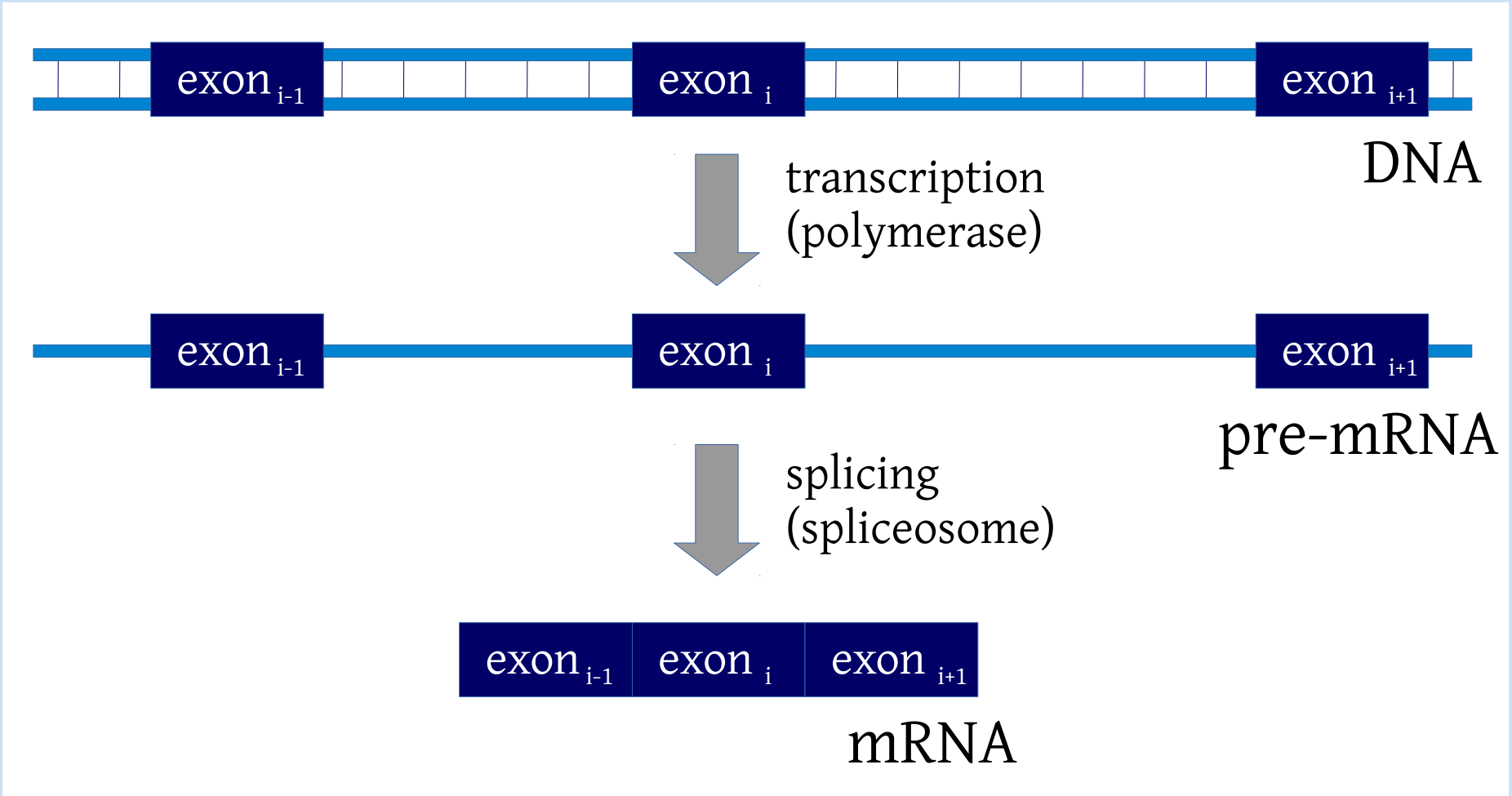
Splicing analysis

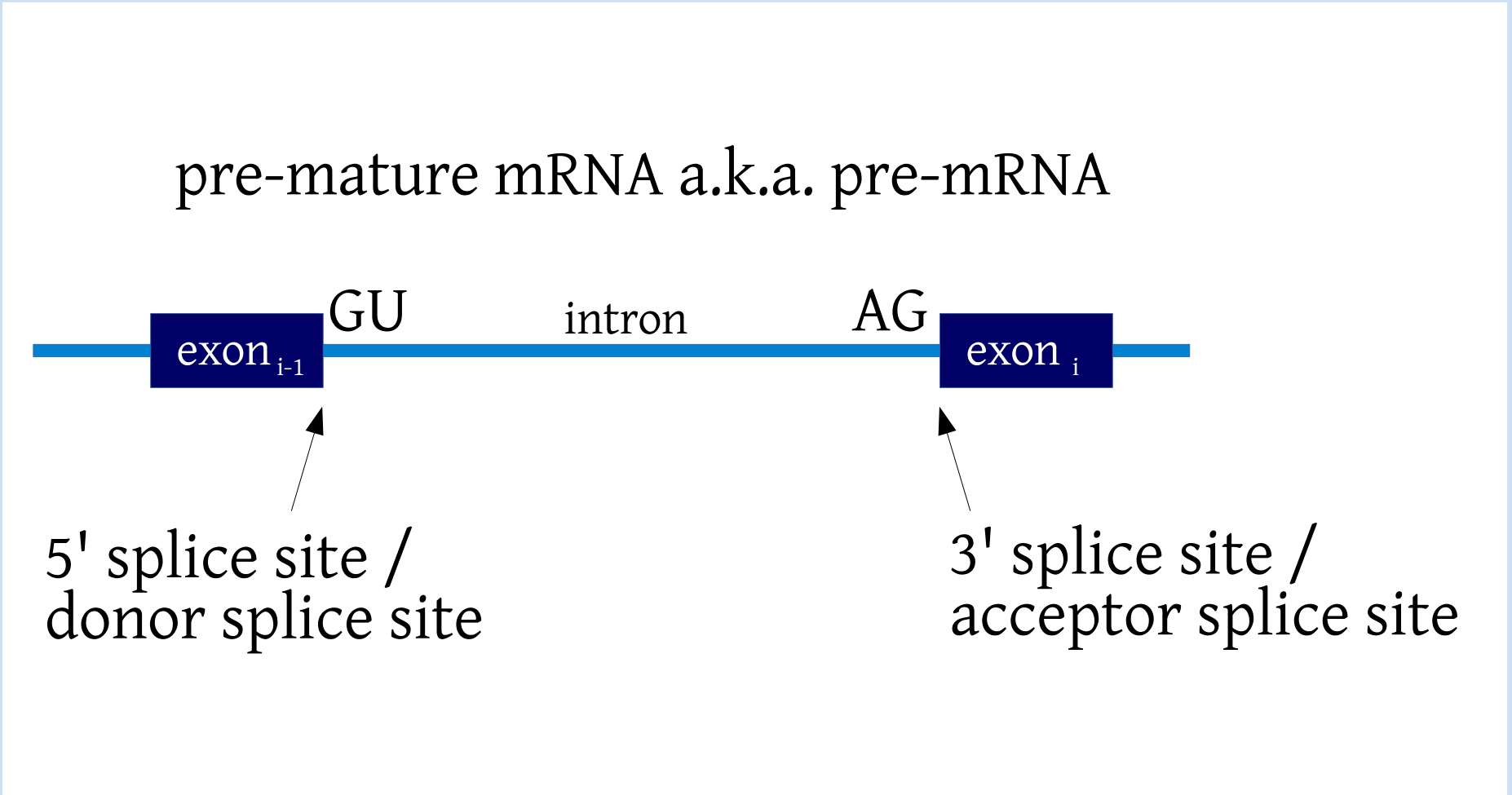
Irina Pulyakhina

WTCHG, RNA-Seq course, 29-April

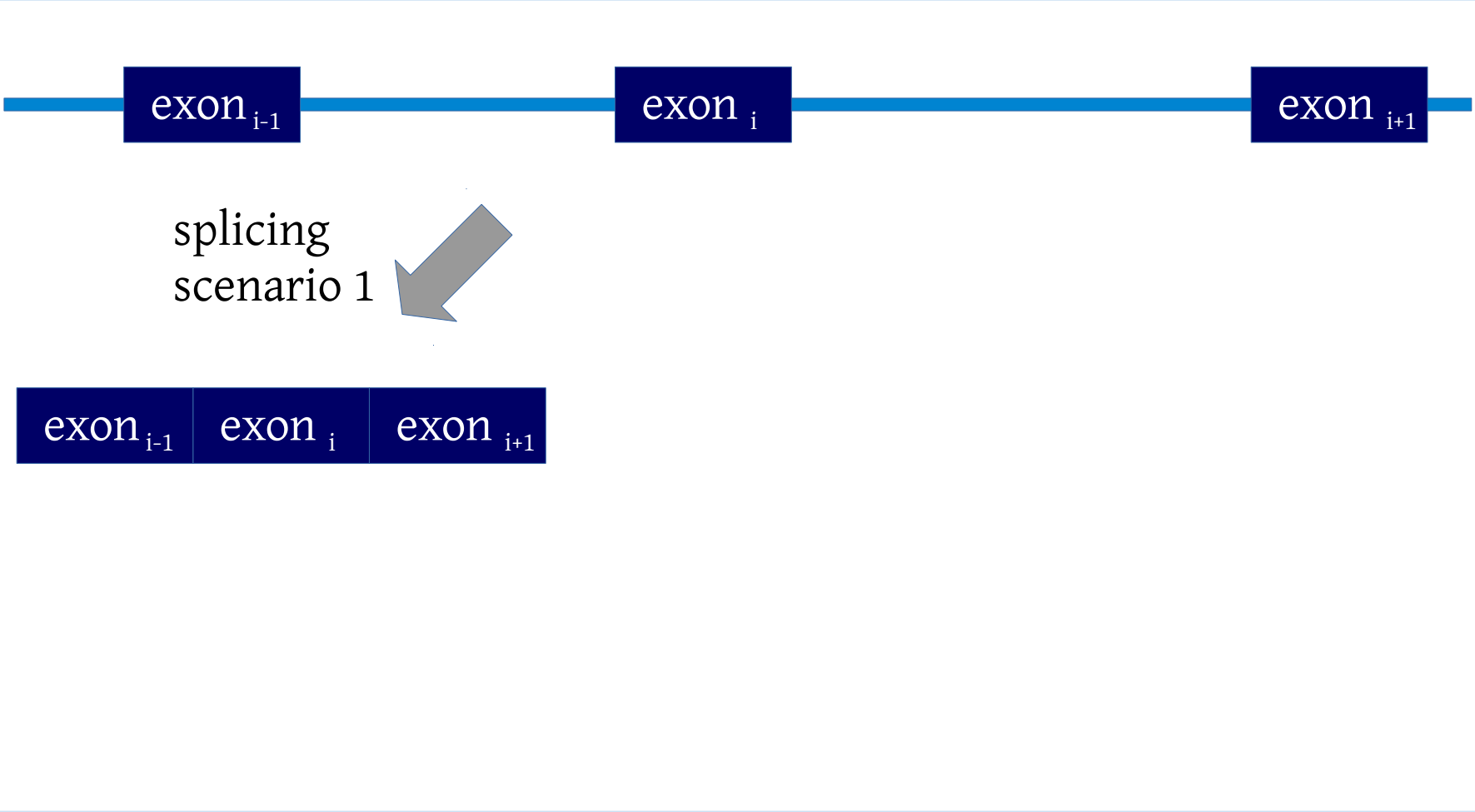
Splicing


exon i-1
exon i
exon i+1
DNA
transcription
(polymerase)
exon i-1
exon i
exon i+1
pre-mRNA
splicing
(spliceosome)
exon i-1
exon i
exon i+1
mRNA

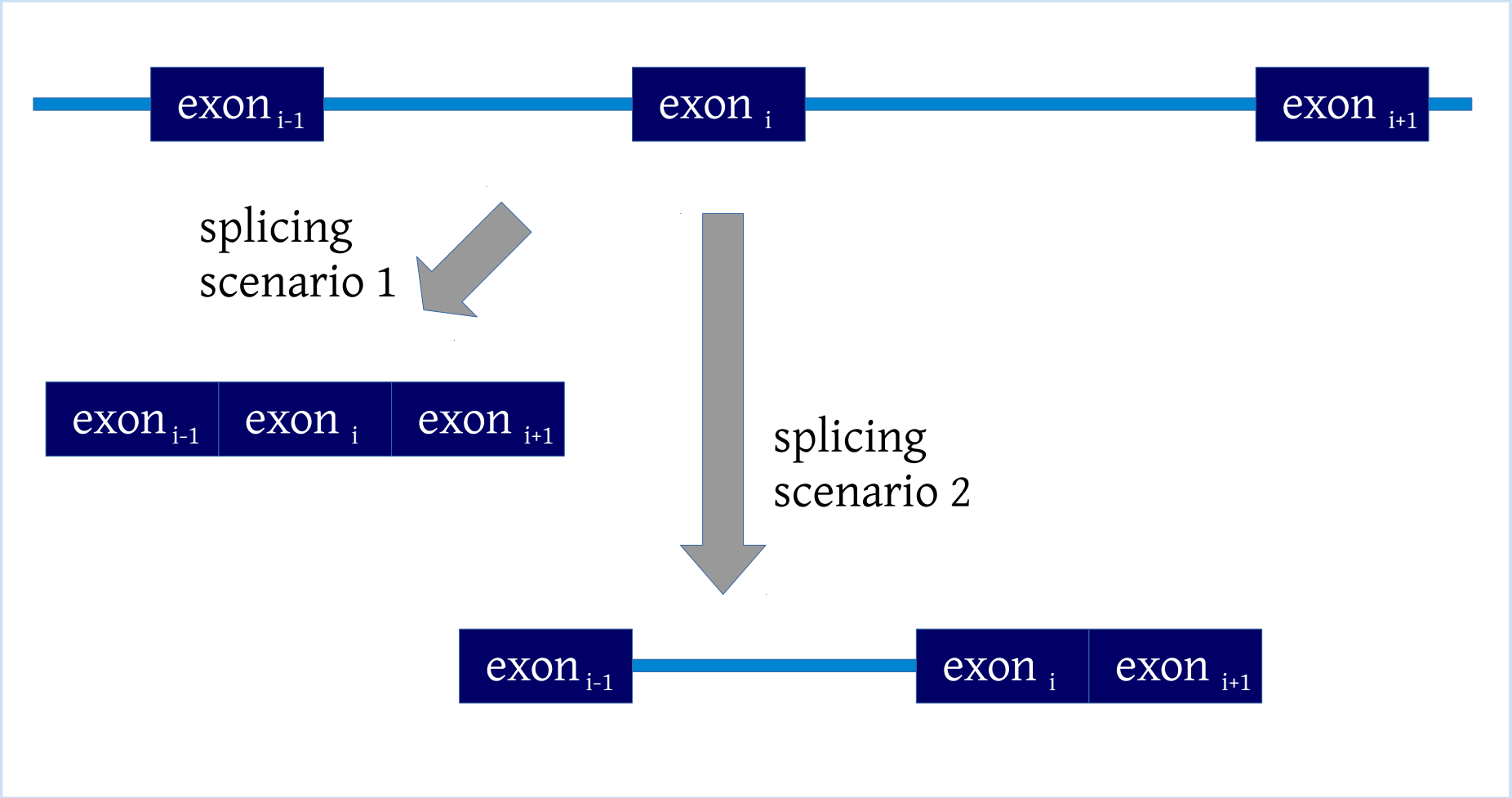
Terminology



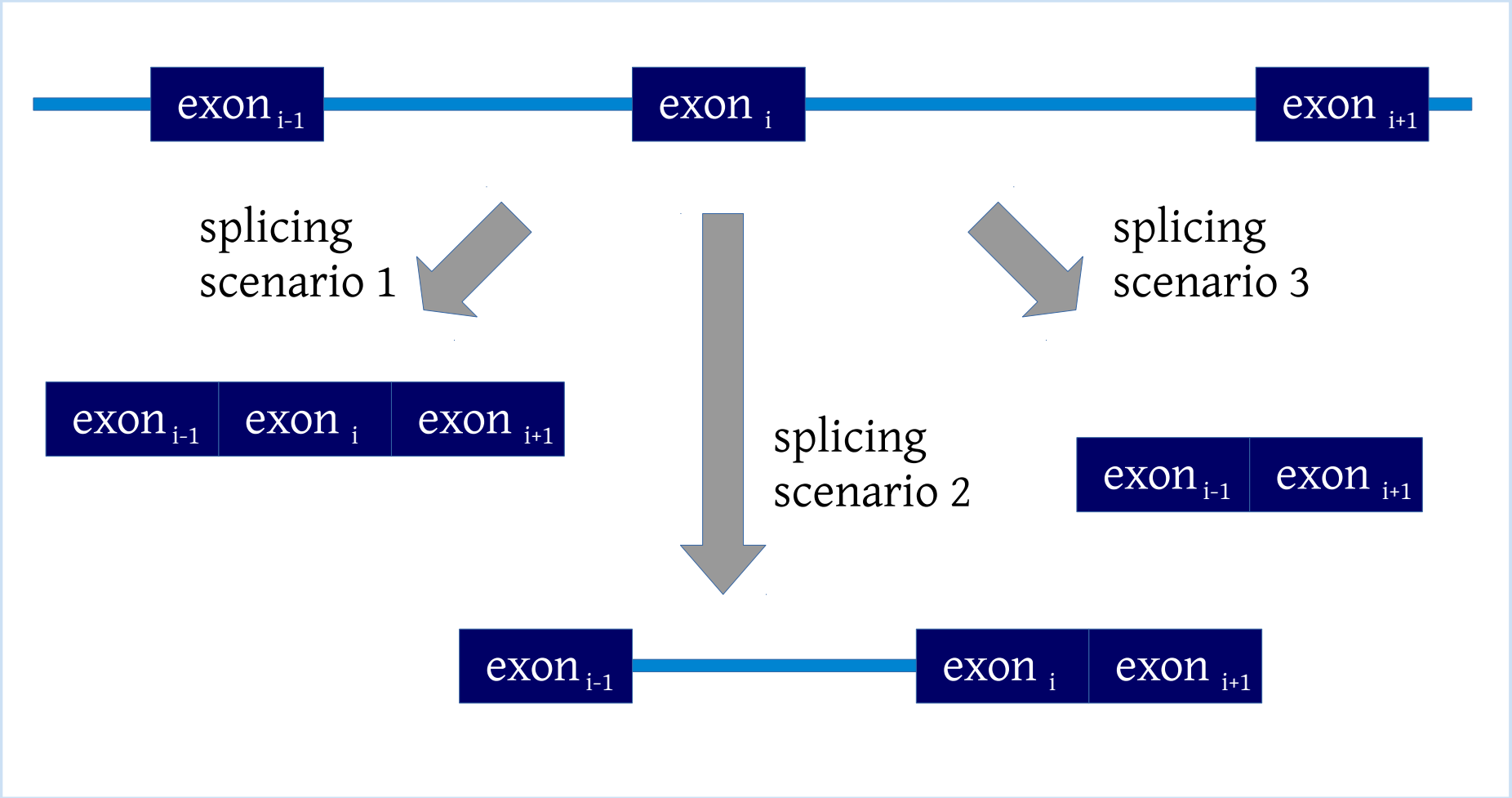
Alternative splicing (1)



Alternative splicing (2)

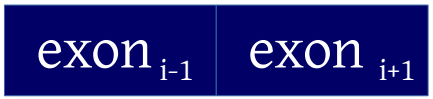


Alternative splicing (3)



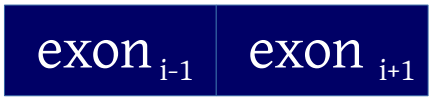
Types of alternative splicing events (1)

exon skipping

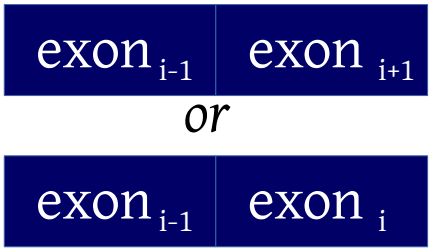


Types of alternative splicing events (2)

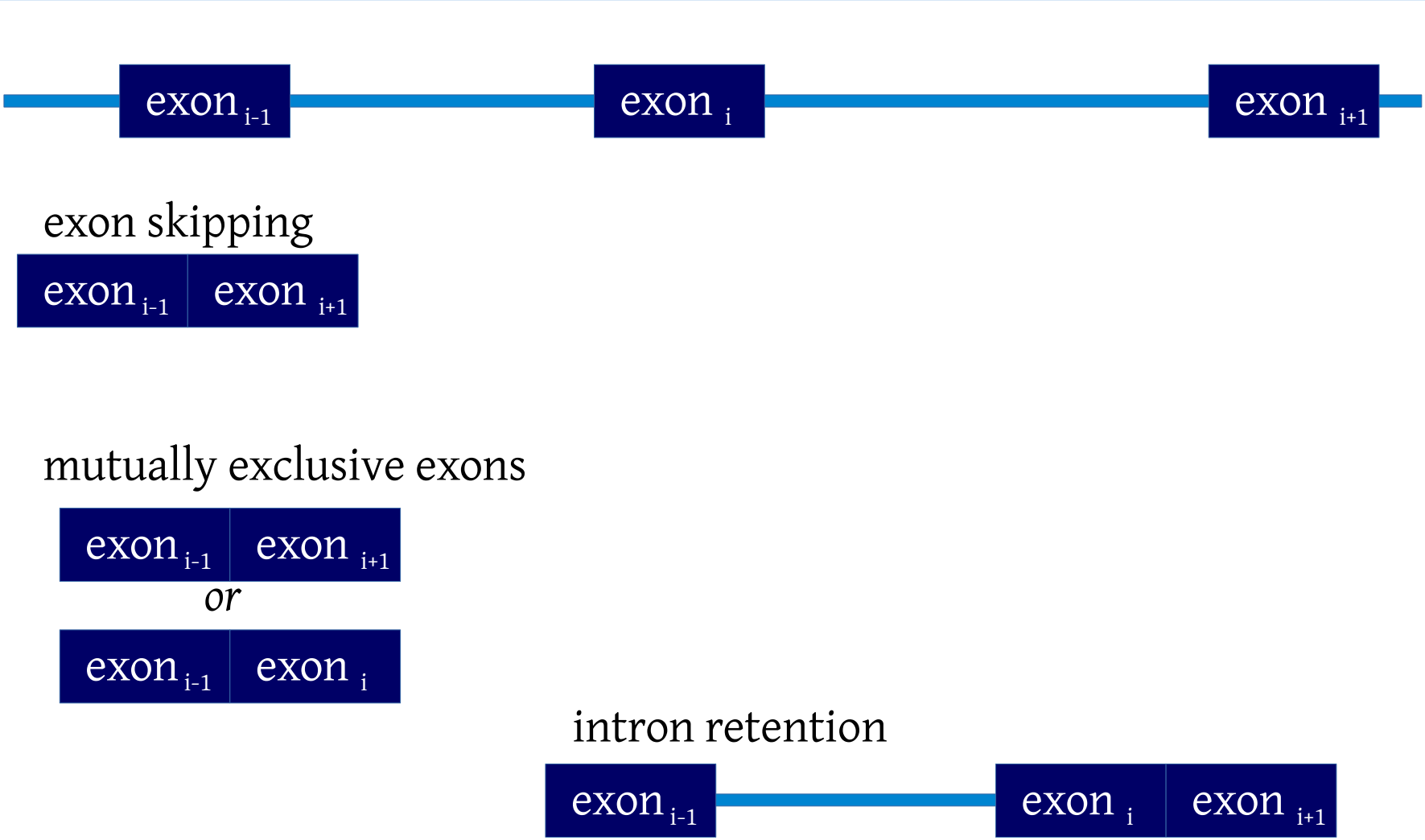
exon skipping



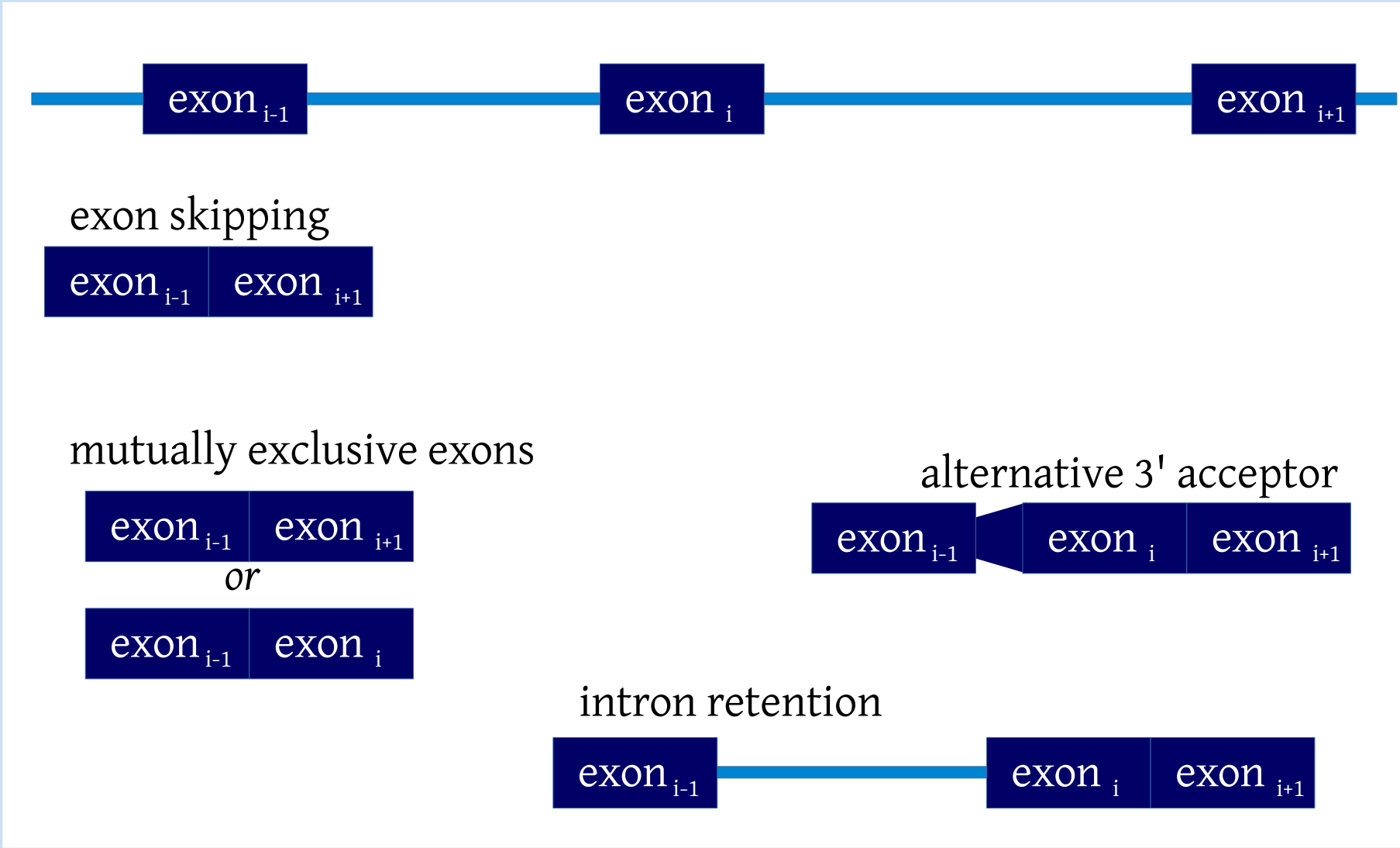
mutually exclusive exons



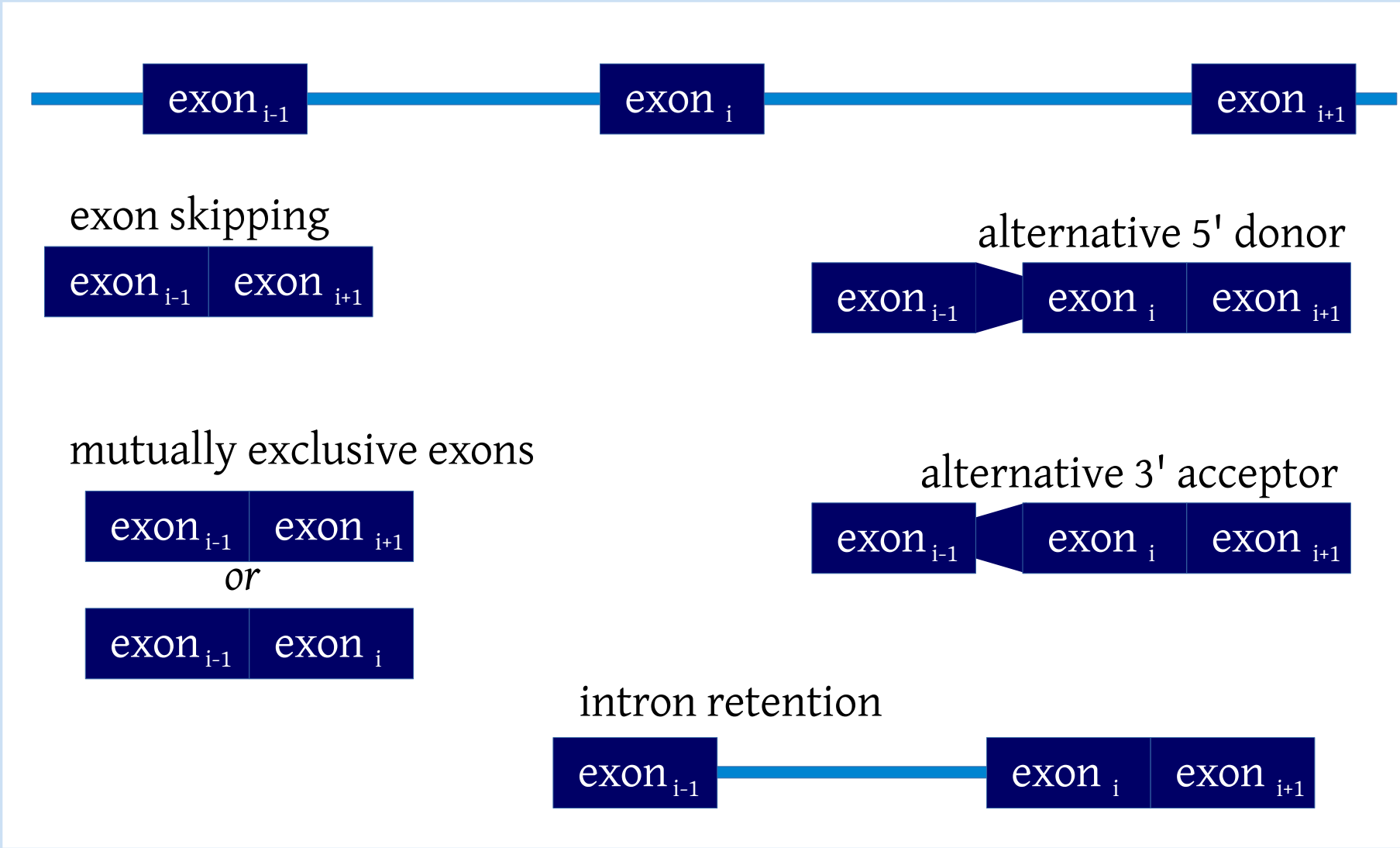
Types of alternative splicing events (3)



Types of alternative splicing events (4)



Types of alternative splicing events (5)



Functional aspects of alternative splicing

- enlarging the pool transcripts from one gene

- may lead to new protein functions

- more often, however, destroys protein function

What can alternative splicing do to mRNA

- introduce a new stop codon (intron retention, alt. 5/3 site)

- skip a stop codon (exon skipping)

- frameshift (any event)

Reference transcript

One transcript is chosen arbitrarily to be considered as a reference, standard splicing scenario.

Other transcripts are considered alternative and will be compared to the reference one.

Reference transcript is usually the first annotated transcript; rarely the one containing all possible exons; rarely the most abundant one.

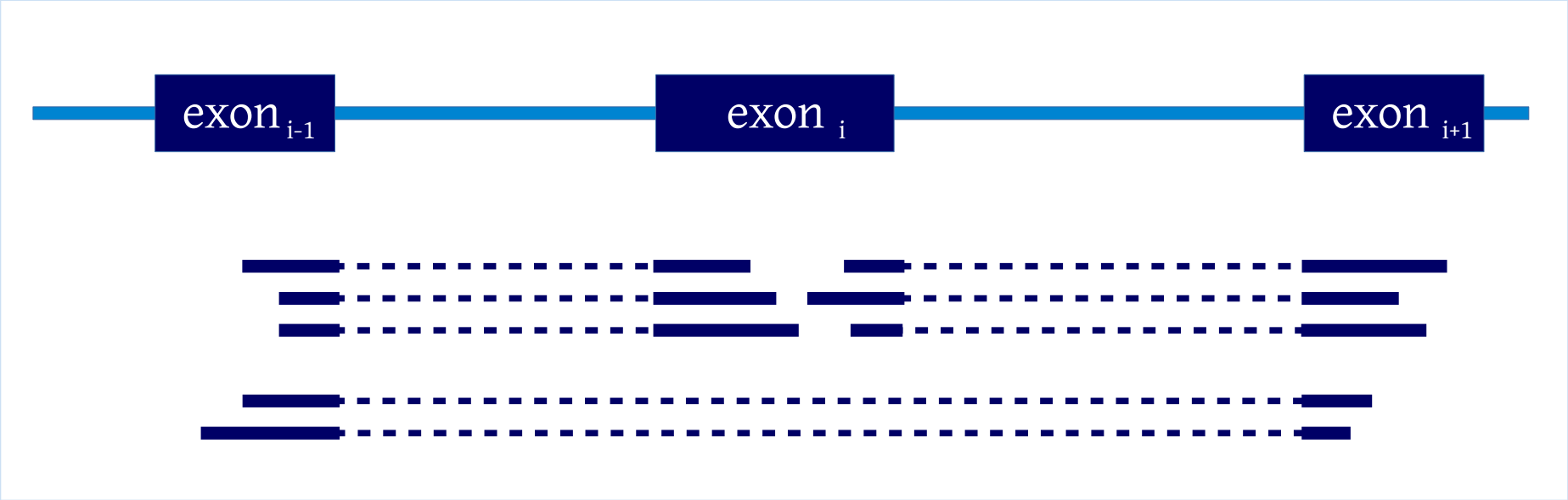
Detecting splicing in NGS data (1)

- Number of reads mapped to an exon-exon junction indicate the abundance of that junction.

- All junctions of a transcript form a baseline of transcript expression.

- A deviation from the baseline indicates an alternative splicing event.

Detecting splicing in NGS data (2)



Detecting splicing in NGS data (3)

- Coverage is never even:
  sequencability bias
  mapability bias
  5'-3' bias

- We rarely observe just one transcript with an alternative splicing event. More often there is a mixture of transcripts present in different abundances.

Detecting splicing in NGS data (4)

Integrated approach:

- using coverage information to calibrate the baseline

- using sequence information to account for sequencability and mapability

- using split reads to estimate the corrected abundance of exon-exon junctions

Does transcript assembly work?

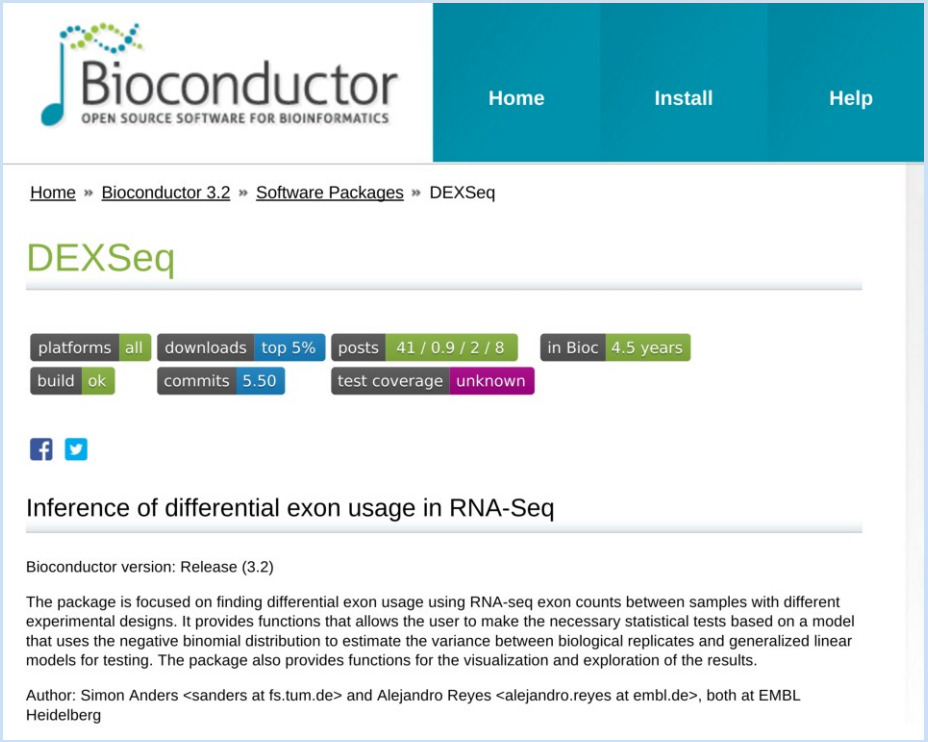
No.

Does transcript assembly work?

No.
All tools reporting full-length transcript structures showed very low overlap and reproducibility (Scripture, Cufflinks, iReckon, DiffSplice).

But we can still detect alternative splicing events and make local transcript structure predictions (~ one-four exons).

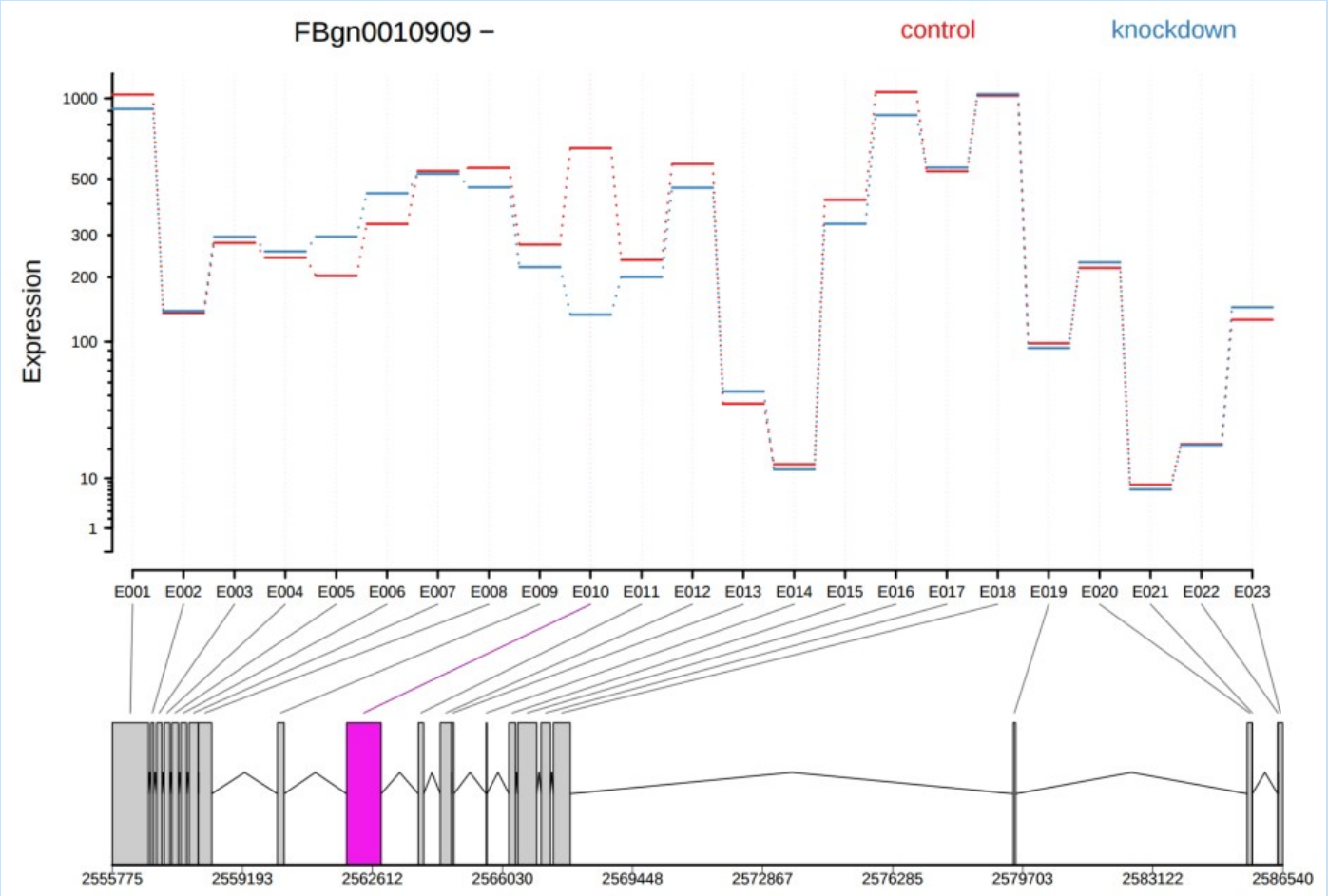
DEXSeq to detect splicing (1)



- adjusts for the average gene expression

- analyzes adjusted exon coverage locally

DEXSeq to detect splicing (2)



From DEXSeq manual

Alternative splicing and diseases

Abnormally spliced mRNAs are highly abundant in tumors and cancerous cells.

Breast cancer cells have elevated levels of the splicing factor SF2/ASF.

Overexpression of a truncated splice variant of the FOSB gene in a specific population of neurons was shown to be the causal mechanism involved in the induction and maintenance of an addiction to drugs and natural rewards.

Manipulating alternative splicing

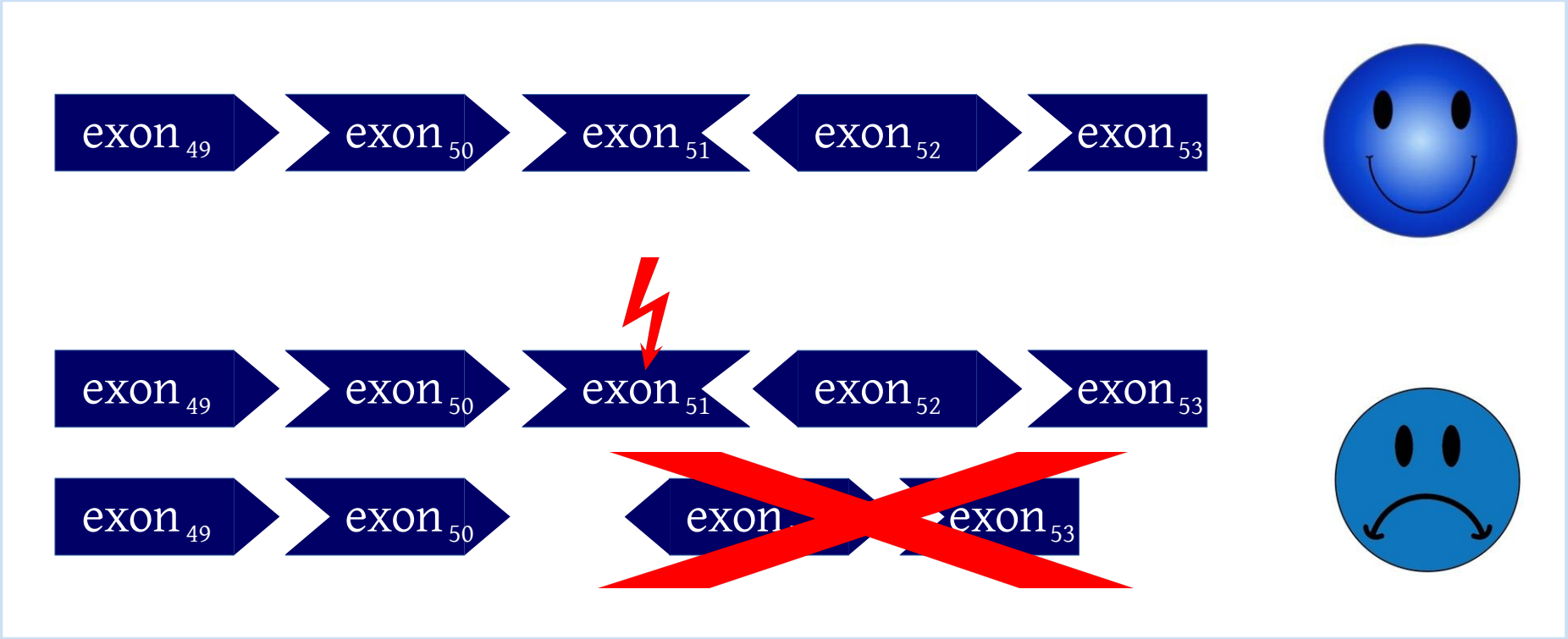
Case of Duchenne Muscular Dystrophy

Muscular protein dystrophin connects the cytoskeleton of a muscle fiber to the surrounding extracellular matrix through the cell membrane.

A mutation in the dystrophin gene (loc Xp21) introduces a frame shift and leads to pre-mature protein truncation.

Manipulating alternative splicing

Case of Duchenne Muscular Dystrophy



Manipulating alternative splicing

Case of Duchenne Muscular Dystrophy

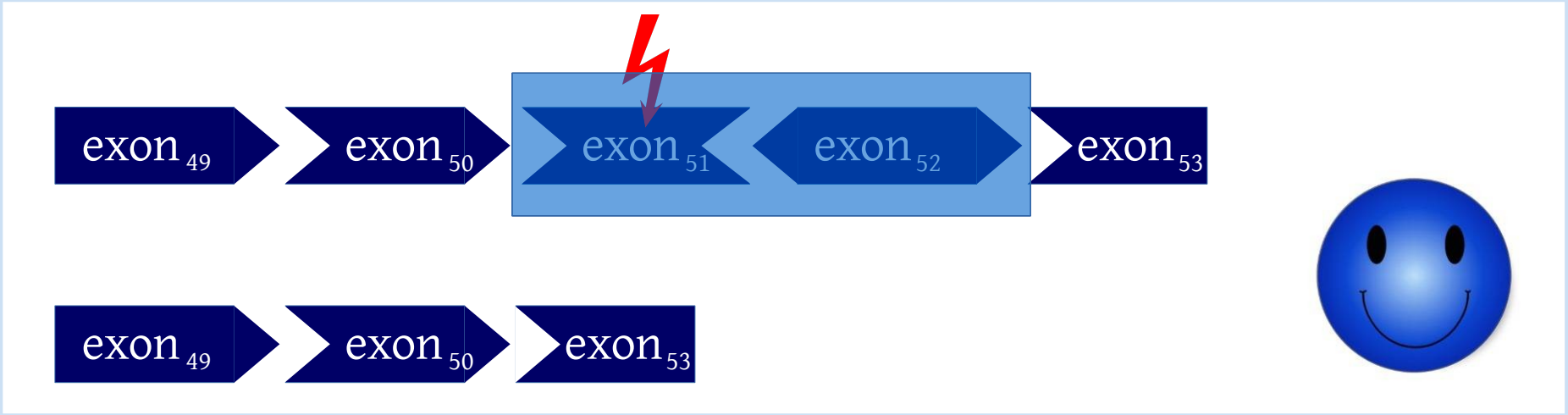
Antisense oligonucleotide masks several exons.

Spliceosome does not recognize this exon and skips it (exon skipping).

Frame is restored, transcript is not truncated (it is only missing a short middle part).

Manipulating alternative splicing

Case of Duchenne Muscular Dystrophy



A functional protein is produces. It is not equal to the wild type, however, it leads to a much milder phenotype.