

SQANTI2 report

Unique Genes: 19312
Unique Isoforms: 165684

Gene classification

Category	# Genes
Annotated Genes	18467
Novel Genes	845

*Characterization of transcripts
based on splice junctions*

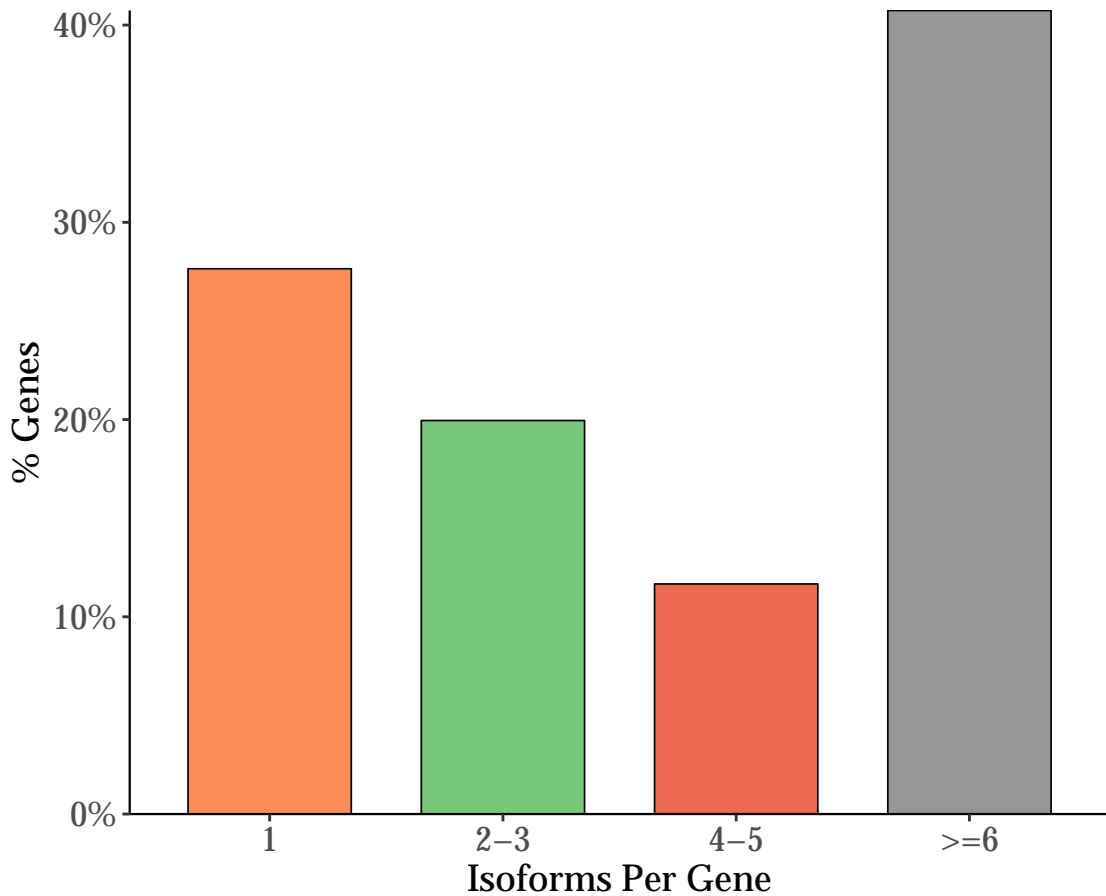
Category	# Isoforms	# Genes
FSM	33746	12579
ISM	18754	6842
NIC	86309	12163
NNC	23889	7384
Genic Genomic	220	181
Antisense	260	219
Fusion	1880	1038
Intergenic	626	626

Splice Junction Classification

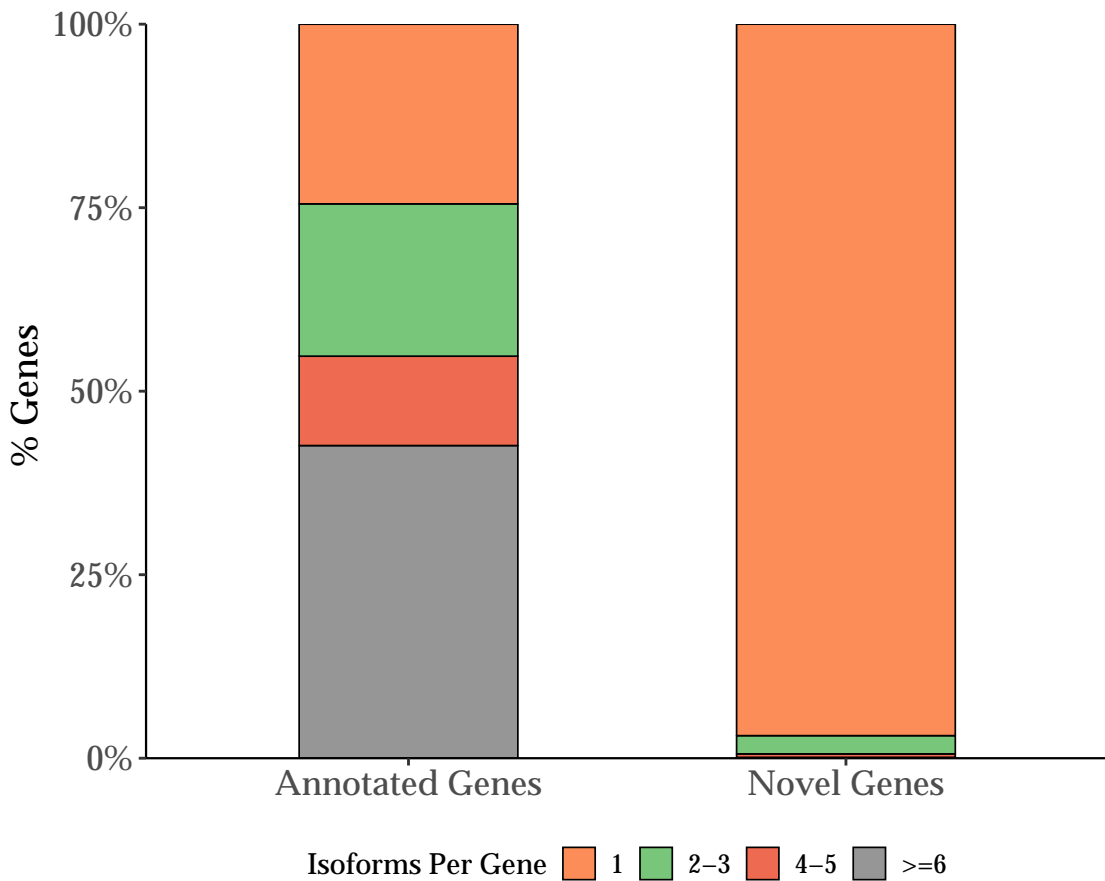
Category	# SJs	Percent
Known canonical	162962	82.43
Known Non-canonical	97	0.05
Novel canonical	34622	17.51
Novel Non-canonical	6	0.00

Gene Characterization

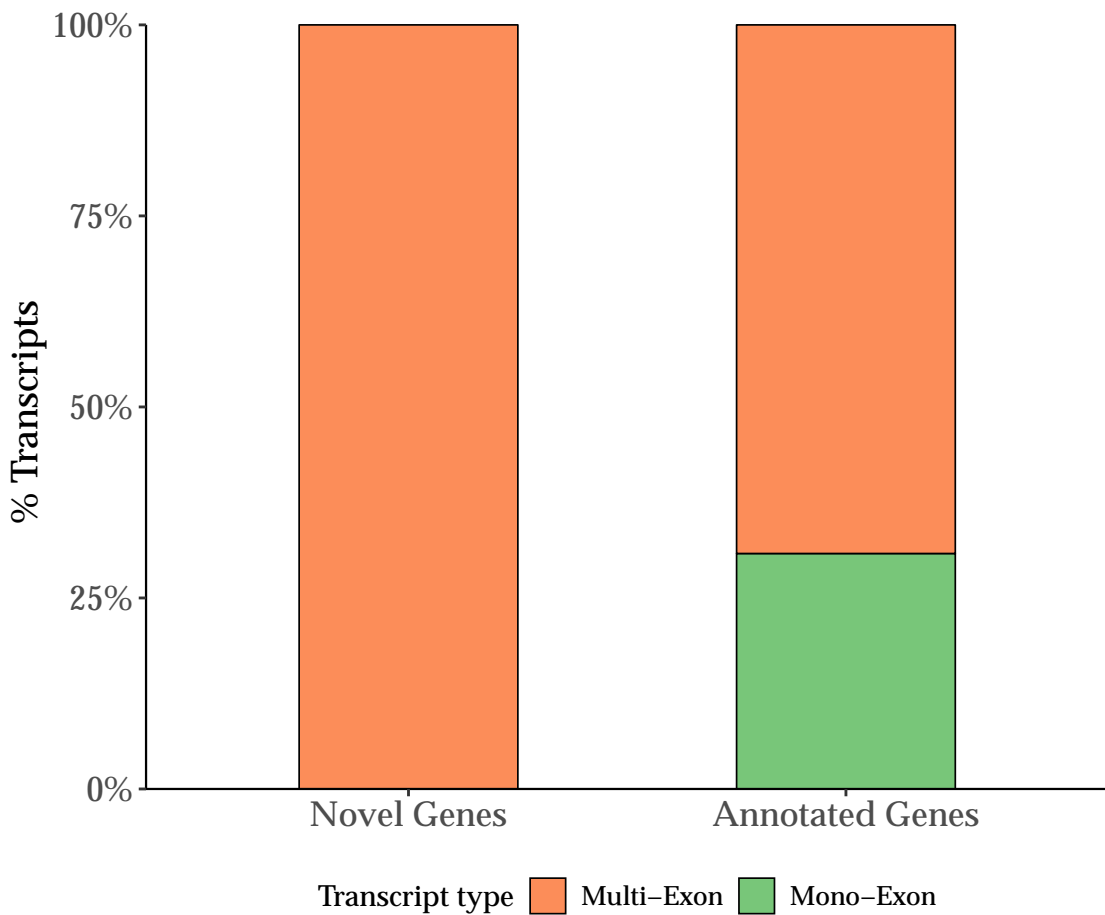
Number of Isoforms per Gene



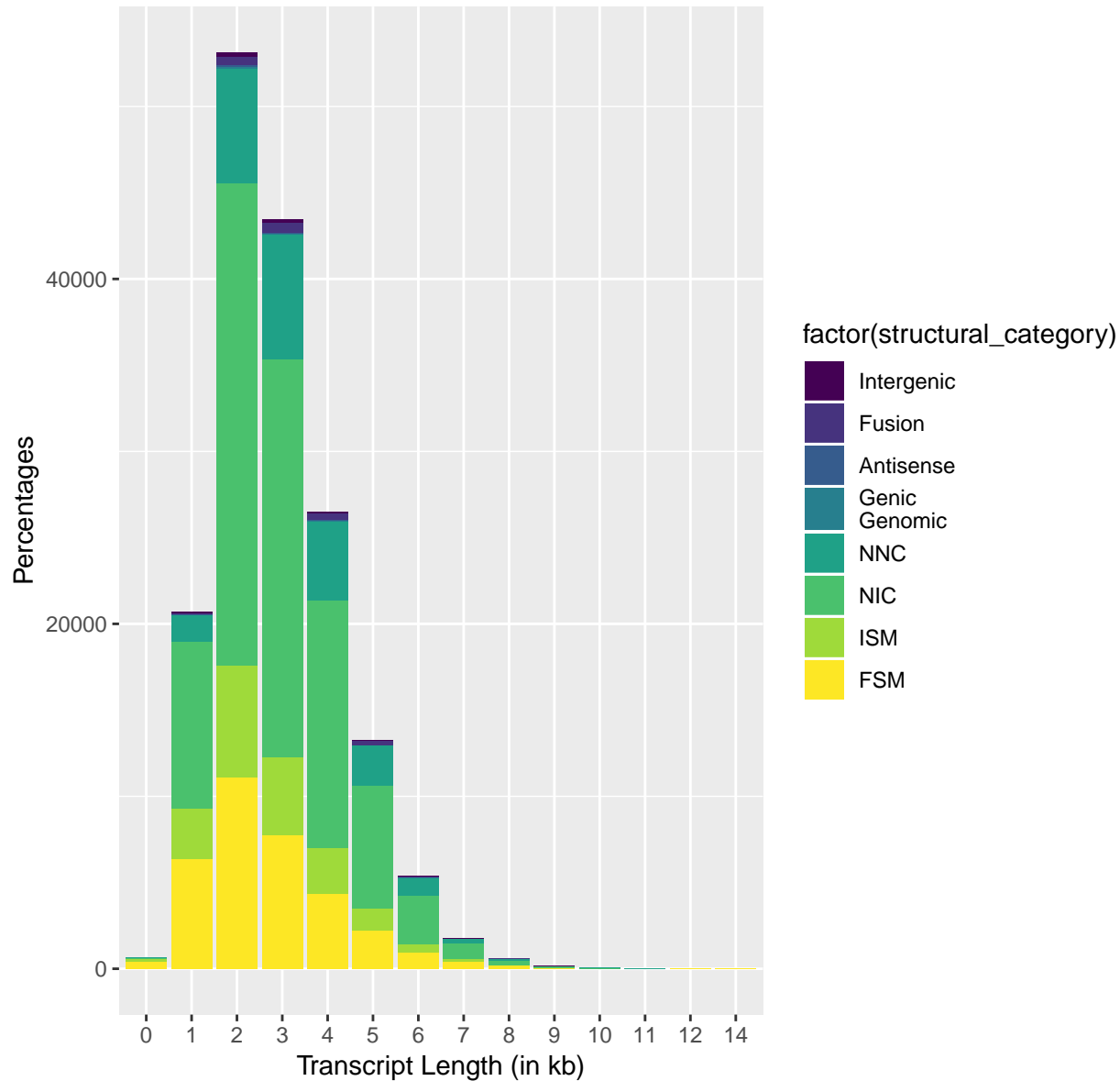
Number of Isoforms per Gene, Novel vs Known Genes



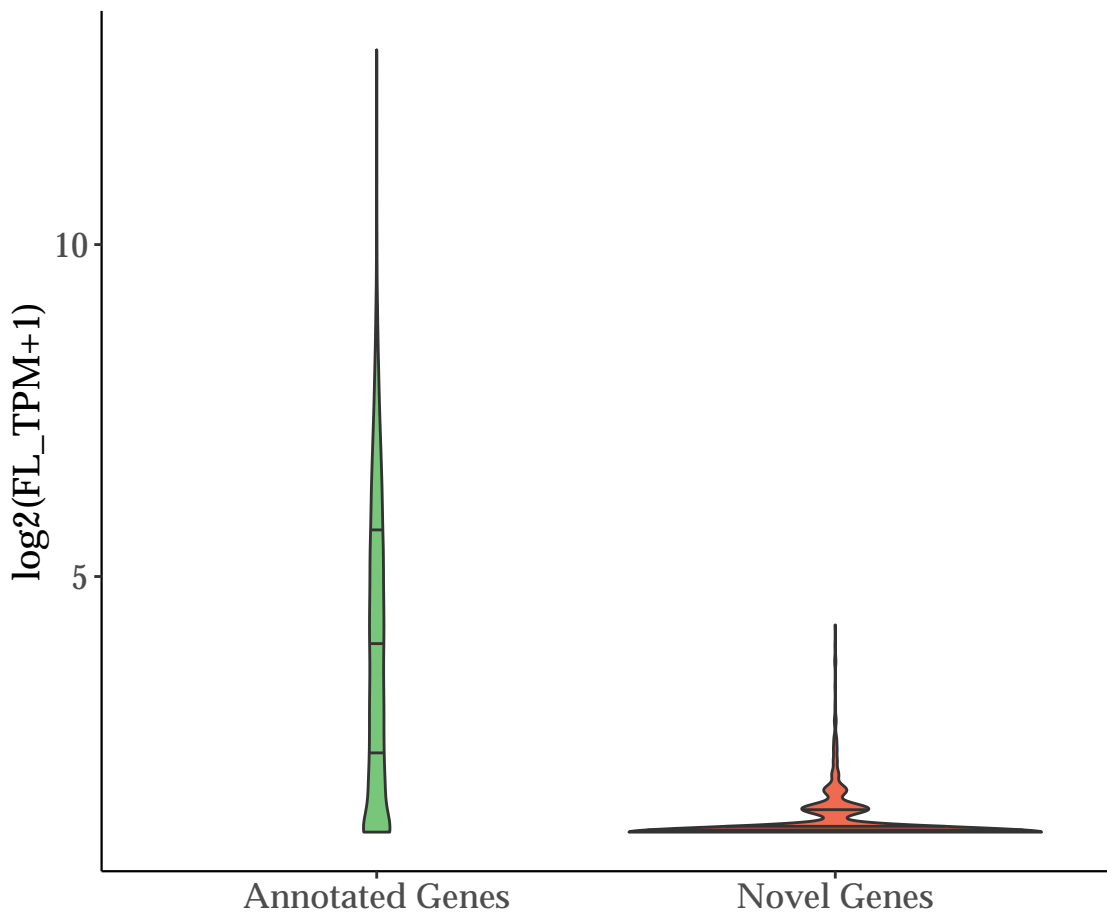
Distribution of Mono- vs Multi-Exon Transcripts



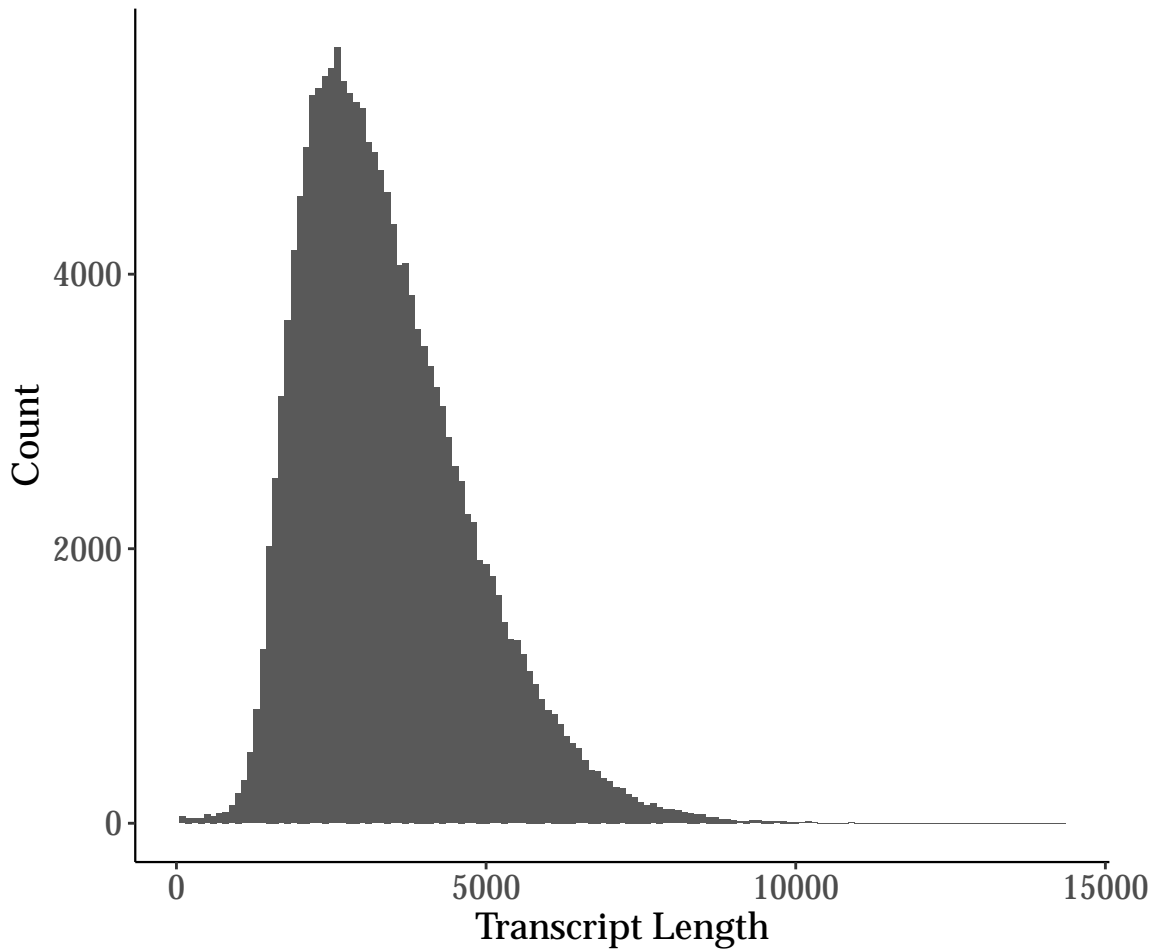
Classifications by Transcript Length



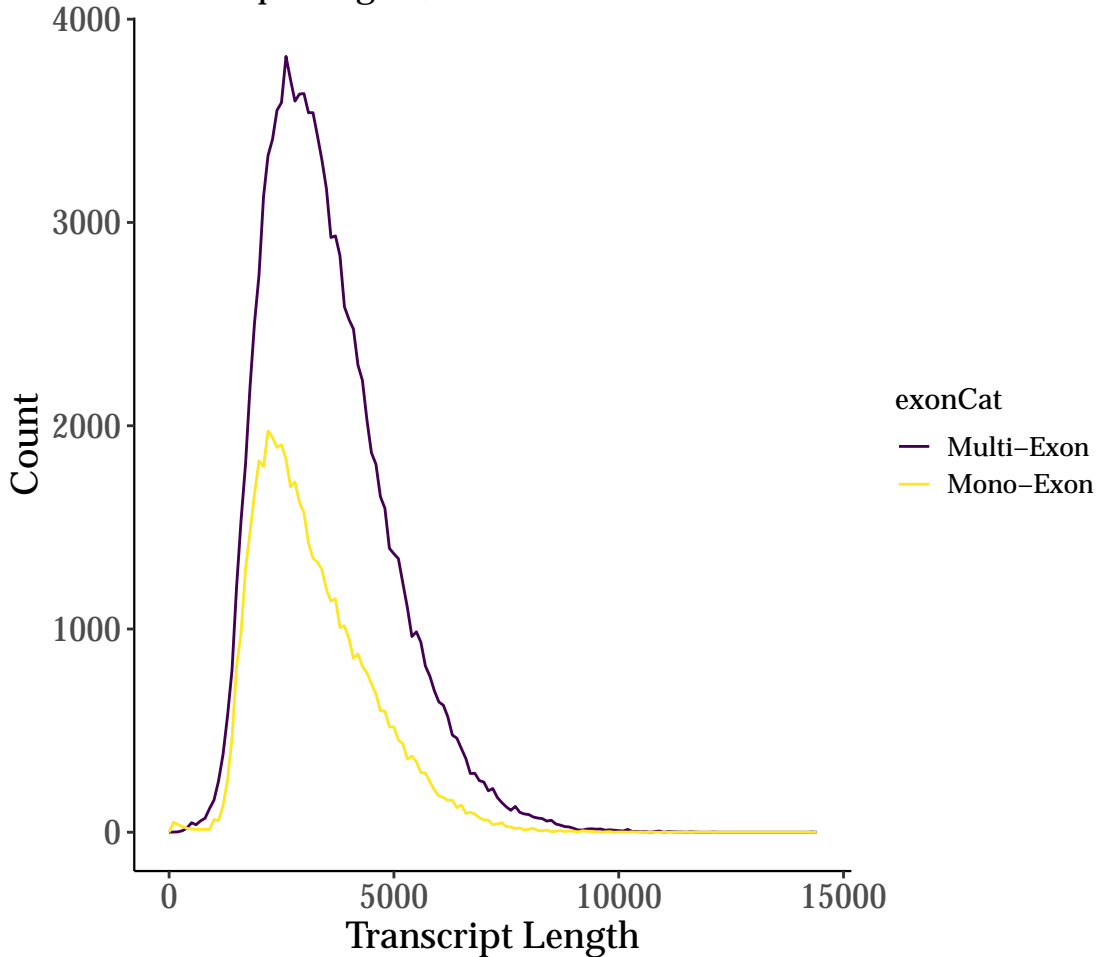
Number of FL reads per Gene by type of gene annotation



Transcript Lengths, all transcripts

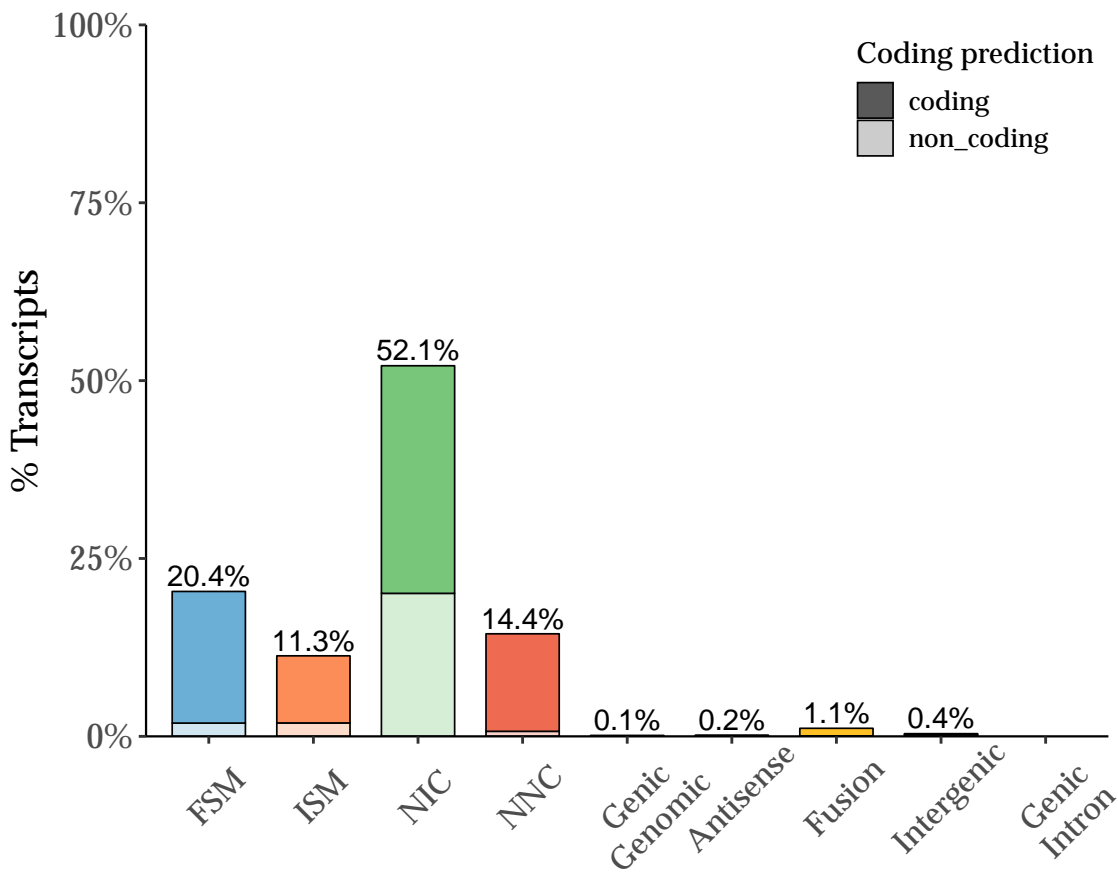


Transcript Lengths, Mono- vs Multi-Exons

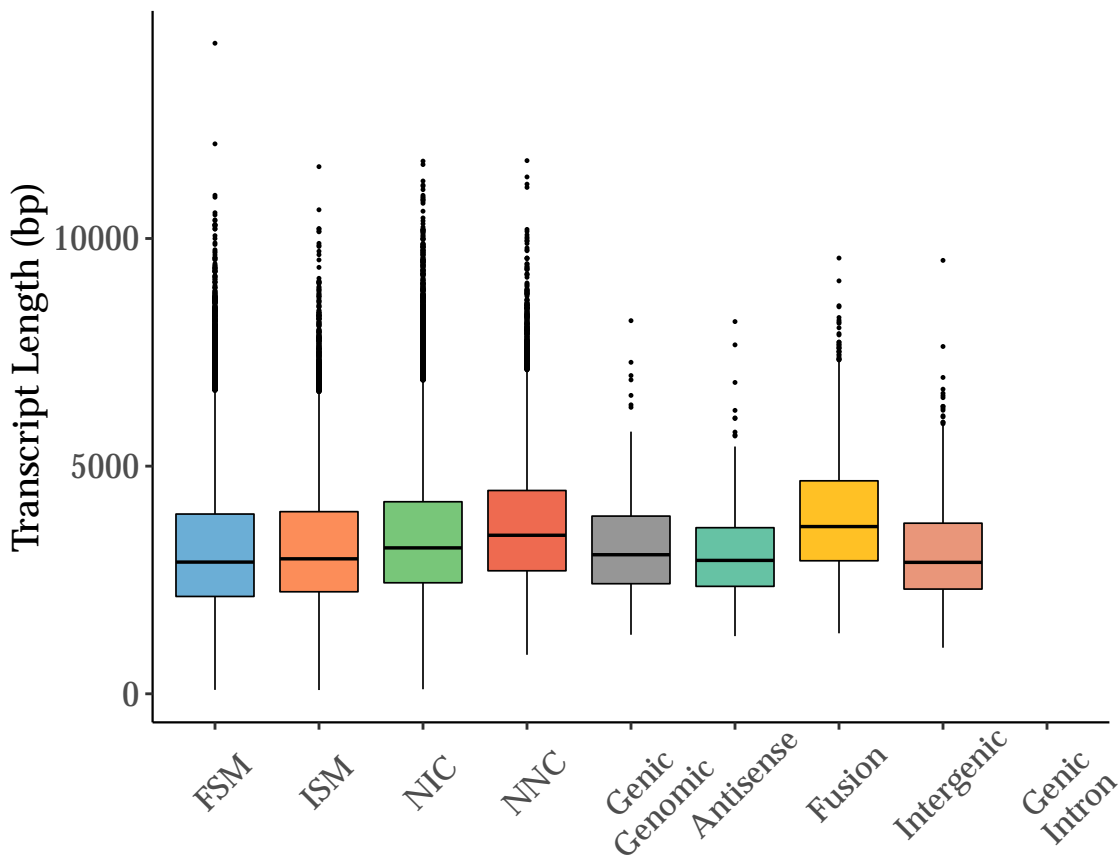


*Structural Isoform Characterization
by Splice Junctions*

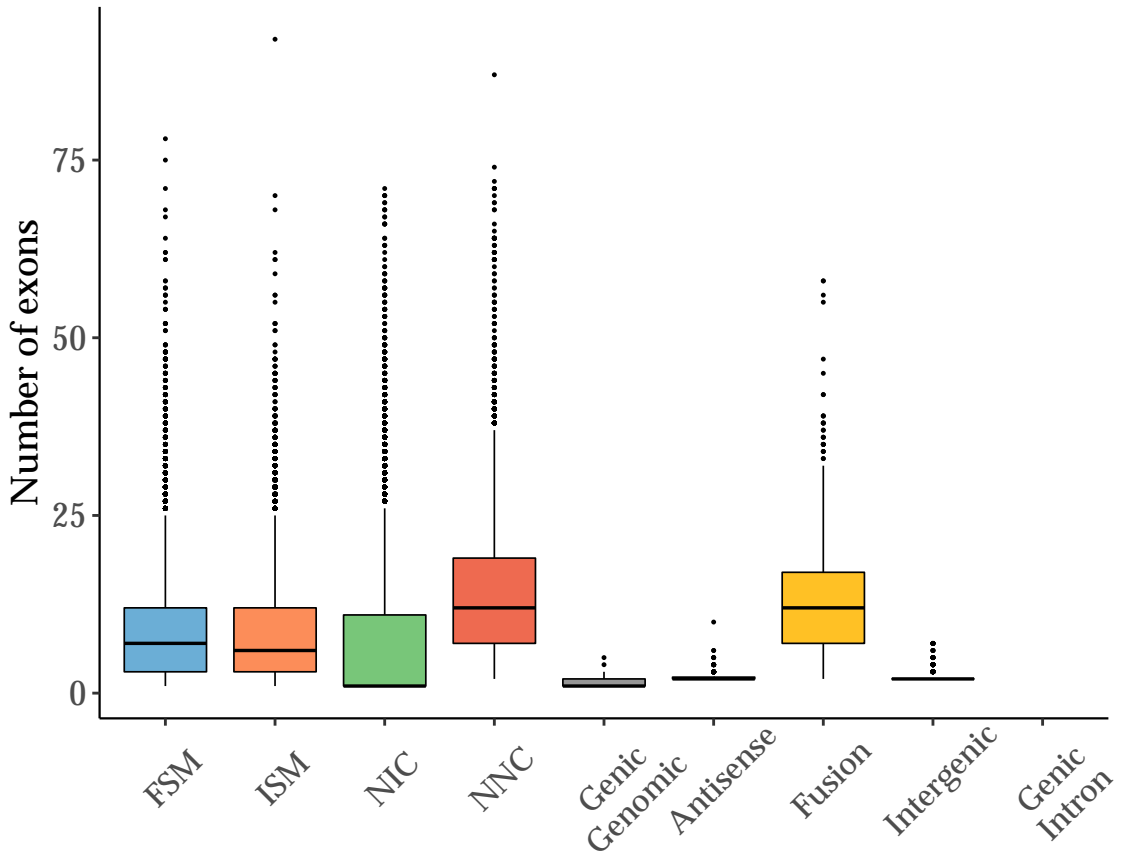
Isoform distribution across structural categories



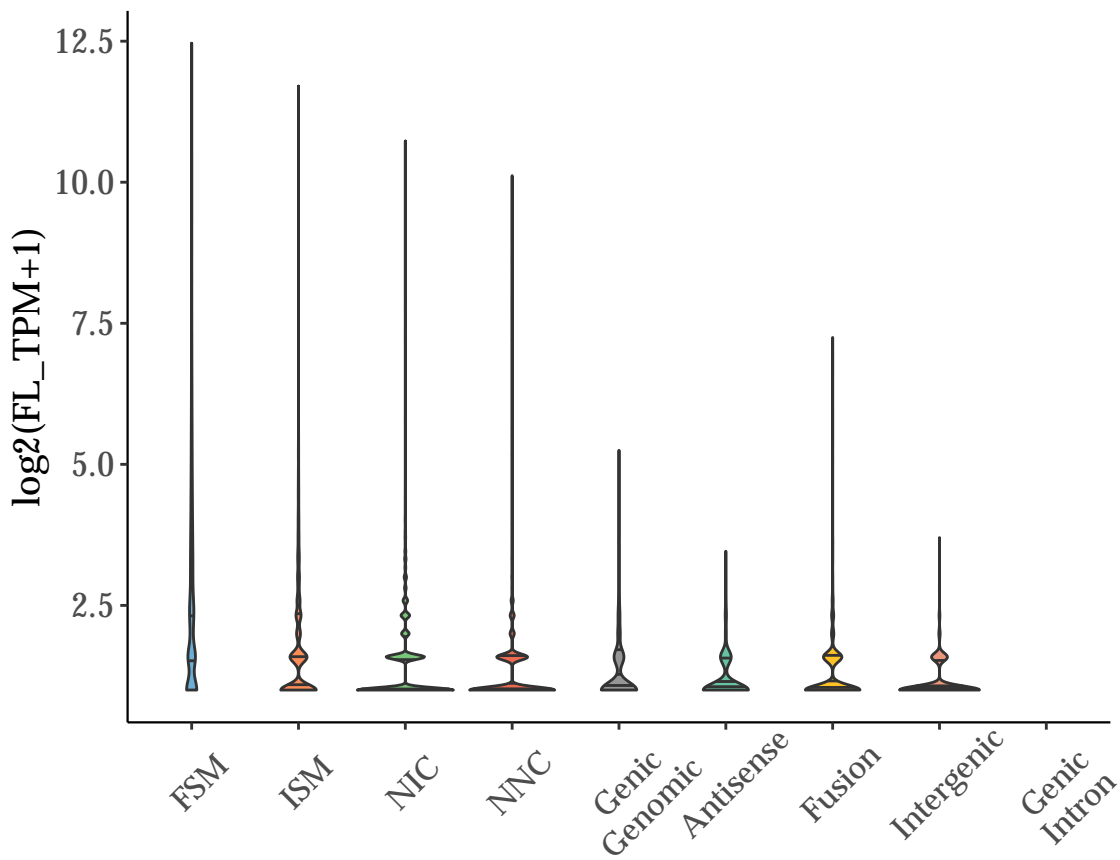
Transcript Lengths by Structural Classification



Exon Counts by Structural Classification



FL Count (normalized) by Structural Category

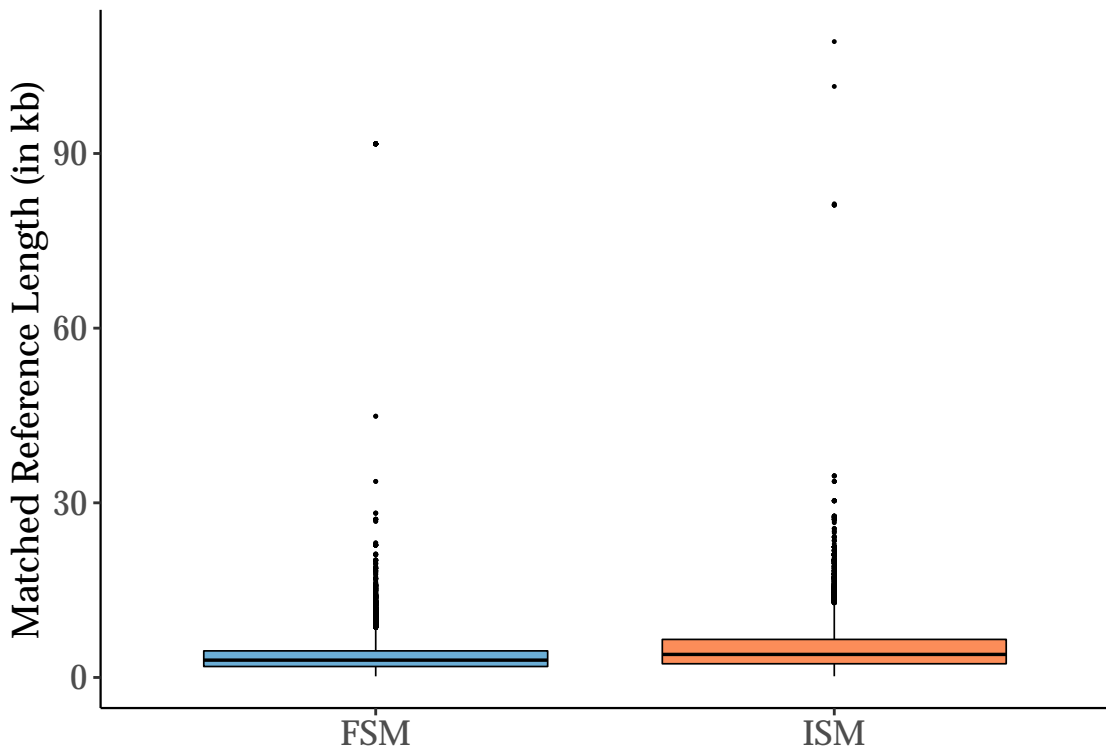


FL counts by category

category	isoforms	FL
FSM	33746	988391
ISM	18754	133048
NIC	86309	411016
NNC	23889	98363
Genic Genomic	220	824
Antisense	260	669
Fusion	1880	6285
Intergenic	626	1442

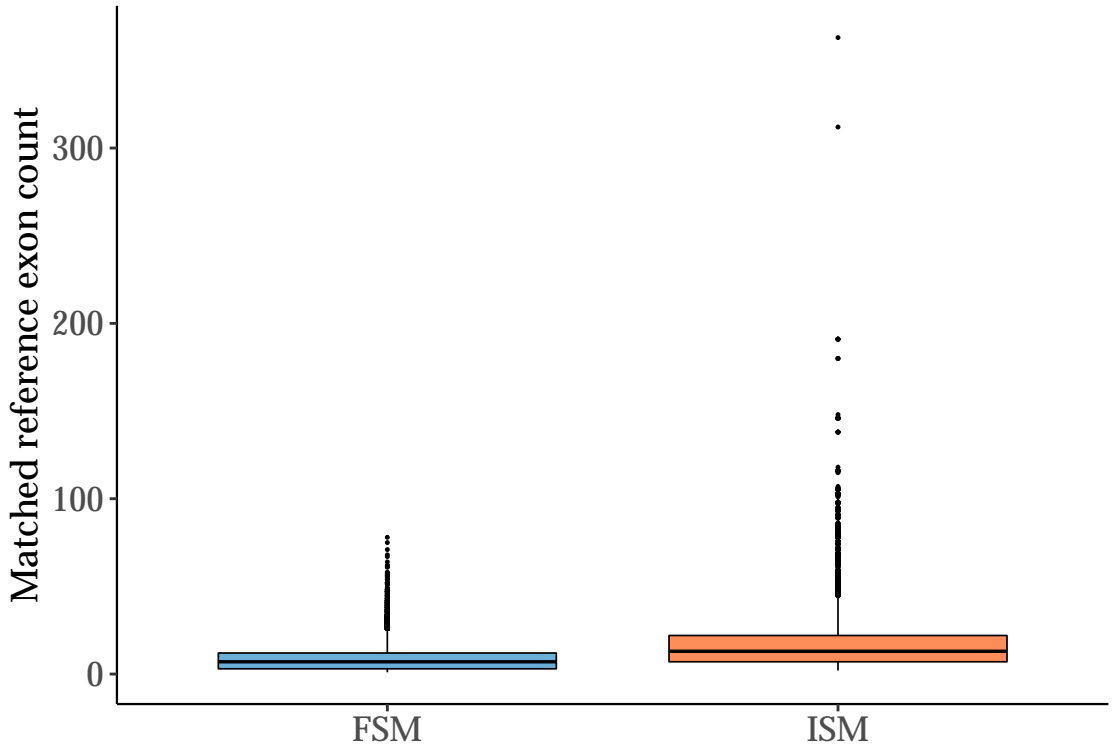
Length Distribution of Matched Reference Transcripts

Applicable only to FSM and ISM categories



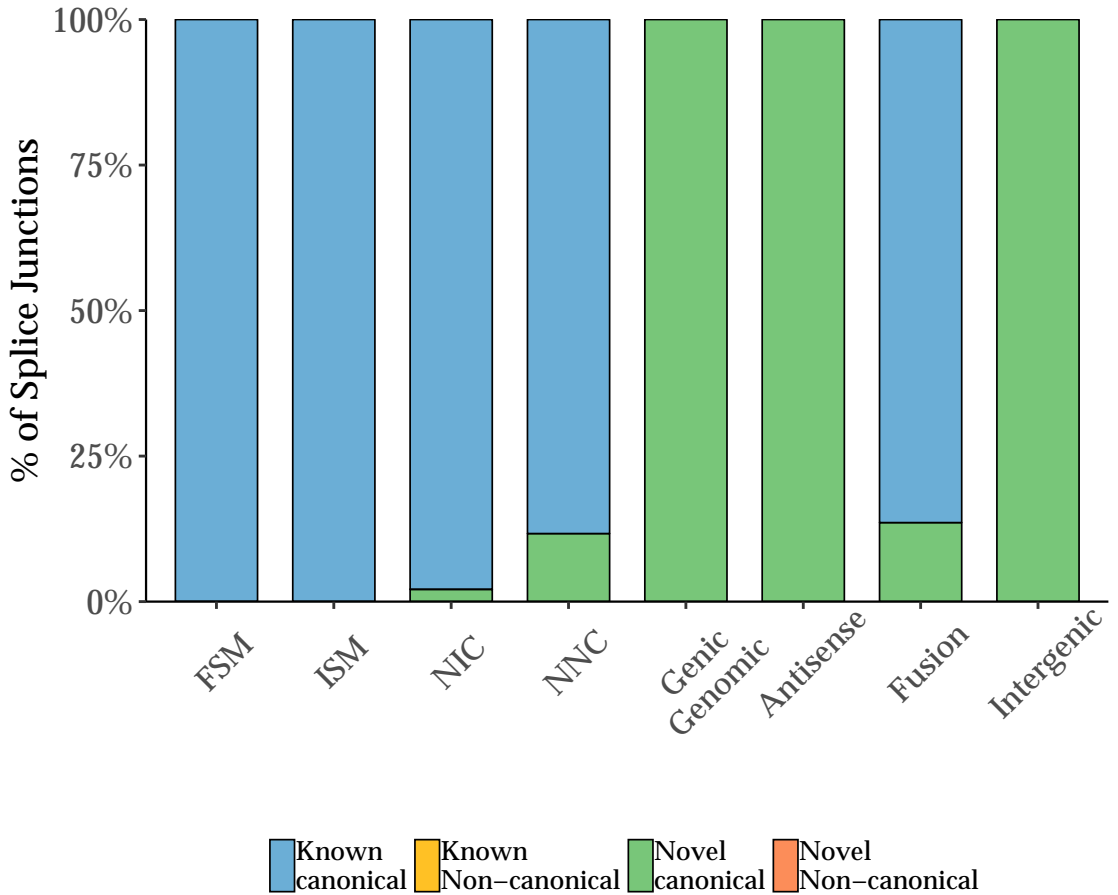
Exon Count Distribution of Matched Reference Transcripts

Applicable only to FSM and ISM categories

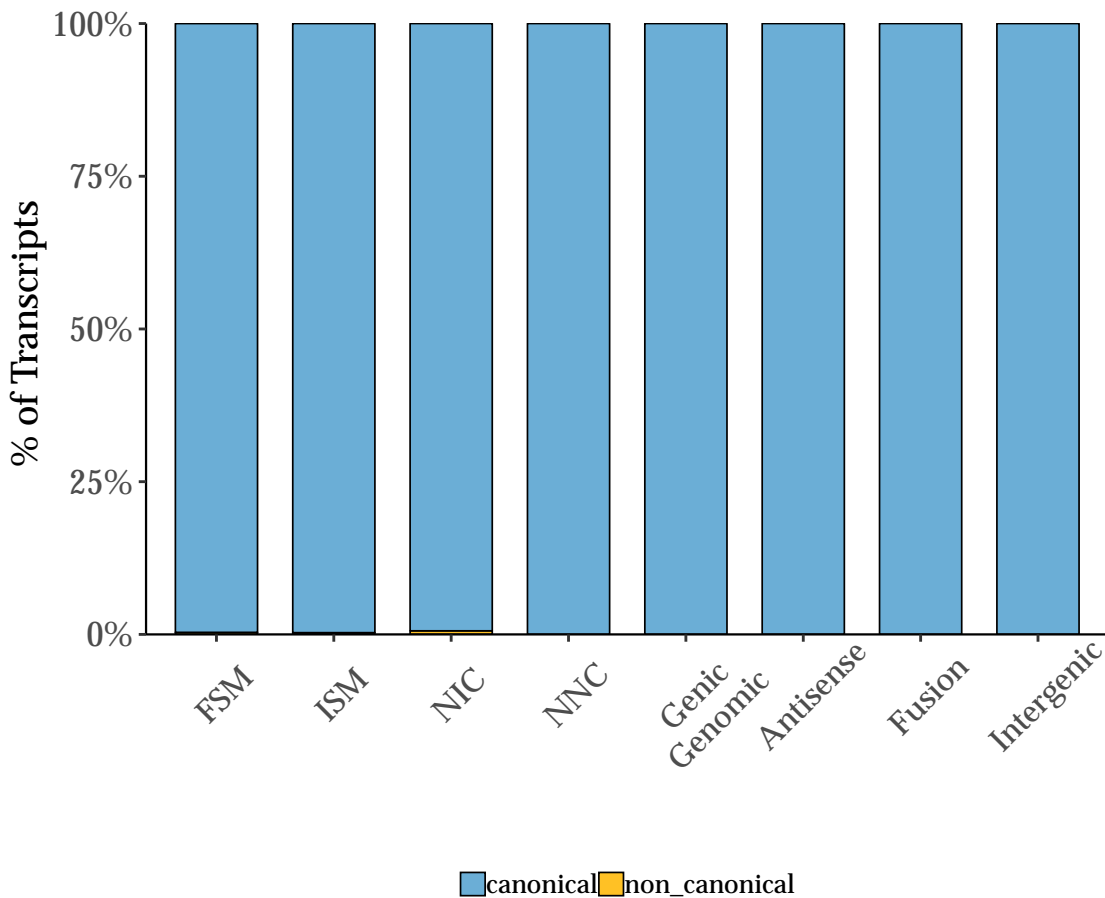


Splice Junction Characterization

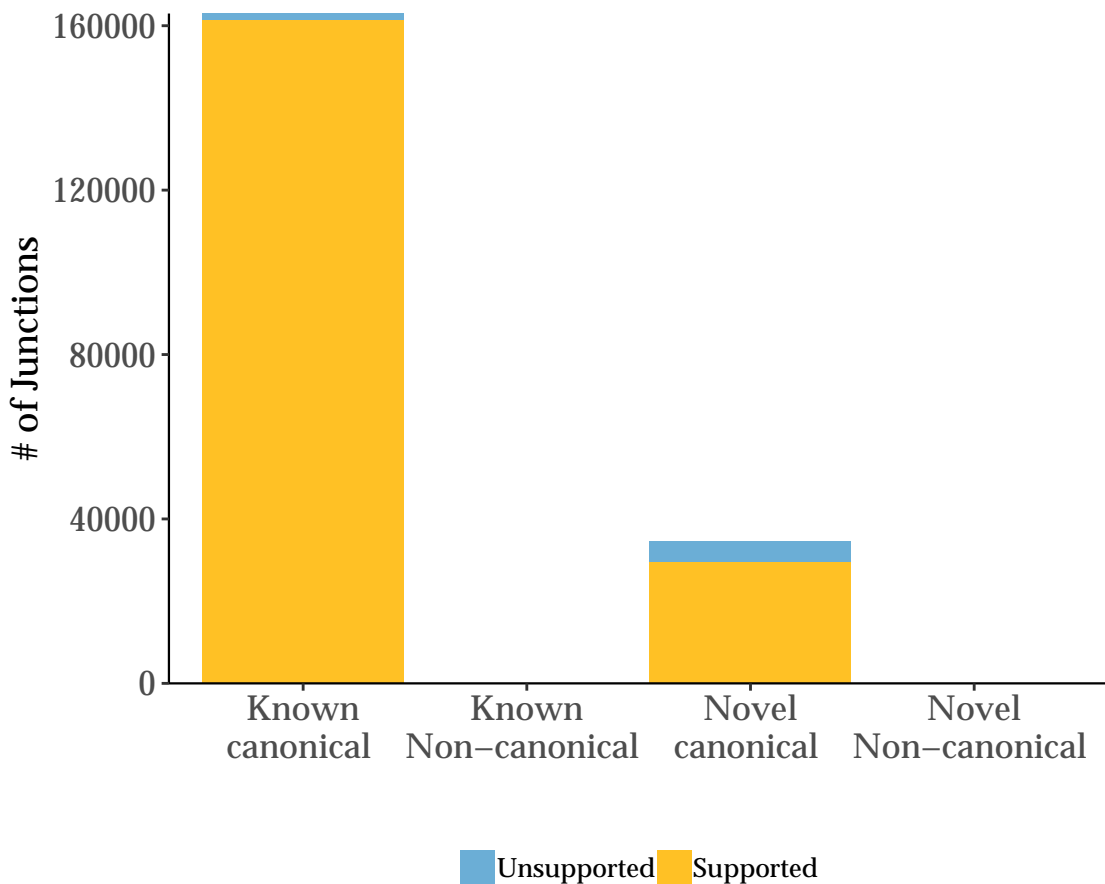
Distribution of Splice Junctions by Structural Classification



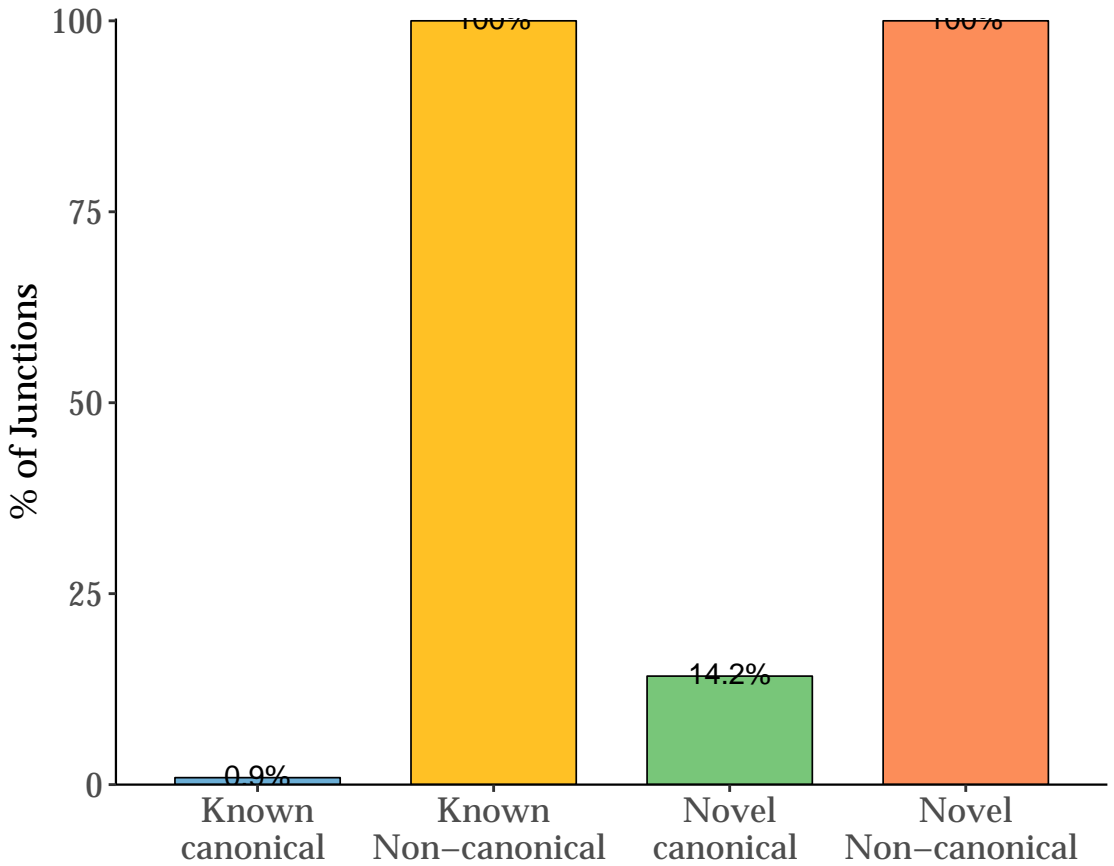
Distribution of Transcripts by Splice Junctions



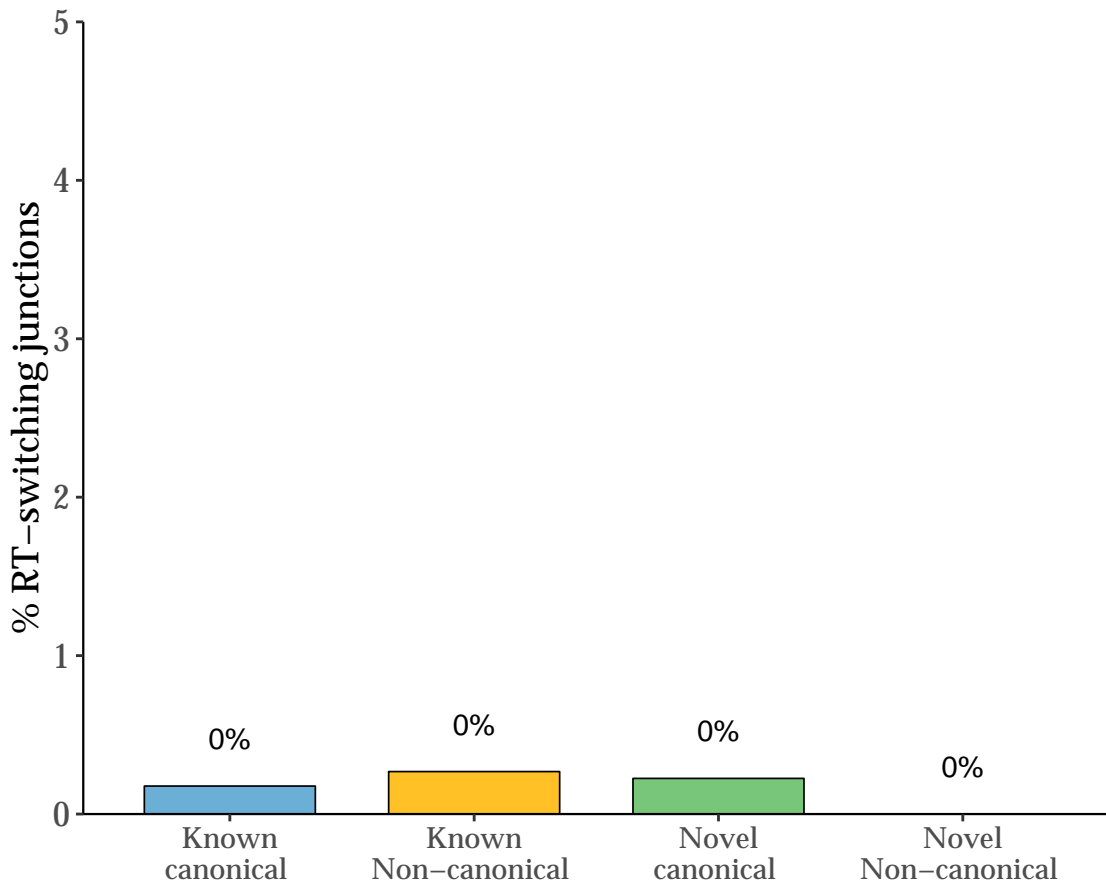
Unique junctions w/ or w/out short read coverage



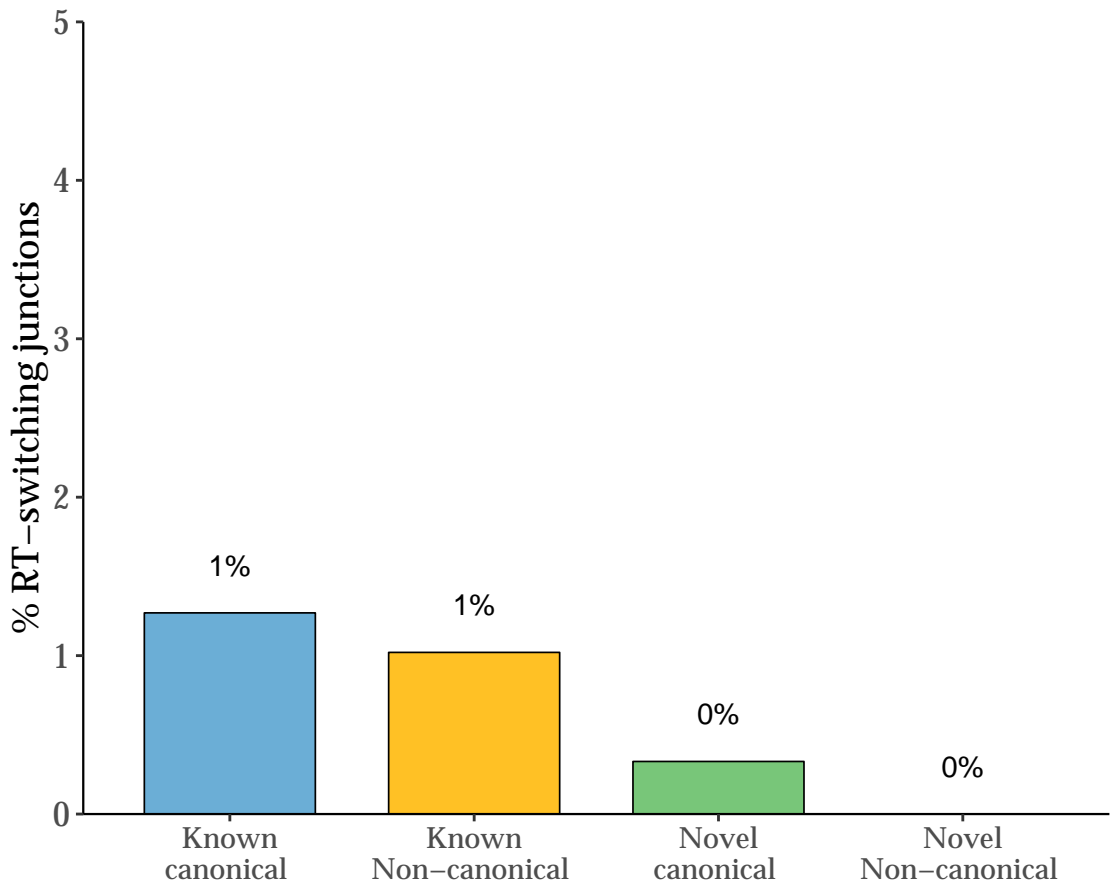
Unique junctions w/out short read coverage (percentage)



RT-switching, all junctions



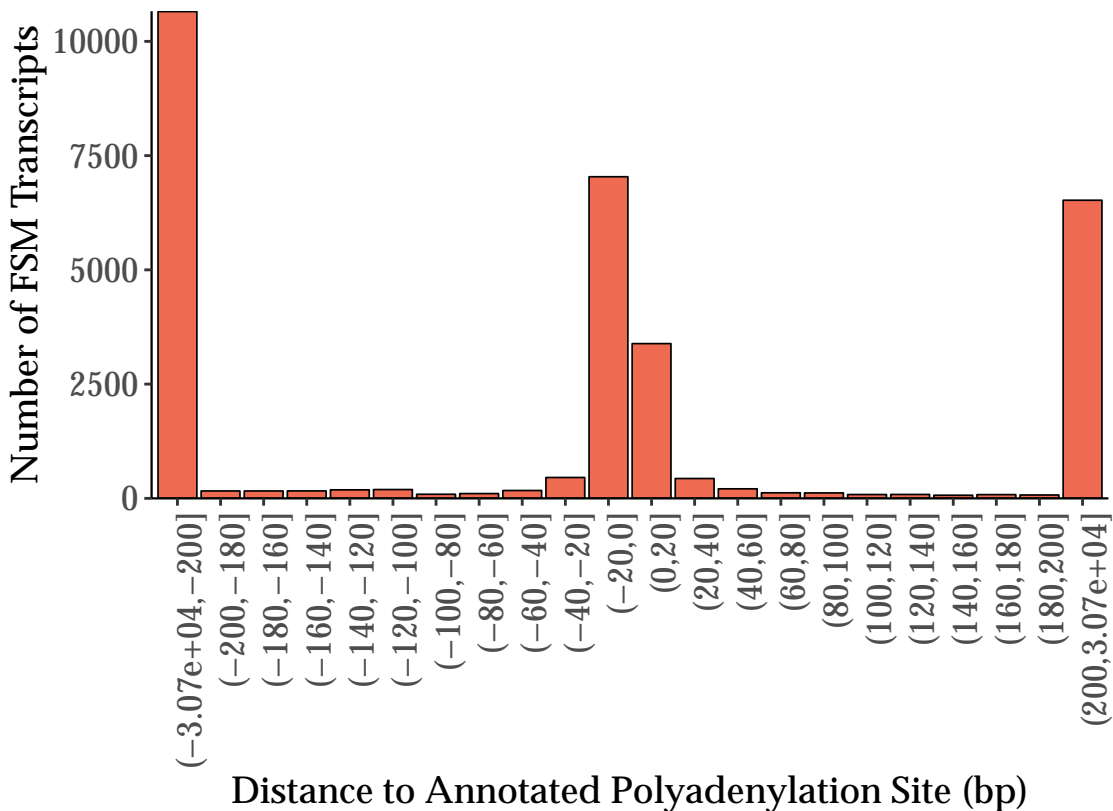
RT-switching, unique junctions



Comparison with Annotated TSS and PolyA Sites

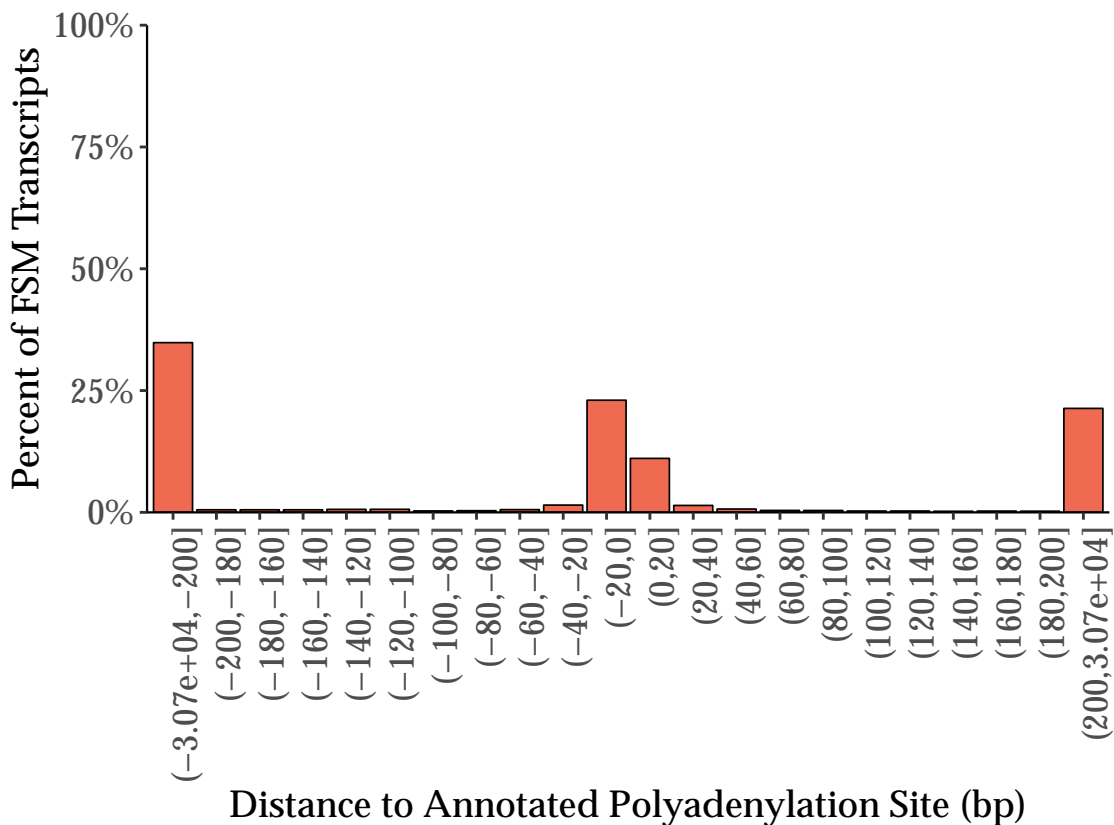
Distance to Annotated Polyadenylation Site, FSM only

Negative values indicate upstream of annotated polyA site



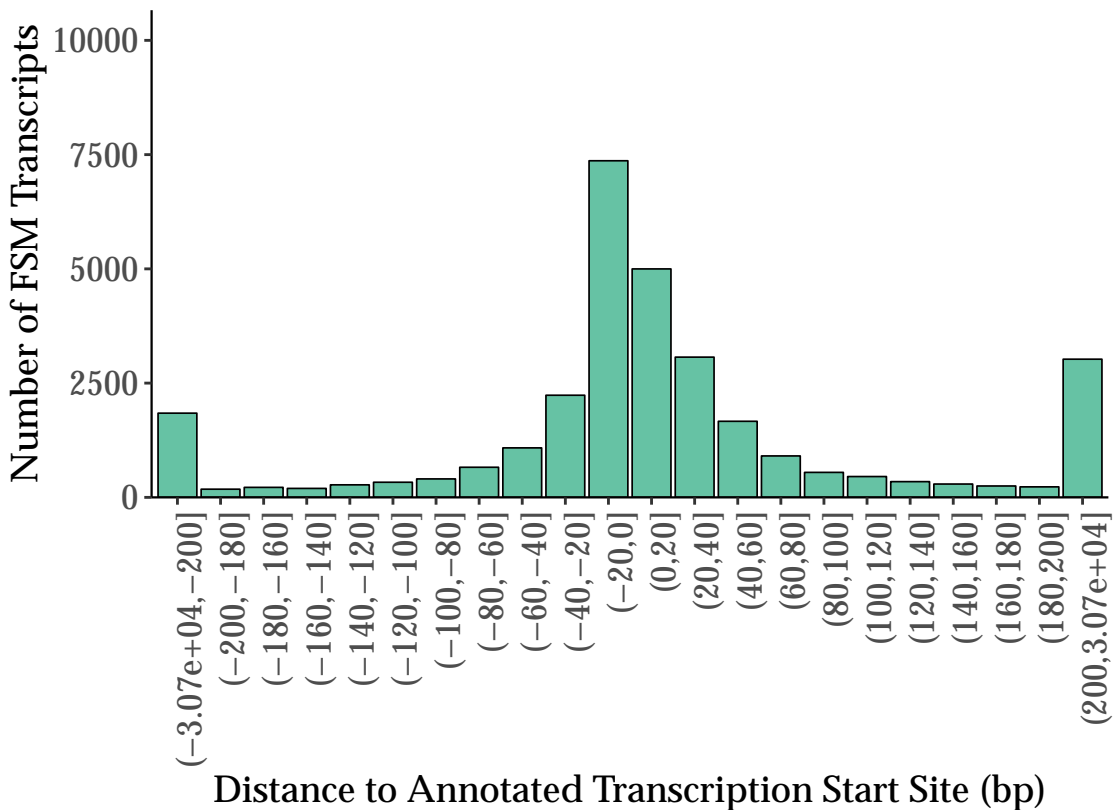
Distance to Annotated Polyadenylation Site, FSM only

Negative values indicate upstream of annotated polyA site



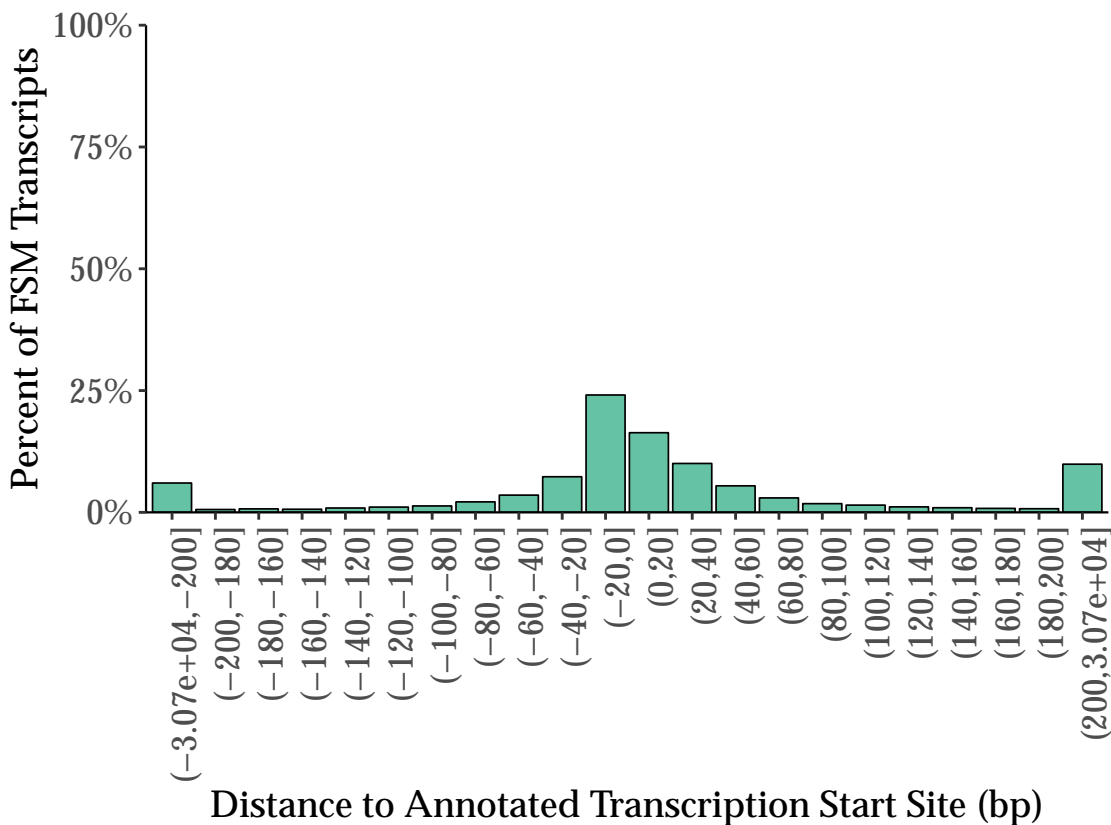
Distance to Annotated Transcription Start Site, FSM only

Negative values indicate downstream of annotated TSS



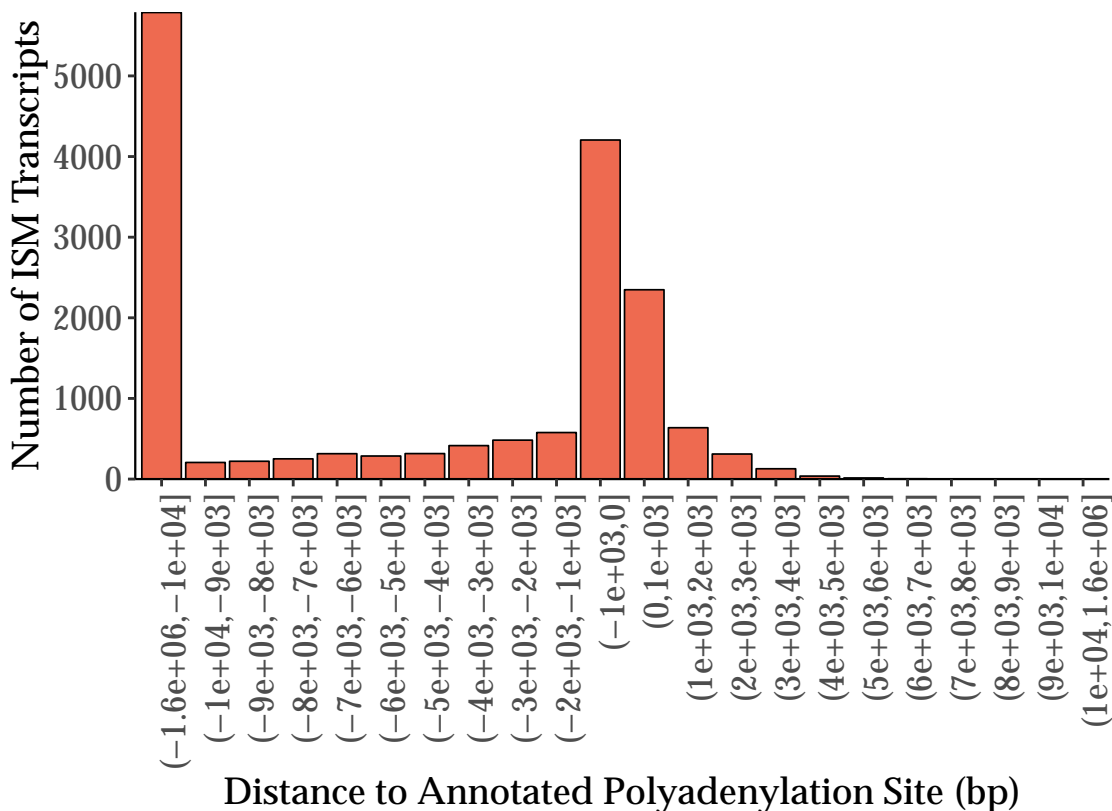
Distance to Annotated Transcription Start Site, FSM only

Negative values indicate downstream of annotated TSS



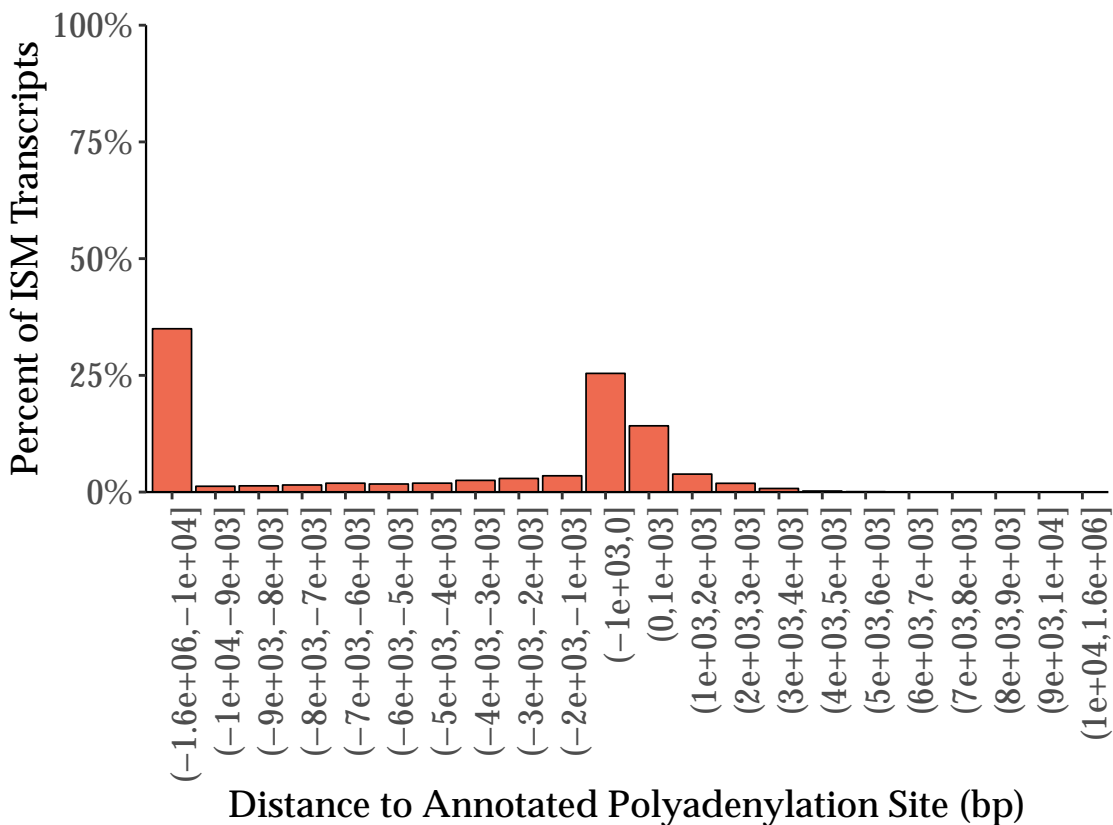
Distance to Annotated Polyadenylation Site, ISM only

Negative values indicate upstream of annotated polyA site



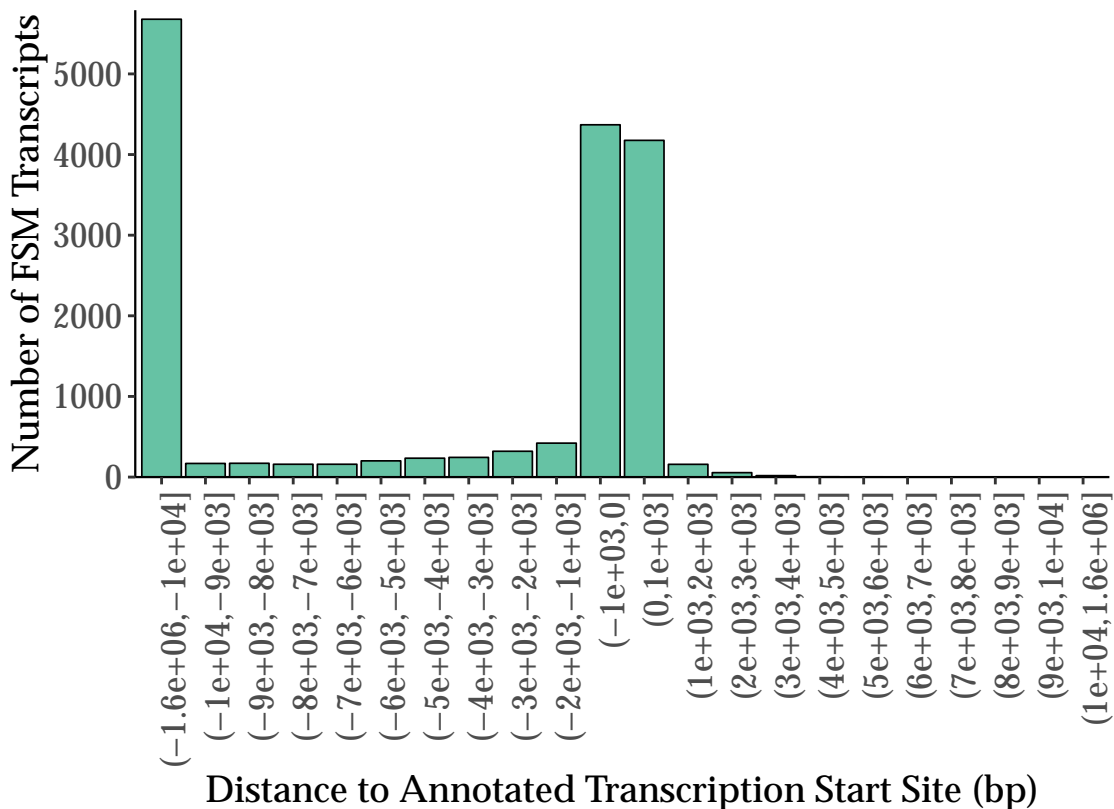
Distance to Annotated Polyadenylation Site, ISM only

Negative values indicate upstream of annotated polyA site



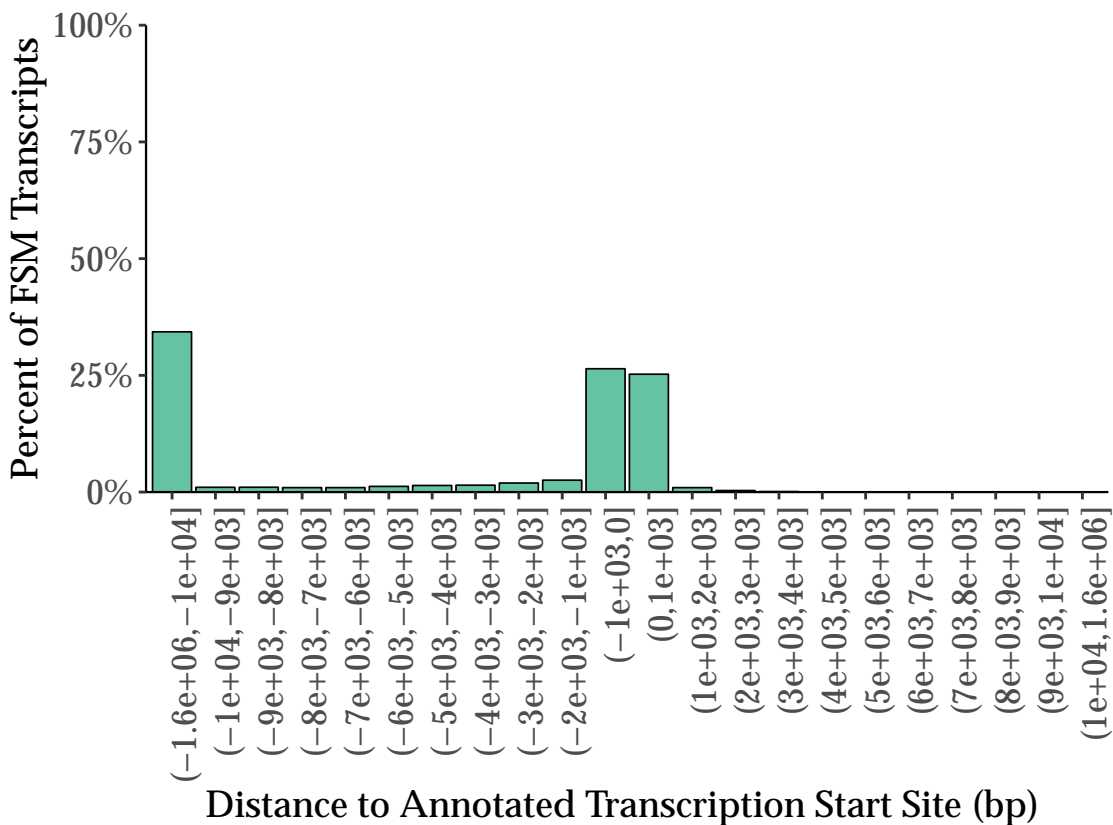
Distance to Annotated Transcription Start Site, ISM only

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

Motif	Count	%
AATAAA	59212	57.9
ATTAAA	14680	14.3
AGTAAA	3518	3.4
TATAAA	3393	3.3
TTTAAA	3079	3.0
AAGAAA	2703	2.6
AATATA	2333	2.3
AAAAAG	2300	2.2
AATACA	1948	1.9
GATAAA	1840	1.8
CATAAA	1835	1.8
AATGAA	1589	1.6
AAAACA	1342	1.3
AATAGA	1200	1.2
ACTAAA	1083	1.1
GGGGCT	250	0.2

Number of polyA Motifs Detected

Category	Count	polyA Detected	%
FSM	33746	24296	72
ISM	18754	11597	62
NIC	86309	47145	55
NNC	23889	17745	74
Genic			
Genomic	220	110	50
Antisense	260	98	38
Fusion	1880	1137	60
Intergenic	626	177	28

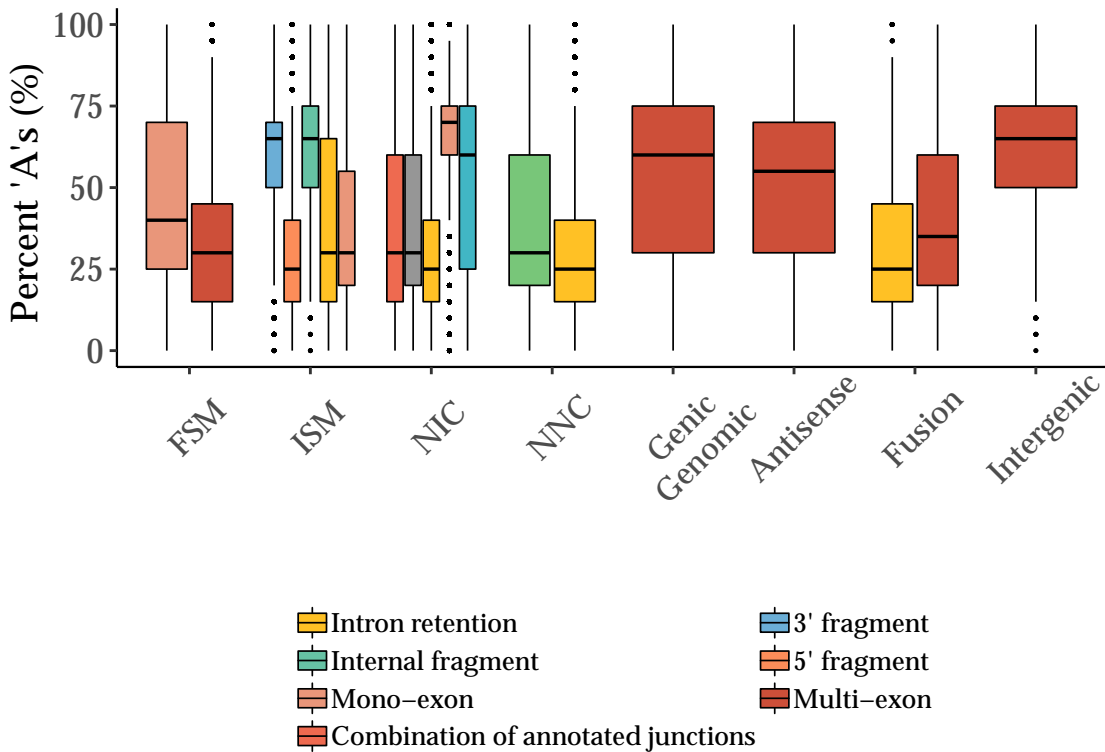
Number of close by CAGE Peaks Detected

Category	Count	Has CAGE peak within 50bp	%
FSM	33746	24278	72
ISM	18754	7277	39
NIC	86309	31215	36
NNC	23889	14524	61
Genic Genomic	220	72	33
Antisense	260	22	8
Fusion	1880	1101	59
Intergenic	626	30	5

Intra-Priming Quality Check

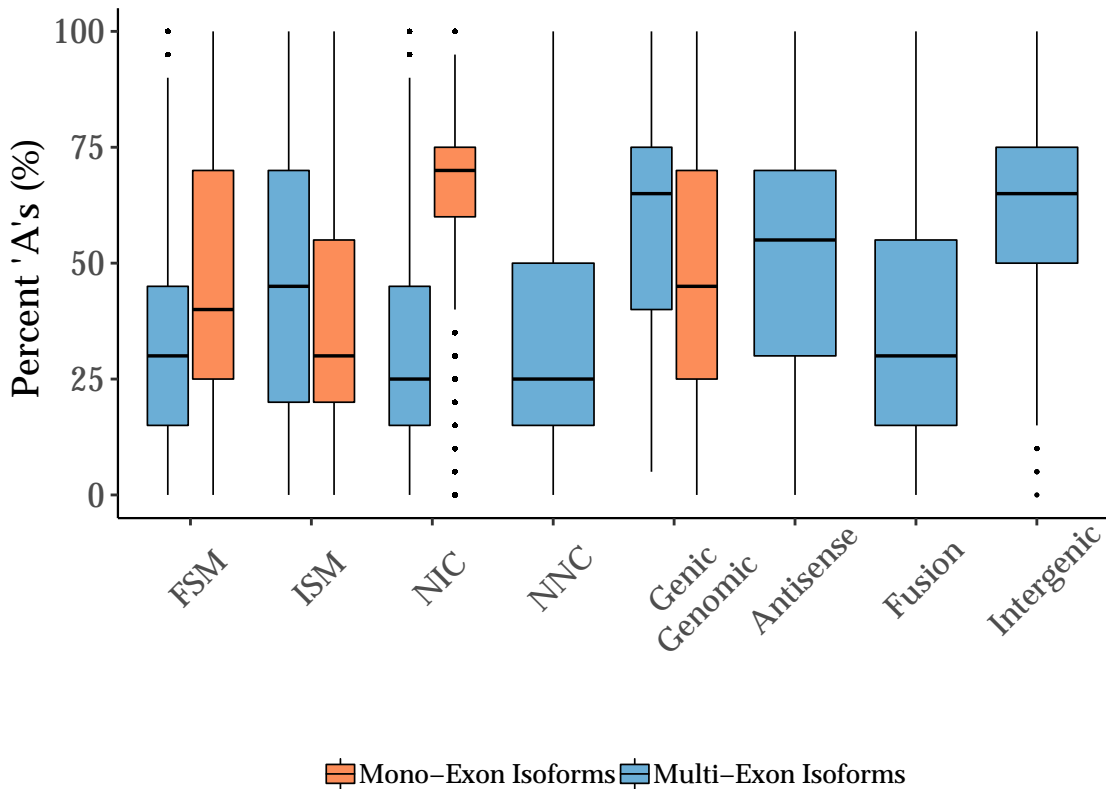
Possible Intra-Priming by Structural Category

Percent of genomic 'A's in downstream 20 bp



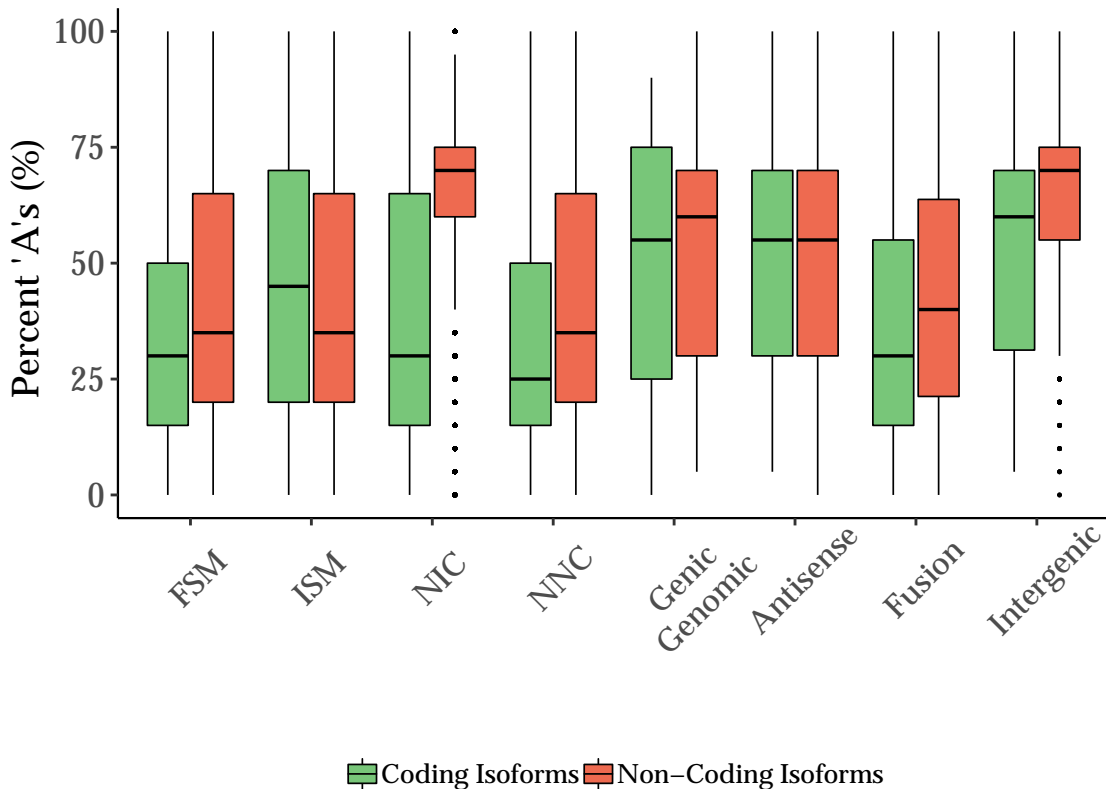
Possible Intra-Priming, Mono- vs Multi-Exon

Percent of genomic 'A's in downstream 20 bp



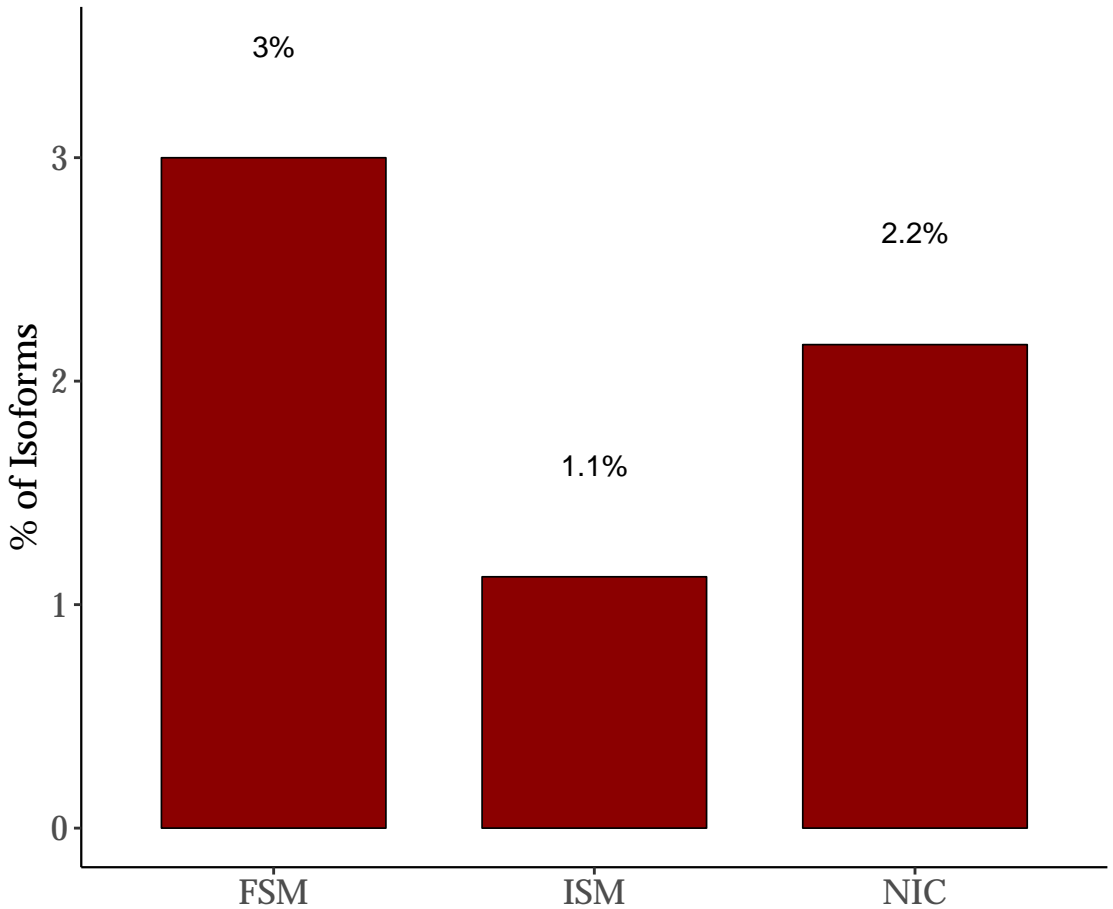
Possible Intra-Priming, Coding vs Non-Coding

Percent of genomic 'A's in downstream 20 bp



Quality Controls

Incidence of RT-switching



Incidence of Non-Canonical Junctions

