

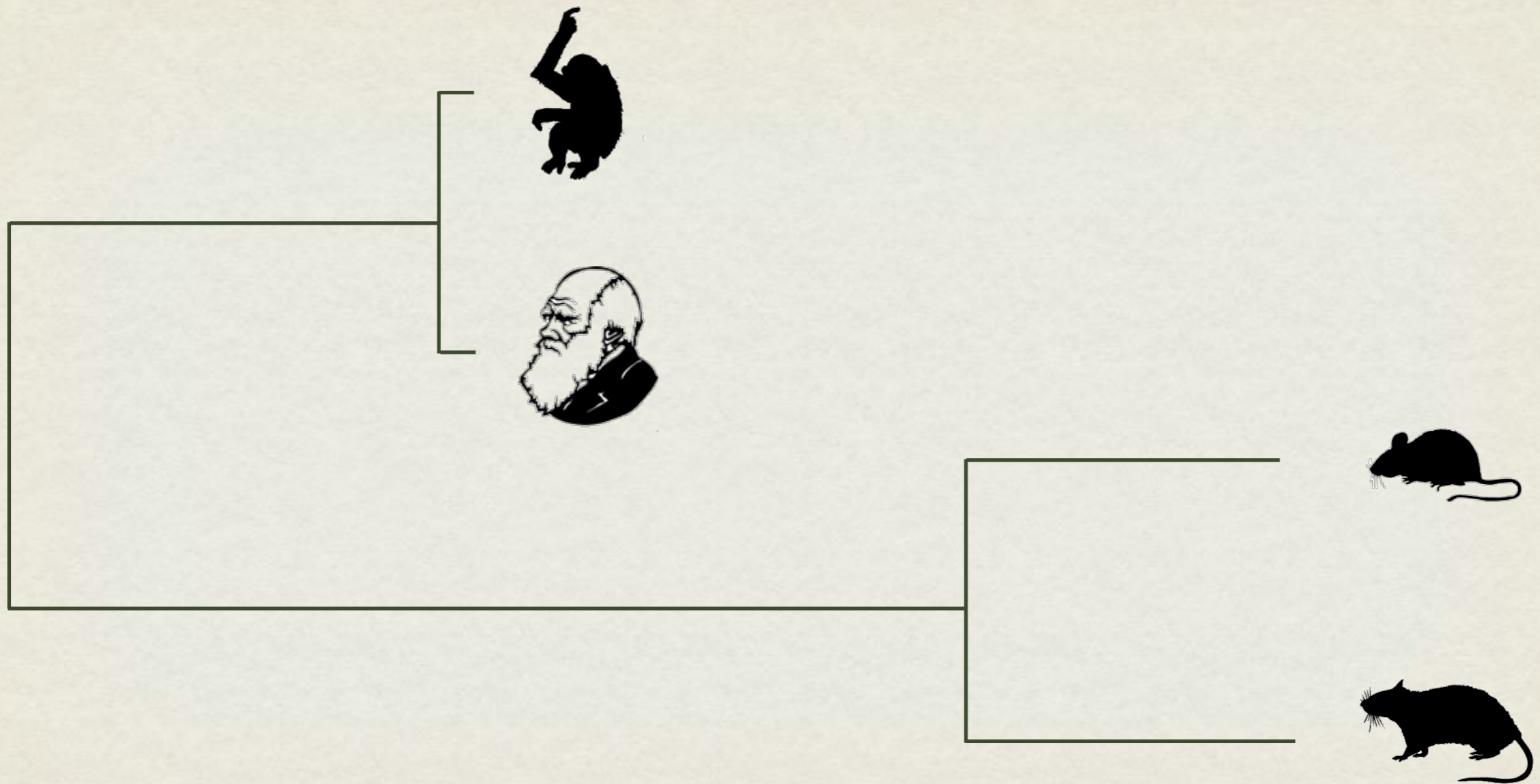


# TREASURE HIDDEN IN THE GENOME

Wojciech Makałowski

Institute of Bioinformatics, University of Münster

# EVOLUTIONARY TREE



# CHIMPANZEE VS HUMAN

Phenotypic level













# CHIMPANZEE VS HUMAN

Molecular level

# CHIMPANZEE VS HUMAN MOLECULAR LEVEL

>gi|454763|gb|U05311.1|

CCAGGAAACCTCTTGGGAGATAATGAAAATATTCTGCATCAGCCAGGCATGGTAGTTTATGGCTATAATC  
CCATCACTGTGGGAGGCAGGAAGATCACTTGAGGCCAGGAGTTCCAAAATATTCTGTATCTTGATTGTGG  
TGGTAGTTACACAACCTGTATAAATTTGTCAAATTCAAAGAACTATACACCTAAAGGGAAGACTTCTACT  
GTACGAAGTATACCTCATAAATAACCAACCAAAAAACCTCTGTTCTCATTTTAGTTTTCTACCCTTAT  
ATTAAAAAAGAGAGGCTTTGGCTGGGCCAGTGGCTCACACCTATAATCCAGCACTCTGGGAGGCCGAG  
GTGGGTGGATCACCCGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGGAACCTGTCTCTACTA  
AAAATACAAAATTAGCTGGGTGCAGTGGCACGCGTCTGTAATCCAAGCTACTTGAGAGGCTGAGGCAGG  
AGAATCGCTTGAACCTCAGGAGGCGGAAGGTGCAGTGGGCCAGATTGCACCCTGCACT

>gi|454760|gb|U05308.1|

GGAAACCTCTTGGGAGATAATGAAAATATTCTGCATCAGCCGGGCATGGTACTTTATGCCTATAATCCCA  
TCACTGTGGGAGGCAGGAAGATCACTTGAGGCCAGGAGTTCCAAAATATTCTGTATCTTGATTGTGGTGG  
TAGTTACACAACCTGTATAAATTTGTCAAATTCAAAGAACTATACACCTAAAGGGAAGACTTCTACTGTA  
CGAAGTATACCTCATAAATAACCAACCAAAAAACCTCTGTTCTCATTTTAGTTTTCTACCCTTATATT  
AAAAAAGAGAGGCTTTGGCTGGGCCAGTGGCTCACACCTATAATCCAGCACTCTGGGAGGGCGAGGTG  
GGTGGATCCACCCGAGGTCAGGAGTTCAAGGCCAGCCTGGCCAACATGGTGGAACCTGTCTCTACTAAA  
AATACAAAATTAGCTGGGTGCAGTGGCACGCGTCTGTAATCCAGCTACTTGAGAGGCTGAGGCAGGAG  
AATCGCTTGAACCTCAGGAGGCGGAAGTTGCAGTGGGCCAGATTGCACCCTGCACT

# CHIMPANZEE VS HUMAN MOLECULAR LEVEL

>gi|454763|gb|U05311.1|HSU05311 **Human** genomic sequence

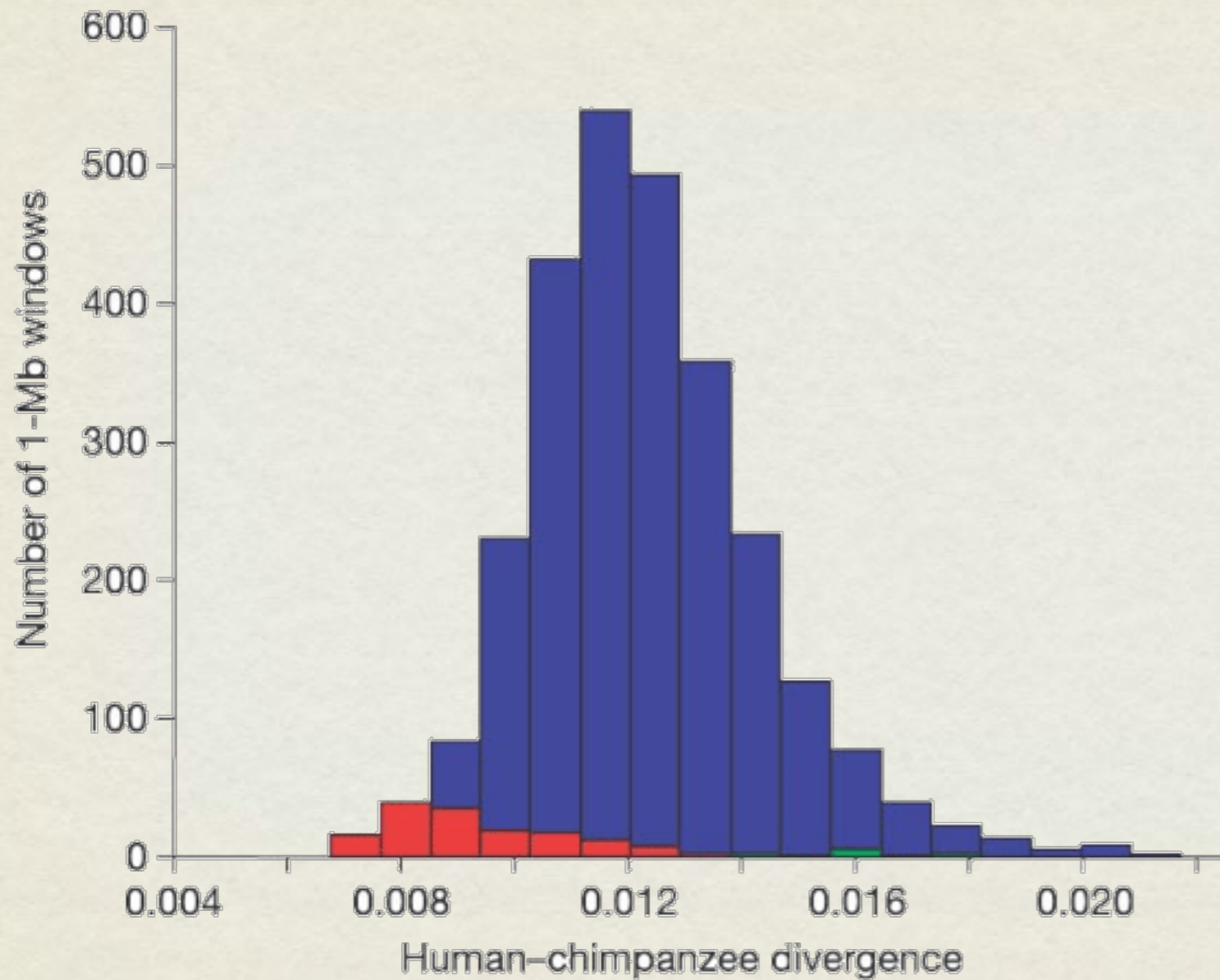
CCAGGAAACCTCTTGGGAGATAATGAAAATATTCTGCATCAGCCAGGCATGGTAGTTTATGGCTATAATC  
CCATCACTGTGGGAGGCAGGAAGATCACTTGAGGCCAGGAGTTCCAAATATTCTGTATCTTGATTGTGG  
TGGTAGTTACACAACCTGTATAAATTTGTCAAATTCAAAGAACTATACACCTAAAGGGAAGACTTCTACT  
GTACGAAGTATACCTCATAAATAACCAACCAAAAAACCTCTGTTCTCATTTTAGTTTTCTACCCTTAT  
ATTAAAAAAGAGAGGCTTTGGCTGGGCCAGTGGCTCACACCTATAATCCAGCACTCTGGGAGGCCGAG  
GTGGGTGGATCACCCGAGGTCAGGAGTTCAGGACCAGCCTGGCCAACATGGTGGAACCTGTCTCTACTA  
AAAATACAAAATTAGCTGGGTGCAGTGGCACGCGTCTGTAATCCAAGCTACTTGAGAGGCTGAGGCAGG  
AGAATCGCTTGAACCTCAGGAGGCGGAAGGTGCAGTGGGCCAGATTGCACCACTGCACT

>gi|454760|gb|U05308.1|PTU05308 **Pan troglodytes** genomic sequence

GGAAACCTCTTGGGAGATAATGAAAATATTCTGCATCAGCCGGGCATGGTACTTTATGCCTATAATCCCA  
TCACTGTGGGAGGCAGGAAGATCACTTGAGGCCAGGAGTTCCAAATATTCTGTATCTTGATTGTGGTGG  
TAGTTACACAACCTGTATAAATTTGTCAAATTCAAAGAACTATACACCTAAAGGGAAGACTTCTACTGTA  
CGAAGTATACCTCATAAATAACCAACCAAAAAACCTCTGTTCTCATTTTAGTTTTCTACCCTTATATT  
AAAAAAGAGAGGCTTTGGCTGGGCCAGTGGCTCACACCTATAATCCAGCACTCTGGGAGGGCGAGGTG  
GGTGGATCCACCCGAGGTCAGGAGTTCAGGACCAGCCTGGCCAACATGGTGGAACCTGTCTCTACTAAA  
AATACAAAATTAGCTGGGTGCAGTGGCACGCGTCTGTAATCCAGCTACTTGAGAGGCTGAGGCAGGAG  
AATCGCTTGAACCTCAGGAGGCGGAAGTTGCAGTGGGCCAGATTGCACCACTGCACT



# HUMAN-CHIMP DIVERGENCE AT THE DNA LEVEL

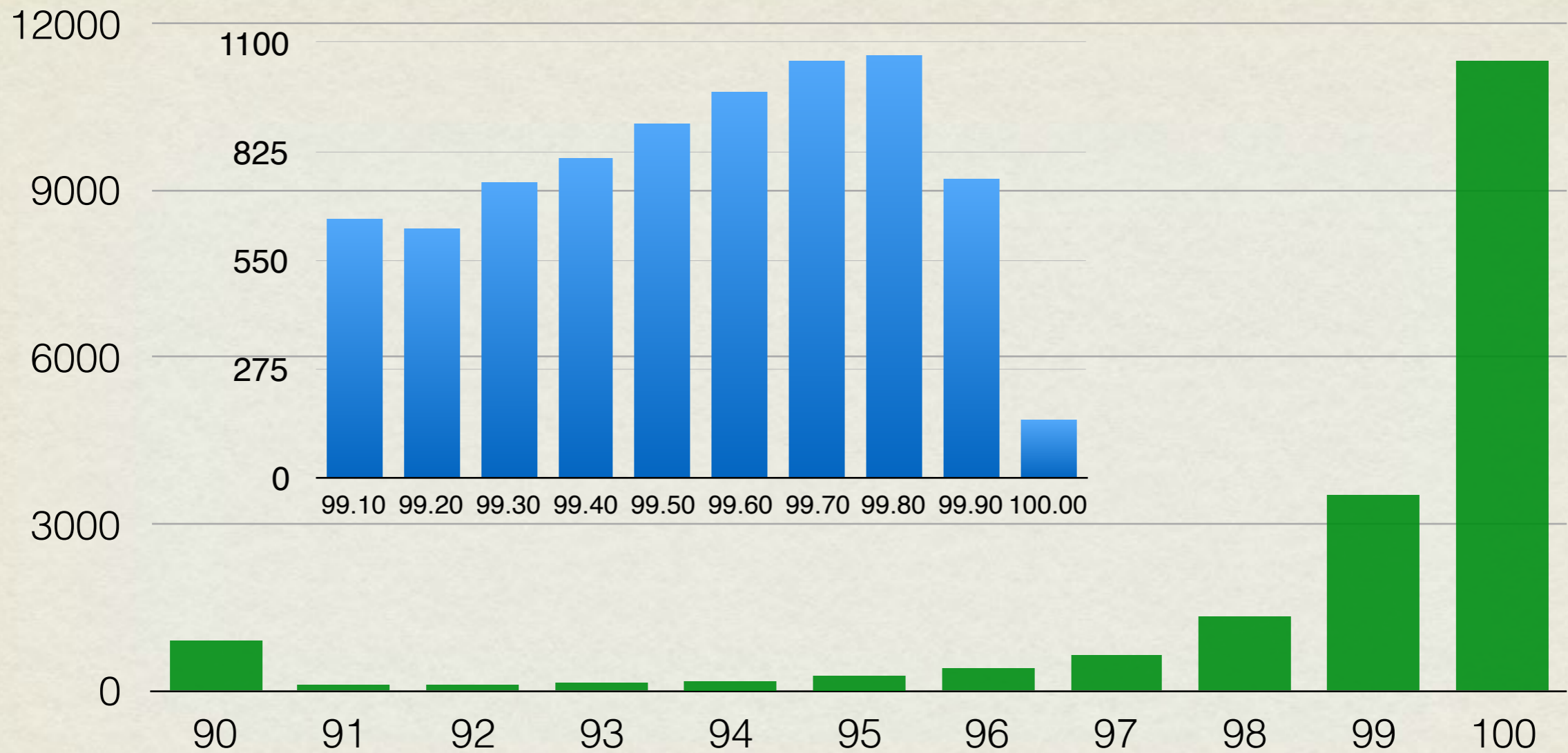


# HUMAN-CHIMP DIVERGENCE AT THE PROTEIN LEVEL



Makalowski, unpublished data

# HUMAN-CHIMP DIVERGENCE AT THE PROTEIN LEVEL



Makalowski, unpublished data

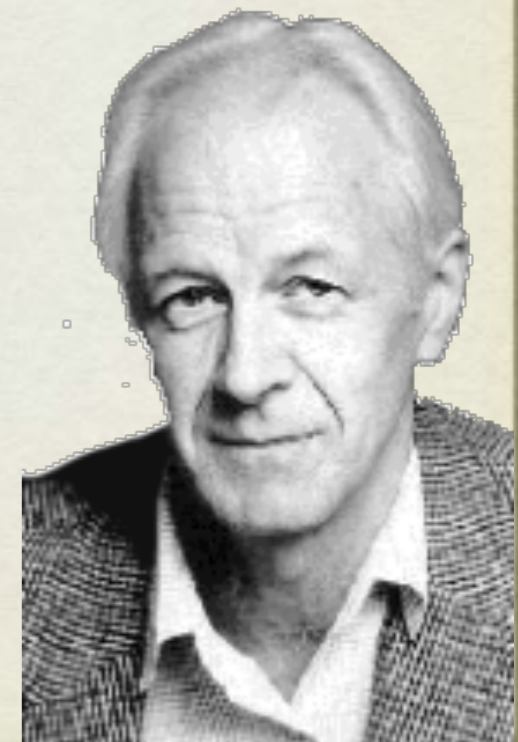
11 April 1975, Volume 188, Number 4184

# SCIENCE

## Evolution at Two Levels in Humans and Chimpanzees

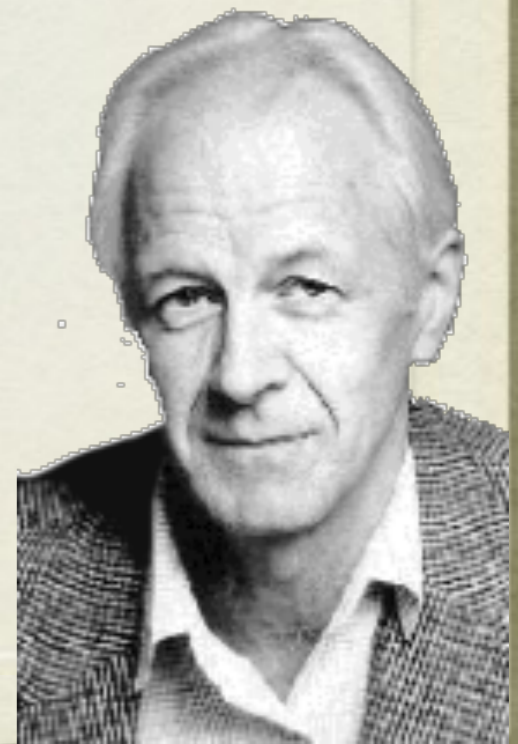
Their macromolecules are so alike that regulatory mutations may account for their biological differences.

Mary-Claire King and A. C. Wilson



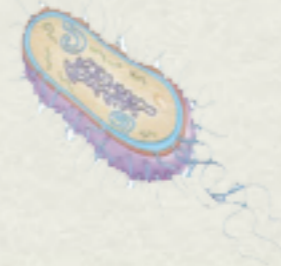


Would be point mutations  
in the regulatory regions  
sufficient?

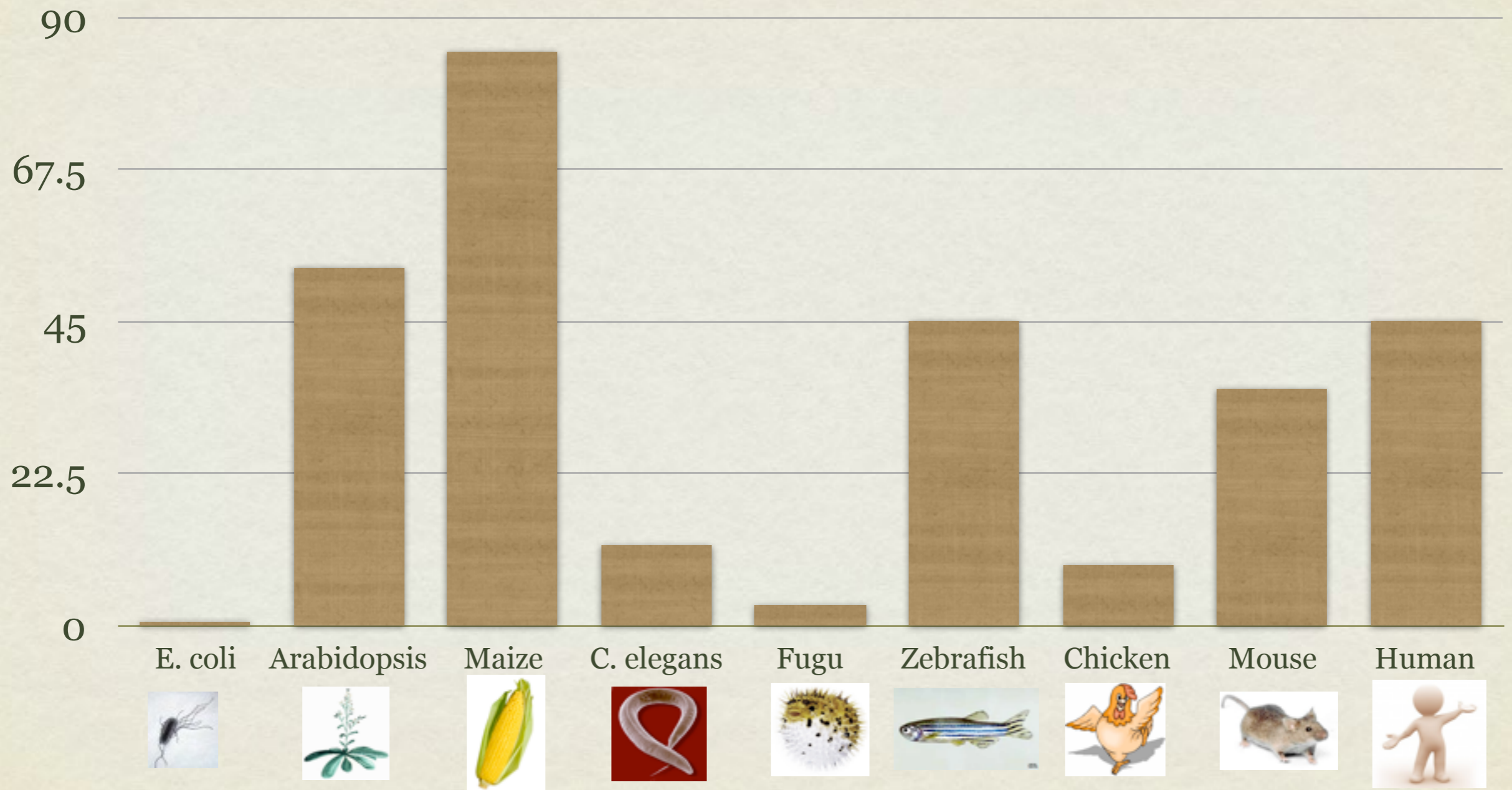




*Transposons*  
invaded all  
organisms



# TE-CONTENT IN DIFFERENT GENOMES

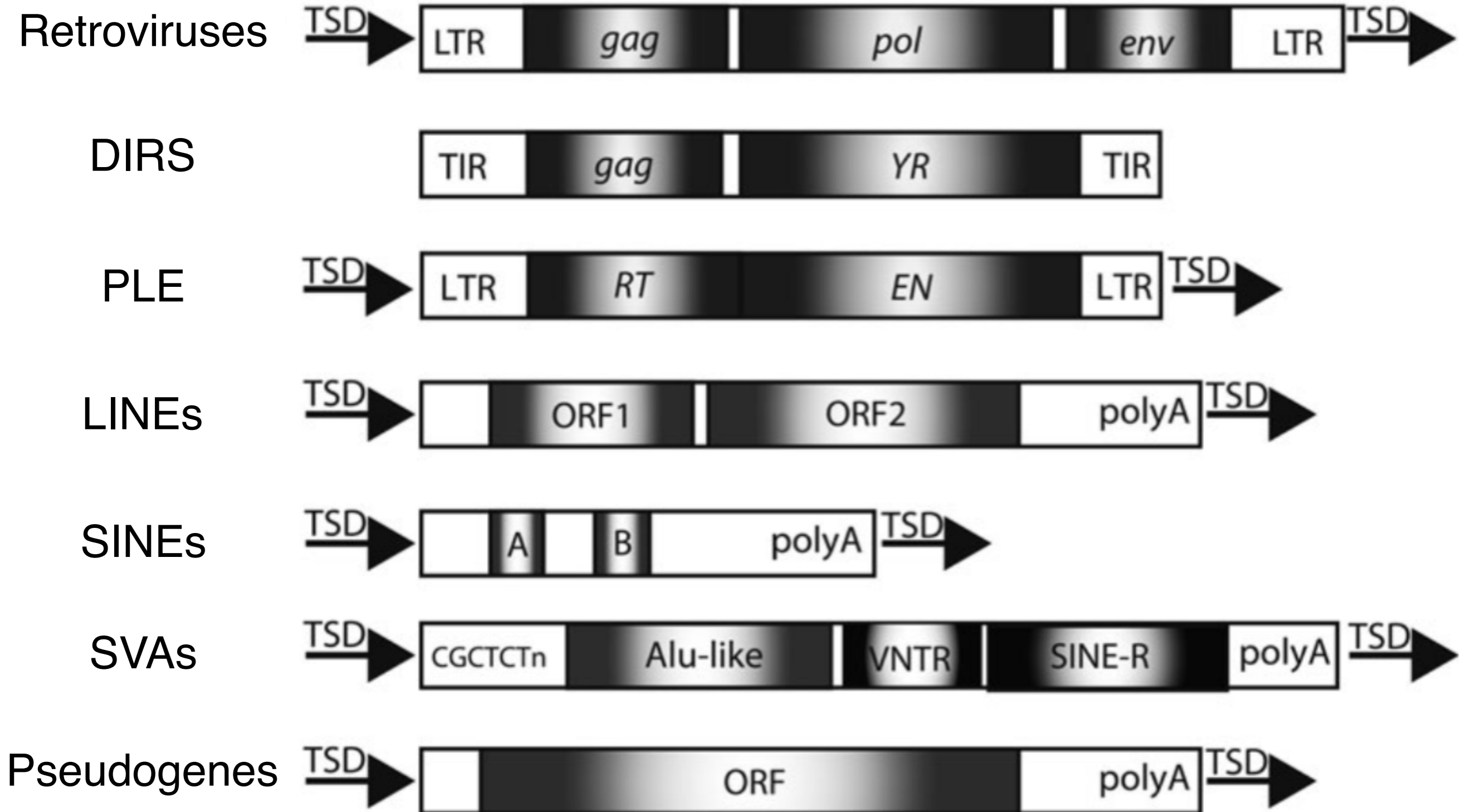


# TWO TYPES DEFINED BASED ON PROPAGATION MODE

- Class I: Retrotransposons
  - “copy-and-paste” transposition
- Class II: DNA transposons
  - “cut-and-paste” transposition
- Both classes are represented by autonomous and non-autonomous elements



# Structure of Class I transposons



# Structure of Class II transposons

“Classical” autonomous DNA transposons



“Classical” non-autonomous DNA transposons








Helitrons



Maverics

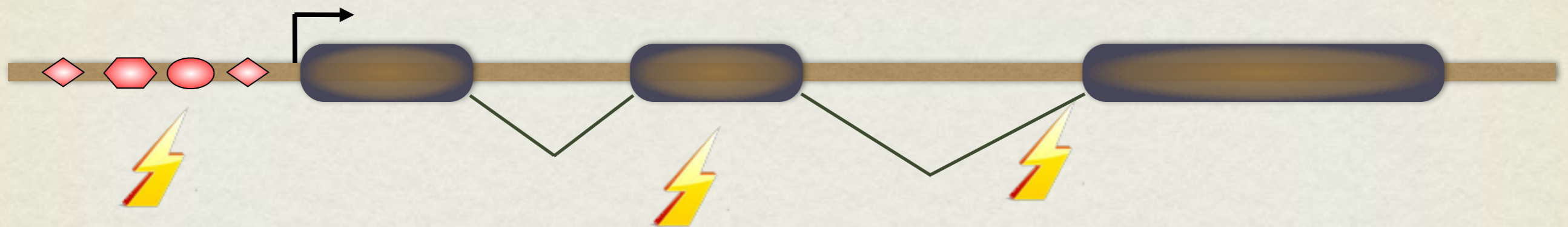


# TRANSPOSON LANDSCAPE OF EUKARYOTIC GENOMES

<i>Genome</i>	<i>Class I</i>	<i>Class II</i>	<i>Total</i>
	84,048	188,498	272,546
	39,131	14,388	53,519
	251,681	24,983	276,664
	3,396,497	159,459	3,555,956
	2,869,000	294,000	3,163,000

# TRANSPOSONS CAN CAUSE DISEASES

## Insertional mutagenesis



disruption of  
gene expression

obesity  
X-linked  
dystonia-  
parkinsonism

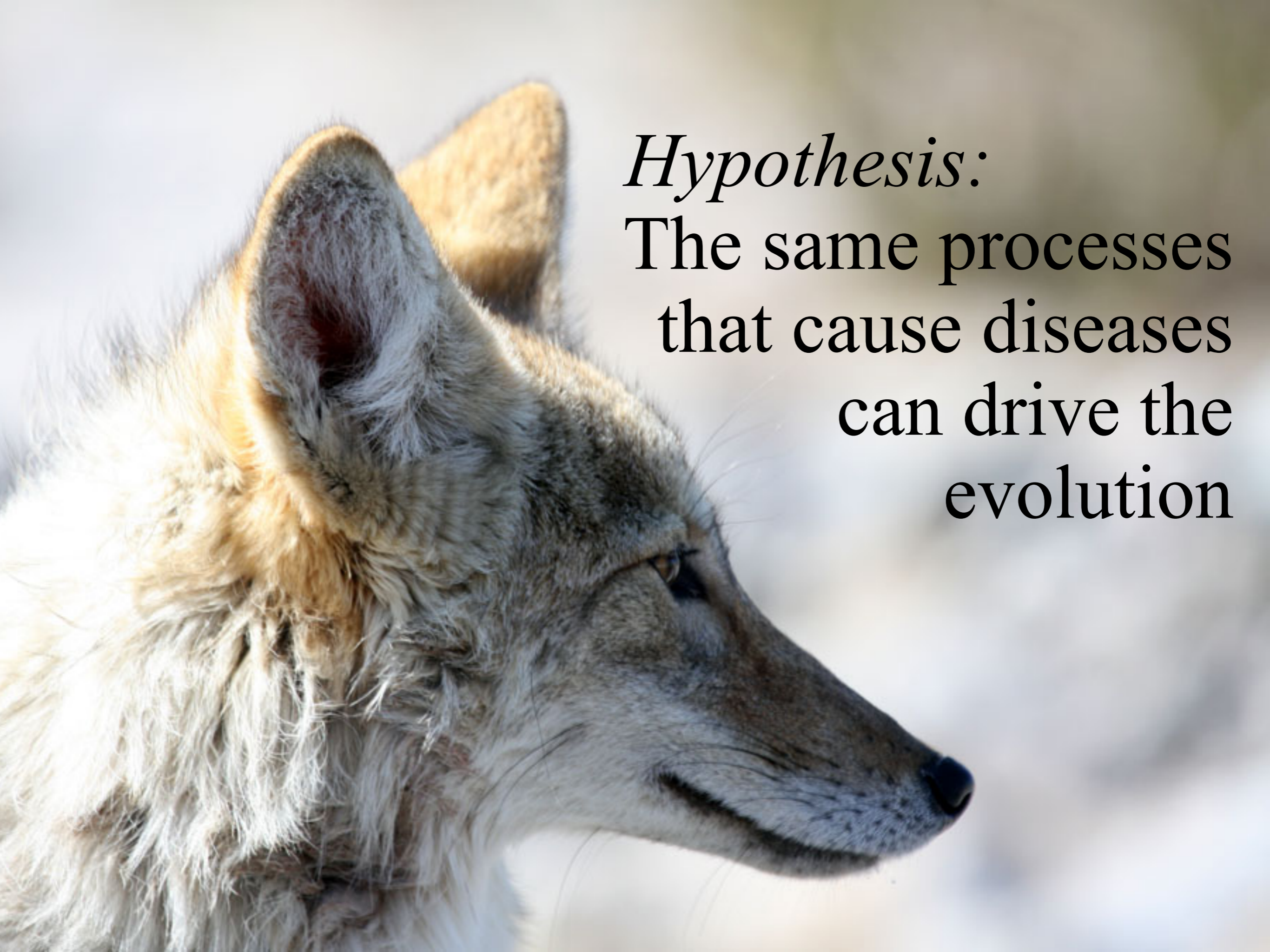
disruption of  
reading frame

gyrate atrophy  
DMD

disruption of  
splicing

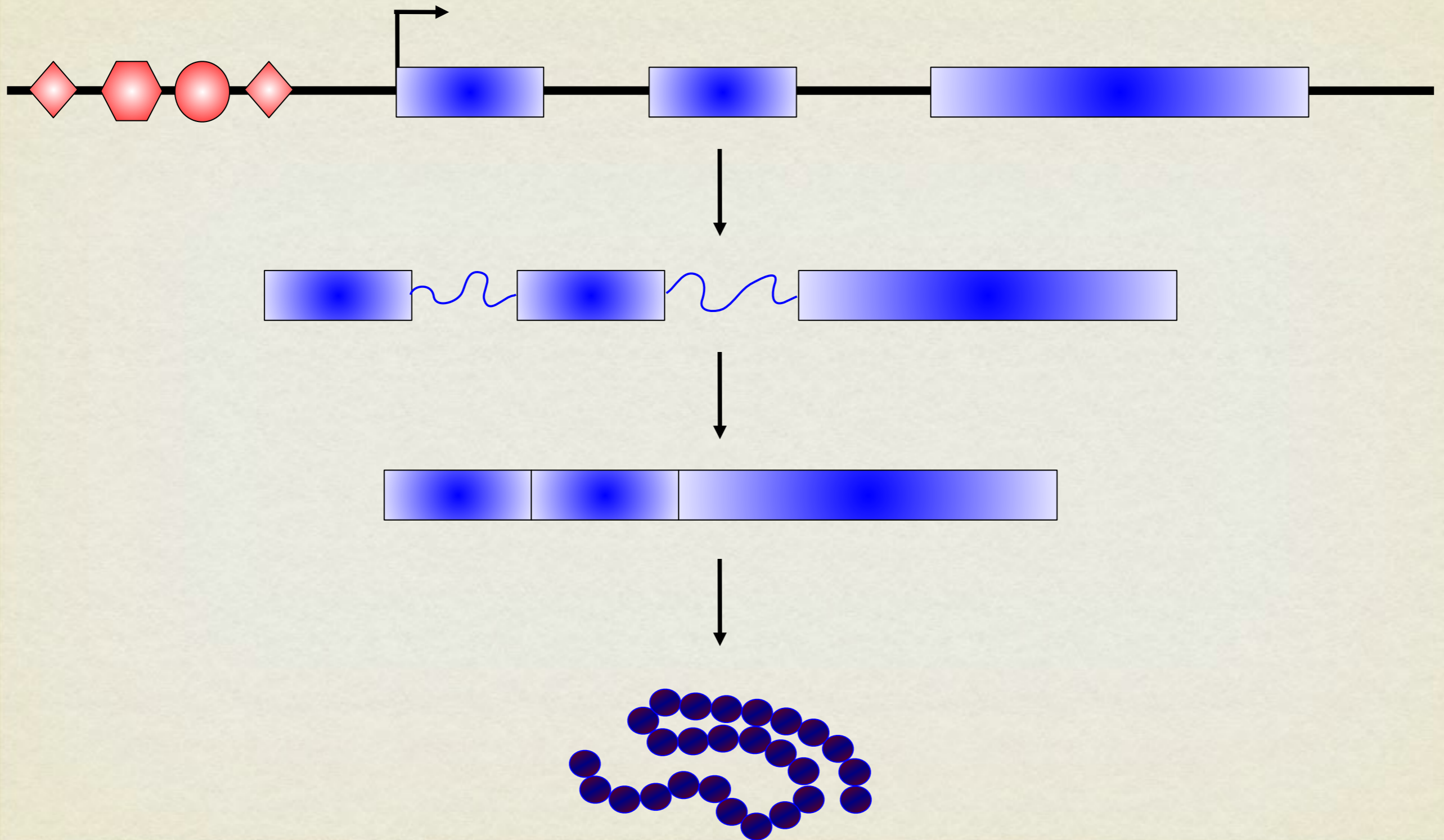
huntington disease

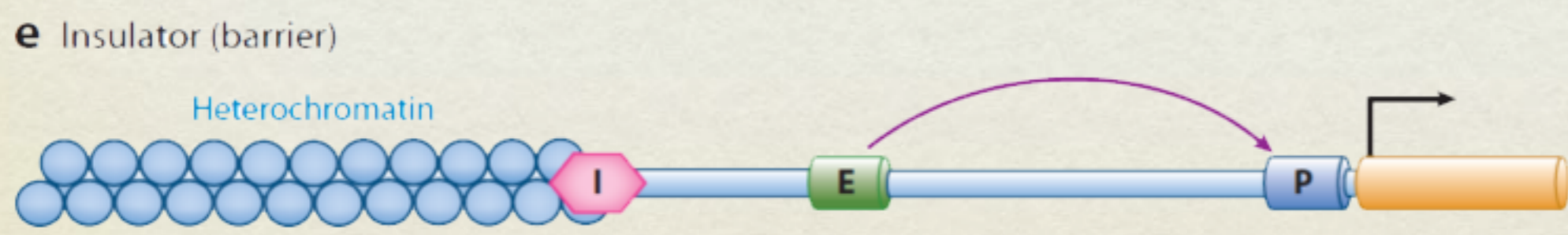
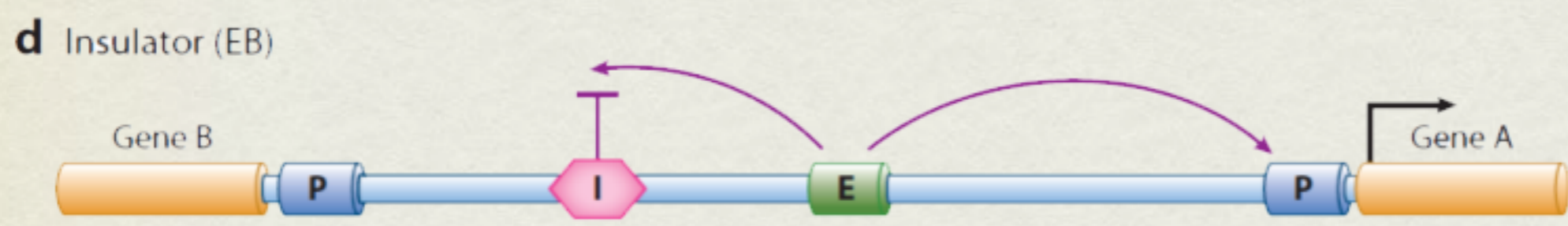
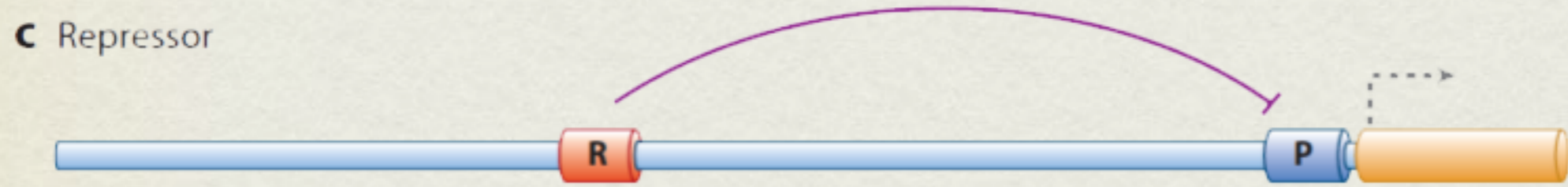
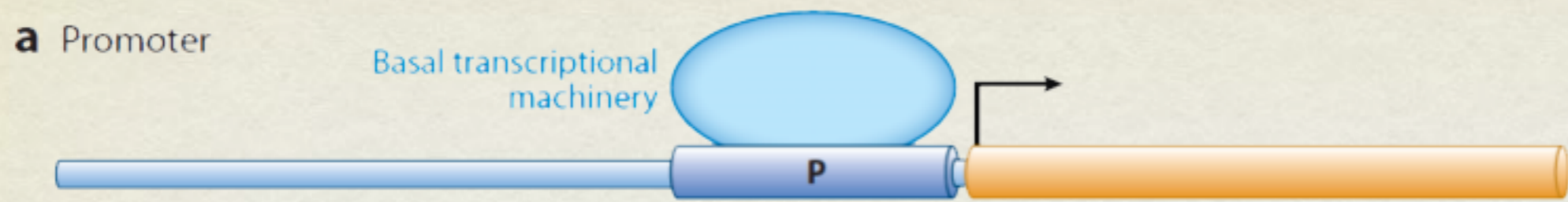




*Hypothesis:*  
The same processes  
that cause diseases  
can drive the  
evolution

# FROM GENE TO PRODUCT



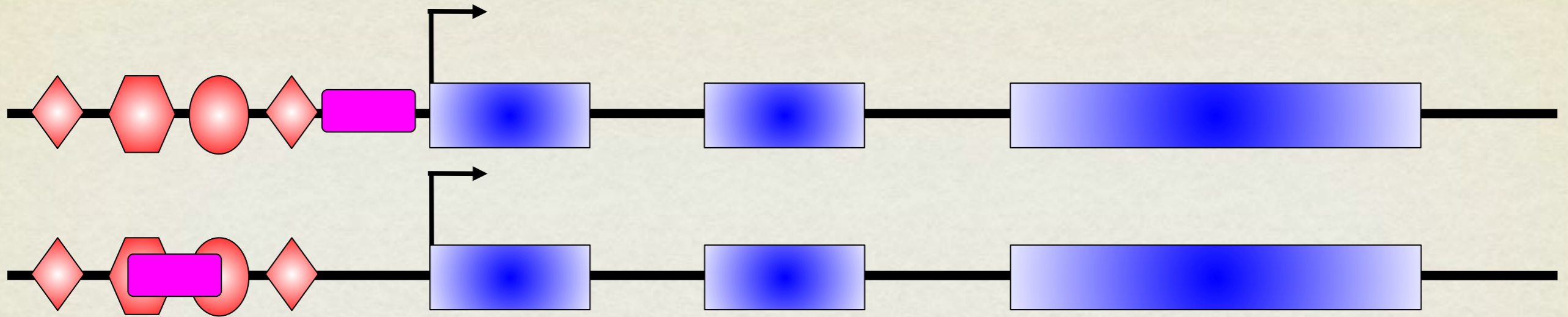


- Repressor
- Insulator
- Promoter
- Enhancer
- Transcribed sequence

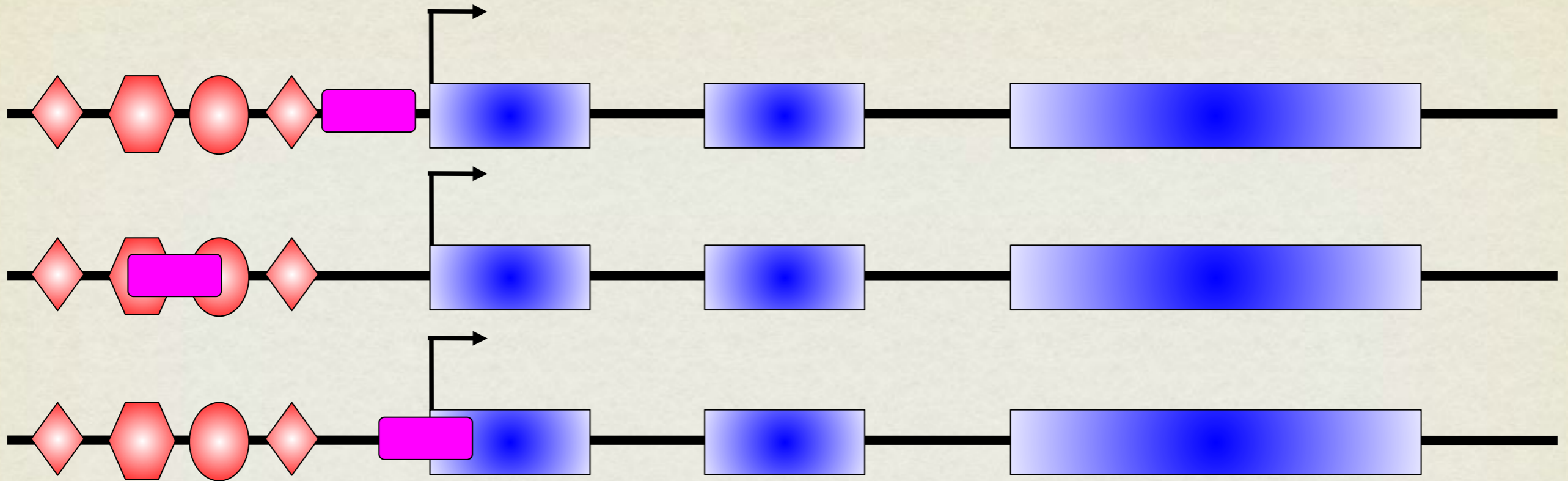
# A NEW *CIS* ELEMENT



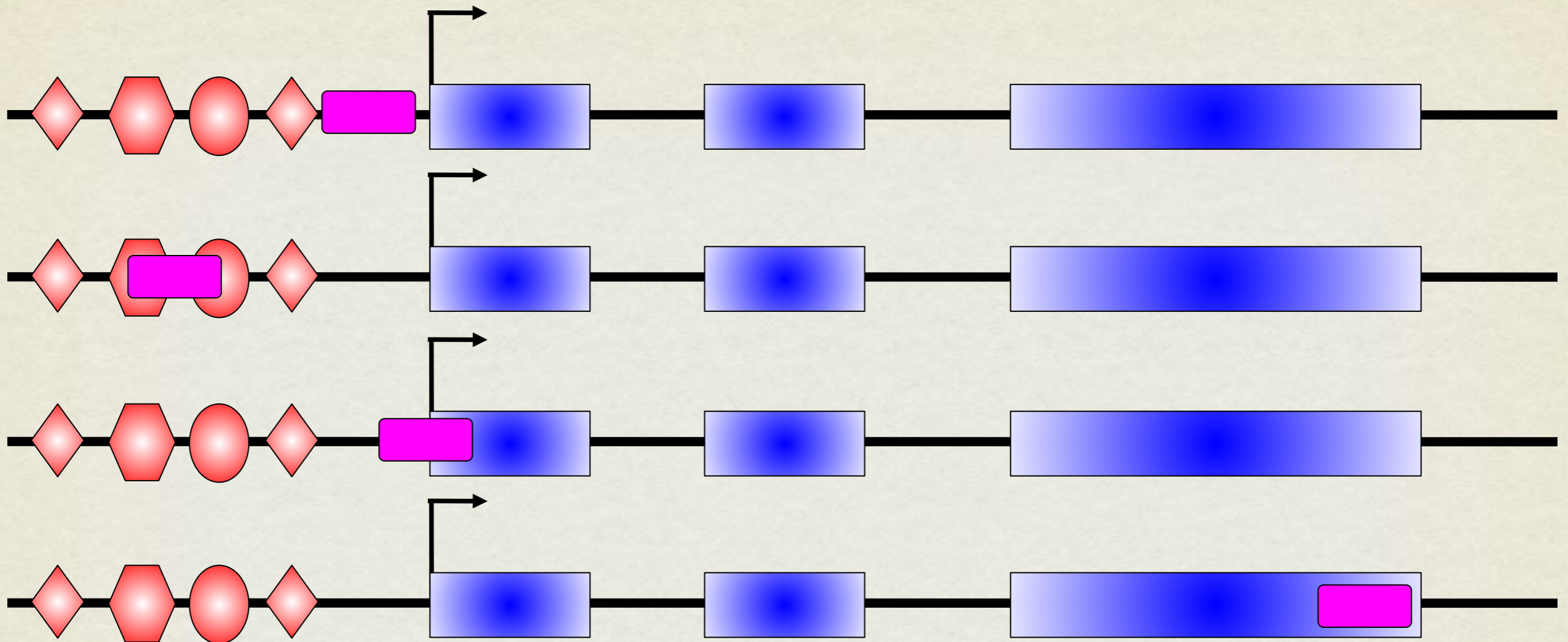
# INTERFERENCE WITH THE EXISTING *CIS* ELEMENTS



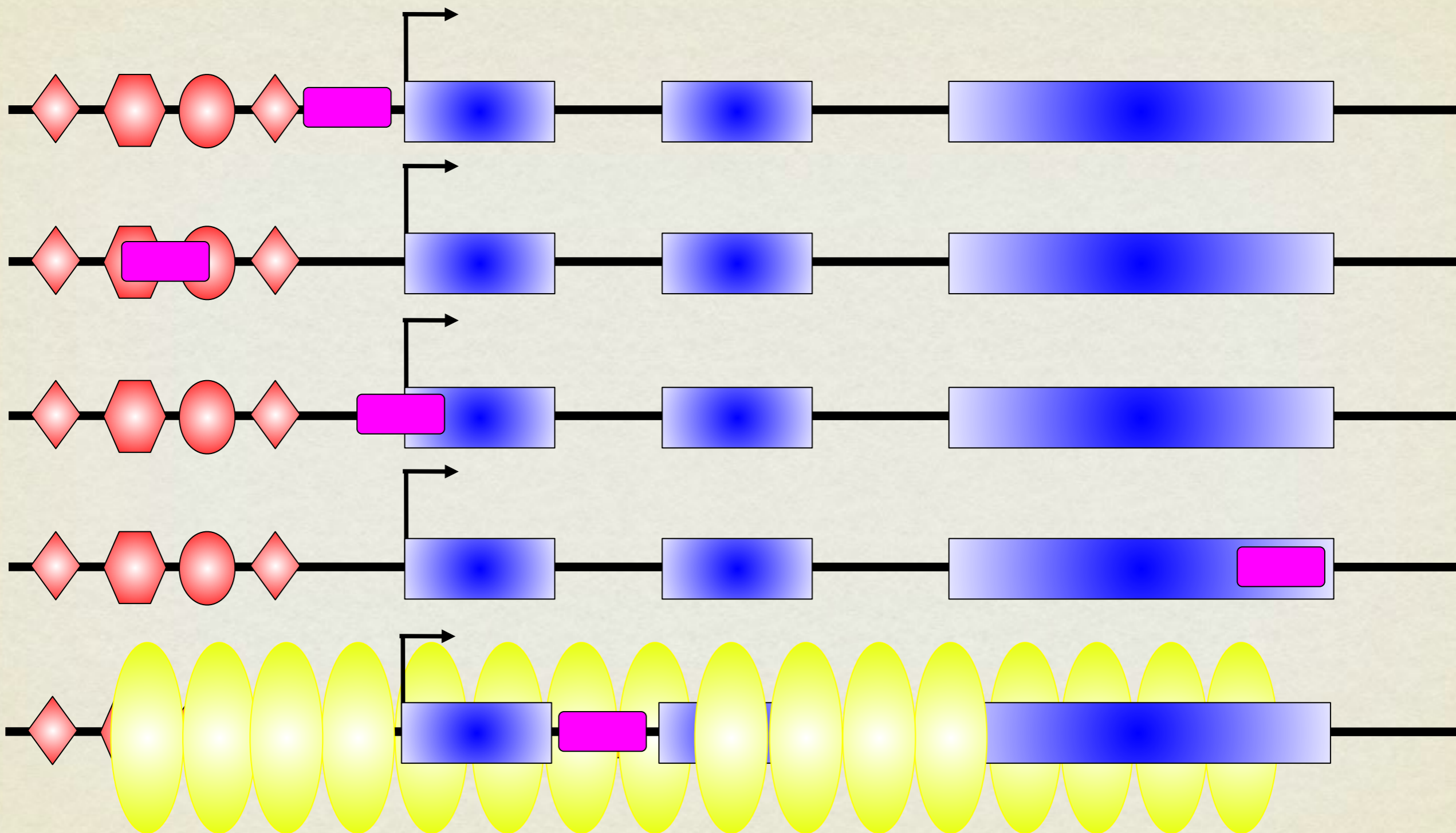
# INTERFERENCE WITH THE EXISTING OR ALTERNATIVE TRANSCRIPTION START



# INTERFERENCE WITH THE SIGNALS IN 3' UTR

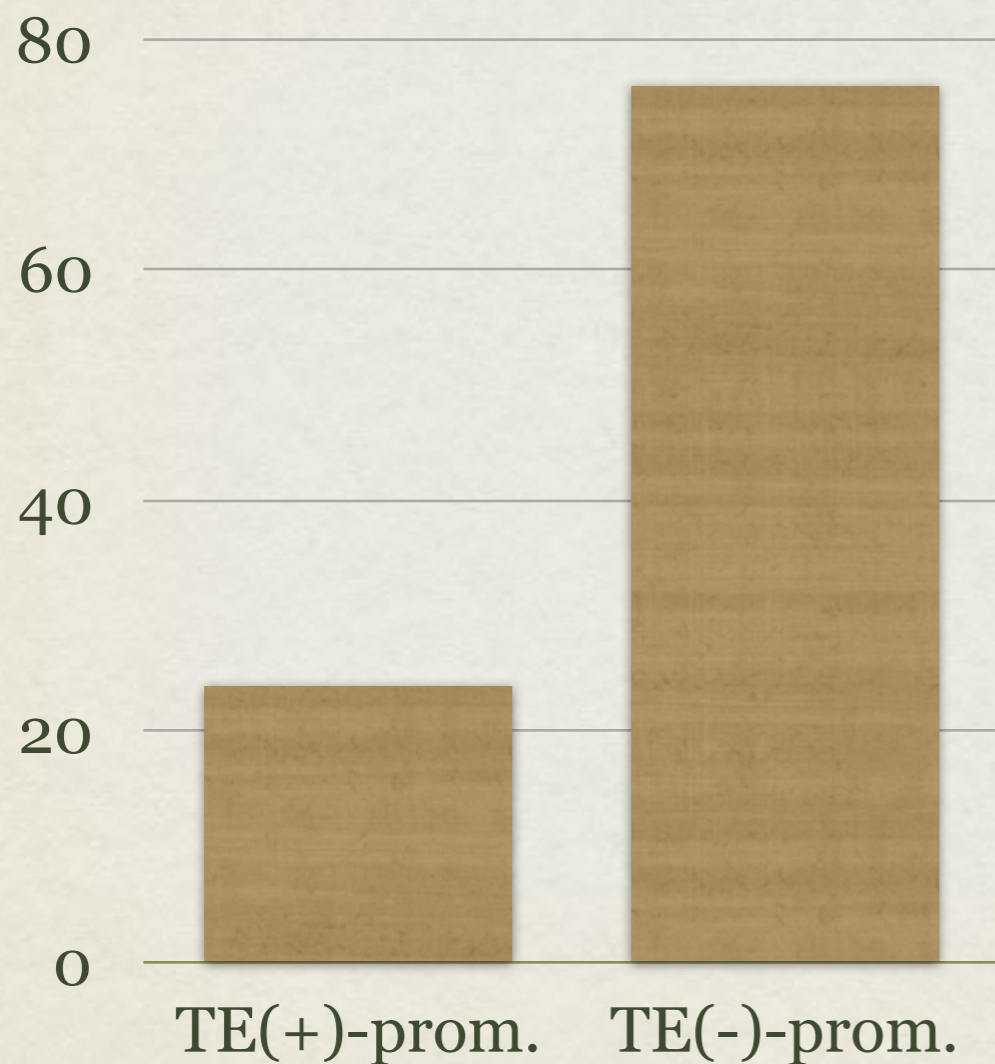
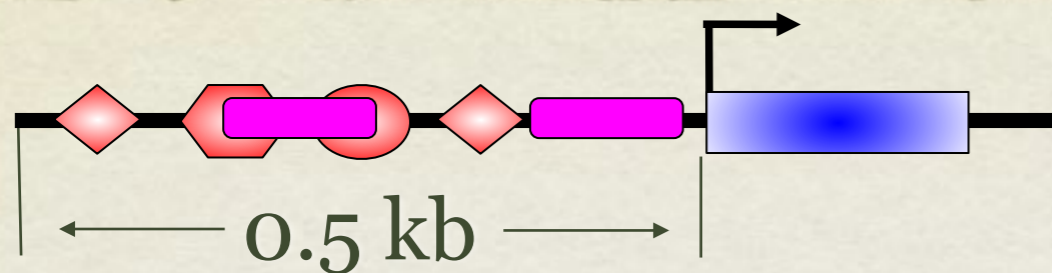


# GENE SILENCING

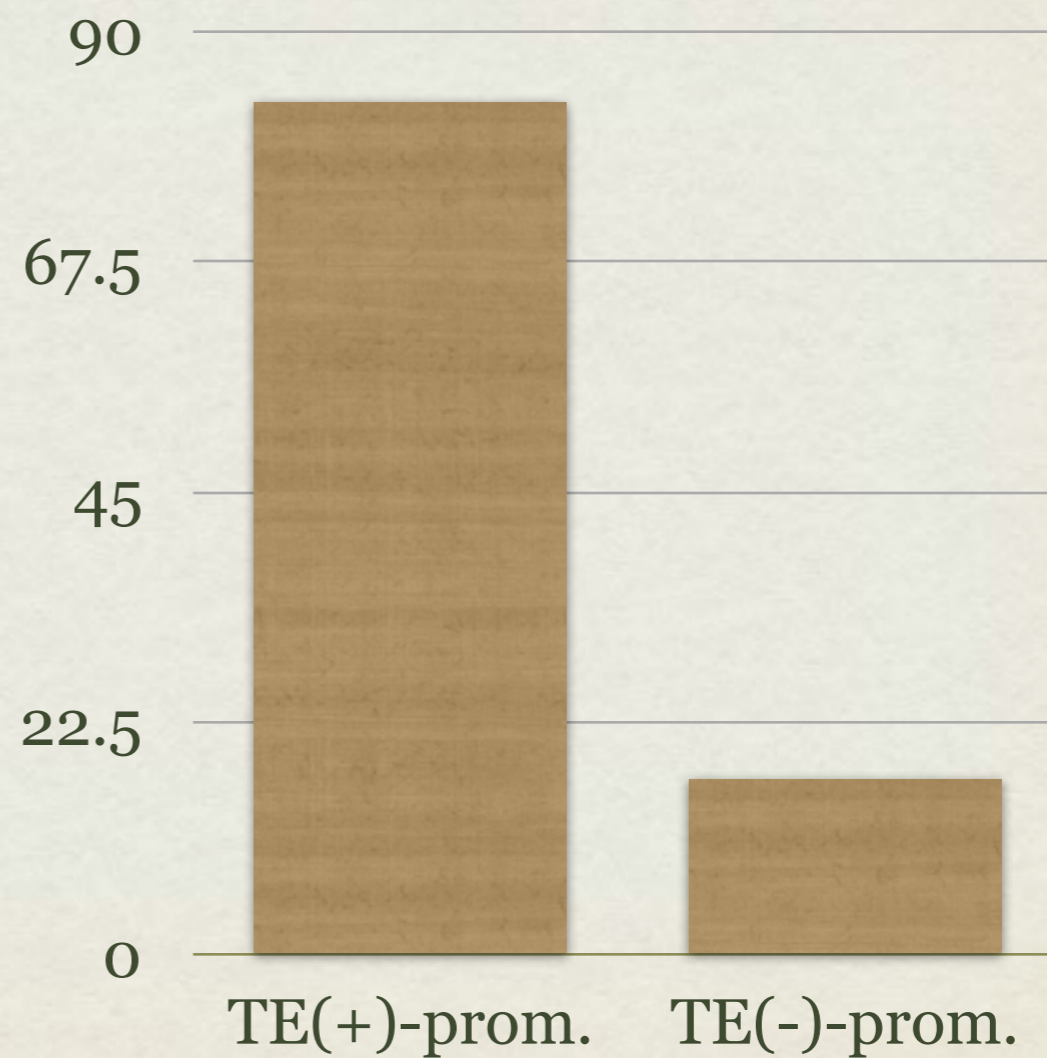
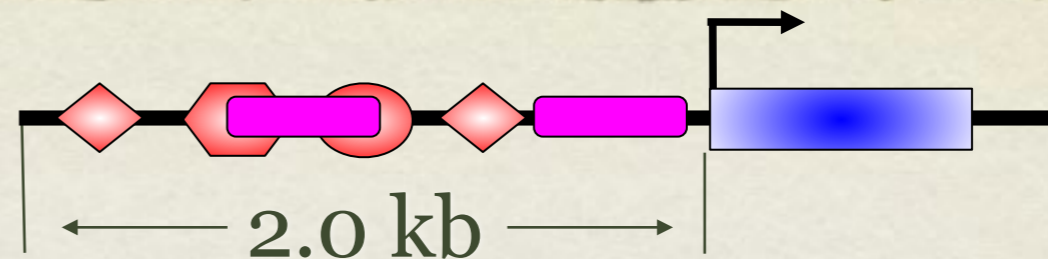




# PROMOTER AREA COVERAGE BY TE TYPE

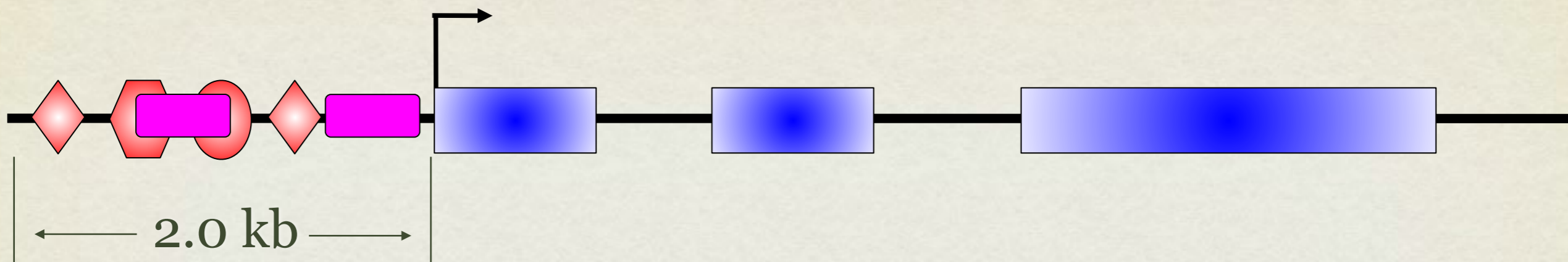


Jordan *et al.* 2003 *Trends Genet.* **19**: 68–72.



Thornburg *et al.* 2006 *Gene*, **365**: 104-110.

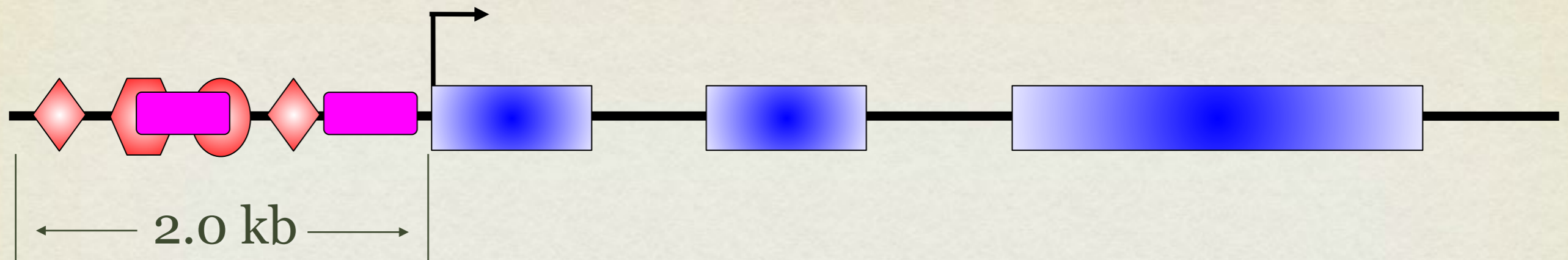
# PROMOTER AREA COVERAGE BY TE TYPE



■ LTR      ■ LINE      ■ SINE      ■ DNA



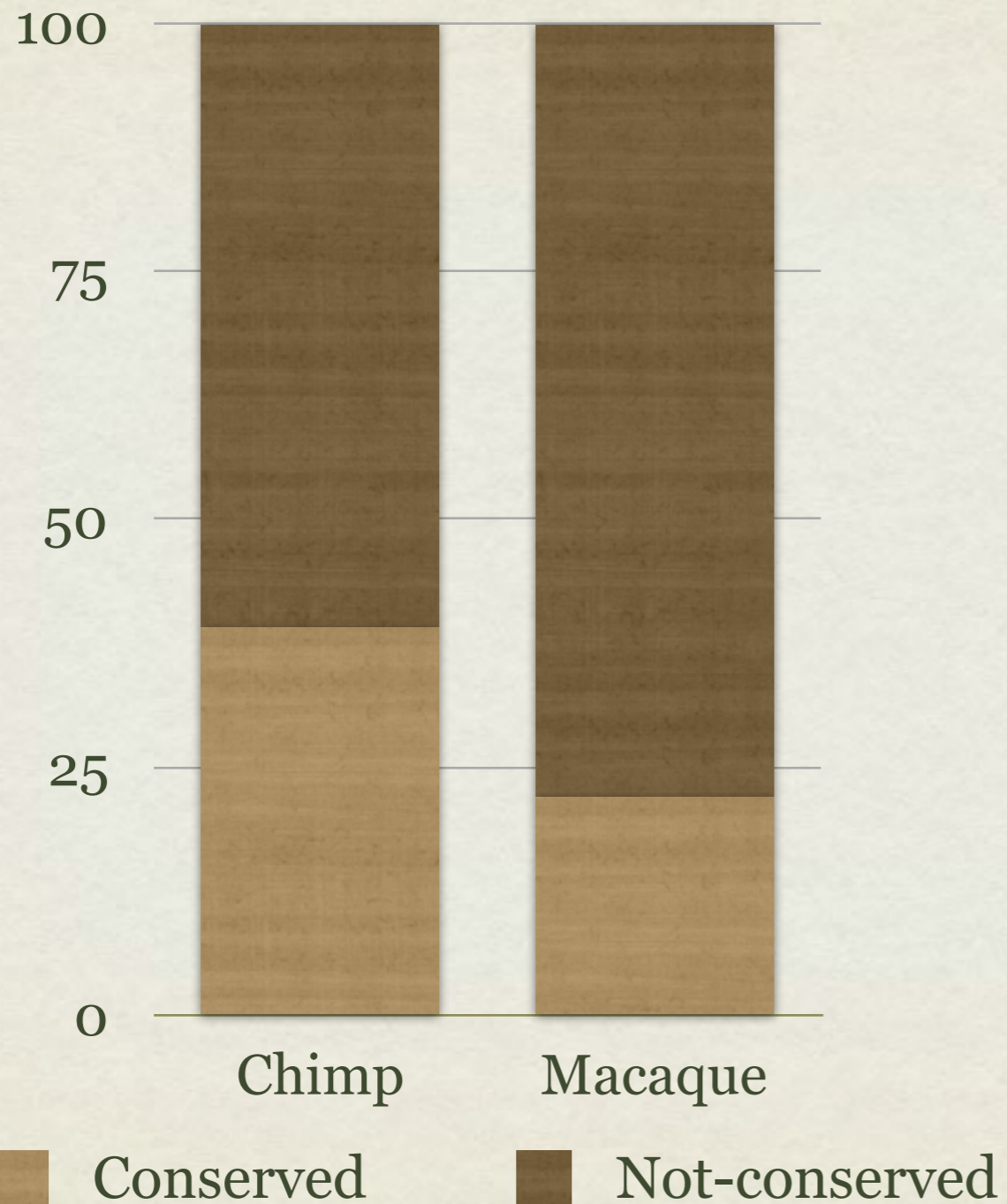
# CONCLUSIONS FROM THORNBURG'S PAPER



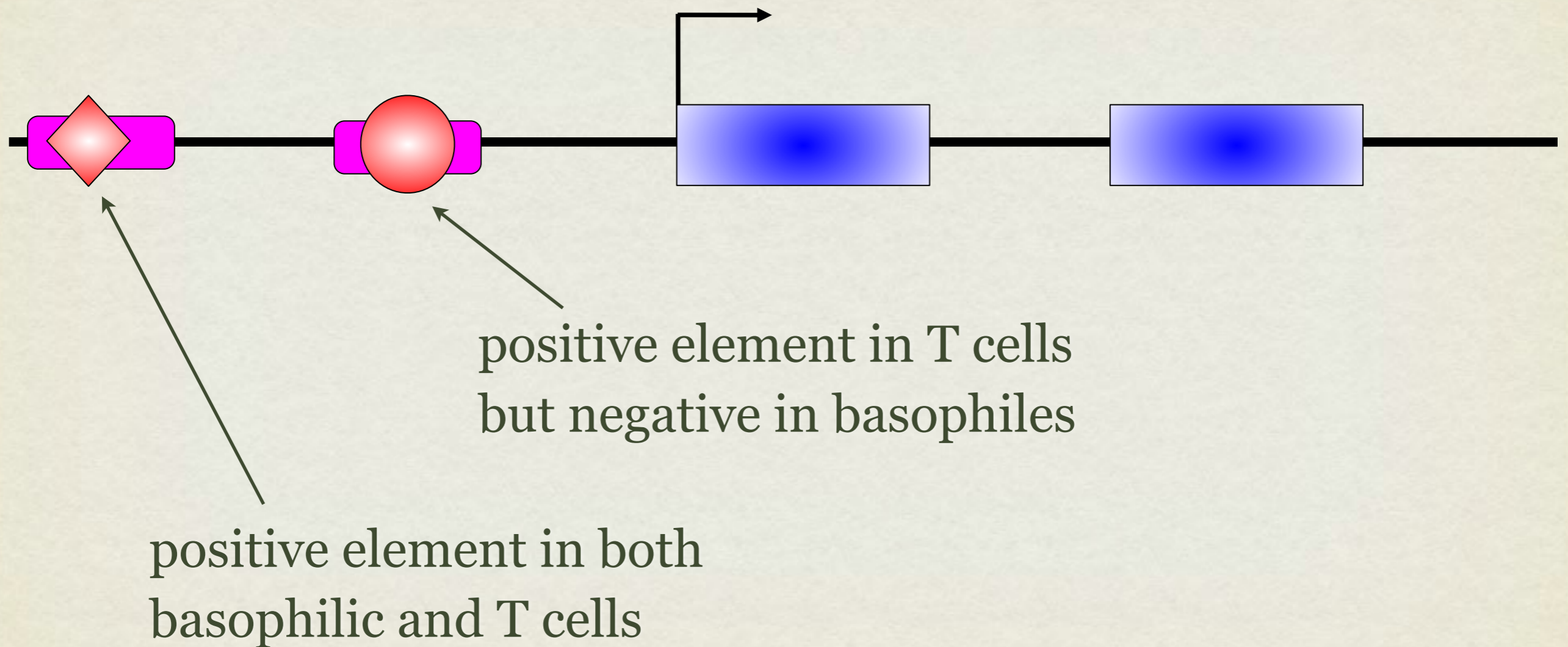
- SINE elements are the most abundant in promoter regions
- SINE elements contain the least binding sites, which might explain their higher retention rates in gene rich regions



# DYNAMIC OF ALU-CONTAINING PROMOTERS

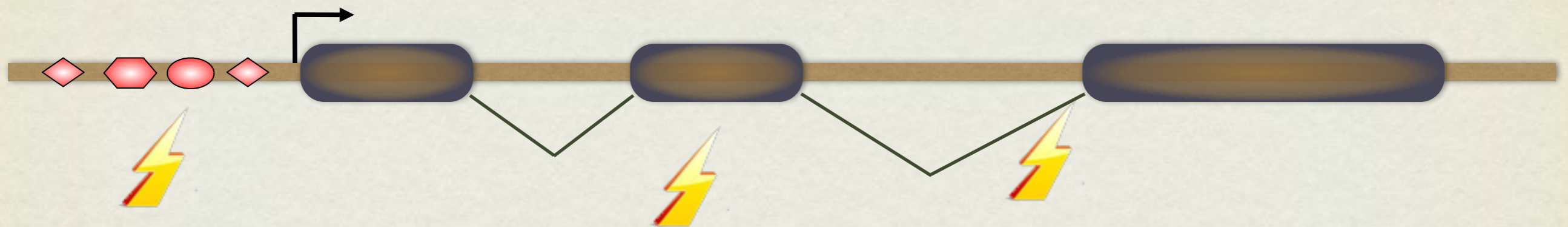


# CELL-SPECIFIC REGULATION OF THE IGE RECEPTOR GENE BY ALU ELEMENTS



# TRANSPOSONS CAN CAUSE DISEASES

## Insertional mutagenesis



disruption of  
gene expression

obesity  
X-linked  
dystonia-  
parkinsonism

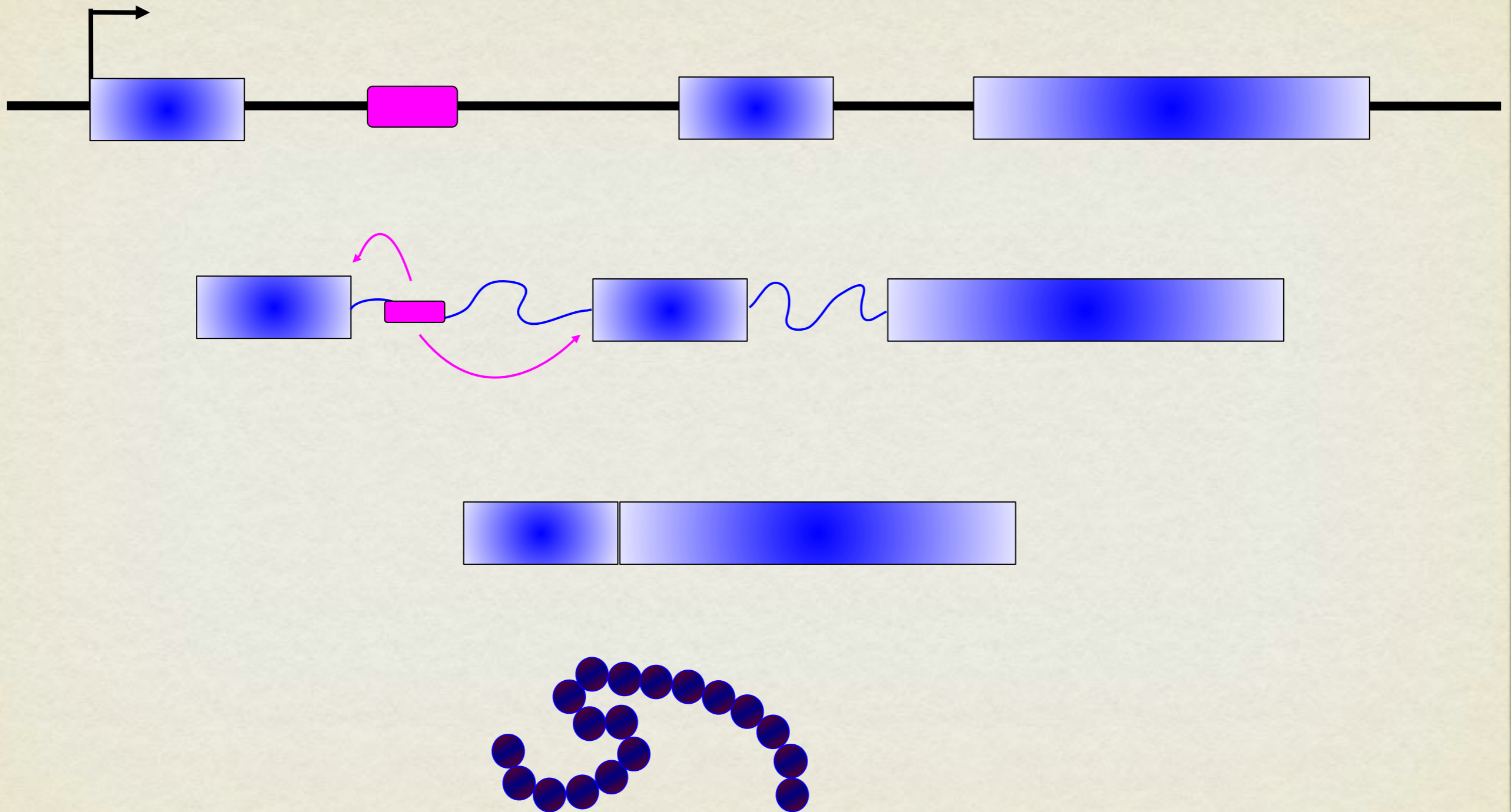
disruption of  
reading frame

gyrate atrophy  
DMD

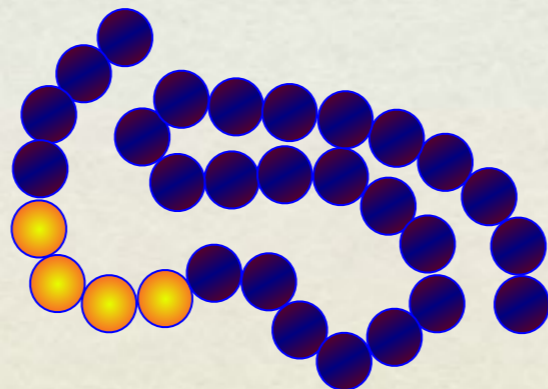
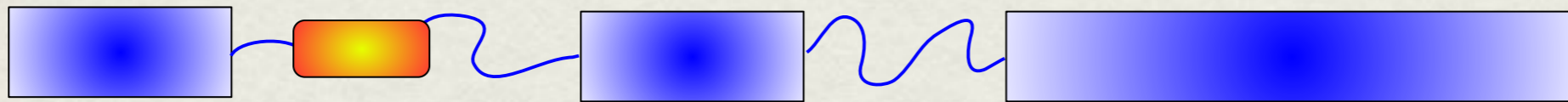
disruption of  
splicing

huntington disease

# ALTERNATIVE SPLICING AND/OR NEW PROTEIN ISOFORMS

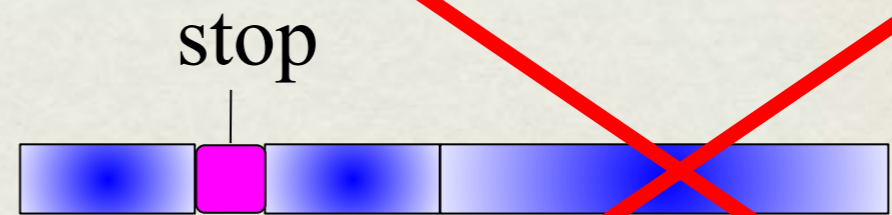
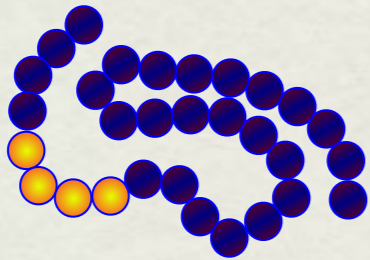


# EXONIZATION





# CONSEQUENCES OF EXONIZATION



Novel protein isoform

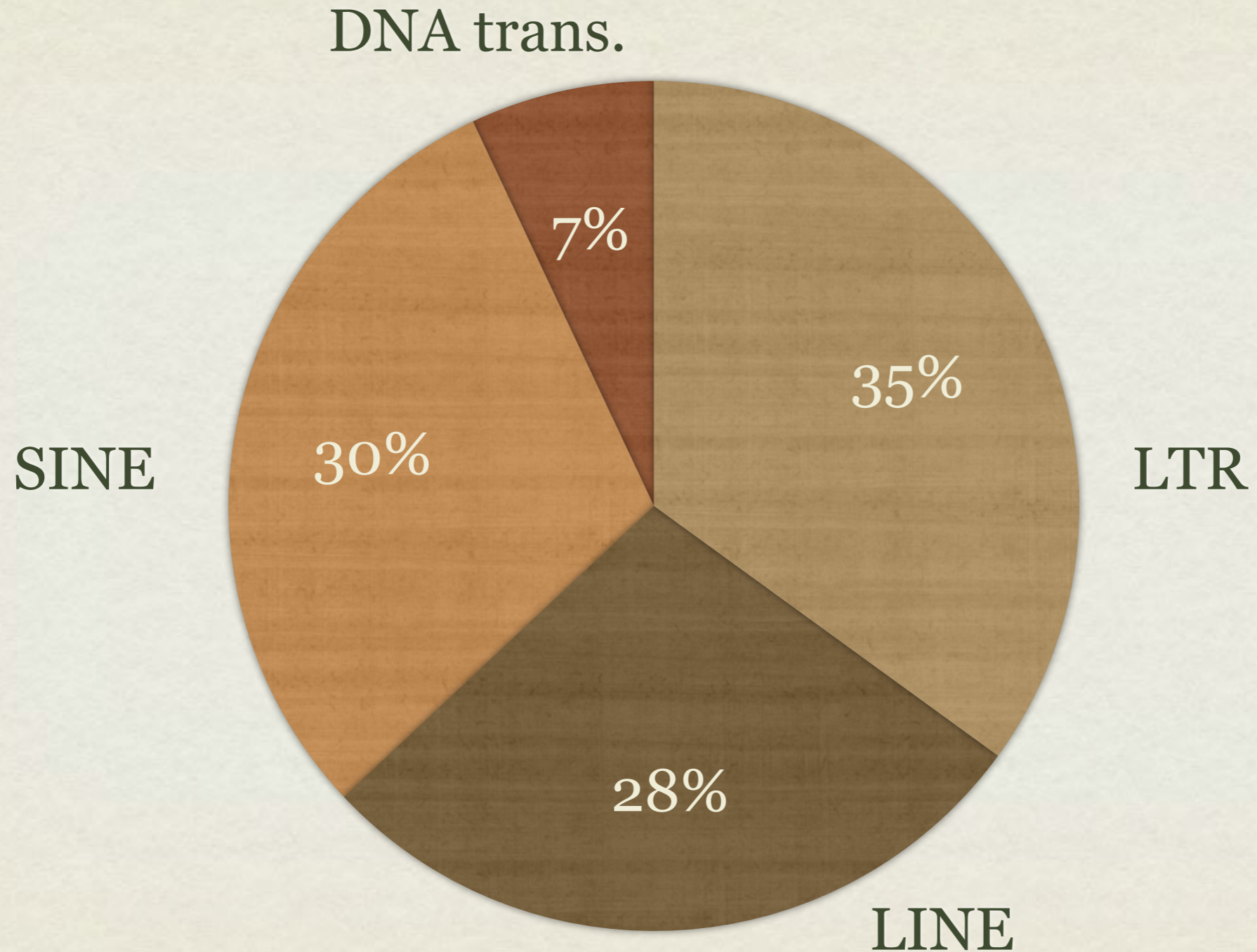
mRNA degradation via NMD

# PROTEIN CODING SEQUENCES

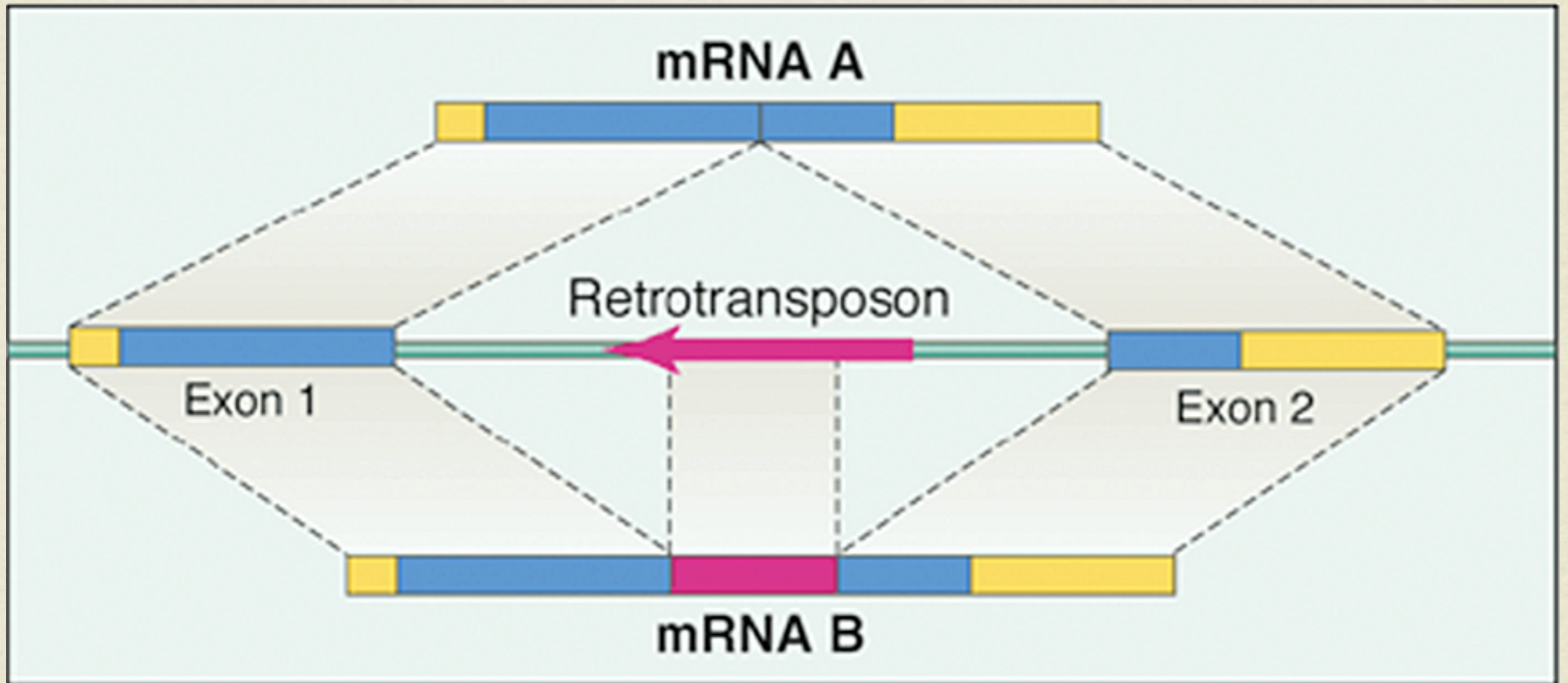
- Different retroposons contribute to protein variability but ...
- Alu elements are predisposed for the inclusion into open reading frames
  - cryptic splicing sites exist in an Alu element
  - over 2000 human mRNAs known to have an Alu in open reading frame
  - an Alu element in an ORF can change protein properties



# TE TYPES IN MAMMALIAN CODING SEQUENCES

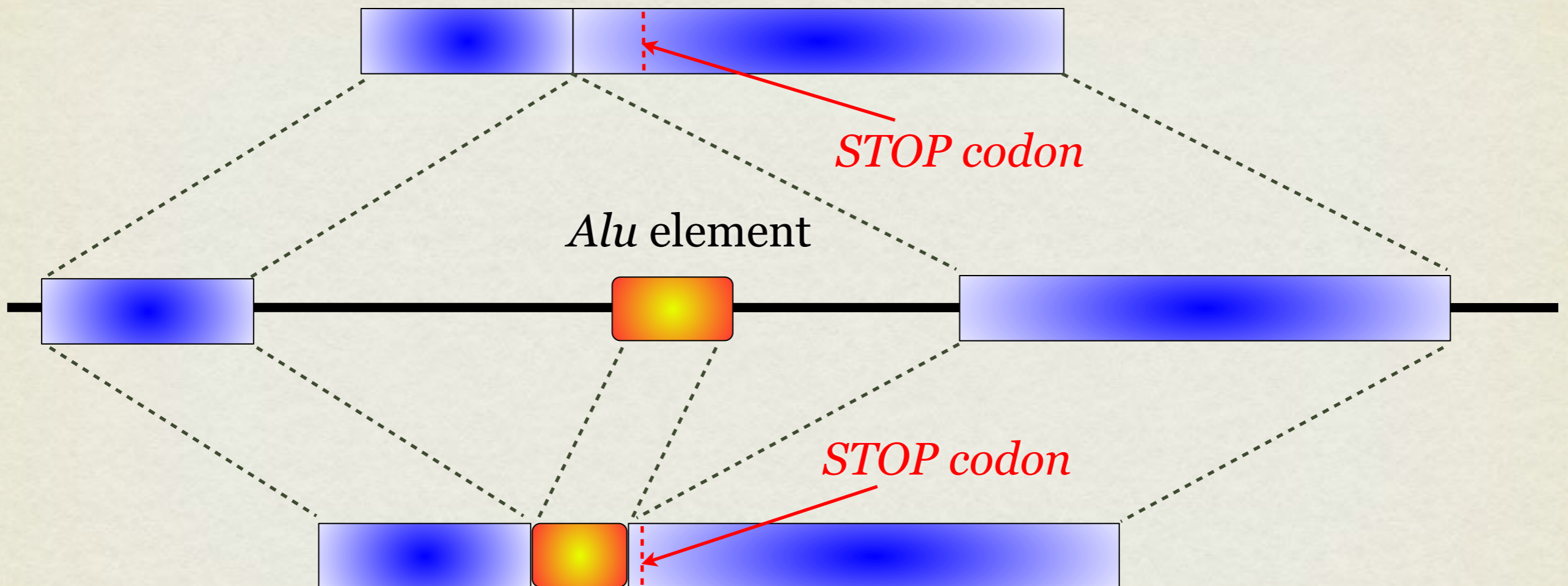


# HOW TO TEST NEW MOTIFS AND KEEP AN OLD FUNCTION



# DECAY ACCELERATING FACTOR

mRNA without an Alu-cassette -> hydrophobic C-terminus



mRNA with an Alu-cassette -> hydrophilic C-terminus

# Ancestral gene

Exon 1

Exon 2

Gene duplication

Gene copy A

Gene copy B

Retrotransposon insertion

Retrotransposon

Exon 1

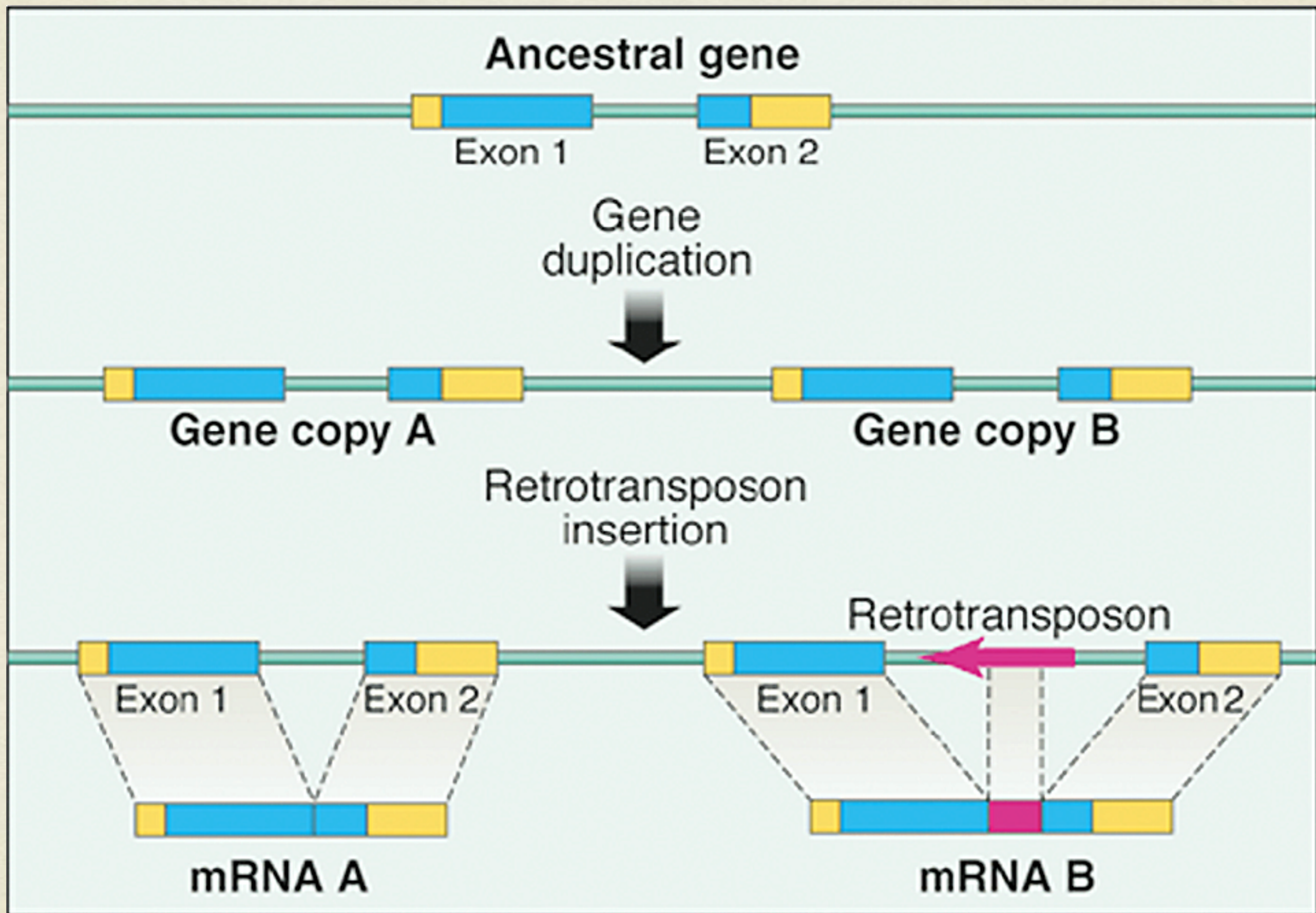
Exon 2

Exon 1

Exon 2

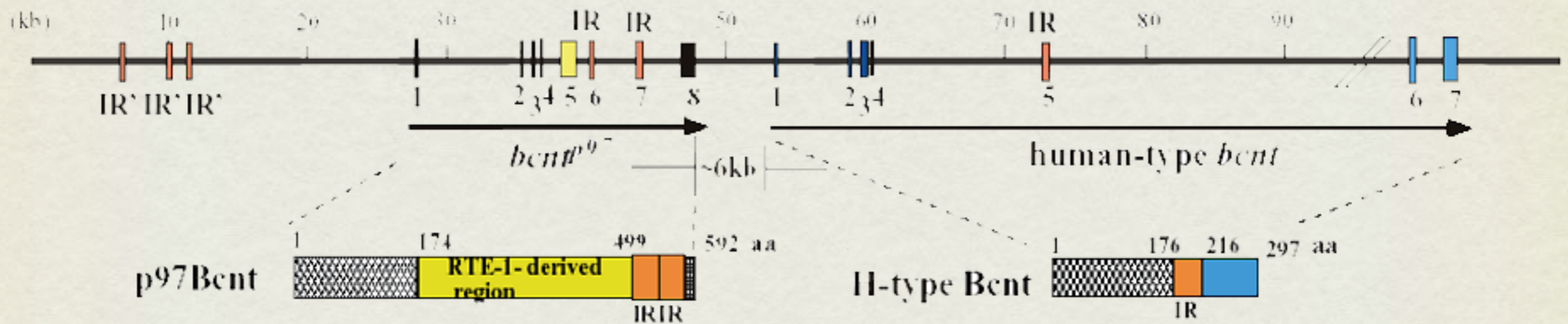
mRNA A

mRNA B

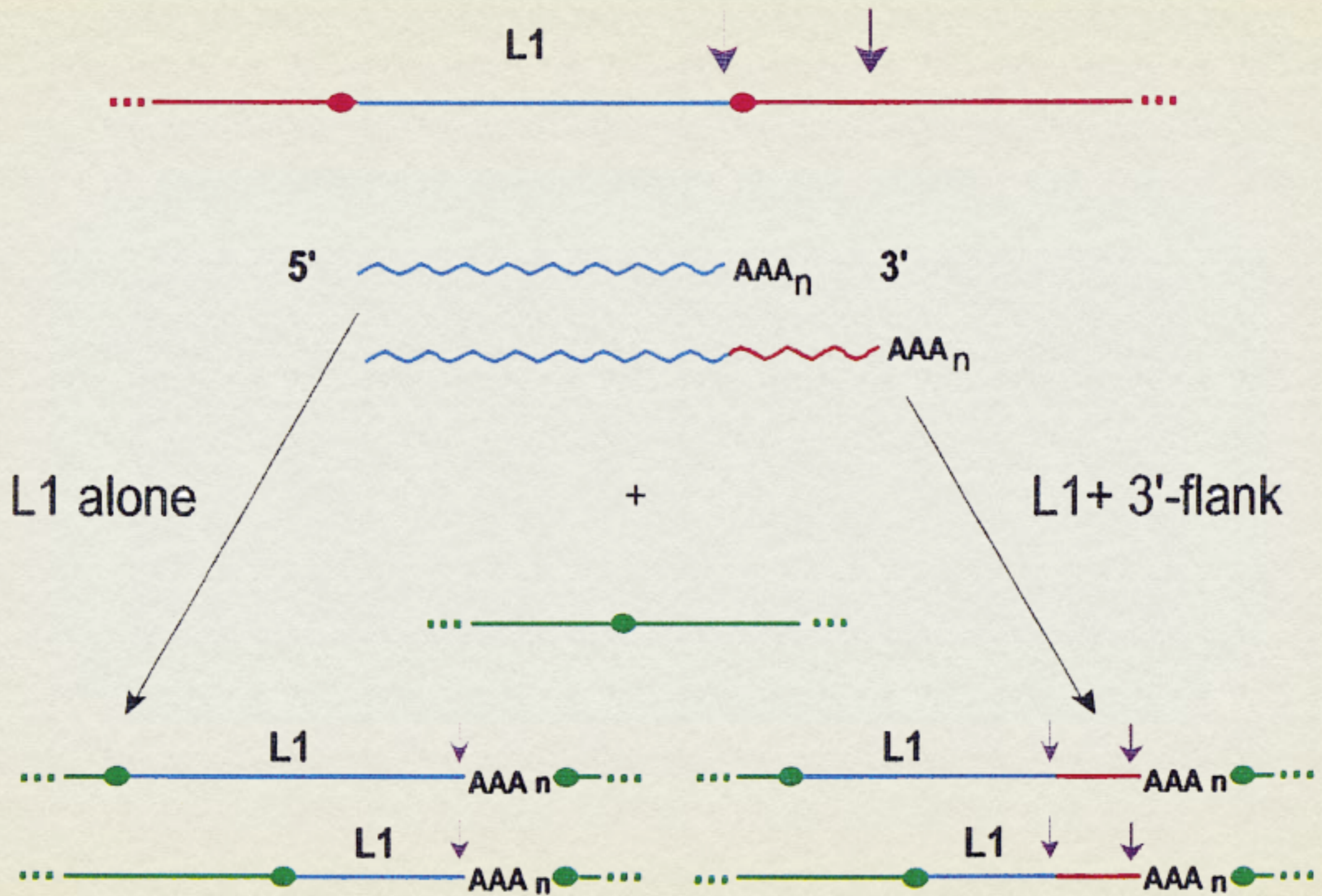


# HOW TO TEST NEW MOTIFS AND KEEP AN OLD FUNCTION

ruminant *Bcnt* gene



# DNA TRANSDUCTION BY L1 TRANSPOSON





**HYPOTHESIS:**

transduction can provide means for genome shuffling

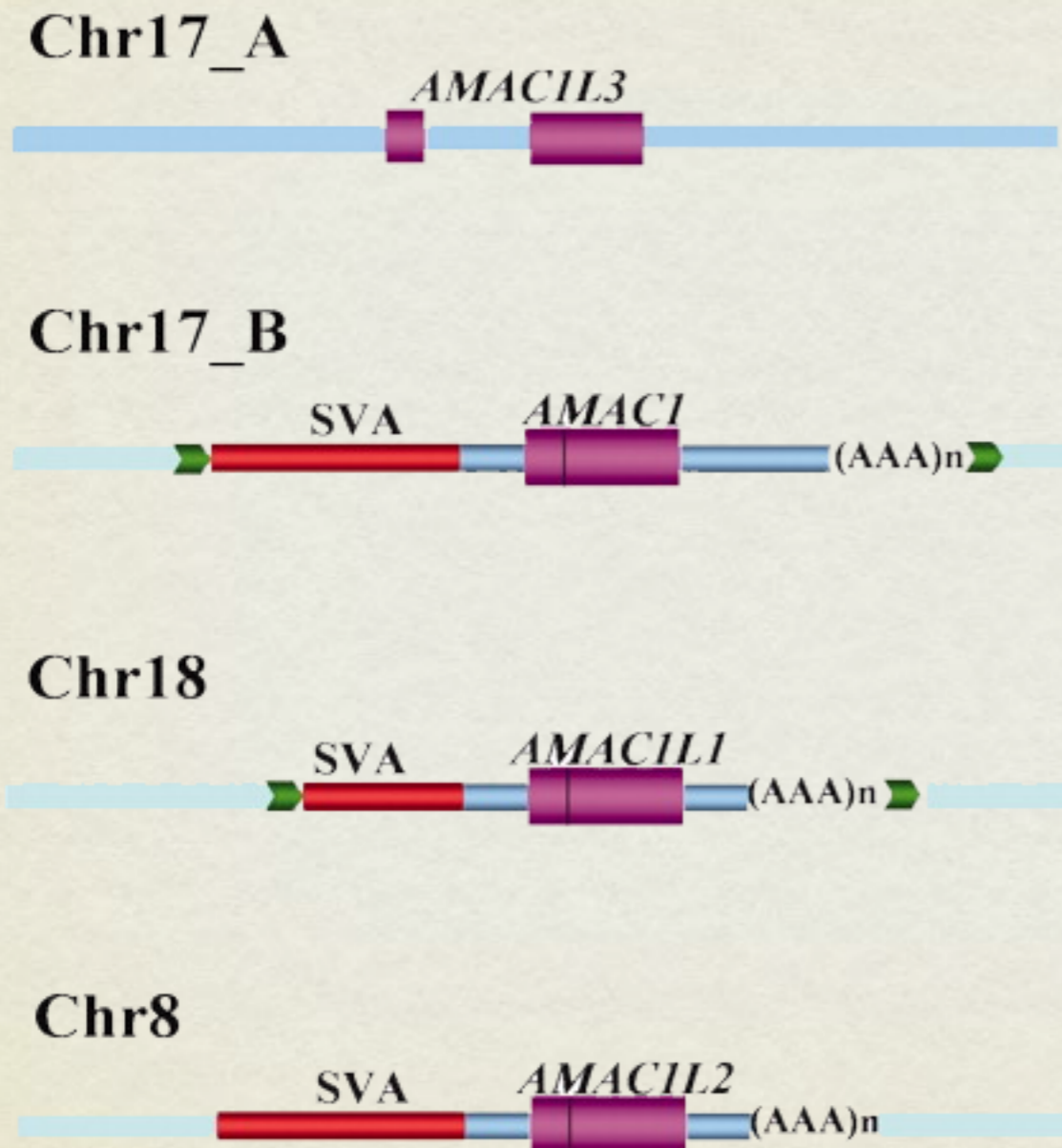


# SOME STATISTICS

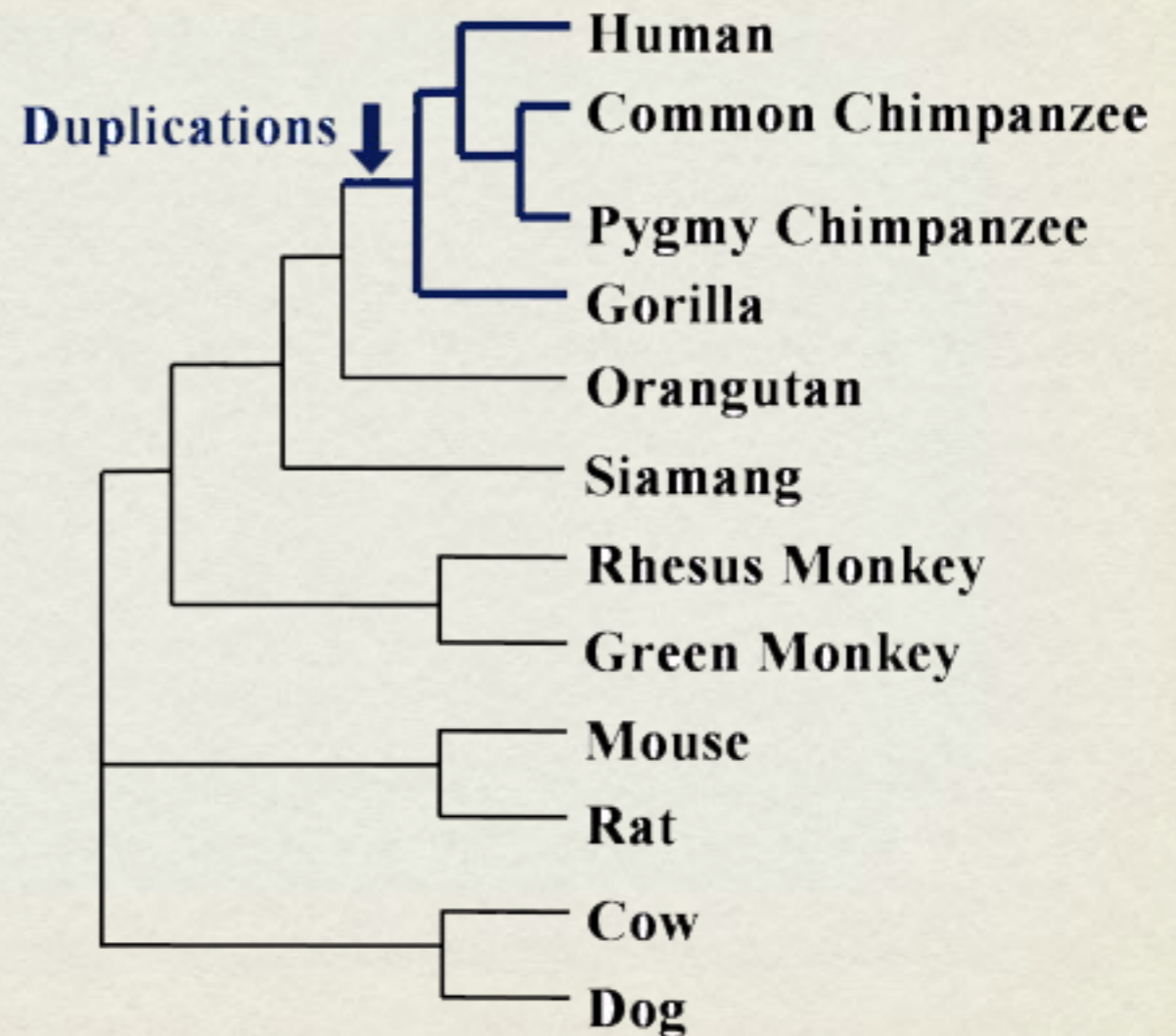
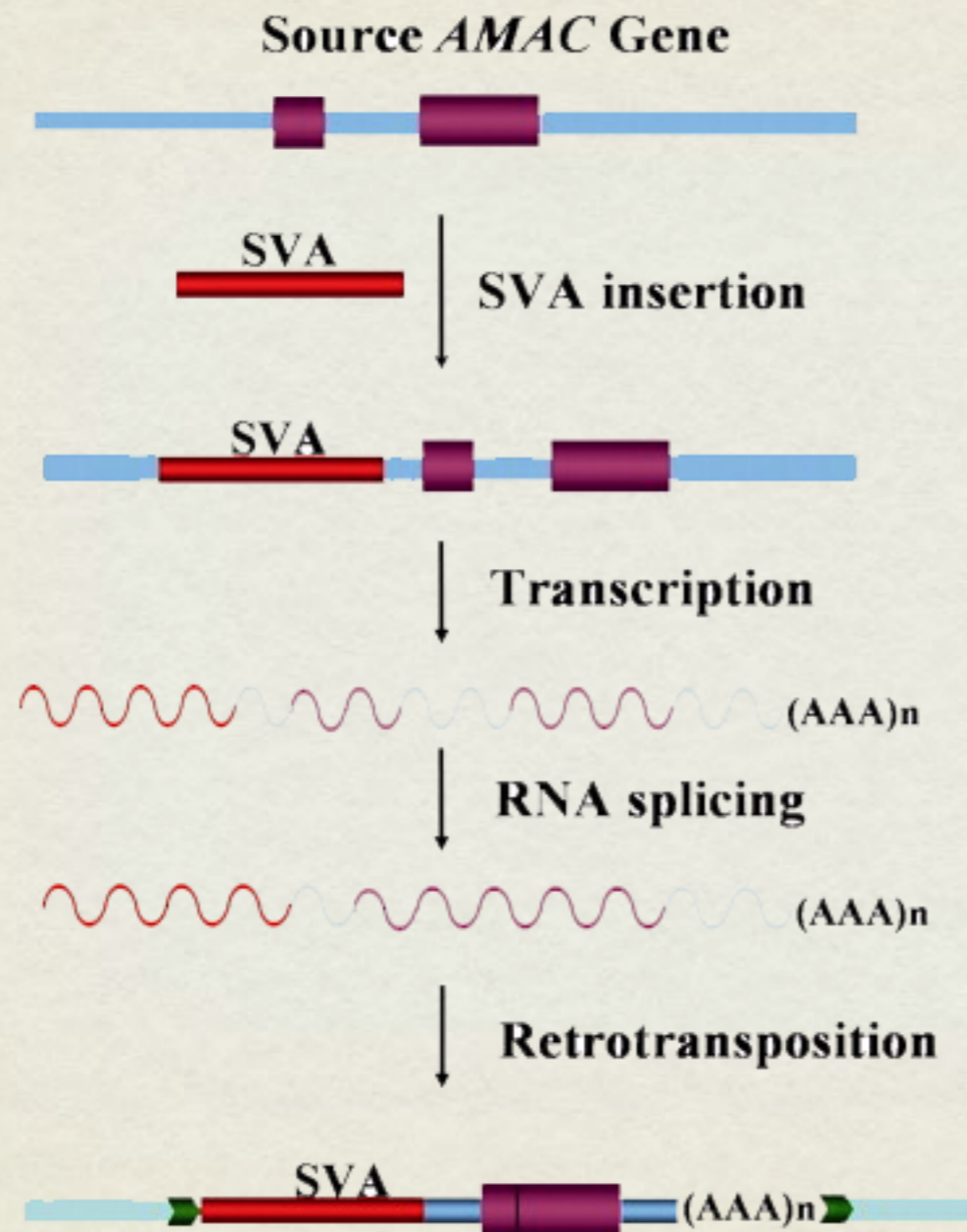
- number of L1 elements in the human genome 72,148
- number of L1 elements that caused 3' transduction of at least 30 nt - 6,178
- transduction frequency - 8.6%
- the longest transduced genomic sequence - 2883 nt
- no protein coding sequence detected in the transduced sequences



# SVA TRANSDUCTION-MEDIATED GENE DUPLICATION



# SVA TRANSDUCTION-MEDIATED GENE DUPLICATION



Evolution takes any chance  
to (re)use existing material





WESTFÄLISCHE  
WILHELMS-UNIVERSITÄT  
MÜNSTER



Search University Sites

Go

▶ [Extended Search](#)

[Home](#) [Publications](#) [Teaching](#) [Conferences](#) [Tools](#) [Internal](#)



- ▶ [Research](#)
- ▶ [People](#)
- ▶ [Presentations](#)
- ▶ [Contact](#)
- ▶ [Links ...](#)

Welcome to the

## Institute of Bioinformatics WWU Muenster

Bioinformatics is a research, development, or application of computational tools and approaches for expanding the use of biological data, including those to acquire, store, organize, archive, analyze, or visualize such data. Its goal is to enable biological discovery based on existing information or in other words transform biological information into knowledge. In recent years bioinformatics became integral part of biology and it is present in almost all biomedical disciplines.



### News

2015-09-02

"TRIM32 modulates pluripotency entry and exit by directly regulating Oct4 stability" by Bahnassawy et al. has been published by Scientific Reports.

▶ [View](#)

SCIENTIFIC  
REPORTS

2015-06-07

Deep sequencing in conjunction with expression and functional analyses reveals

20th  
1995-2015  
Clinical  
Cancer  
Research  
Anniversary

<http://bioinformatics.uni-muenster.de/>

[wojmak@uni-muenster.de](mailto:wojmak@uni-muenster.de)