

SQANTI3 report

Unique Genes: 248
Unique Isoforms: 1609

Gene classification

Category	# Genes
Annotated Genes	224
Novel Genes	24

Splice Junction Classification

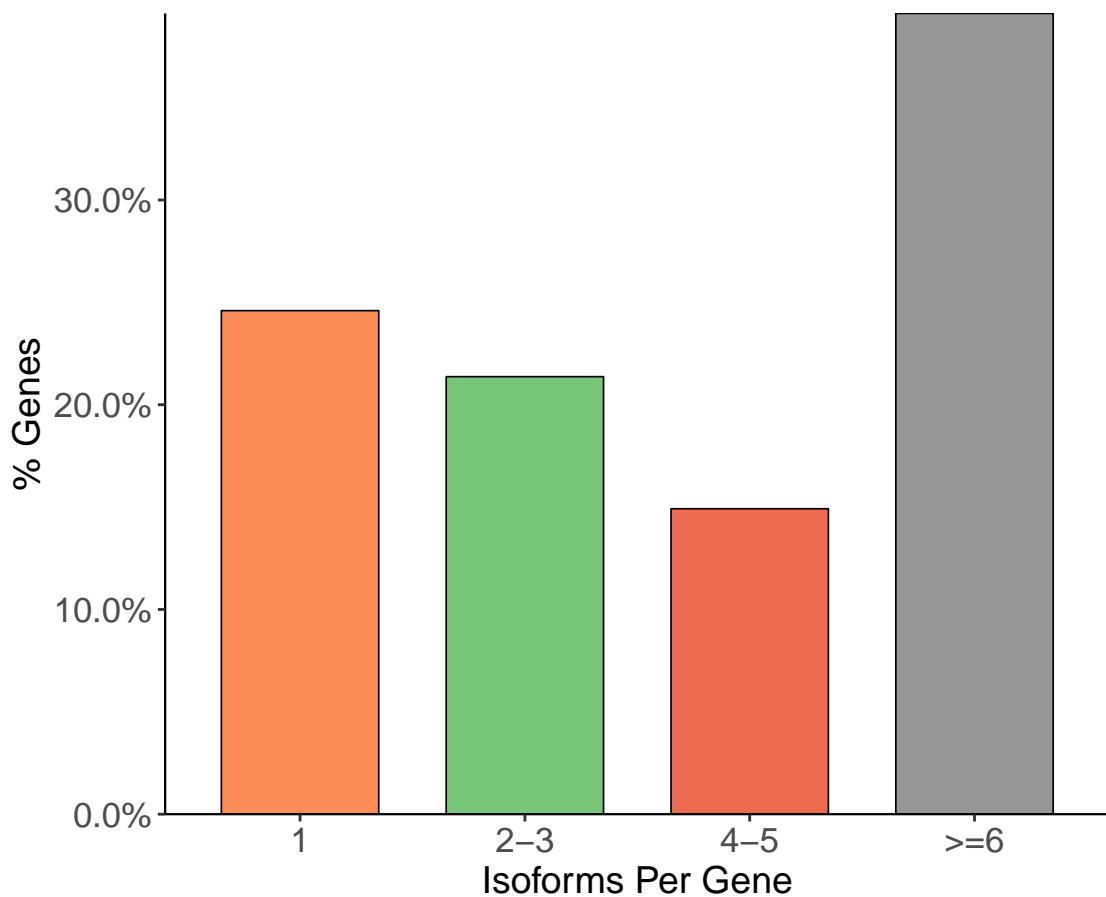
Category	# SJs	Percent
Known canonical	2465	82.86
Known Non-canonical	3	0.10
Novel canonical	443	14.89
Novel Non-canonical	64	2.15

*Characterization of transcripts
based on splice junctions*

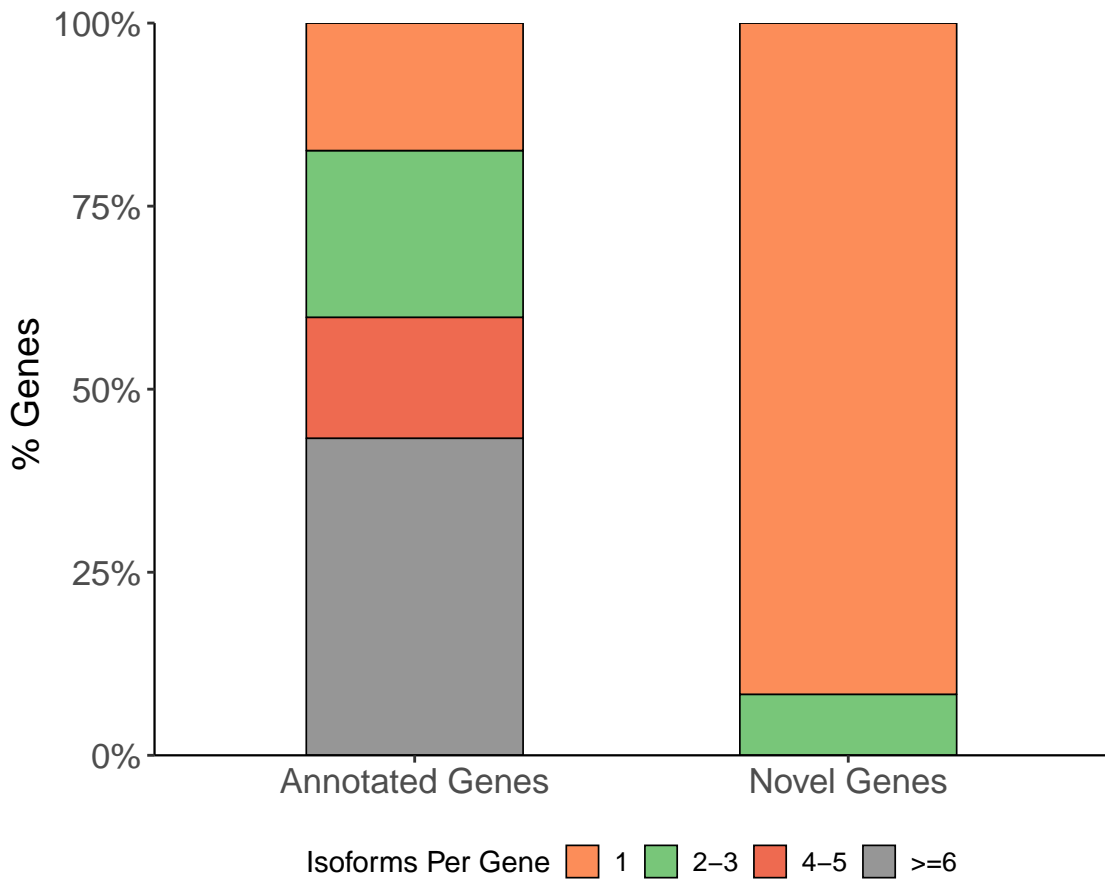
Category	# Isoforms
FSM	576
ISM	156
NIC	439
NNC	395
Genic Genomic	11
Antisense	9
Fusion	6
Intergenic	17
Genic Intron	0

Gene Characterization

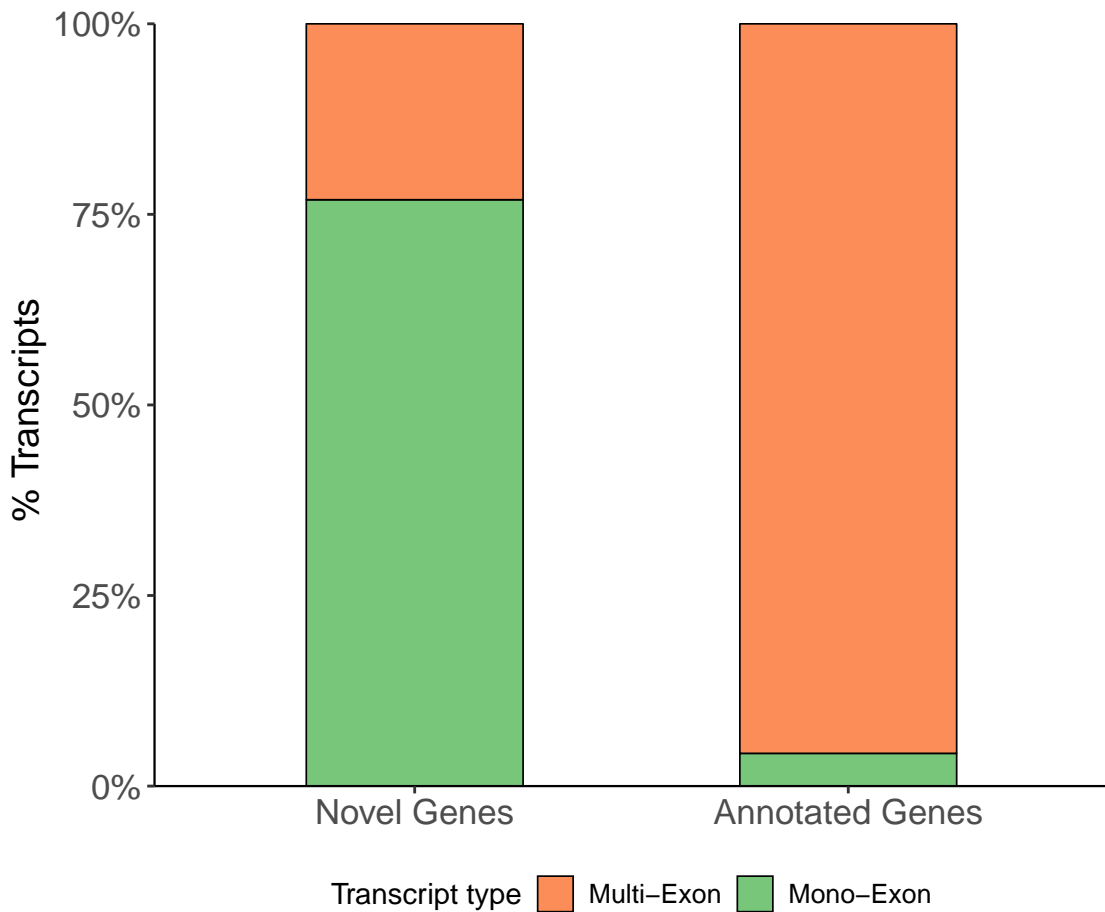
Number of Isoforms per Gene



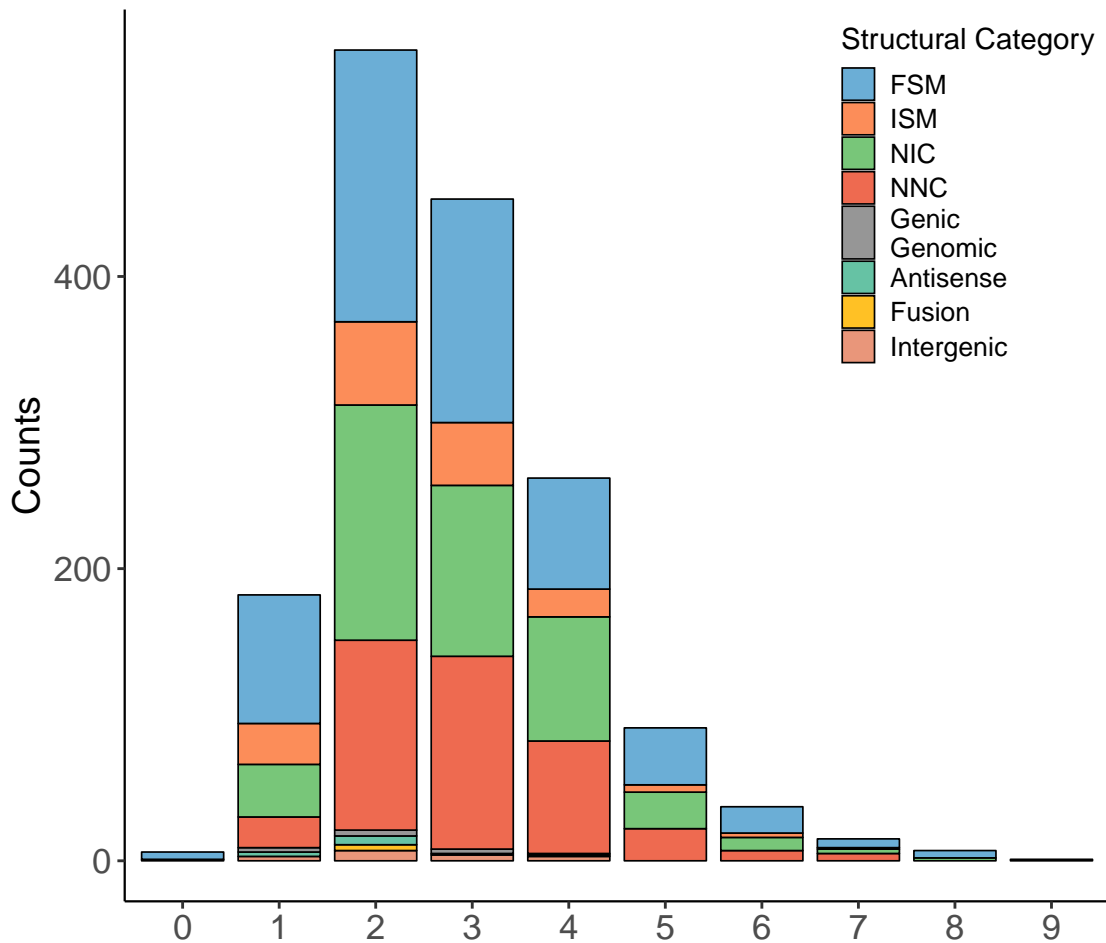
Number of Isoforms per Gene, Known vs Novel Genes



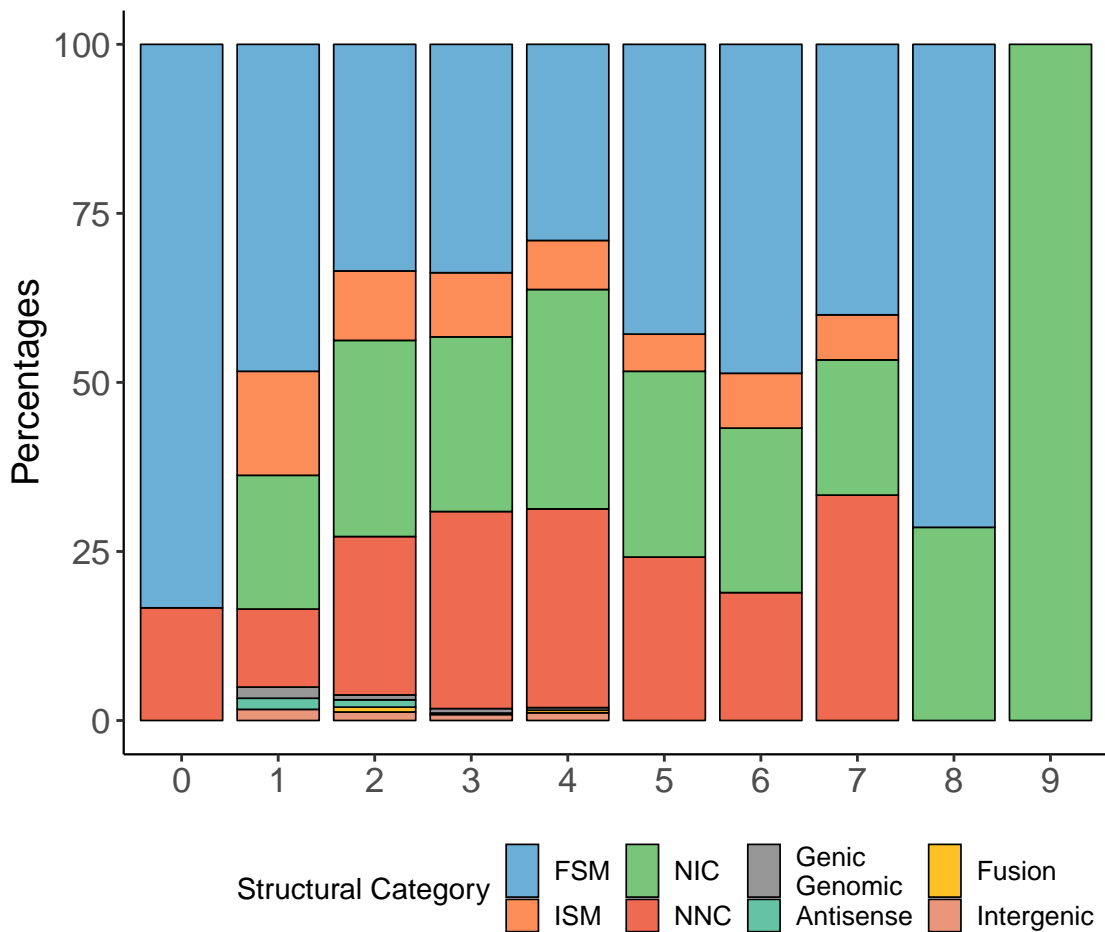
Distribution of Mono- vs Multi-Exon Transcripts



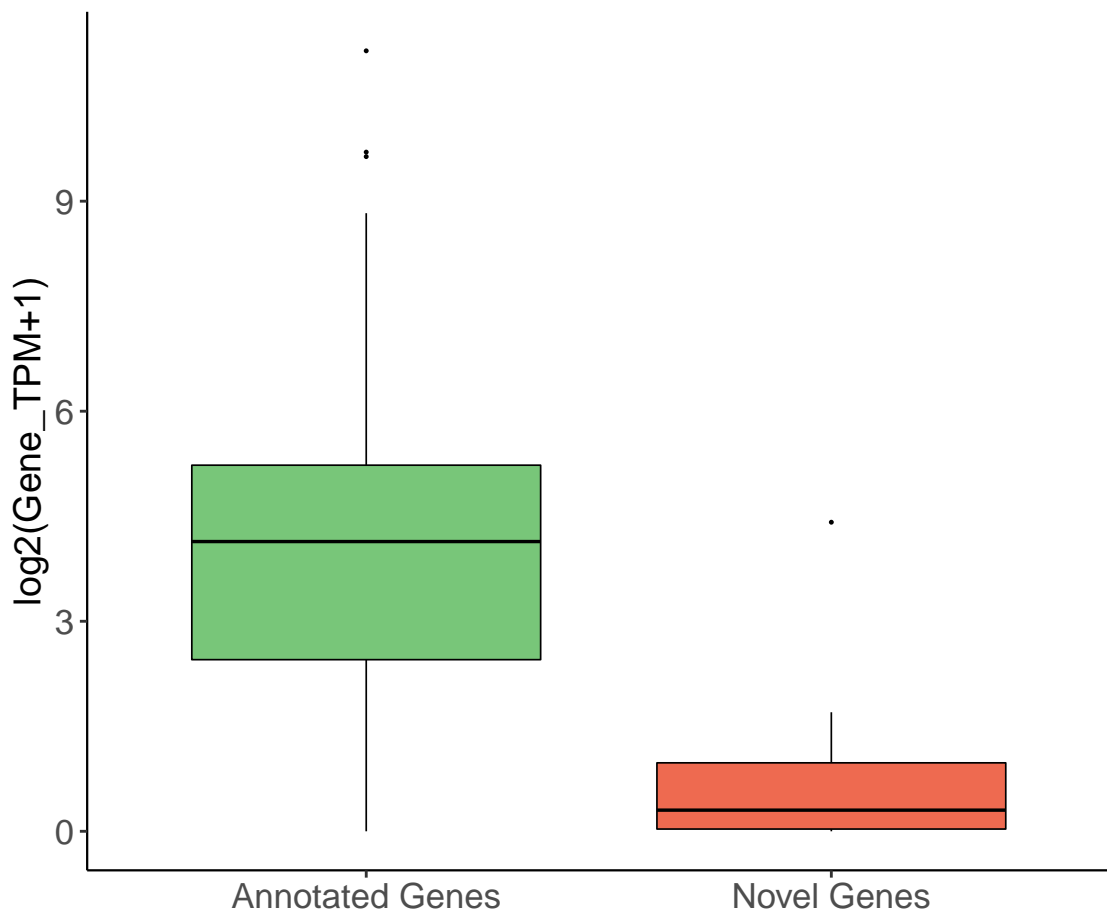
Classifications by Transcript Length



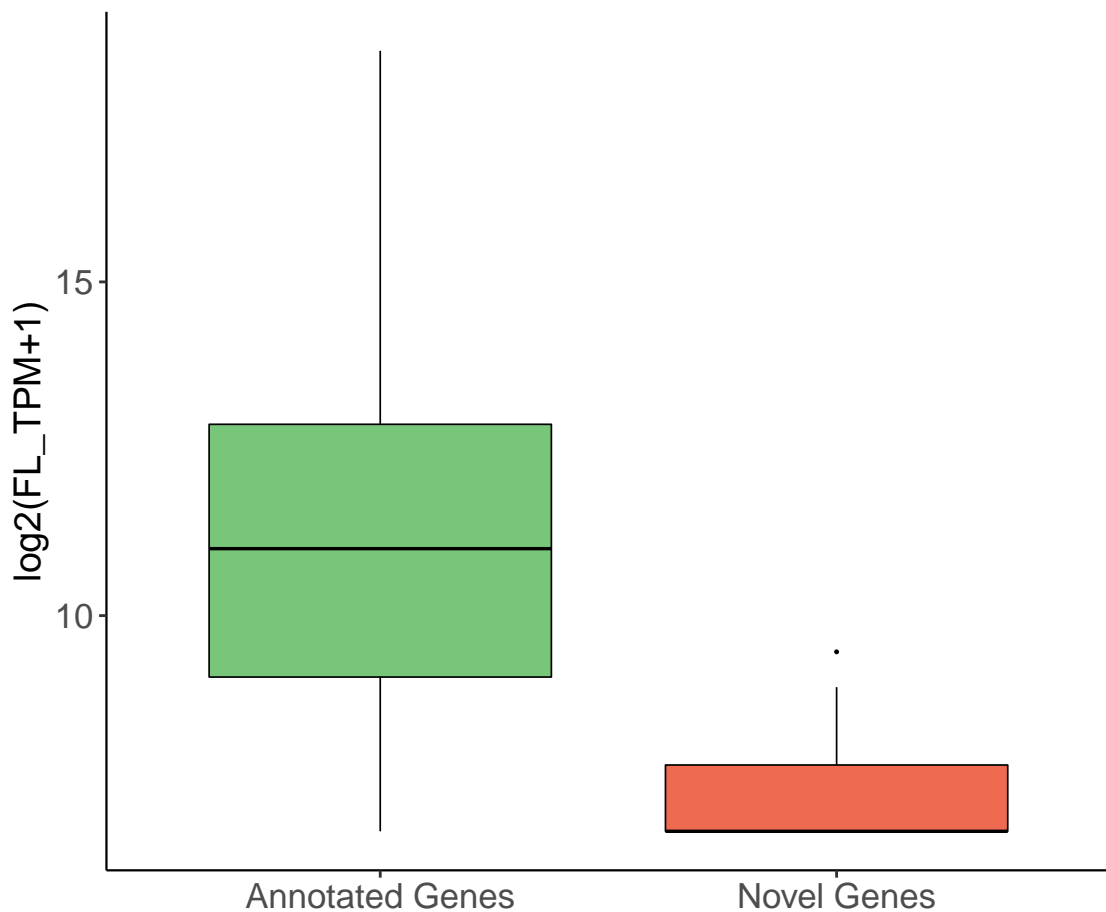
Classifications by Transcript Length, normalized



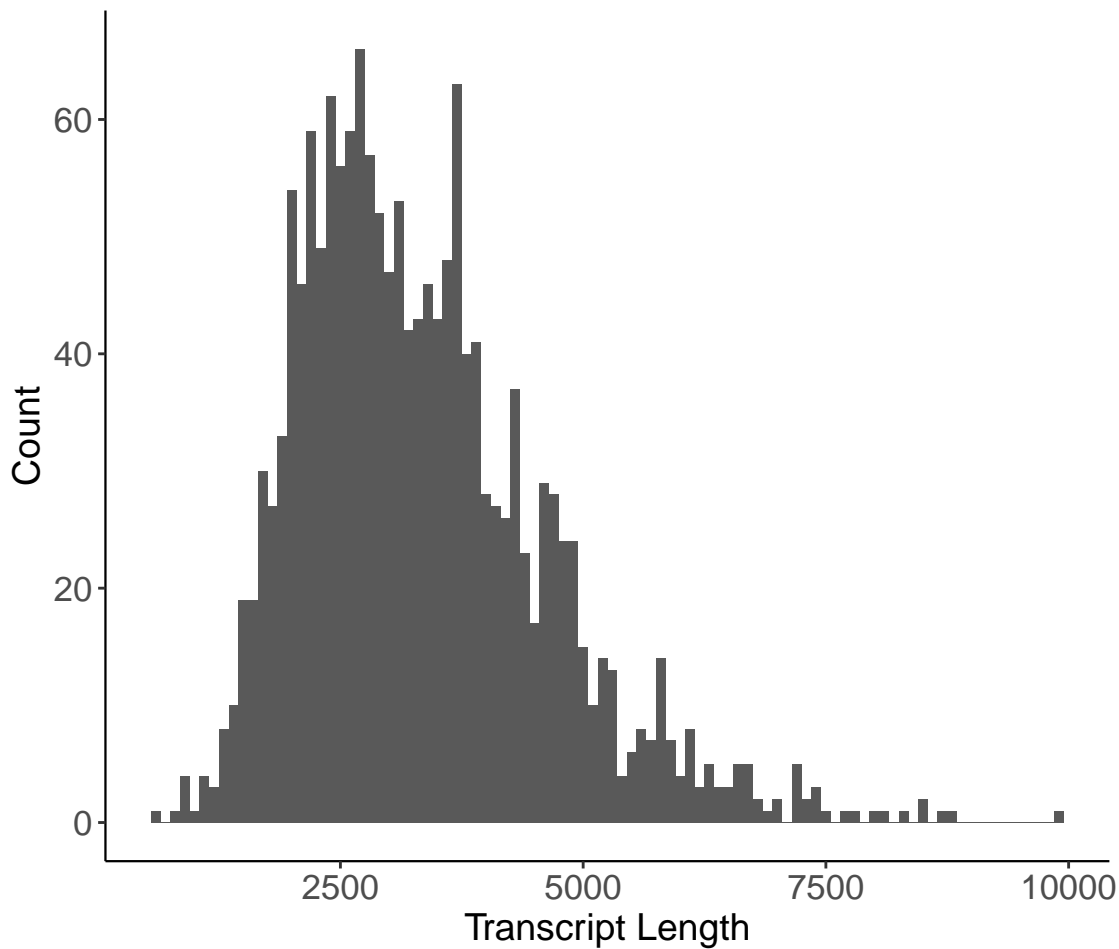
Gene Expression, Annotated vs Novel



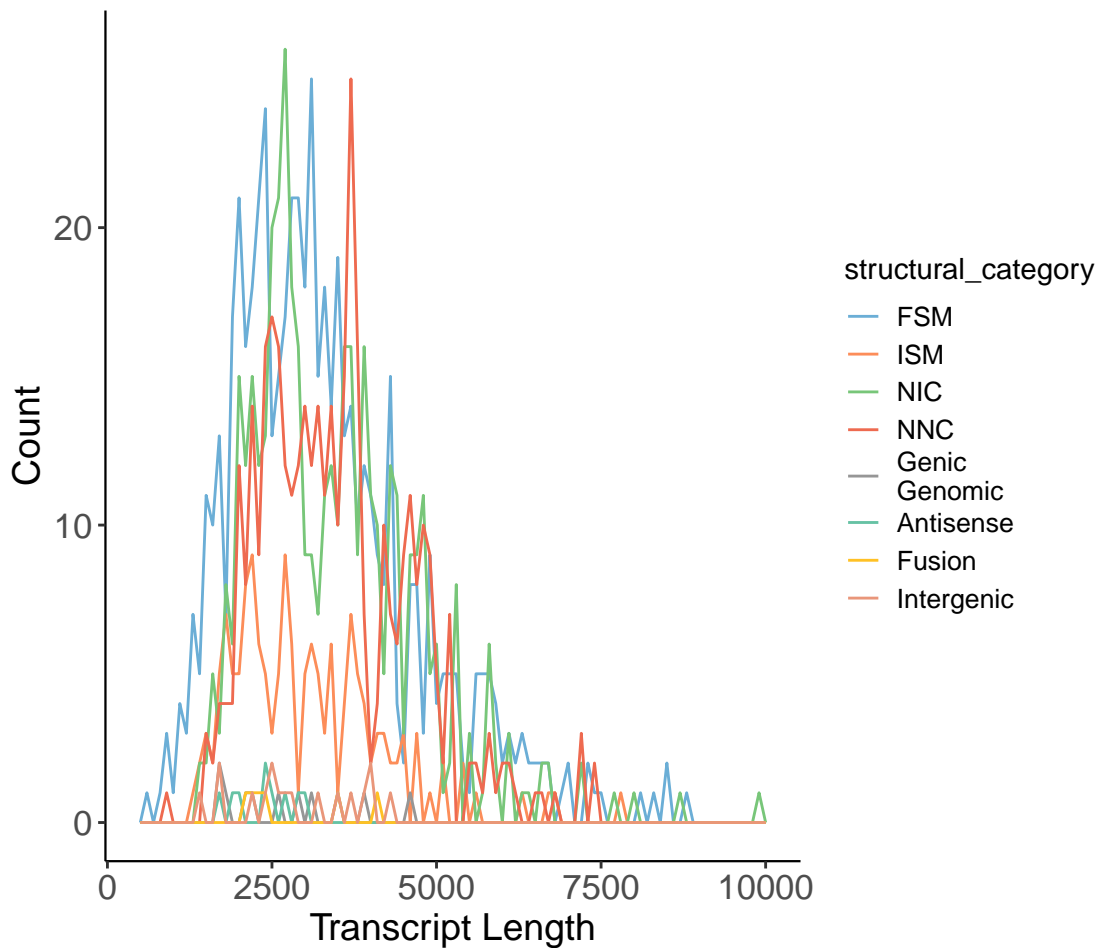
Number of FL reads per Gene by type of gene annotation



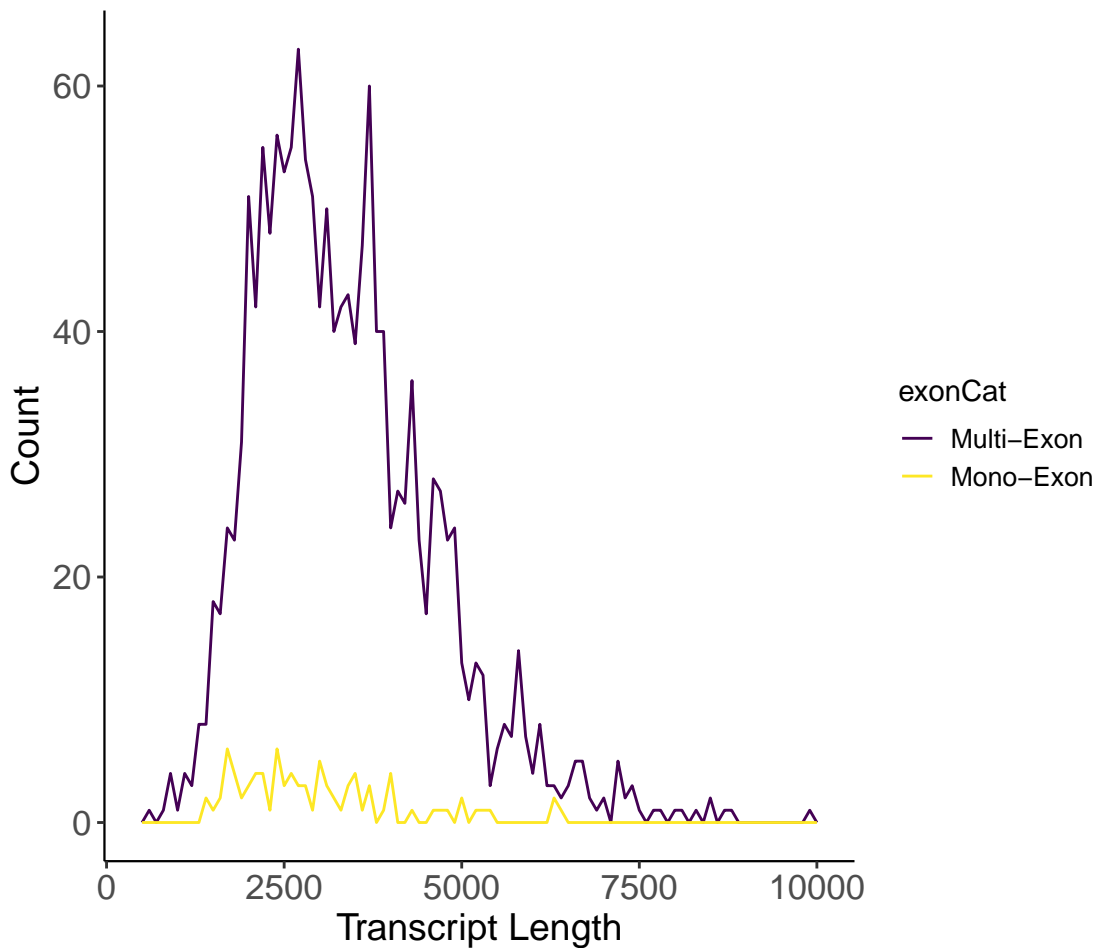
Transcript Lengths, all transcripts



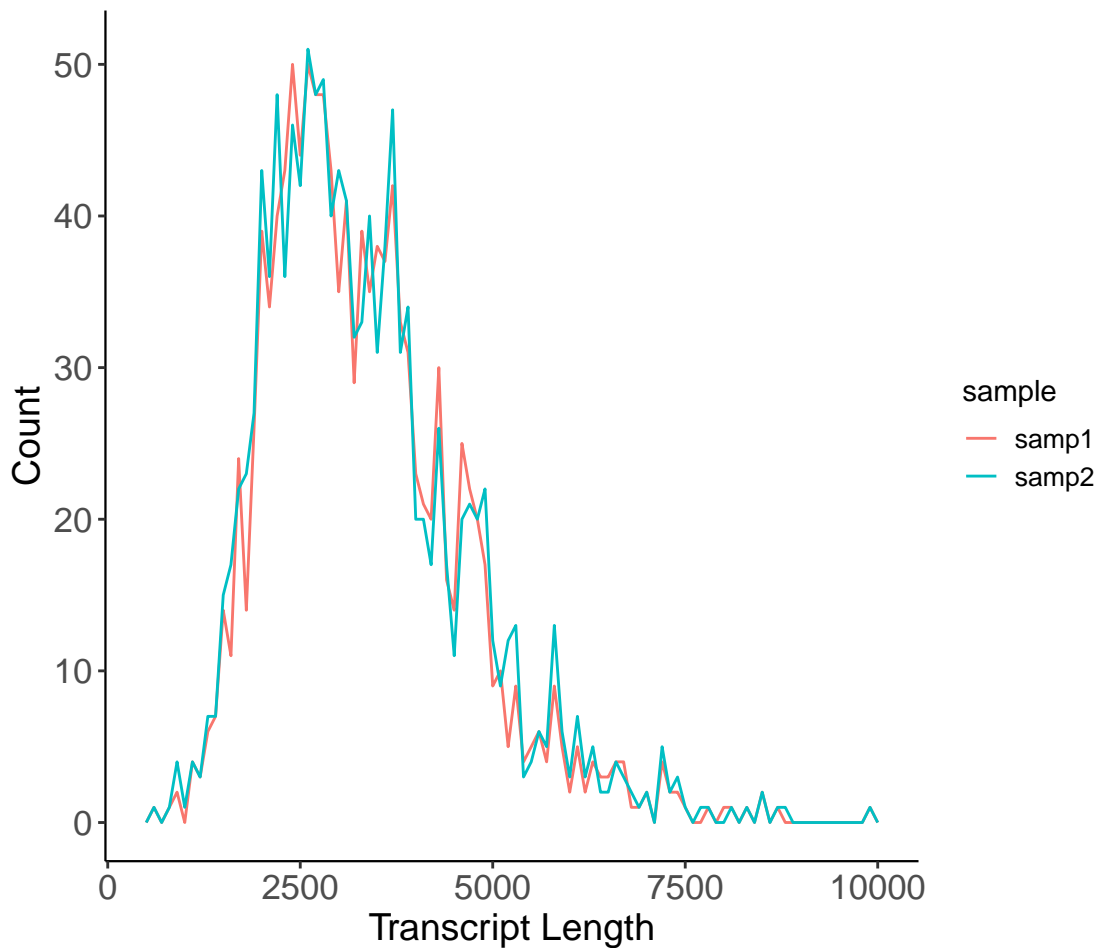
Transcript Lengths, by structural category



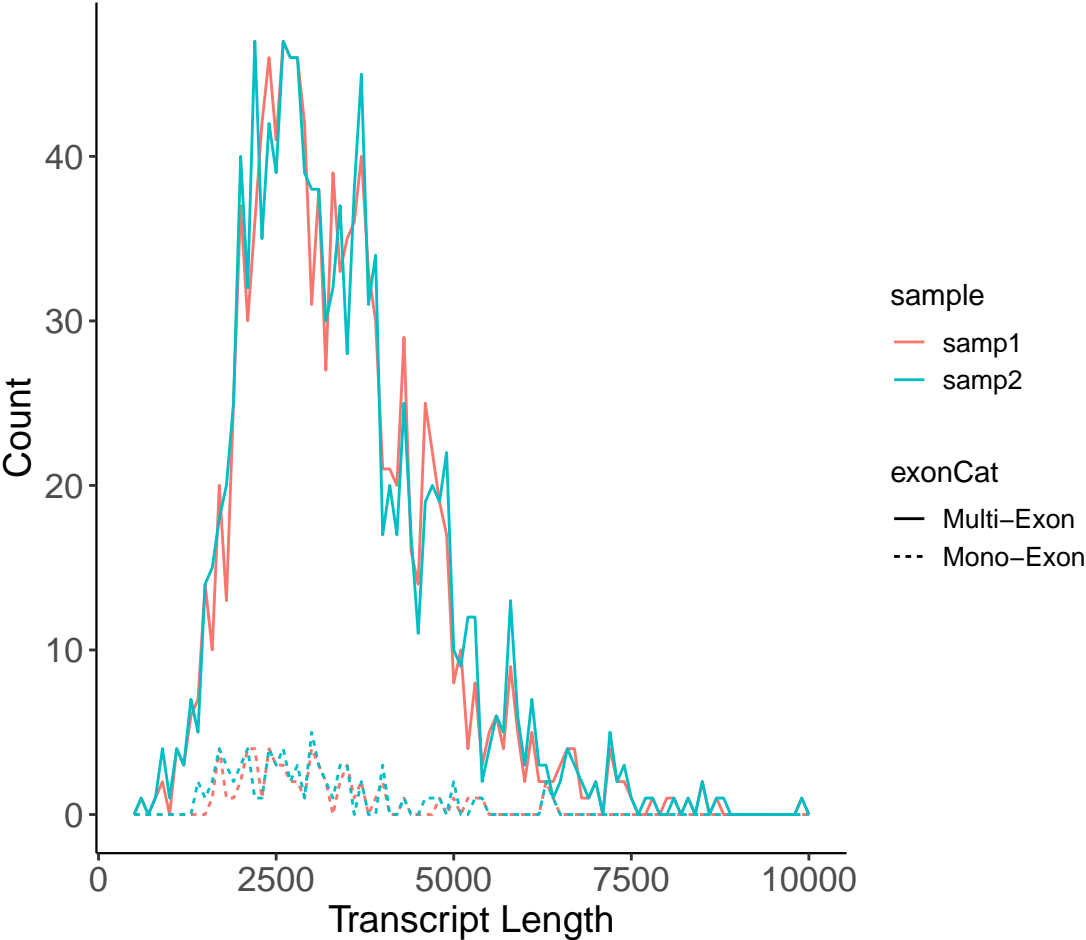
Transcript Lengths, Mono- vs Multi-Exons



Transcript Lengths, By Sample

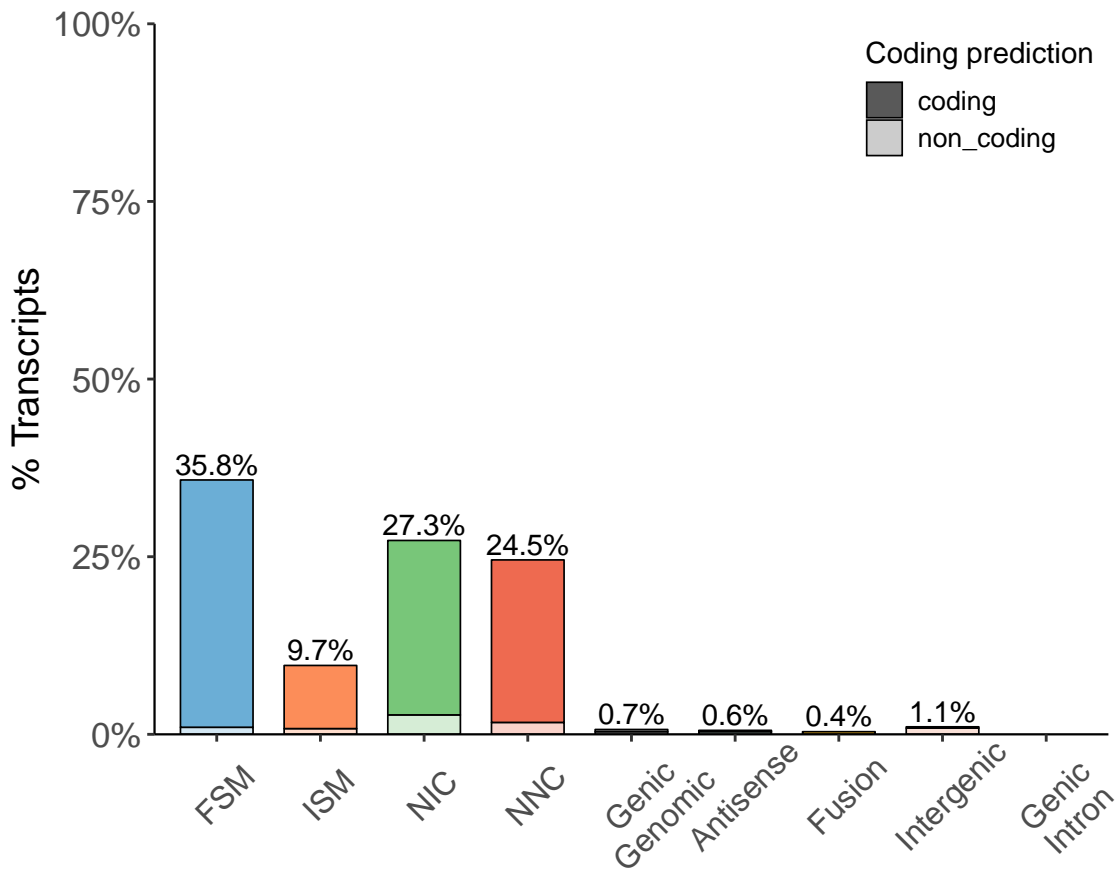


Transcript Lengths, Mono- vs Multi-Exons, By Sample

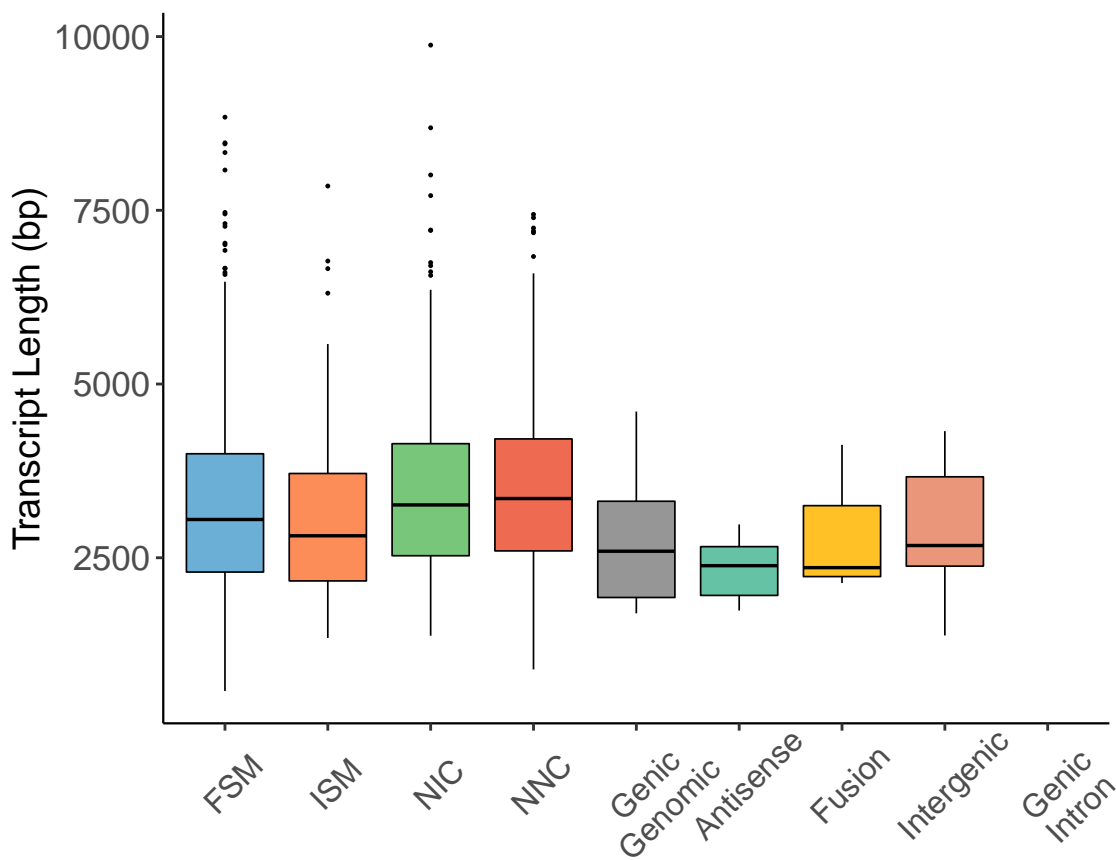


*Structural Isoform Characterization
by Splice Junctions*

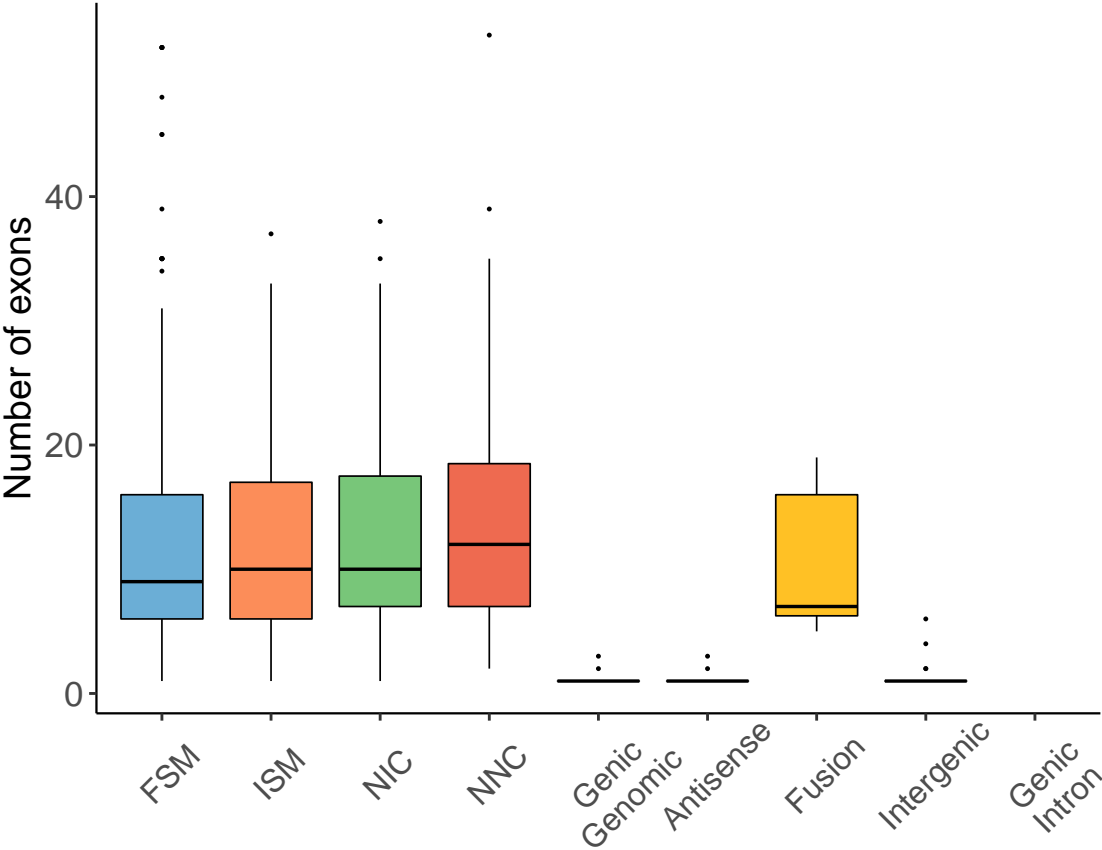
Isoform distribution across structural categories



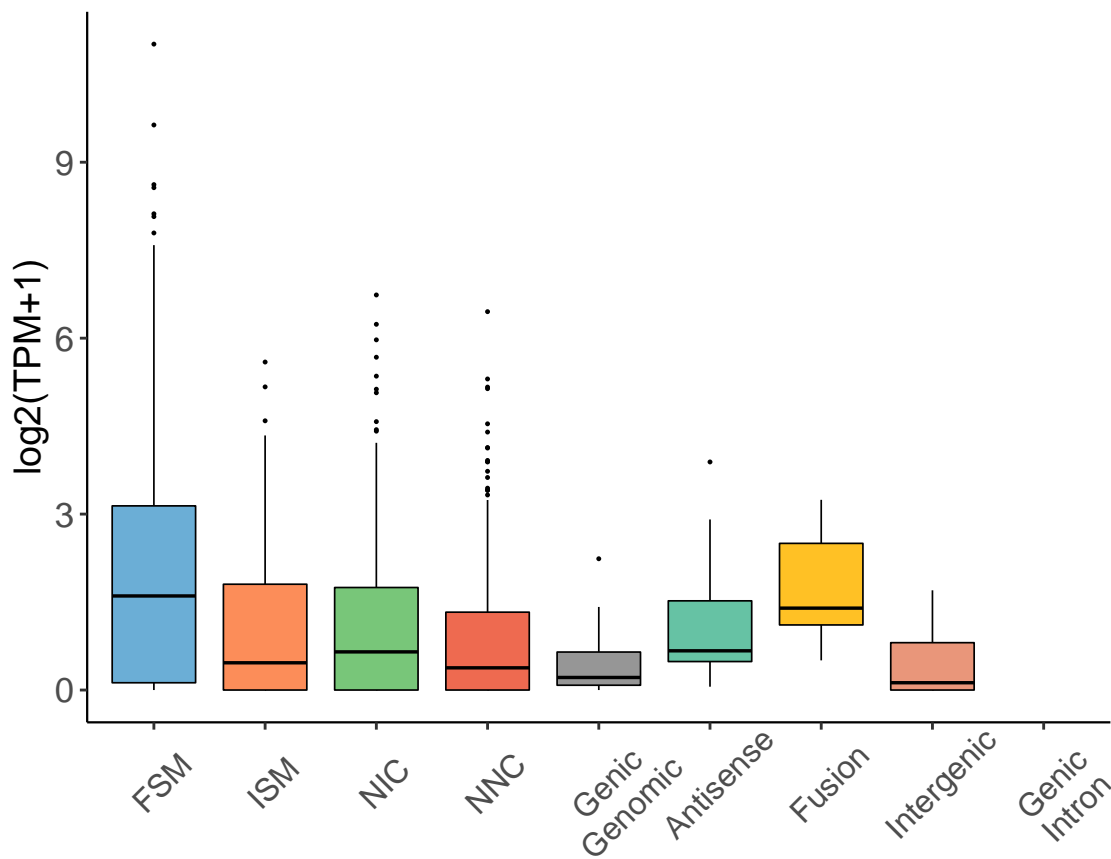
Transcript Lengths by Structural Classification



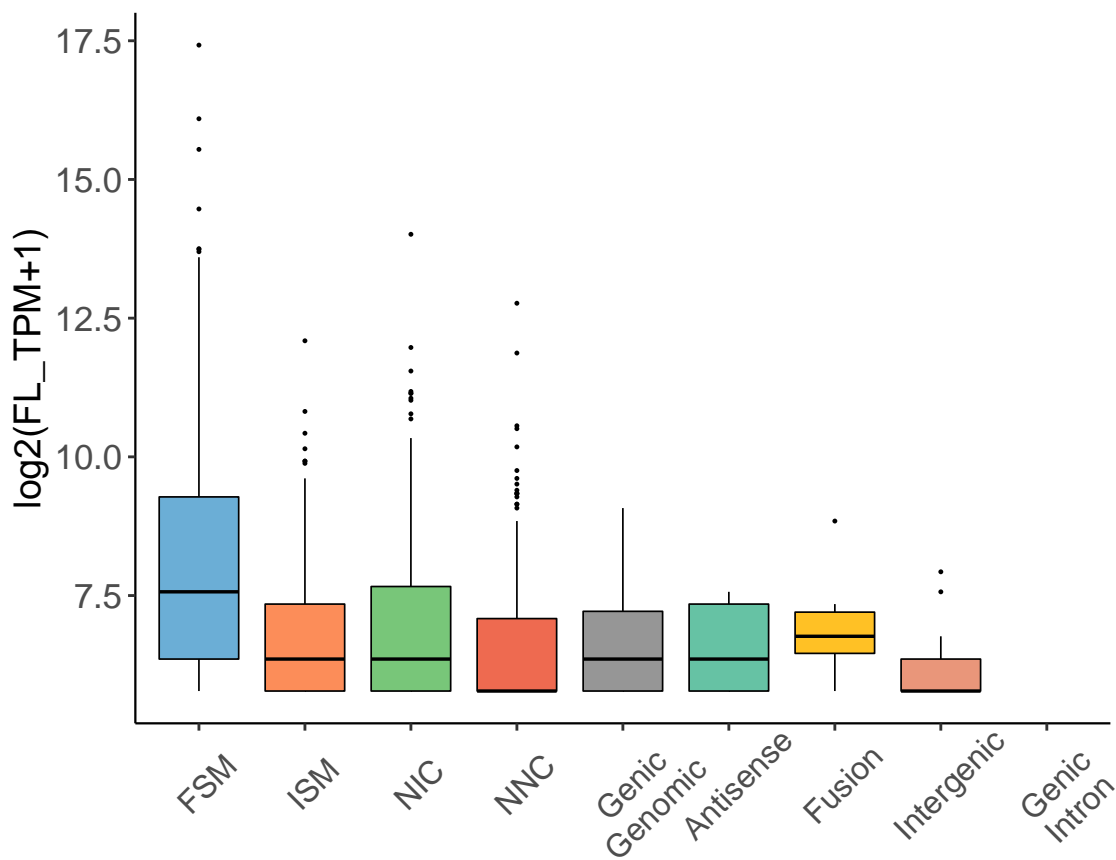
Exon Counts by Structural Classification



Transcript Expression by Structural Category

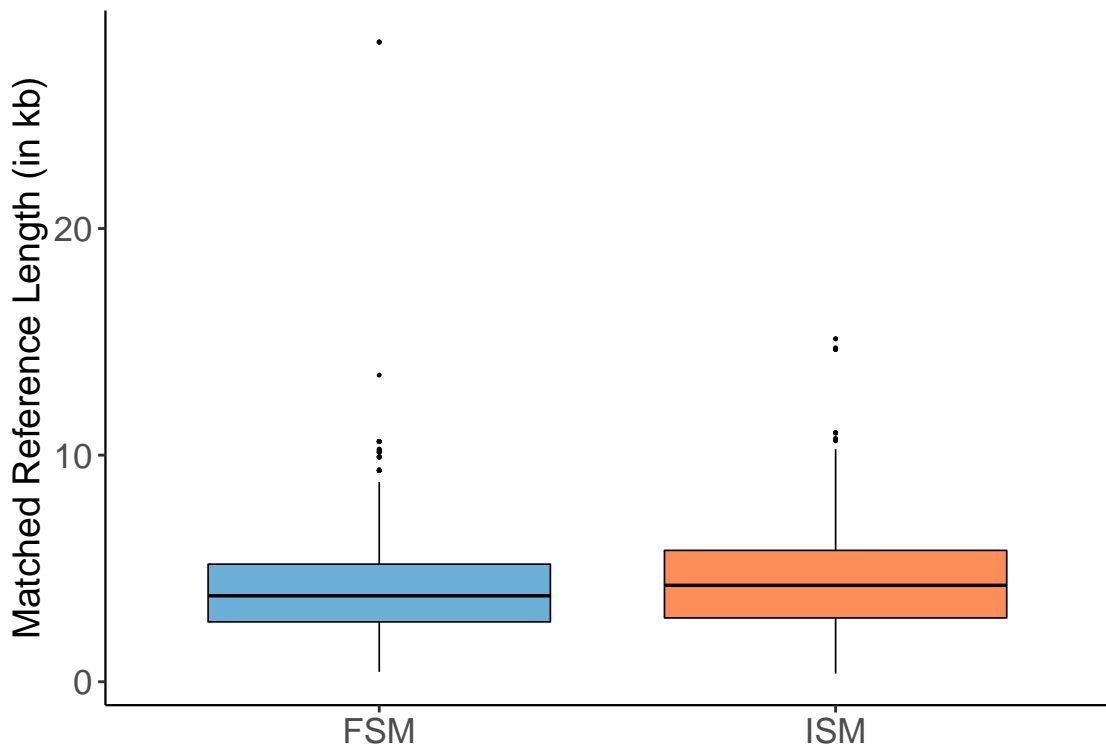


FL Count (normalized) by Structural Category



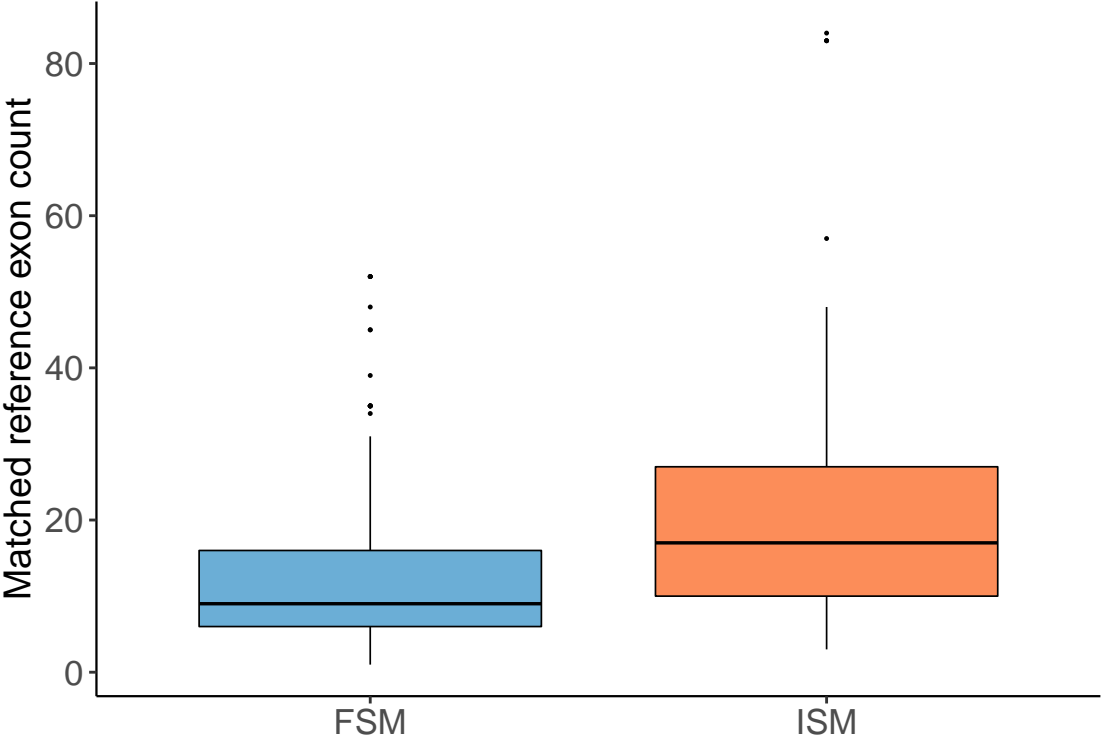
Length Distribution of Matched Reference Transcripts

Applicable only to FSM and ISM categories

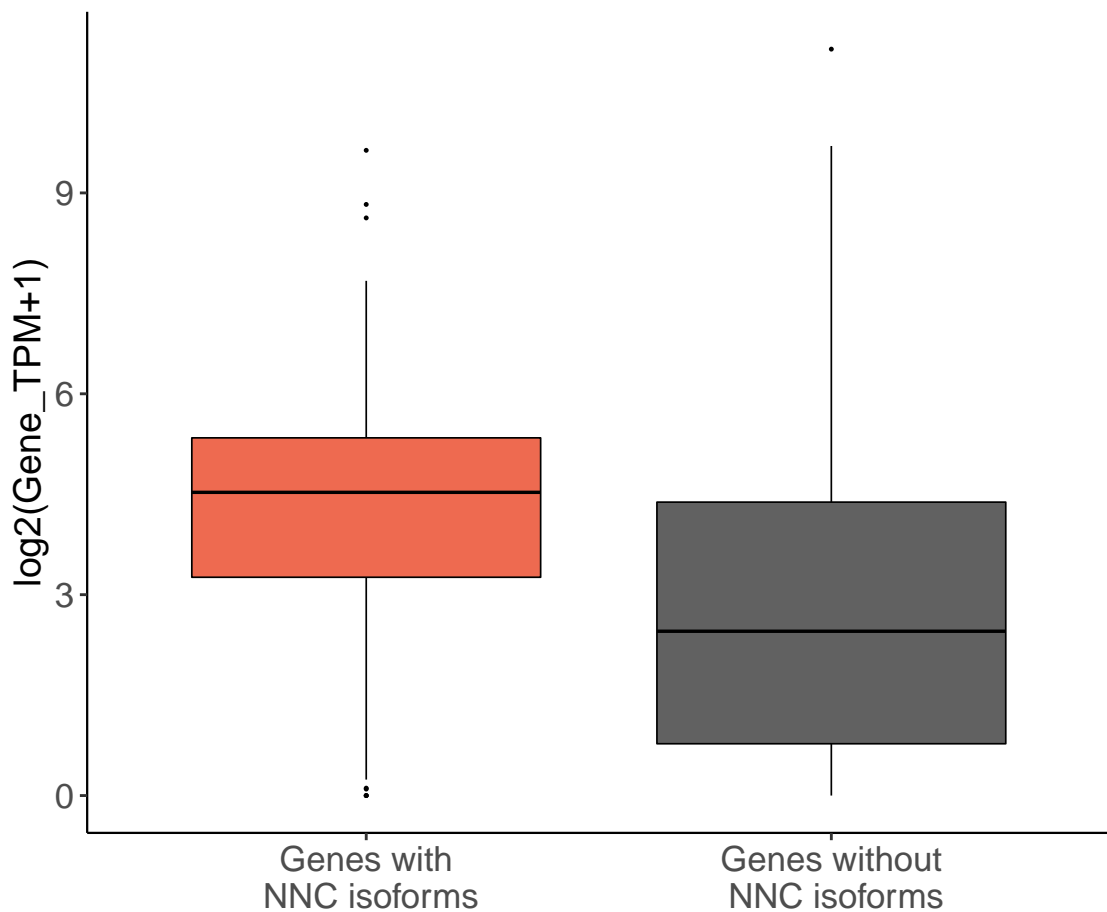


Exon Count Distribution of Matched Reference Transcripts

Applicable only to FSM and ISM categories

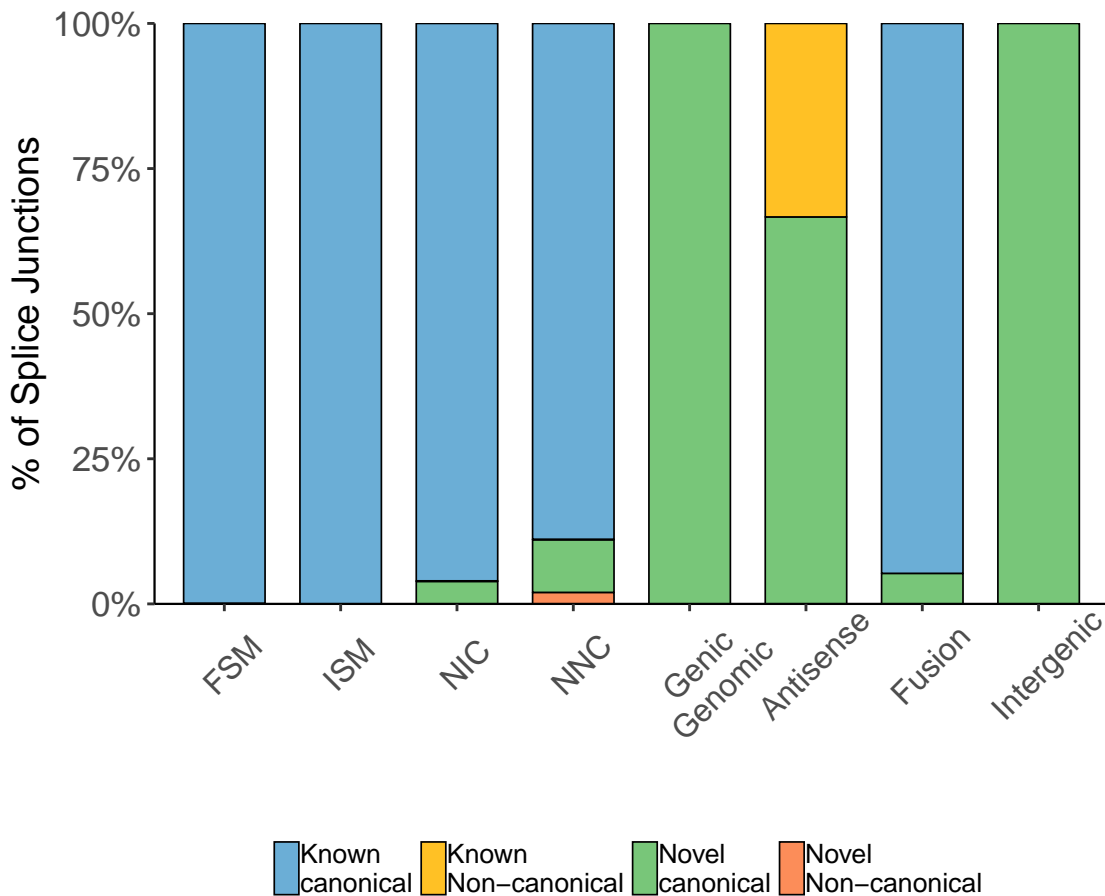


Gene Expression between NNC and not NNC containing Genes

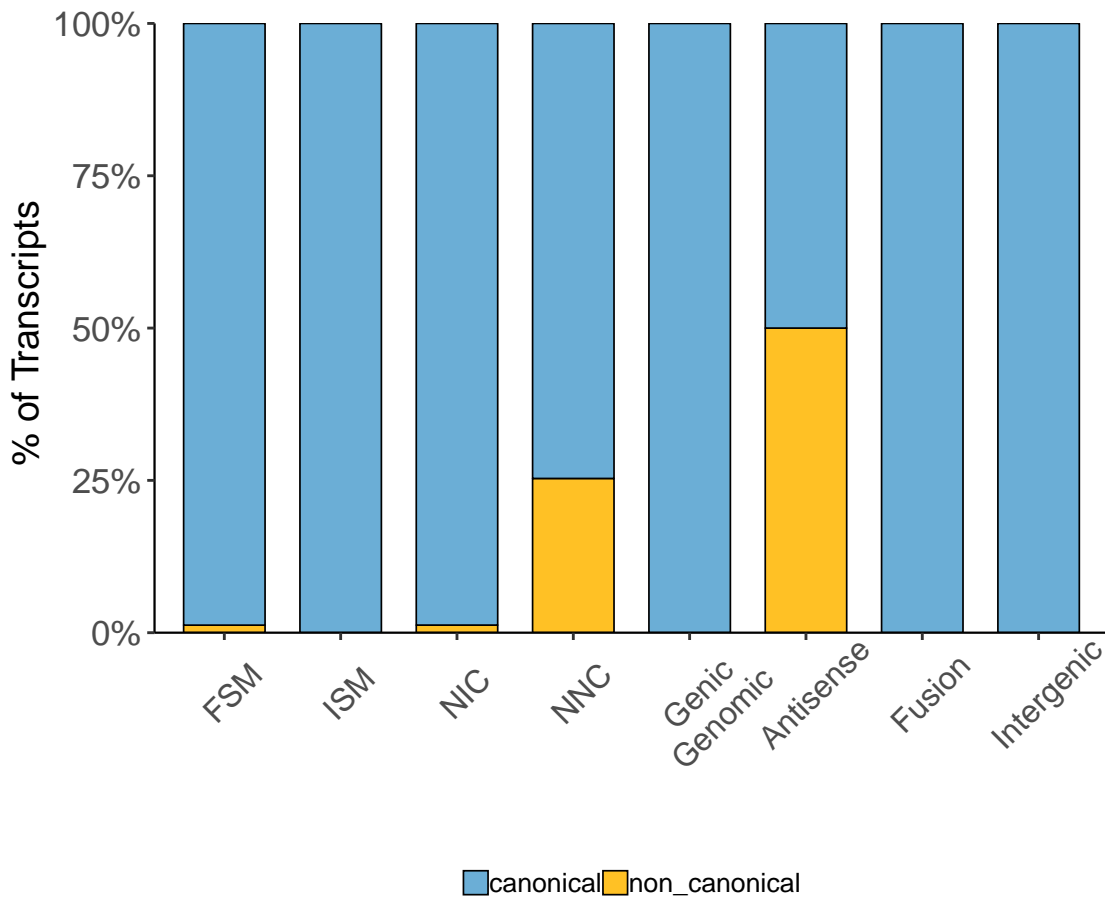


Splice Junction Characterization

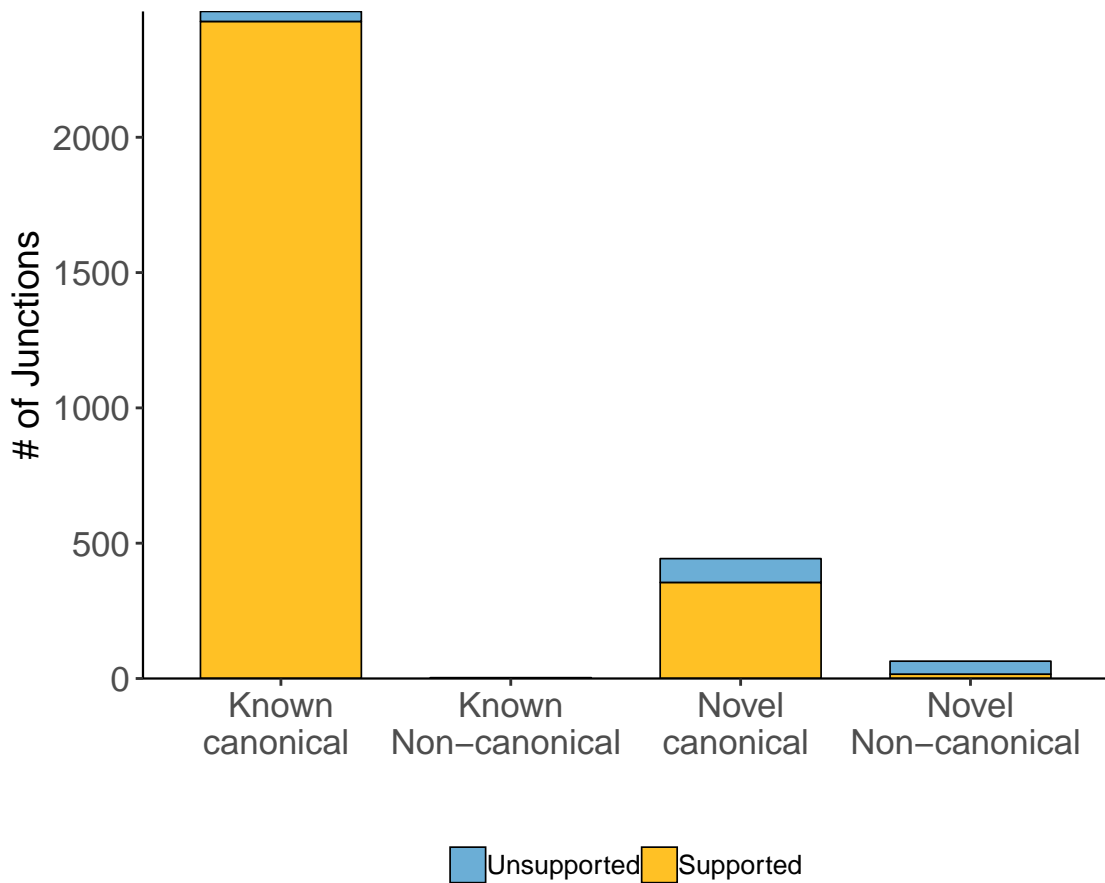
Distribution of Splice Junctions by Structural Classification



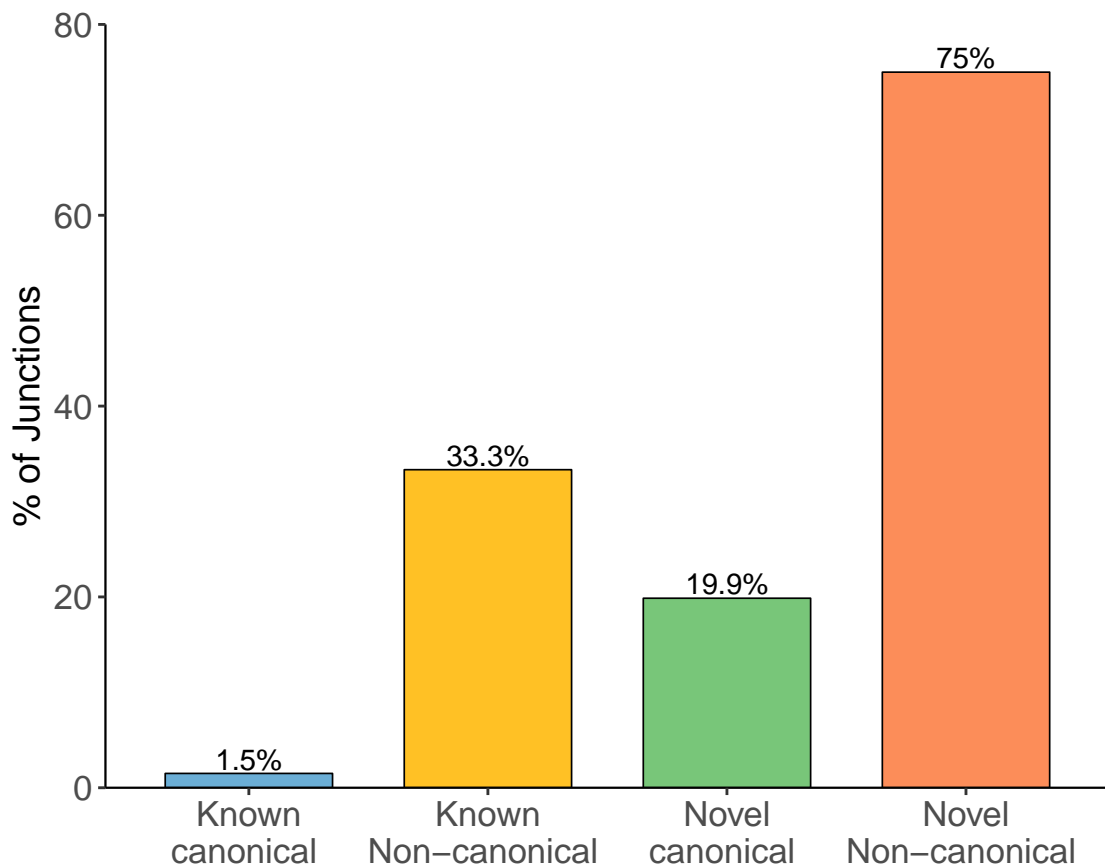
Distribution of Transcripts by Splice Junctions



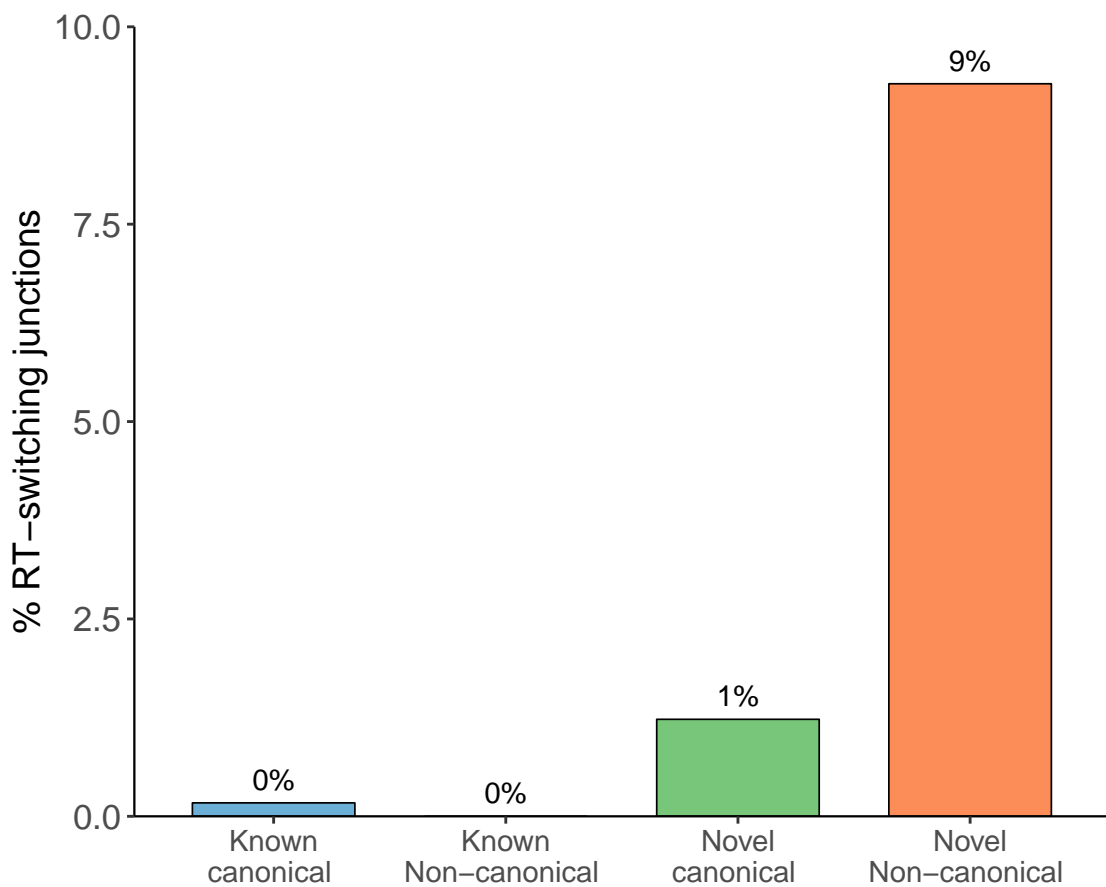
Unique junctions w/ or w/out short read coverage



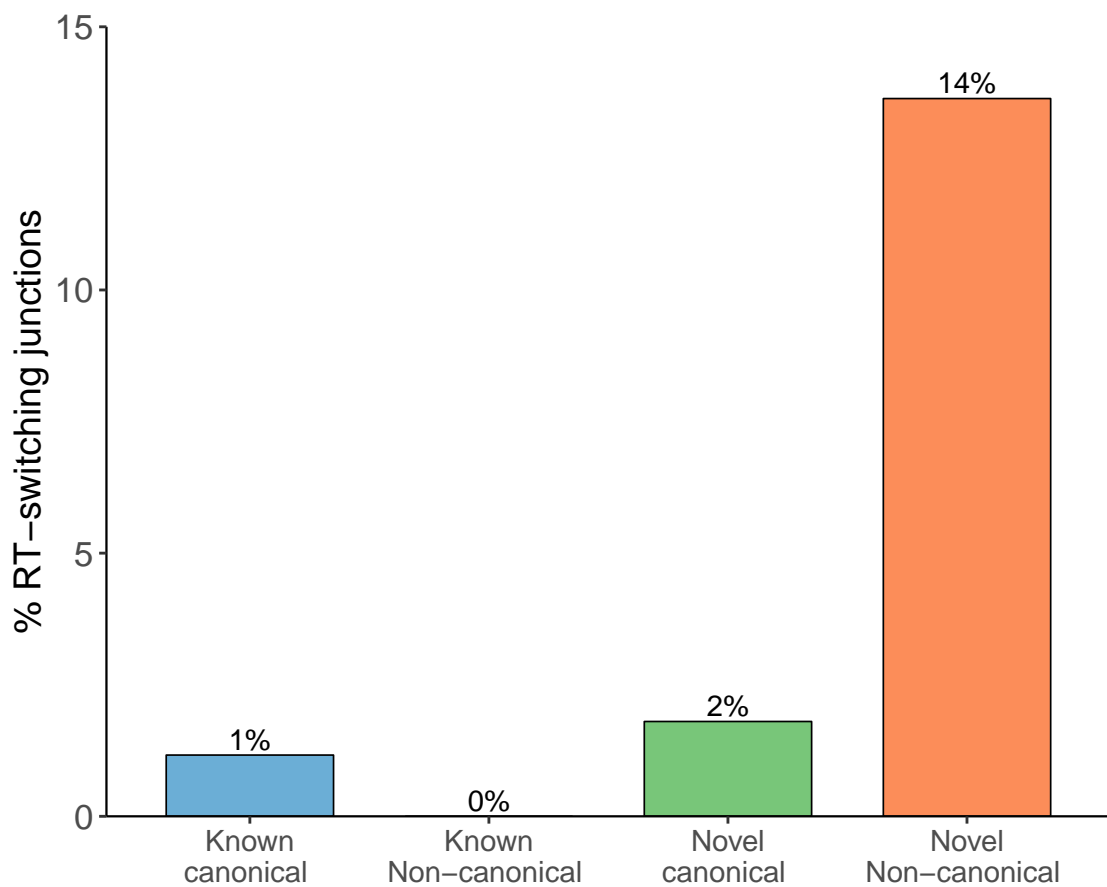
Unique junctions w/out short read coverage (percentage)



RT-switching, all junctions



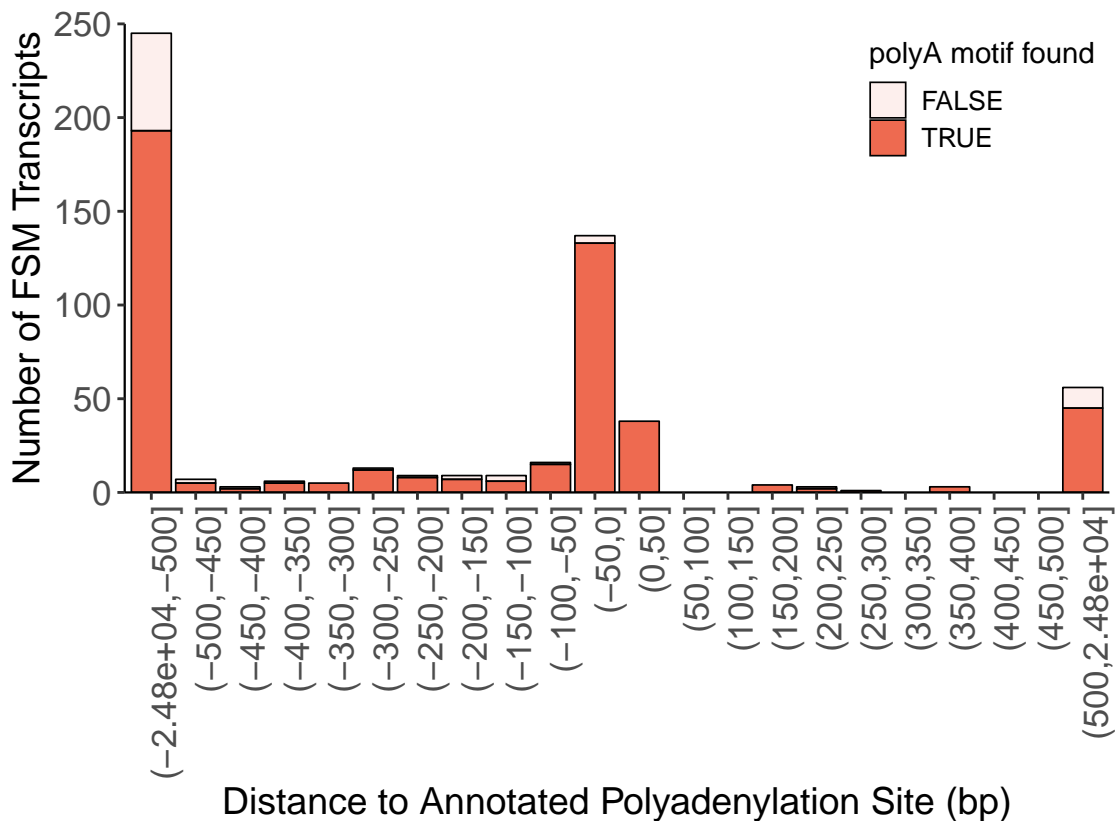
RT-switching, unique junctions



Comparison with Annotated TSS and PolyA Sites

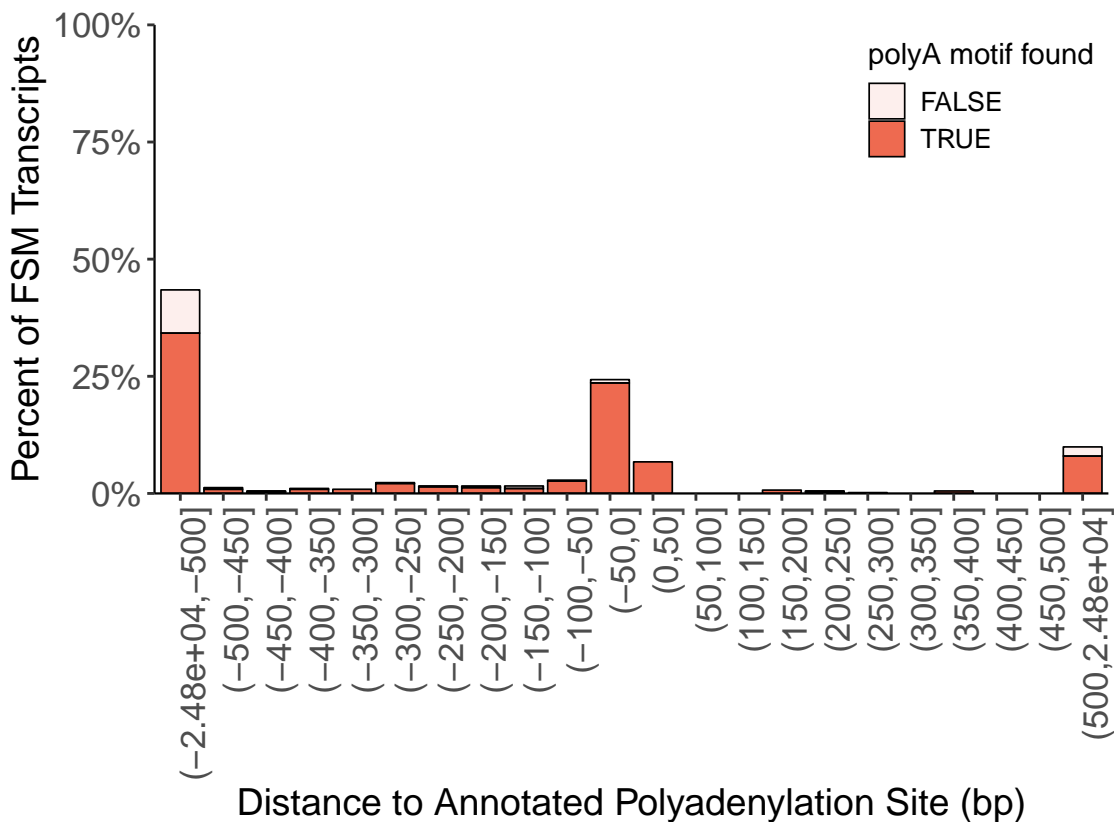
Distance to Annotated Polyadenylation Site, FSM only

Negative values indicate upstream of annotated polyA site



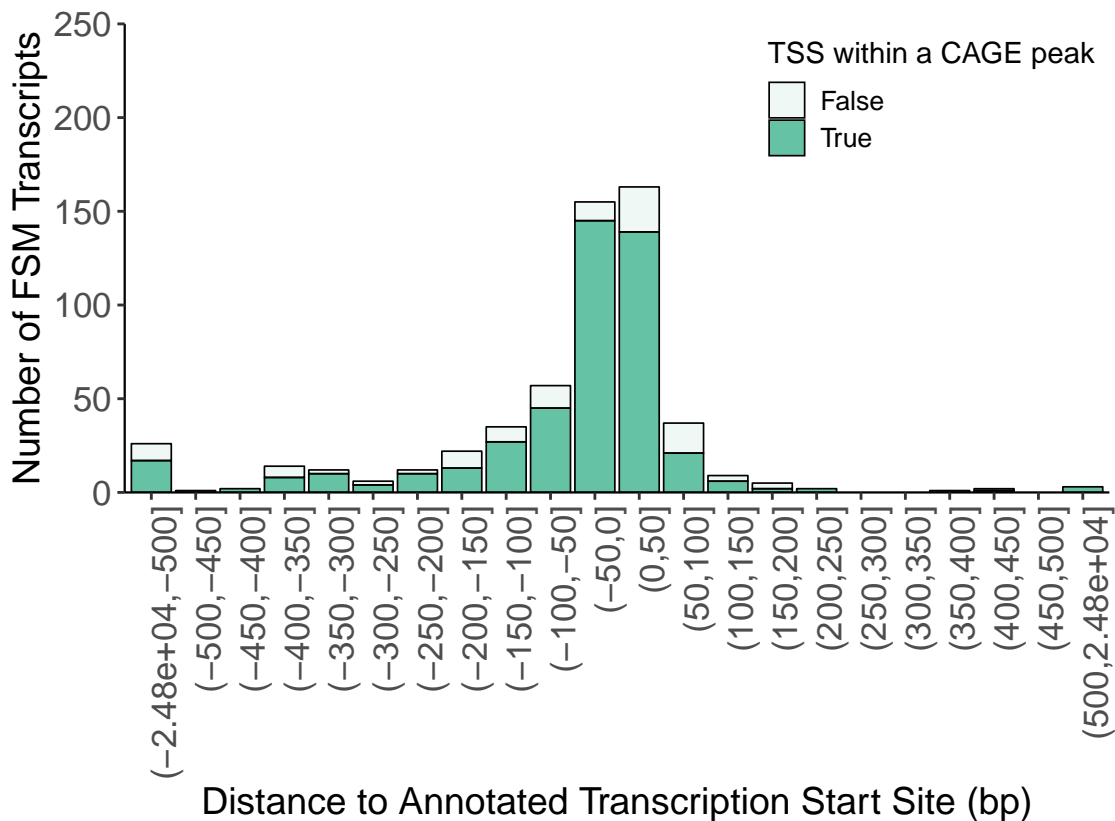
Distance to Annotated Polyadenylation Site, FSM only

Negative values indicate upstream of annotated polyA site



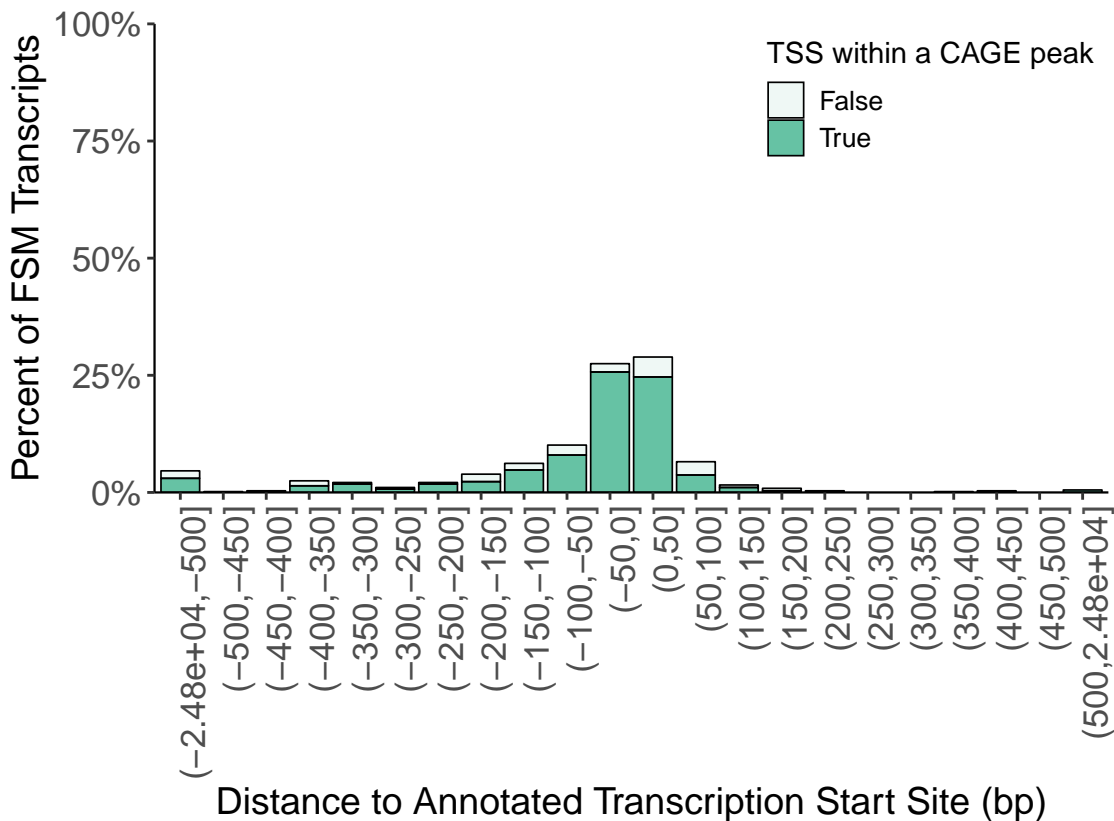
Distance to Annotated Transcription Start Site, FSM only

Negative values indicate downstream of annotated TSS



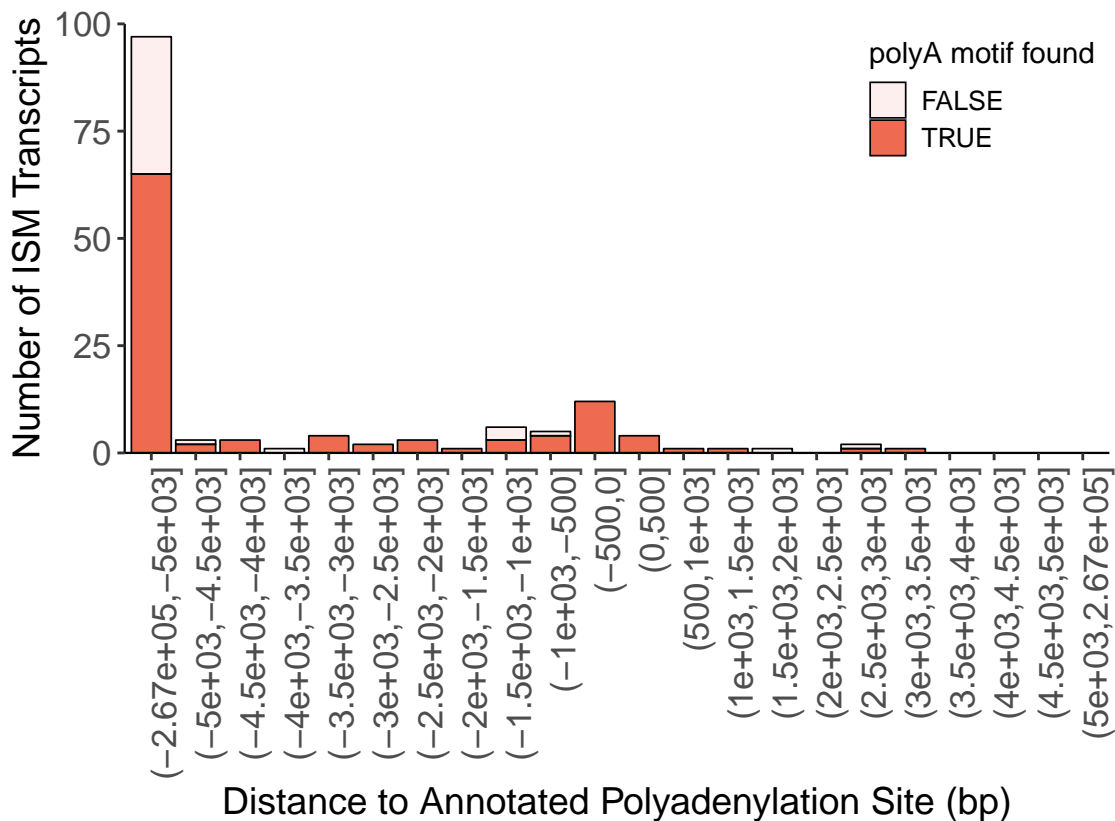
Distance to Annotated Transcription Start Site, FSM only

Negative values indicate downstream of annotated TSS



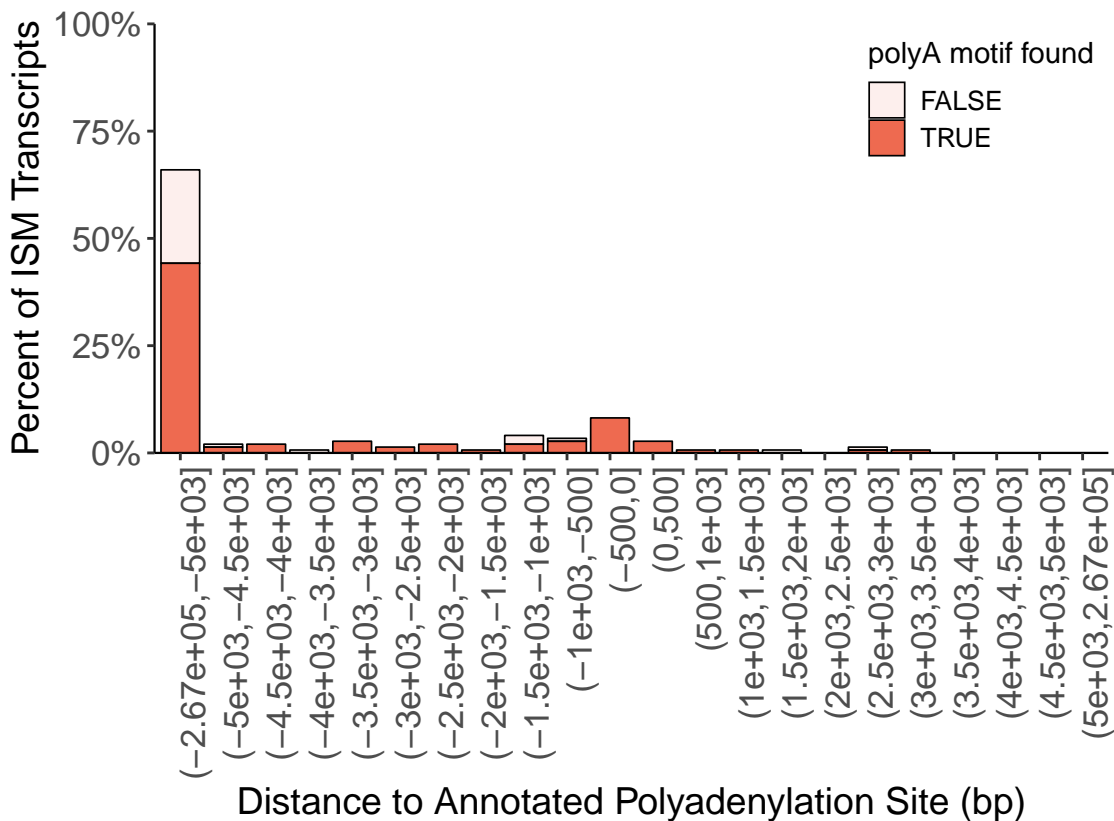
Distance to Annotated Polyadenylation Site, ISM only

Negative values indicate upstream of annotated polyA site



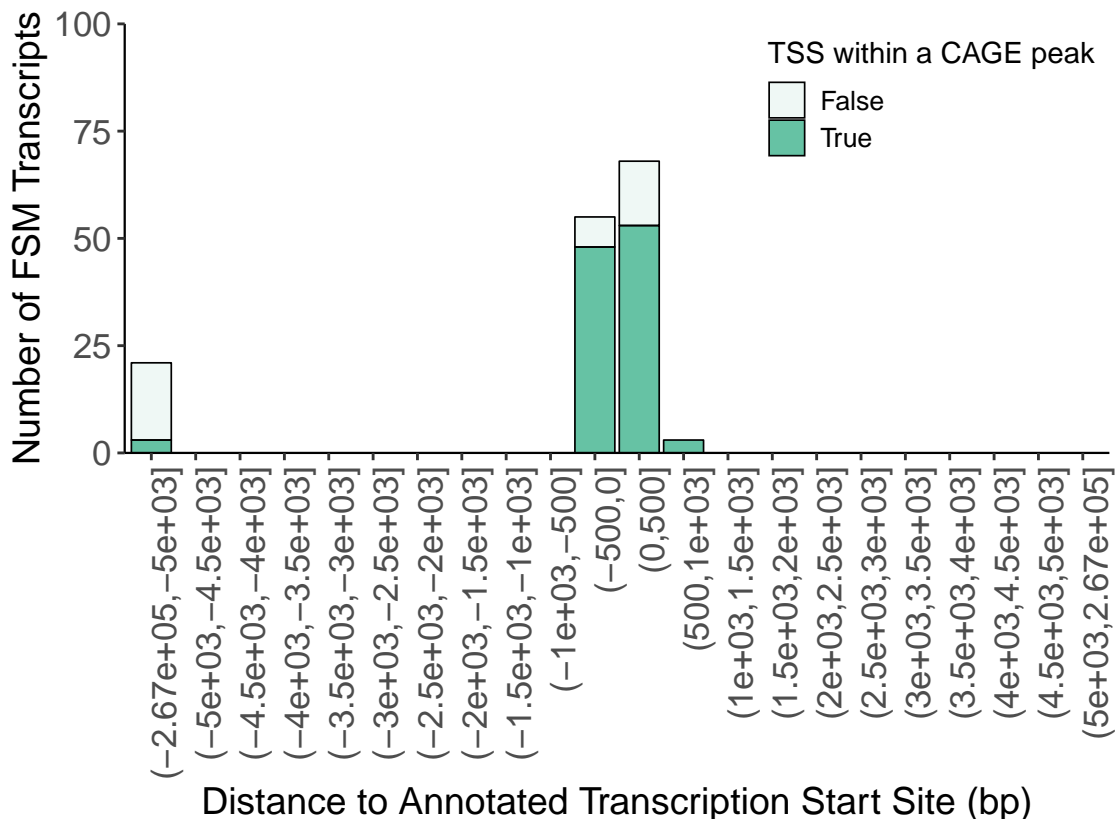
Distance to Annotated Polyadenylation Site, ISM only

Negative values indicate upstream of annotated polyA site



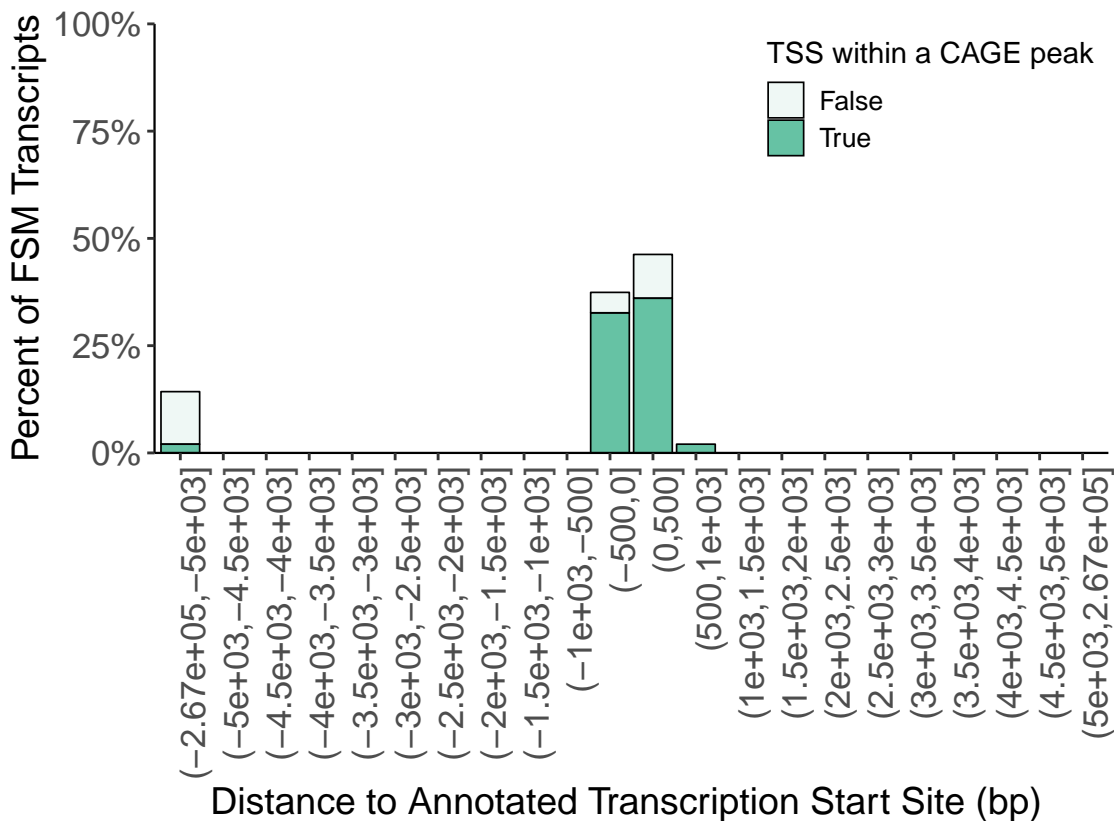
Distance to Annotated Transcription Start Site, ISM only

Negative values indicate downstream of annotated TSS



Distance to Annotated Transcription Start Site, ISM only

Negative values indicate downstream of annotated TSS



Frequency of polyA motifs

Number of polyA Motifs Detected

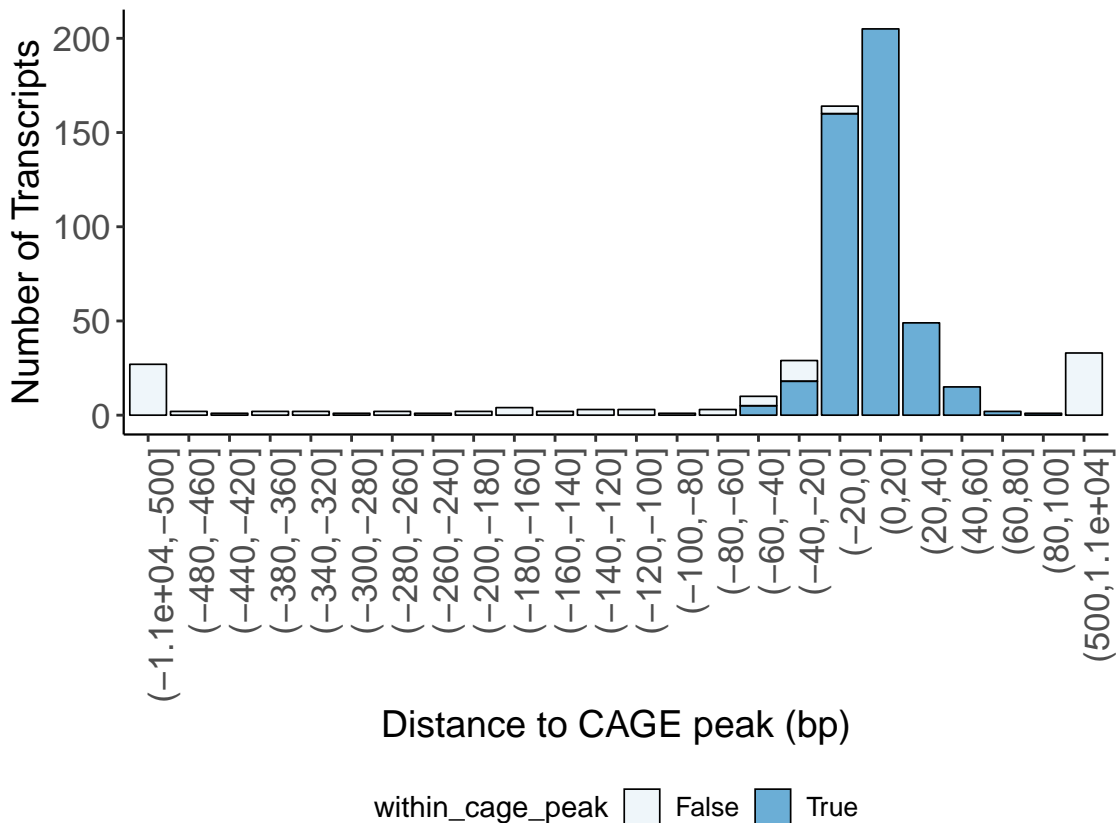
Category	Count	polyA Detected	%
FSM	576	492	85
ISM	156	114	73
NIC	439	358	82
NNC	395	362	92
Genic Genomic	11	6	55
Antisense	9	5	56
Fusion	6	6	100
Intergenic	17	8	47

Motif	Count	%
AATAAA	844	62.5
ATTAAA	255	18.9
AATACA	38	2.8
TATAAA	34	2.5
AGTAAA	29	2.1
TTTAAA	28	2.1
AAGAAA	25	1.9
AAAAAG	20	1.5
AATATA	18	1.3
CATAAA	18	1.3
GATAAA	12	0.9
AATGAA	9	0.7
ACTAAA	8	0.6
AATAGA	7	0.5
AAAACA	6	0.4

CAGE distances analysis

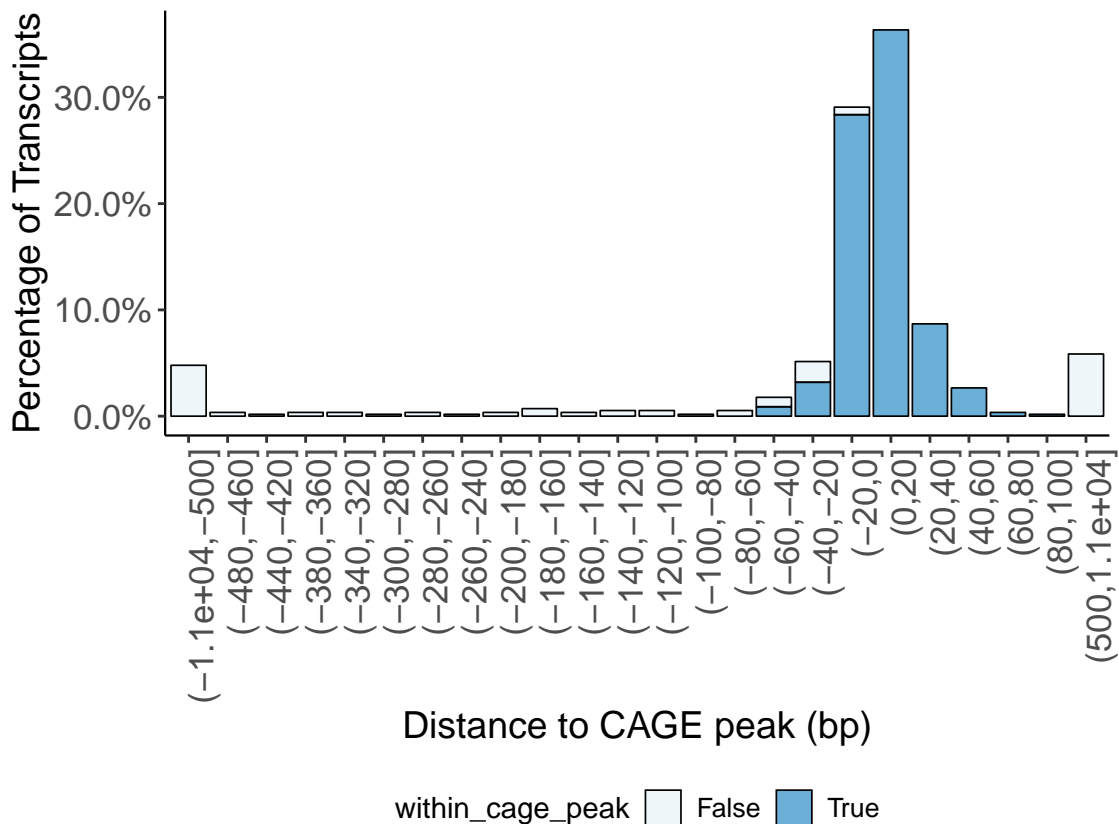
Distance to CAGE peak of multi-exonic FSM

Negative values indicate downstream of annotated CAGE peak



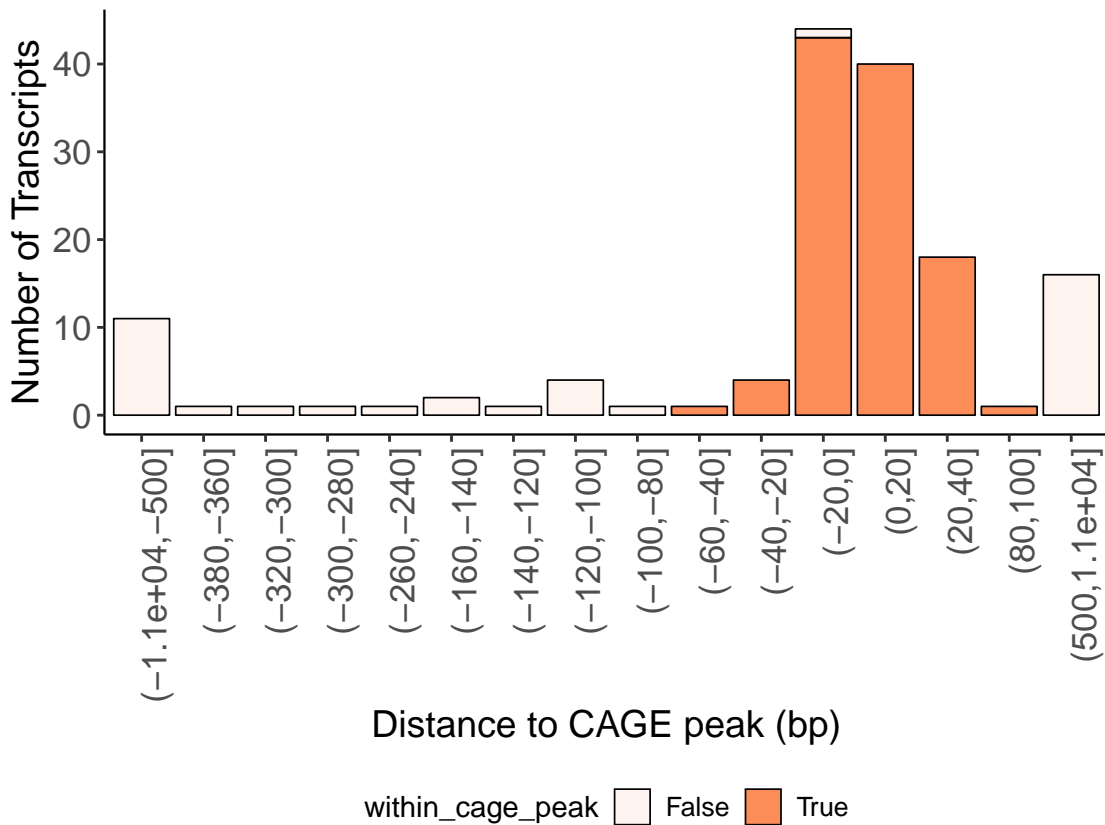
Distance to CAGE peak of multi-exonic FSM

Negative values indicate downstream of annotated CAGE peak



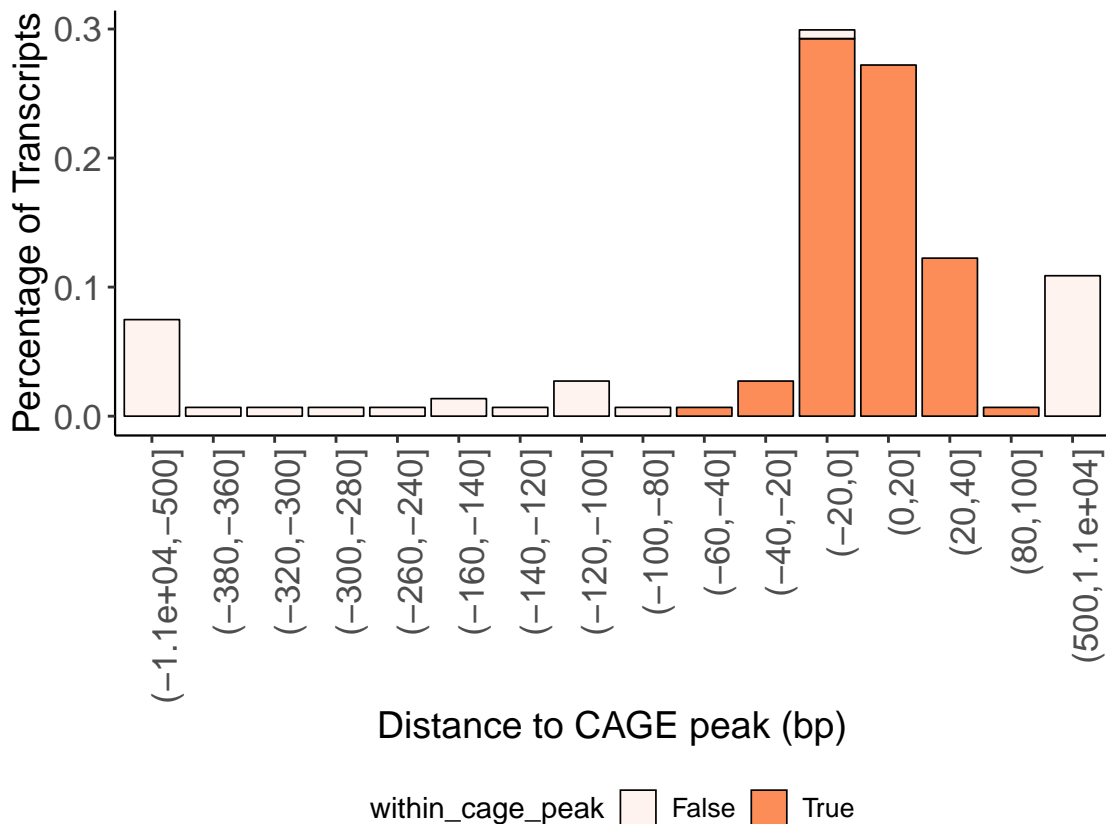
Distance to CAGE peak of multi-exonic ISM

Negative values indicate downstream of annotated CAGE peak

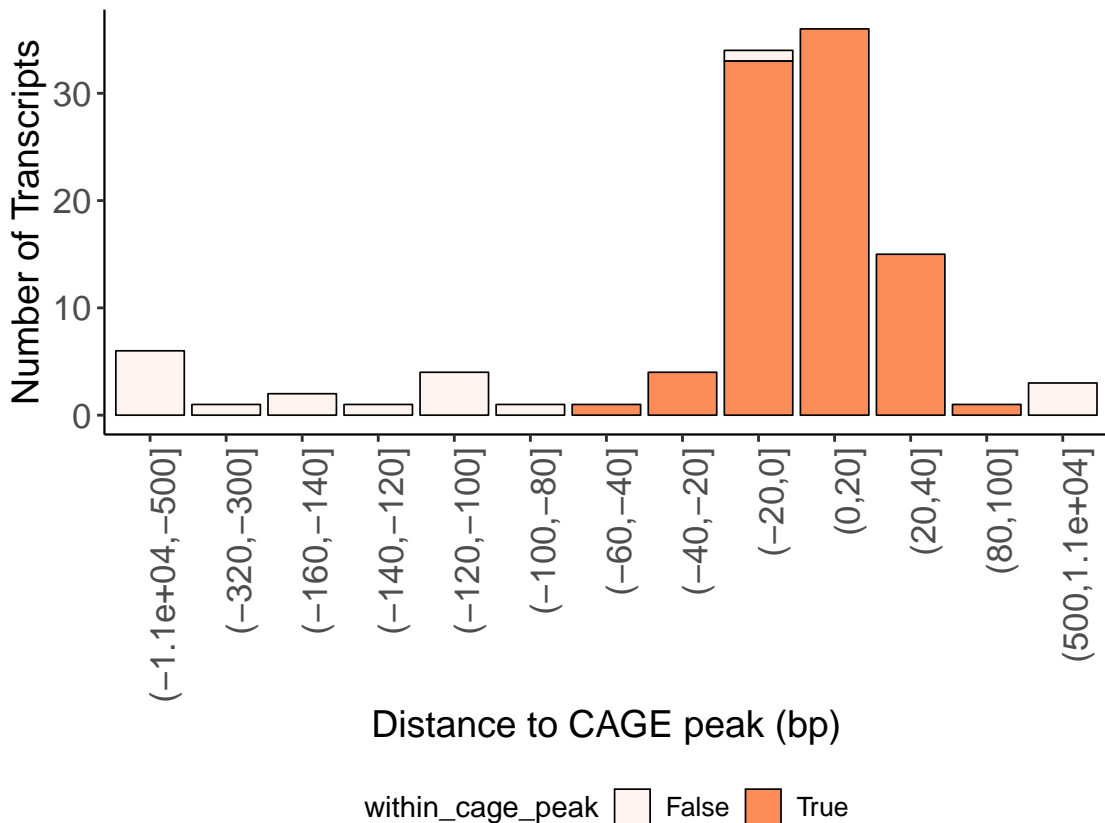


Distance to CAGE peak of multi-exonic ISM

Negative values indicate downstream of annotated CAGE peak

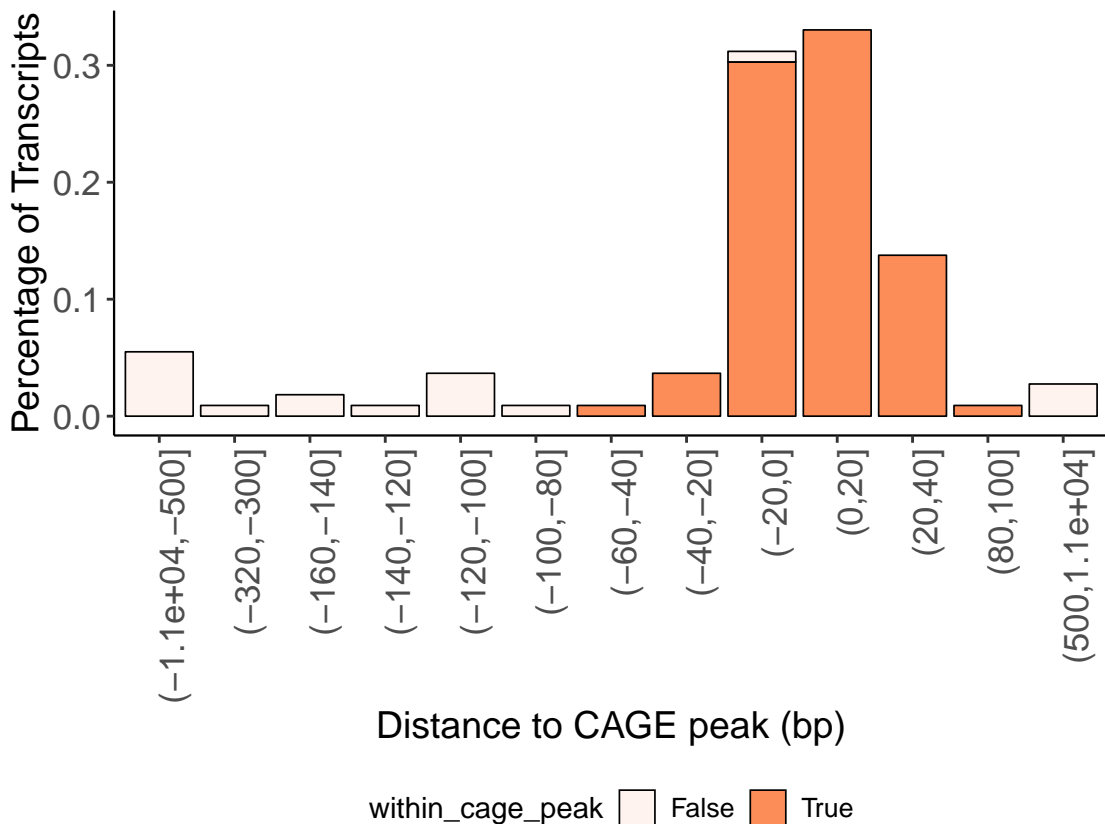


Distance to CAGE peak of multi-exonic ISM 3prime fragments
Negative values indicate downstream of annotated CAGE peak

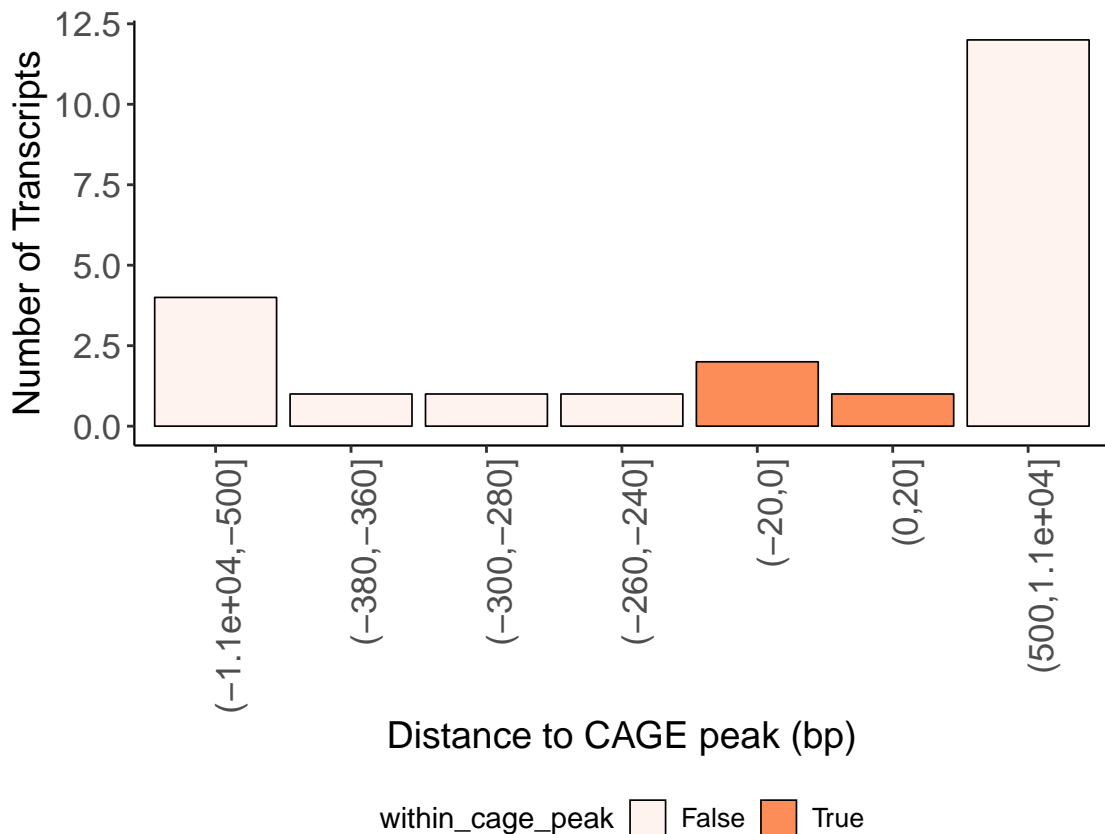


Distance to CAGE peak of multi-exonic ISM 3prime fragments

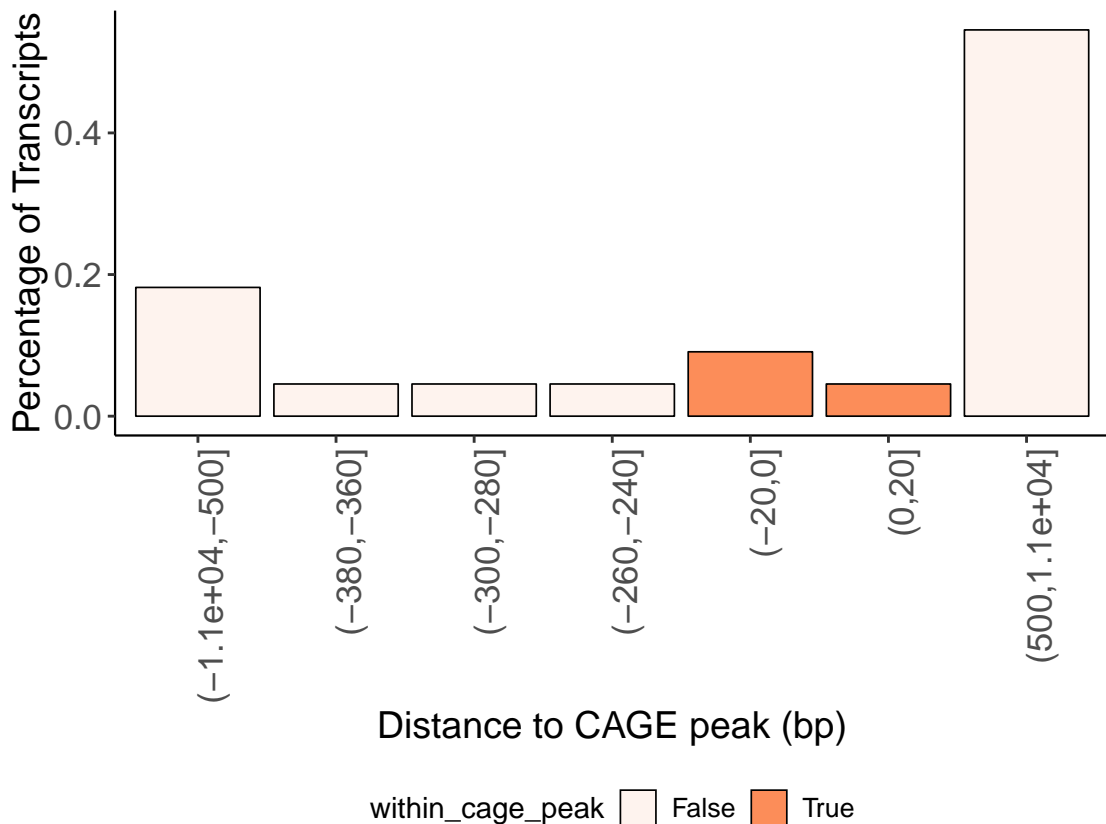
Negative values indicate downstream of annotated CAGE peak



Distance to CAGE peak of multi-exonic ISM 5prime fragments
Negative values indicate downstream of annotated CAGE peak

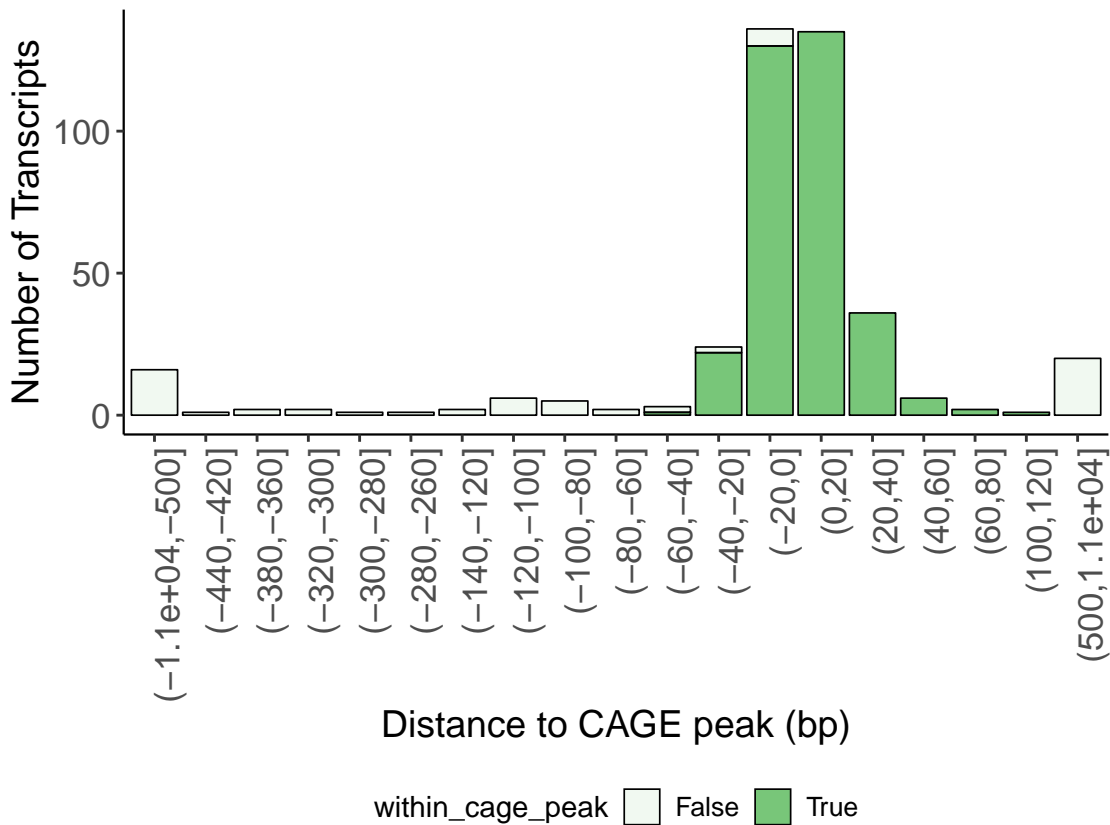


Distance to CAGE peak of multi-exonic ISM 5prime fragments
Negative values indicate downstream of annotated CAGE peak



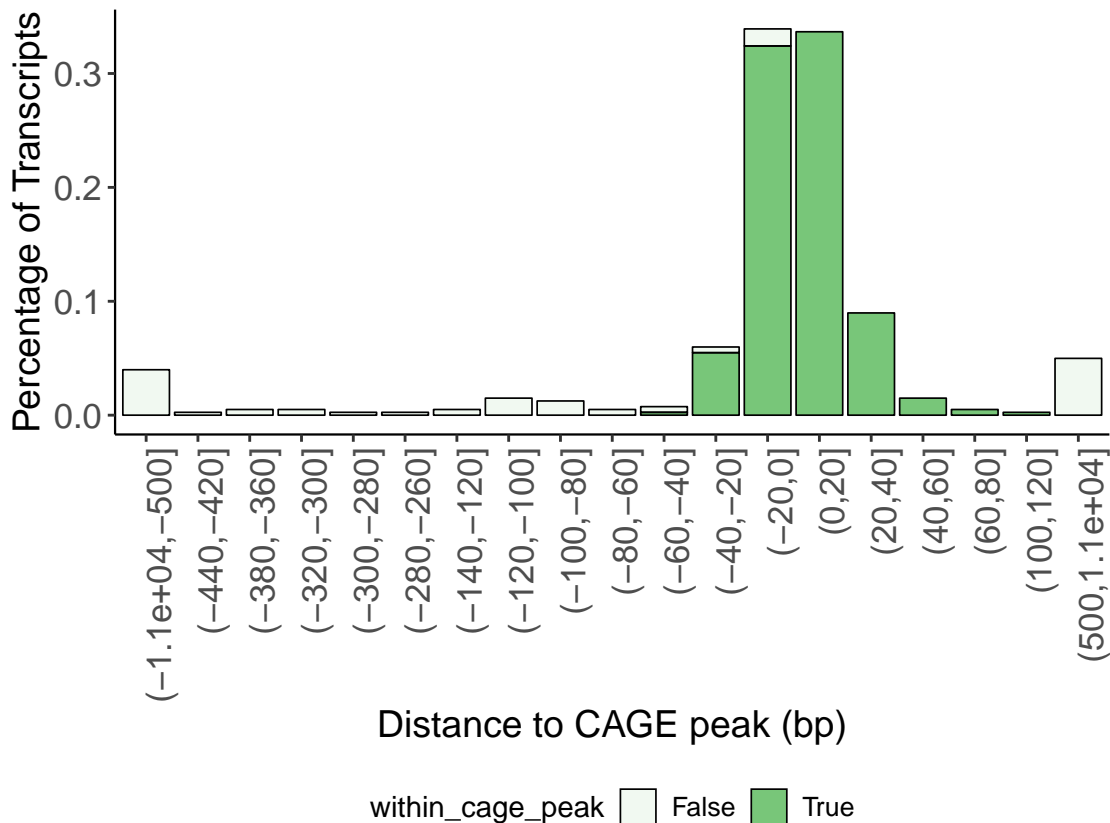
Distance to CAGE peak of multi-exonic NIC

Negative values indicate downstream of annotated CAGE peak



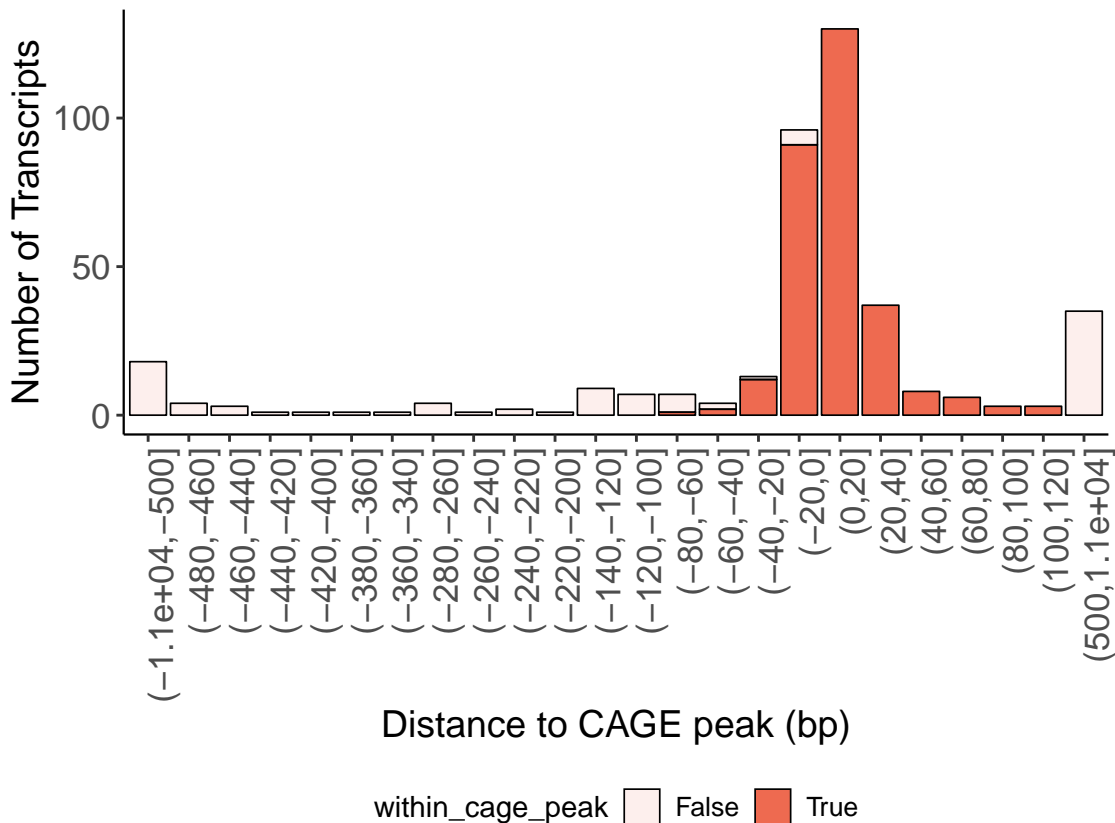
Distance to CAGE peak of multi-exonic NIC

Negative values indicate downstream of annotated CAGE peak



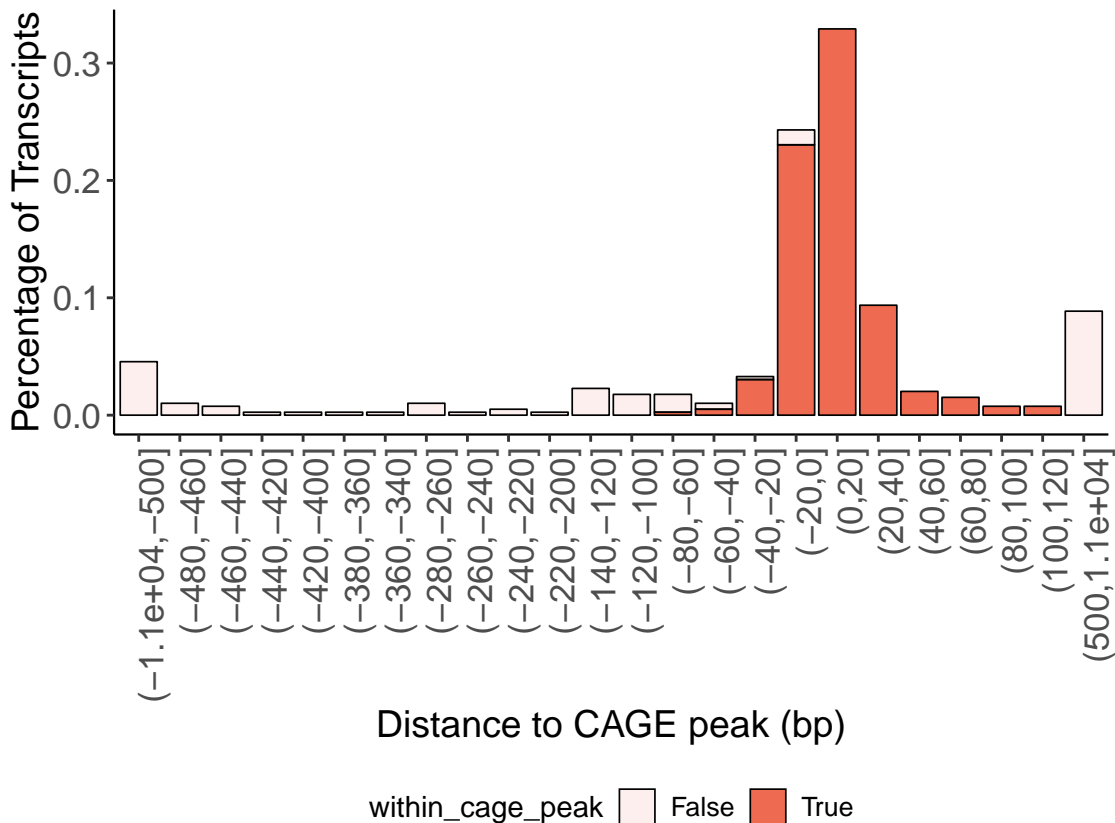
Distance to CAGE peak of multi-exonic NNC

Negative values indicate downstream of annotated CAGE peak



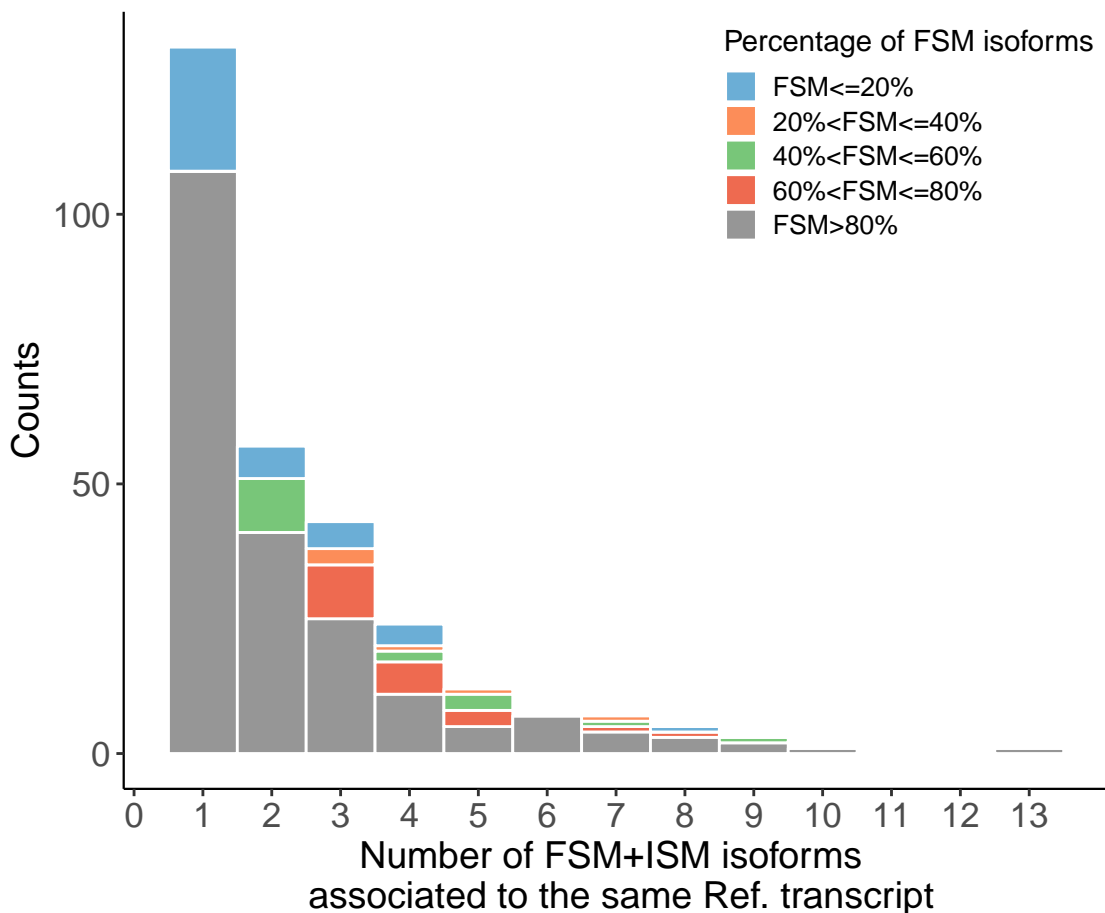
Distance to CAGE peak of multi-exonic NNC

Negative values indicate downstream of annotated CAGE peak



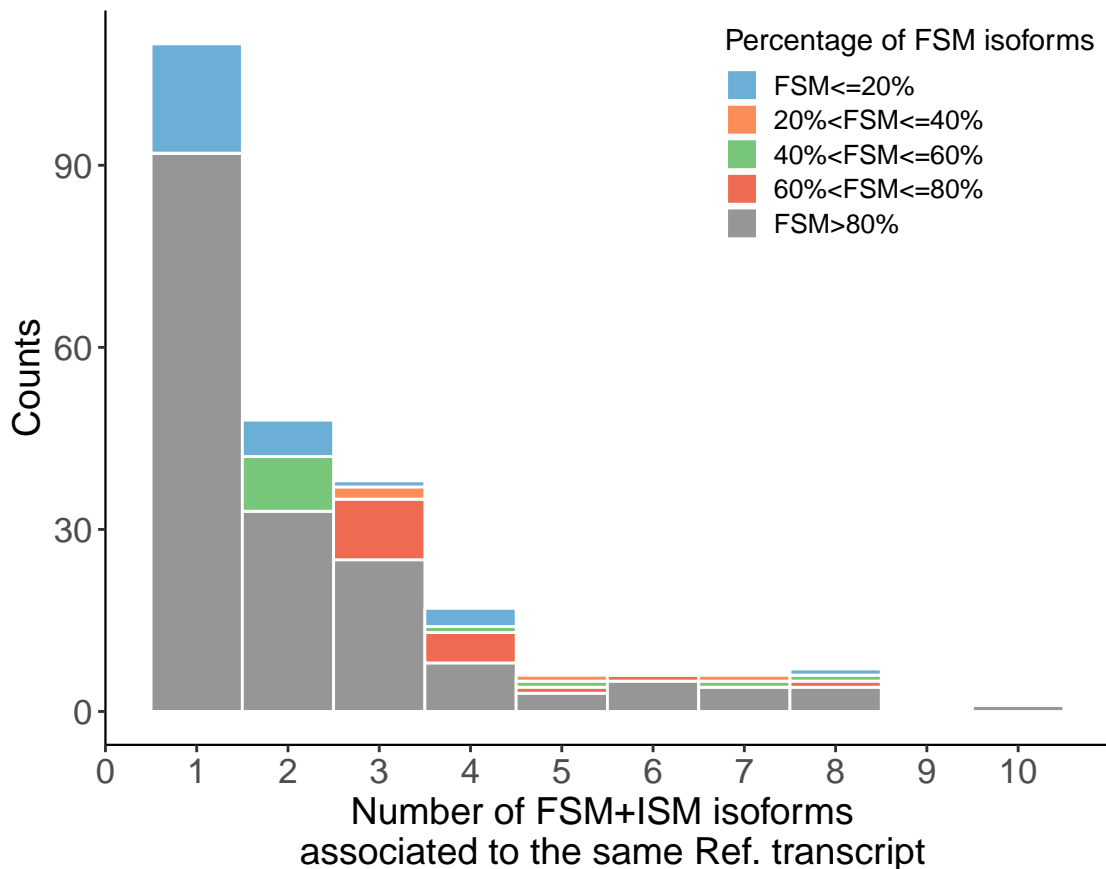
*Accumulation of FSM and ISM
to the same reference transcript*

Accumulation of FSM and ISM isoforms
associated to the same reference transcript.



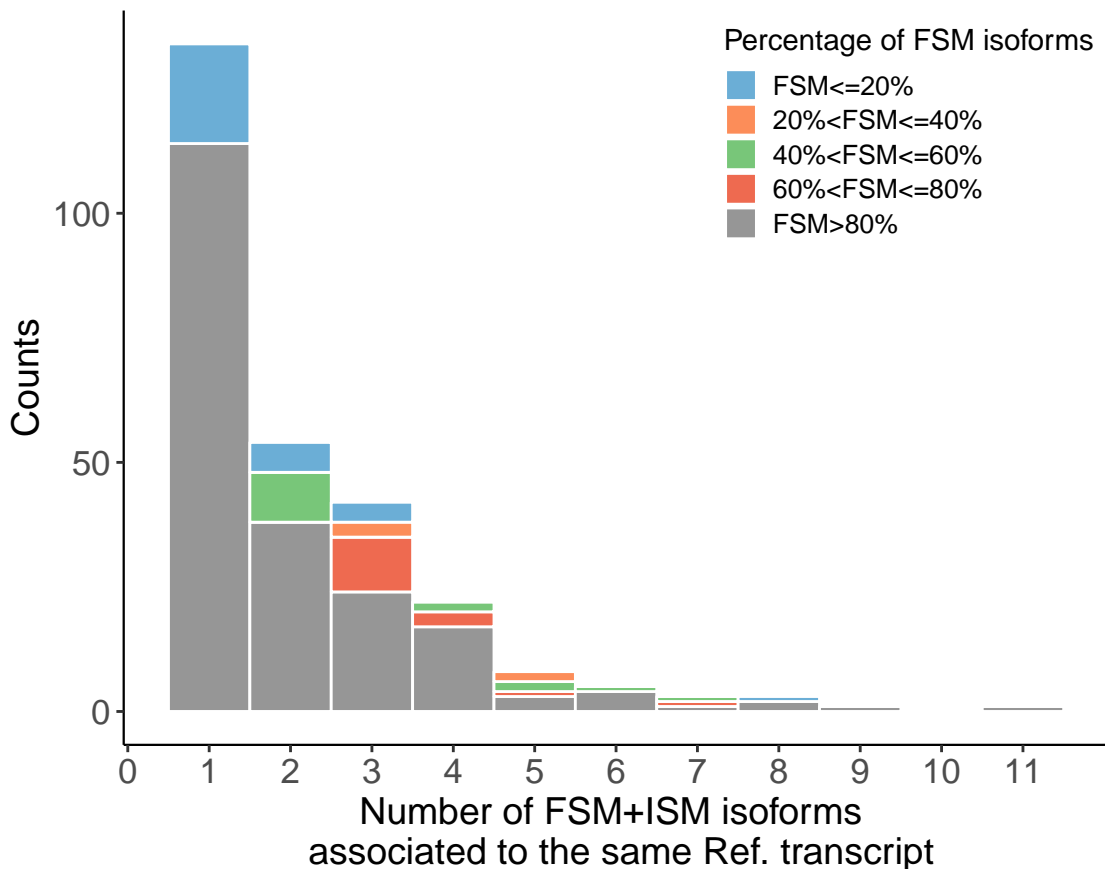
Accumulation of FSM and ISM isoforms associated to the same reference transcript.

ONLY polyA motif and CAGE + isoforms



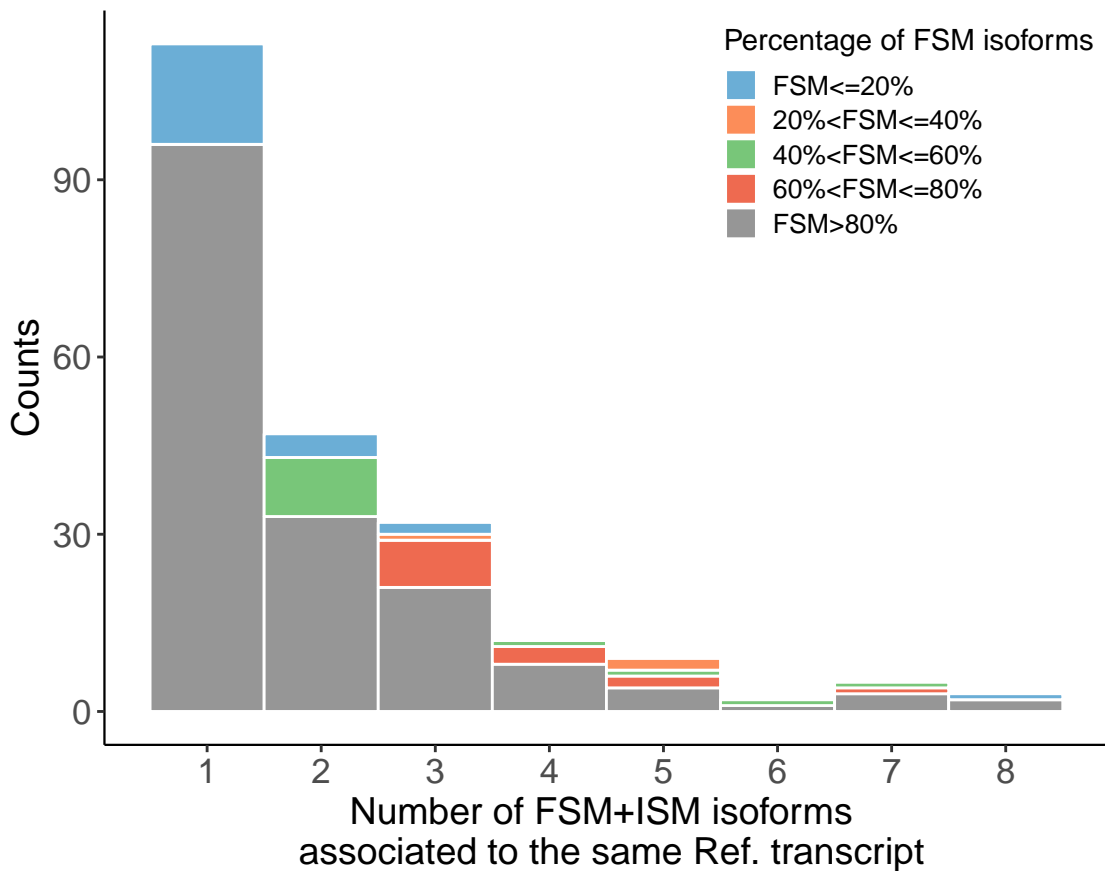
Accumulation of FSM and ISM isoforms
associated to the same reference transcript.

ONLY polyA motif and CAGE + isoforms



Accumulation of FSM and ISM isoforms associated to the same reference transcript.

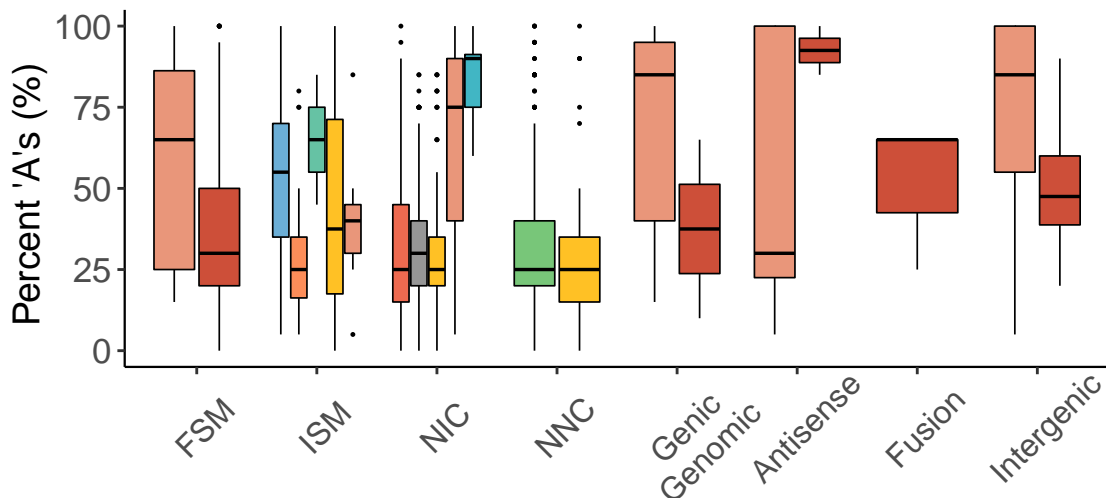
ONLY polyA motif and CAGE + isoforms



Intra-Priming Quality Check

Possible Intra-Priming by Structural Category

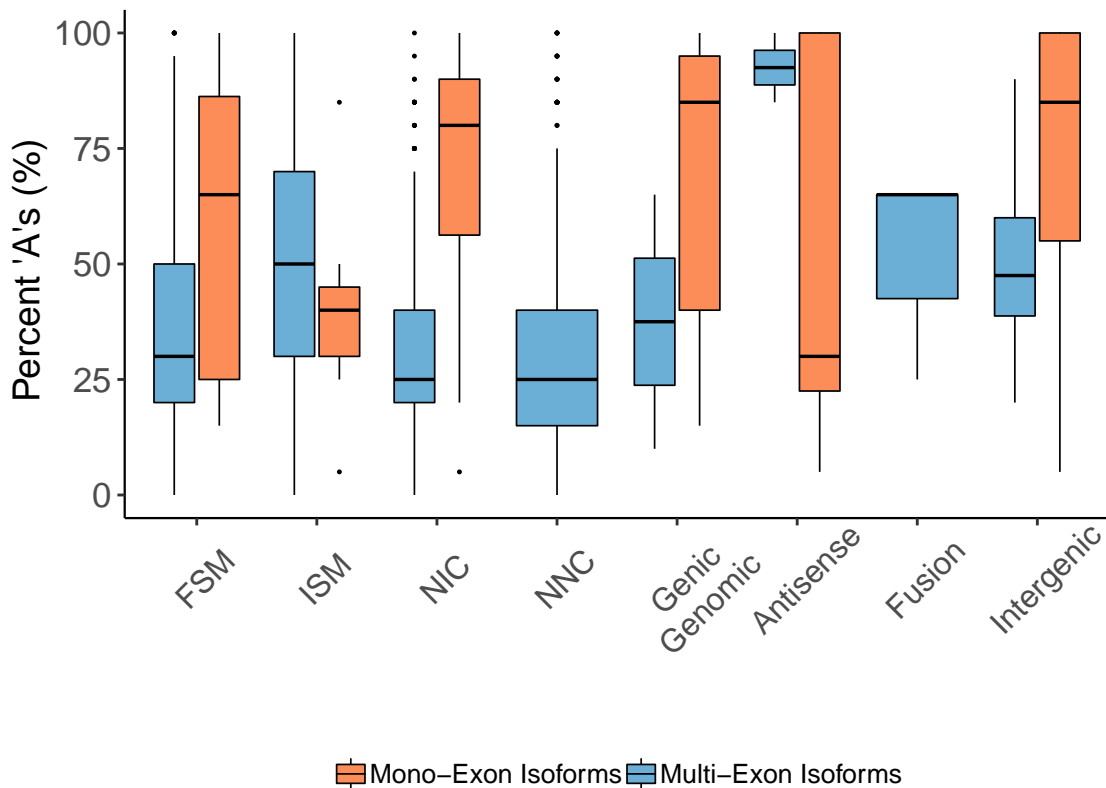
Percent of genomic 'A's in downstream 20 bp



- Intron retention
- Internal fragment
- Mono-exon
- Combination of annotated junctions
- Mono-exon by intron retention
- 3' fragment
- 5' fragment
- Multi-exon
- Combination of annotated splice sites
- At least one annotated donor/acceptor

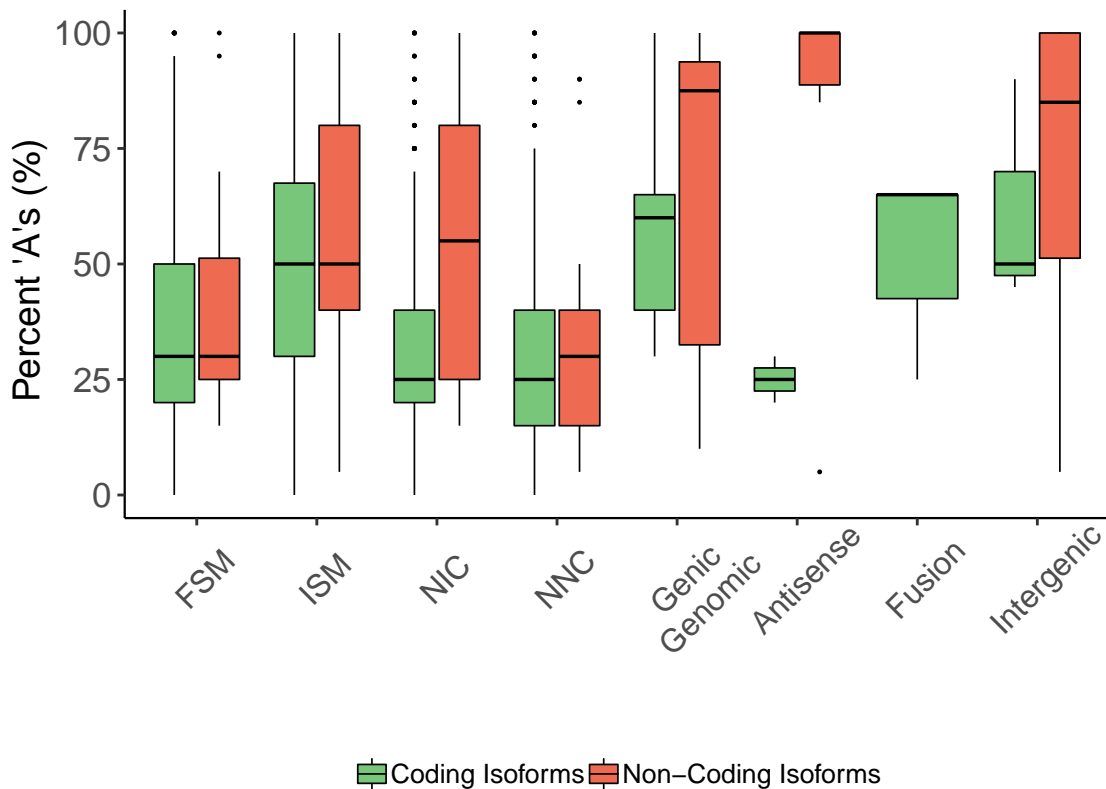
Possible Intra-Priming, Mono- vs Multi-Exon

Percent of genomic 'A's in downstream 20 bp



Possible Intra-Priming, Coding vs Non-Coding

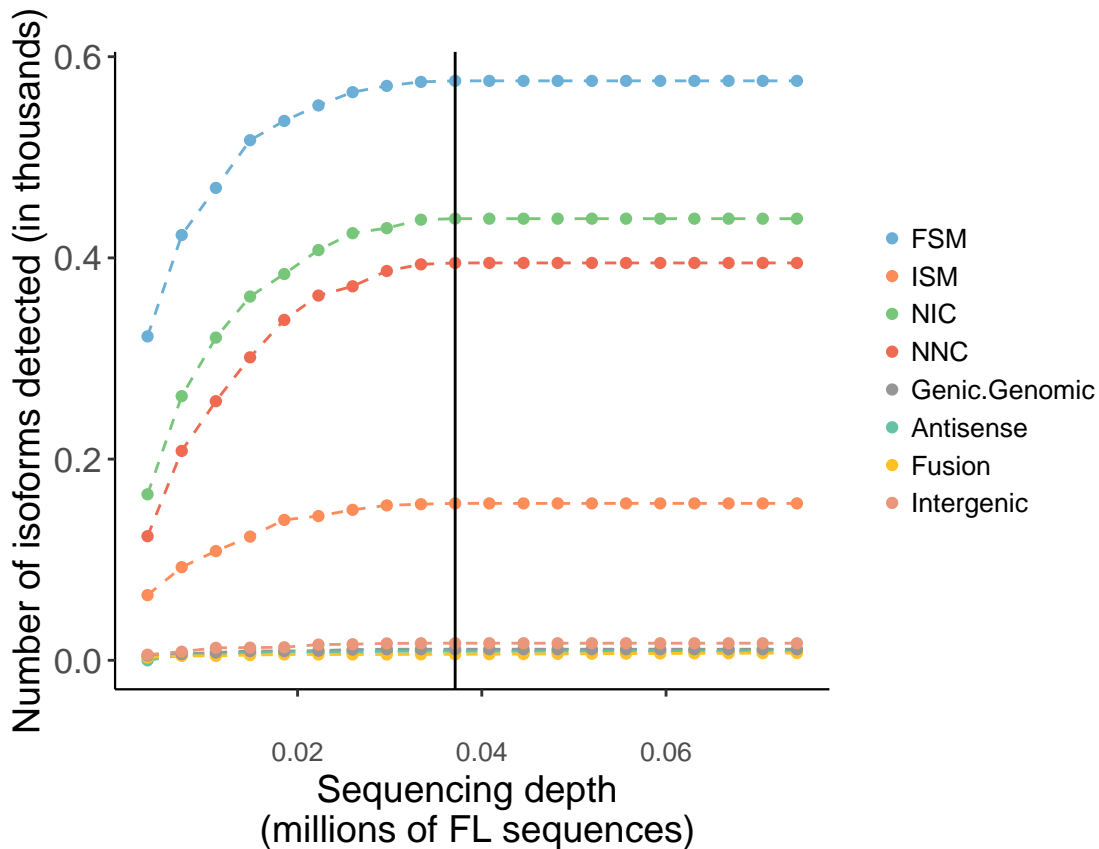
Percent of genomic 'A's in downstream 20 bp



Saturation curves

Saturation plot per structural category

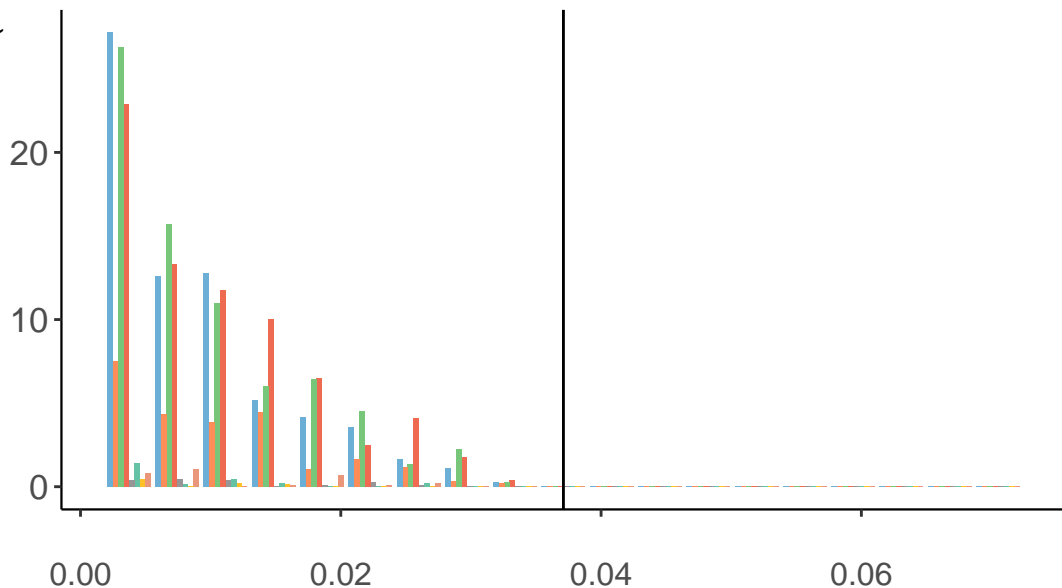
Minimum number of FL-reads required to call an isoform = 1



Absolute increment of isoforms detected (in thousands)

Increments of detected isoforms per structural category

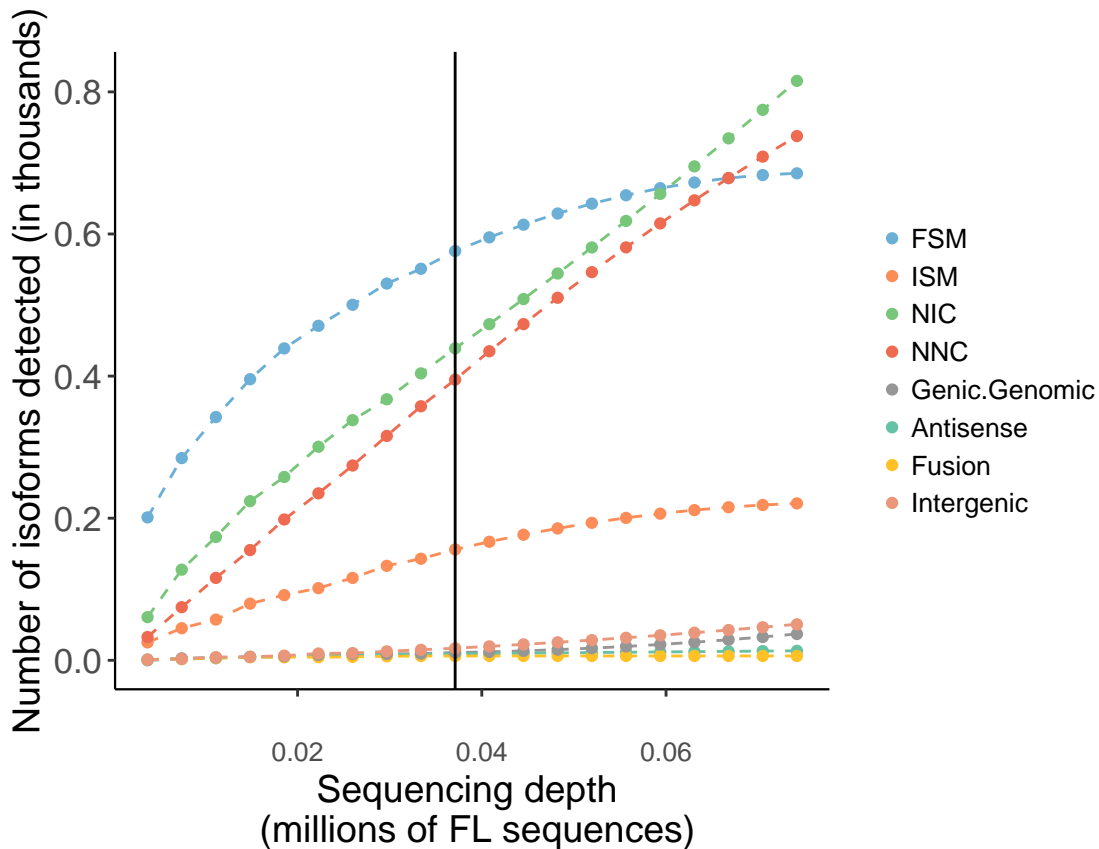
Minimum number of FL-reads required to call an isoform = 1

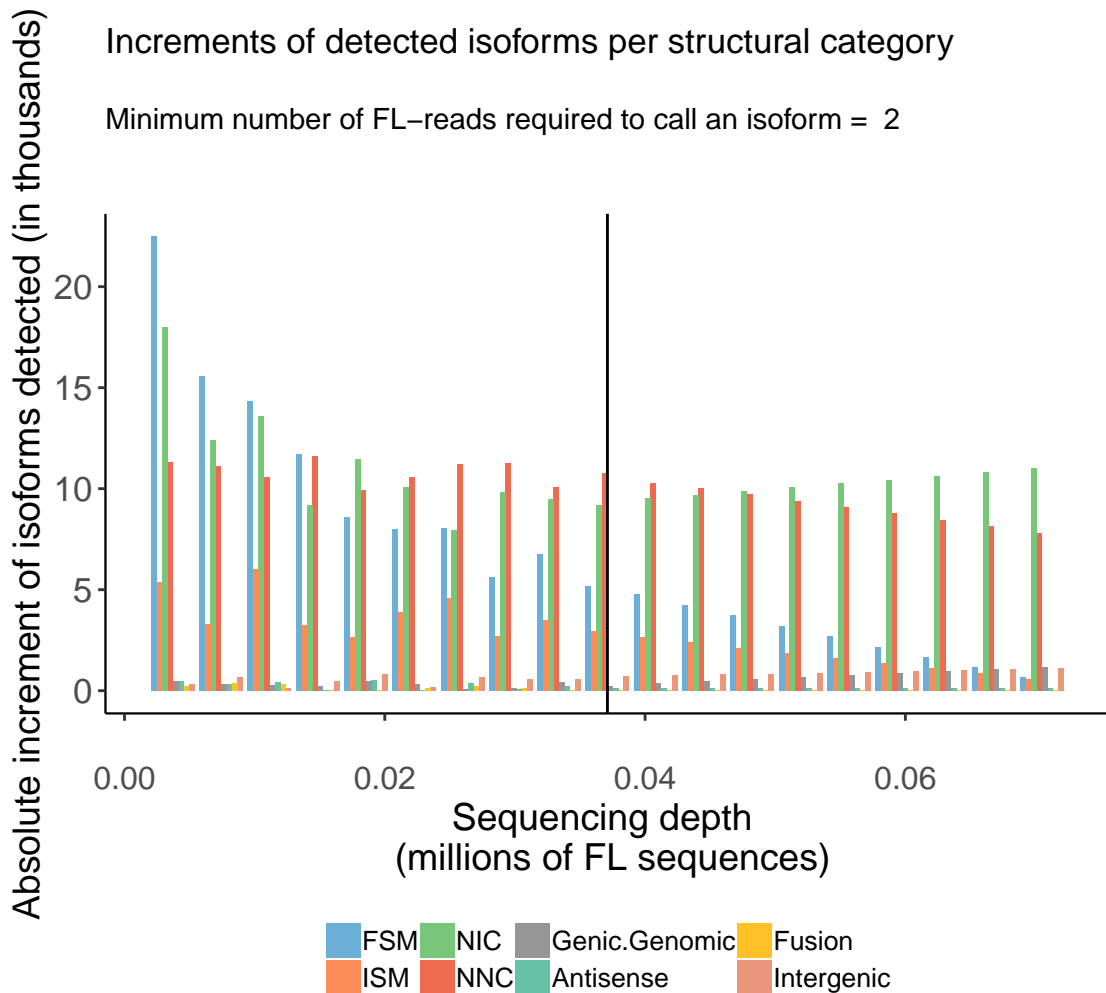


FSM NIC Genic.Genomic Fusion
ISM NNC Antisense Intergenic

Saturation plot per structural category

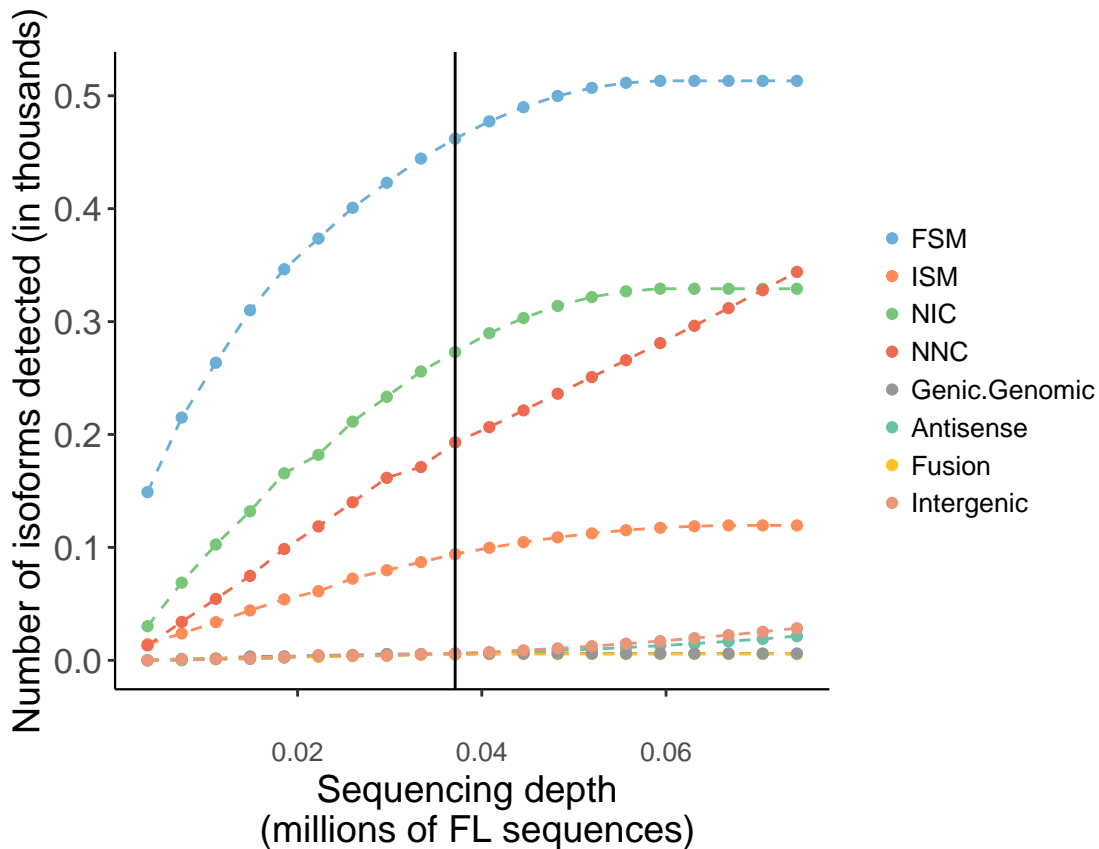
Minimum number of FL-reads required to call an isoform = 2

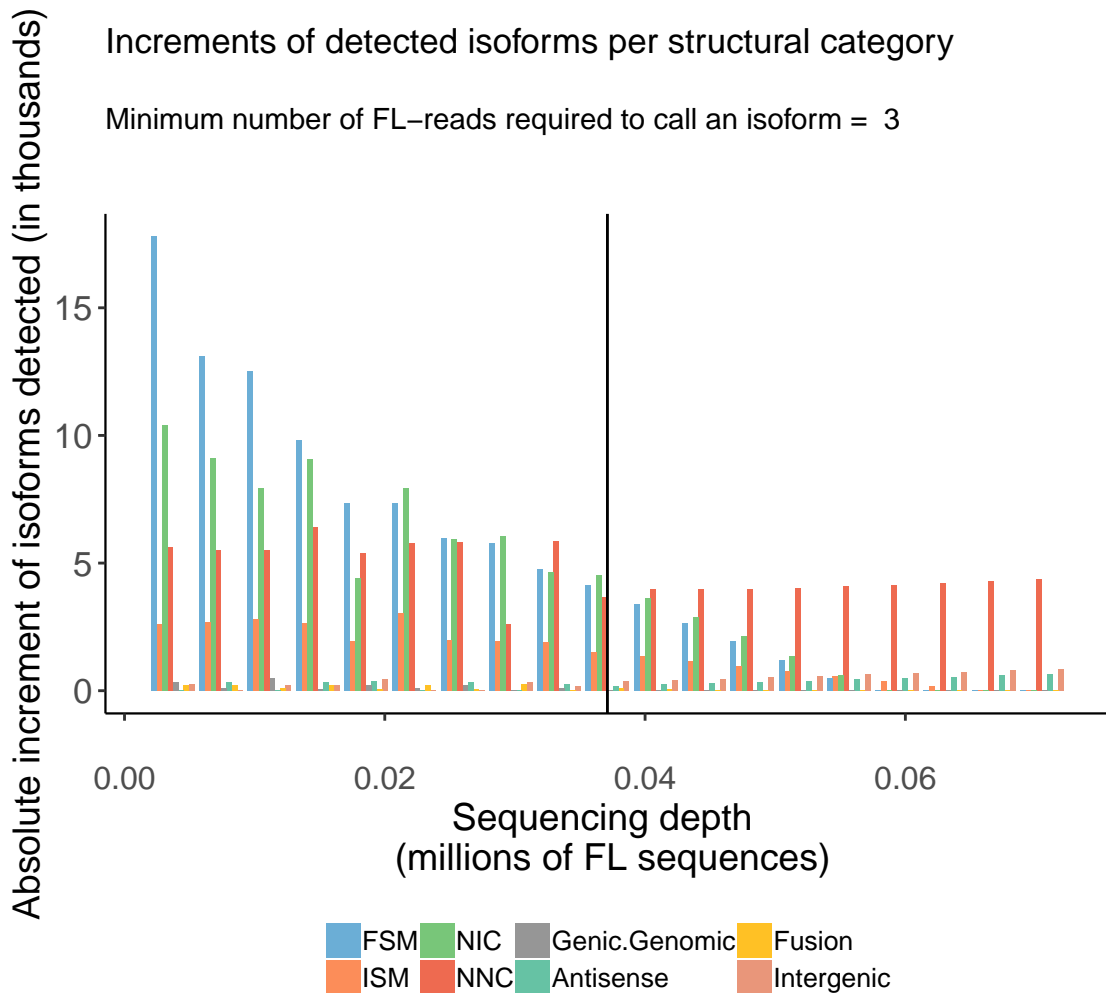




Saturation plot per structural category

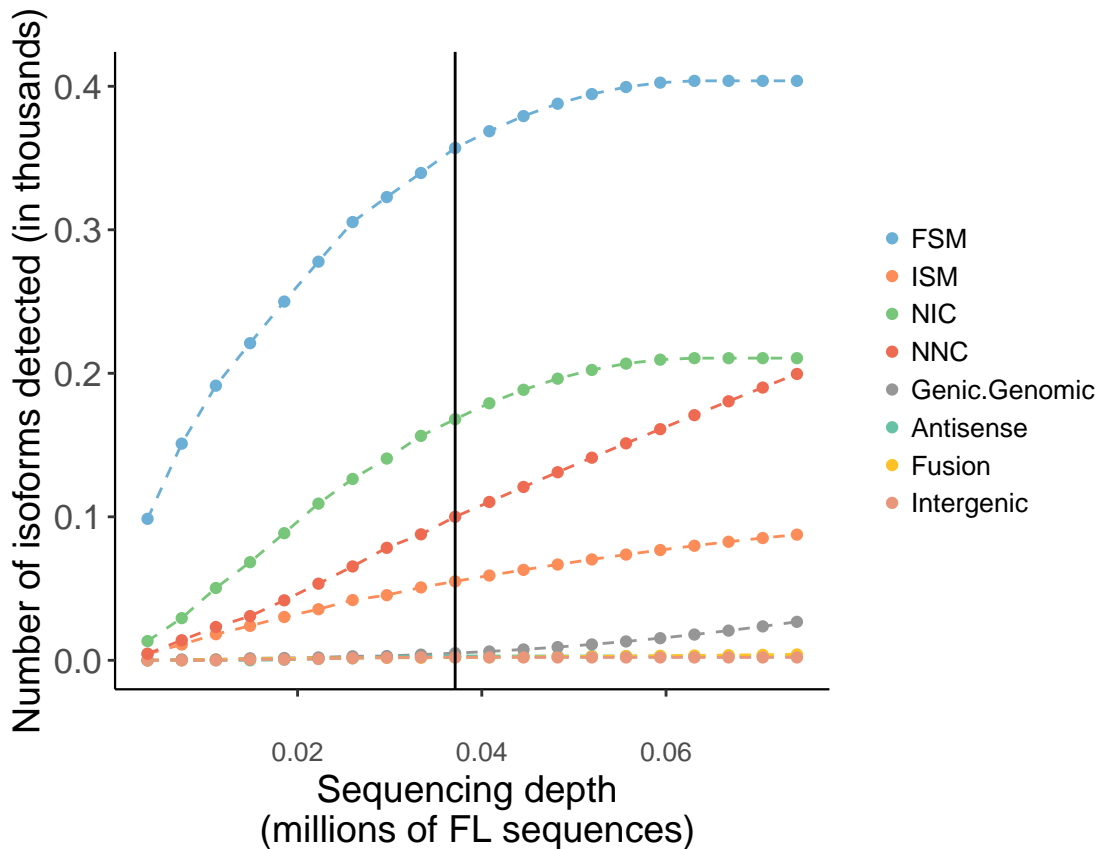
Minimum number of FL-reads required to call an isoform = 3

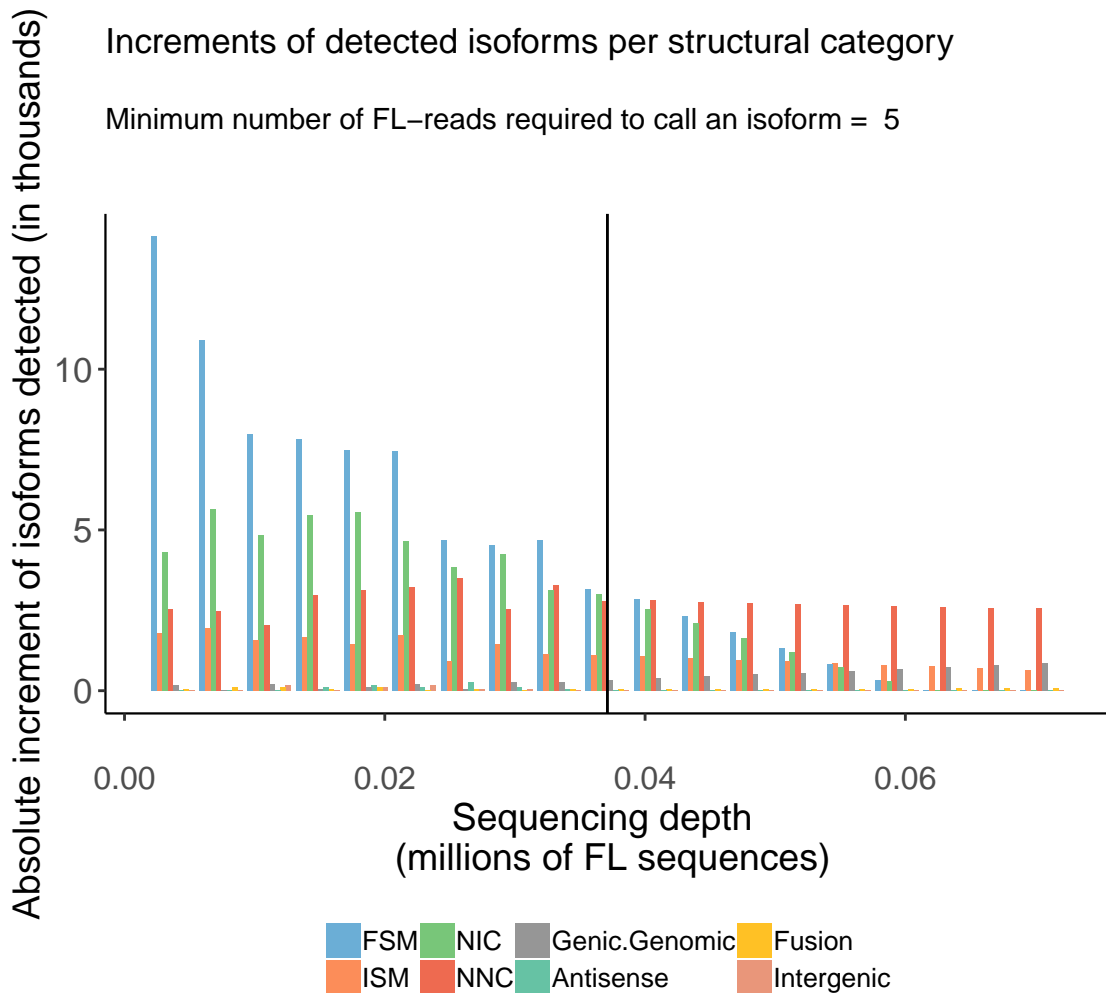




Saturation plot per structural category

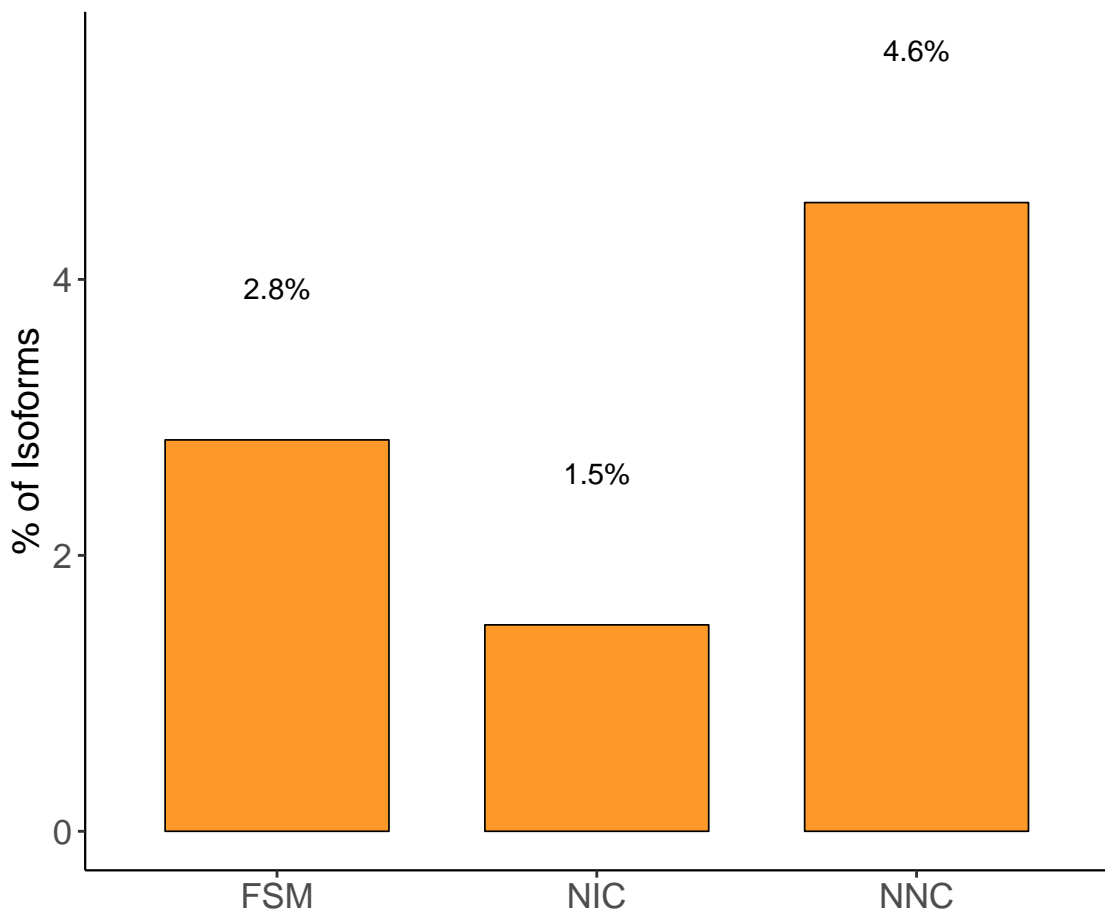
Minimum number of FL-reads required to call an isoform = 5



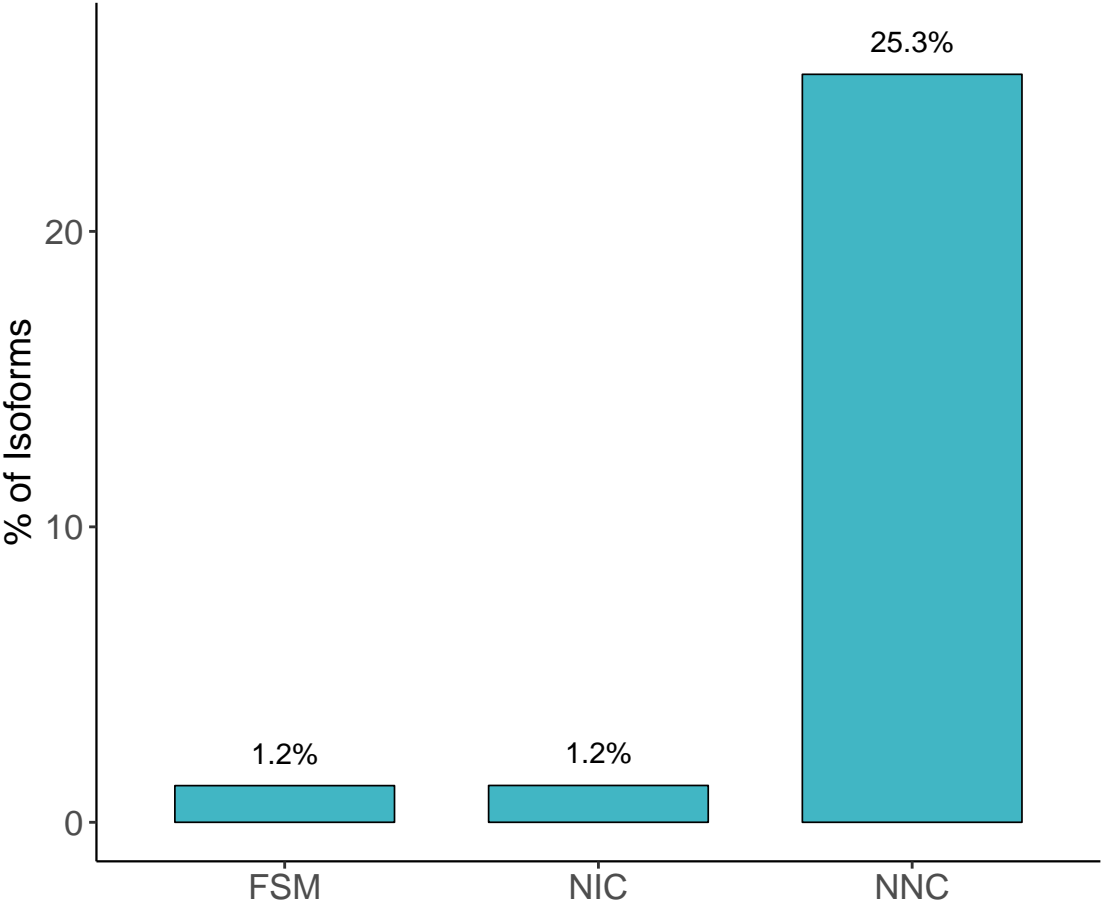


Quality Controls

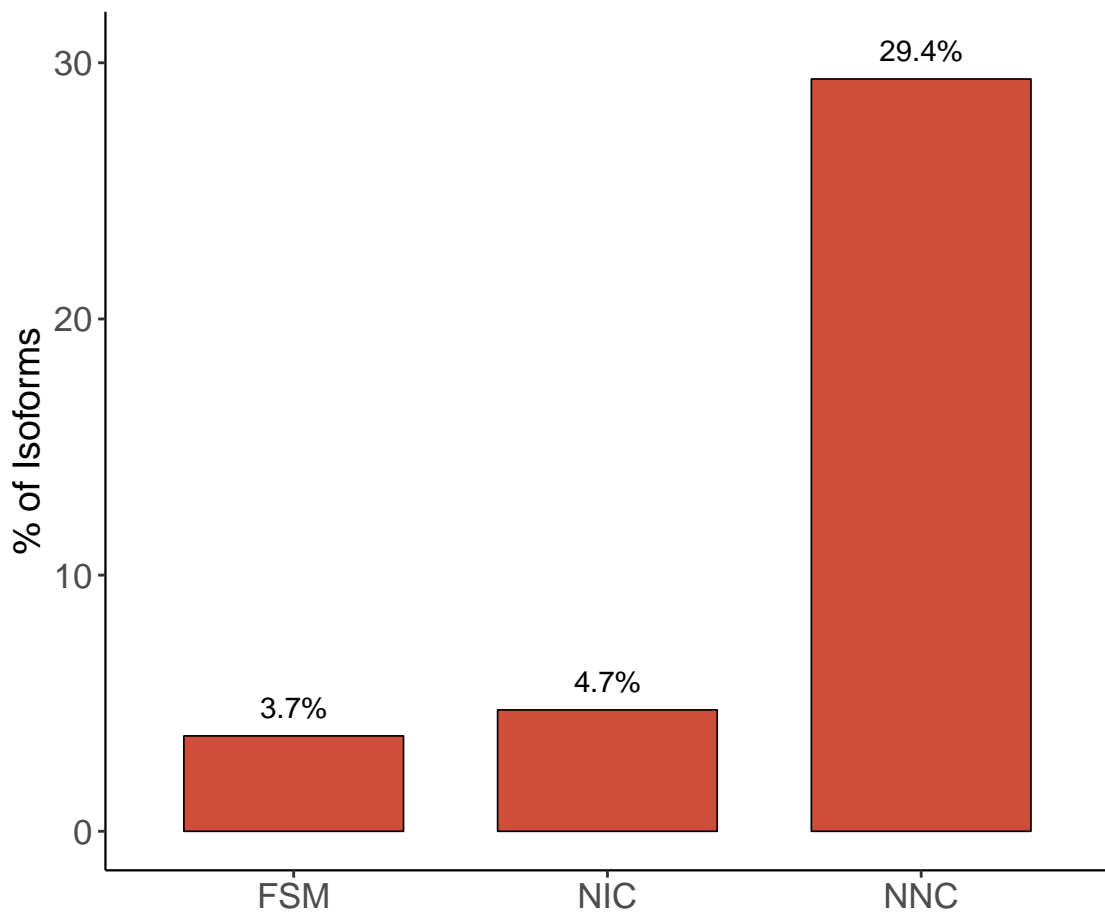
Incidence of RT-switching



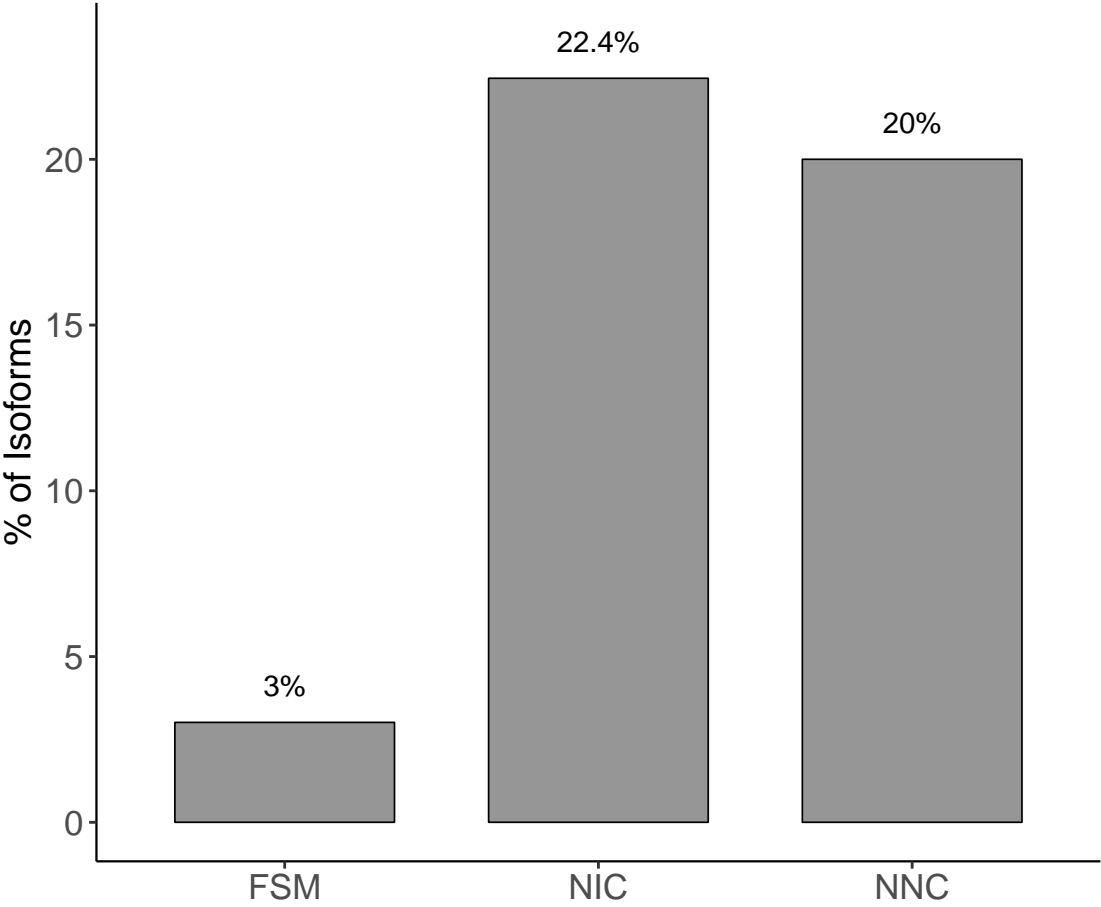
Incidence of Non-Canonical Junctions



Incidence of SJ without SR coverage Junctions



Incidence of NMD by structural category



Quality control attributes across structural categories

