

SQANTI2 report

Unique Genes: 41200
Unique Isoforms: 262424

Gene classification

Category	# Genes
Annotated Genes	20874
Novel Genes	20326

*Characterization of transcripts
based on splice junctions*

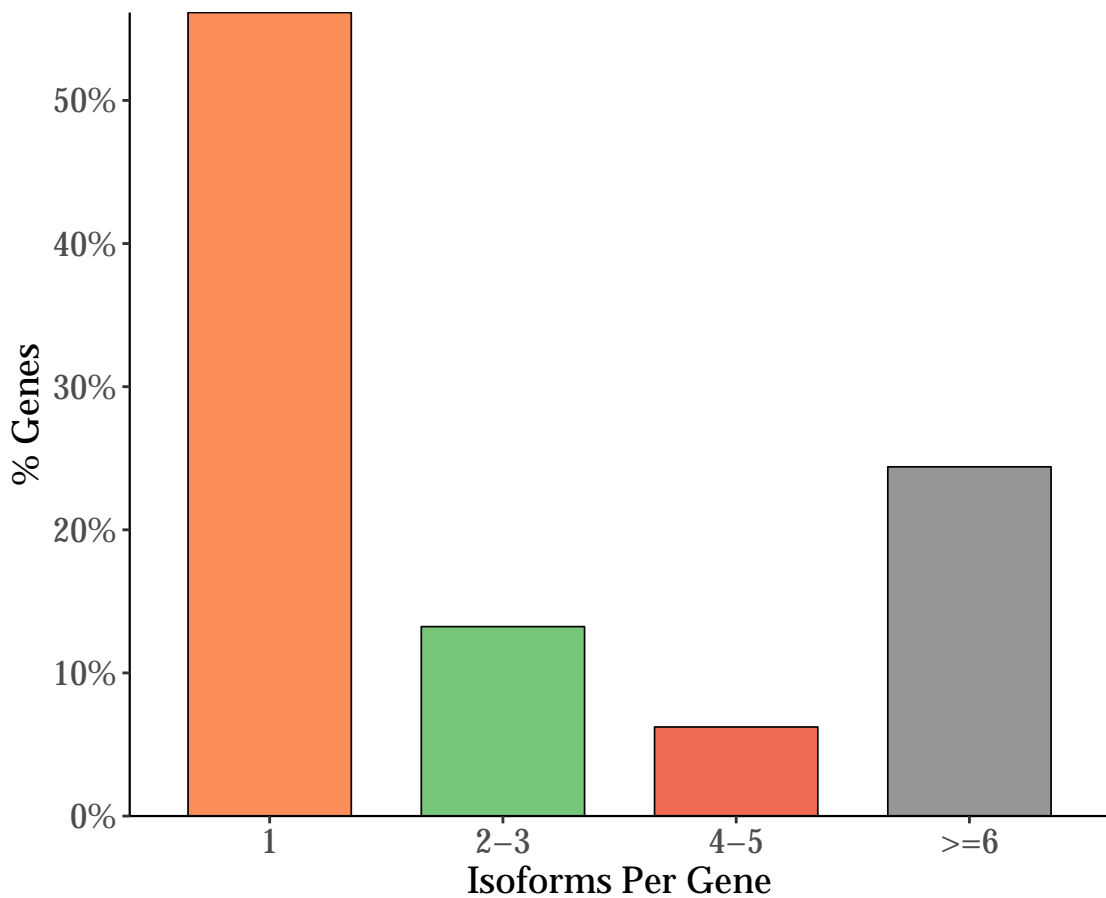
Category	# Isoforms	# Genes
FSM	36495	13096
ISM	23178	7568
NIC	121124	13188
NNC	47109	10584
Genic Genomic	4835	3466
Antisense	11807	5381
Fusion	2931	1503
Intergenic	14945	14945

Splice Junction Classification

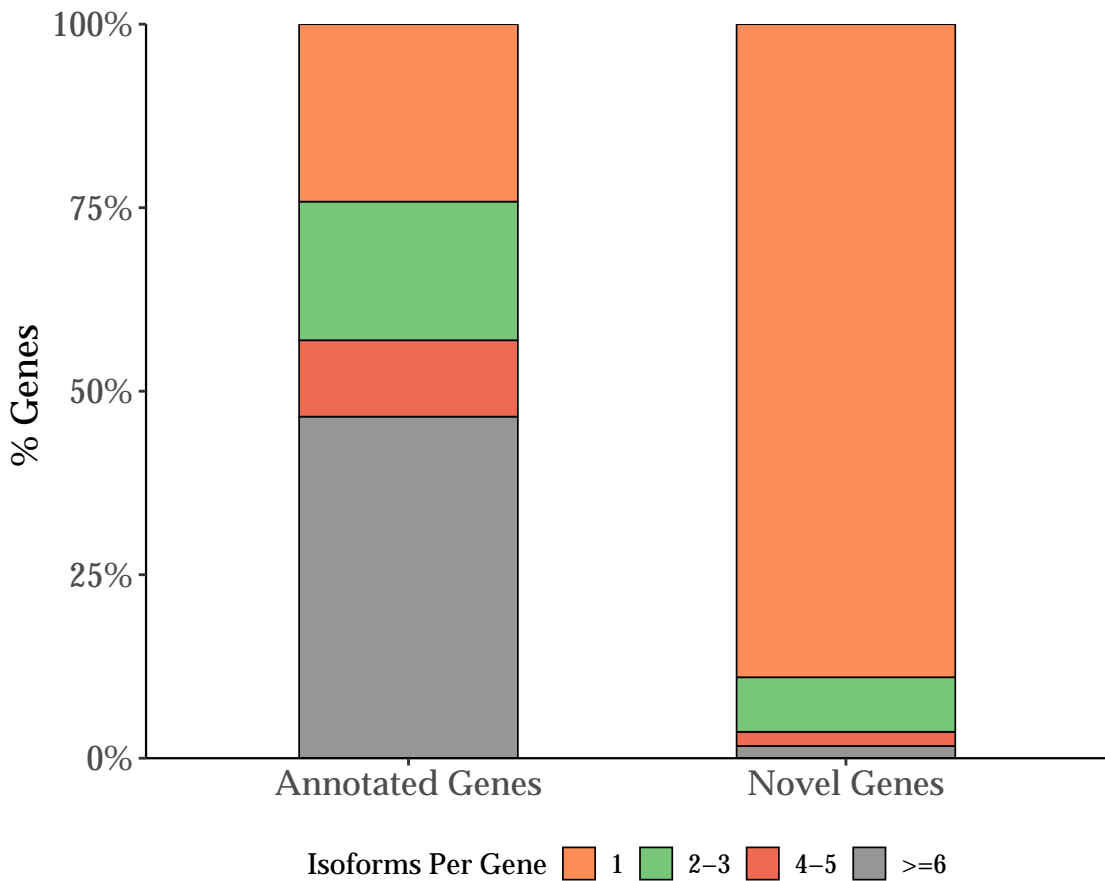
Category	# SJs	Percent
Known canonical	170956	69.94
Known Non-canonical	657	0.27
Novel canonical	44994	18.41
Novel Non-canonical	27827	11.38

Gene Characterization

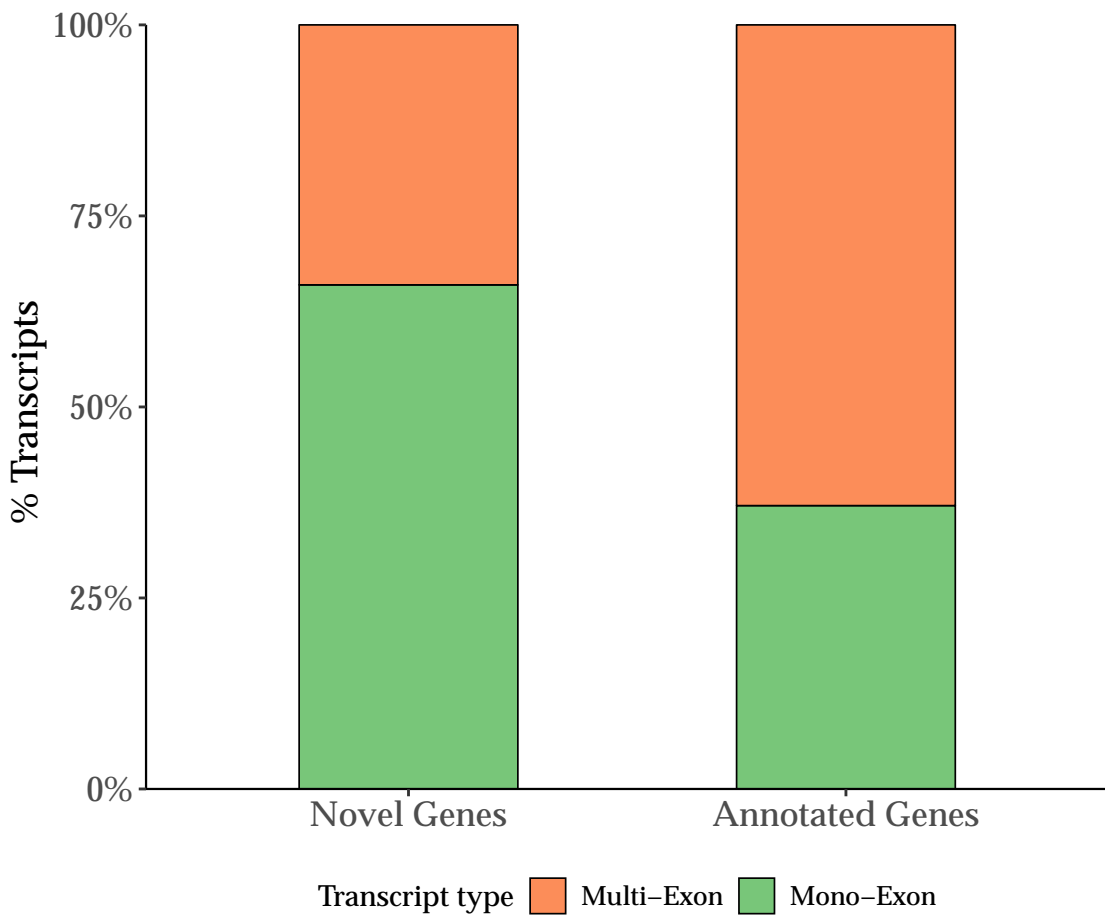
Number of Isoforms per Gene



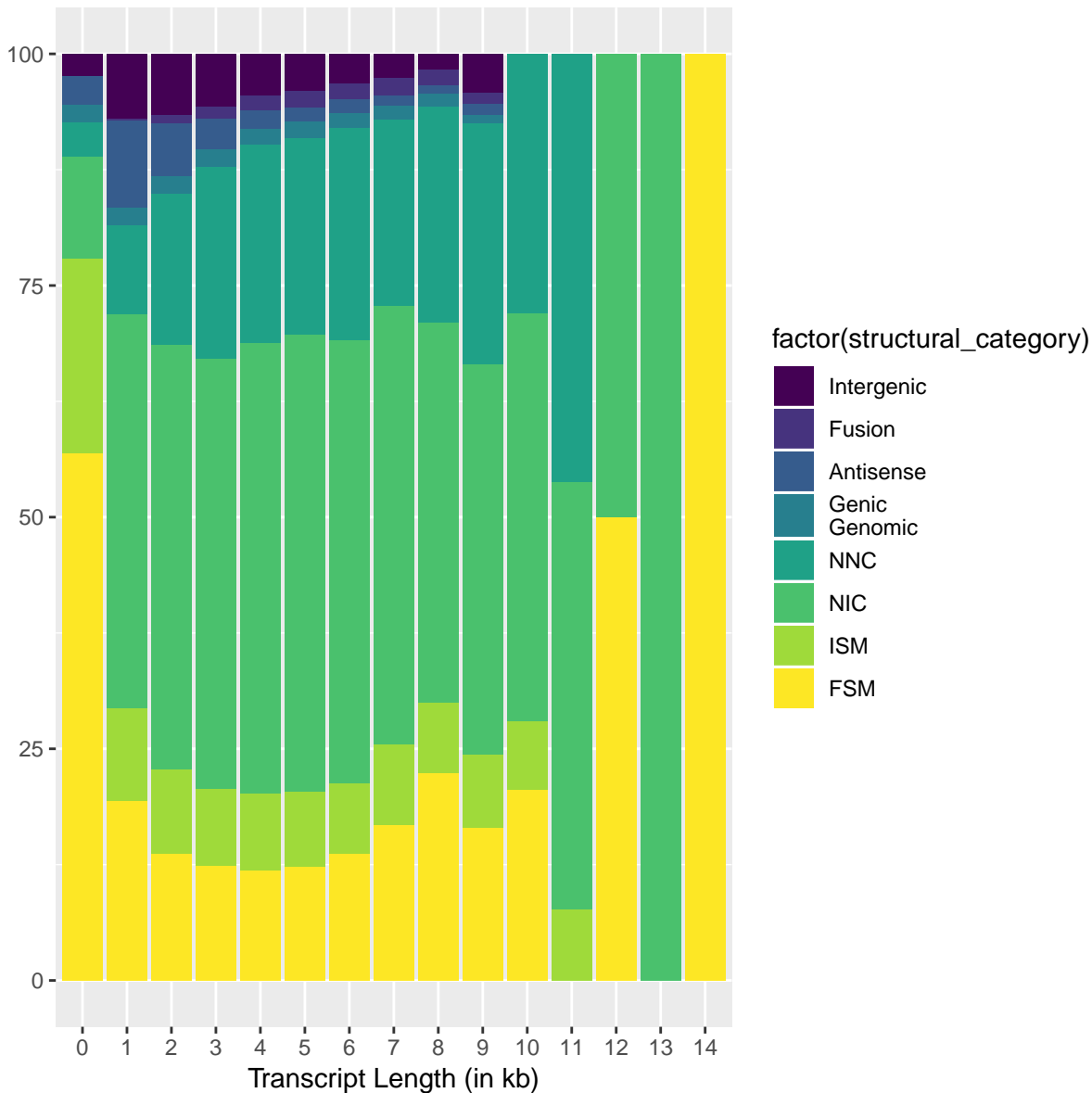
Number of Isoforms per Gene, Novel vs Known Genes



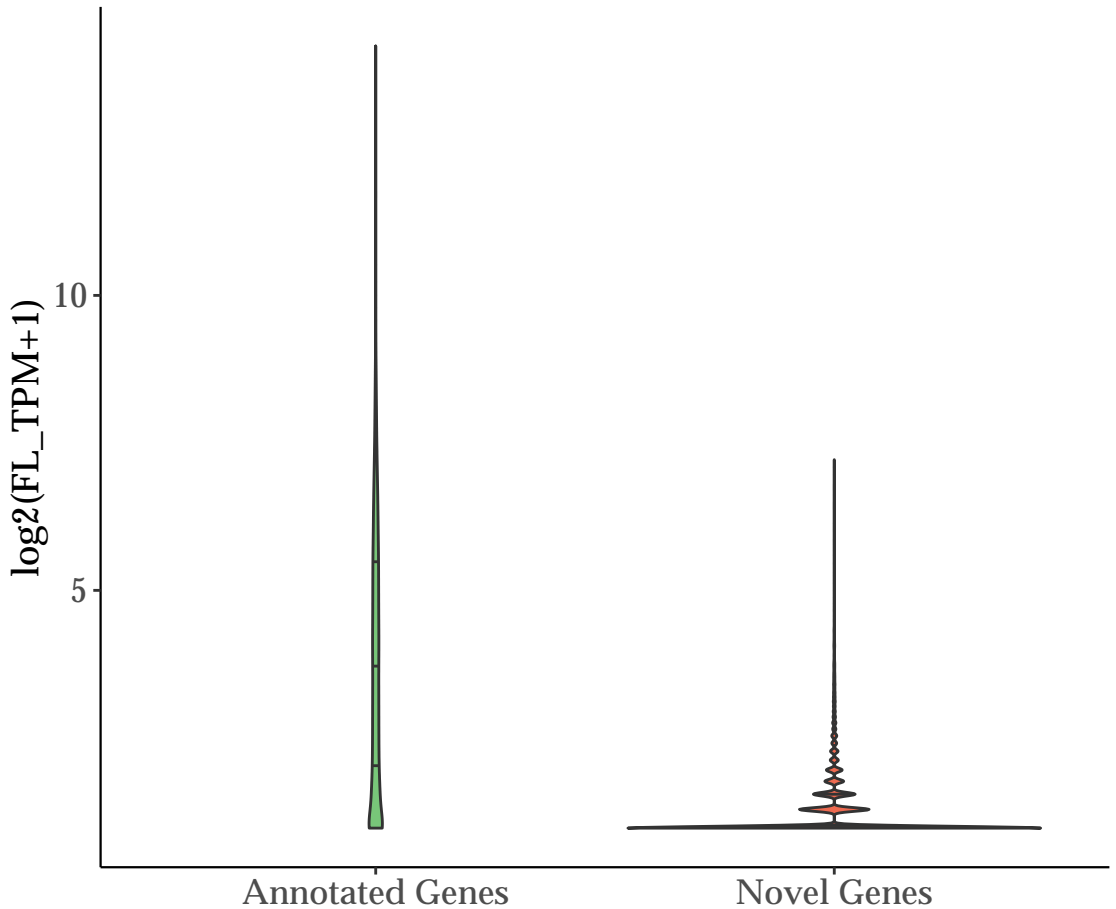
Distribution of Mono- vs Multi-Exon Transcripts



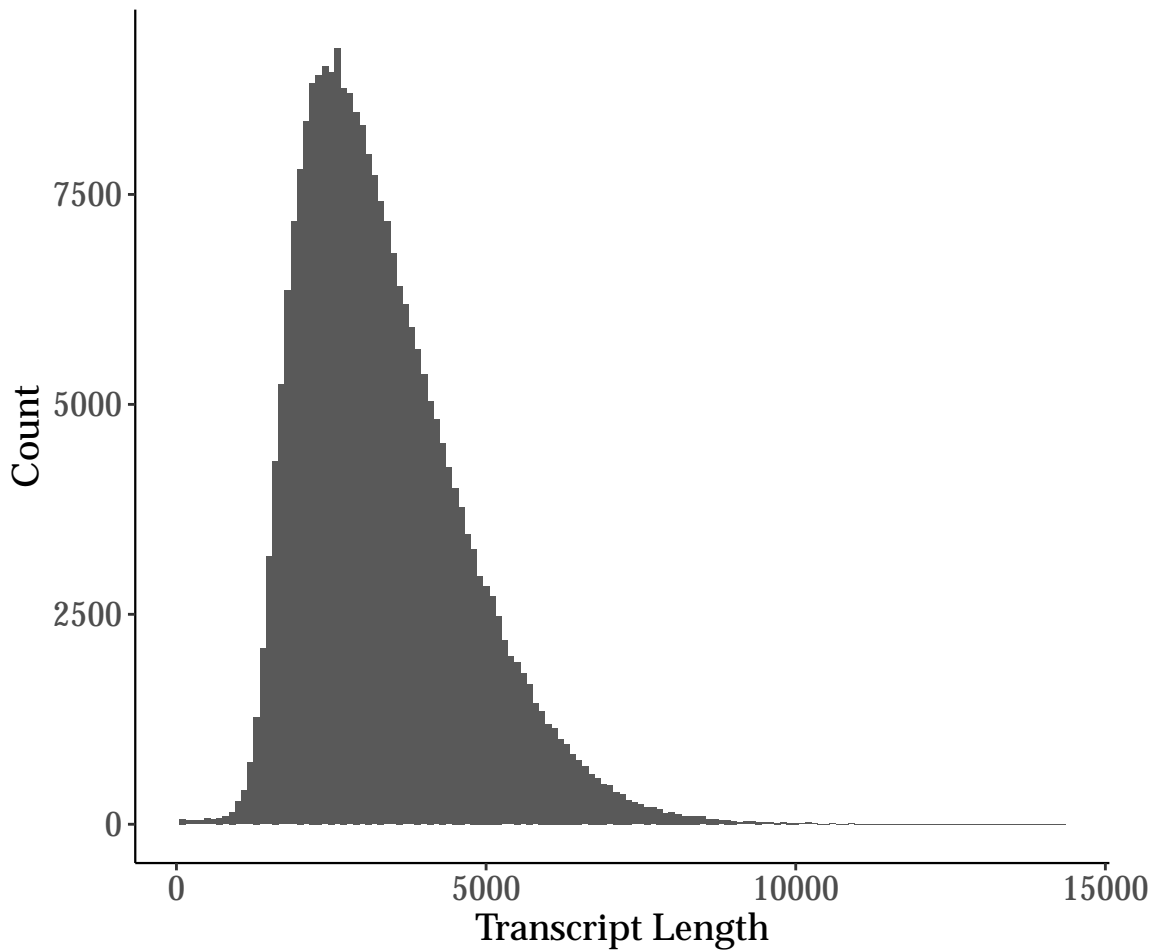
Classifications by Transcript Length, normalized



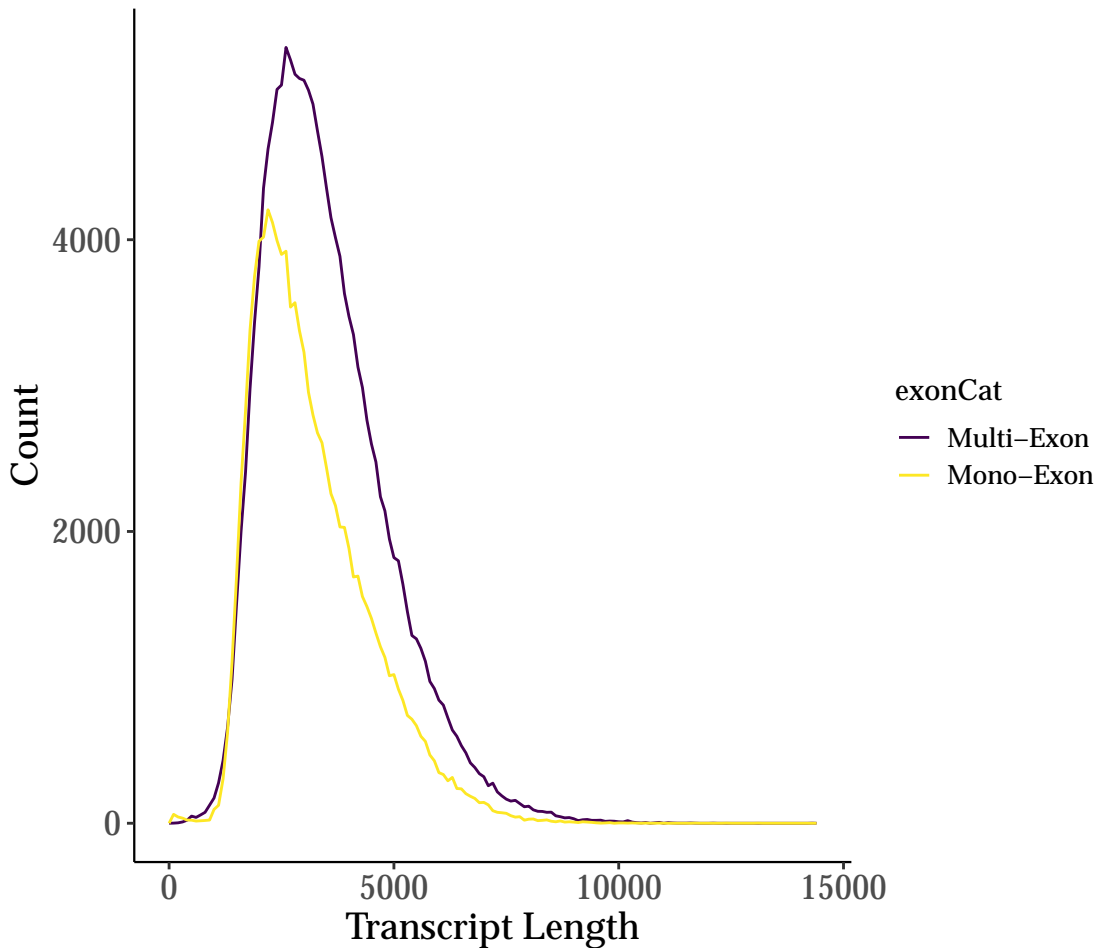
Number of FL reads per Gene by type of gene annotation



Transcript Lengths, all transcripts

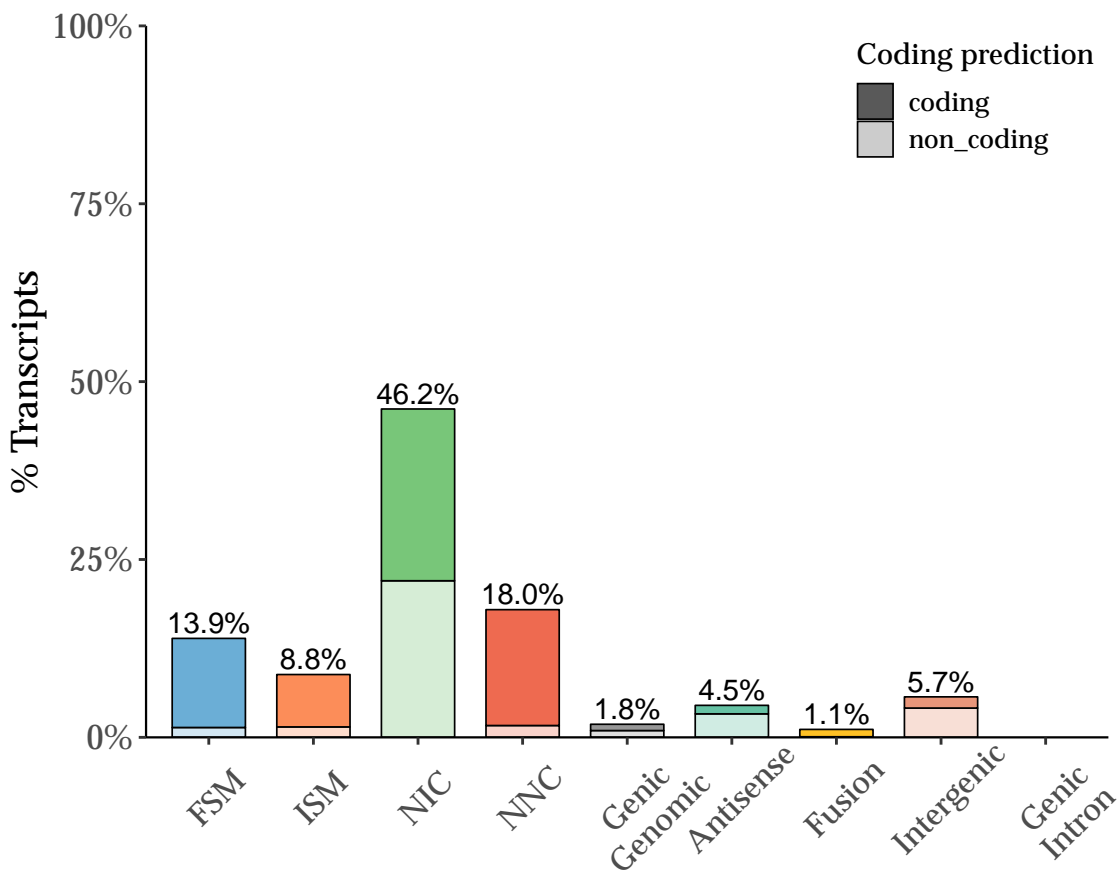


Transcript Lengths, Mono- vs Multi-Exons

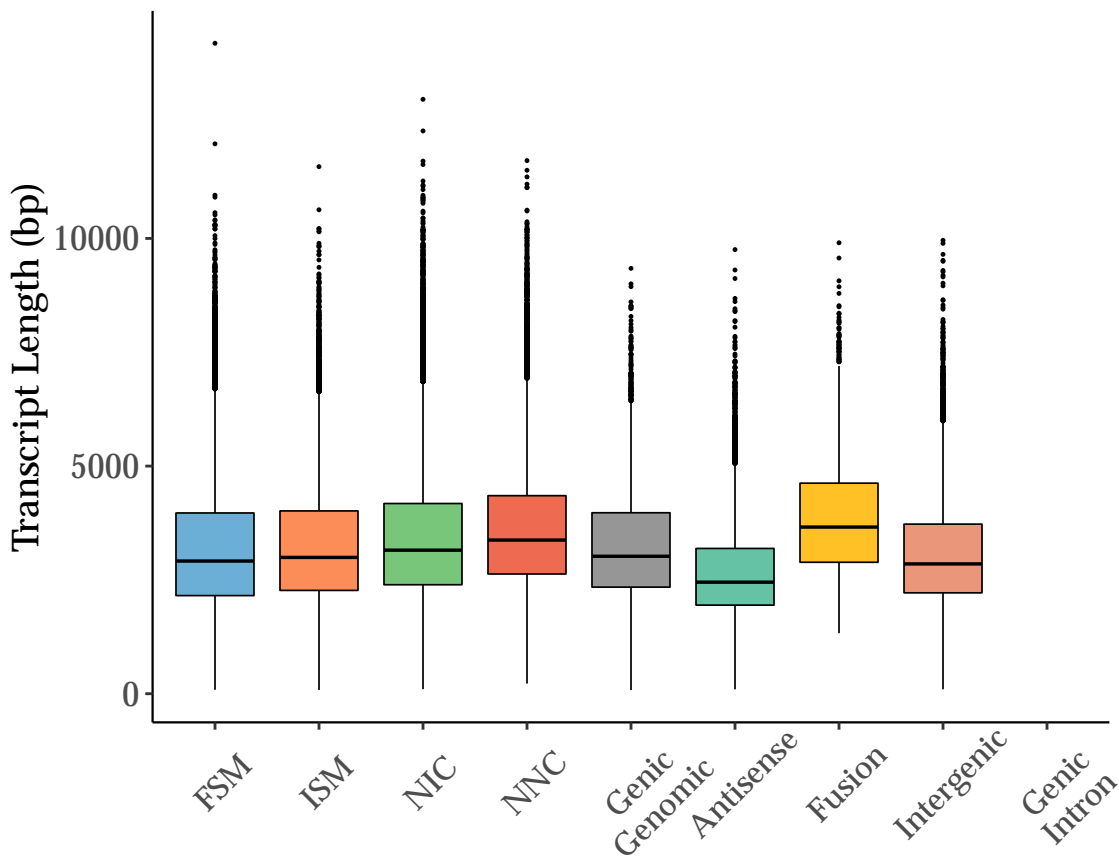


*Structural Isoform Characterization
by Splice Junctions*

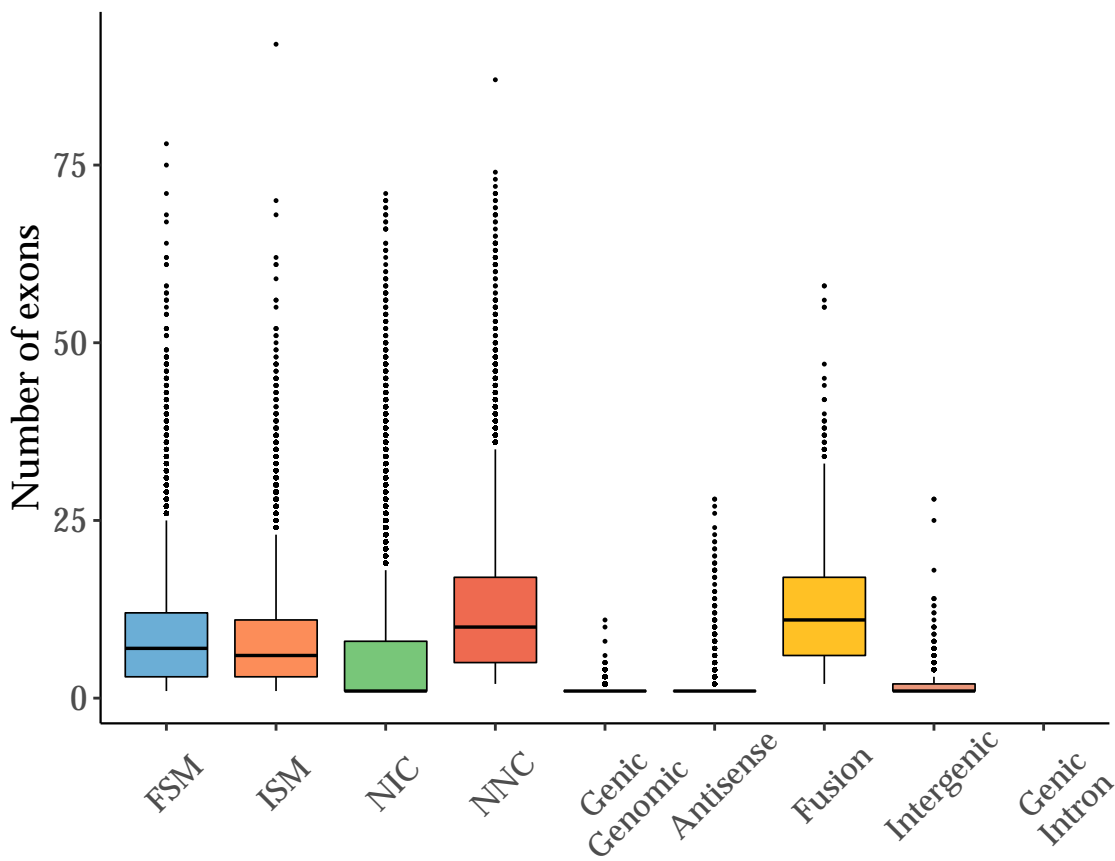
Isoform distribution across structural categories



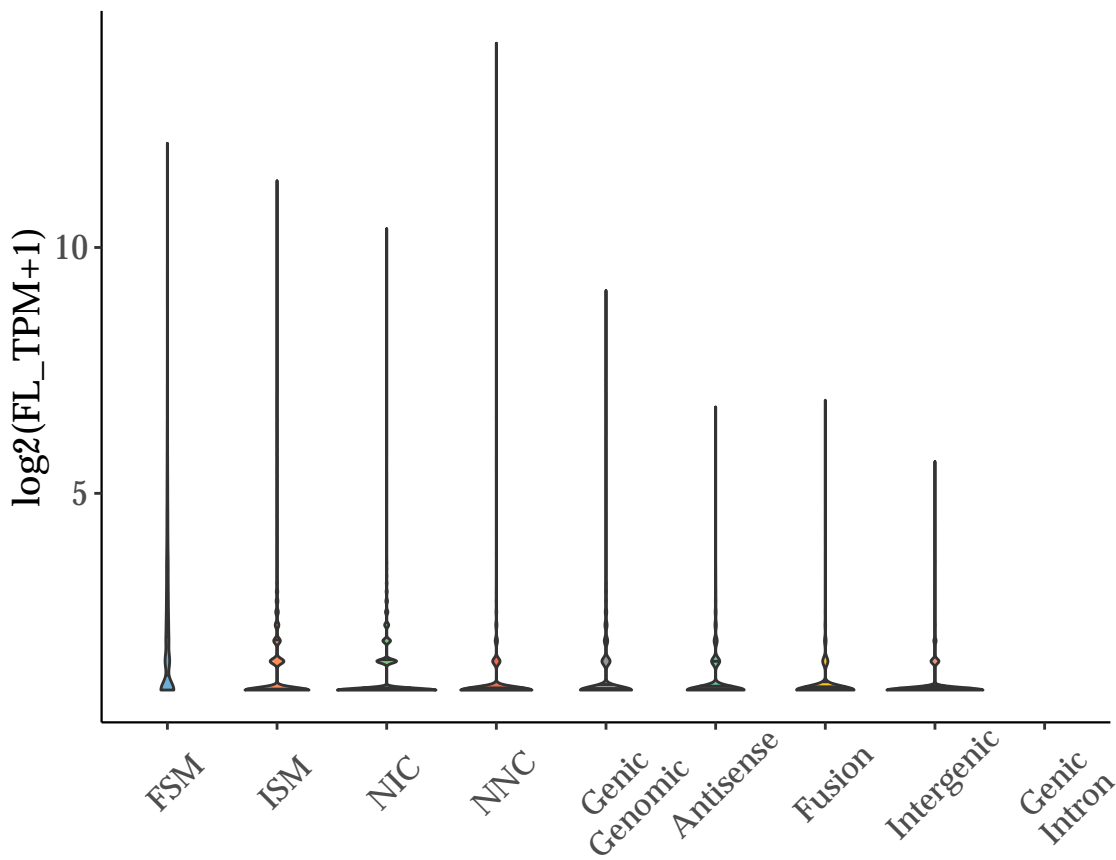
Transcript Lengths by Structural Classification



Exon Counts by Structural Classification



FL Count (normalized) by Structural Category

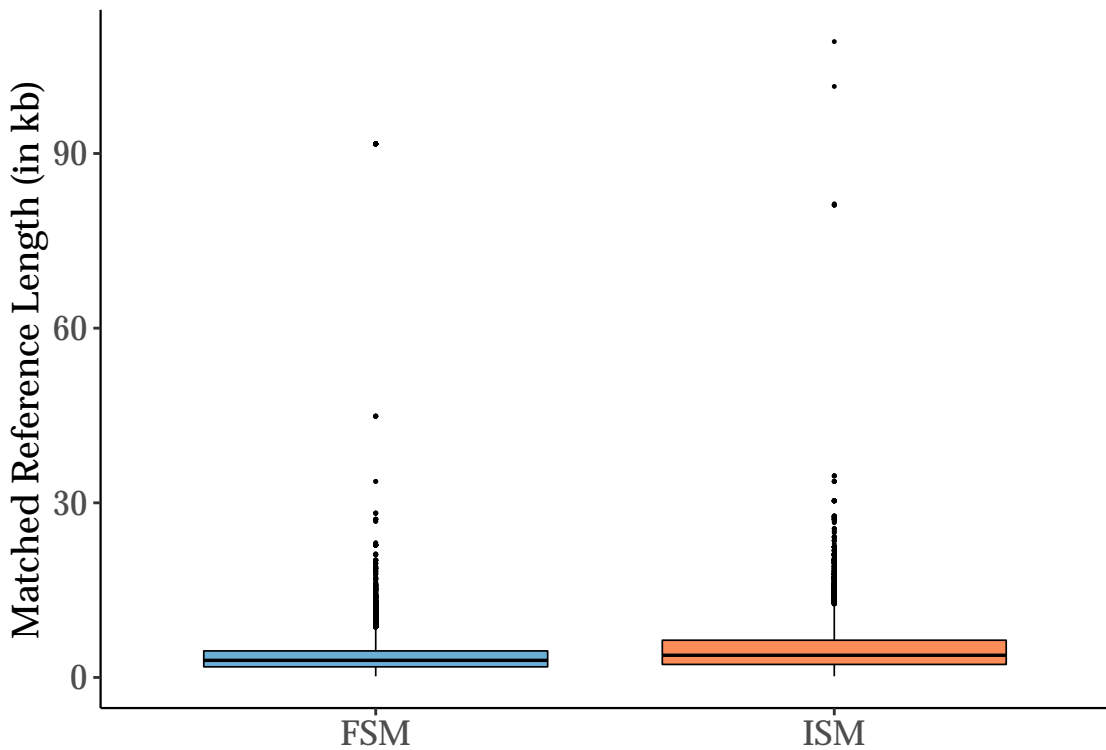


FL counts by category

category	isoforms	FL
FSM	36495	1025294
ISM	23178	155878
NIC	121124	554723
NNC	47109	226996
Genic Genomic	4835	33776
Antisense	11807	40897
Fusion	2931	9084
Intergenic	14945	40787

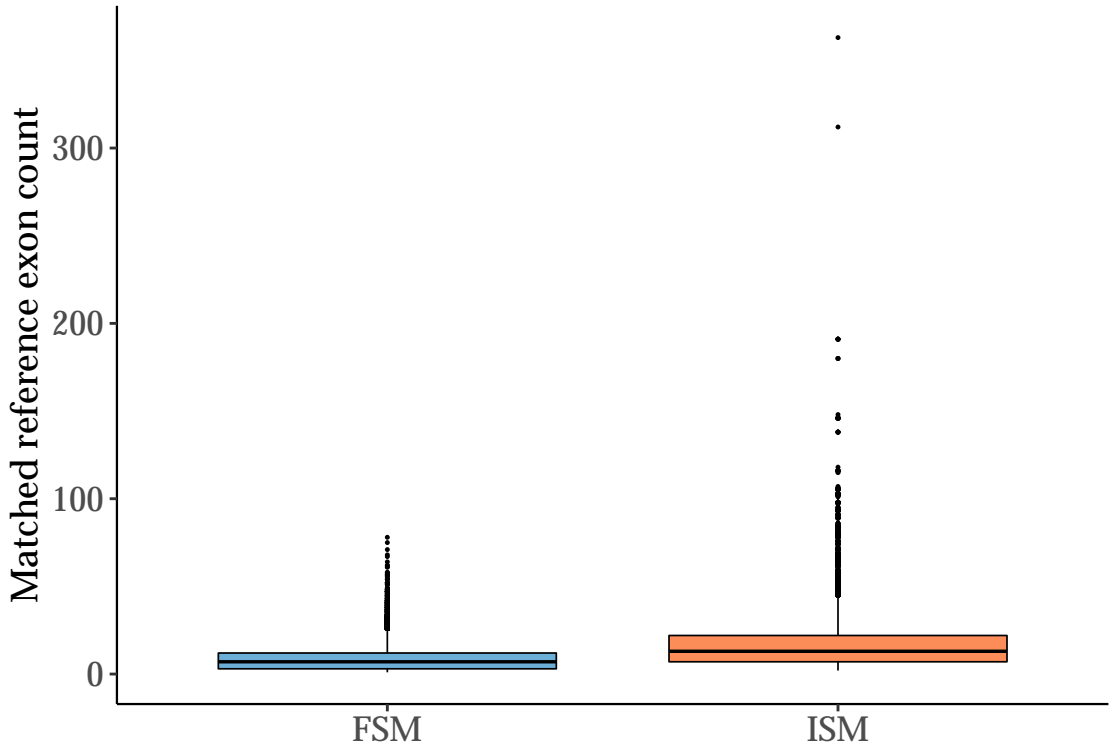
Length Distribution of Matched Reference Transcripts

Applicable only to FSM and ISM categories



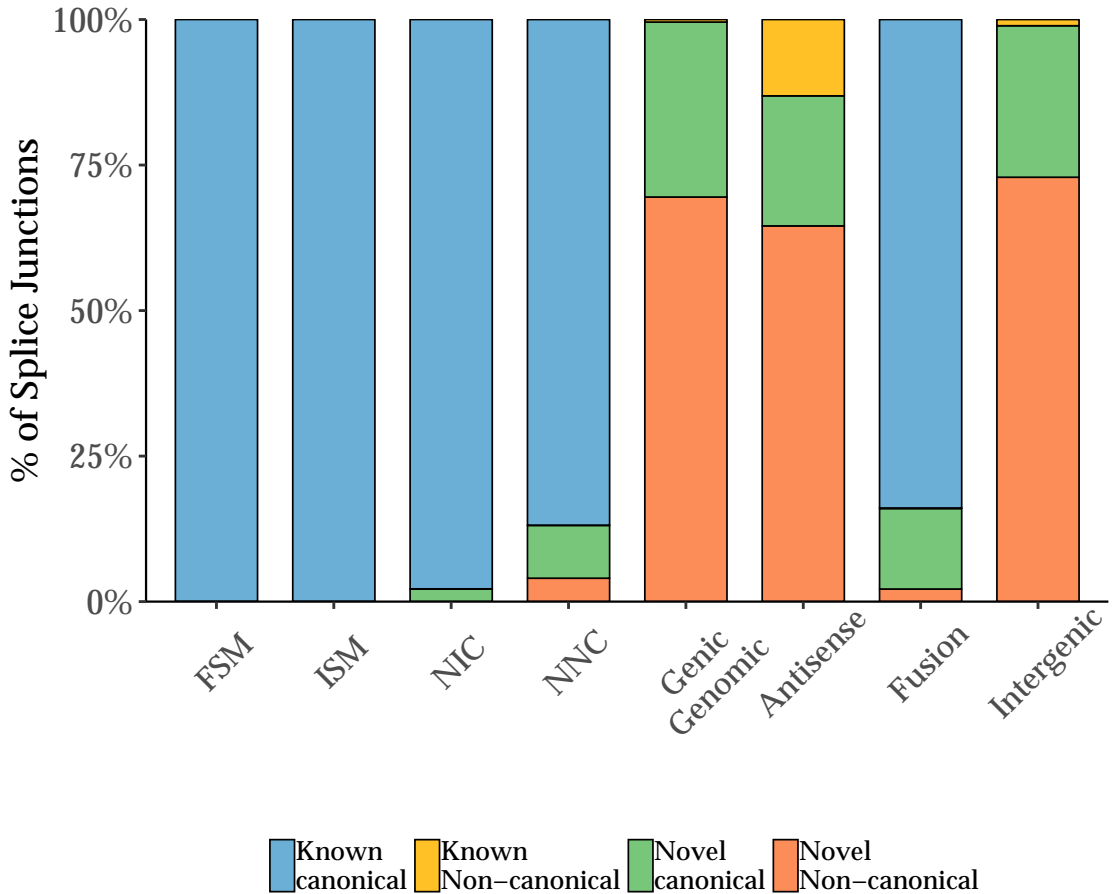
Exon Count Distribution of Matched Reference Transcripts

Applicable only to FSM and ISM categories

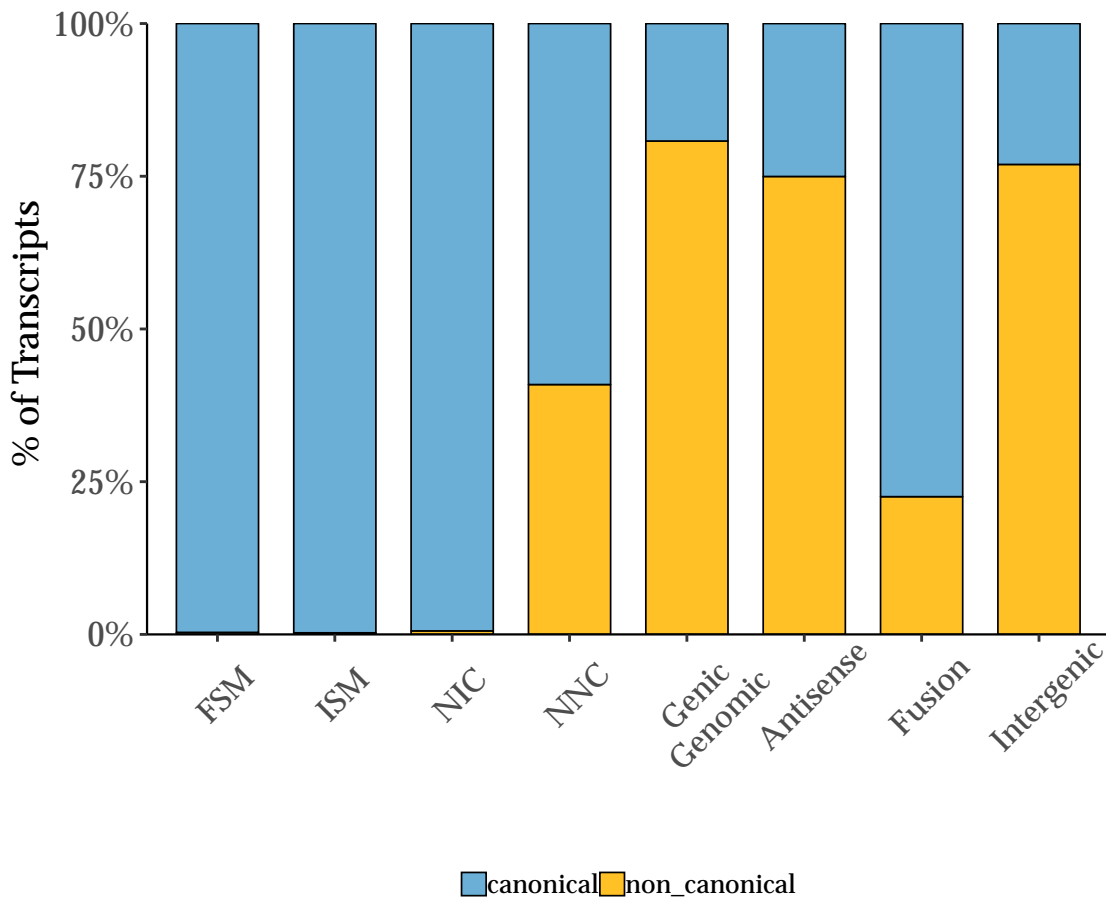


Splice Junction Characterization

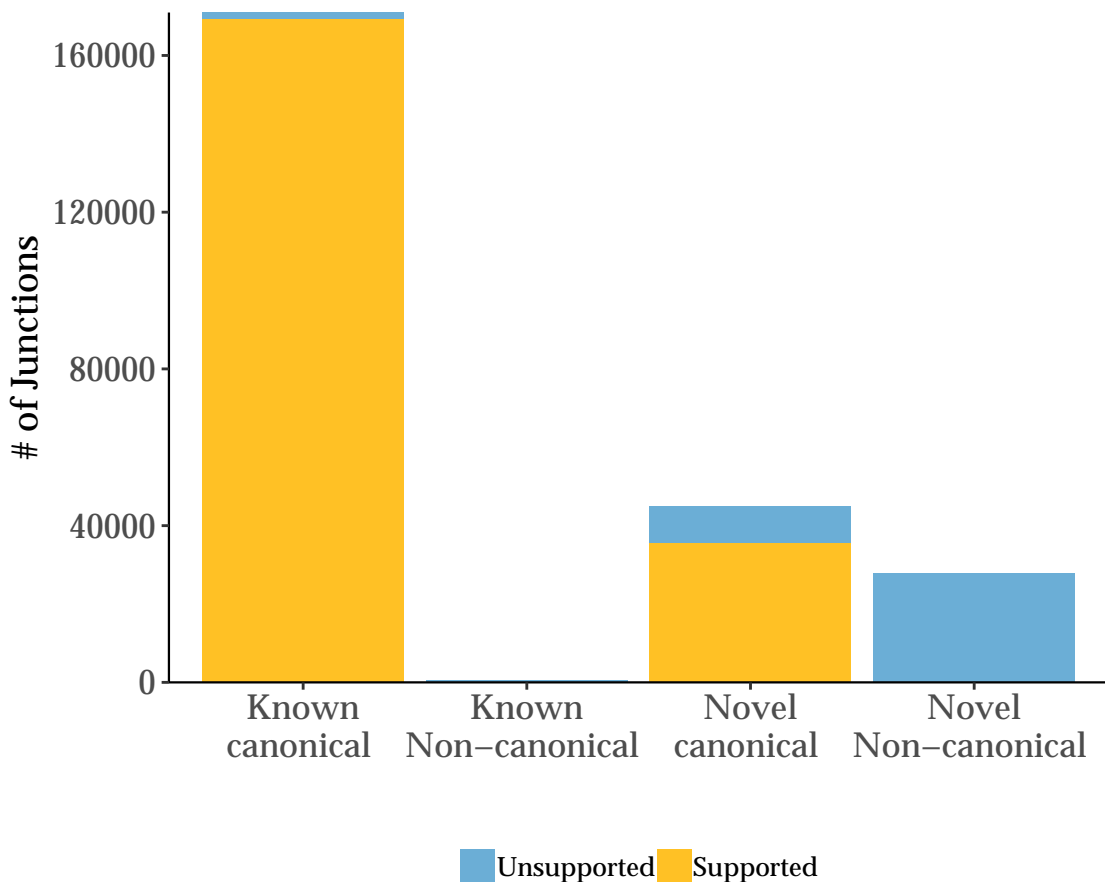
Distribution of Splice Junctions by Structural Classification



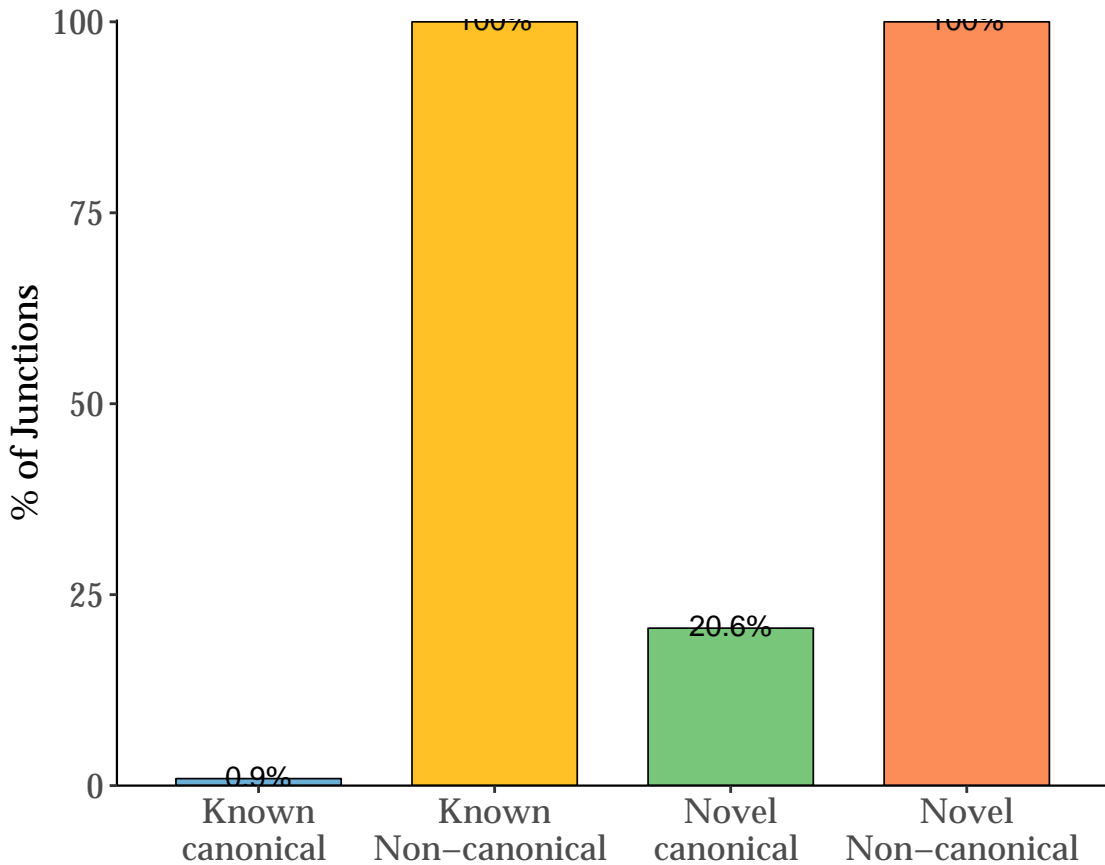
Distribution of Transcripts by Splice Junctions



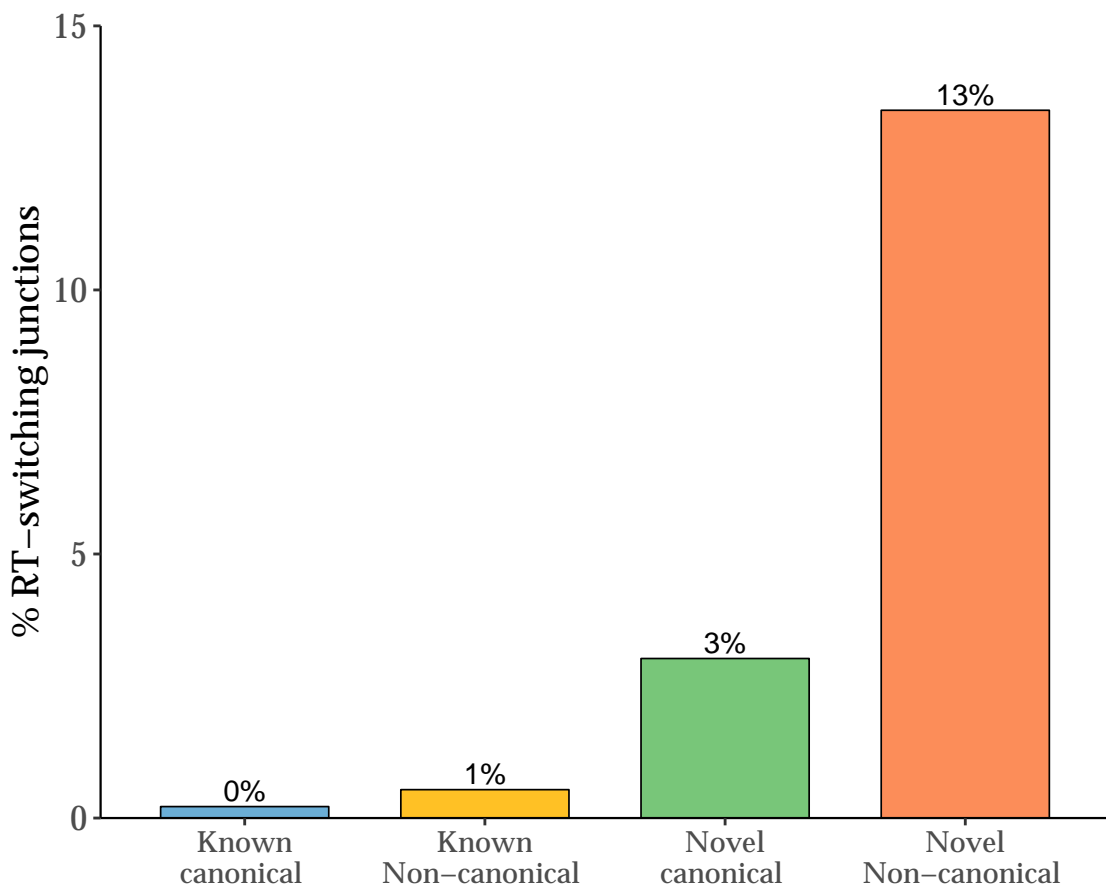
Unique junctions w/ or w/out short read coverage



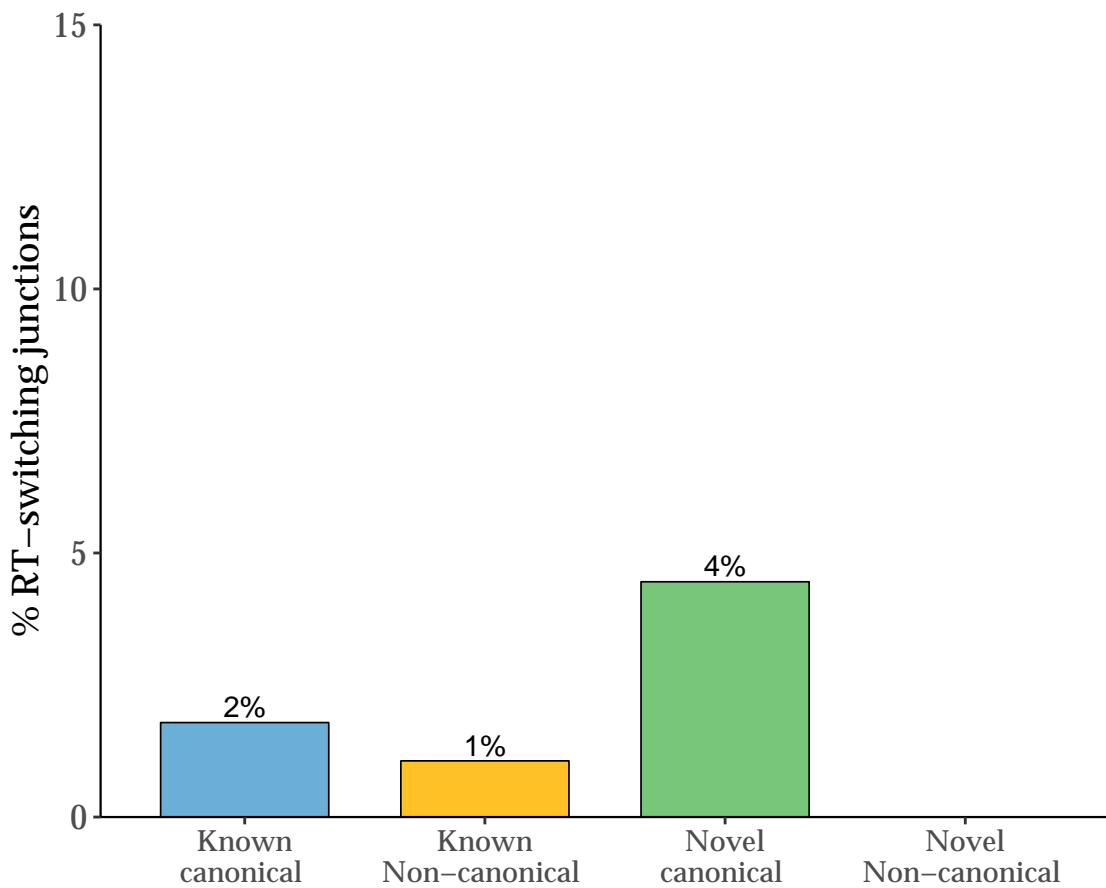
Unique junctions w/out short read coverage (percentage)



RT-switching, all junctions



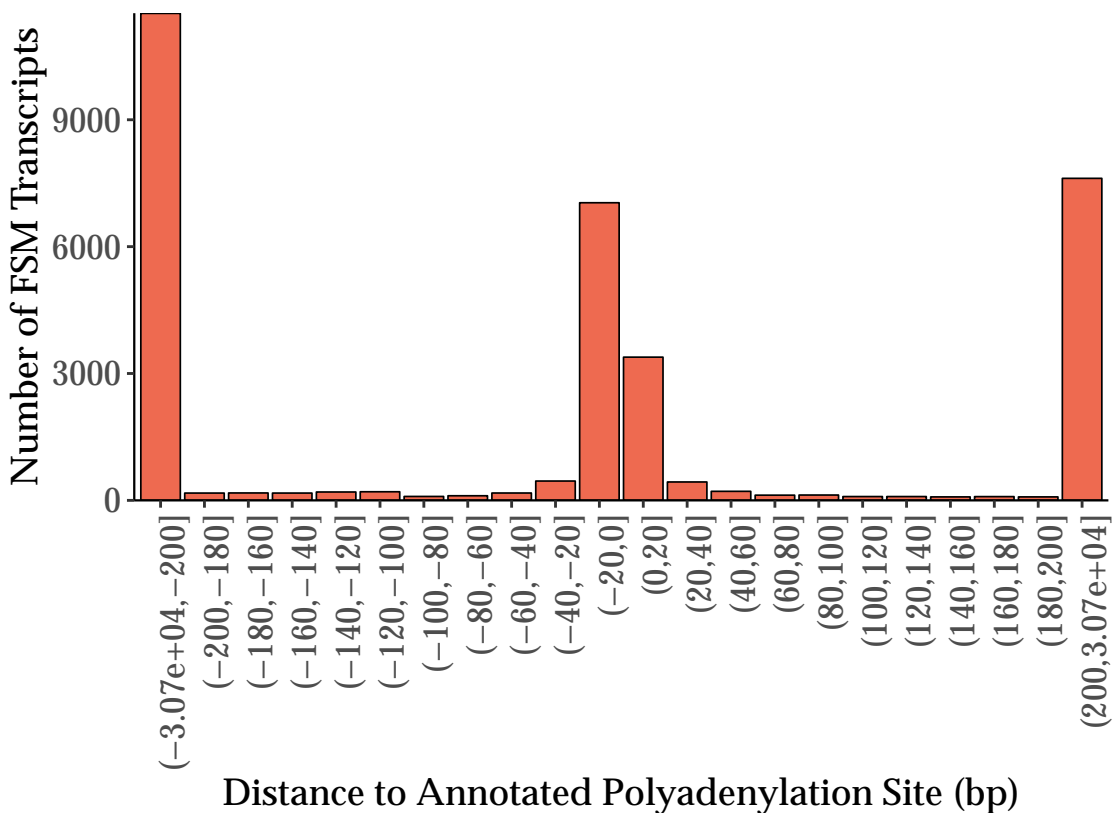
RT-switching, unique junctions



Comparison with Annotated TSS and PolyA Sites

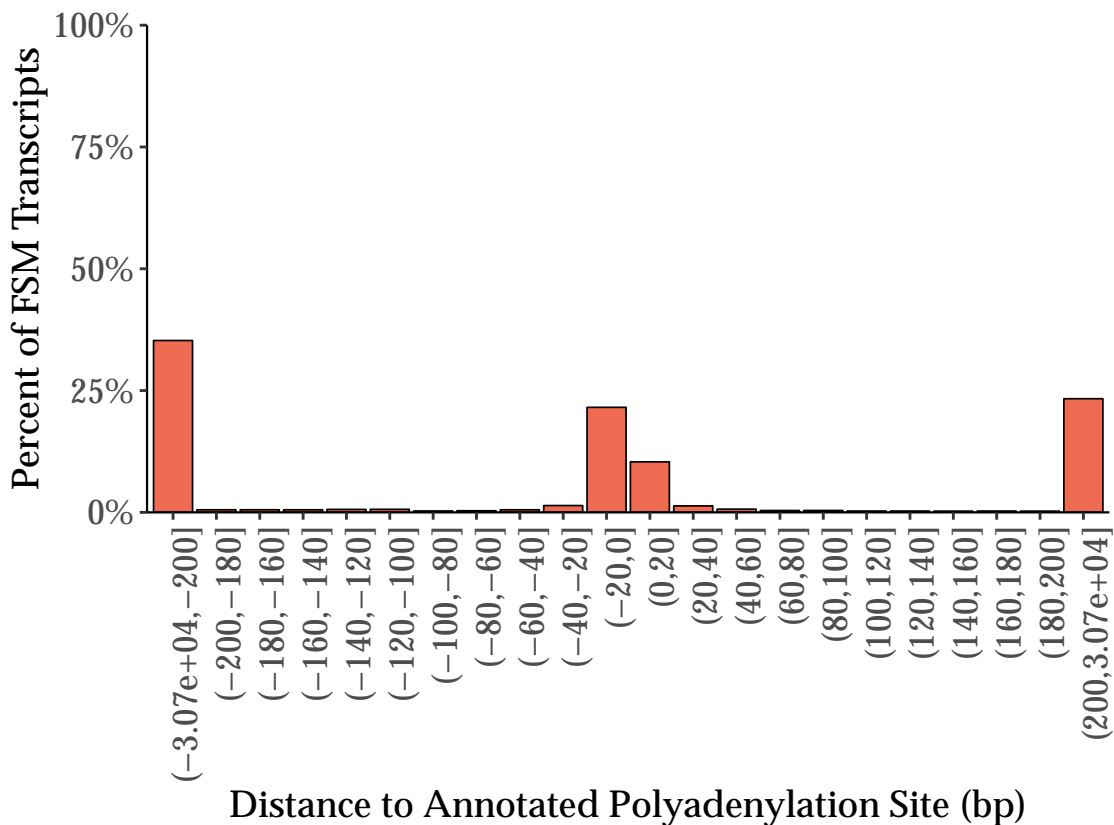
Distance to Annotated Polyadenylation Site, FSM only

Negative values indicate upstream of annotated polyA site



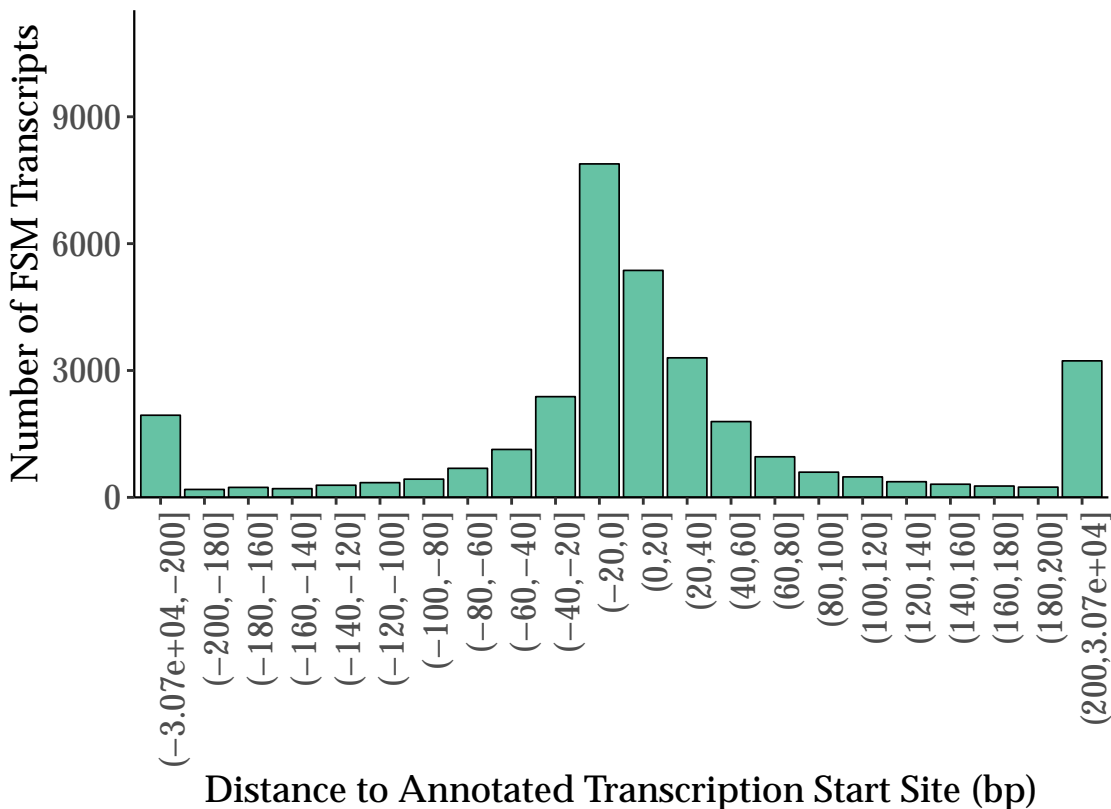
Distance to Annotated Polyadenylation Site, FSM only

Negative values indicate upstream of annotated polyA site



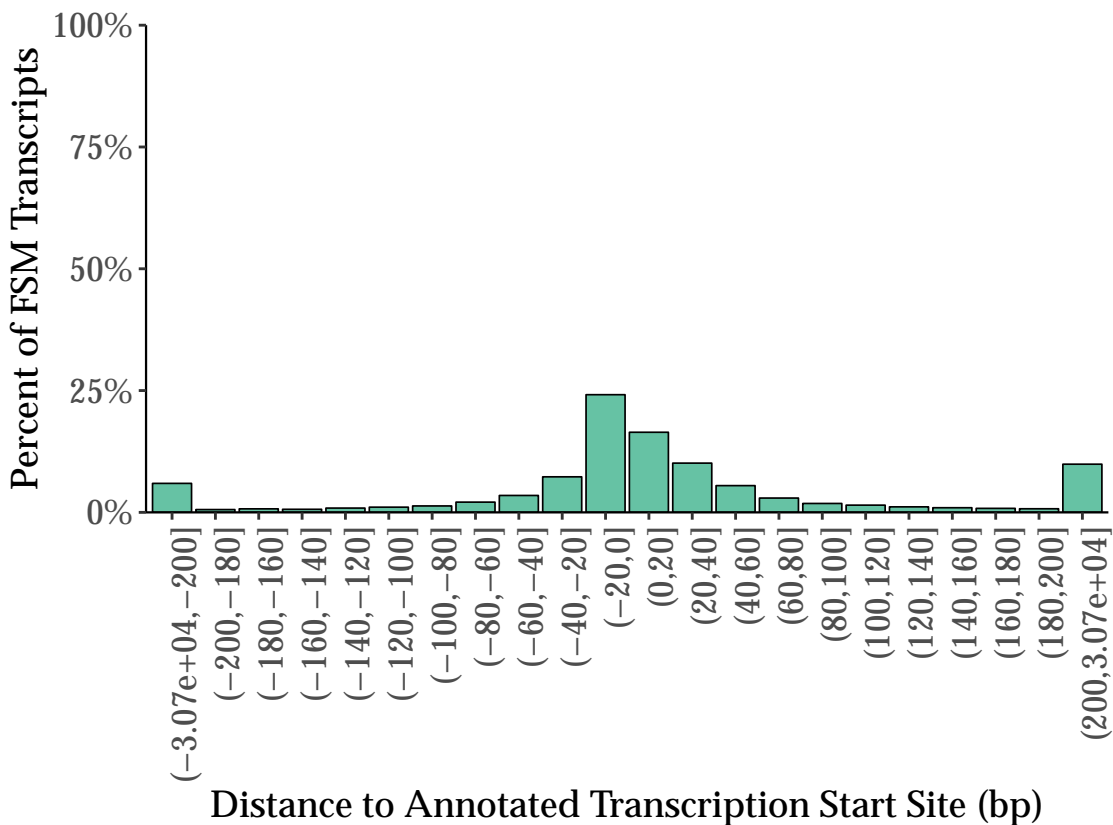
Distance to Annotated Transcription Start Site, FSM only

Negative values indicate downstream of annotated TSS



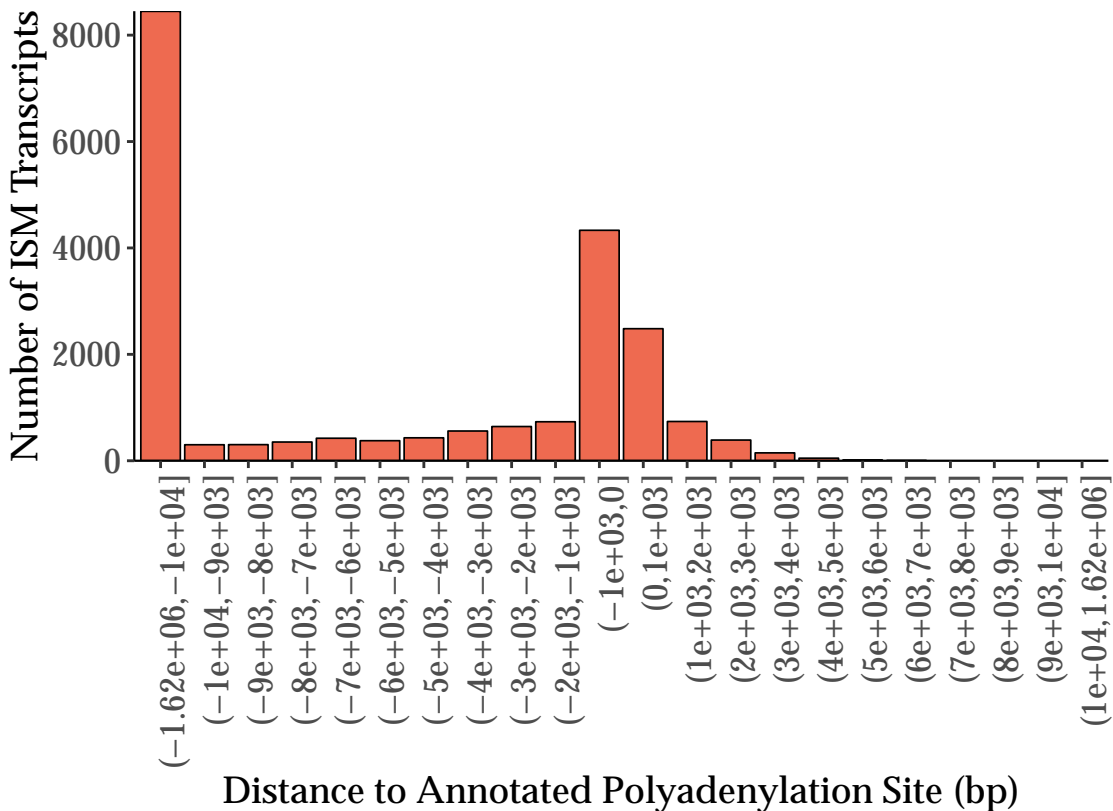
Distance to Annotated Transcription Start Site, FSM only

Negative values indicate downstream of annotated TSS



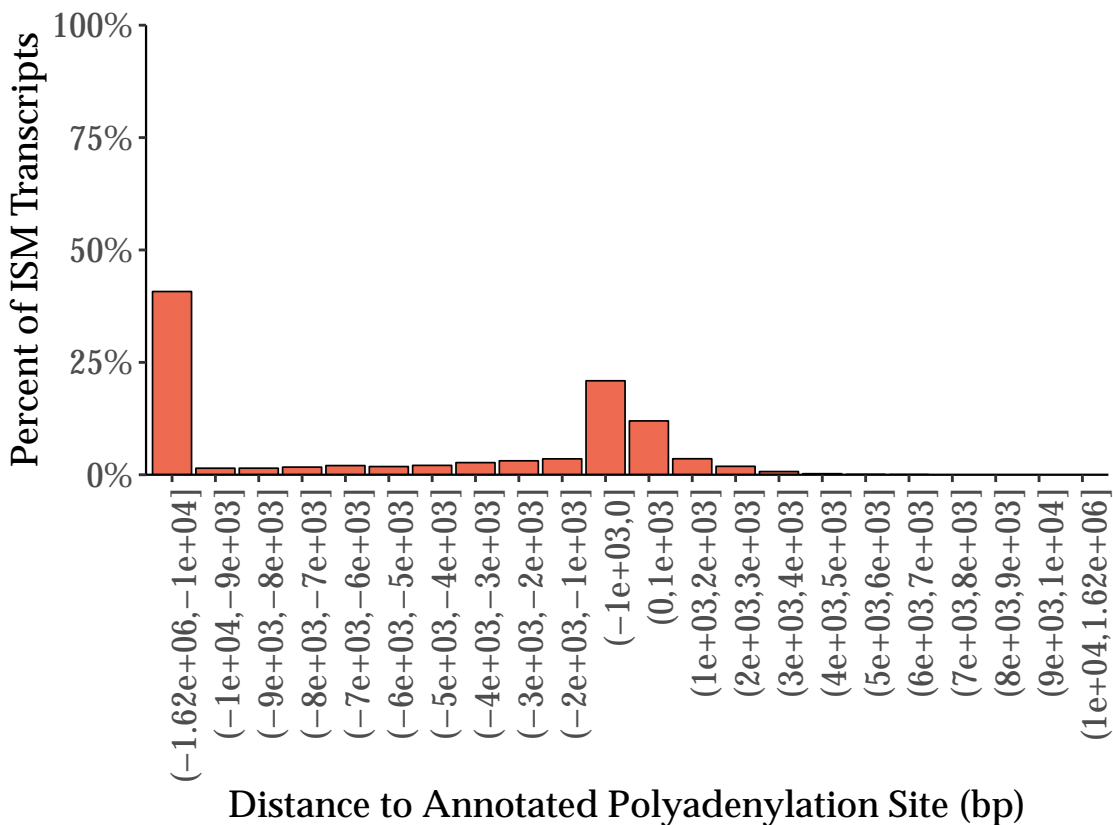
Distance to Annotated Polyadenylation Site, ISM only

Negative values indicate upstream of annotated polyA site



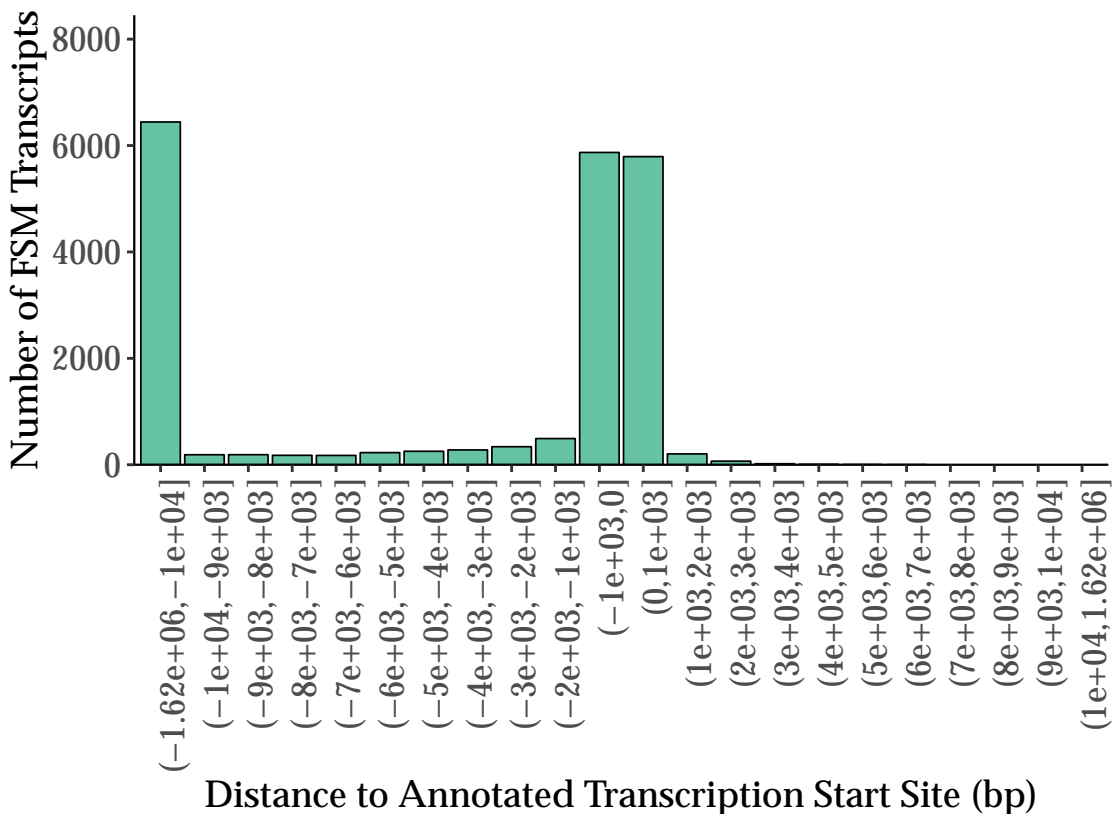
Distance to Annotated Polyadenylation Site, ISM only

Negative values indicate upstream of annotated polyA site



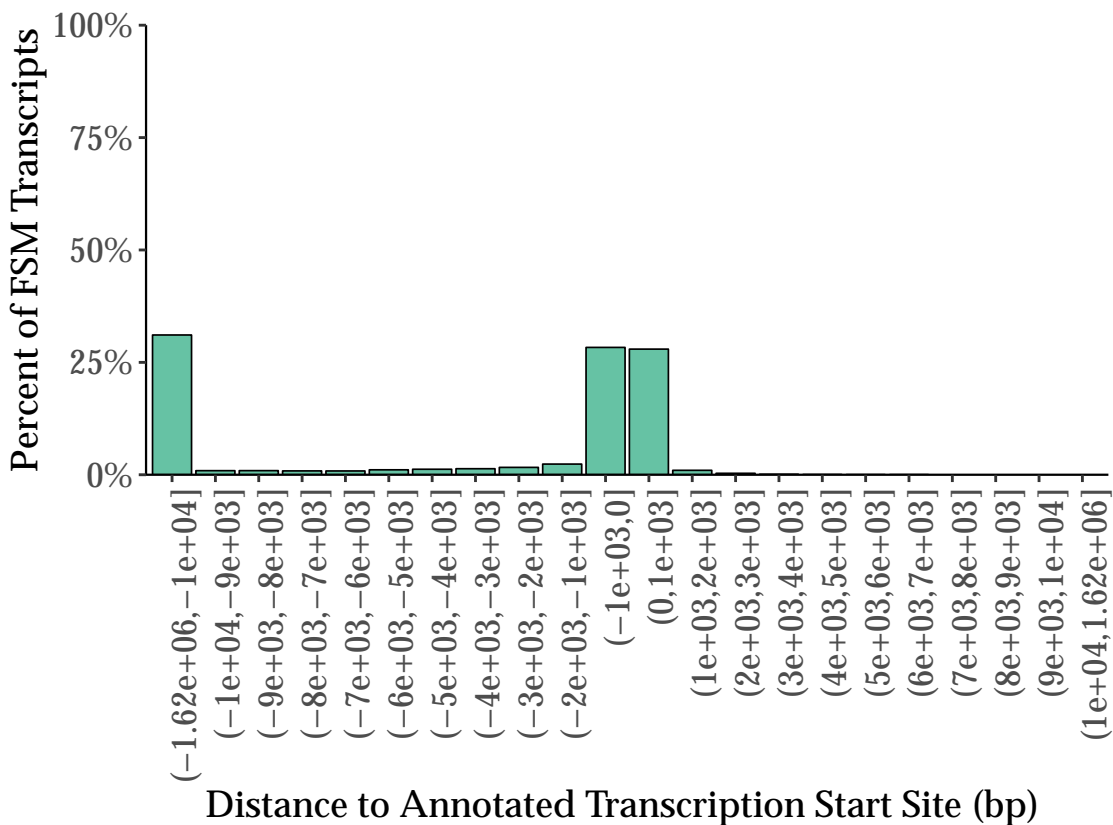
Distance to Annotated Transcription Start Site, ISM only

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	68914	57.6
ATTAAA	17064	14.3
AGTAAA	4057	3.4
TATAAA	4001	3.3
TTTAAA	3694	3.1
AAGAAA	3242	2.7
AAAAAG	2770	2.3
AATATA	2739	2.3
AATACA	2323	1.9
CATAAA	2180	1.8
GATAAA	2150	1.8
AATGAA	1859	1.6
AAAACA	1672	1.4
AATAGA	1438	1.2
ACTAAA	1264	1.1
GGGGCT	297	0.2

Number of polyA Motifs Detected

Category	Count	polyA Detected	%
FSM	36495	24296	67
ISM	23178	11597	50
NIC	121124	47145	39
NNC	47109	29299	62
Genic Genomic	4835	1648	34
Antisense	11807	1965	17
Fusion	2931	1534	52
Intergenic	14945	2180	15

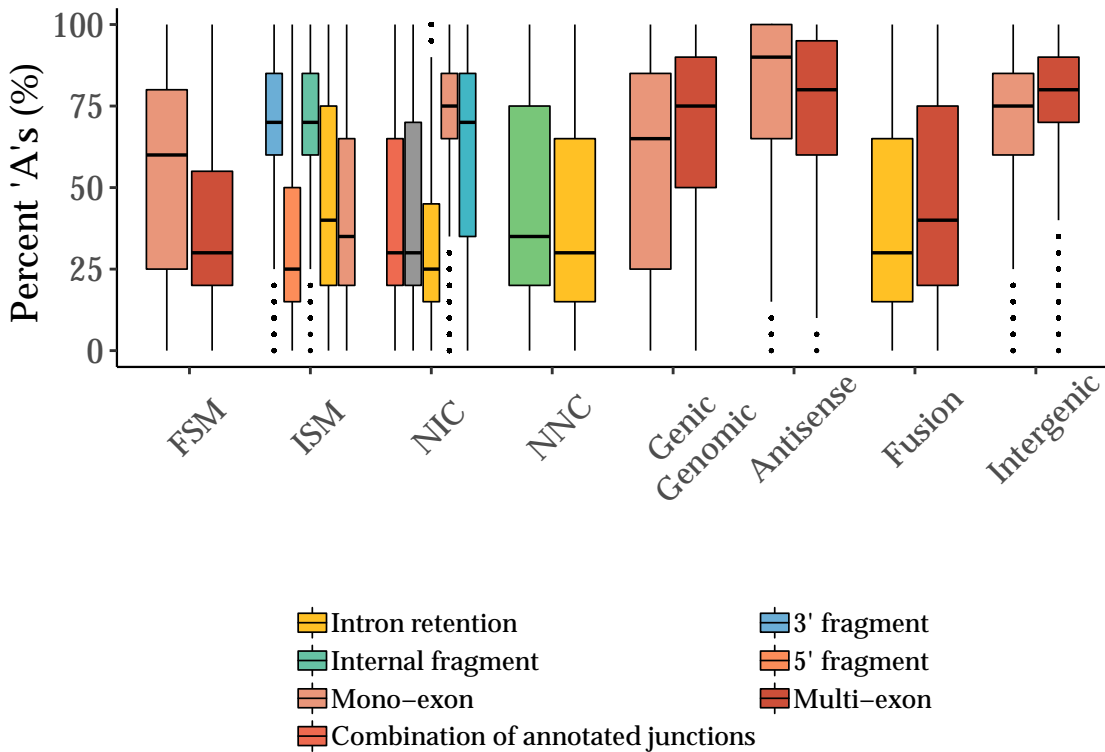
Number of close by CAGE Peaks Detected

Category	Count	Has CAGE peak within 50bp	%
FSM	36495	26056	71
ISM	23178	9942	43
NIC	121124	34318	28
NNC	47109	28252	60
Genic Genomic	4835	1344	28
Antisense	11807	190	2
Fusion	2931	1708	58
Intergenic	14945	202	1

Intra-Priming Quality Check

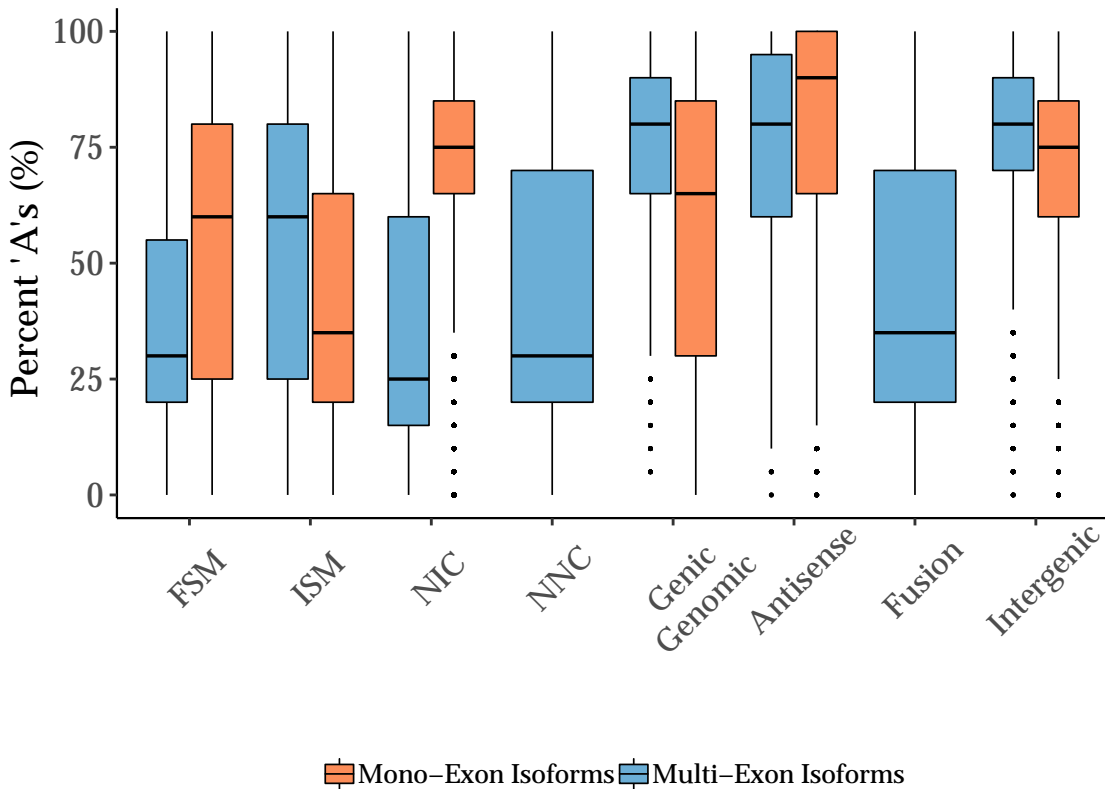
Possible Intra-Priming by Structural Category

Percent of genomic 'A's in downstream 20 bp



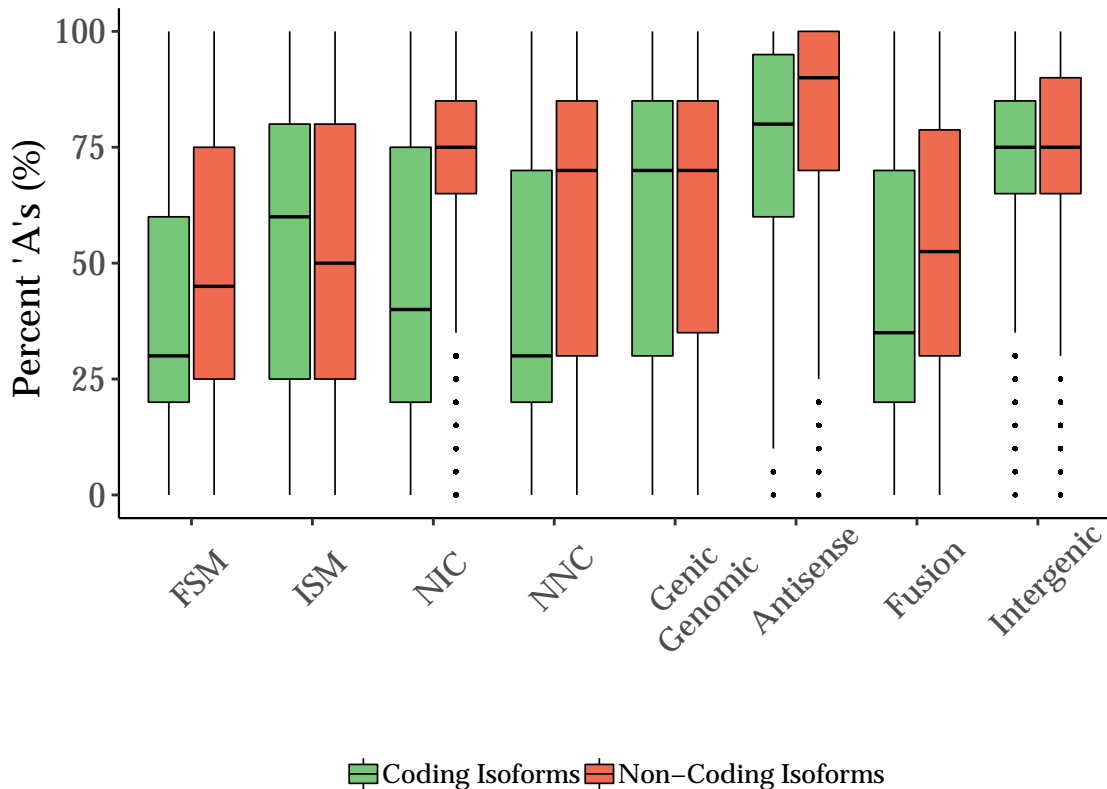
Possible Intra-Priming, Mono- vs Multi-Exon

Percent of genomic 'A's in downstream 20 bp



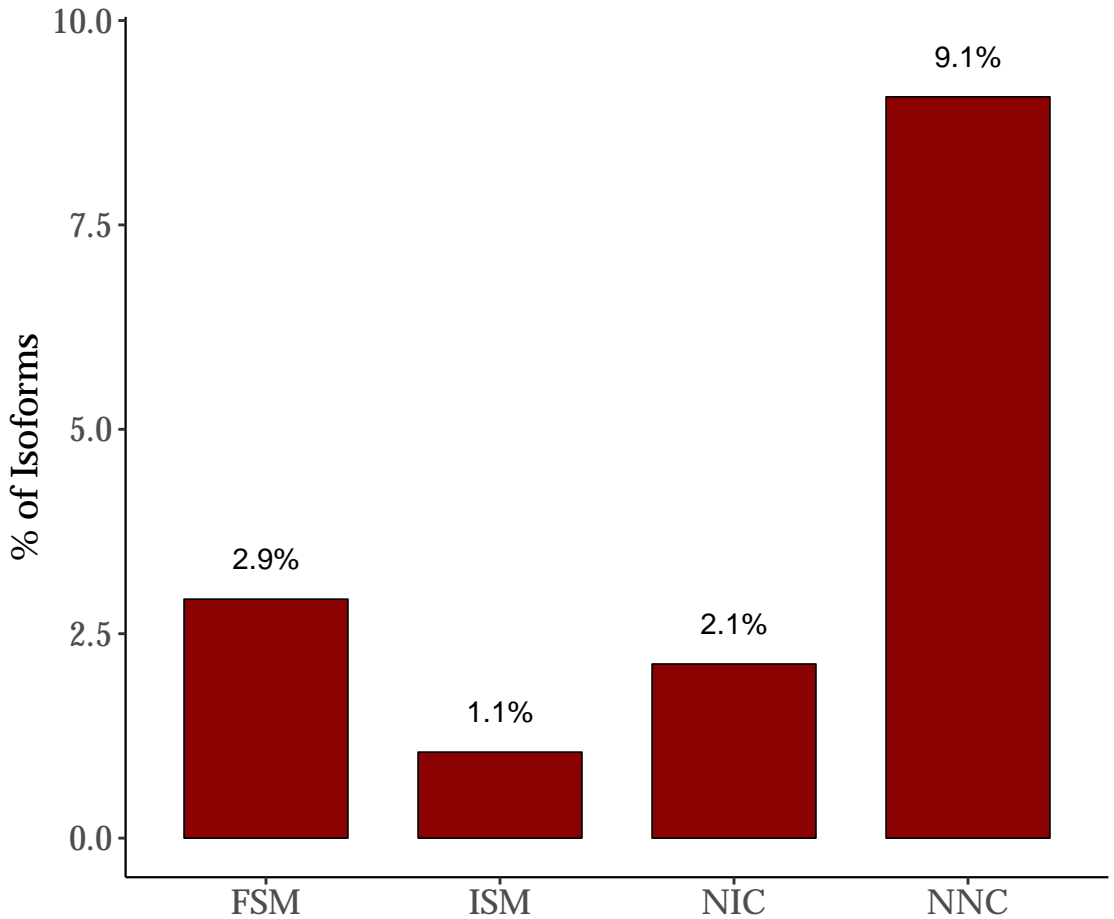
Possible Intra-Priming, Coding vs Non-Coding

Percent of genomic 'A's in downstream 20 bp



Quality Controls

Incidence of RT-switching



Incidence of Non-Canonical Junctions

