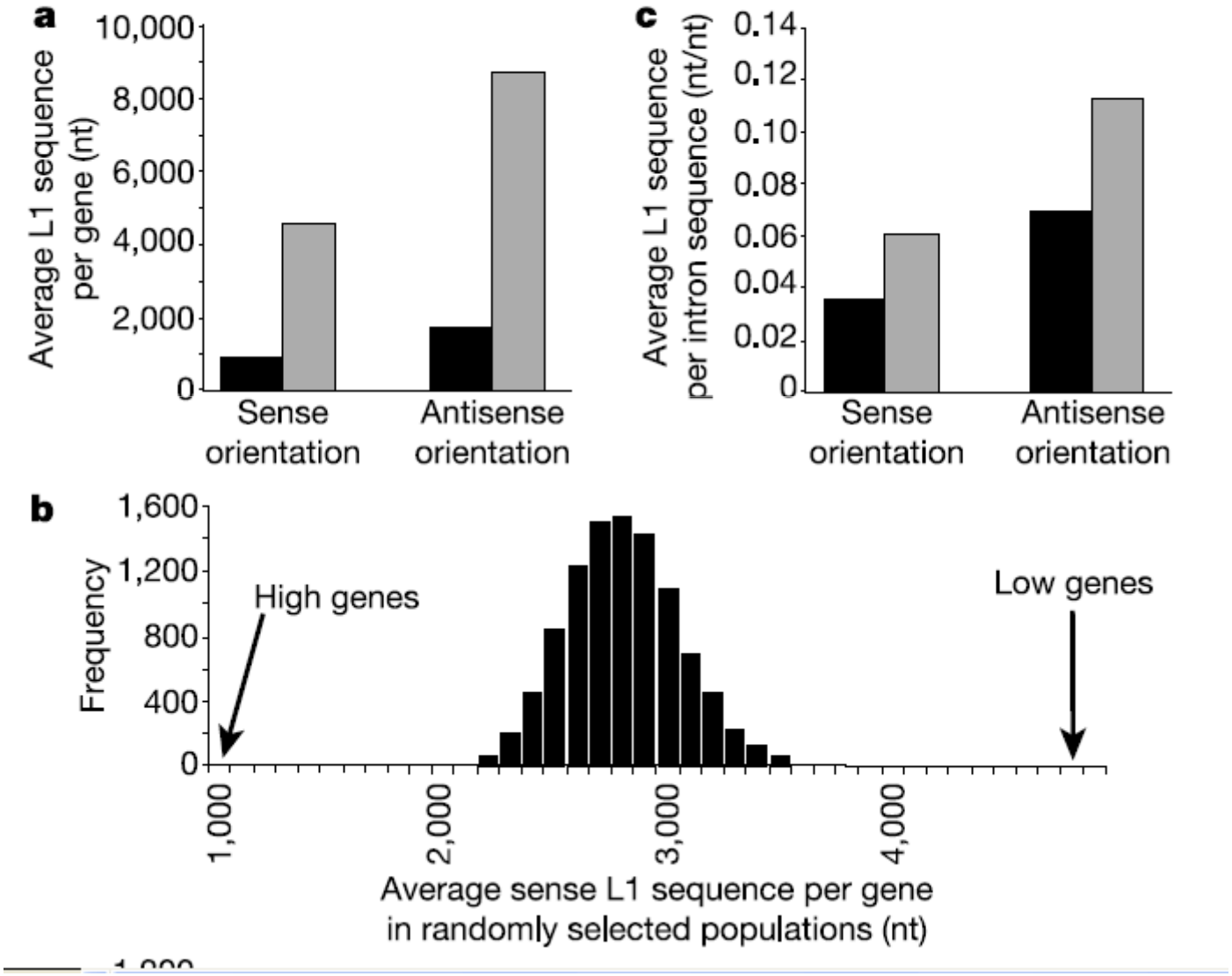
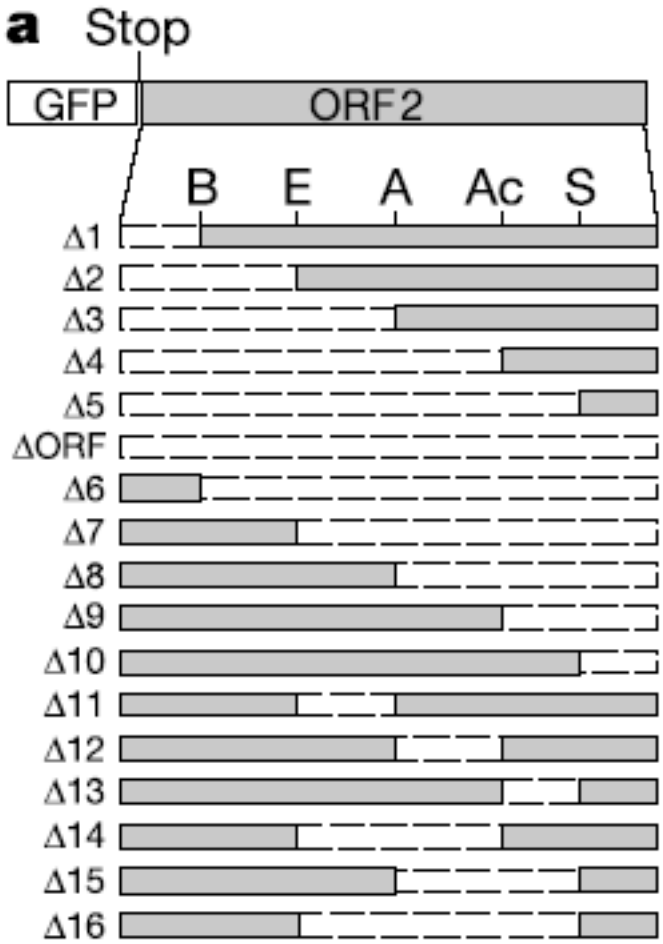
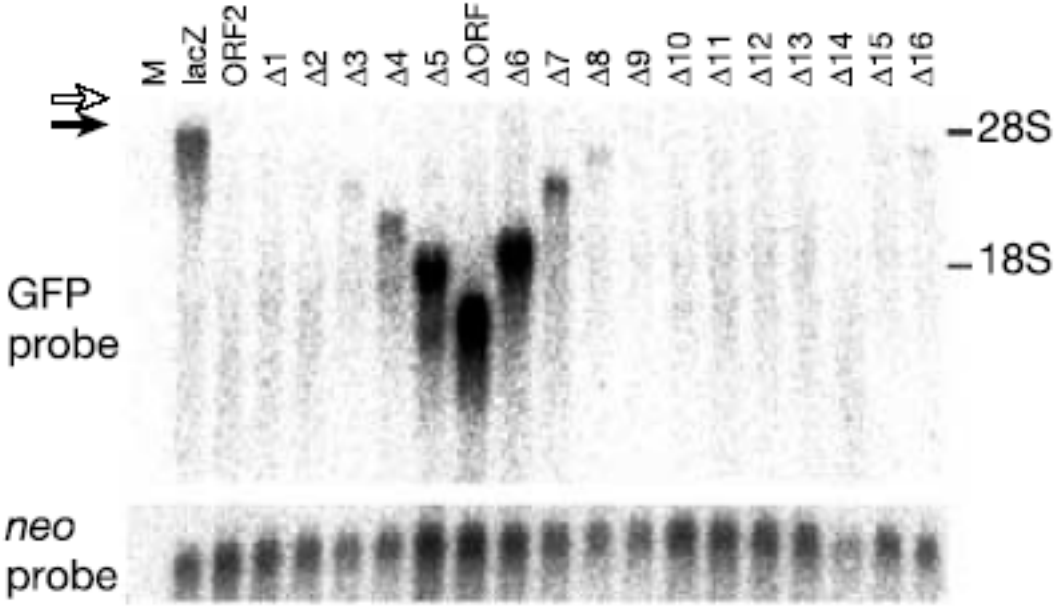


Figure 5 Bioinformatic analysis of L1 content in genes. **a**, Average L1 content of genomic loci of sets of highly (black bars) and poorly (grey bars) expressed genes (see Methods). **b**, Average L1 content in sets of randomly selected populations of genes (see Methods). Positions where the highly 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.

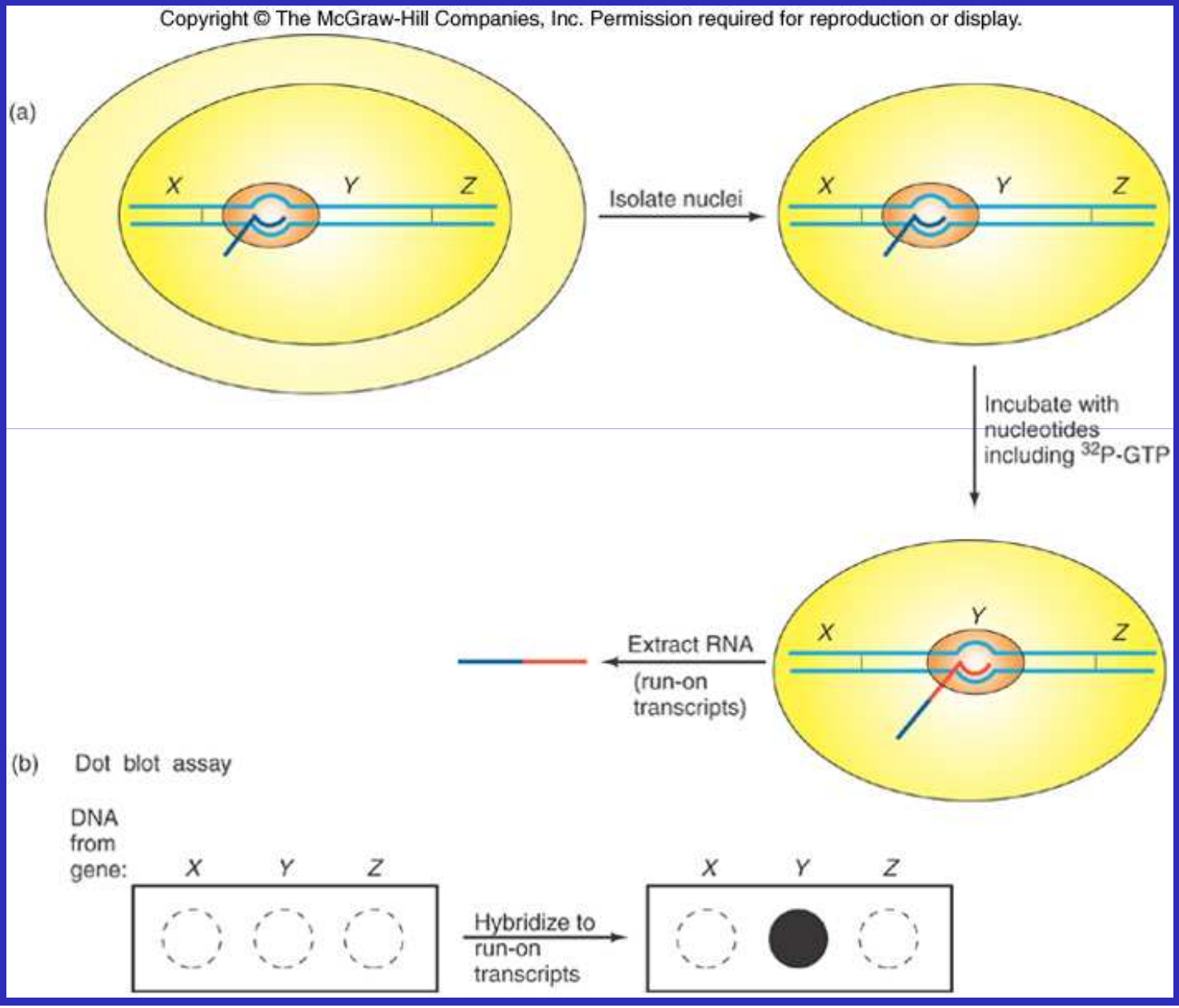
Nuclear Run-On Transcription

- 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

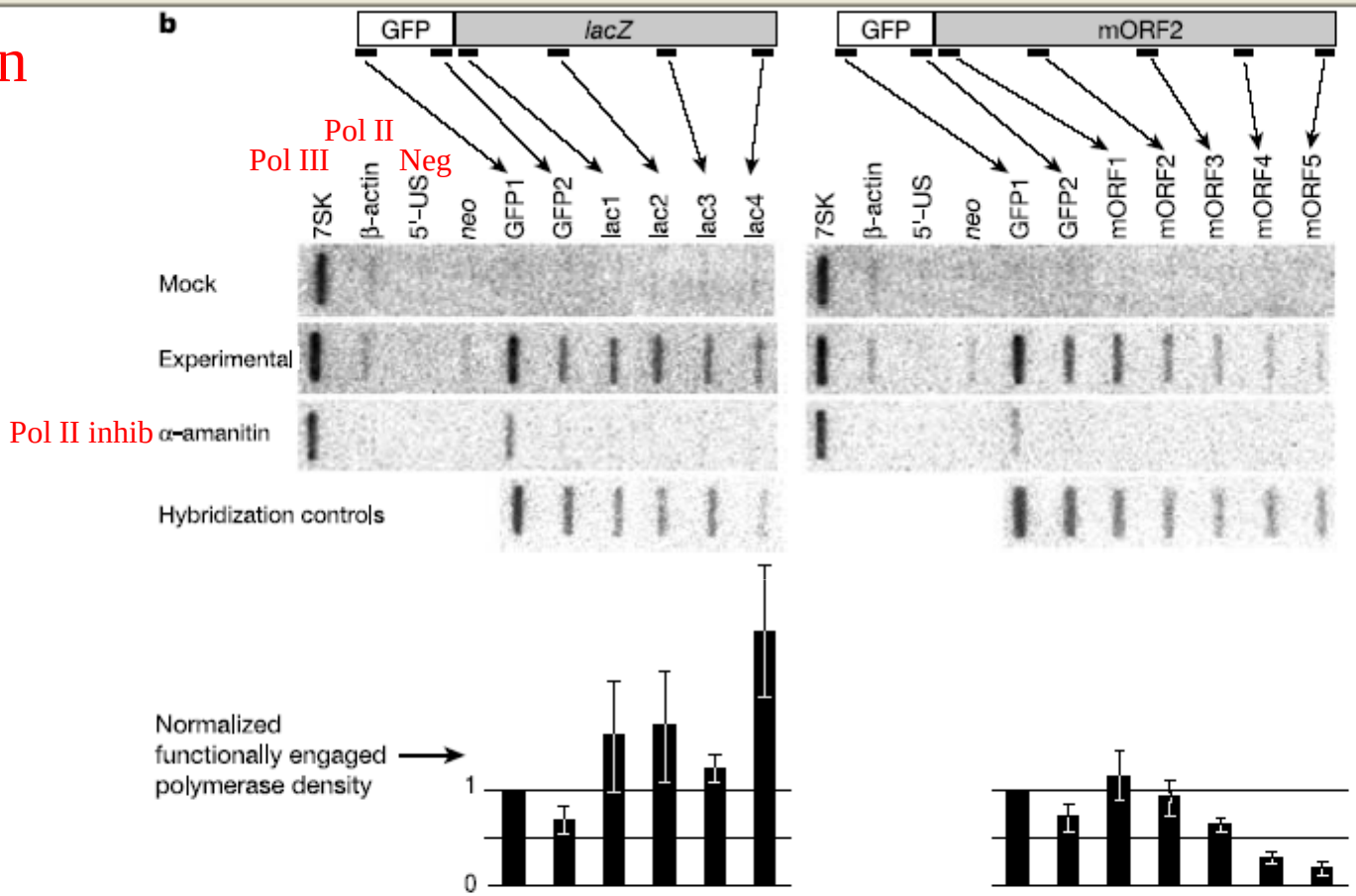


Figure 4 Analysis of ORF2 stability and transcription. **a**, Half-life measurements. In the left panel, transfected HeLa cells were treated with actinomycin D 48 h after transfection. Total RNA was collected 0, 30, 40, 120, 240 and 480 min after treatment and quantified by blotting. In the right panel, the half-lives ($t_{1/2}$) of GFPlacZ or GFPORF2 (relative to the $t_{1/2}$ of the *neo* control) were calculated and compared with $t_{1/2,GFPlacZ}/t_{1/2,neo}$ set to 1. **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 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.

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.

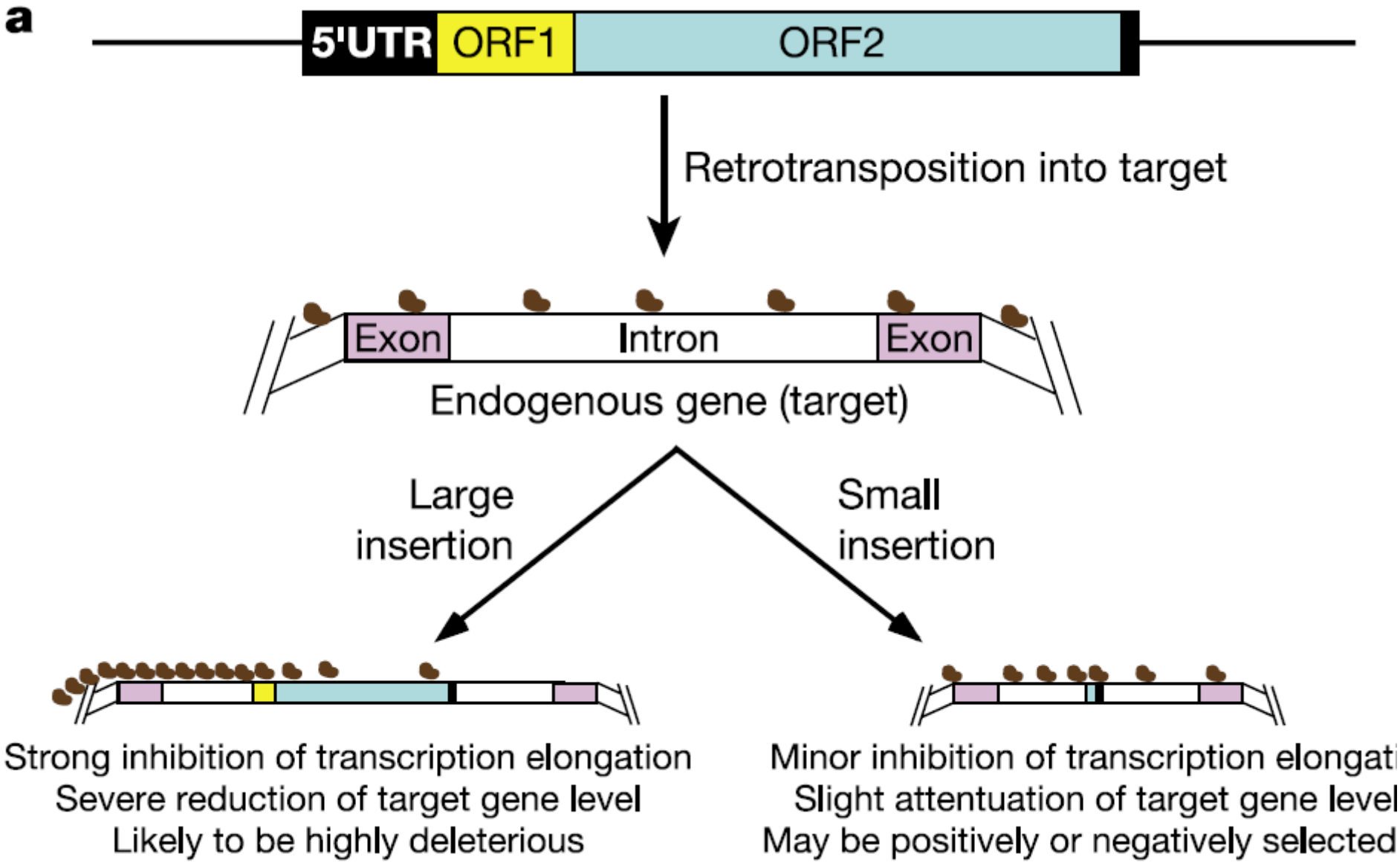

a
5'UTR
ORF1
ORF2
Retrotransposition into target
Exon
Intron
Exon
Endogenous gene (target)
Large insertion
Small insertion
Strong inhibition of transcription elongation
Severe reduction of target gene level
Likely to be highly deleterious
Minor inhibition of transcription elongati
Slight attentuation 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 frames were substituted for the wild-type open reading frames in an established retrotransposition assay[4], transposition levels increased more than 200-fold. This indicates that there are

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 protein levels. Remarkably, when the synthetic open reading frames were substituted for the wild-type open reading frames in an established retrotransposition assay[4], transposition levels increased more than 200-fold. This indicates that there are

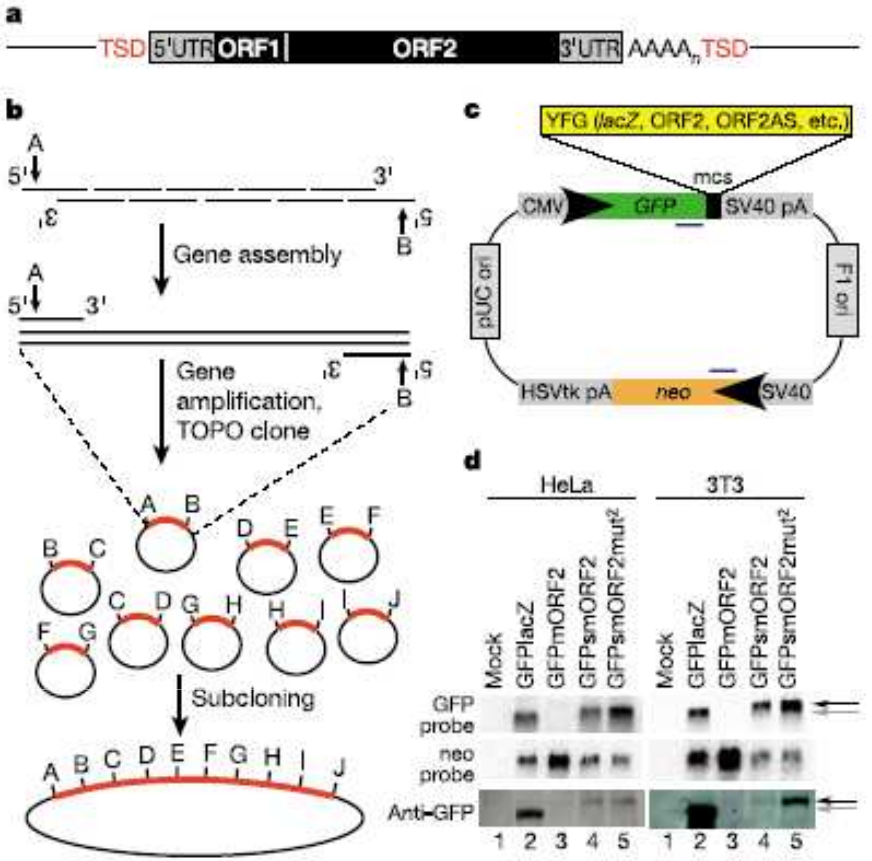
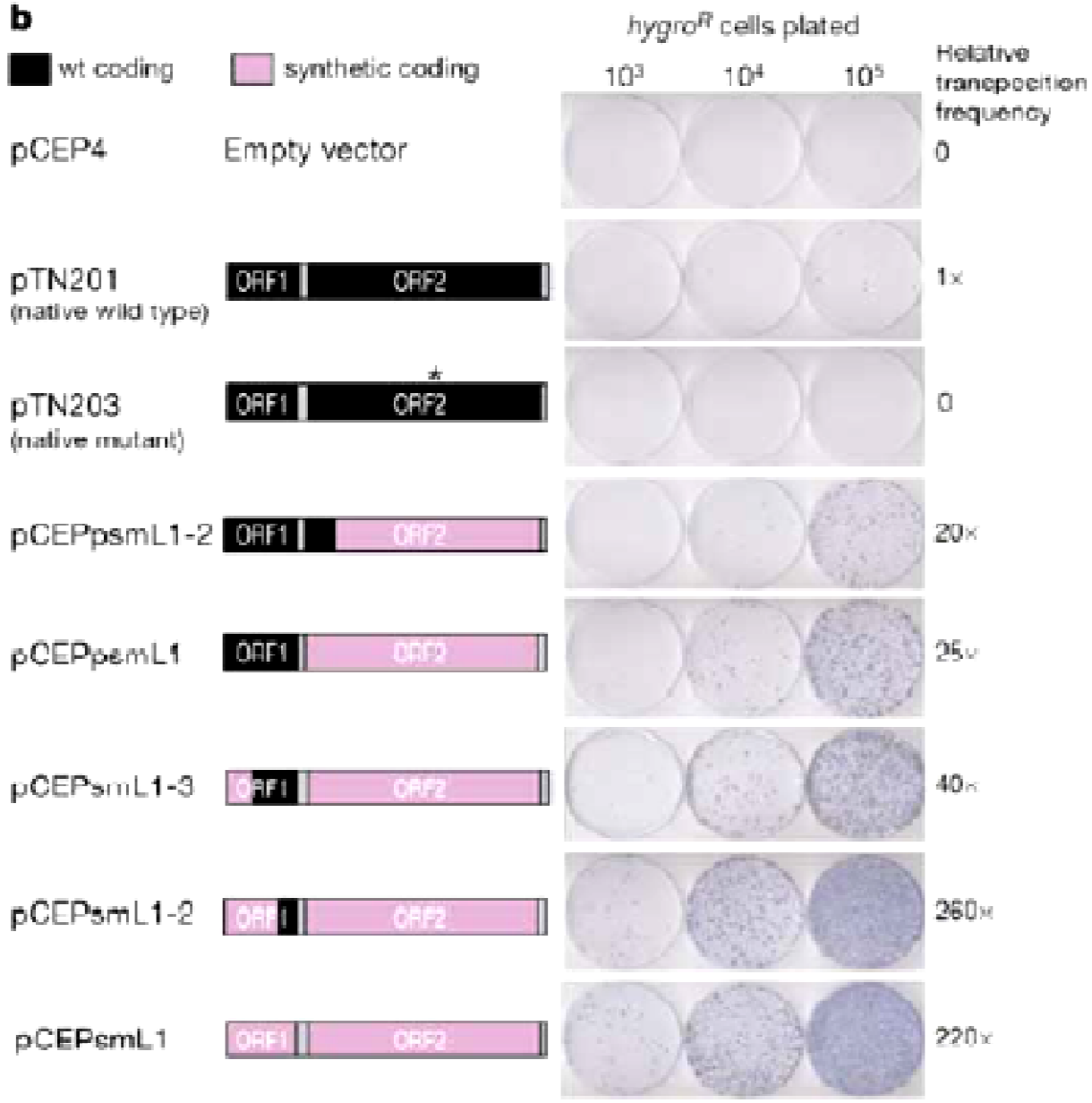


Figure 1 Synthesis and expression of synthetic mouse ORF2. **a**, L1 structure. TSD, target site duplication; UTR, untranslated region. **b**, Overview of gene synthesis. Oligonucleotides encoding each fragment were mixed in a PCR assembly reaction and subsequently used as template amplification. Amplification 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 expression of GFPlacZ, GFPmORF2 and GFPsmORF2. Middle, RNA expression of loading control. Bottom, protein expression of GFPlacZ, GFPmORF2 and GFPsmORF2. Grey and black arrows indicate the expected sizes of GFPlacZ and GFPmORF2/GFPsmORF2, respectively.


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×

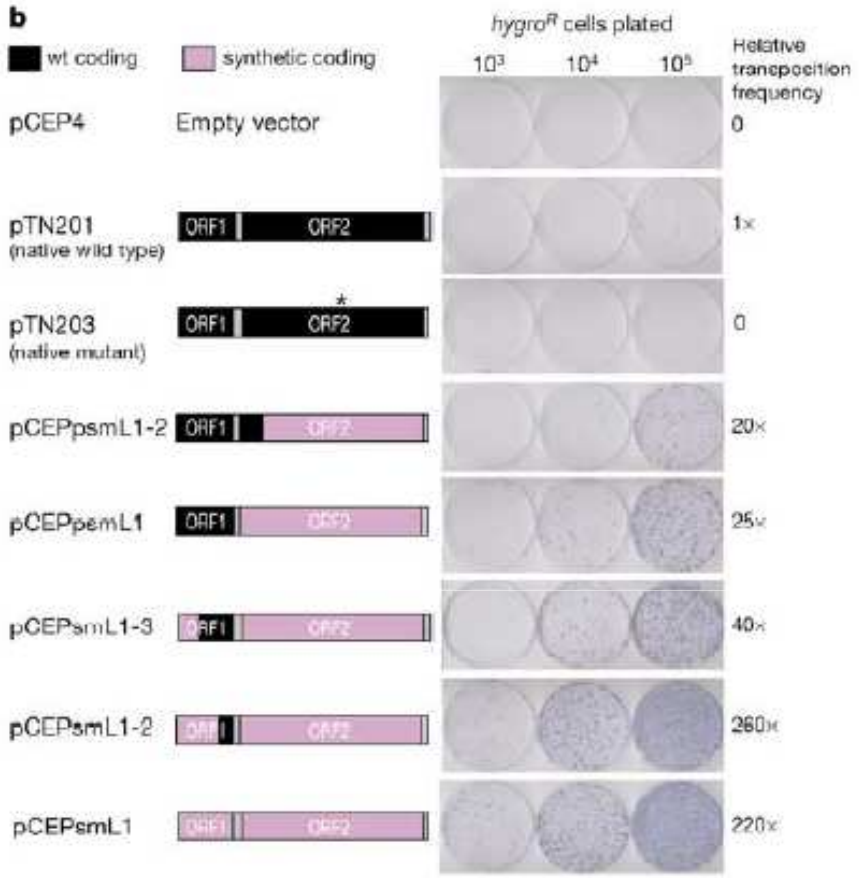


Figure 2 Retrotransposition of synthetic mL1. **a**, The retrotransposition assay. The L1 element contains an intron-interrupted *neo* reporter in the 3′ untranslated region with its own promoter and polyadenylation signal. Only when *neo* is transcribed from the L1 promoter, spliced, reverse transcribed 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.

further increases in retrotransposition, reaching a maximum of more than 200-fold increase over wild type (Fig. 2b) in the element with two fully synthetic ORFs.

To verify that these smL1 G418-resistant colonies resulted from authentic L1 retrotransposition, we characterized six smL1 insertions. The mutant loci were identified by inverse polymerase chain reaction (PCR), enabling the amplification of each complete insertion and flanking sequence. For each primer pair, parental HeLa cells produced only empty site products (Fig. 3a, odd-numbered lanes), whereas the respective G418-resistant clones produced both empty site and filled smL1 insertion products of predicted sizes (Fig. 3a, even-numbered lanes). Amplicons were cloned and sequenced to determine their general structures and genomic flanks, summarized in Fig. 3b. All showed a properly spliced *neo* gene, a poly(A) tail, and most (five of six) had target site duplications 5–108 bp long. Insertion no. 10 had a 10-bp target deletion and insertion no. 18 had a 5′ L1 inversion, features commonly found in L1 insertions[13–15]. In addition, various chromosomes served as targets, and the endonuclease cleavage sites inferred from target site duplications matched the previously reported degenerate consensus (5′-TTTT/AA-3′ on the bottom strand)[3,16] (Fig. 3c). Taken together, these results suggest that the synthetic L1 retrotransposes

Table 1 **High-frequency retrotransposition in mouse cells**

Plasmid	Relative transposition frequency		
	HeLa	3T3	L
pCEP4 (empty vector)	0	0	0
pTN201 (native mouse wild type)	<0.005	<0.002	<0.002
pTN203 (native mouse mutant)	0	0	0
pJM101L1 (native human wild type)	0.13	0.017	0.07
pCEPsmL1 (synthetic mouse wild type)	1	1	1
pCEPsmL1mut[2] (synthetic mouse mutant)	0	0	<0.002

With the use of the transient assay[17], synthetic mouse L1 (pCEPsmL1) retrotransposition frequency was compared with that of wild-type native human L1 and wild-type native mouse L1 ($N = 3$). The average absolute numbers of colonies of pJM101L1rp (colonies per 10^6 transfected cells) for HeLa, 3T3 and L cells were 2,904, 108 and 1,568, respectively.

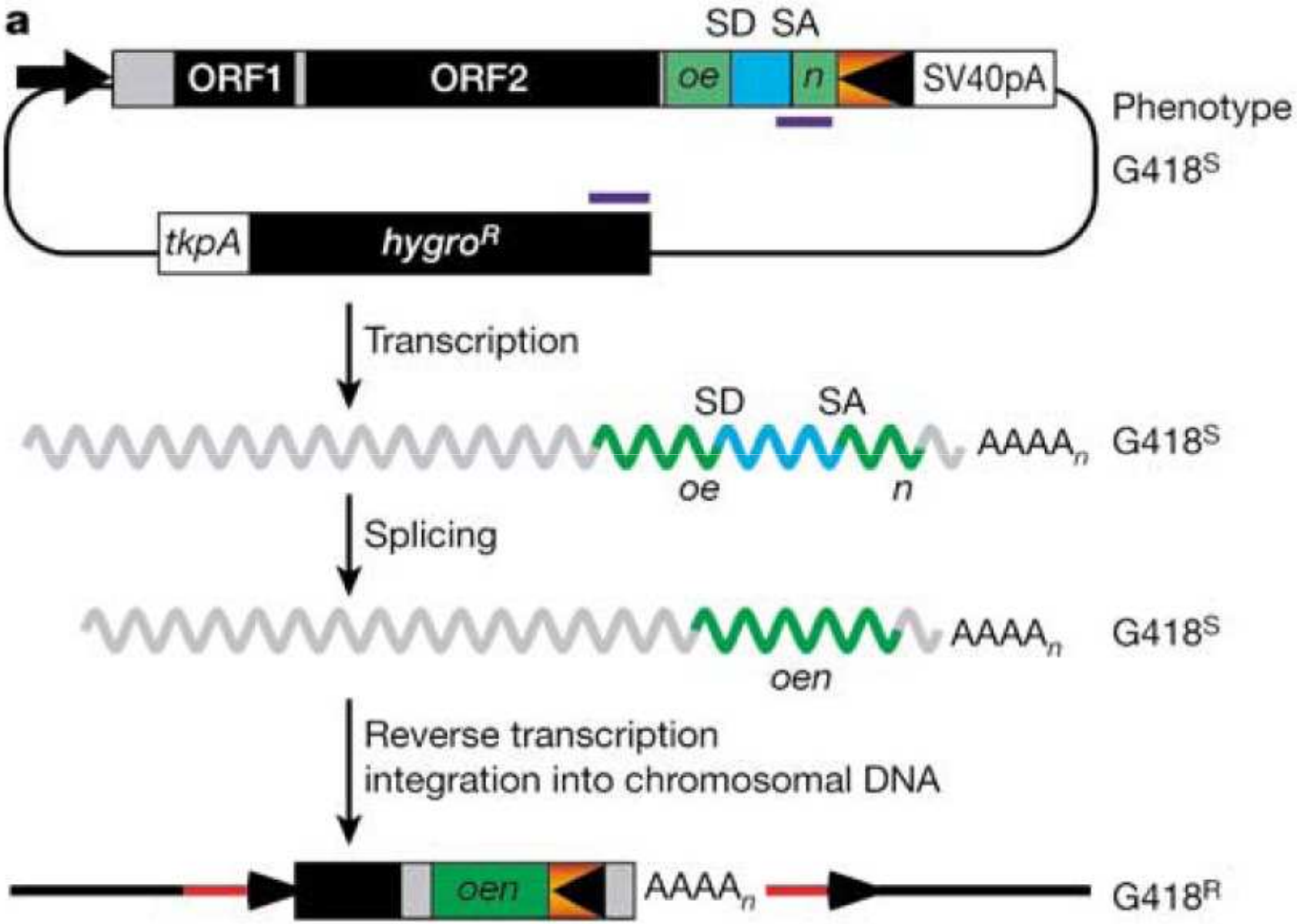
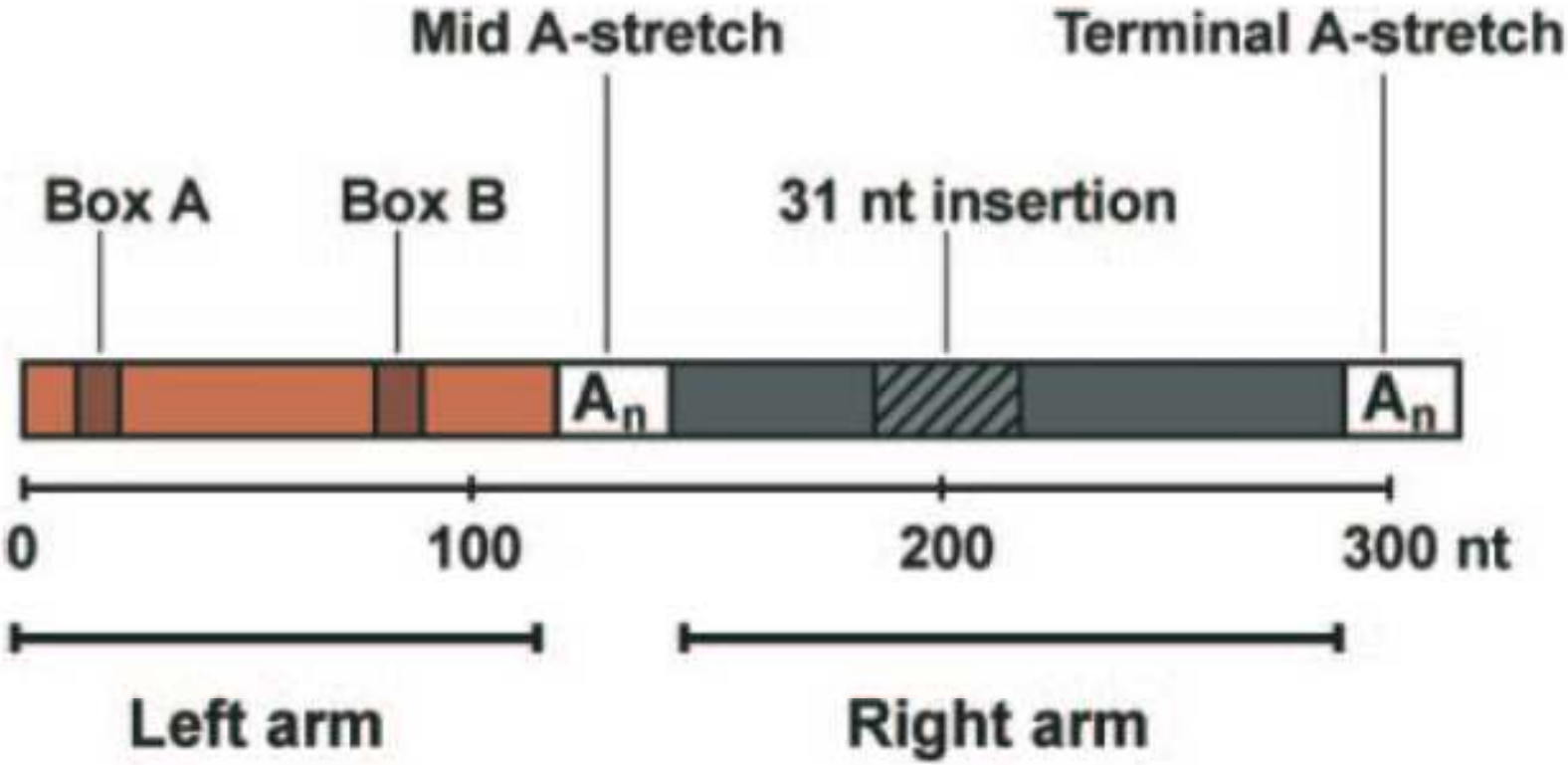

a
SD SA
ORF1
ORF2
oe
n
SV40pA
Phenotype
G418S
tkpA
hygroR
Transcription
SD
SA
AAAAn
G418S
oe
n
Splicing
AAAAn
G418S
oen
Reverse transcription
integration into chromosomal DNA
oen
AAAAn
G418R

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.

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.


Mid A-stretch
Terminal A-stretch
Box A
Box B
31 nt insertion
An
An
0
100
200
300 nt
Left arm
Right arm

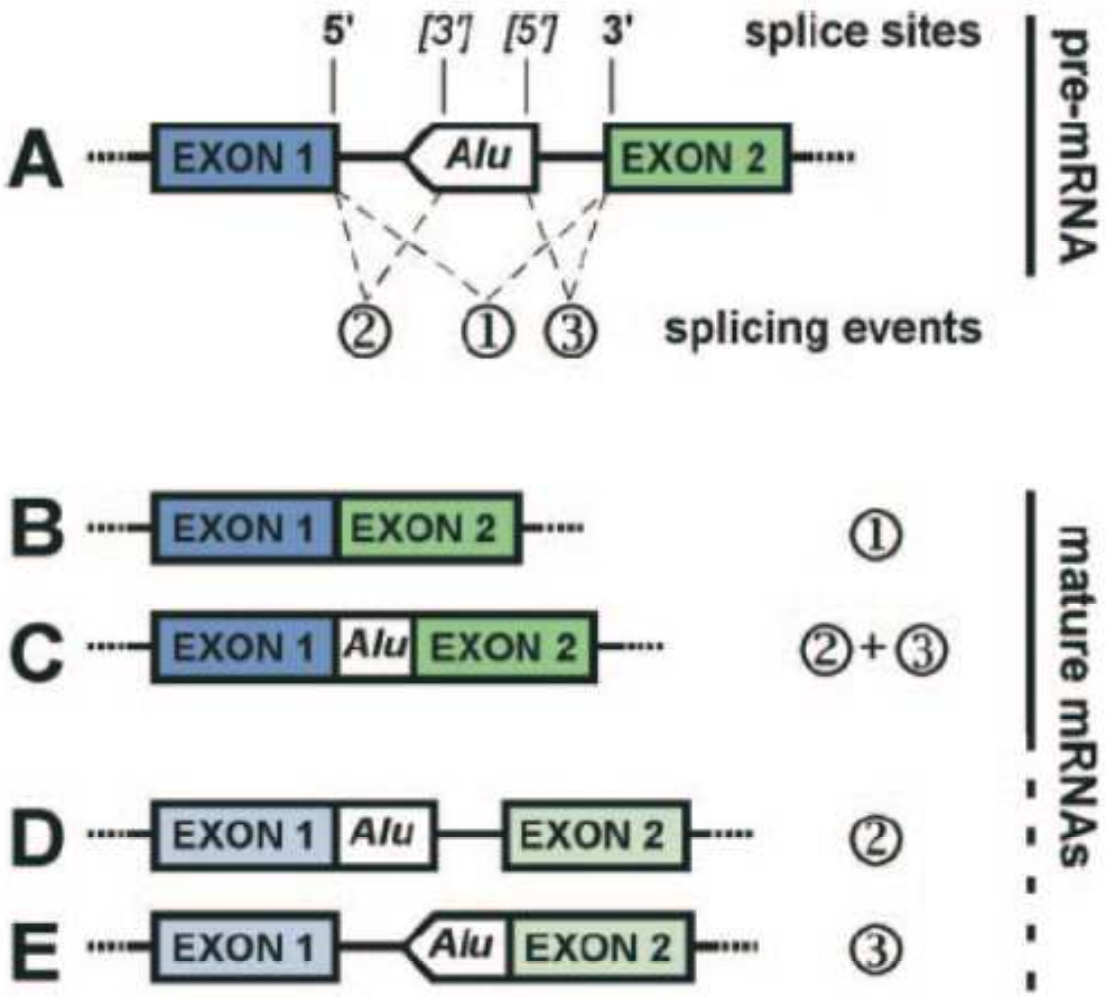


Figure 2. Exonization of intronic *Alu* elements. Hypothetical *Alu* element inserted in an orientation that opposes the sense of transcription in an intronic region. This element has a major 3′ splice site near position 275 and a major 5′ splice site near position 158 as described in the text. The use of alternative splice sites of the *Alu* element can lead to a variety of mature mRNAs. (**A**) Pre-mRNA with possible splicing events (1; 2; 3). 5′ and 3′splice sites are indicated. Alternative splice sites of the *Alu* element are under brackets. (**B**) Regular splicing without *Alu* exonization resulting from splicing event 1. (**C**) Exonization of the intronic *Alu* element by use of its 5′ and 3′ splice sites

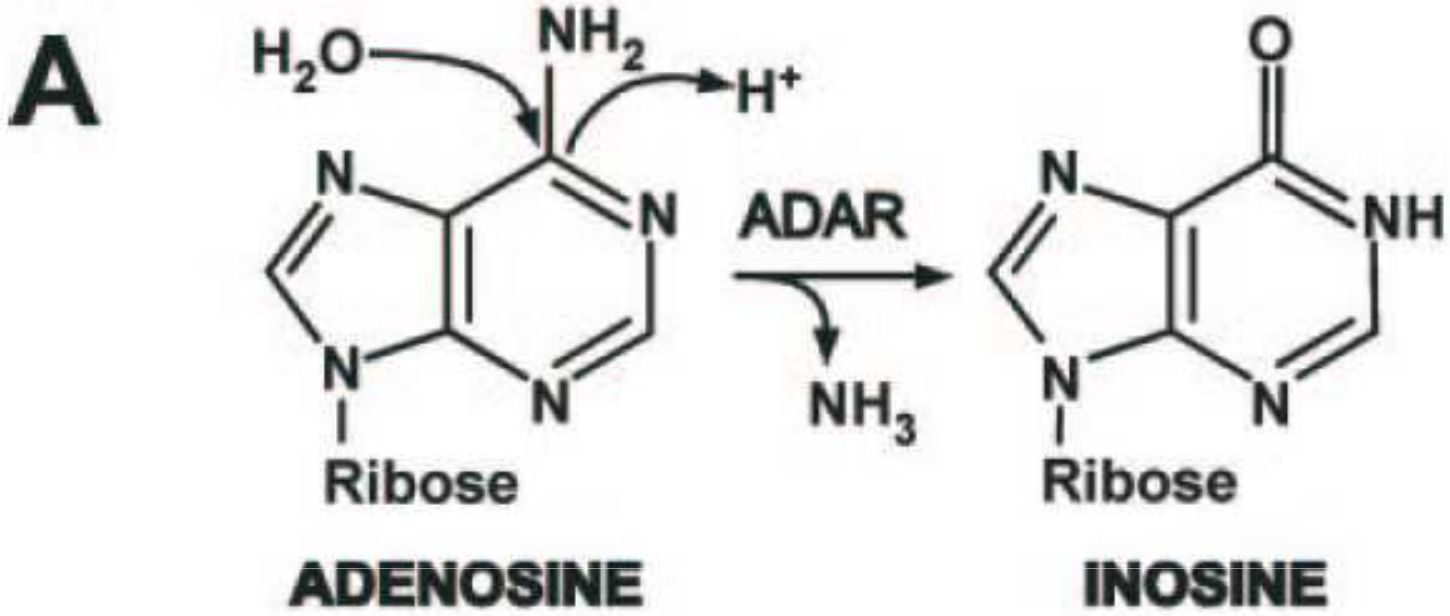
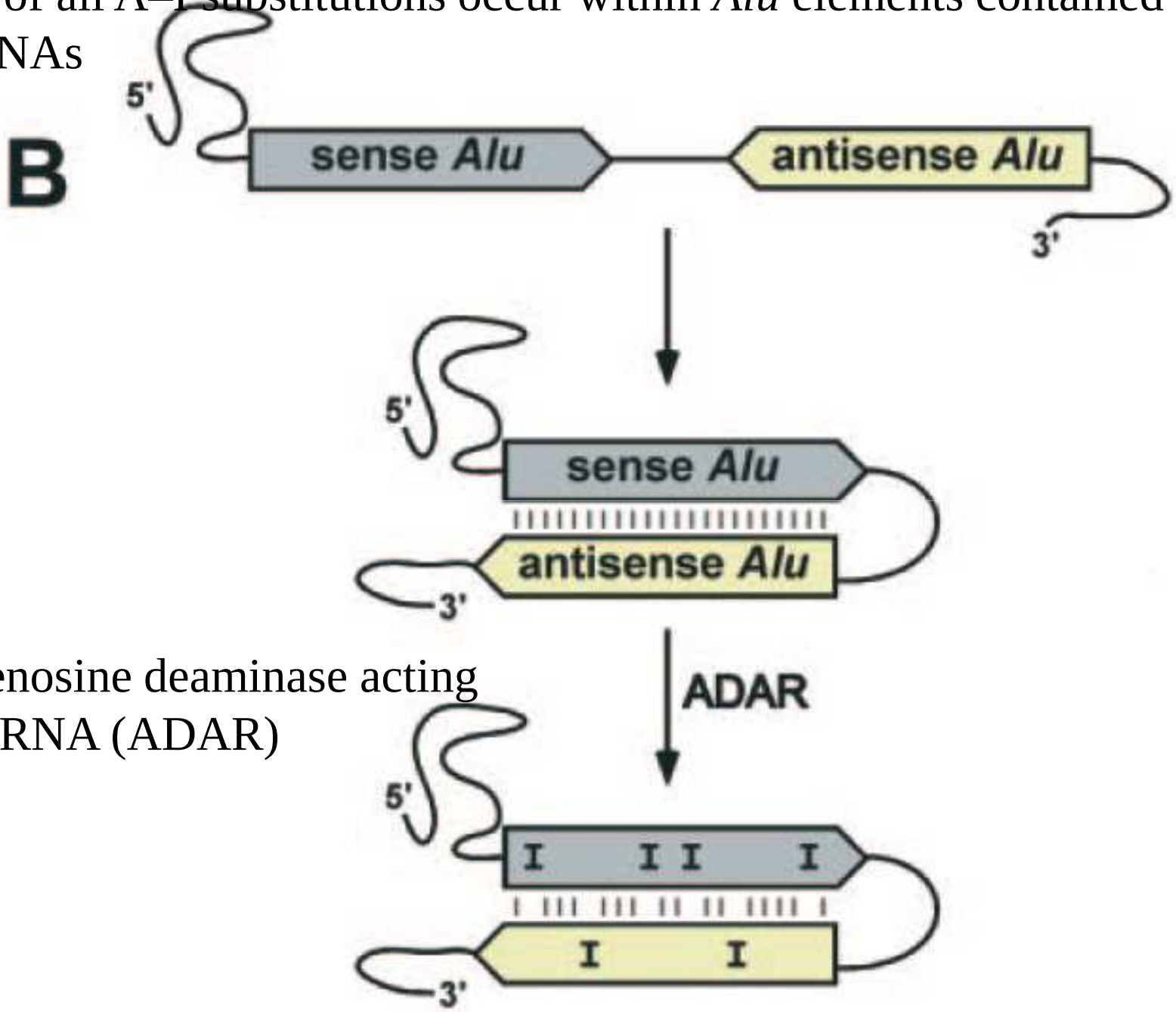


Figure 3. A-to-I editing in *Alu* elements. (**A**) Deamination of adenosine by ADAR leading to the production of inosine. (**B**) Intramolecular base paring of a mRNA containing two *Alu* elements in opposite orientation. Base pairing between the two *Alu* elements leads to the formation of a long stable double-stranded RNA region in which ADAR performs A-to-I substitutions (marked as I).

>90% of all A–I substitutions occur within *Alu* elements contained in mRNAs



adenosine deaminase acting on RNA (ADAR)

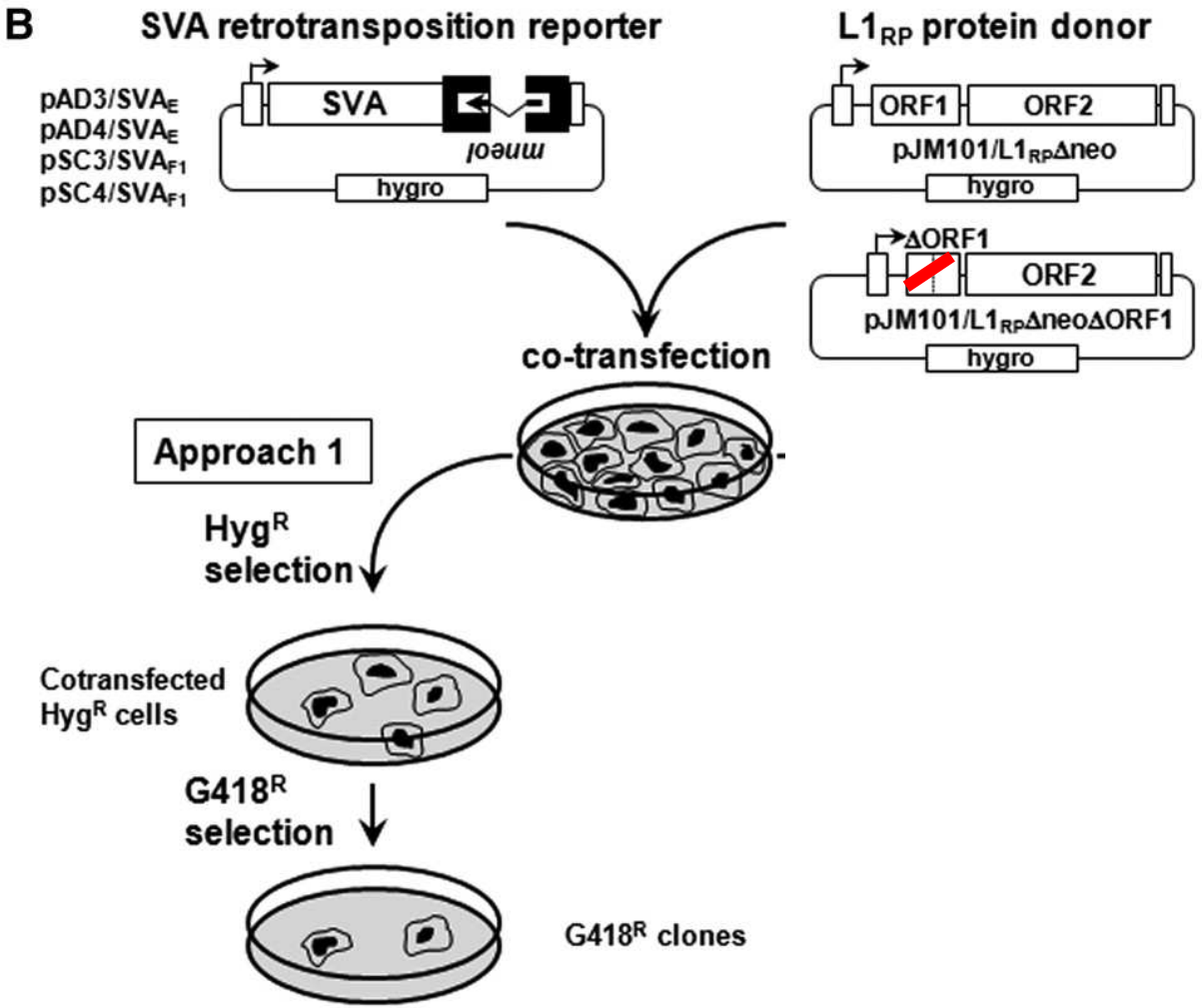
In summary, *Alu* elements play an important role in editing of the human transcriptome by providing ideal templates to the ADAR family of enzymes. The large number of *Alu* elements present in mRNAs and their relatively low divergence explains why they are more prone to be edited than other sequences, and why widespread editing is specific to primates (39). Although the precise role of RNA editing is still speculative, it might affect gene expression at several steps. As inosine does not base pair with uracyl but with cytosine, editing might influence the stability of RNA molecules by creating and disrupting secondary structures. At another level, as inosine is recognized as guanosine by the translation and splicing machineries, A–I editing could lead to amino acid substitutions in the coding sequence, or to modification of splice sites in introns that could induce premature termination or frame-shifts. Knowing that aberrant editing is found in several neurological disorders (40,41), it is highly probable that the phenomenon is of great physiological importance.

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

Rationale of the SVA trans-mobilization assay.

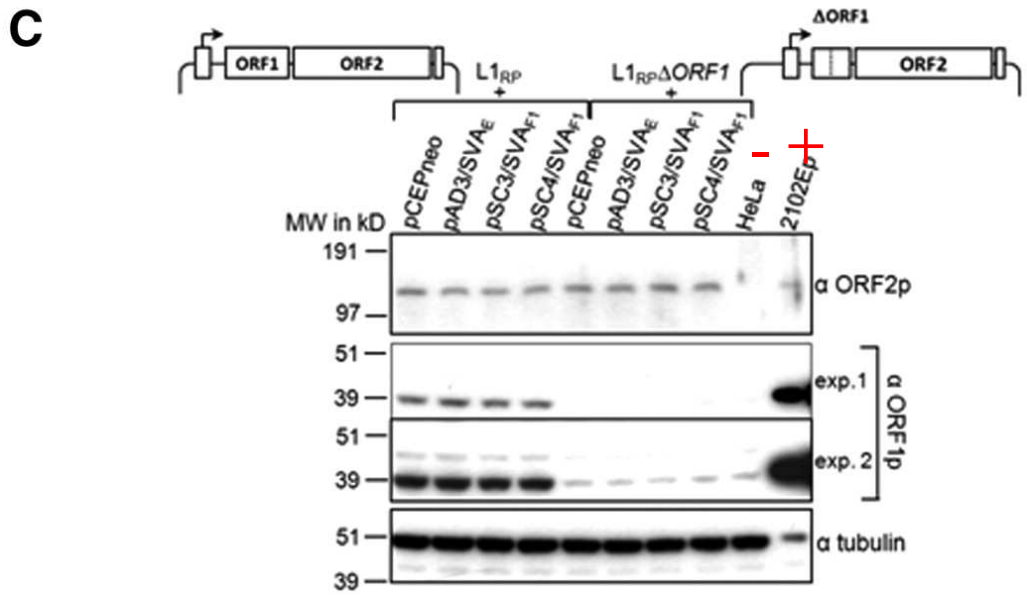


Raiz J et al. Nucl. Acids Res. 2011;nar.gkr863

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.

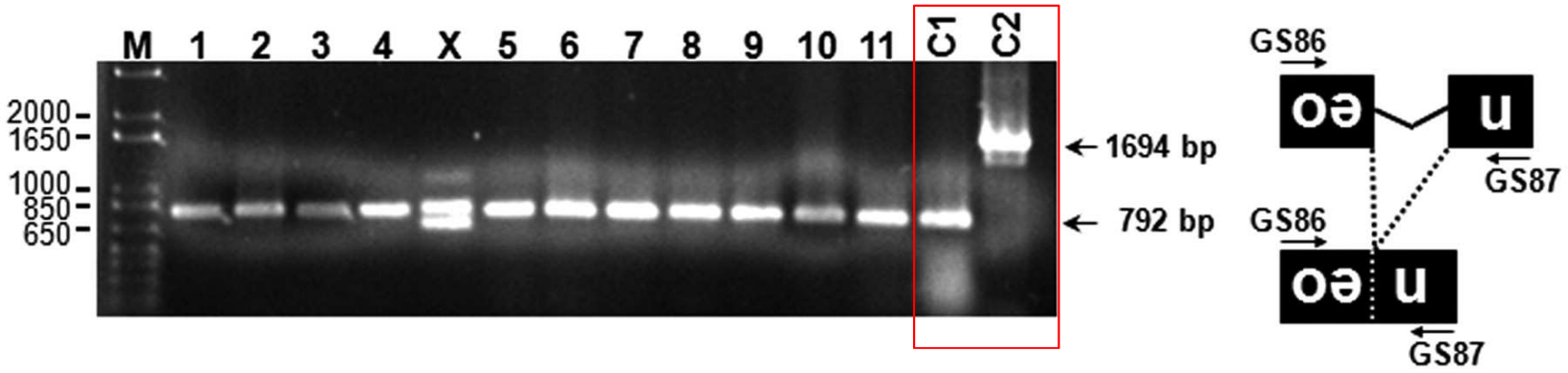


Raiz J et al. Nucl. Acids Res. 2011;nar.gkr863

Nucleic Acids Research

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

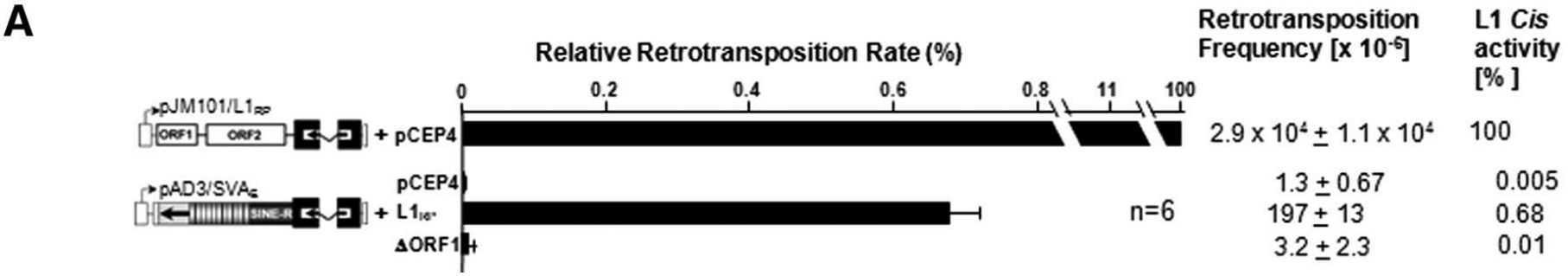
PCR to test for correct splicing of the intron from the mneoI cassette



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

Nucleic Acids Research

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



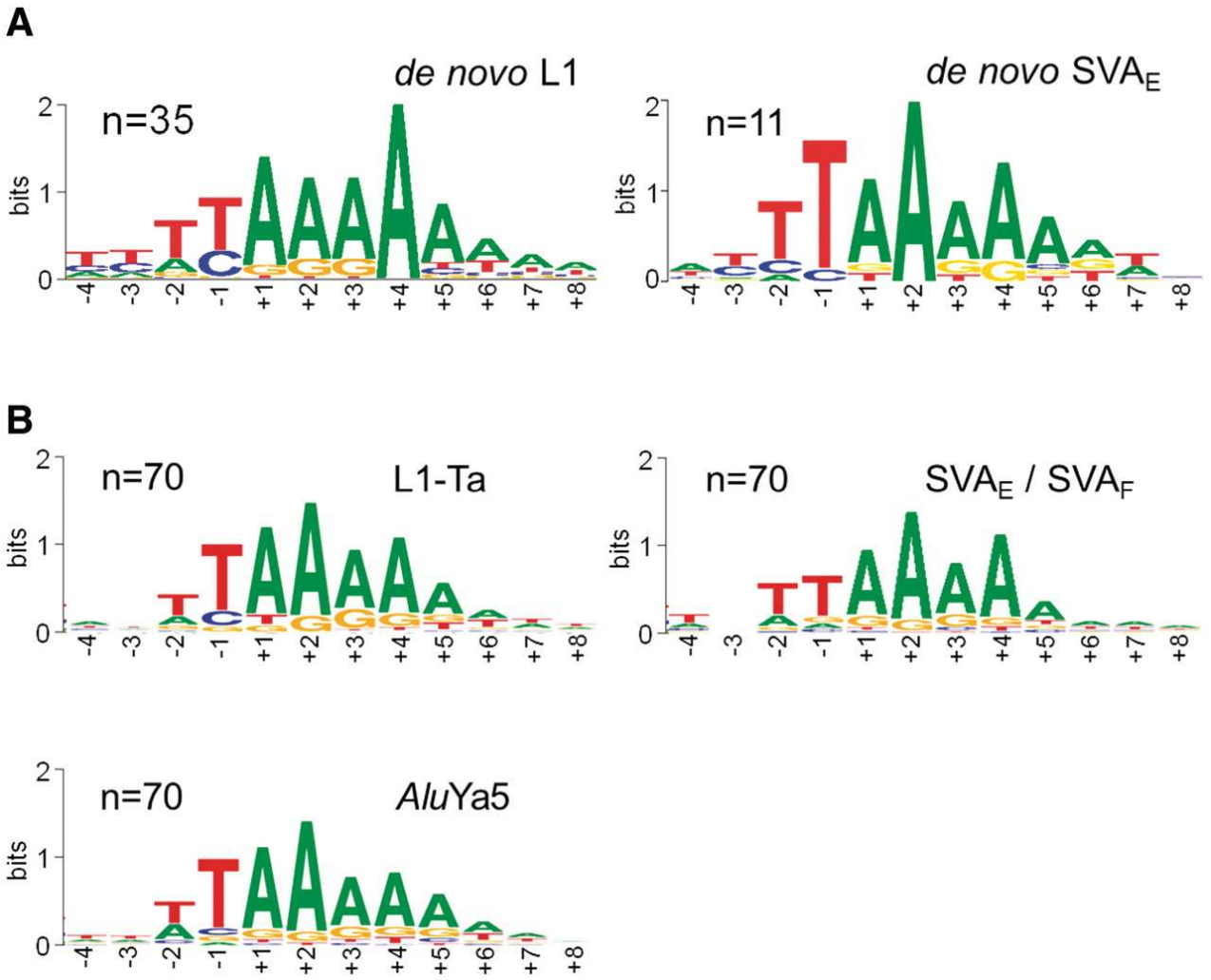
pJM101/L1RP =L1 cis activity Controllo positivo 100%

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

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

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

Nucleic Acids Research