

From isoforms to splice events: Interpretable splicing analysis with splicellogic

Beatriz Campillo^{1,2}, Michael I. Love^{2,3}

¹ Bioinformatics and Computational Biology Curriculum, UNC-Chapel Hill · ² Department of Genetics, UNC-Chapel Hill · ³ Department of Biostatistics, UNC-Chapel Hill



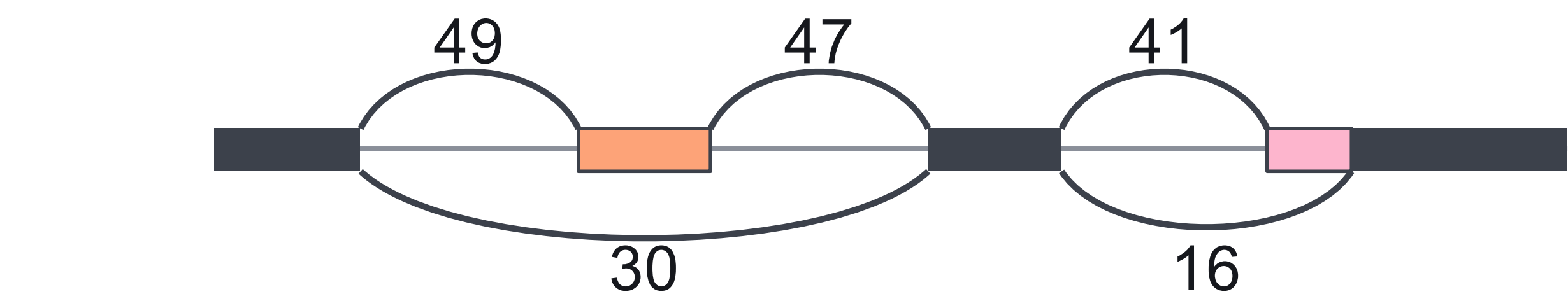
The University of North Carolina at Chapel Hill

BACKGROUND

1 Why long reads?

Short reads → junctions

Reads counted on each splice junction



Exon skipping PSI 0.62 Alt 3' splice site PSI 0.72

One event at a time, e.g. PSI = 41 / (41 + 16) = 0.72

Long reads → transcripts

Two whole transcripts flagged by a DTU test

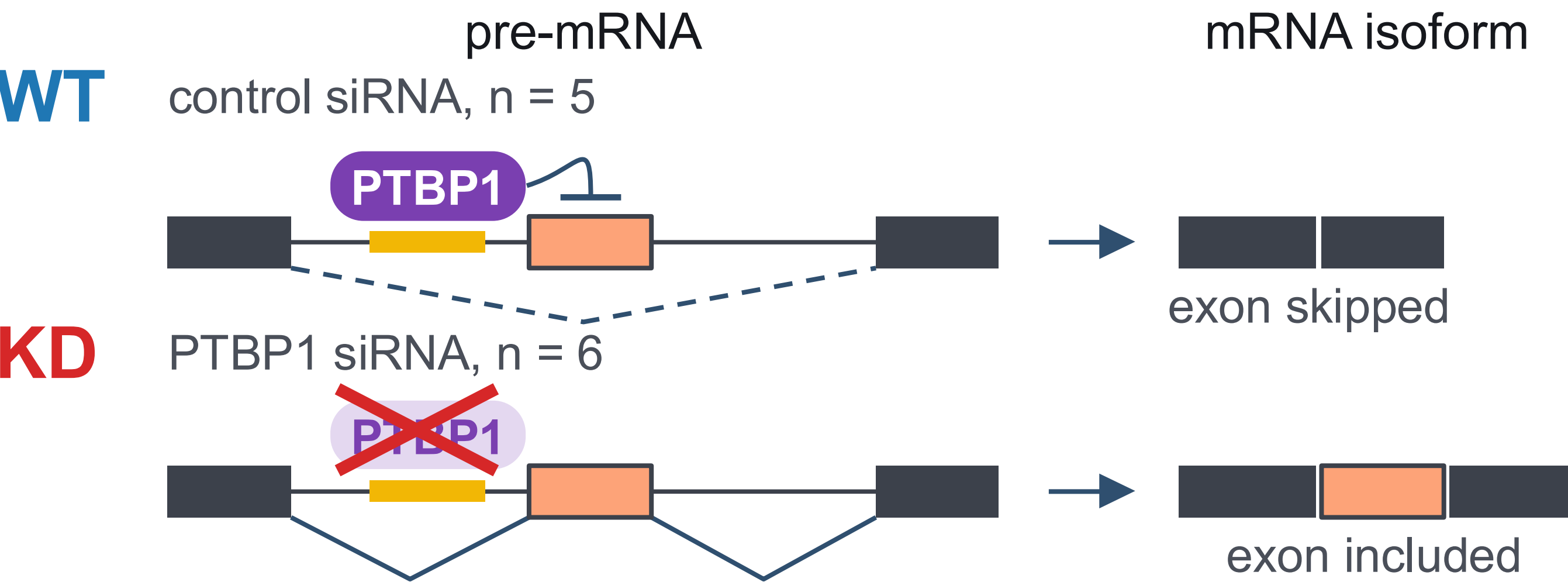


Alt first exon Exon skipping Alt 3' splice site

Every difference named at once, tied to T1 → T2

2 PTBP1 promotes exon skipping when binding upstream regions rich in CU

Data: PTBP1 siRNA knockdown in GTEx skin fibroblasts, Nanopore cDNA (Glinos et al., 2022) 11 libraries from 4 donors; with donor as a model covariate.



Hypothesis for the knockdown: skipped exons regulated by PTBP1 are included more and carry CU-rich sequence **upstream**.

WT vs knockdown

385,669 transcripts

GENCODE V48 oarfish v0.9.4

satuRn v1.18.0

Novel vs annotated

4,402 novel + 385,669 annotated

GENCODE V48

bambu v3.12.1



Computation: bambu and oarfish ran on AnVIL. **Mixed references:**

tximeta v1.30.0 imports GENCODE + novel quantifications

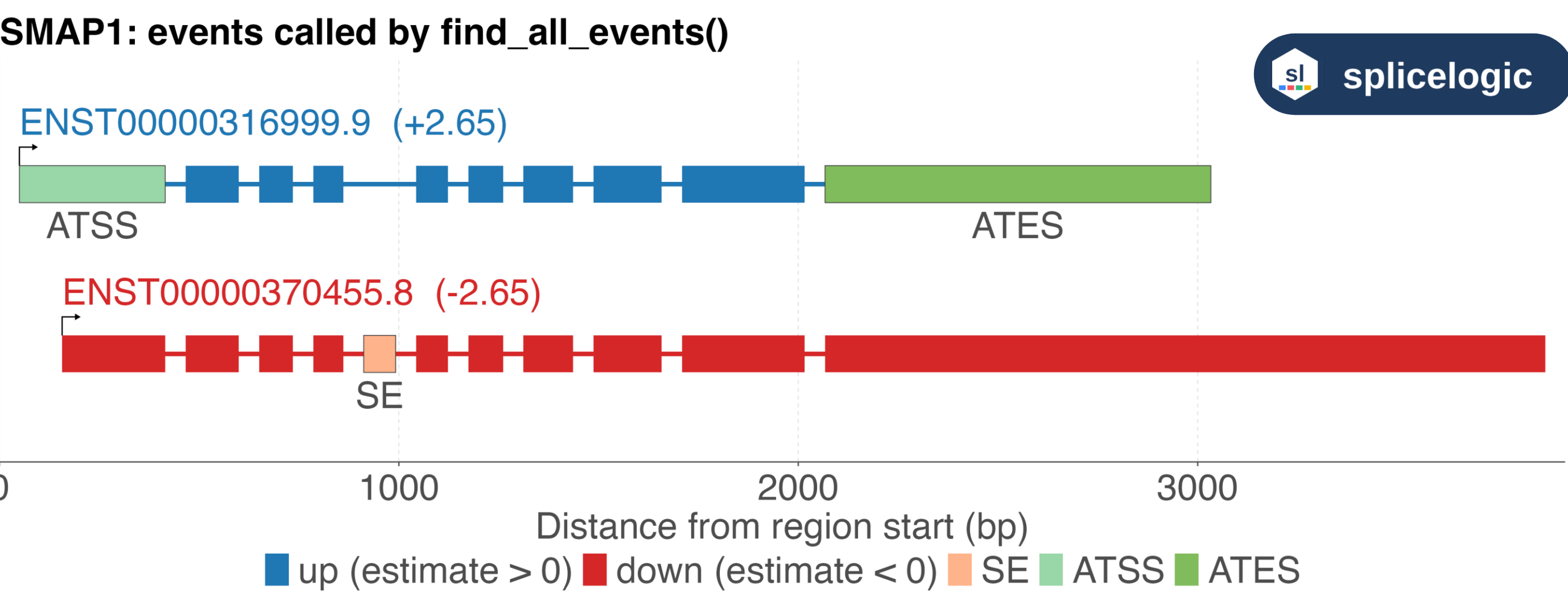
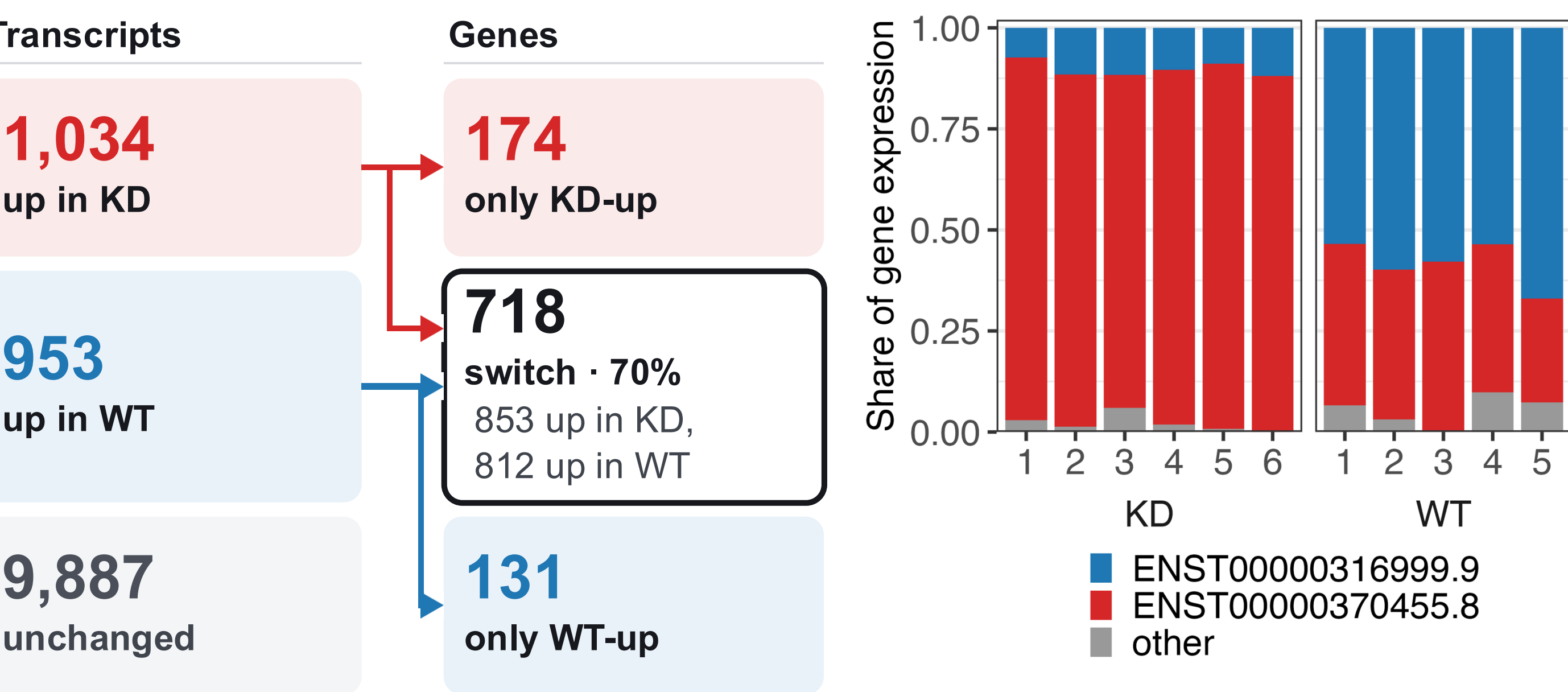
(oarfish --annotated / --novel), keeping the metadata for each part.

Glinos DA et al. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature* 608, 353–359 (2022).

PTBP1 knockdown · WT vs KD

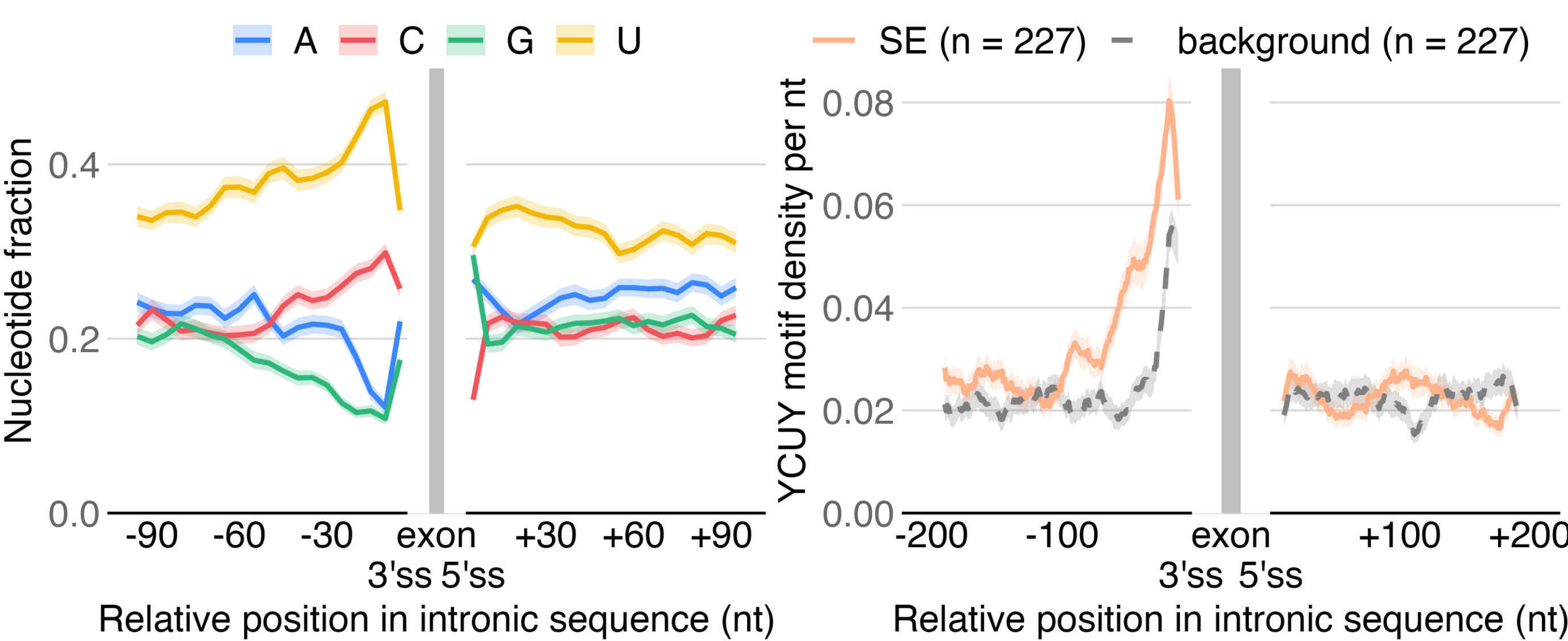
3 PTBP1 knockdown leads to isoform switches in 718 genes

Of 11,874 transcripts in multi-isoform genes (≥ 10 reads in ≥ 6 samples), **1,987** in **1,023 genes** change their share of the gene after PTBP1 knockdown (satuRn, FDR < 0.1). **Blue** = higher usage in WT, **red** = higher usage in KD.



4 Intronic sequences around skipped exons show PTBP1's known binding motif

There are **227 exons in 190 genes** (303 events) that are skipped in WT but kept in KD, like SMAP1 exon 5. We compare their flanking intronic sequence with 227 background exons from unchanged transcripts.



Novel vs annotated transcripts

5 Collapsed detected events reveal where novel transcripts differ from the annotation

Of 4,402 novel transcripts bambu found in the same 11 libraries, **1,069** in **879 genes** pass the filter (≥ 10 reads in ≥ 6 samples); those at annotated genes are compared with that gene's annotated transcripts.

Transcripts	Genes
19,847 annotated	11,387 annotated
1,069 novel	879 with a novel transcript
699 at new loci · dropped	541 new loci
370 at annotated genes · kept	338 annotated genes

Collapsed novel-transcript events by type

Event codes

SE skipped exon

MXE mutually exclusive

RI retained intron

SI spliced intron

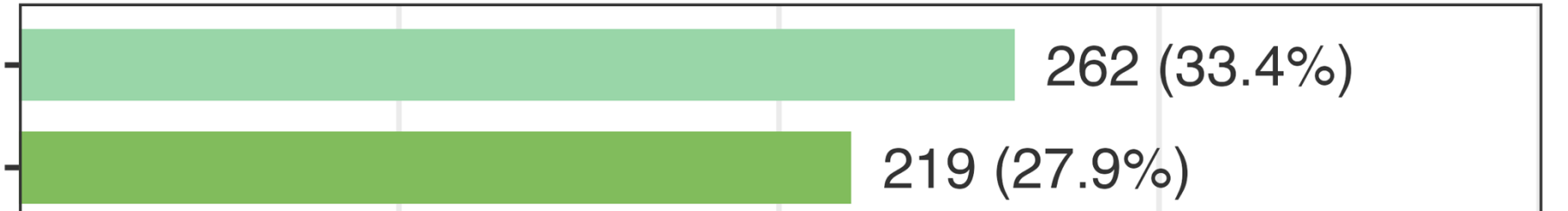
A5SS alt. 5' splice site

A3SS alt. 3' splice site

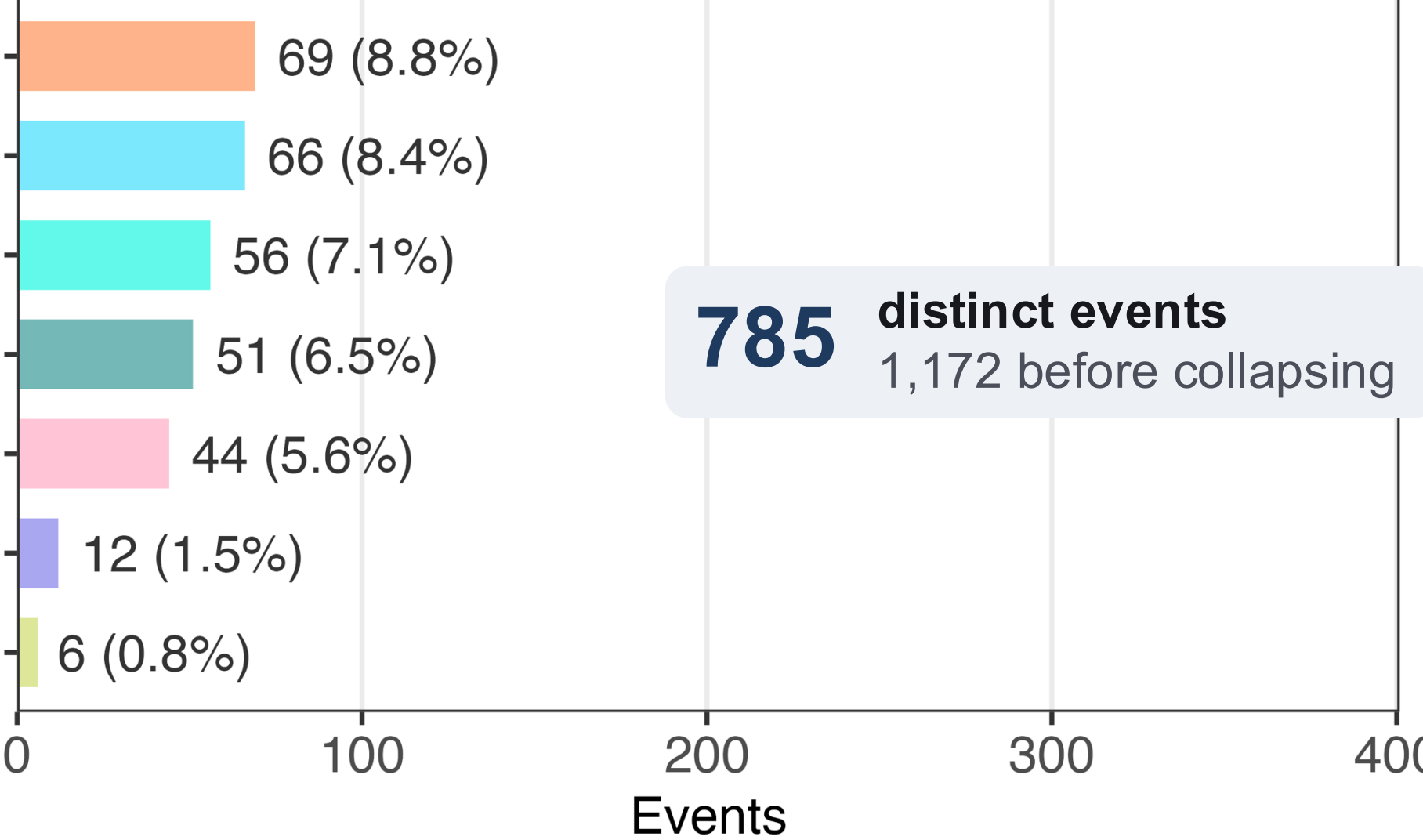
ATSS alt. start

ATES alt. end

Transcript ends: 481 (61.3%)



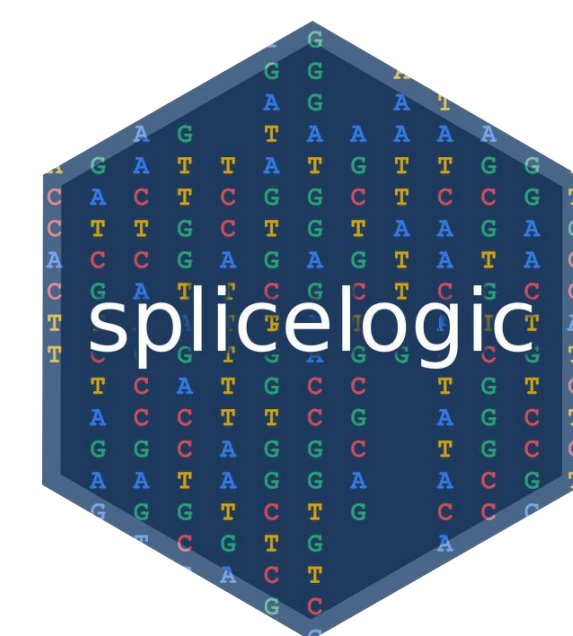
Internal splicing: 304 (38.7%)



785 distinct events
1,172 before collapsing

Conclusions

- splicellogic turns any two transcript sets into named splice events that can go into Bioconductor and tidyomics workflows.
- PTBP1 knockdown shifts the usage of 1,987 transcripts; detected skipped exons carry CU-rich sequence in upstream intronic regions, matching PTBP1's known repressive binding.
- Novel transcripts at known genes mostly differ at their starts and ends.
- Try splicellogic! **FEEDBACK WELCOME**



ANALYSED WITH
splicellogic

Compare any two transcript sets and get named splice events as ranges ready for plotting and sequence analysis.

thelovelab/splicellogic

