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).

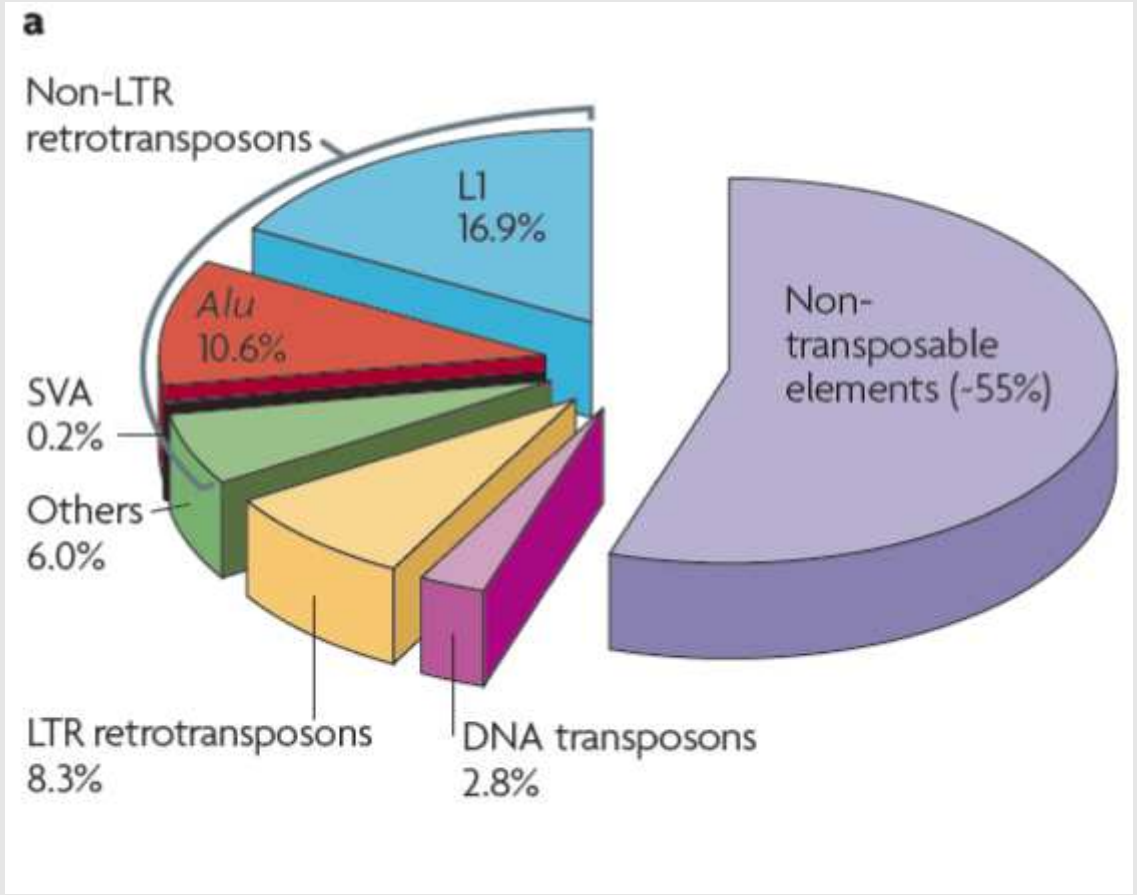
LTR retrotransposons and retroviruses are

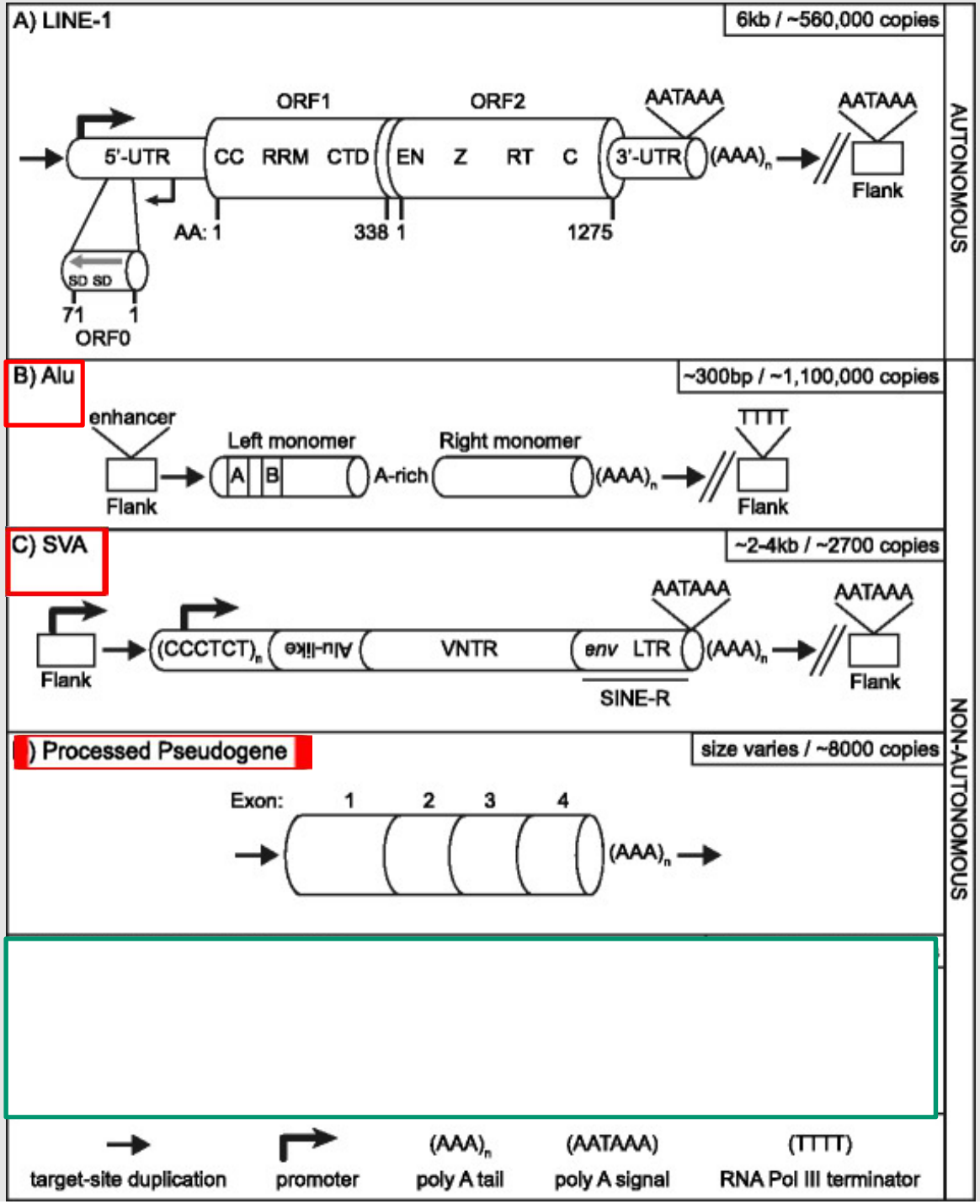
Watson BM Gene Capitolo 11

Roles for retrotransposon insertions in human disease

Over evolutionary time, the dynamic nature of a genome is driven, in part, by the activity of transposable elements (TE) such as retrotransposons.

On a shorter time scale it has been established that new TE insertions can result in single-gene disease in an individual.

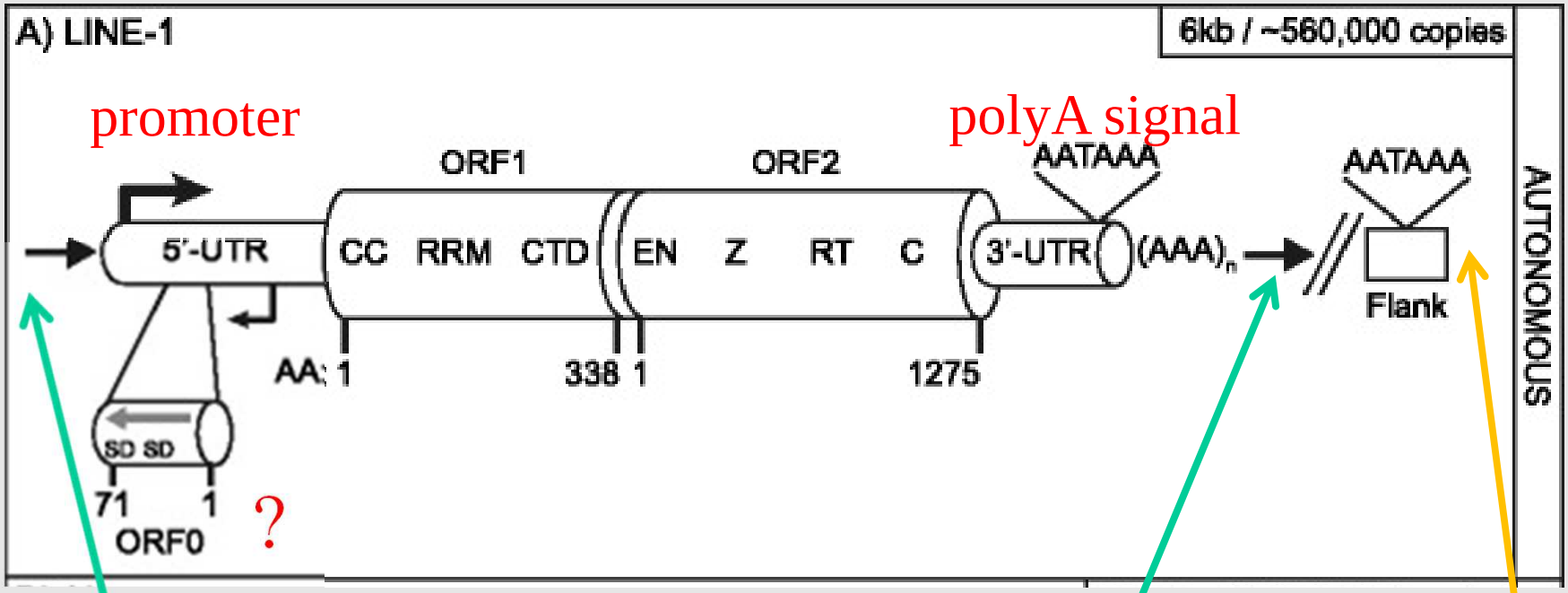

a
Non-LTR retrotransposons
L1 16.9%
Alu 10.6%
SVA 0.2%
Others 6.0%
LTR retrotransposons 8.3%
DNA transposons 2.8%
Non-transposable elements (~55%)


A) LINE-1
6kb / ~560,000 copies
ORF1
ORF2
AATAAA
AATAAA
5'-UTR
CC
RRM
CTD
EN
Z
RT
C
3'-UTR
(AAA)n
Flank
AA: 1
338 1
1275
SD SD
71
1
ORF0
AUTONOMOUS
B) Alu
~300bp / ~1,100,000 copies
enhancer
Left monomer
Right monomer
TTTT
A
B
A-rich
(AAA)n
Flank
Flank
C) SVA
~2-4kb / ~2700 copies
AATAAA
AATAAA
Flank
(CCCTCT)n
Alu-like
VNTR
env
LTR
(AAA)n
Flank
SINE-R
) Processed Pseudogene
size varies / ~8000 copies
Exon:
1
2
3
4
(AAA)n
NON-AUTONOMOUS
target-site duplication
promoter
(AAA)n
poly A tail
(AATAAA)
poly A signal
(TTTT)
RNA Pol III terminator

LINE-1

The non-LTR retrotransposon Long INterspersed Element-1 (or L1) is the only active autonomous TE.

In addition to mobilizing its own RNA to new genomic locations via a “copy-and-paste” mechanism, LINE-1 is able to retrotranspose other RNAs including Alu, SVA, and occasionally cellular RNAs.



flanking target-site duplications

LINE-1 frequently bypassed its own polyA signal in favor of a downstream one

Trascrizione L1

- RNApolymeraseII
- Transcription factors

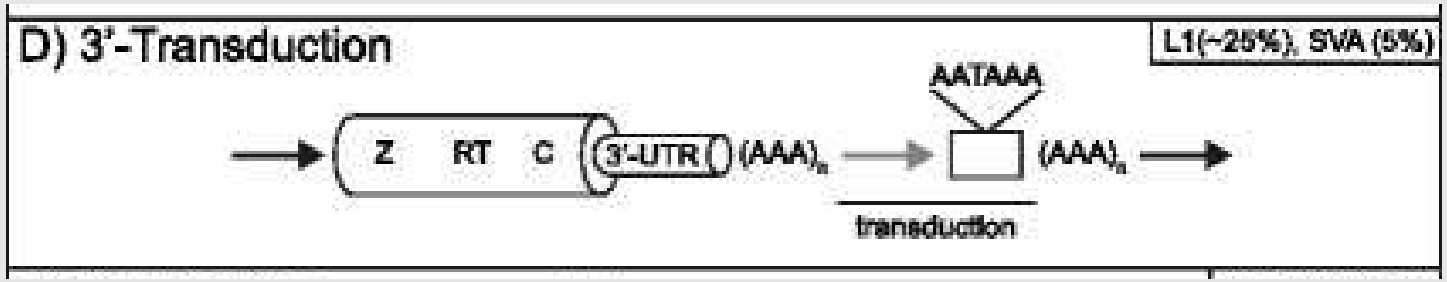
SOX11

YY1

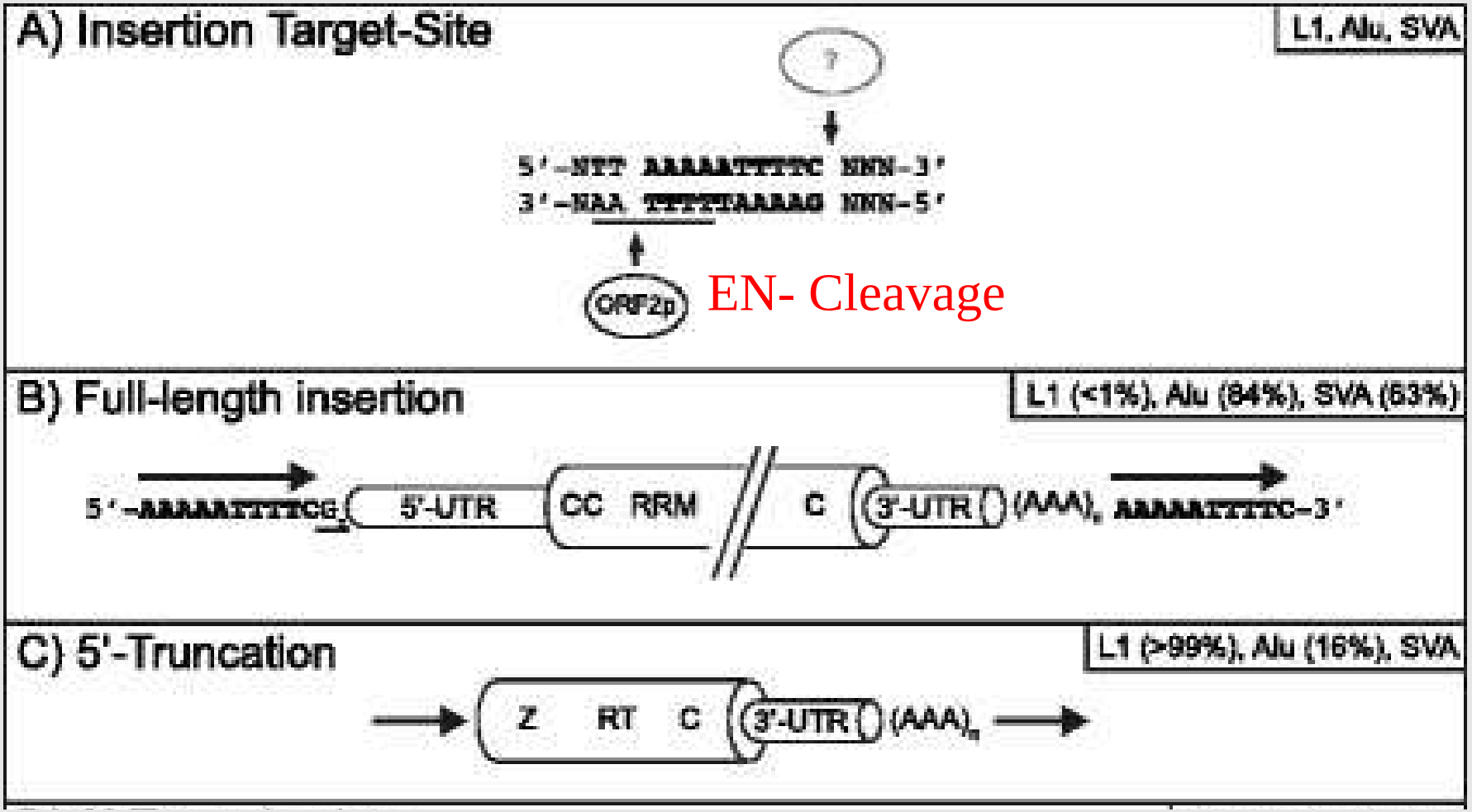
RUNX3

p53?

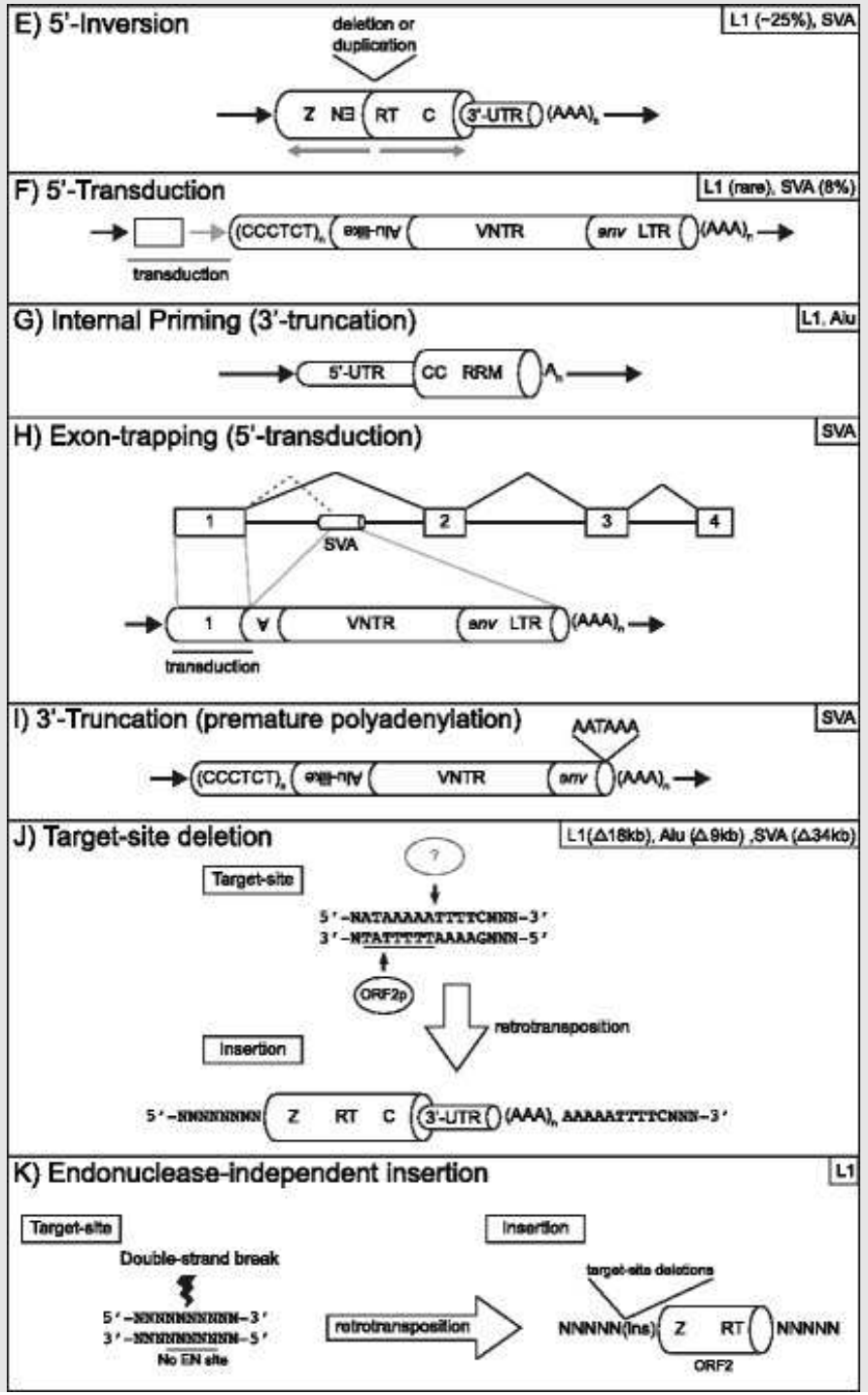
LINE-1 frequently bypassed its own polyA signal (AATAAA) in favor of a downstream one (AATAAA)



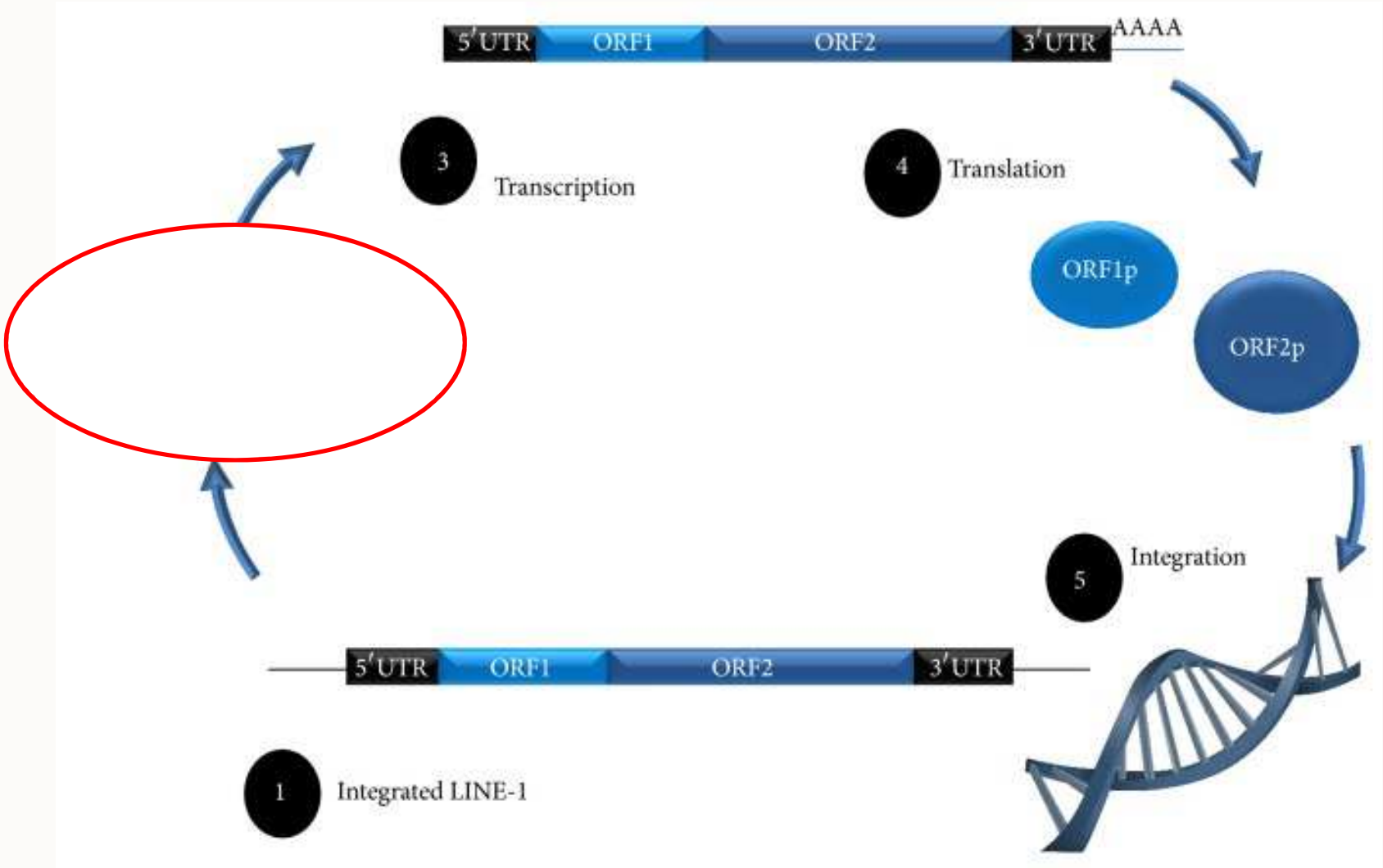
retrotransposon insertions



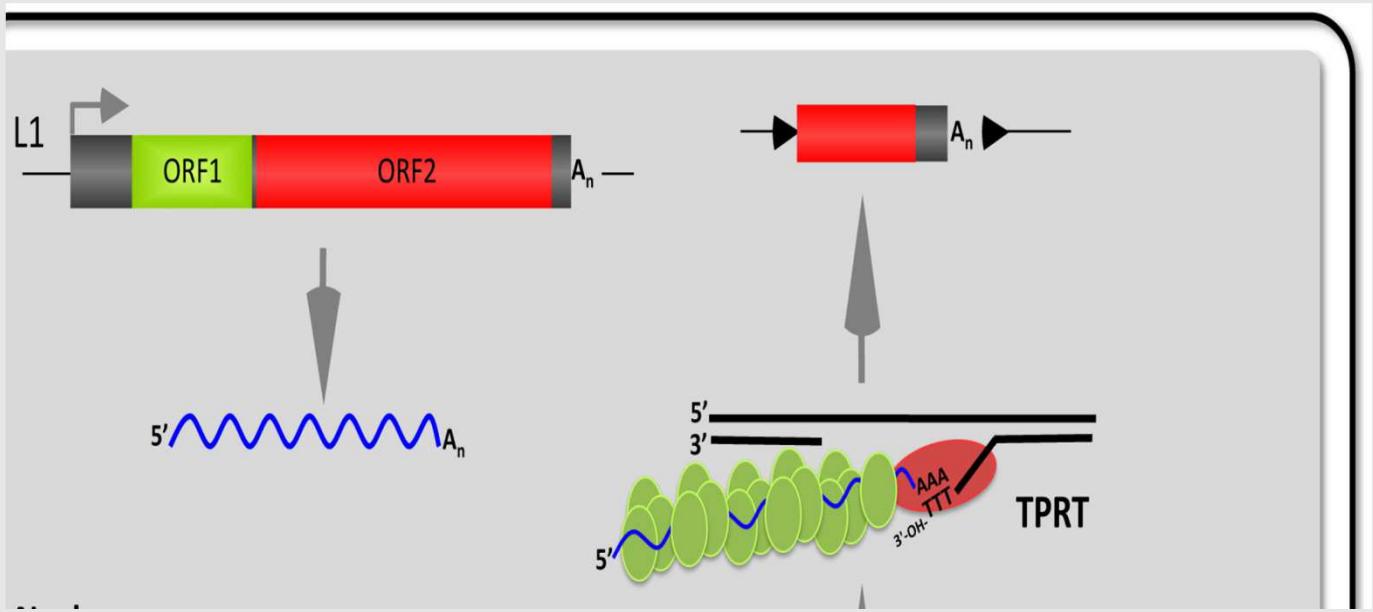
Numerosi altri eventi mutazionali

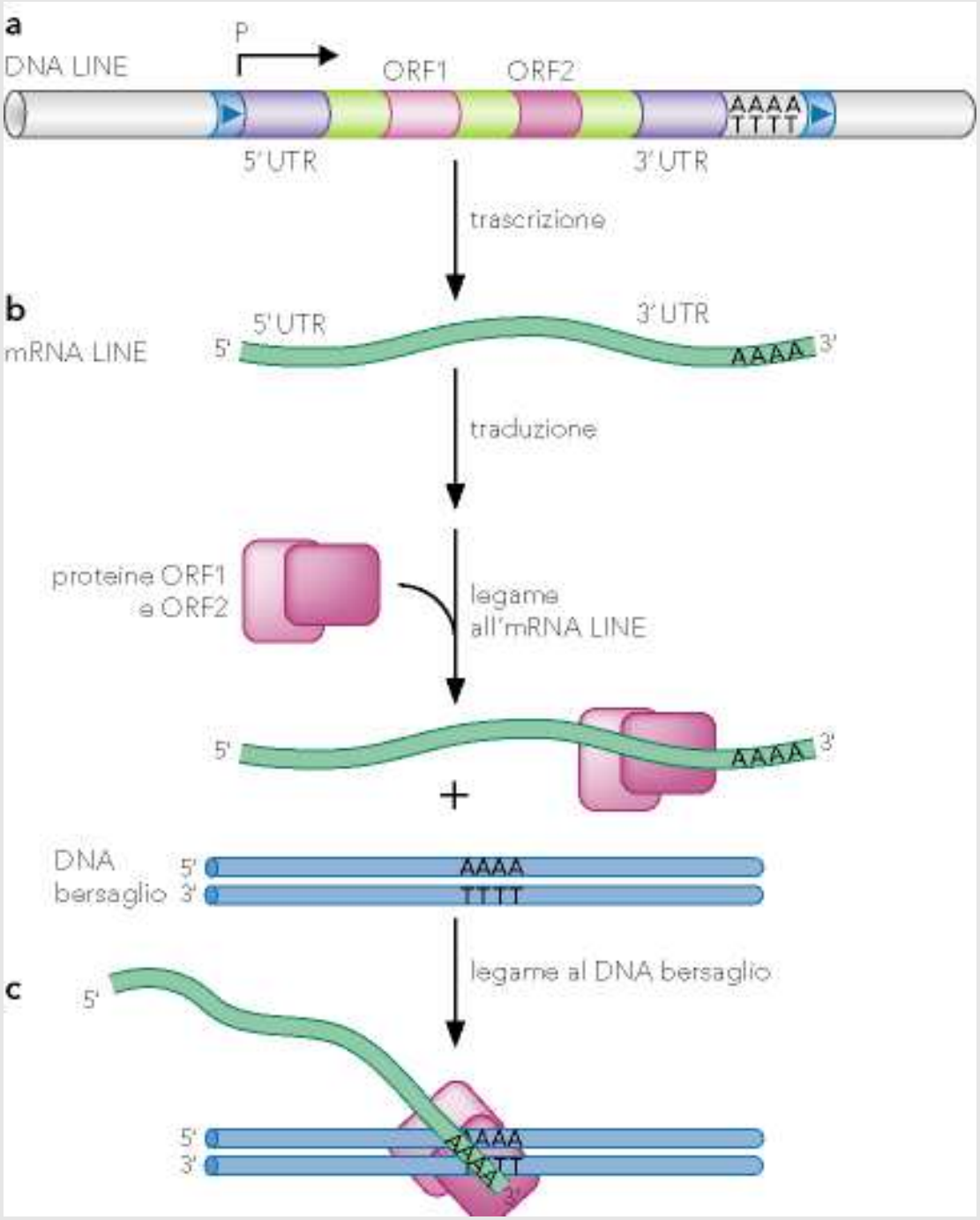


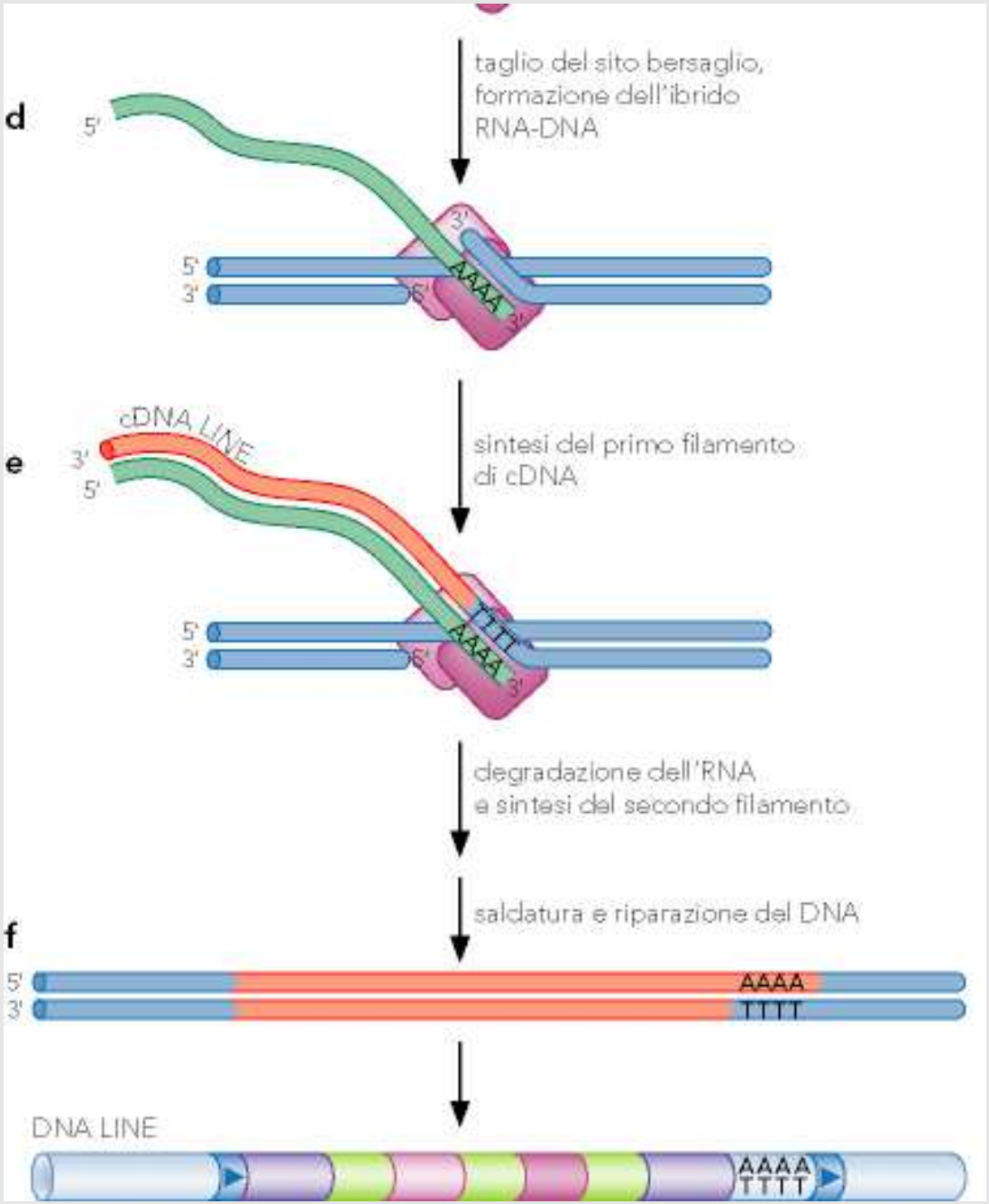
la trasposizione


5′UTR
ORF1
ORF2
3′UTR
AAAA
3
Transcription
4
Translation
ORF1p
ORF2p
5
Integration
5′UTR
ORF1
ORF2
3′UTR
1
Integrated LINE-1

target-site primed reverse transcription

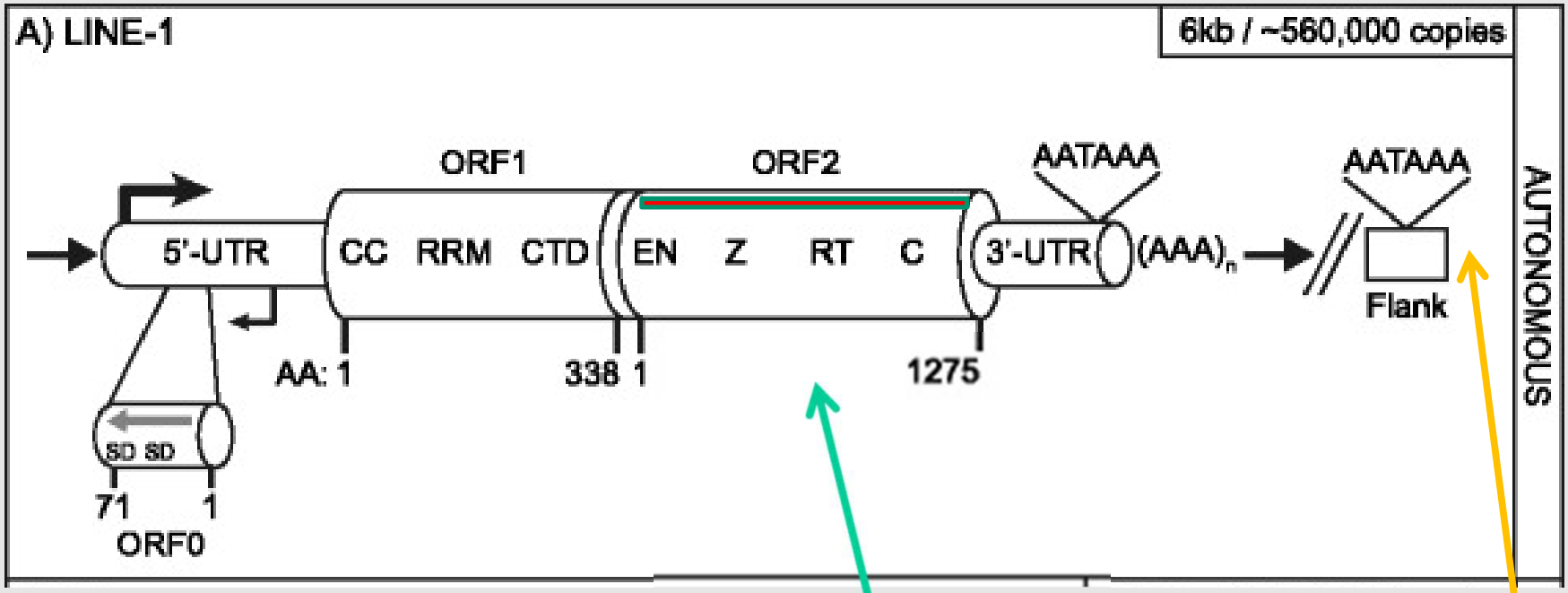



a
DNA LINE
P
ORF1
ORF2
AAAA
TTTT
5' UTR
3' UTR
trascrizione
b
mRNA LINE
5' UTR
3' UTR
5'
AAAA
3'
traduzione
proteine ORF1
e ORF2
legame
all'mRNA LINE
5'
AAAA
3'
+
DNA
bersaglio
5'
3'
AAAA
TTTT
legame al DNA bersaglio
c
5'
5'
3'
AAAA
TT
3'

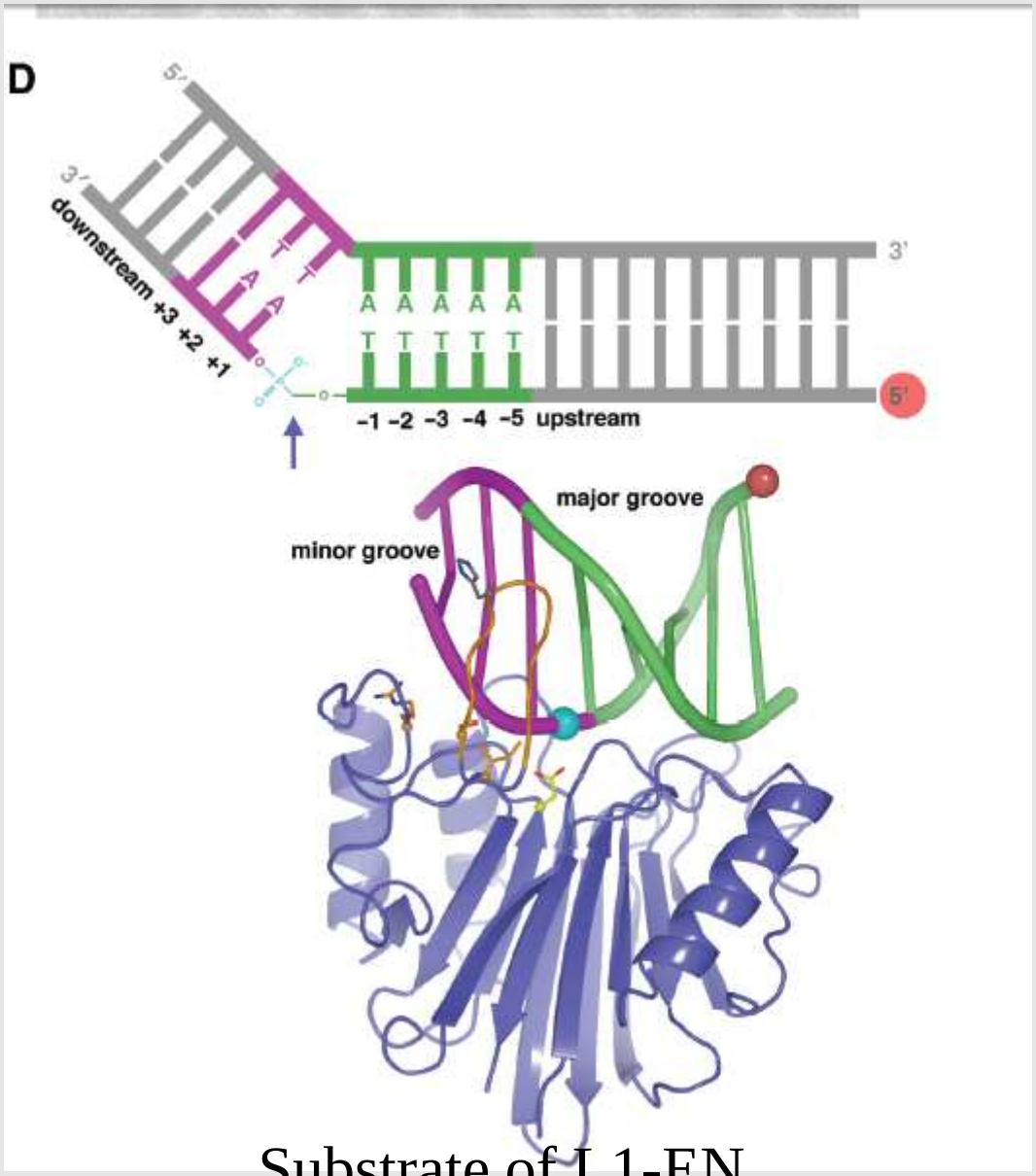


DNA damage response

Le proteine ed ENZIMI della trasposizione

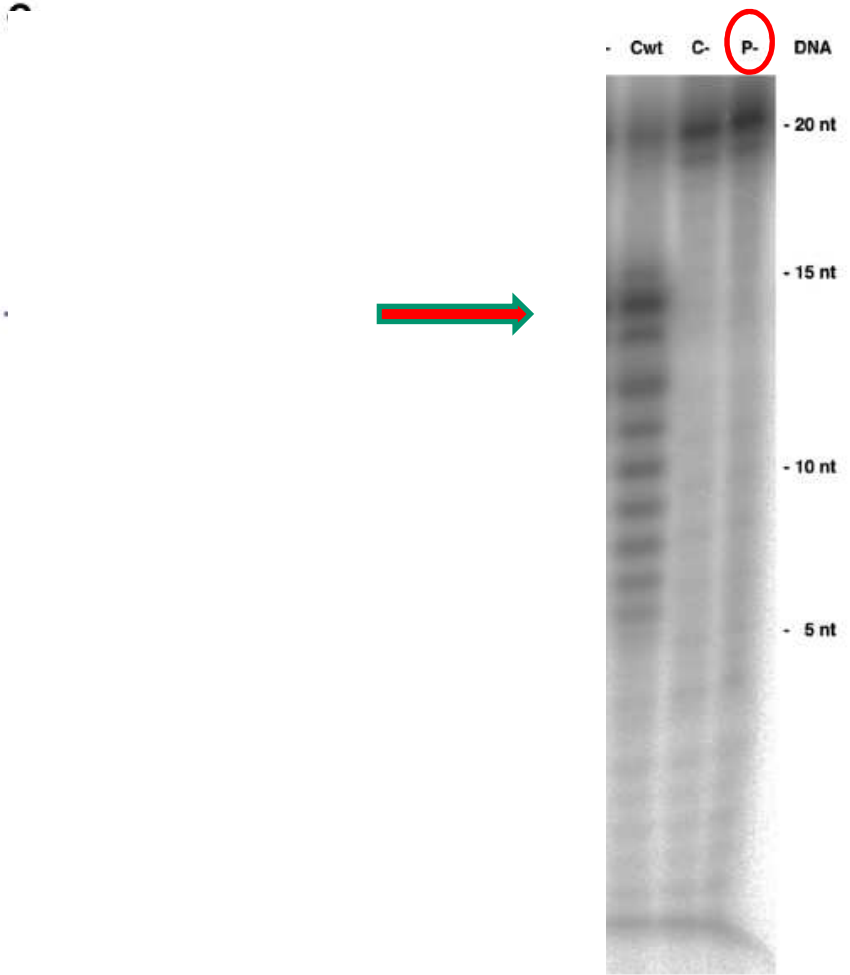


EN-endonuclease
Z domain
RT-reverse transcriptase
C-cysteine-rich.



Substrate of L1-EN

L1-EN cleavage



(P-): Cwt without L1-EN protein

Specificity of L1-EN

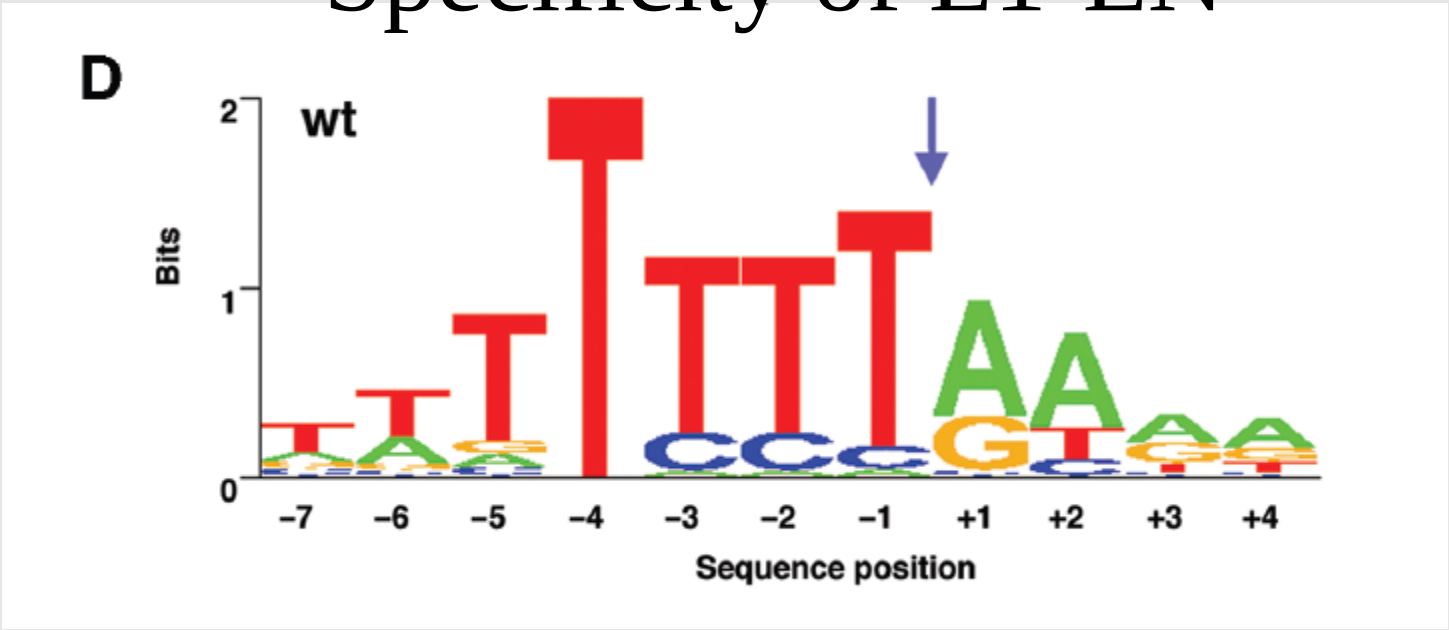
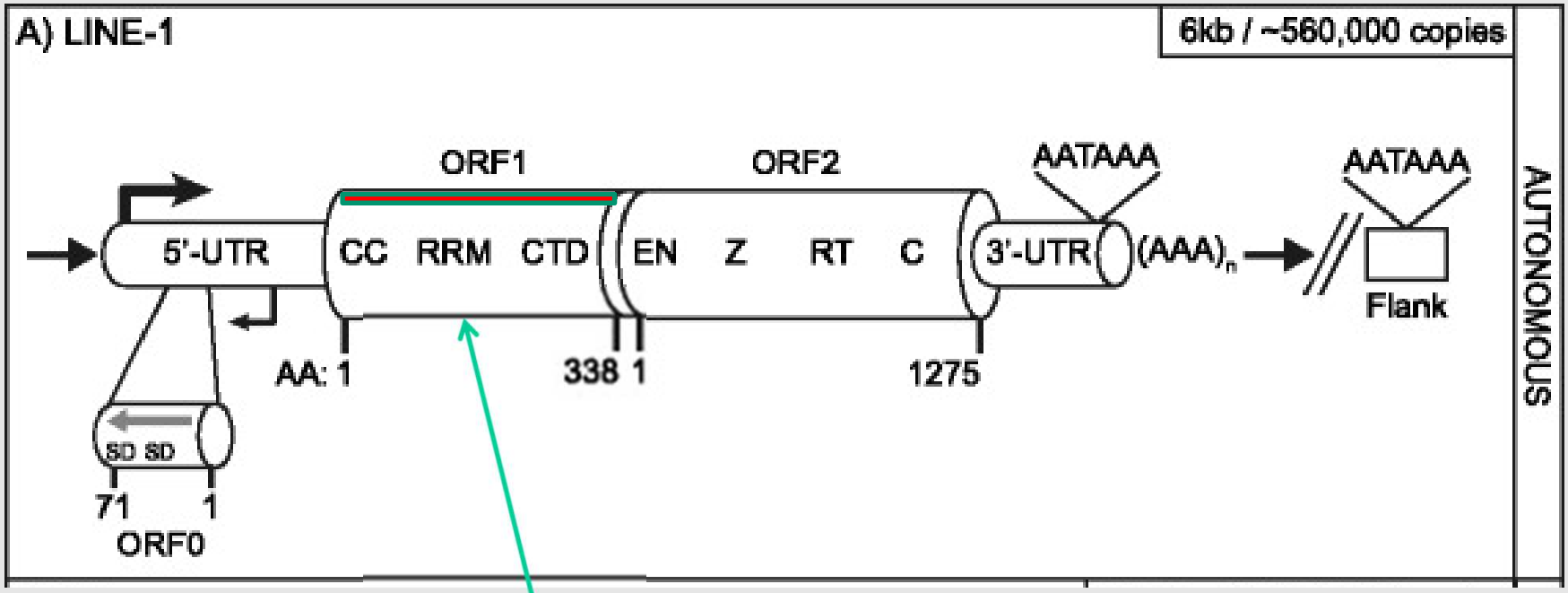


Table 1. Comparison of retrotransposition frequencies *in vivo* and plasmid nicking activities *in vitro*

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

EN mutants

Proporzionalità !!



CC-coiled coiled domain

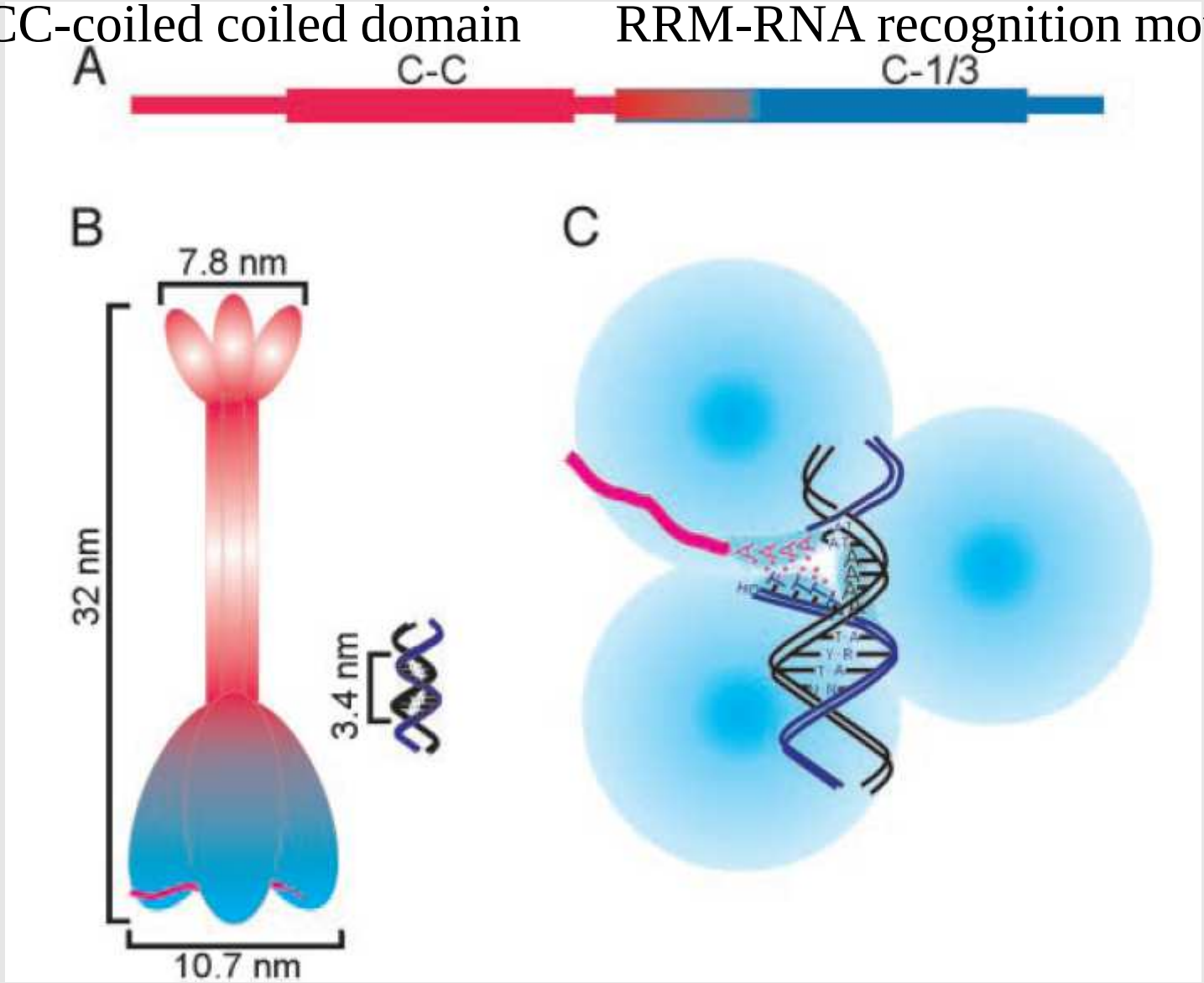
RRM-RNA recognition motif .

Trimeric structure for an essential protein in L1 retrotransposition

Sandra L. Martin*†,

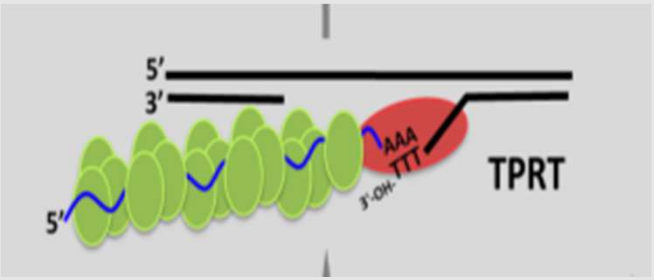
acidic(red) basic (blue)

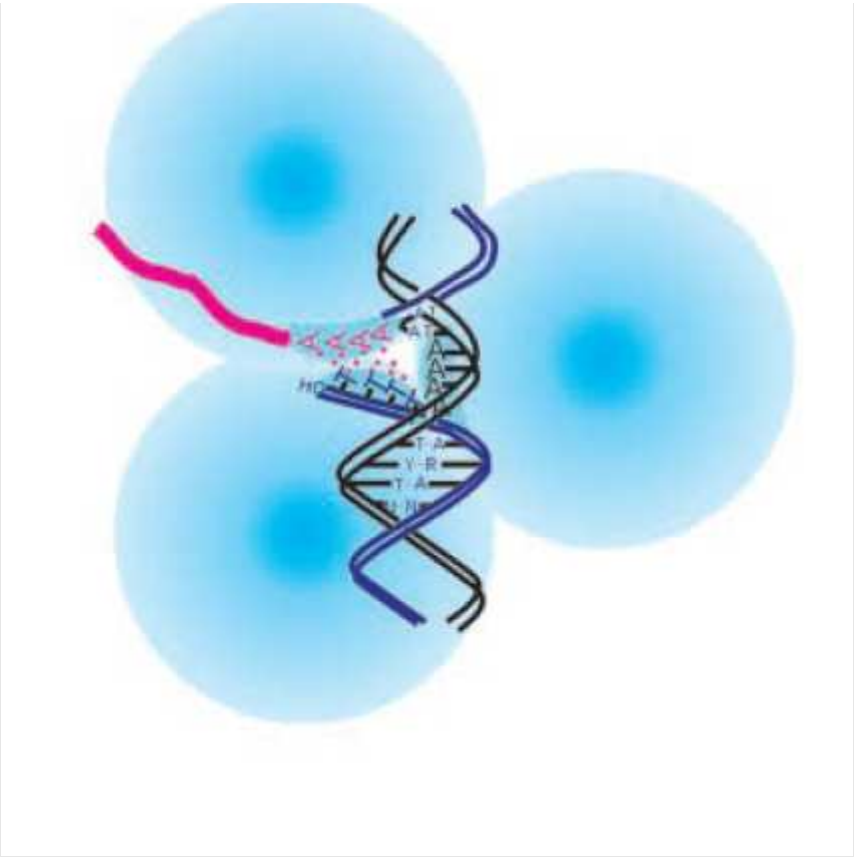
CC-coiled coiled domain RRM-RNA recognition motif



ORF1p binds single-stranded nucleic acids (L1 RNA and DNA) And functions as a nucleic acid chaperone.

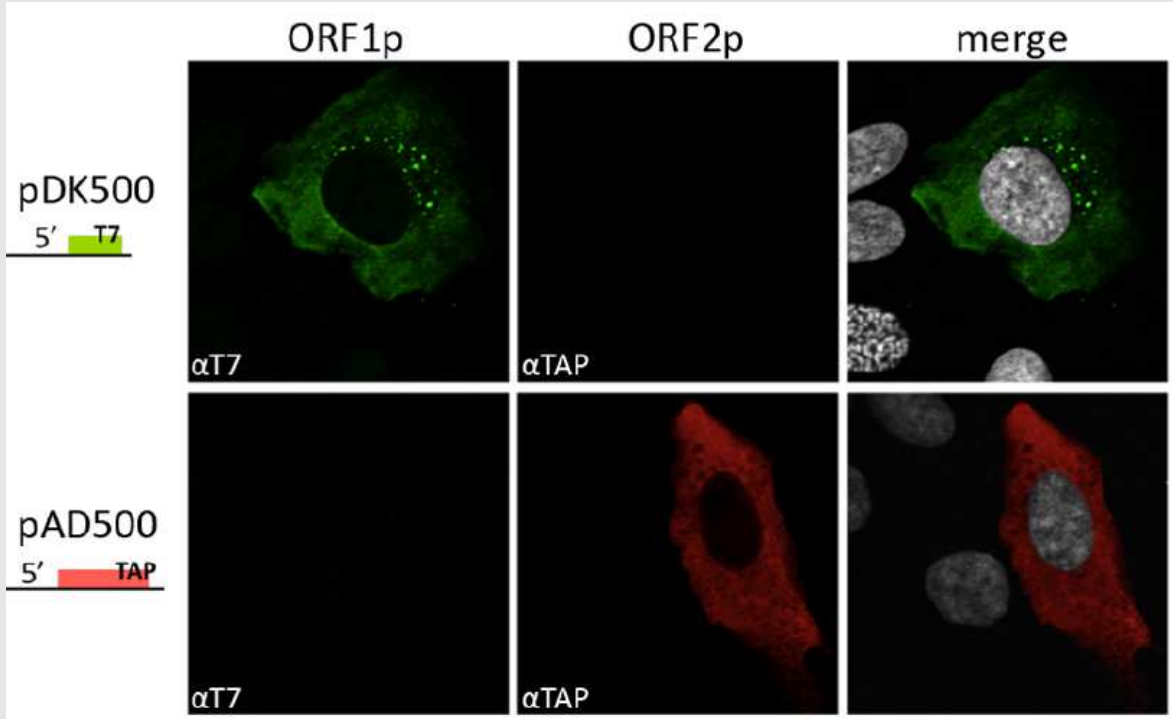
1 Each subunit of the trimer contains one single-stranded nucleic acid binding interface which is bound with one of the DNA target strands or the polyA tail of the L1 RNA (red). The double-stranded regions of the target are not bound

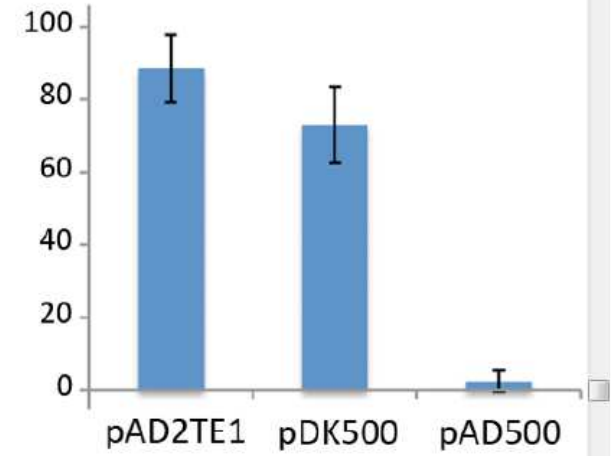




2 ORF1p coats the entire 7-kb L1 RNA to form a ribonucleoprotein particle
The nucleic acid chaperone activity of ORF1p melts the DNA and then facilitates formation of the RNA:DNA hybrid

ORF1p is necessary and sufficient for L1 cytoplasmic foci formation

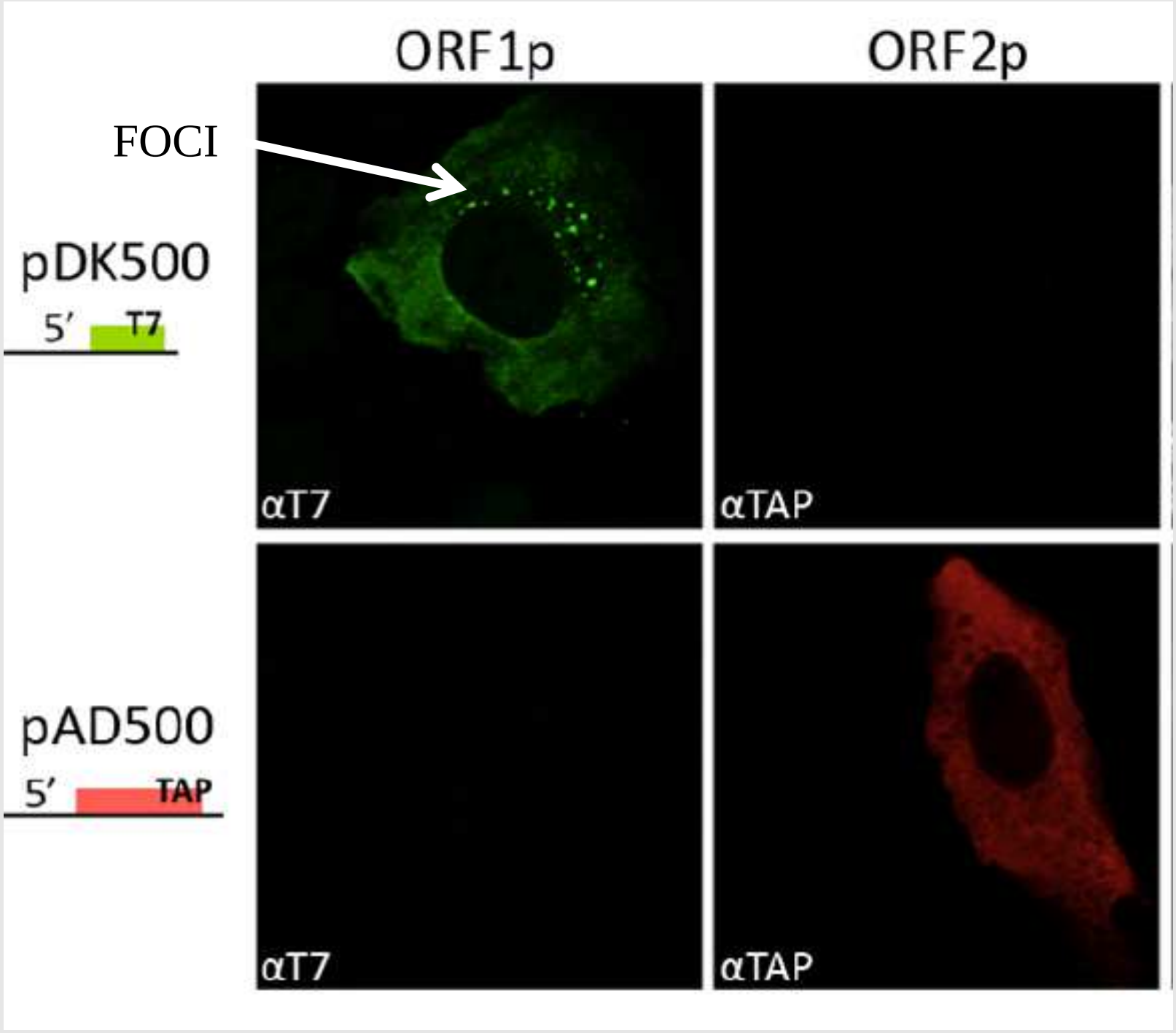




positive control ORF1p ORF2p

T7-tagged ORF1p green

TAP-tagged ORF2p red;


ORF1p
ORF2p
FOCI
pDK500
5′
T7
αT7
αTAP
pAD500
5′
TAP
αT7
αTAP

MUTAGENESI INSERZIONALE

LINE-1

To date in humans, 124 LINE-1-mediated insertions which result in genetic diseases have been reported.

Hancks and Kazazian Mobile DNA (2016) 7:9

Disease causing LINE-1 insertions have provided a wealth of insight and the foundation for valuable tools to study these genomic parasites.

POCHISSIMI ELEMENTI L1 SONO ATTIVI

it is estimated, on the basis of full-length L1 elements with preserved open reading frames and activity in in vitro retrotransposition assays, that there are 50 to 120 currently active L1 repeats in the human genome, of which a small number are highly active -“hot-L1s”

EVENTI PATOLOGICI RARI

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

Haemophilia A

Haemophilia A resulting from de novo insertion ofL1 sequences represents a novel mechanism for mutation in man.

Kazazian 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.

Haemophilia A

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.