SQANTI3 report

Unique Genes: 23271

Unique Isoforms: 359188

Gene classification

Category	# Genes
Annotated Genes	21220
Novel Genes	2051

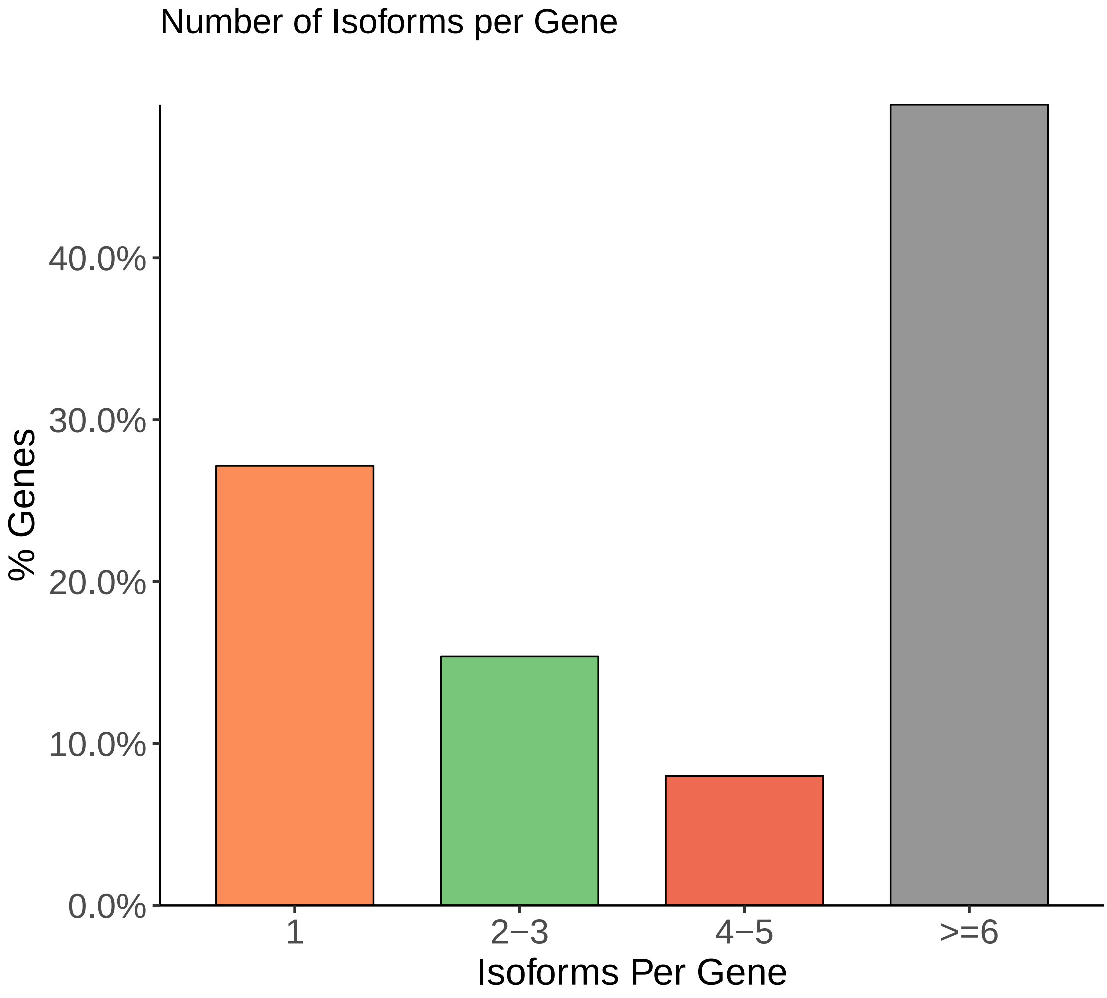
Characterization of transcripts based on splice junctions

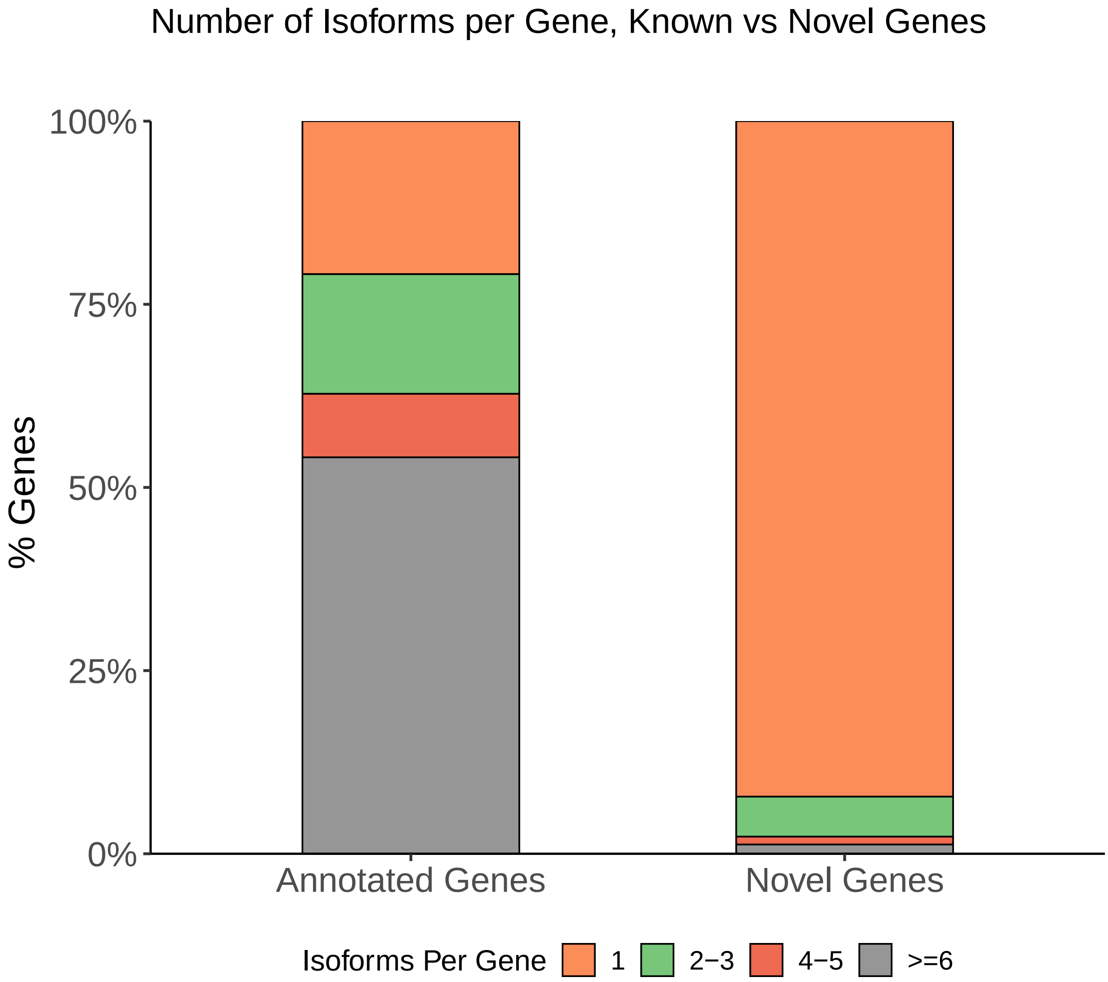
Category	# Isoforms
FSM	72322
ISM	32762
NIC	141721
NNC	103728
Genic Genomic	679
Antisense	1075
Fusion	5527
Intergenic	1374
Genic Intron	0

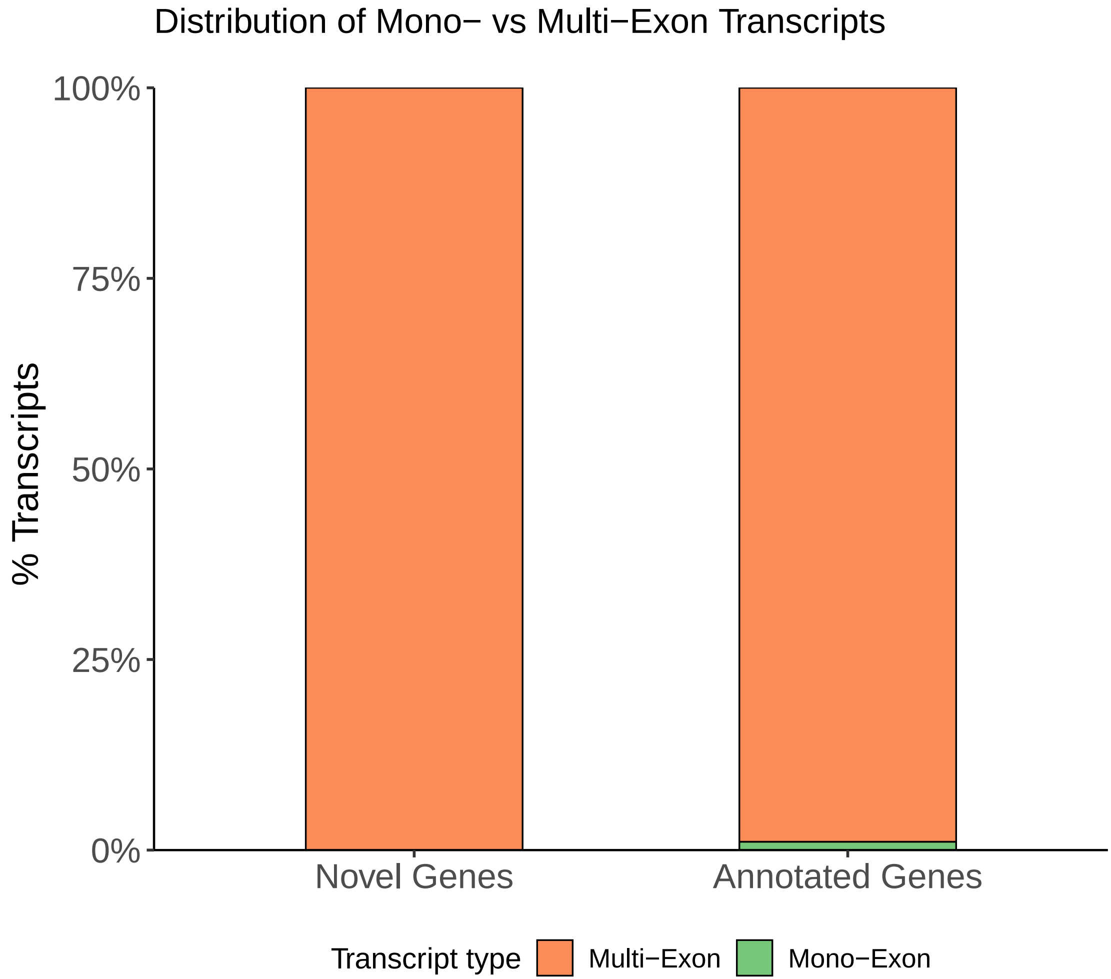
Splice Junction Classification

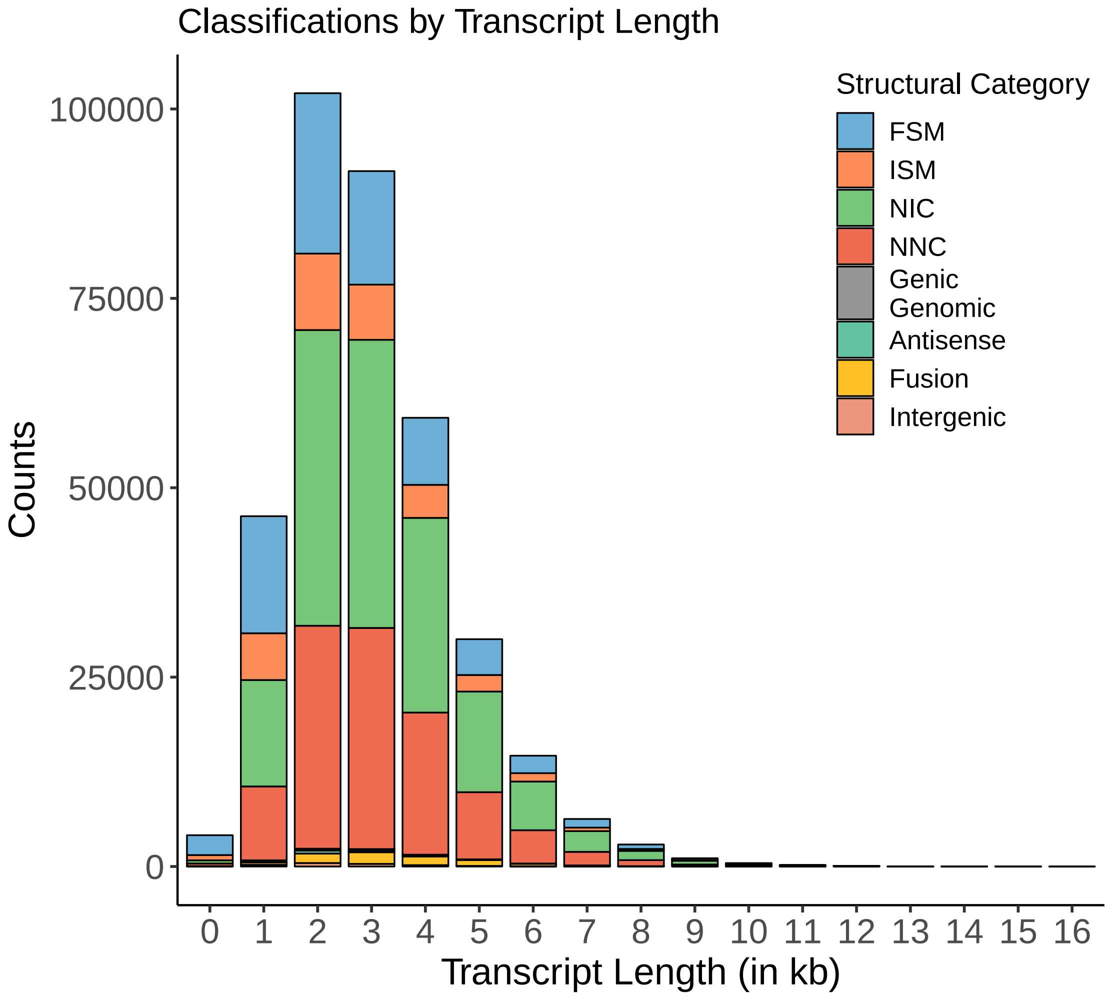
Category	# SJs	Percent
Known canonical	205698	63.87
Known Non−canonical	84	0.03
Novel canonical	116271	36.10
Novel Non−canonical	0	0.00

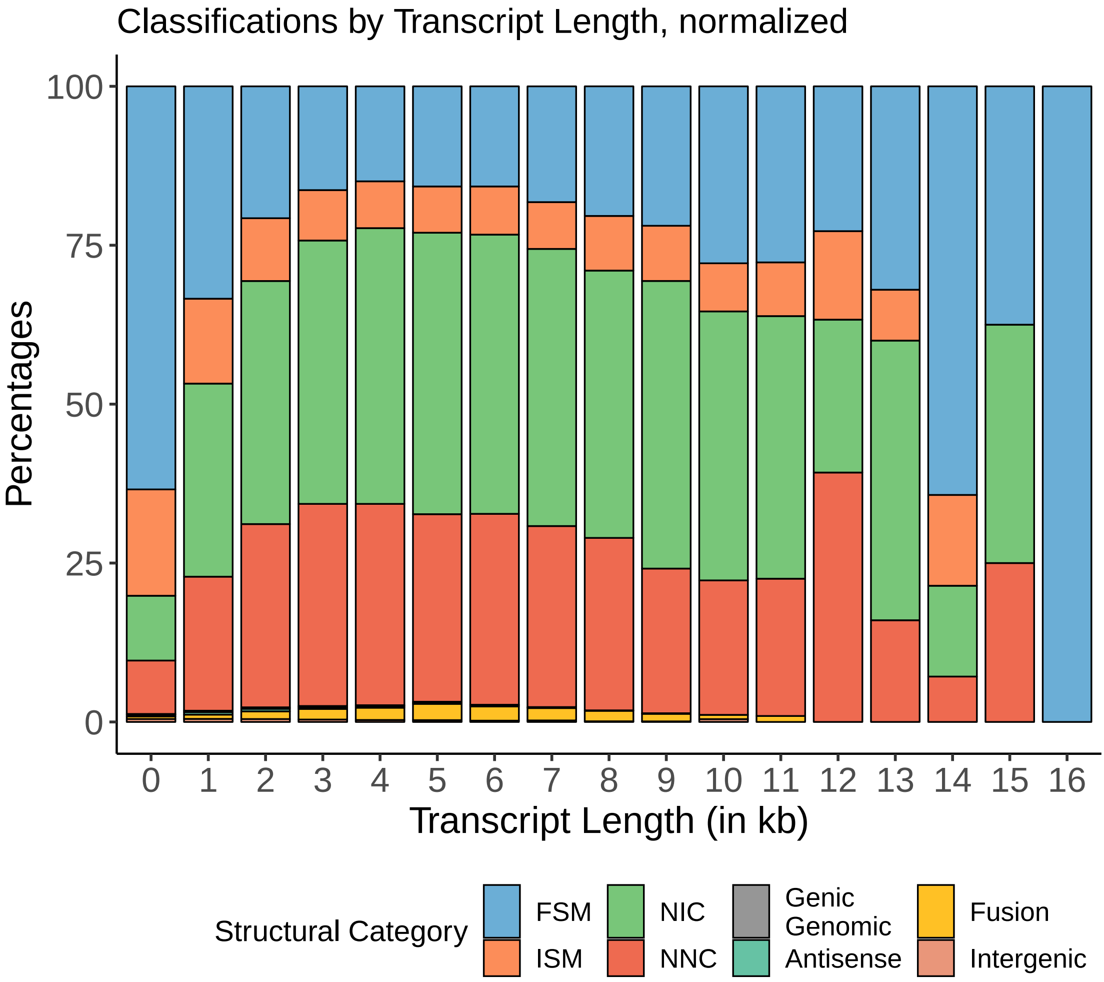
Gene Characterization

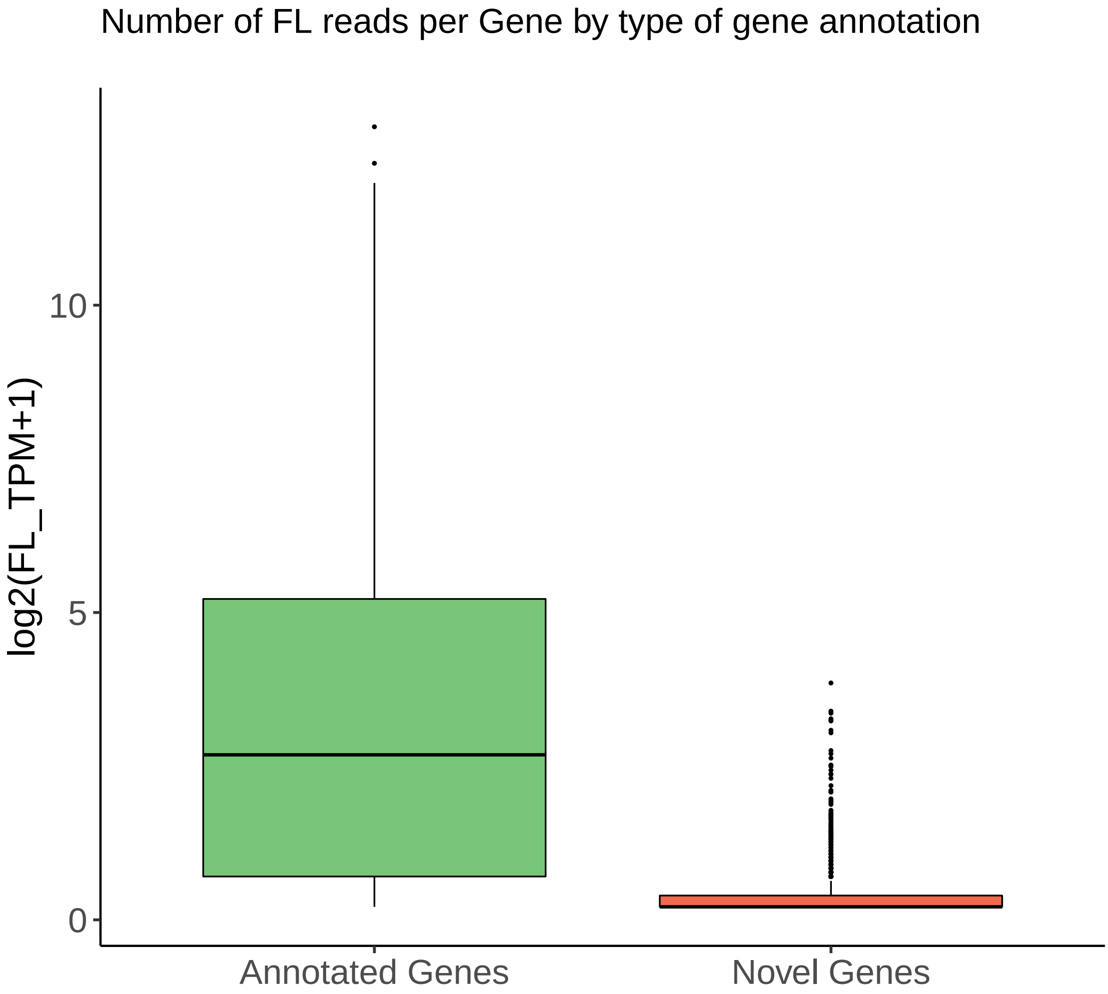

Number of Isoforms per Gene
% Genes
40.0%
30.0%
20.0%
10.0%
0.0%
1
2−3
4−5
>=6
Isoforms Per Gene


Number of Isoforms per Gene, Known vs Novel Genes
100%
75%
50%
25%
0%
% Genes
Annotated Genes
Novel Genes
Isoforms Per Gene
1
2−3
4−5
>=6

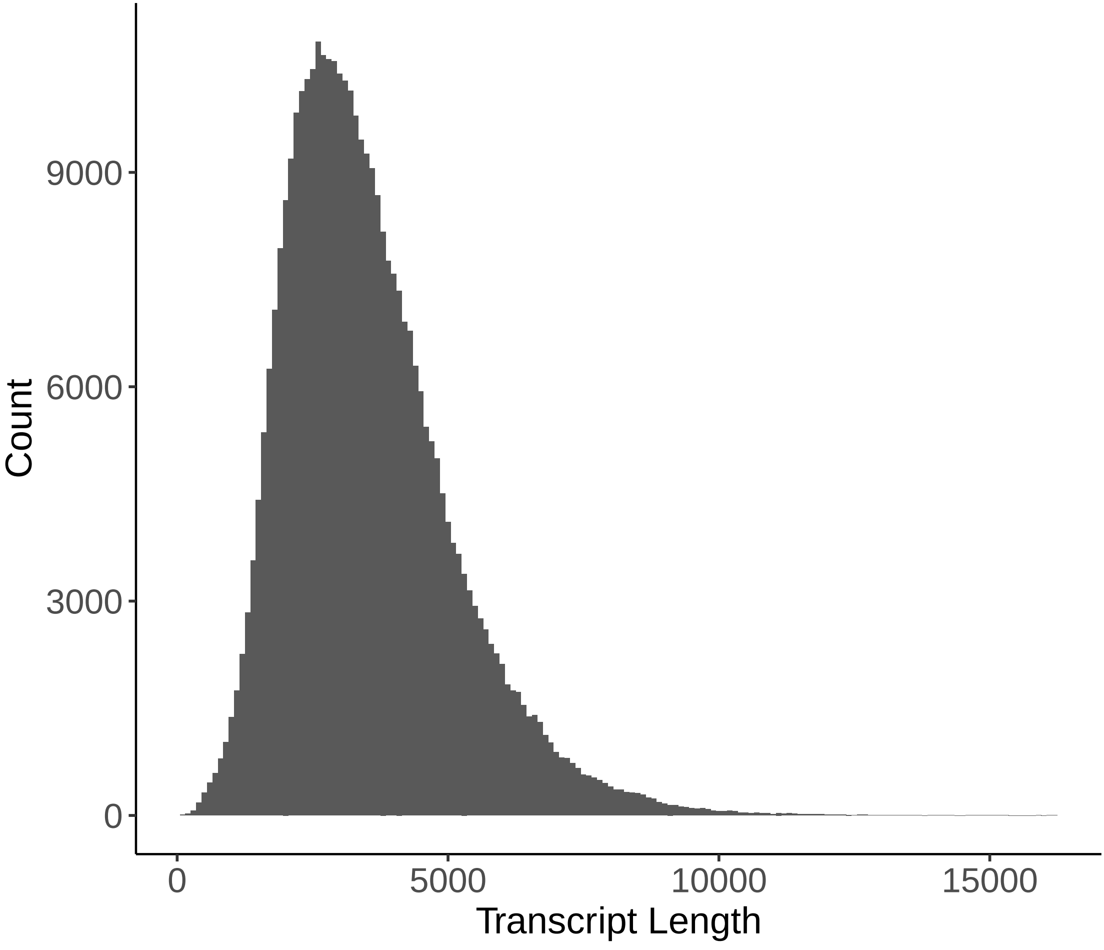

Distribution of Mono− vs Multi−Exon Transcripts
% Transcripts
100%
75%
50%
25%
0%
Novel Genes
Annotated Genes
Transcript type
Multi−Exon
Mono−Exon

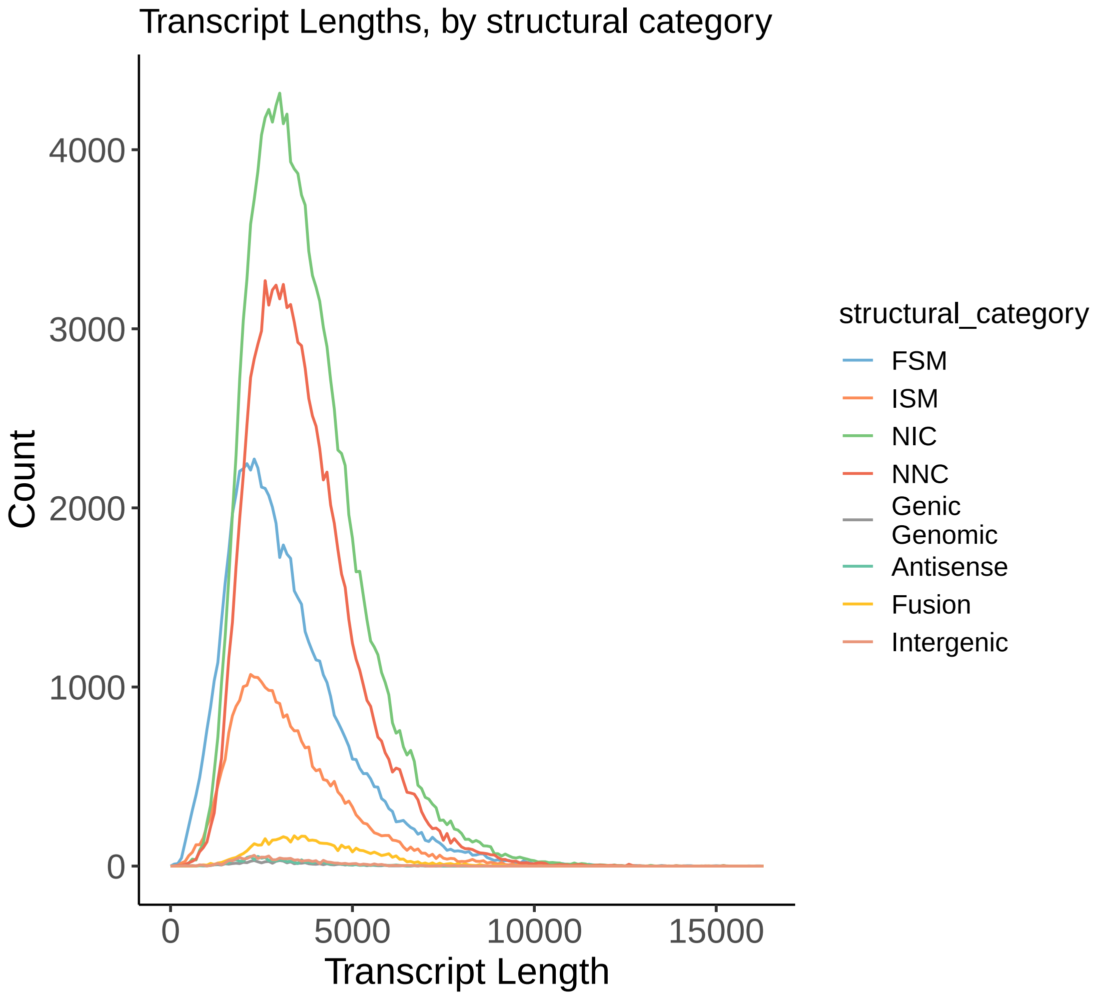

Classifications by Transcript Length
Structural Category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Counts
100000
75000
50000
25000
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Transcript Length (in kb)

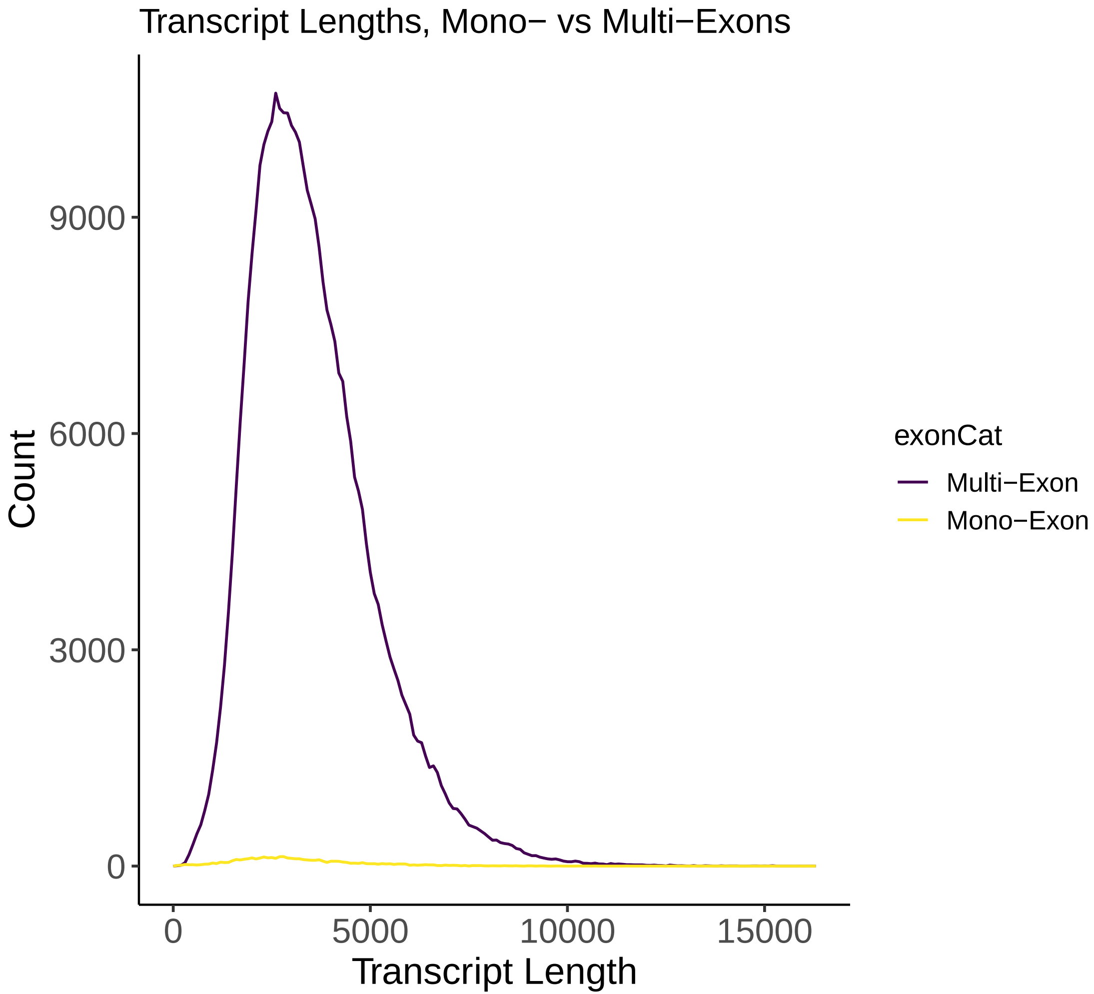

Classifications by Transcript Length, normalized
Percentages
100
75
50
25
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Transcript Length (in kb)
Structural Category
FSM
ISM
NIC
NNC
Genic
Genomic
Antisense
Fusion
Intergenic


Number of FL reads per Gene by type of gene annotation
log2(FL_TPM+1)
10
5
0
Annotated Genes
Novel Genes

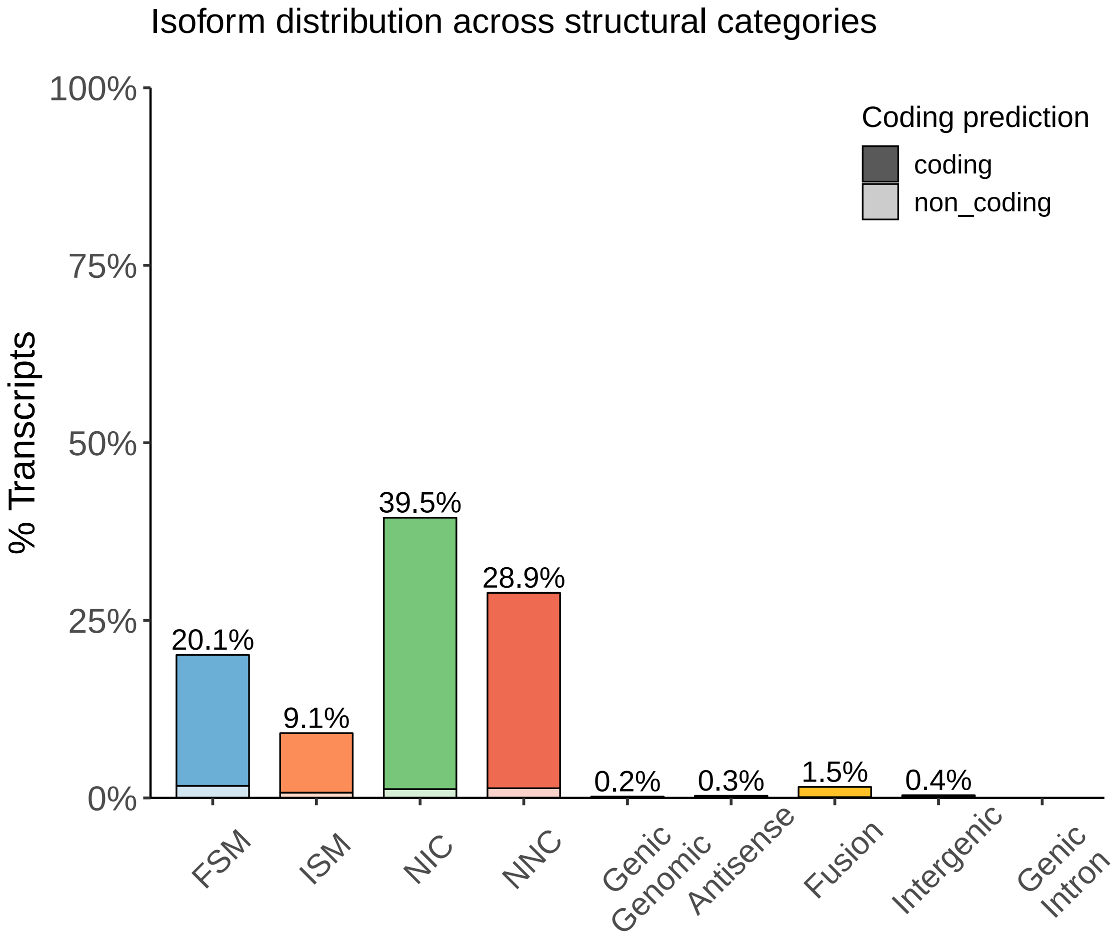
Transcript Lengths, all transcripts

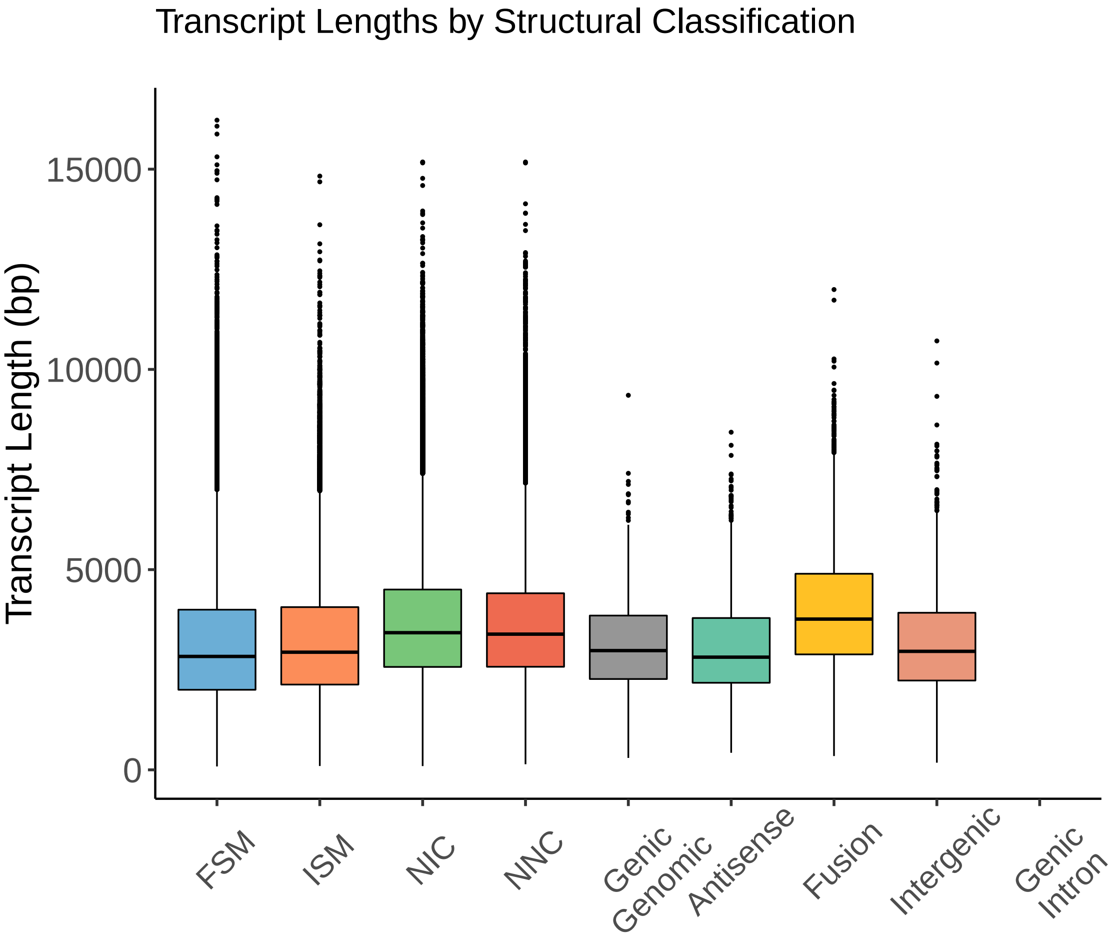


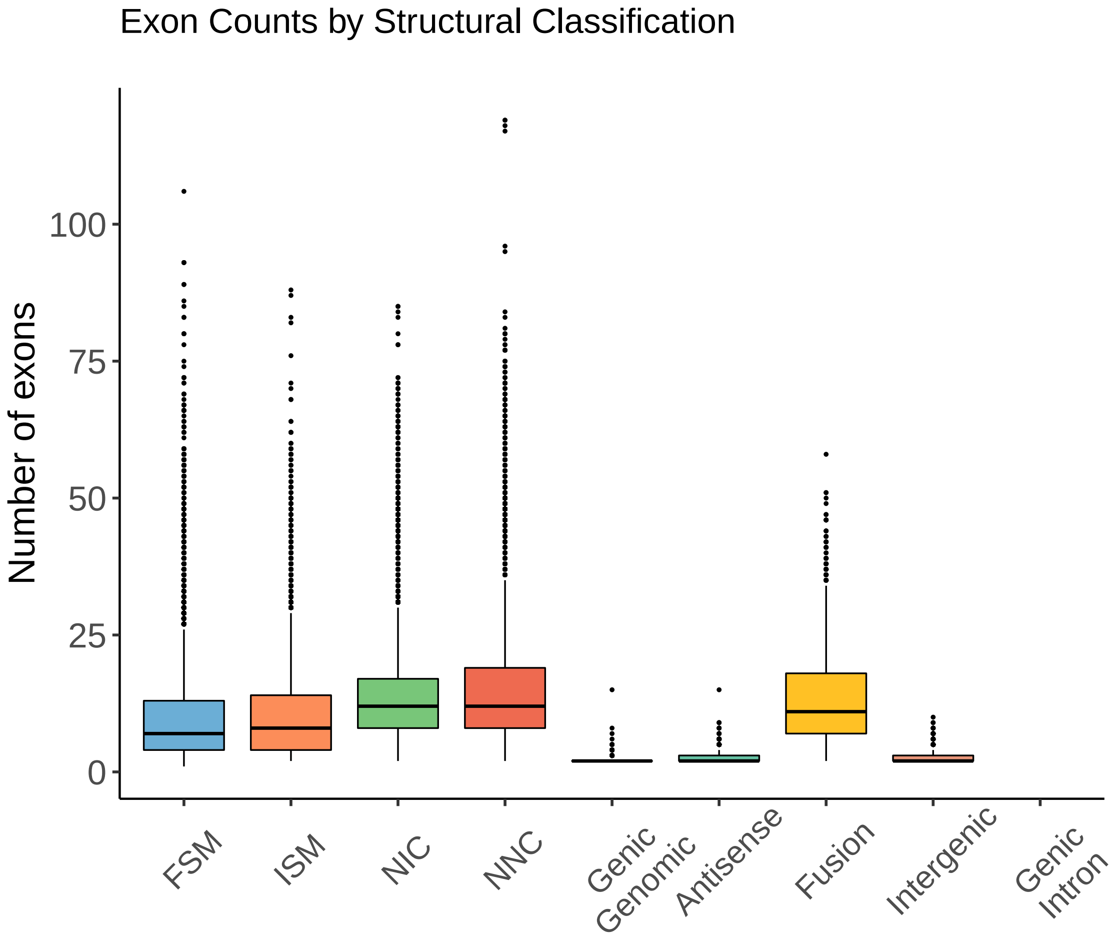

Transcript Lengths, by structural category
Count
4000
3000
2000
1000
0
0
5000
10000
15000
Transcript Length
structural_category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic


Transcript Lengths, Mono− vs Multi−Exons
Count
9000
6000
3000
0
0
5000
10000
15000
Transcript Length
exonCat
Multi−Exon
Mono−Exon

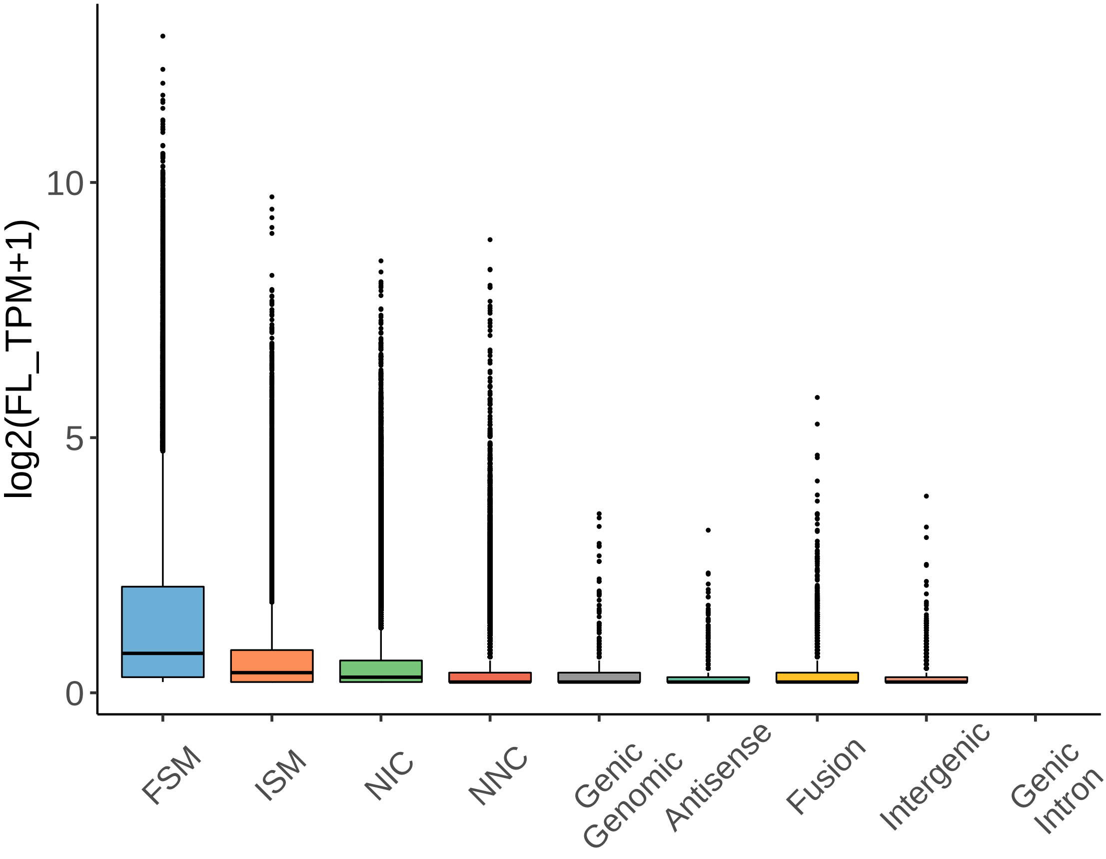
Structural Isoform Characterization by Splice Junctions


Isoform distribution across structural categories
% Transcripts
100%
75%
50%
25%
0%
Coding prediction
coding
non_coding
20.1%
9.1%
39.5%
28.9%
0.2%
0.3%
1.5%
0.4%
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron


Transcript Lengths by Structural Classification
Transcript Length (bp)
15000
10000
5000
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

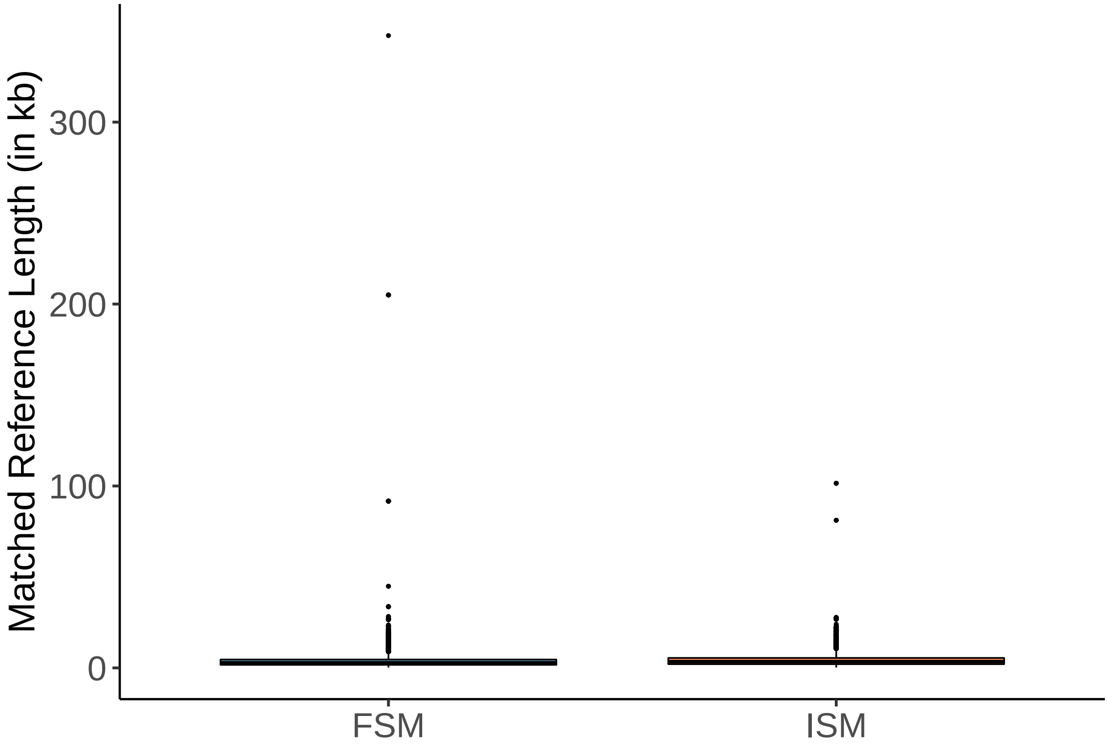

Exon Counts by Structural Classification
Number of exons
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

FL Count (normalized) by Structural Category



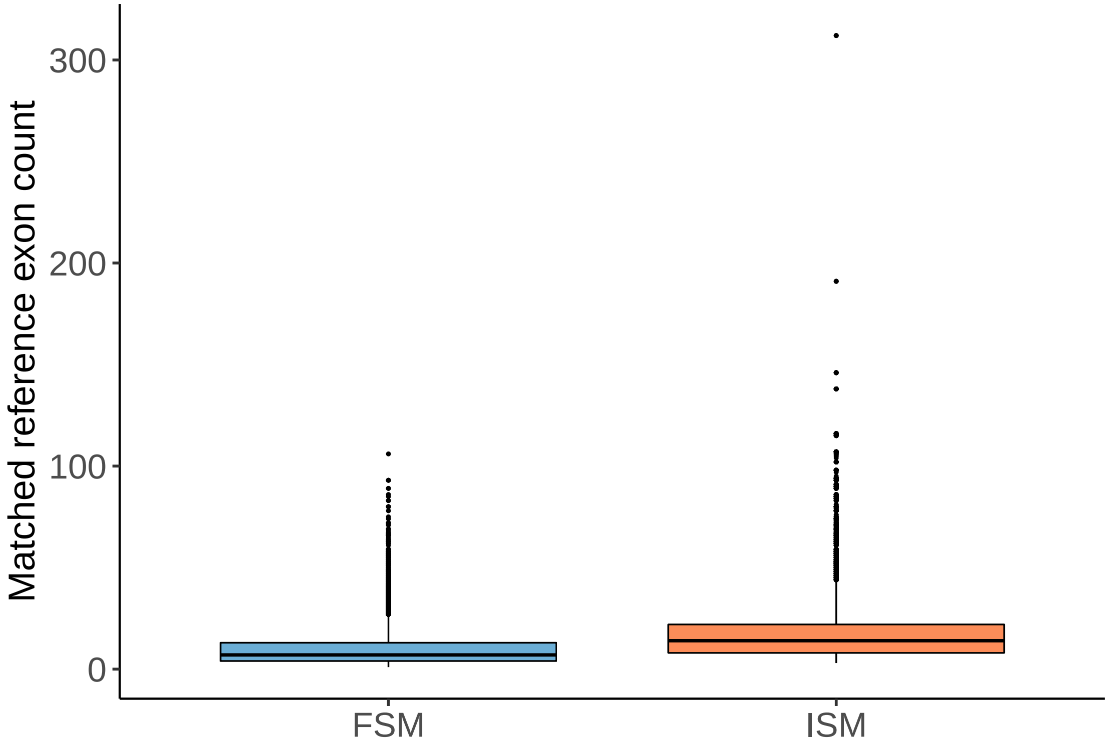
Length Distribution of Matched Reference Transcripts

Applicable only to FSM and ISM categories

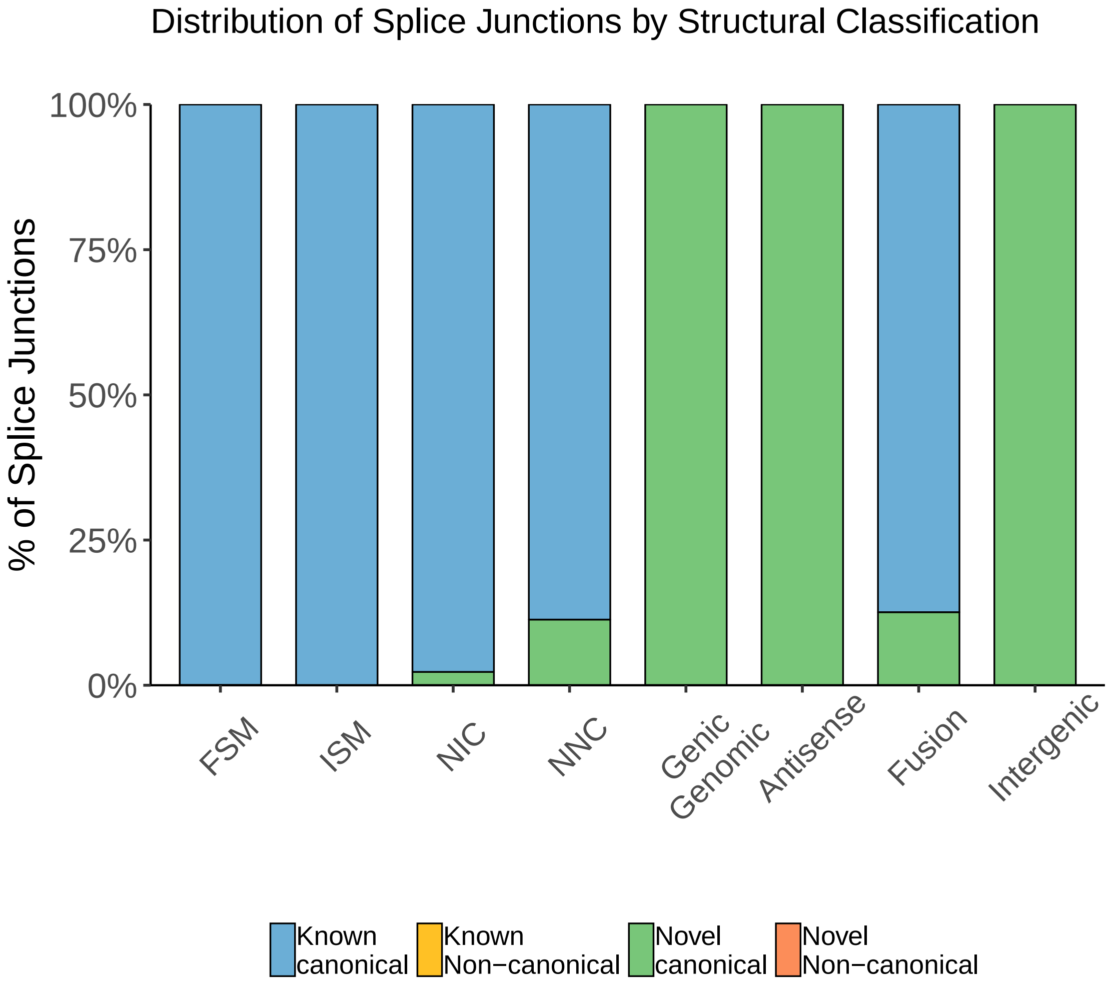


Exon Count Distribution of Matched Reference Transcripts

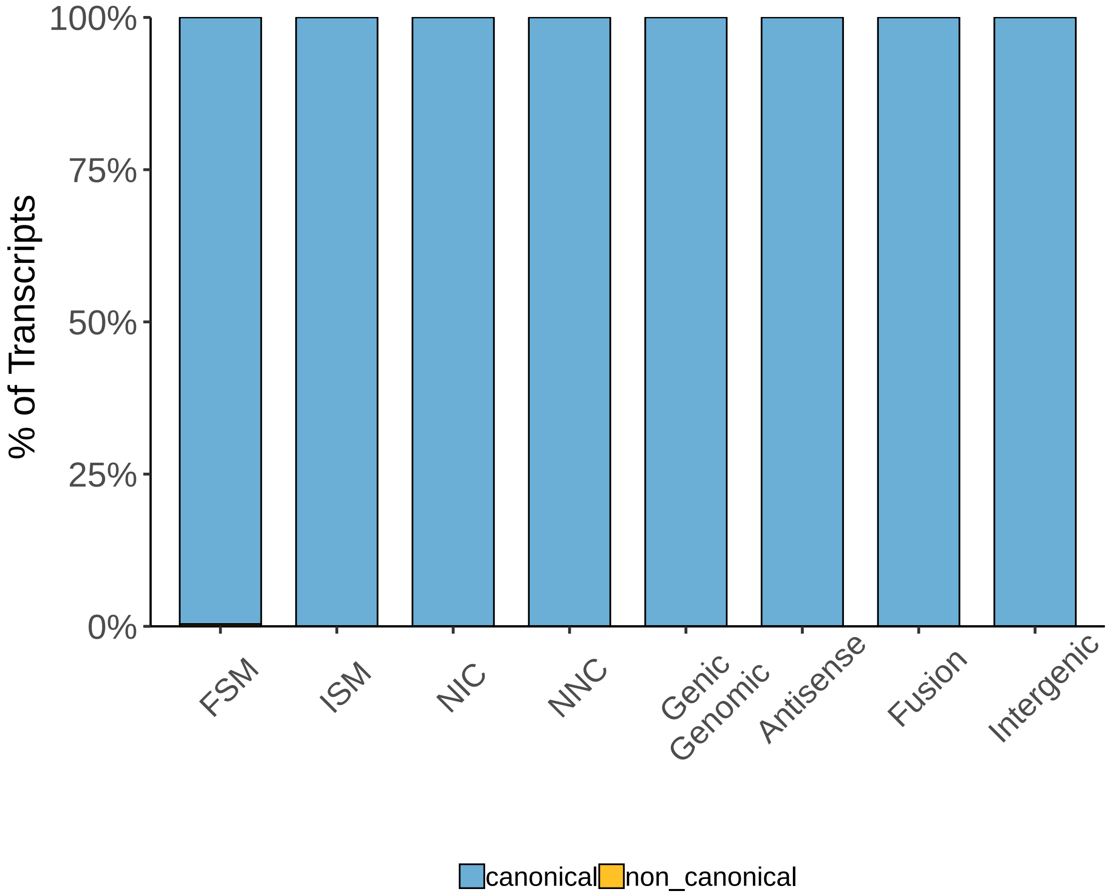
Applicable only to FSM and ISM categories

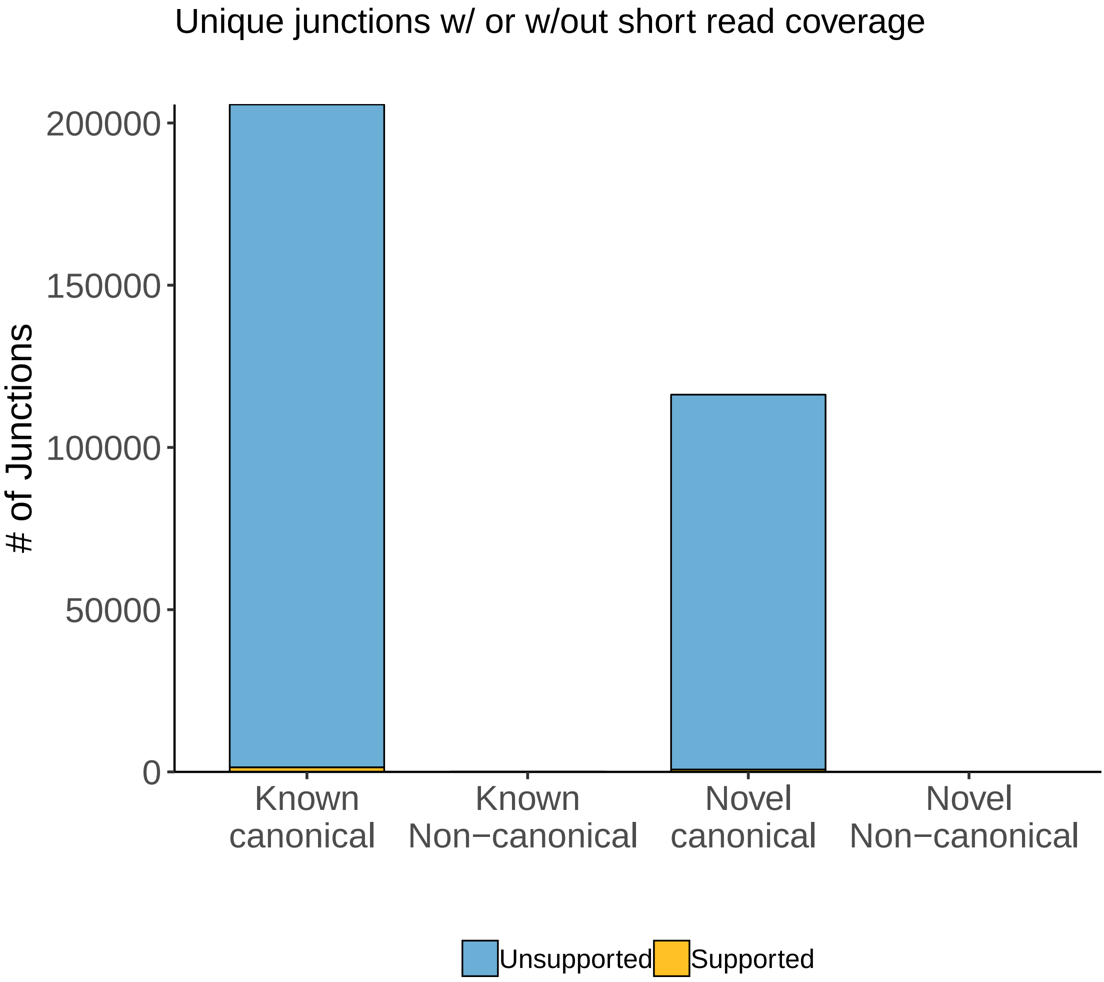


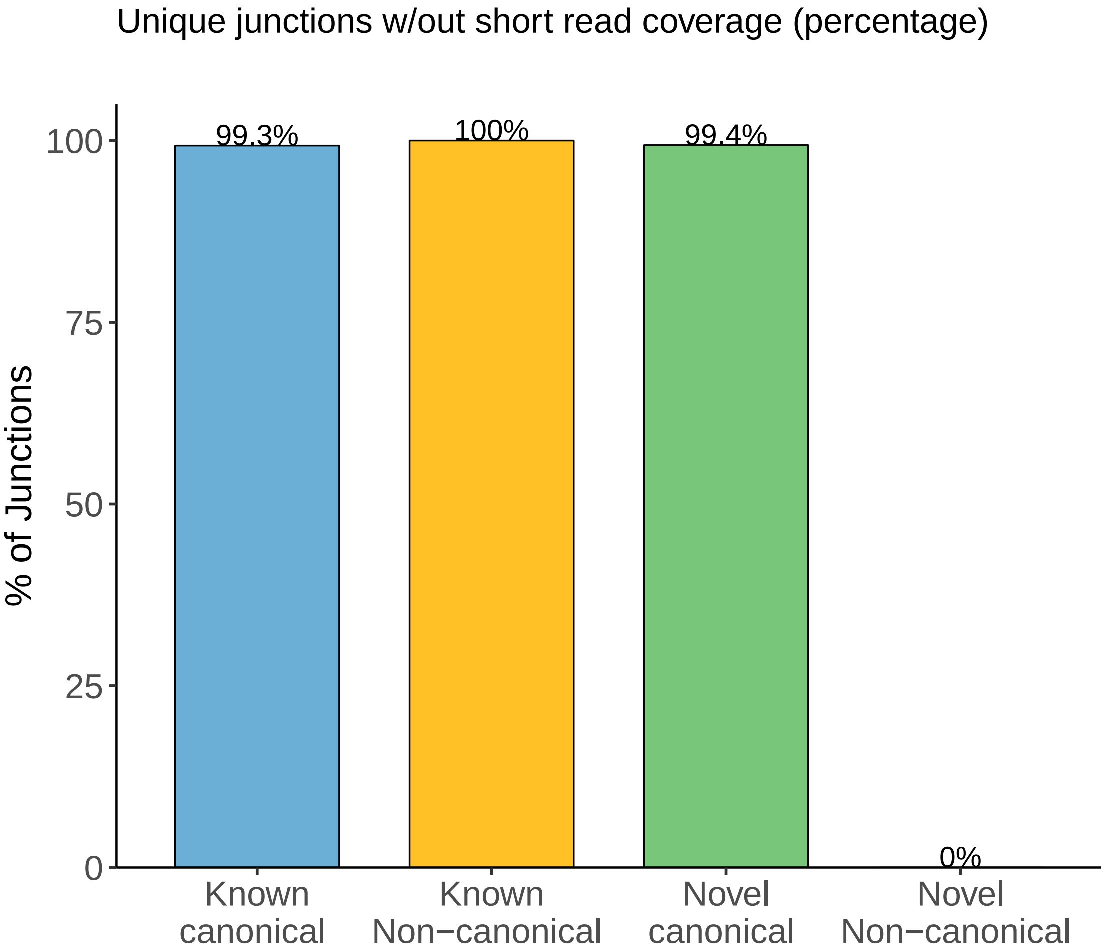
Splice Junction Characterization

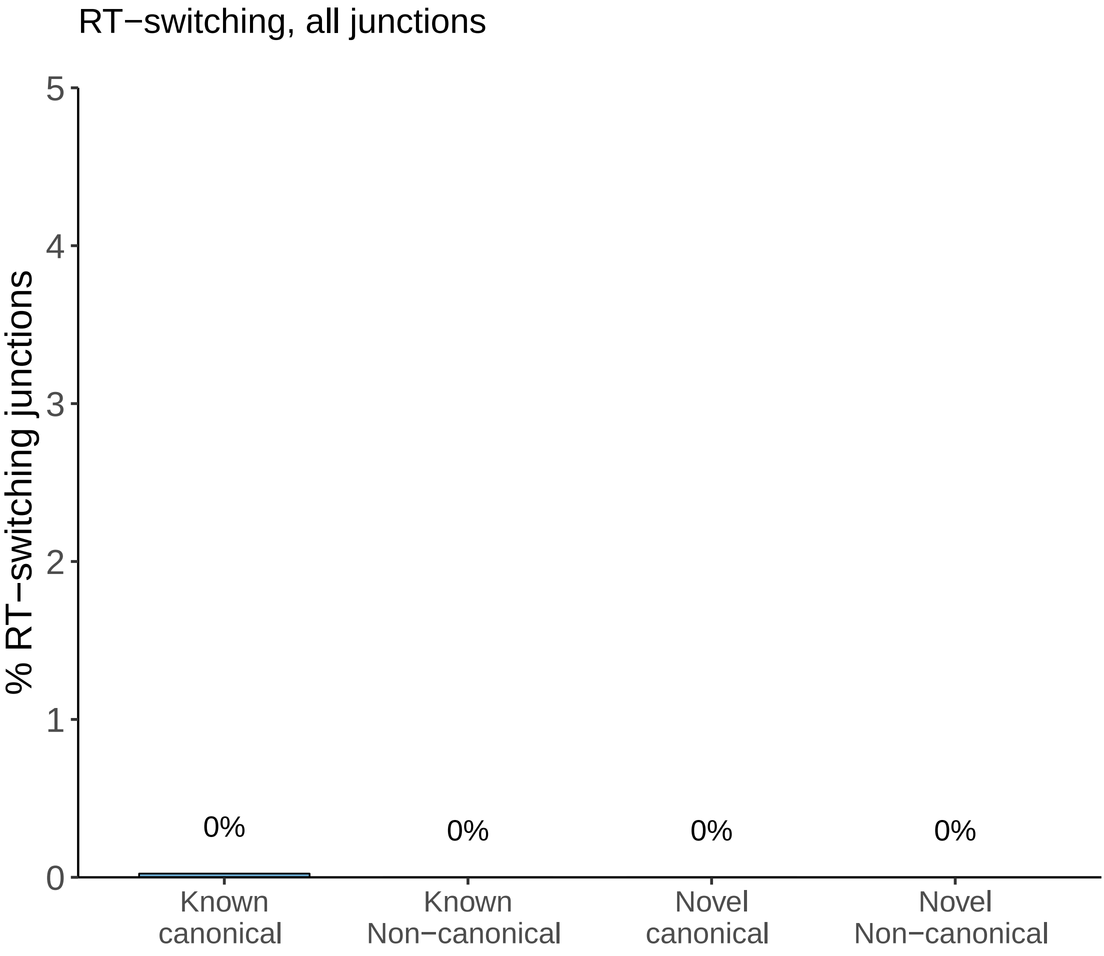

Distribution of Splice Junctions by Structural Classification
% of Splice Junctions
100%
75%
50%
25%
0%
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

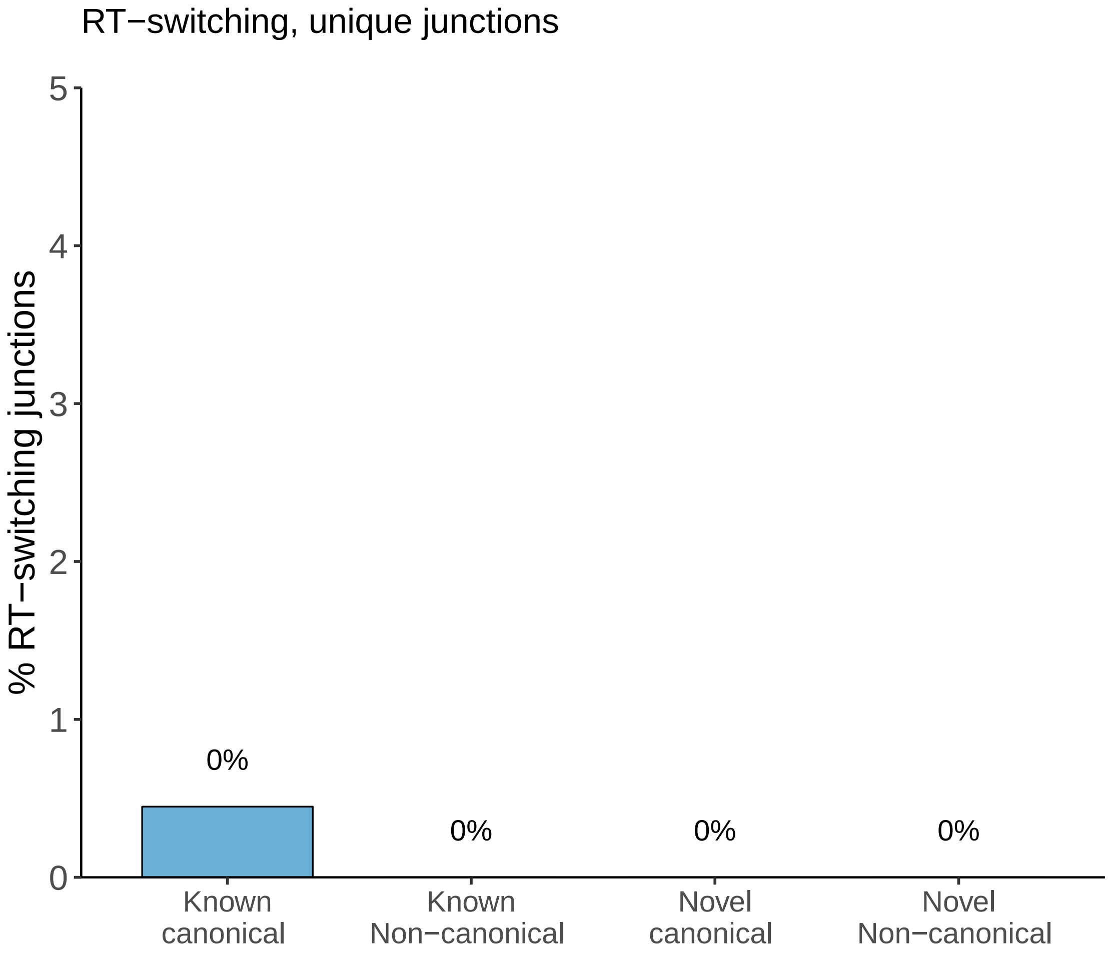
Distribution of Transcripts by Splice Junctions



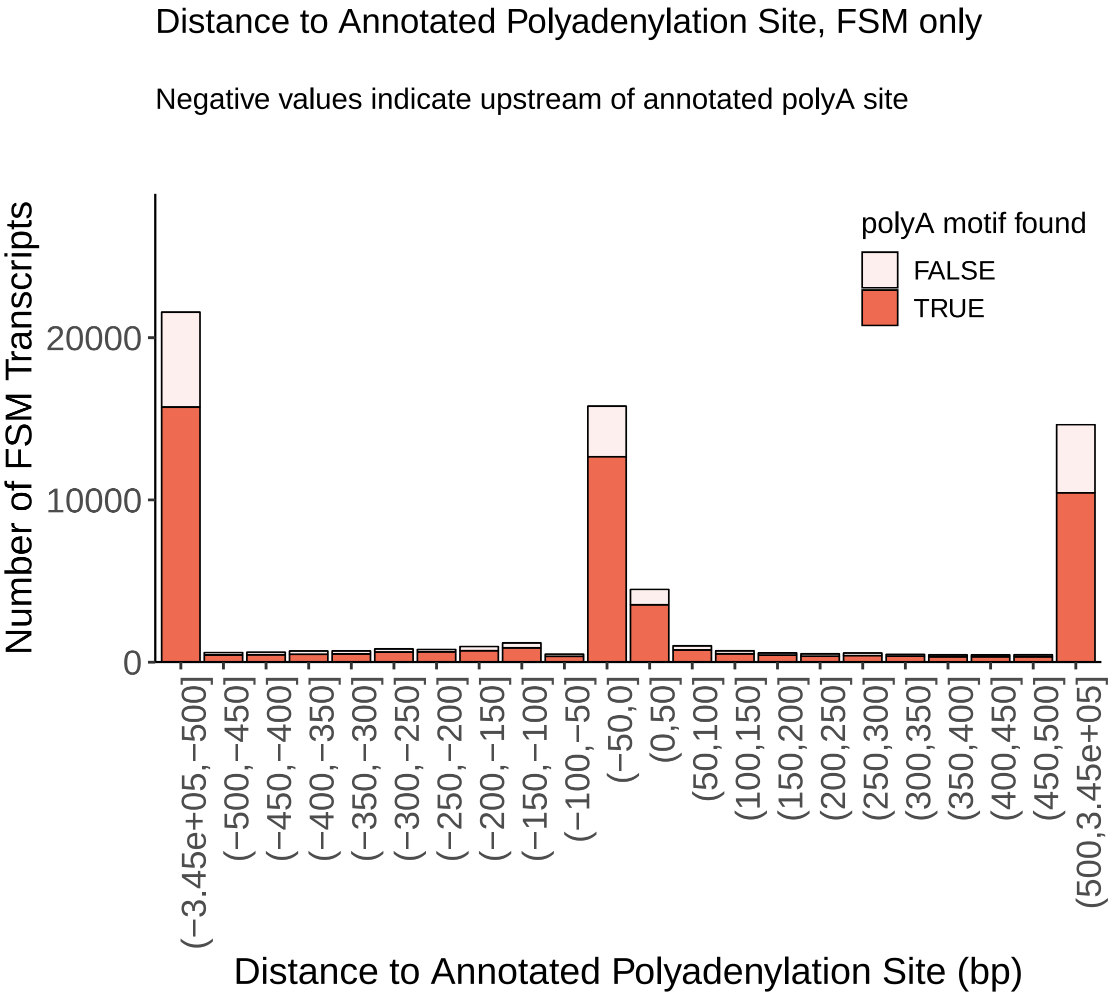

Unique junctions w/ or w/out short read coverage
# of Junctions
200000
150000
100000
50000
0
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical
Unsupported
Supported


Unique junctions w/out short read coverage (percentage)
% of Junctions
100
75
50
25
0
99.3%
100%
99.4%
0%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical


RT−switching, all junctions
% RT−switching junctions
5
4
3
2
1
0
0%
0%
0%
0%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

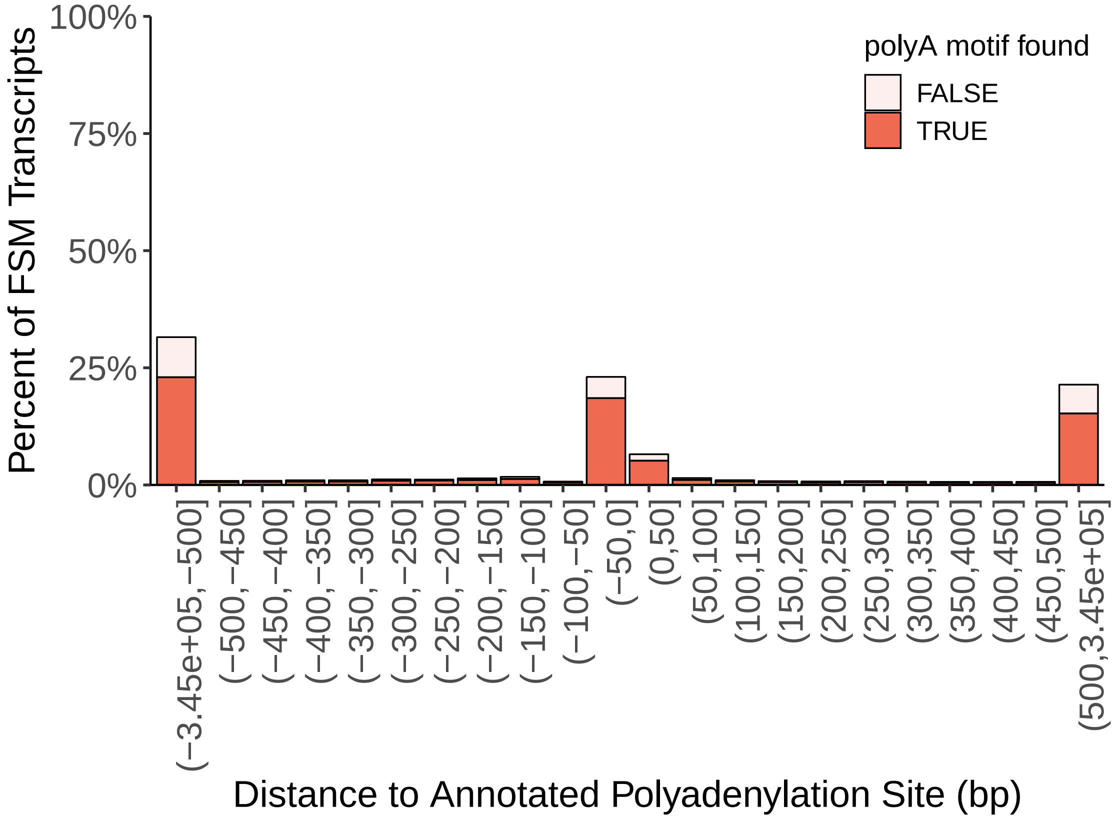

RT−switching, unique junctions
% RT−switching junctions
5
4
3
2
1
0
0%
0%
0%
0%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

Comparison with Annotated TSS and PolyA Sites



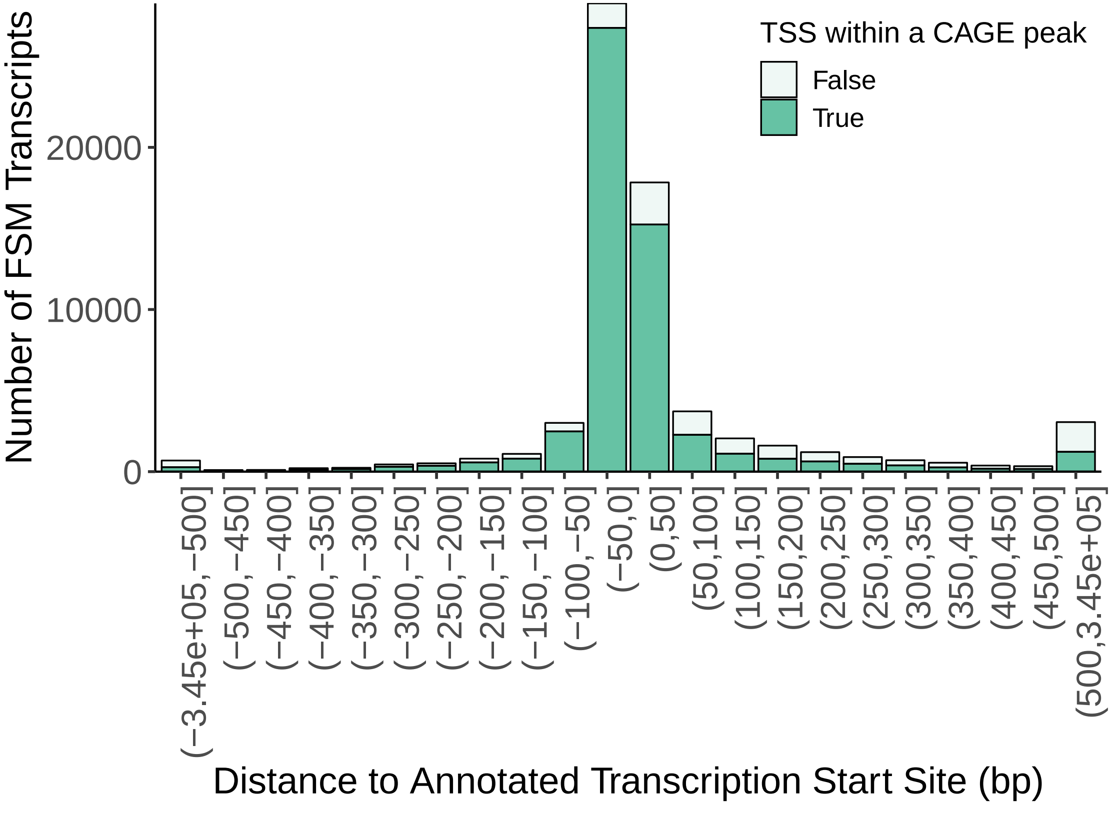
Distance to Annotated Polyadenylation Site, FSM only

Negative values indicate upstream of annotated polyA site



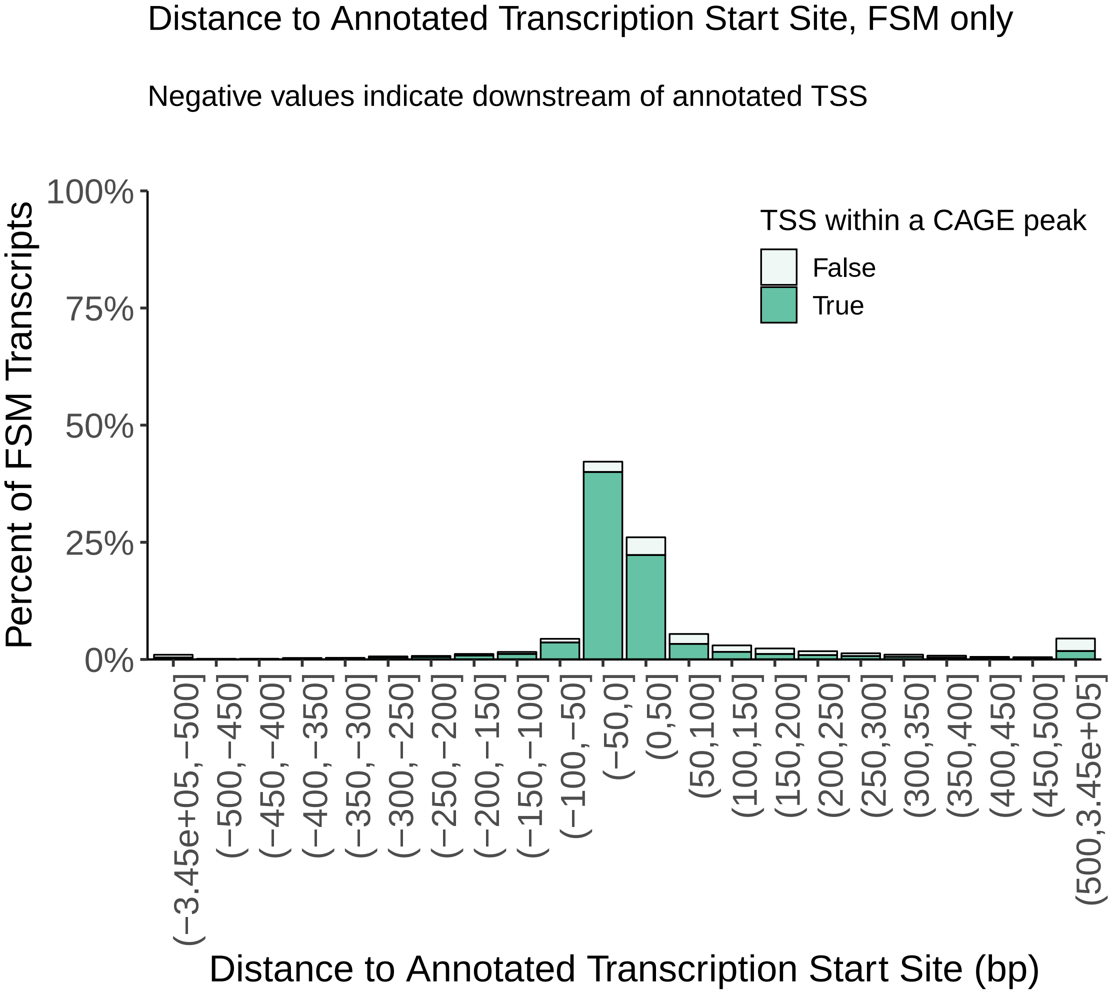
Distance to Annotated Polyadenylation Site, FSM only

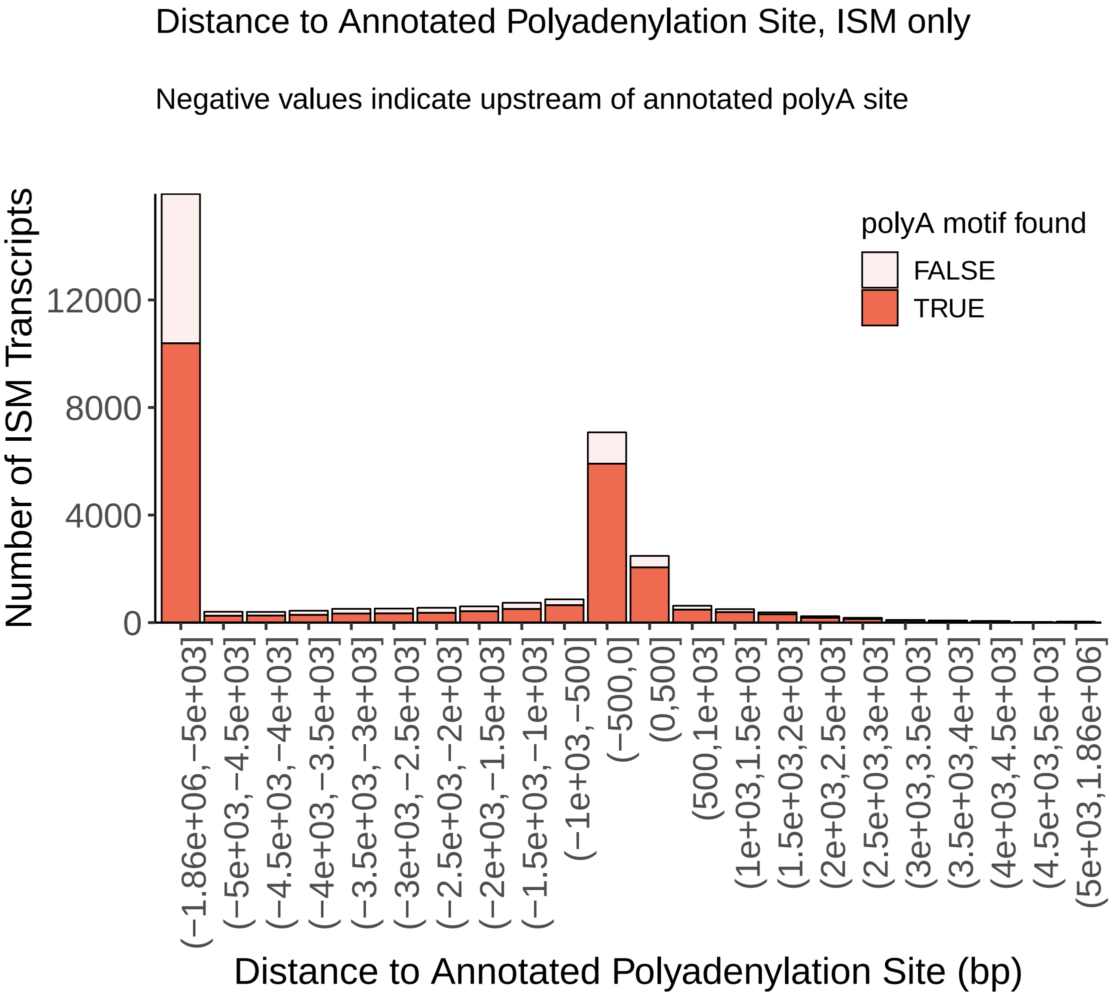
Negative values indicate upstream of annotated polyA site



Distance to Annotated Transcription Start Site, FSM only

Negative values indicate downstream of annotated TSS

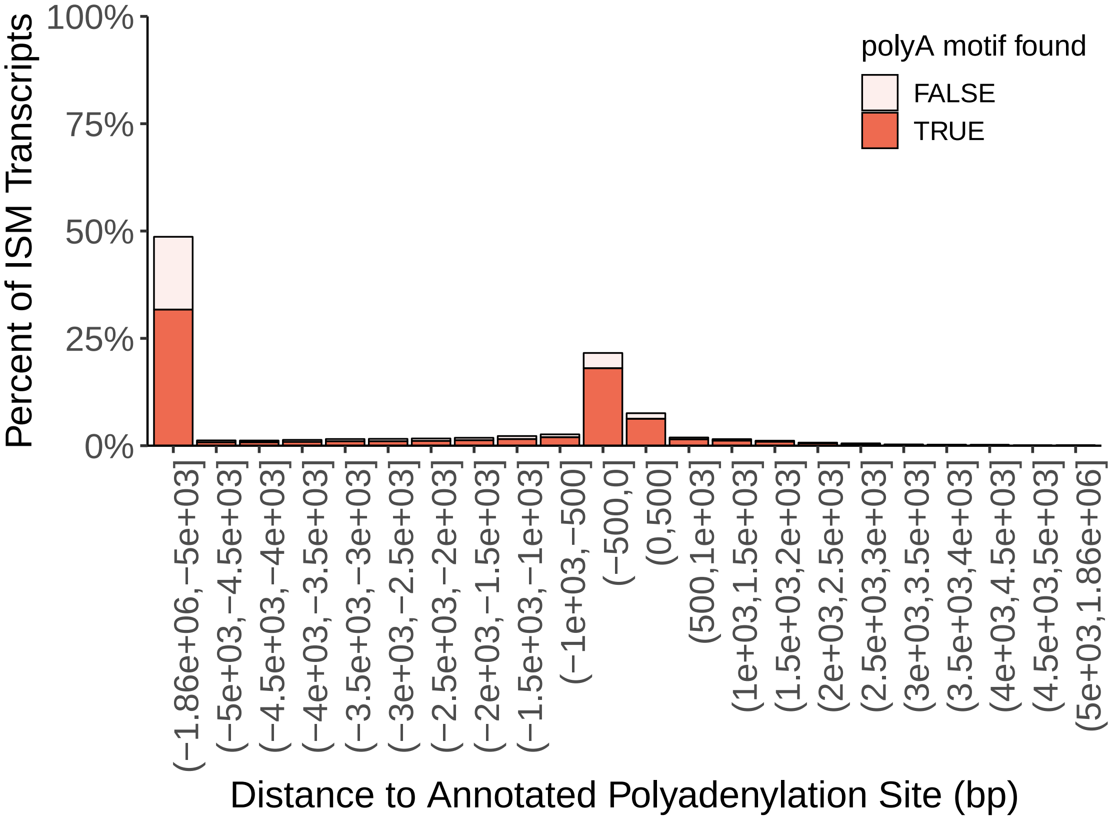

Distance to Annotated Transcription Start Site, FSM only
Negative values indicate downstream of annotated TSS
Percent of FSM Transcripts
100%
75%
50%
25%
0%
TSS within a CAGE peak
False
True
(−3.45e+05,−500]
(−500,−450]
(−450,−400]
(−400,−350]
(−350,−300]
(−300,−250]
(−250,−200]
(−200,−150]
(−150,−100]
(−100,−50]
(−50,0]
(0,50]
(50,100]
(100,150]
(150,200]
(200,250]
(250,300]
(300,350]
(350,400]
(400,450]
(450,500]
(500,3.45e+05]
Distance to Annotated Transcription Start Site (bp)

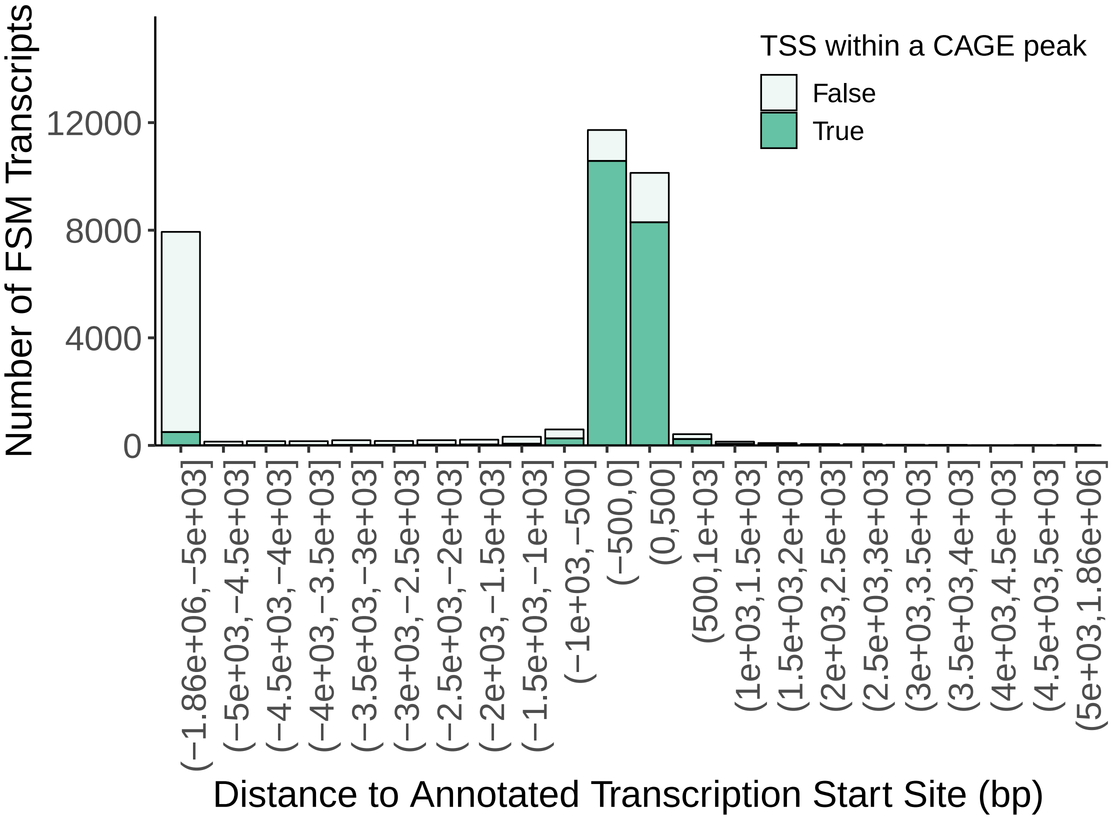

Distance to Annotated Polyadenylation Site, ISM only

Negative values indicate upstream of annotated polyA site

Distance to Annotated Polyadenylation Site, ISM only

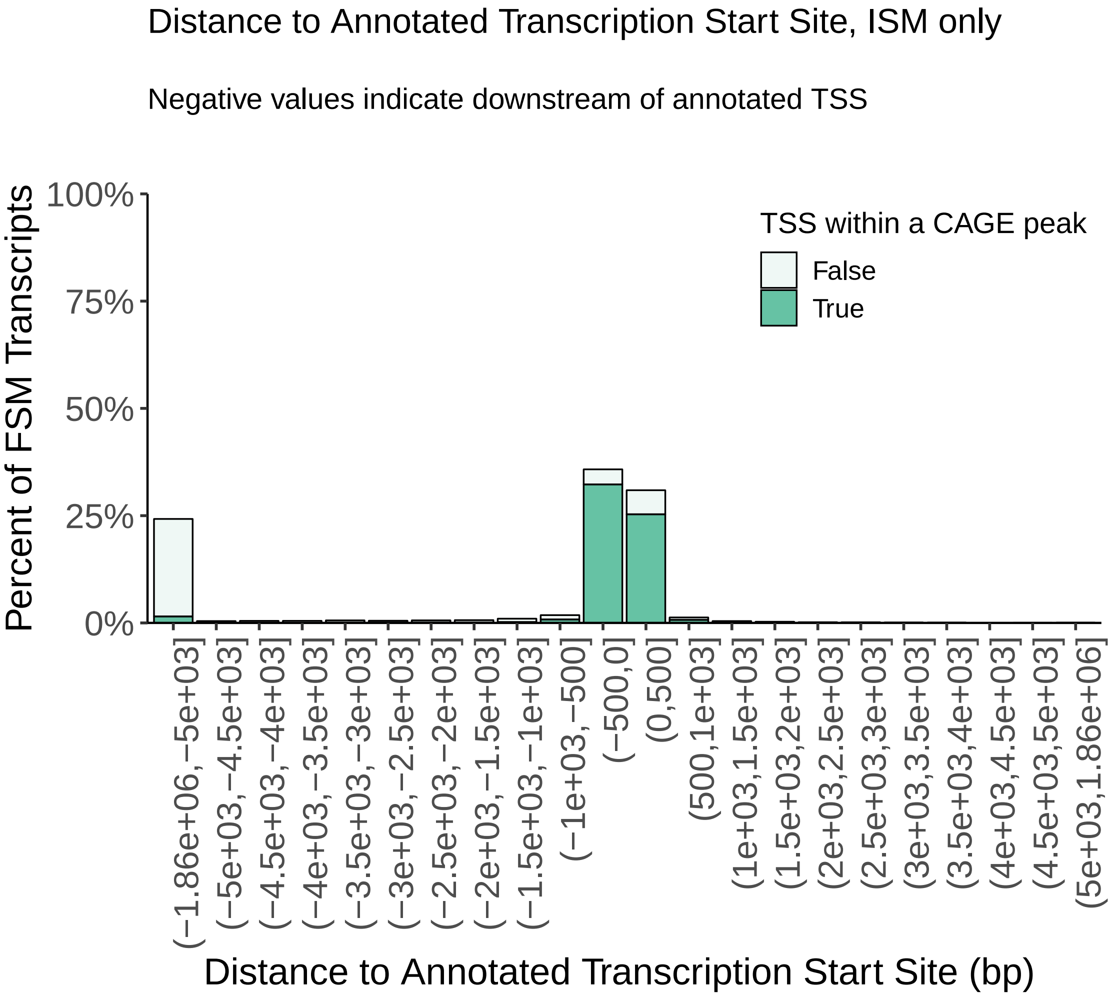
Negative values indicate upstream of annotated polyA site

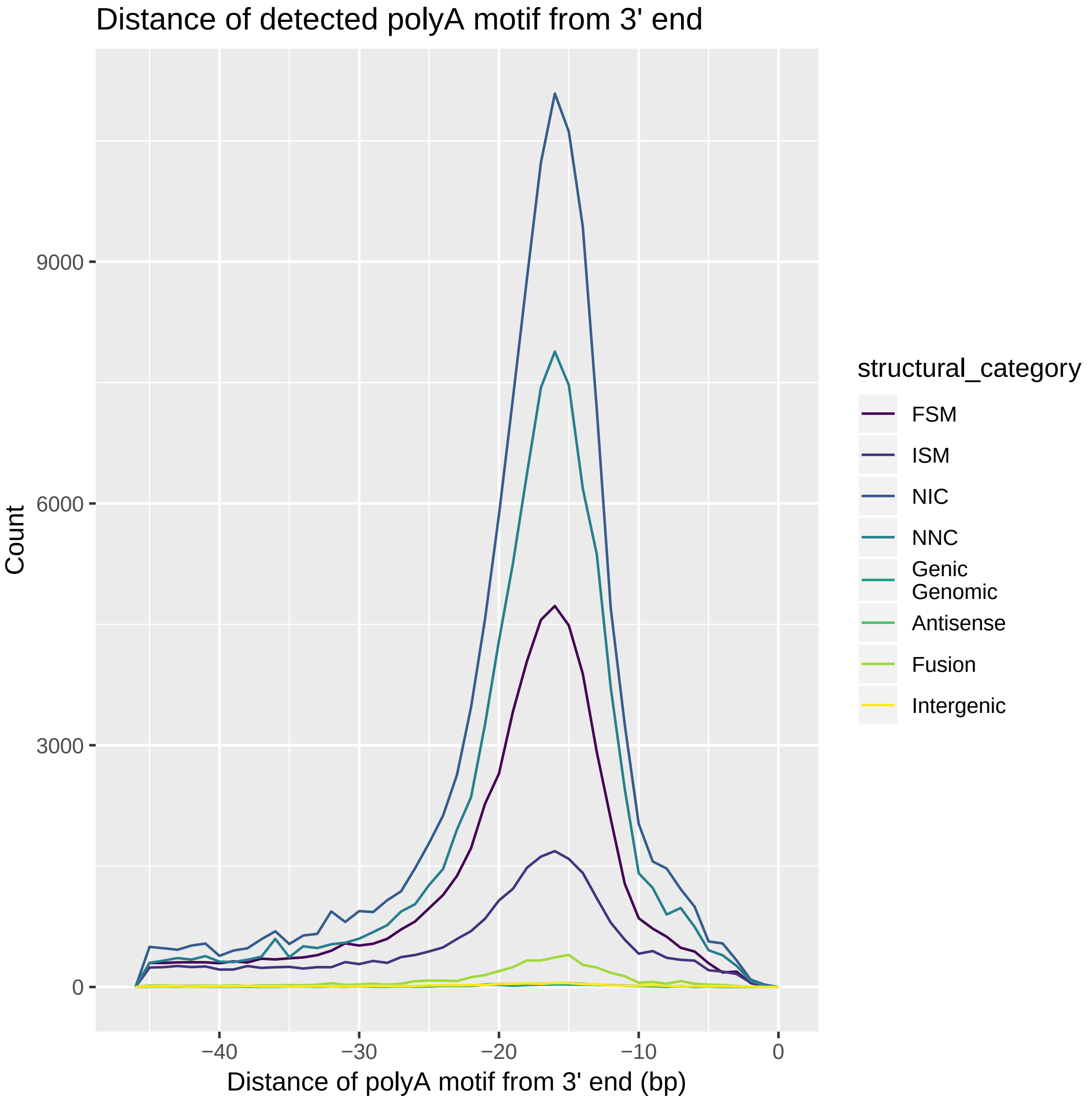




Distance to Annotated Transcription Start Site, ISM only

Negative values indicate downstream of annotated TSS


Distance to Annotated Transcription Start Site, ISM only
Negative values indicate downstream of annotated TSS
Percent of FSM Transcripts
100%
75%
50%
25%
0%
TSS within a CAGE peak
False
True
(−1.86e+06,−5e+03]
(−5e+03,−4.5e+03]
(−4.5e+03,−4e+03]
(−4e+03,−3.5e+03]
(−3.5e+03,−3e+03]
(−3e+03,−2.5e+03]
(−2.5e+03,−2e+03]
(−2e+03,−1.5e+03]
(−1.5e+03,−1e+03]
(−1e+03,−500]
(−500,0]
(0,500]
(500,1e+03]
(1e+03,1.5e+03]
(1.5e+03,2e+03]
(2e+03,2.5e+03]
(2.5e+03,3e+03]
(3e+03,3.5e+03]
(3.5e+03,4e+03]
(4e+03,4.5e+03]
(4.5e+03,5e+03]
(5e+03,1.86e+06]
Distance to Annotated Transcription Start Site (bp)


Distance of detected polyA motif from 3' end
Count
9000
6000
3000
0
−40
−30
−20
−10
0
Distance of polyA motif from 3' end (bp)
structural_category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic

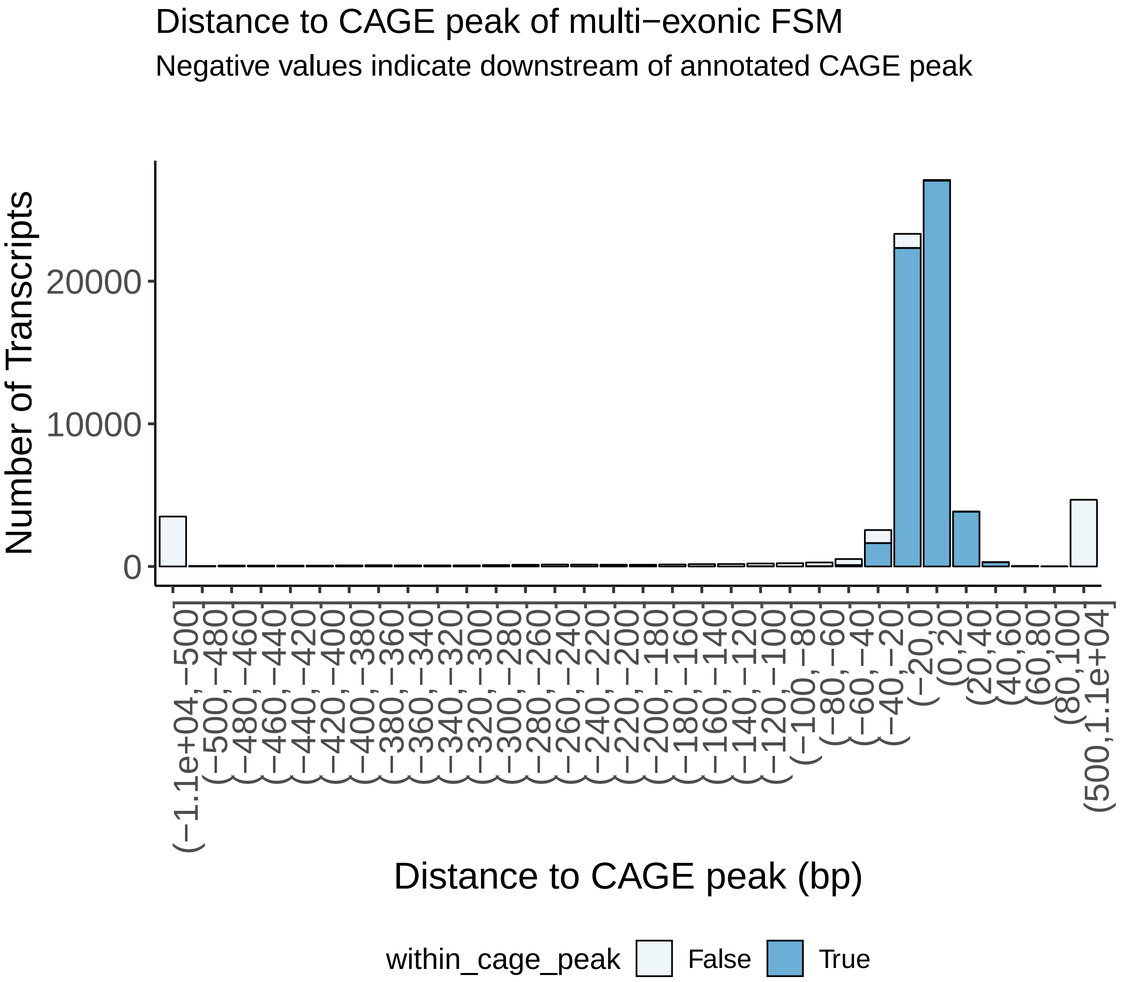
Frequency of polyA motifs

Motif	Count	%
AATAAA	170353	60.3
ATTAAA	45627	16.1
AGTAAA	10146	3.6
TATAAA	9319	3.3
AAGAAA	6112	2.2
TTTAAA	5972	2.1
AATATA	5174	1.8
CATAAA	4907	1.7
AATACA	4779	1.7
AAAAAG	4622	1.6
GATAAA	4489	1.6
AATGAA	3607	1.3
AAAACA	2554	0.9
AATAGA	2269	0.8
ACTAAA	2171	0.8
GGGGCT	486	0.2

Number of polyA Motifs Detected

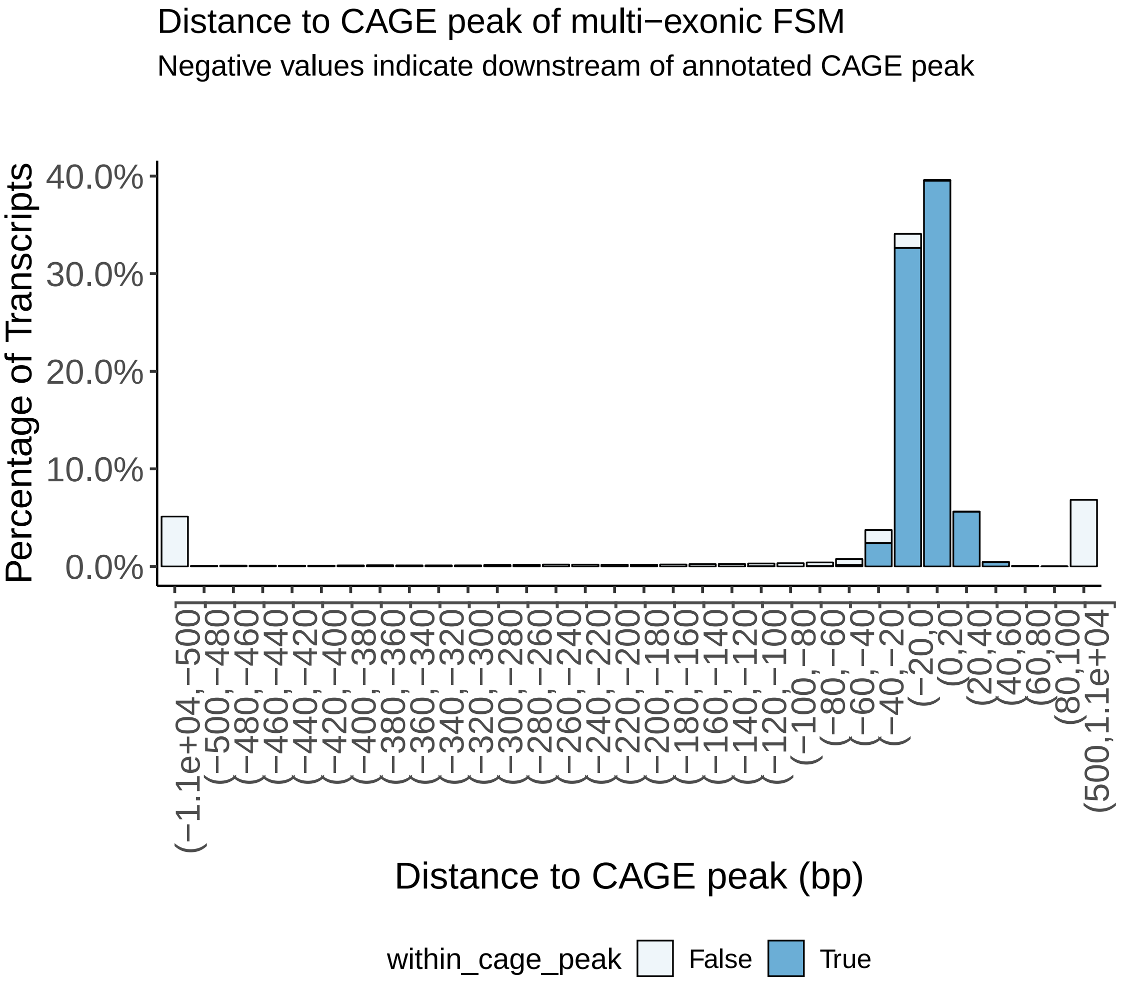
Category	Count	polyA Detected	%
FSM	72322	53797	74
ISM	32762	23522	72
NIC	141721	116110	82
NNC	103728	83289	80
Genic Genomic	679	441	65
Antisense	1075	611	57
Fusion	5527	4055	73
Intergenic	1374	762	55

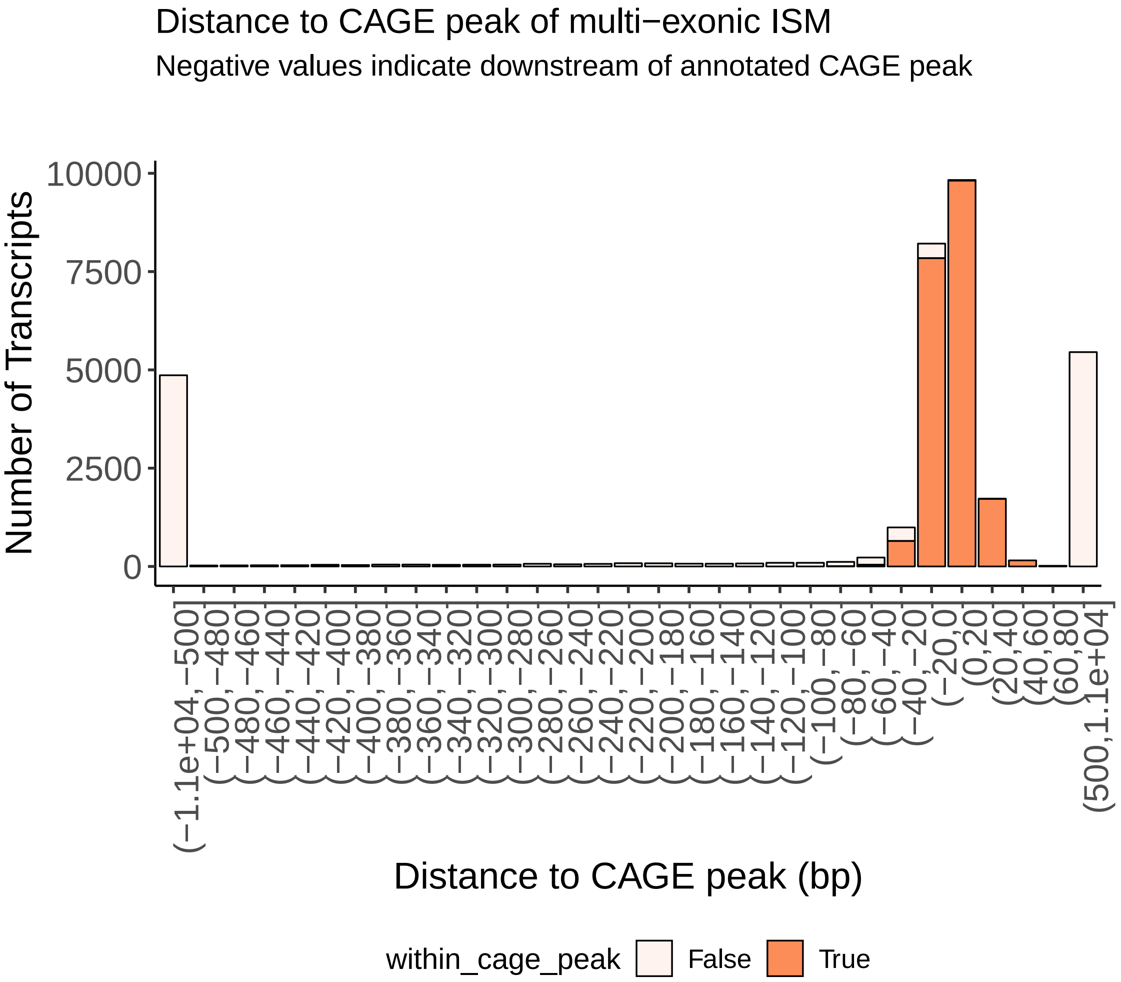
CAGE distances analysis



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
Percentage of Transcripts
40.0%
30.0%
20.0%
10.0%
0.0%
(−1.1e+04,−500]
(−500,−480]
(−480,−460]
(−460,−440]
(−440,−420]
(−420,−400]
(−400,−380]
(−380,−360]
(−360,−340]
(−340,−320]
(−320,−300]
(−300,−280]
(−280,−260]
(−260,−240]
(−240,−220]
(−220,−200]
(−200,−180]
(−180,−160]
(−160,−140]
(−140,−120]
(−120,−100]
(−100,−80]
(−80,−60]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(40,60]
(60,80]
(80,100]
(500,1.1e+04]
Distance to CAGE peak (bp)
within_cage_peak
False
True

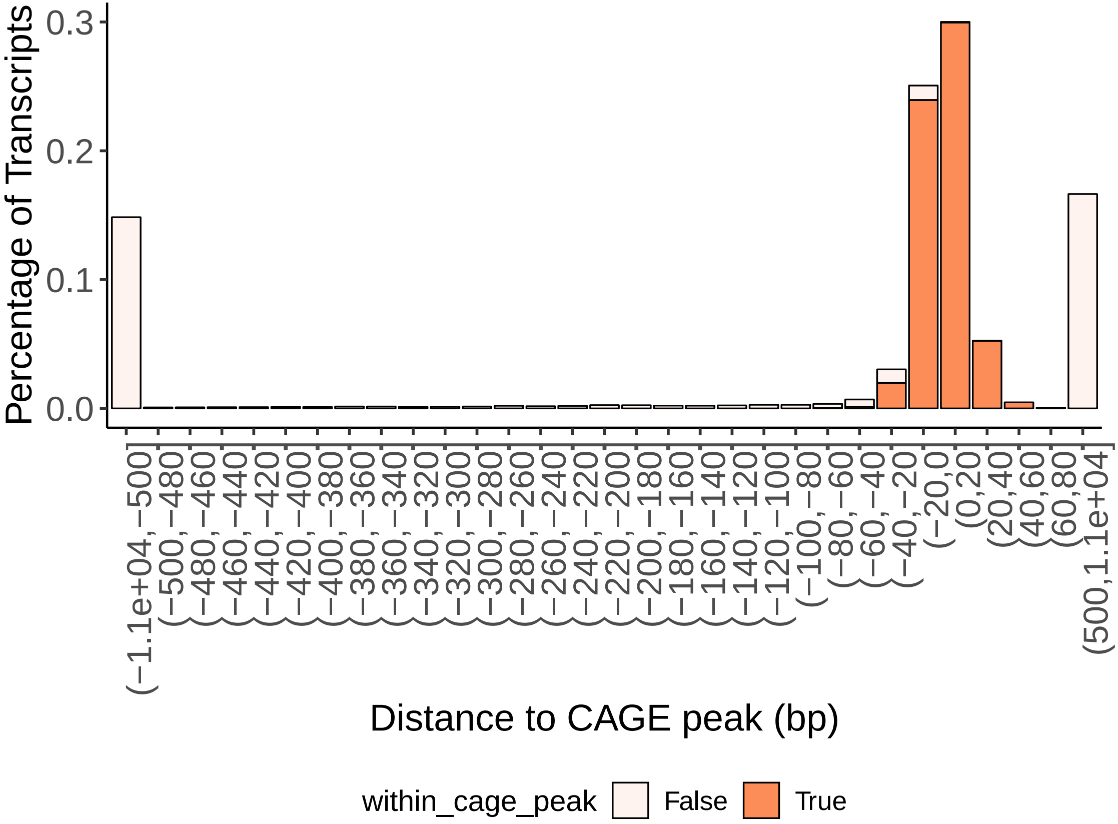


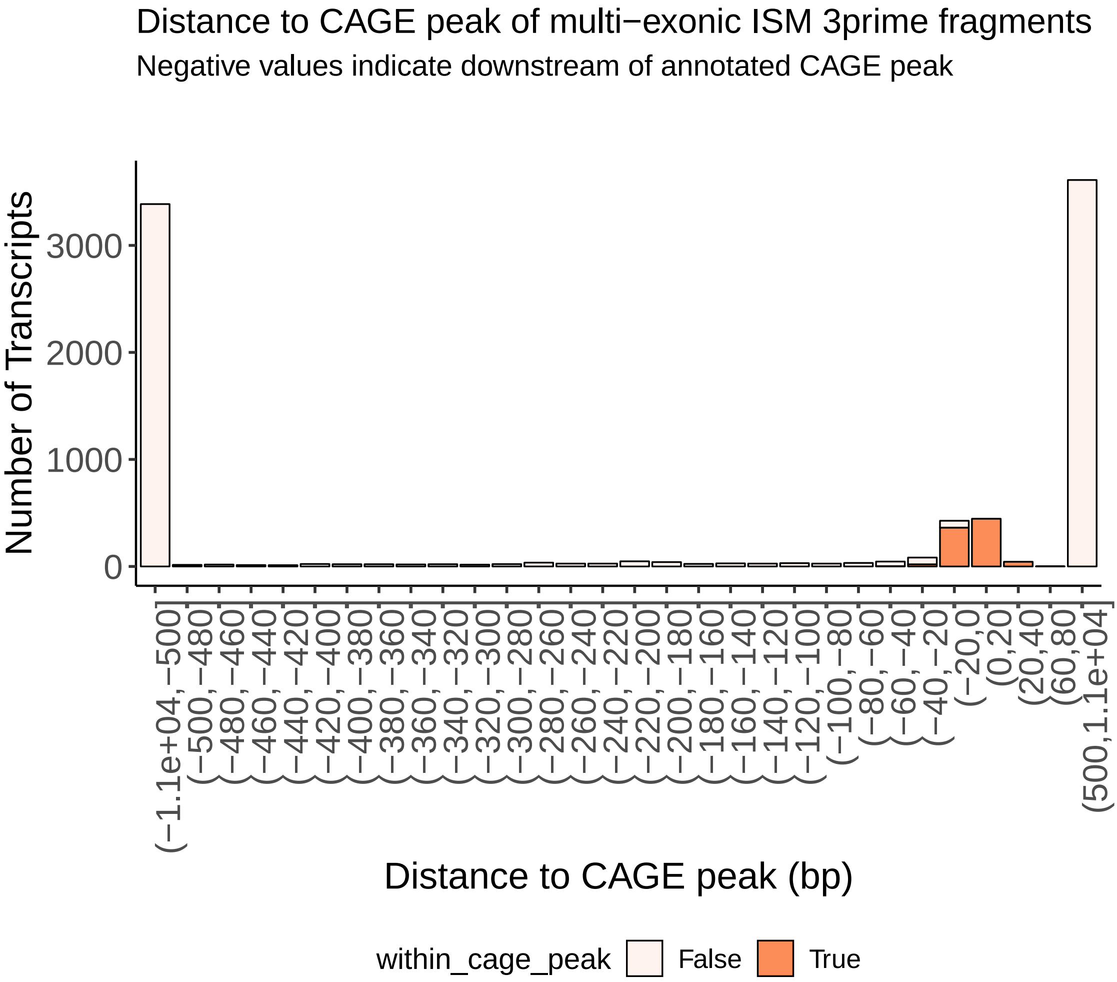
Distance to CAGE peak of multi−exonic ISM

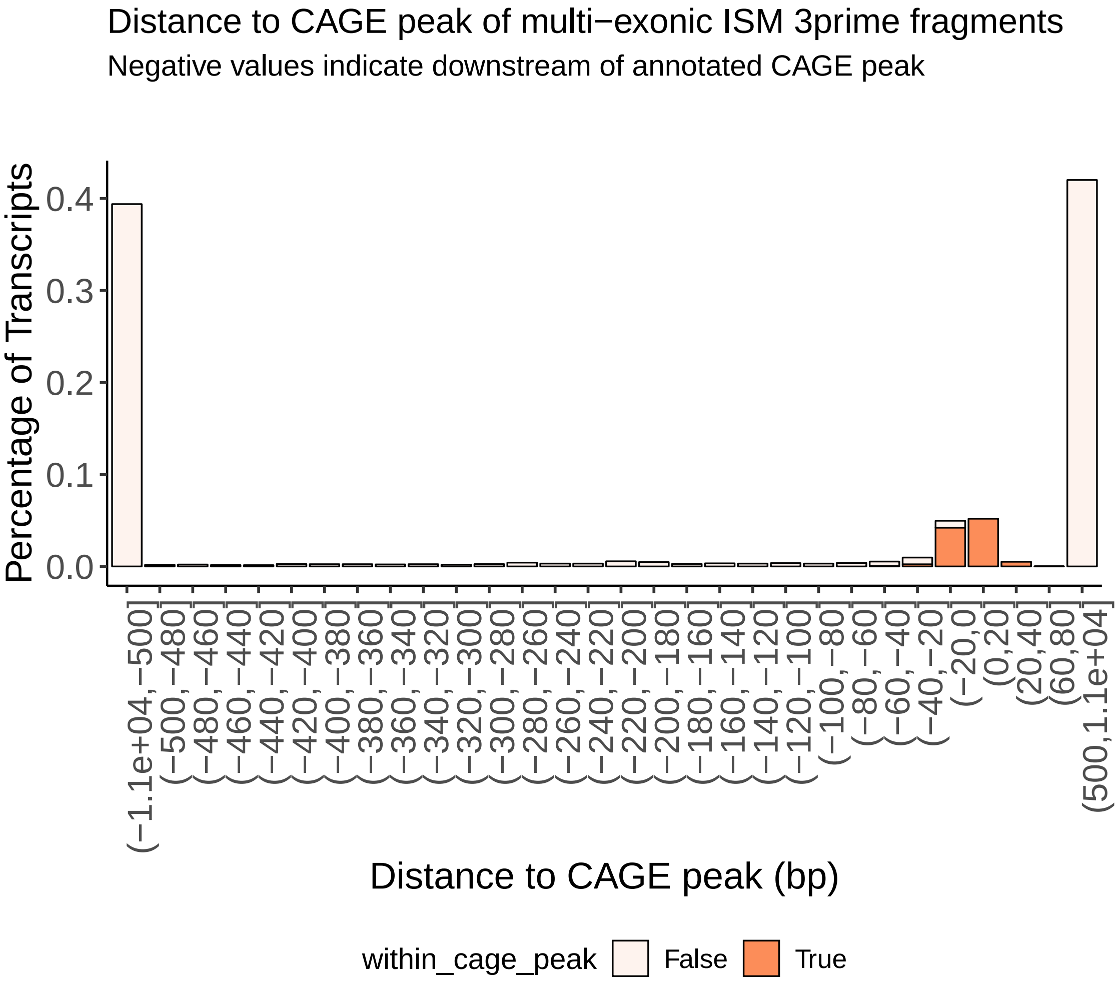
Negative values indicate downstream of annotated CAGE peak

Distance to CAGE peak of multi−exonic ISM

Negative values indicate downstream of annotated CAGE peak

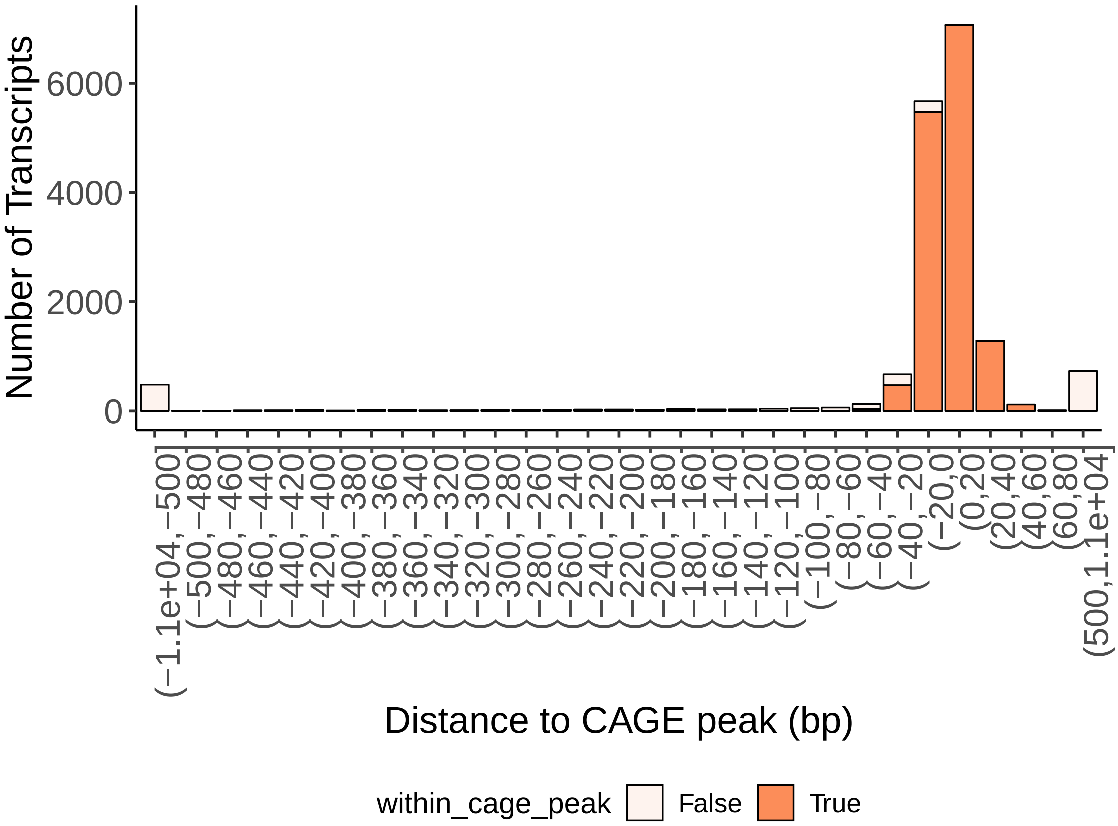



Distance to CAGE peak of multi−exonic ISM 3prime fragments
Negative values indicate downstream of annotated CAGE peak
Number of Transcripts
0
1000
2000
3000
(−1.1e+04,−500]
(−500,−480]
(−480,−460]
(−460,−440]
(−440,−420]
(−420,−400]
(−400,−380]
(−380,−360]
(−360,−340]
(−340,−320]
(−320,−300]
(−300,−280]
(−280,−260]
(−260,−240]
(−240,−220]
(−220,−200]
(−200,−180]
(−180,−160]
(−160,−140]
(−140,−120]
(−120,−100]
(−100,−80]
(−80,−60]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(60,80]
(500,1.1e+04]
Distance to CAGE peak (bp)
within_cage_peak
False
True


Distance to CAGE peak of multi−exonic ISM 3prime fragments
Negative values indicate downstream of annotated CAGE peak
Percentage of Transcripts
0.4
0.3
0.2
0.1
0.0
(−1.1e+04,−500]
(−500,−480]
(−480,−460]
(−460,−440]
(−440,−420]
(−420,−400]
(−400,−380]
(−380,−360]
(−360,−340]
(−340,−320]
(−320,−300]
(−300,−280]
(−280,−260]
(−260,−240]
(−240,−220]
(−220,−200]
(−200,−180]
(−180,−160]
(−160,−140]
(−140,−120]
(−120,−100]
(−100,−80]
(−80,−60]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(60,80]
(500,1.1e+04]
Distance to CAGE peak (bp)
within_cage_peak
False
True

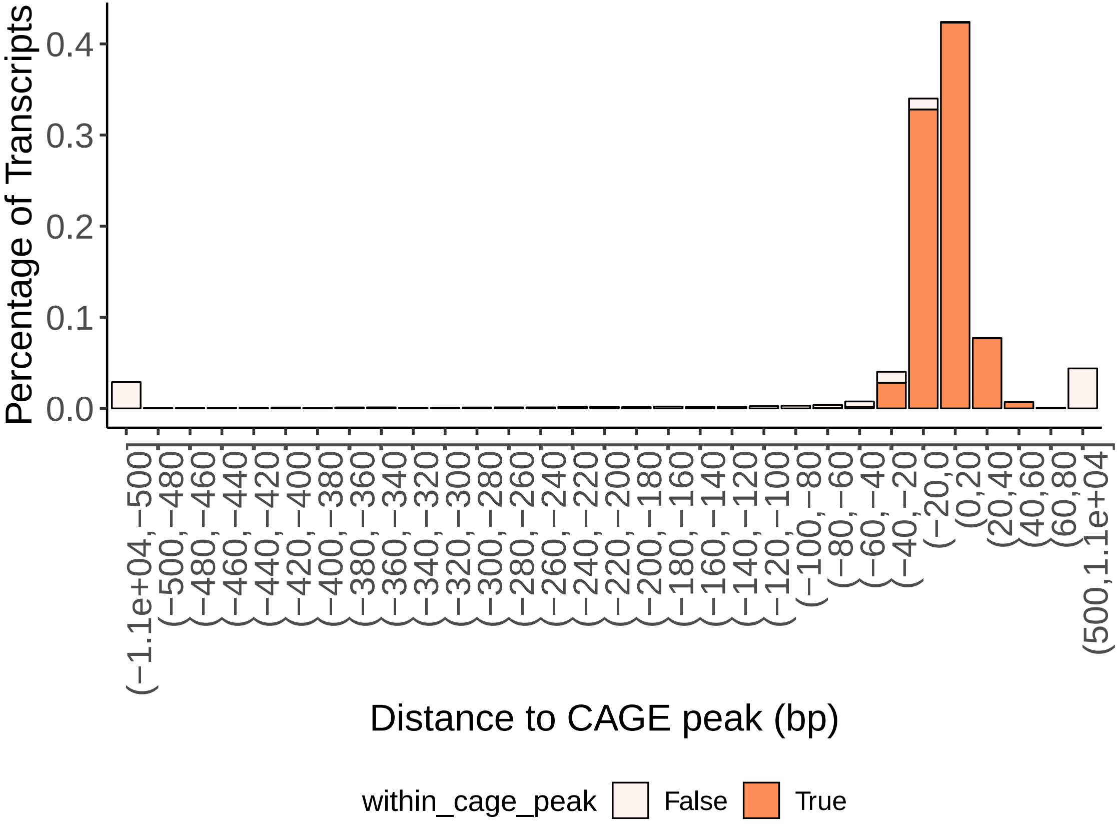
Distance to CAGE peak of multi−exonic ISM 5prime fragments

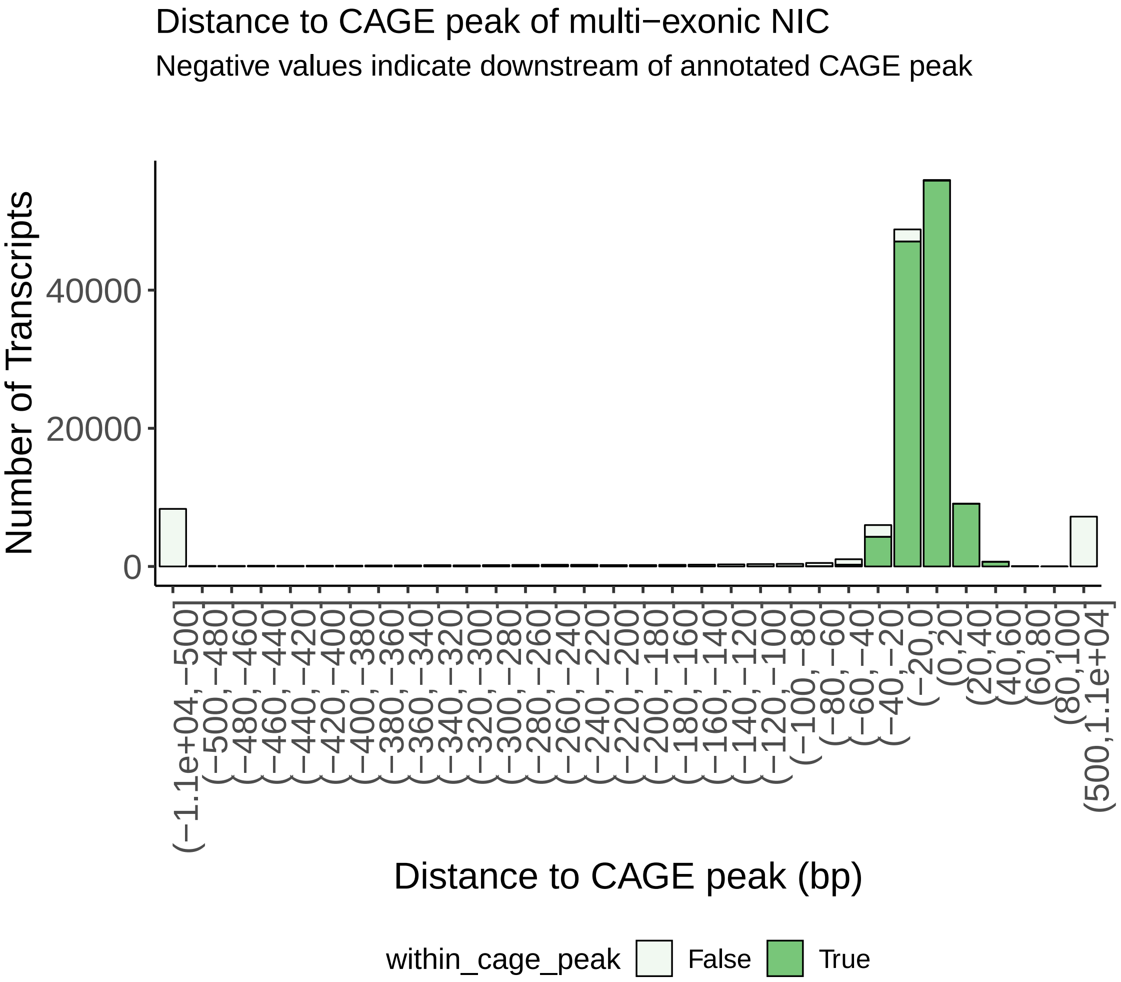
Negative values indicate downstream of annotated CAGE peak



Distance to CAGE peak of multi−exonic ISM 5prime fragments

Negative values indicate downstream of annotated CAGE peak



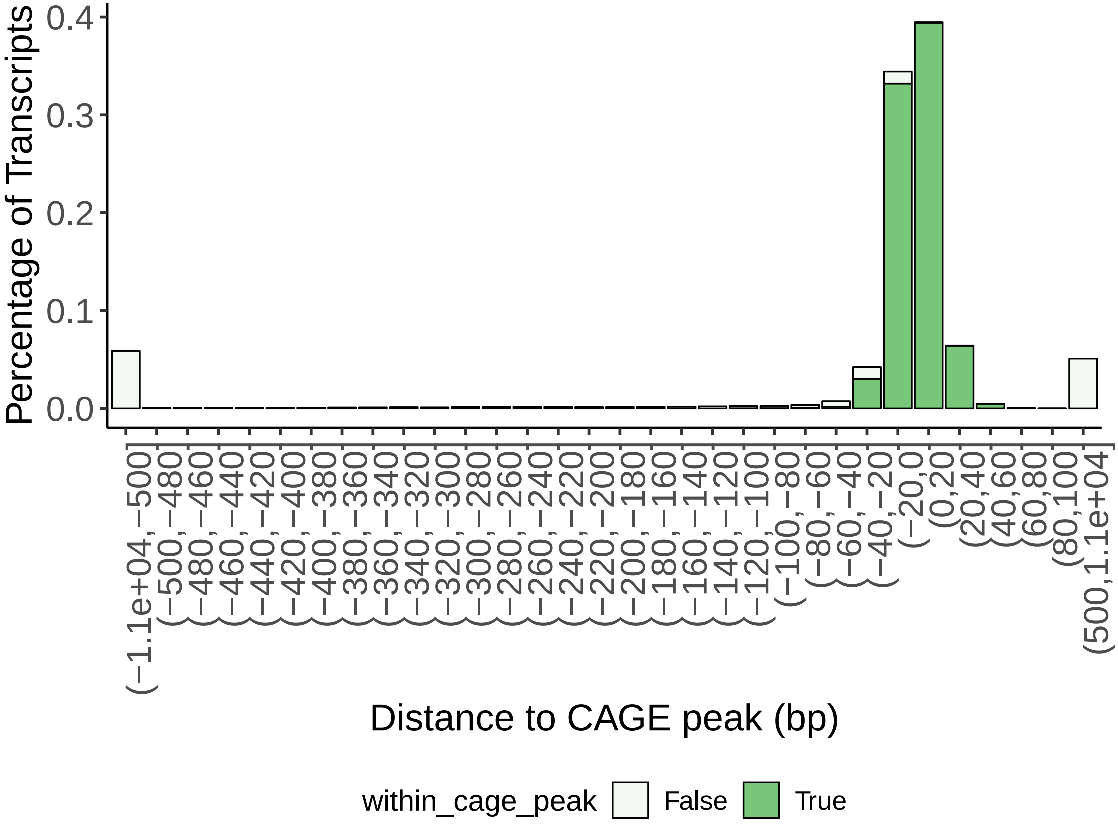


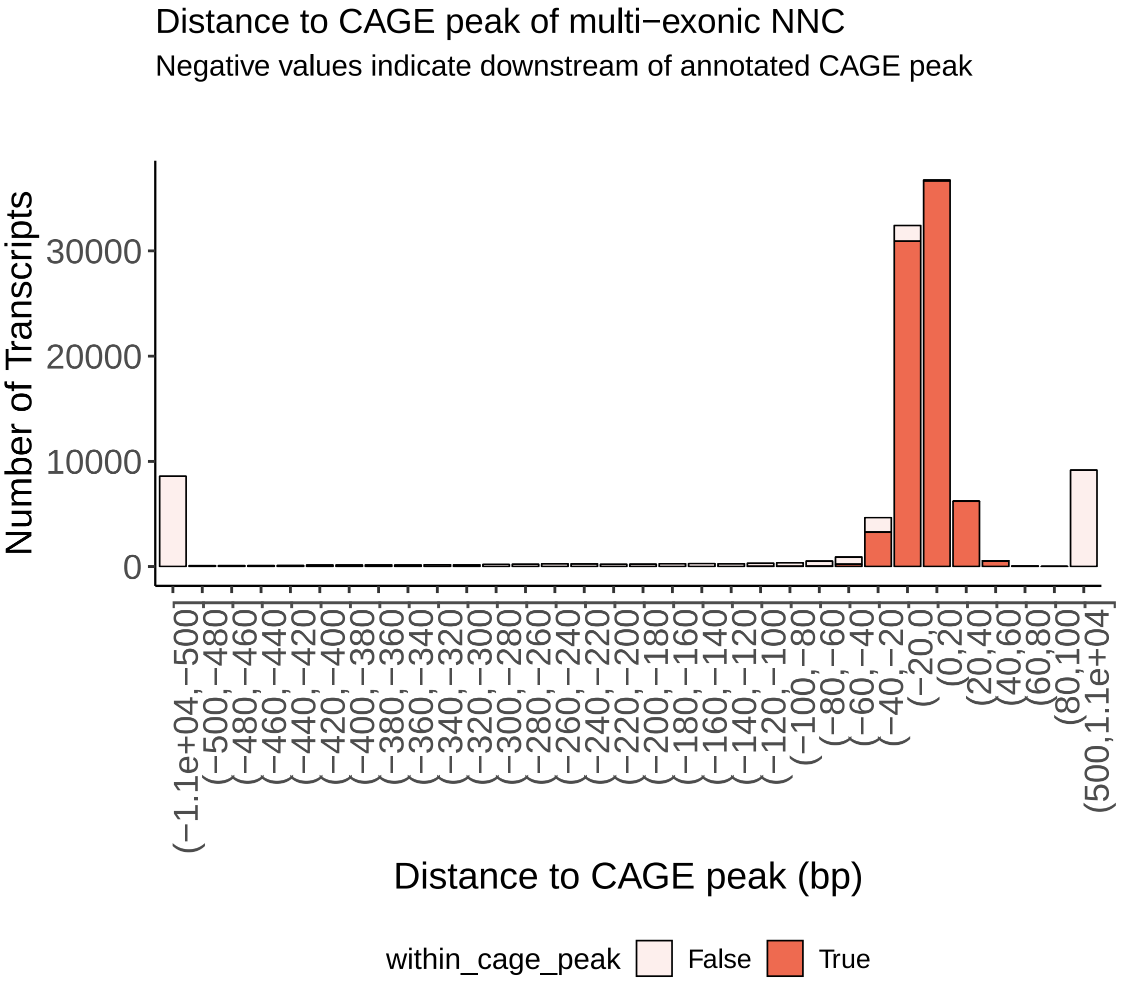
Distance to CAGE peak of multi−exonic NIC

Negative values indicate downstream of annotated CAGE peak

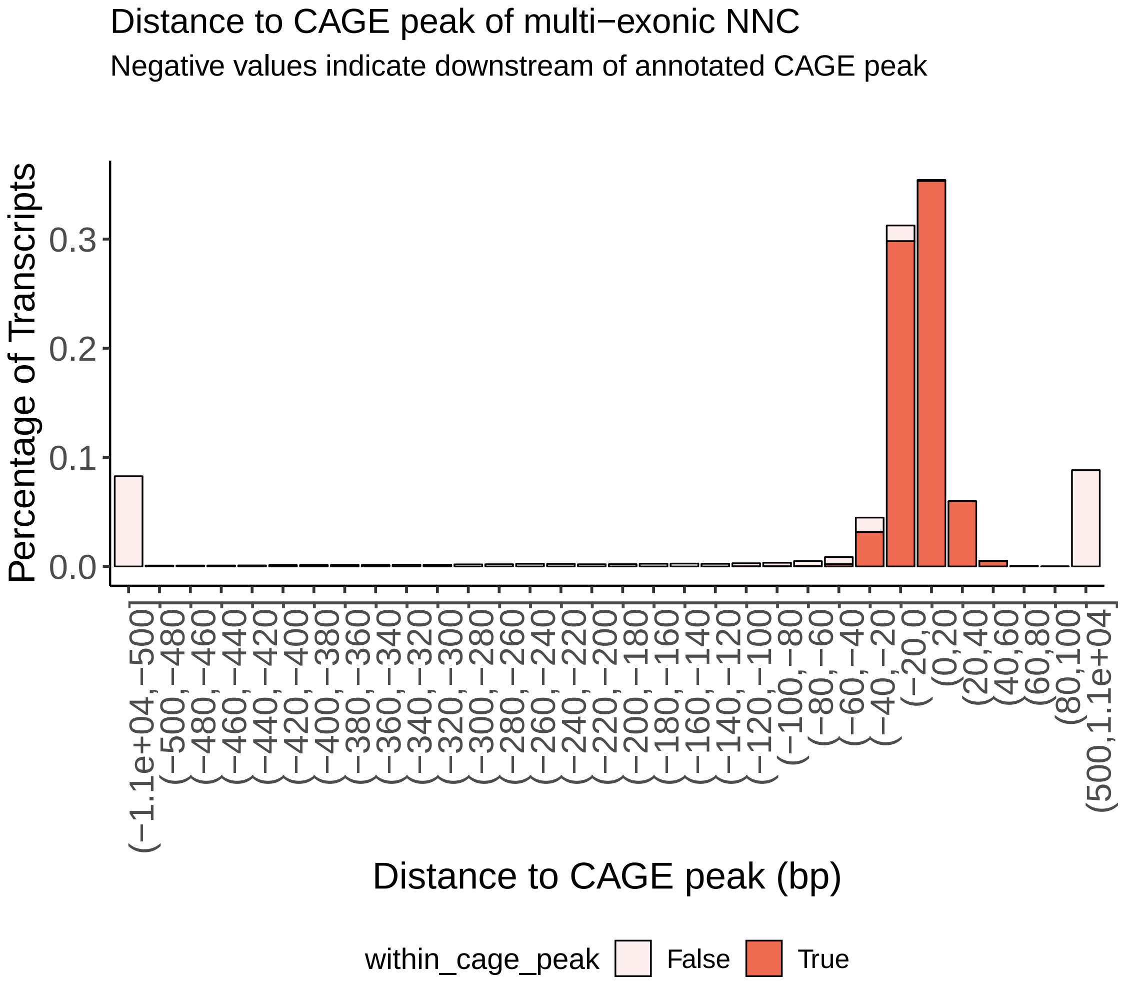
Distance to CAGE peak of multi−exonic NIC

Negative values indicate downstream of annotated CAGE peak




Distance to CAGE peak of multi−exonic NNC

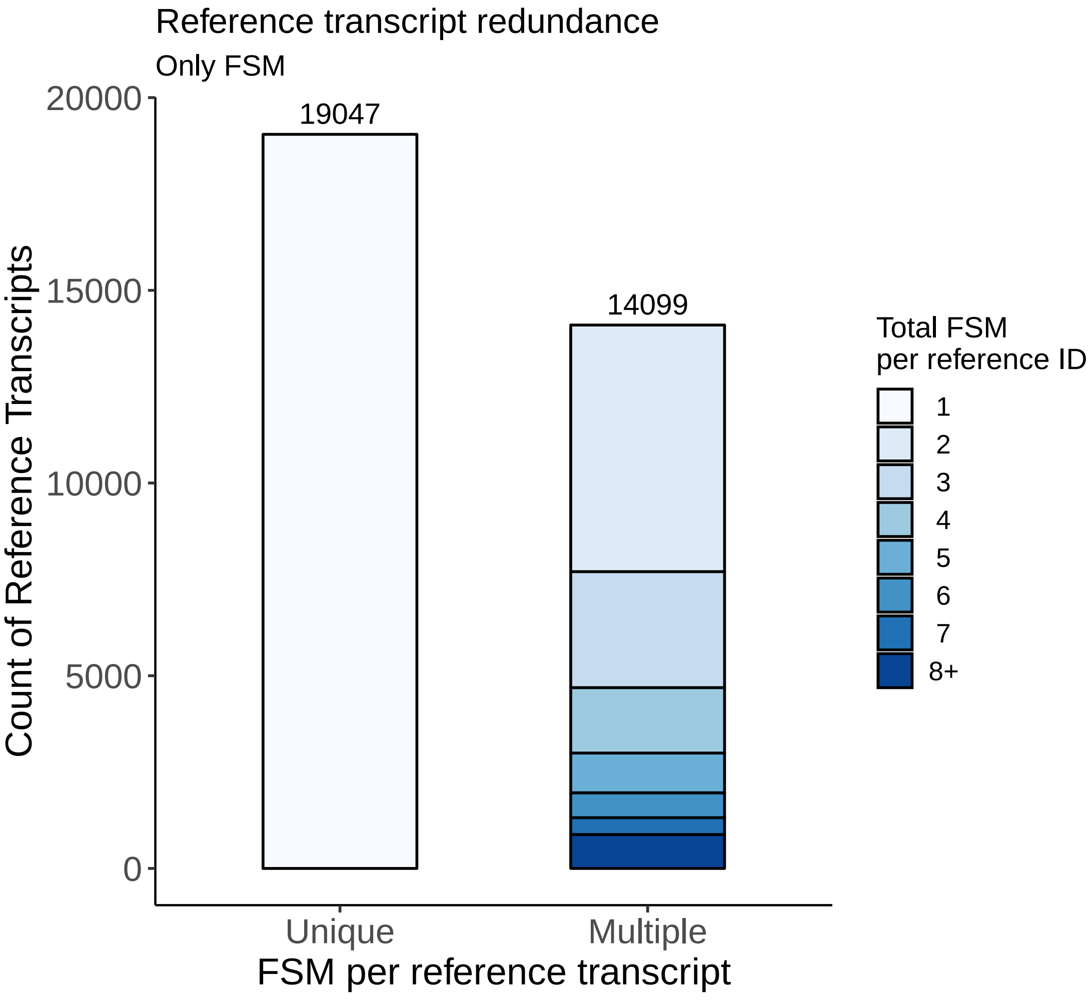
Negative values indicate downstream of annotated CAGE peak

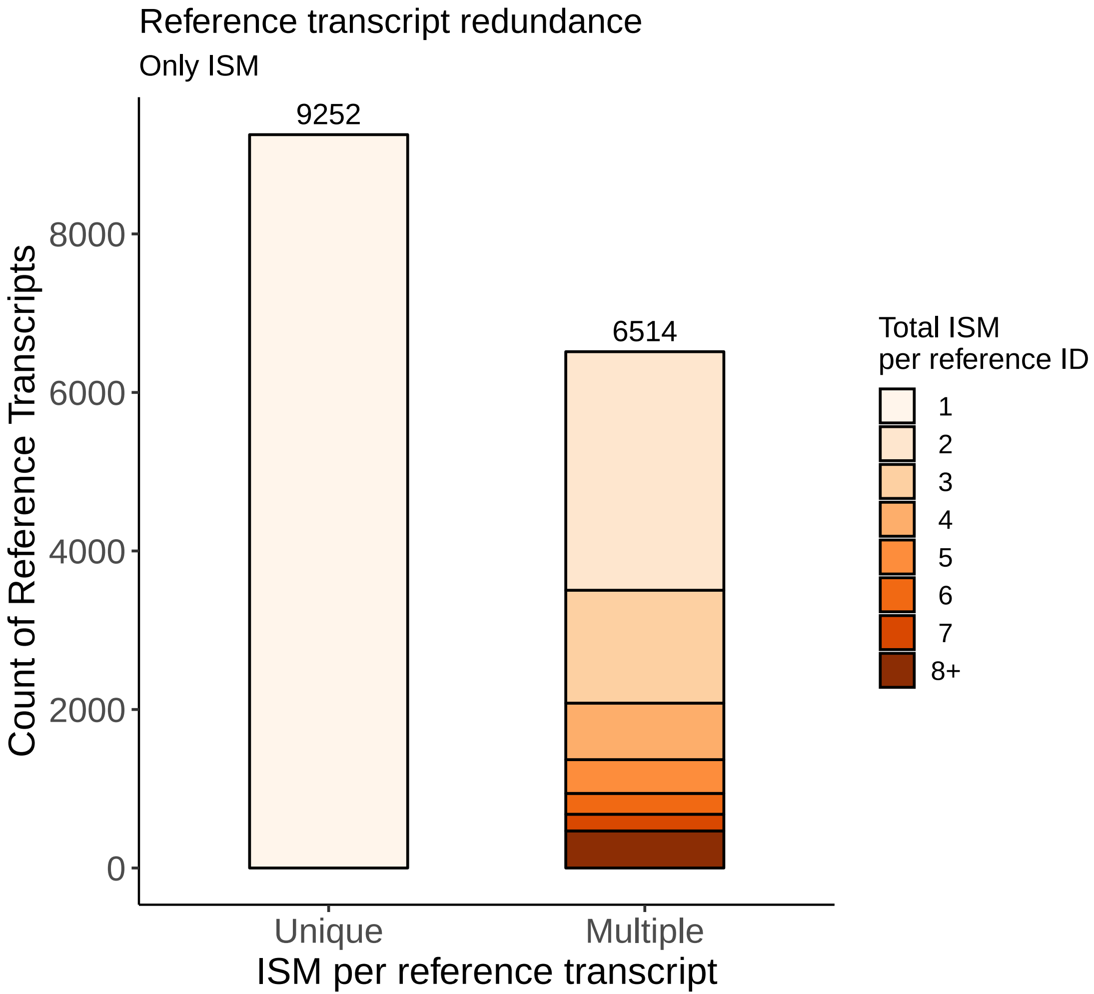


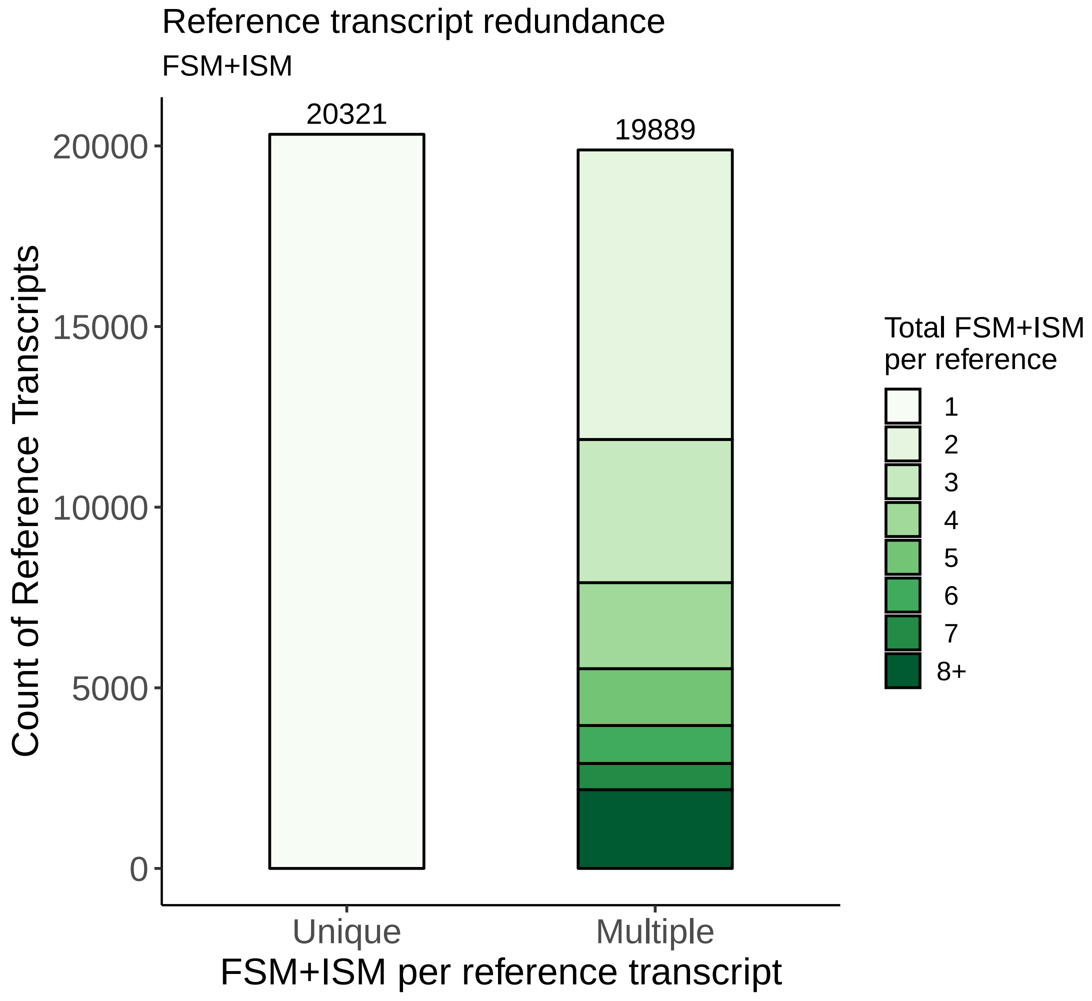
Distance to CAGE peak of multi−exonic NNC

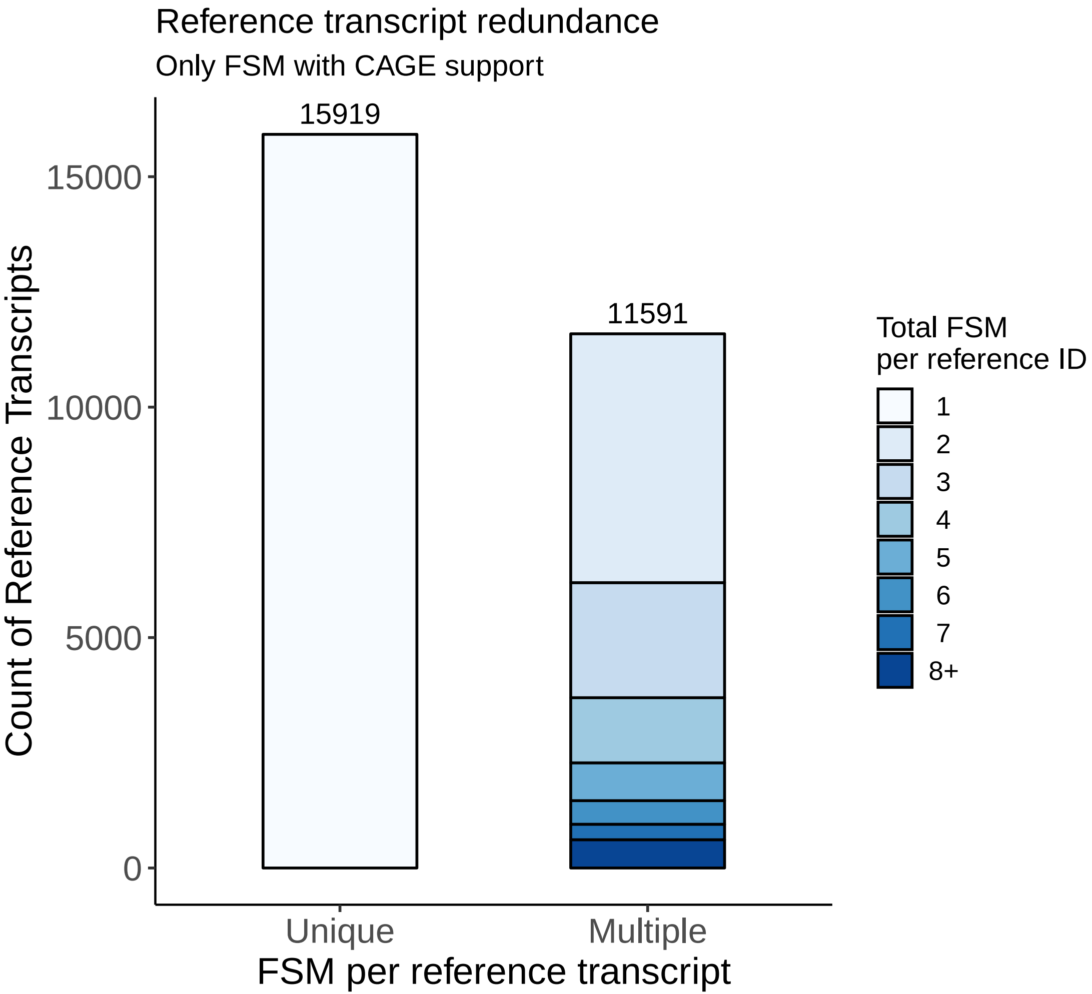
Negative values indicate downstream of annotated CAGE peak

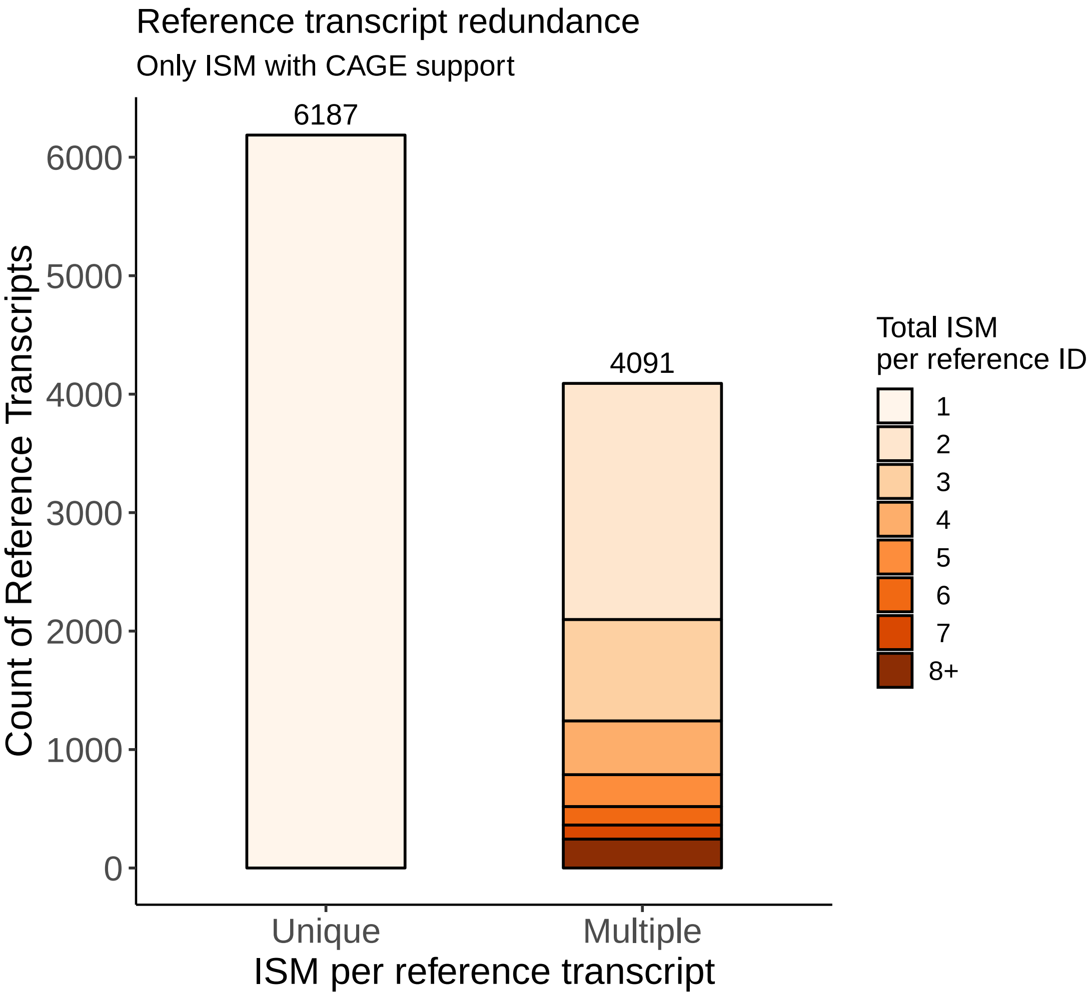
Accumulation of FSM and ISM
to the same reference transcript

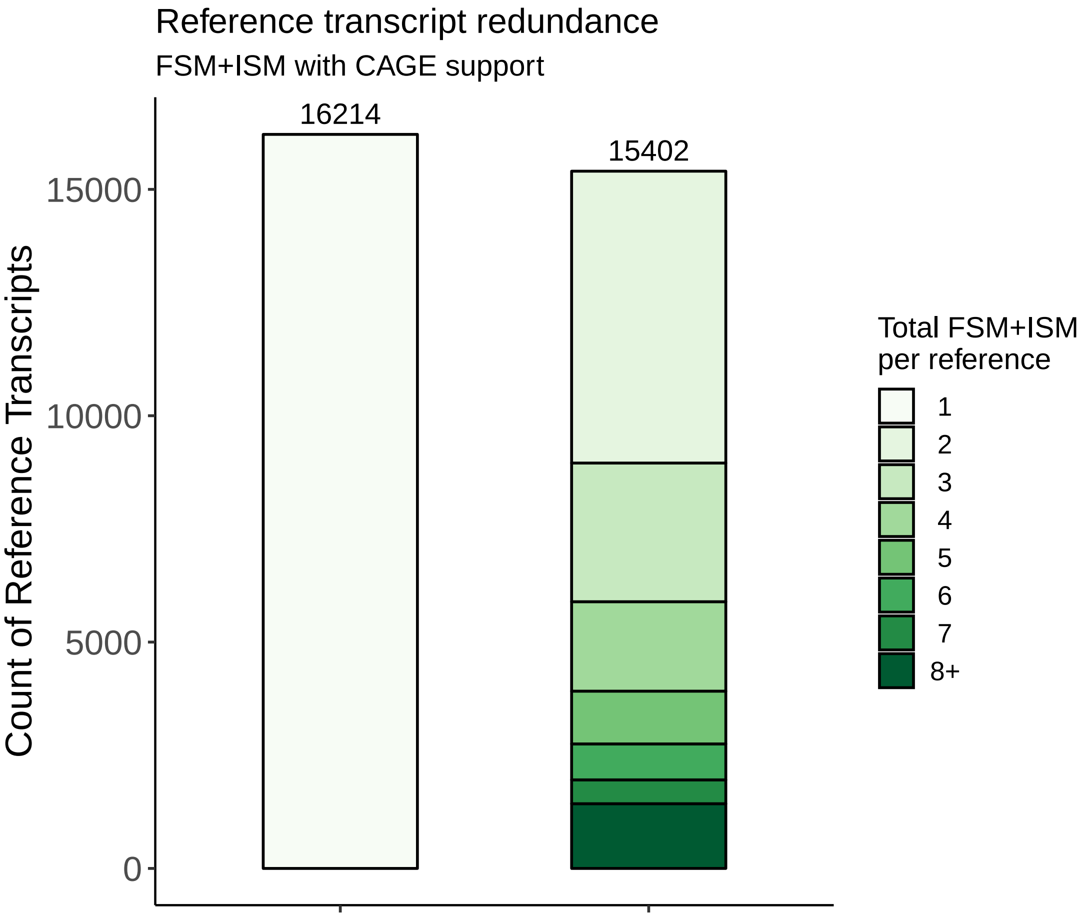

Reference transcript redundance
Only FSM
Count of Reference Transcripts
20000
15000
10000
5000
0
19047
14099
Unique
Multiple
FSM per reference transcript
Total FSM
per reference ID
1
2
3
4
5
6
7
8+

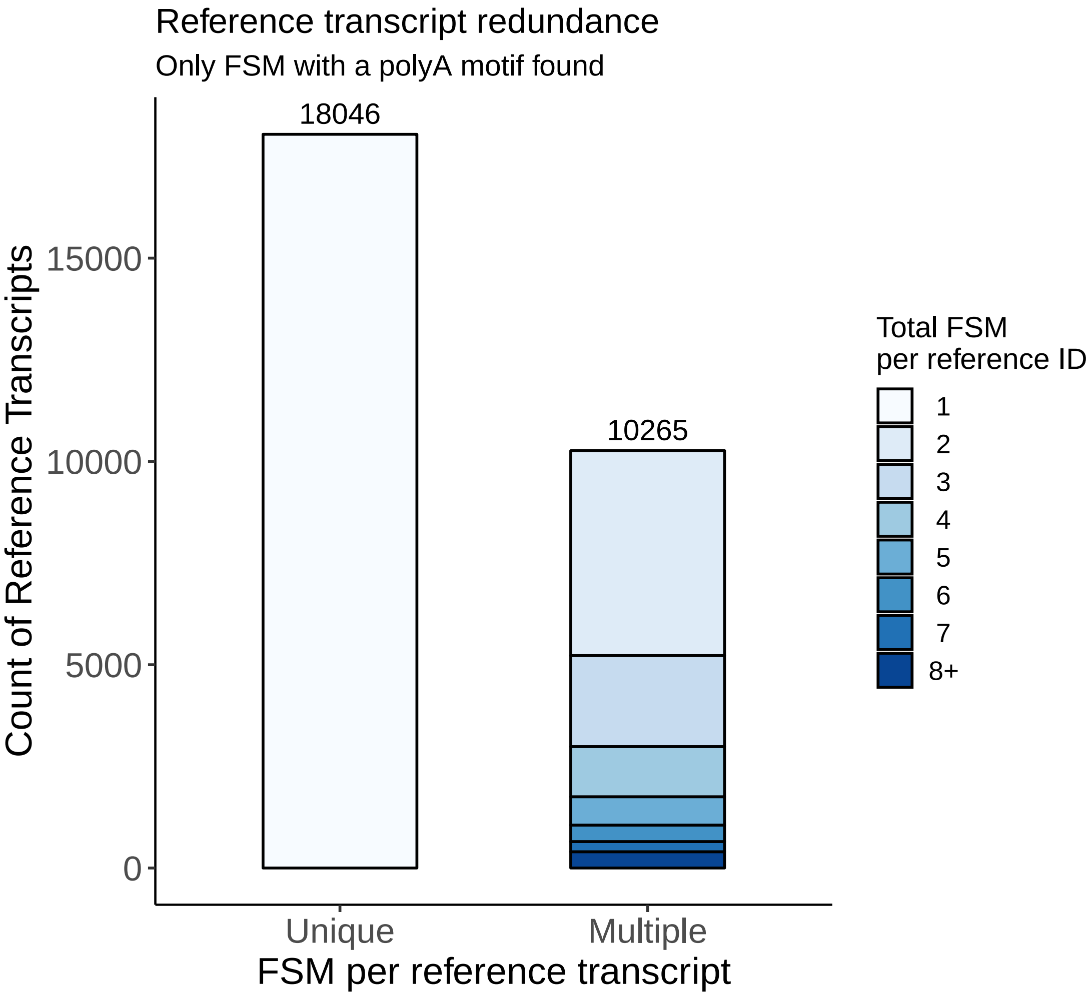

Reference transcript redundance
Only ISM
9252
6514
8000
6000
4000
2000
0
Count of Reference Transcripts
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
5
6
7
8+

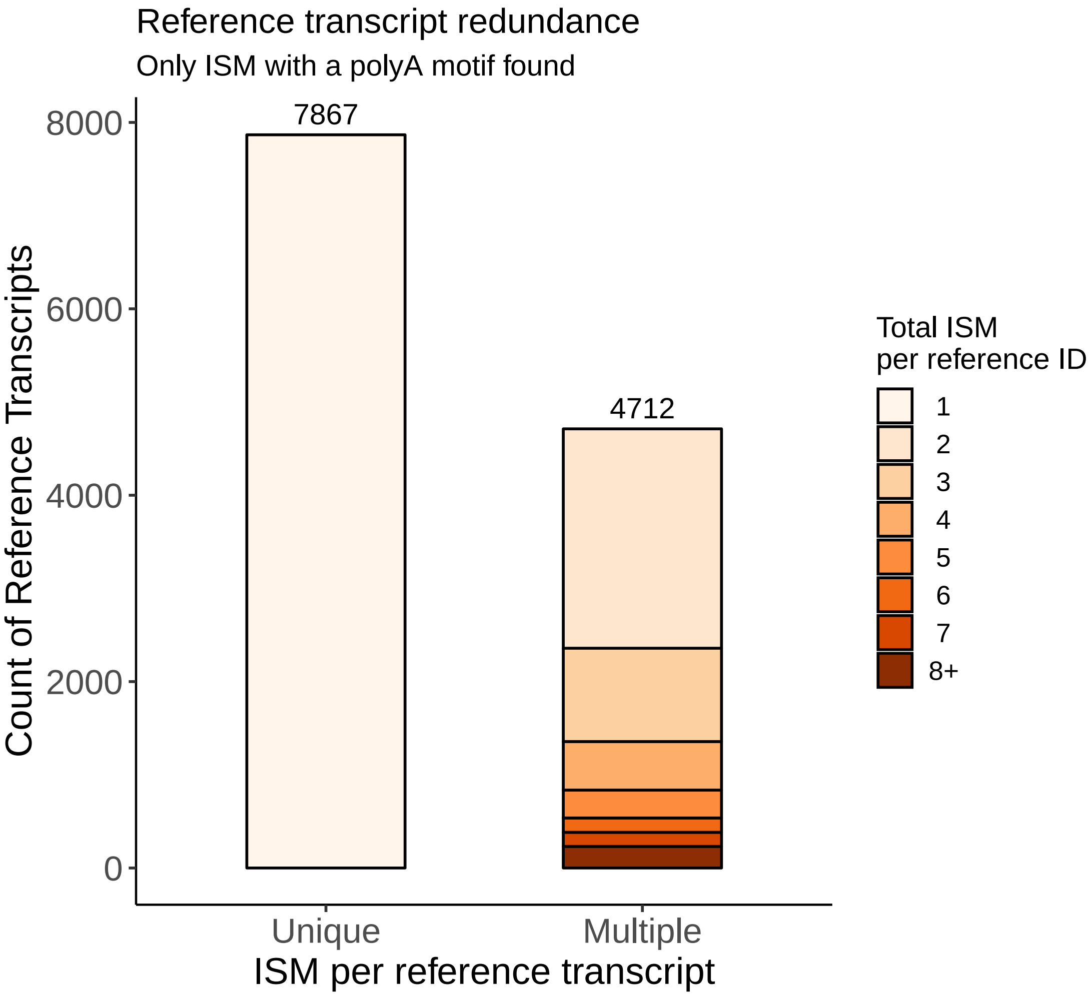

Reference transcript redundance
FSM+ISM
20321
19889
20000
15000
10000
5000
0
Count of Reference Transcripts
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+

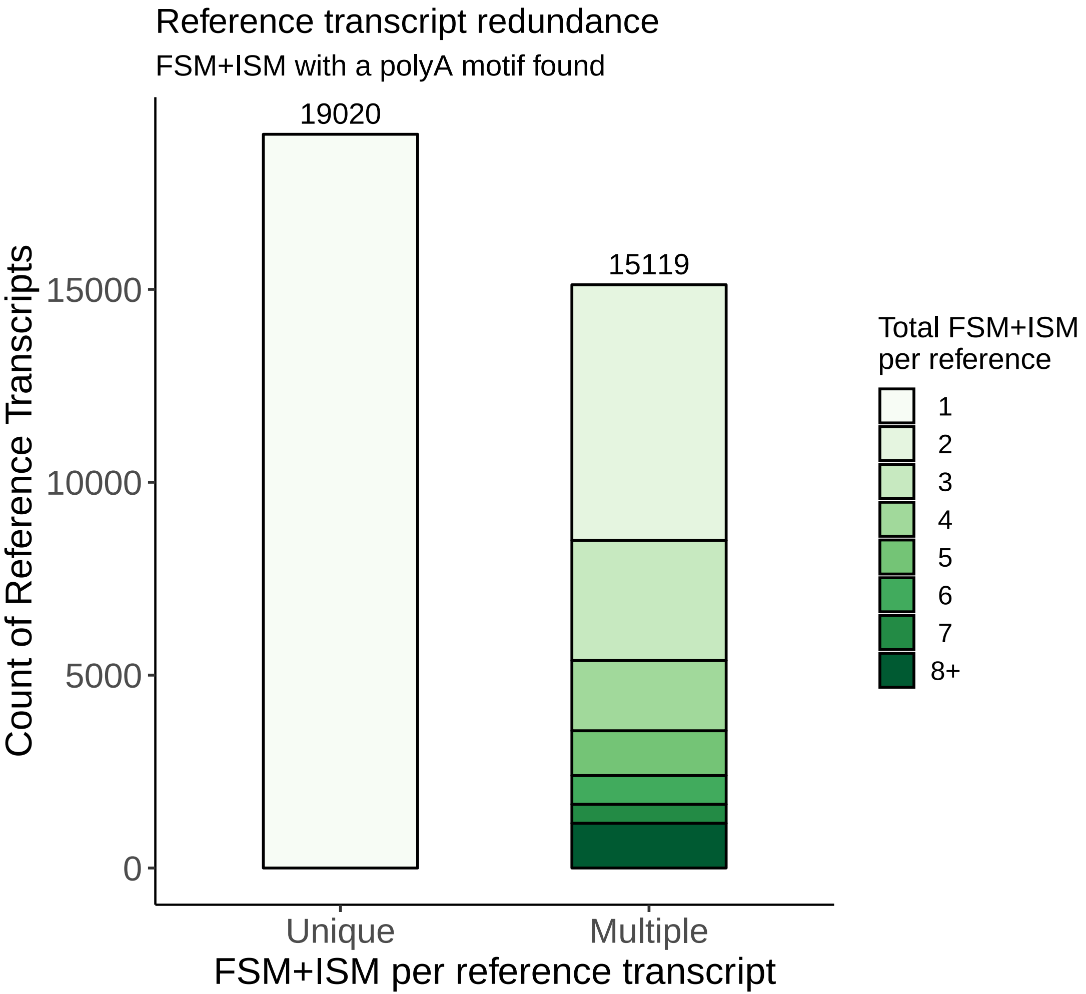

Reference transcript redundance
Only FSM with CAGE support
15919
11591
15000
10000
5000
0
Count of Reference Transcripts
Unique
Multiple
FSM per reference transcript
Total FSM
per reference ID
1
2
3
4
5
6
7
8+

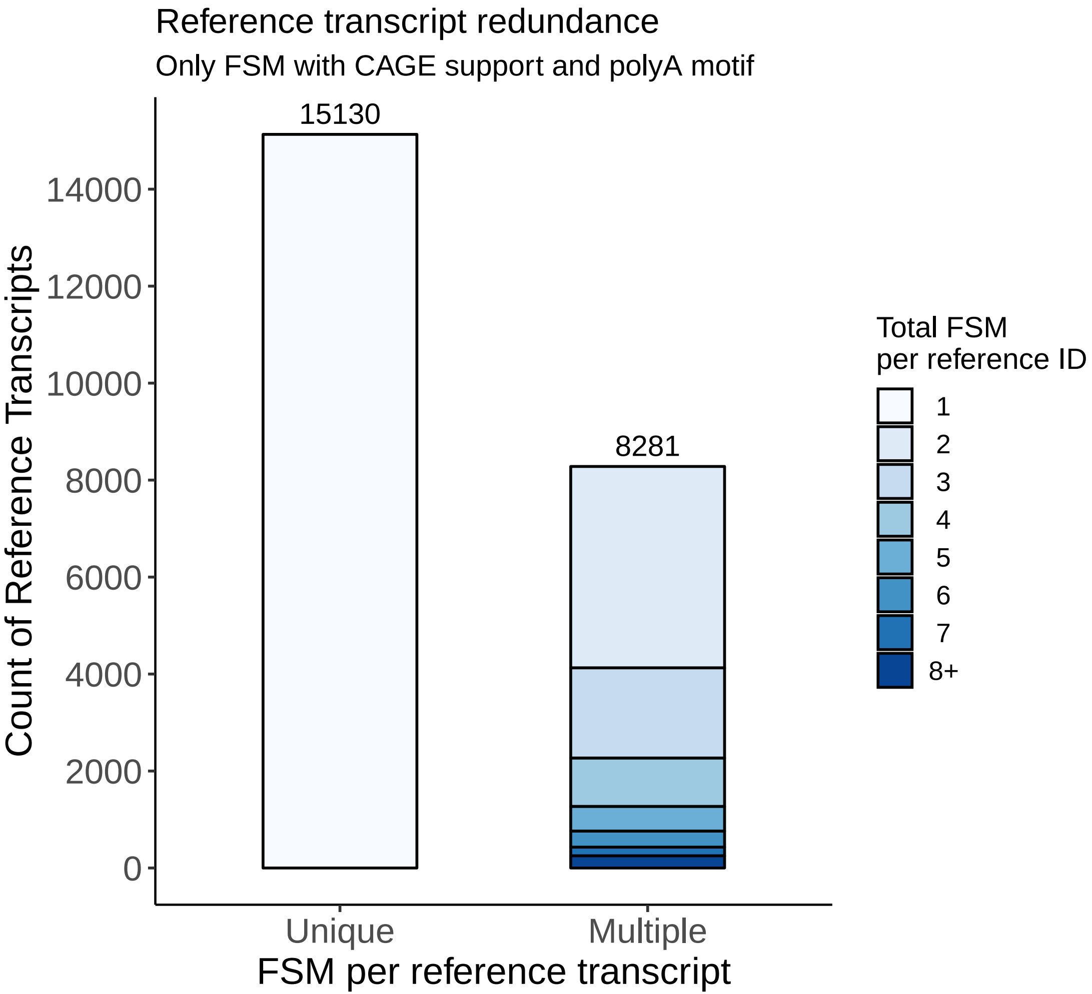

Reference transcript redundance
Only ISM with CAGE support
6187
4091
6000
5000
4000
3000
2000
1000
0
Count of Reference Transcripts
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
5
6
7
8+

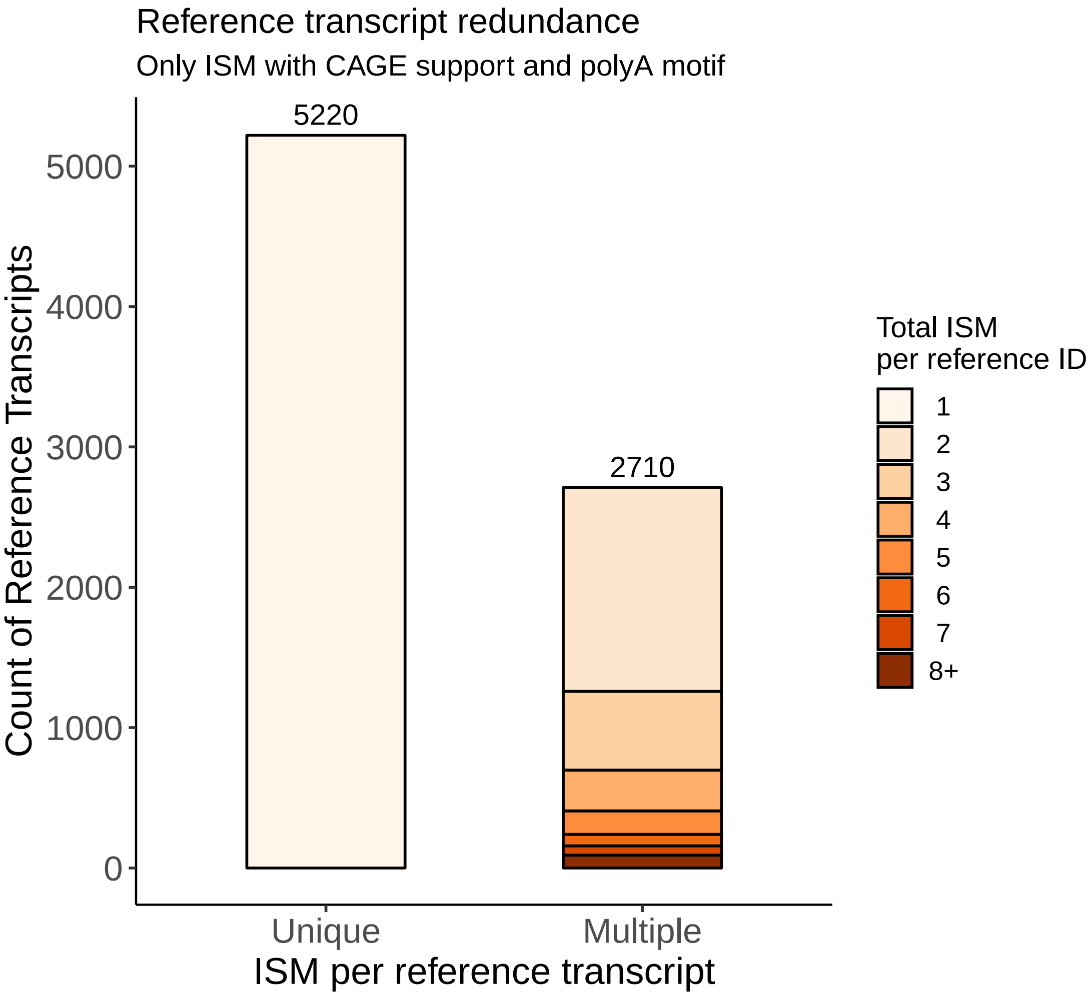

Reference transcript redundance
FSM+ISM with CAGE support
16214
15402
15000
10000
5000
0
Count of Reference Transcripts
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+

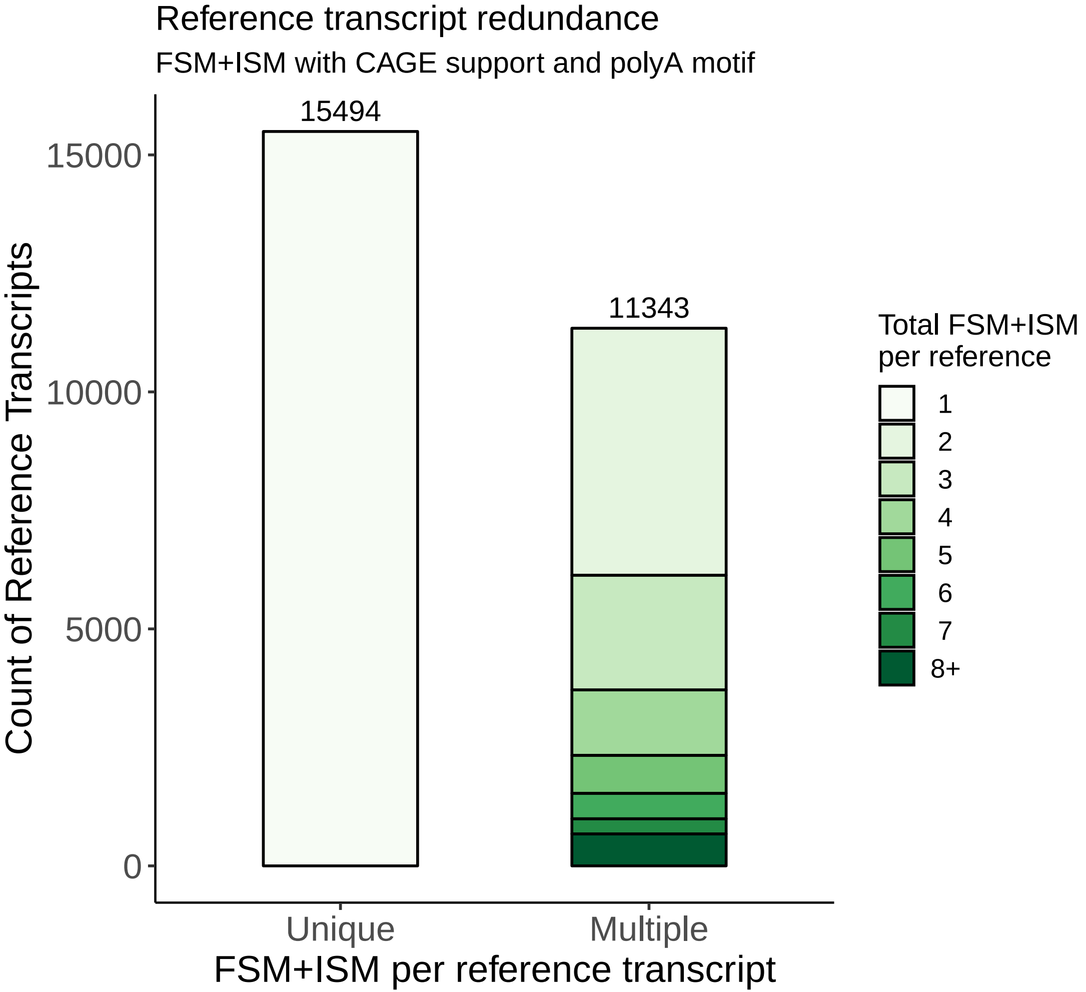

Reference transcript redundance
Only FSM with a polyA motif found
18046
10265
15000
10000
5000
0
Count of Reference Transcripts
Unique
Multiple
FSM per reference transcript
Total FSM
per reference ID
1
2
3
4
5
6
7
8+


Reference transcript redundance
Only ISM with a polyA motif found
Count of Reference Transcripts
8000
6000
4000
2000
0
7867
4712
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
5
6
7
8+

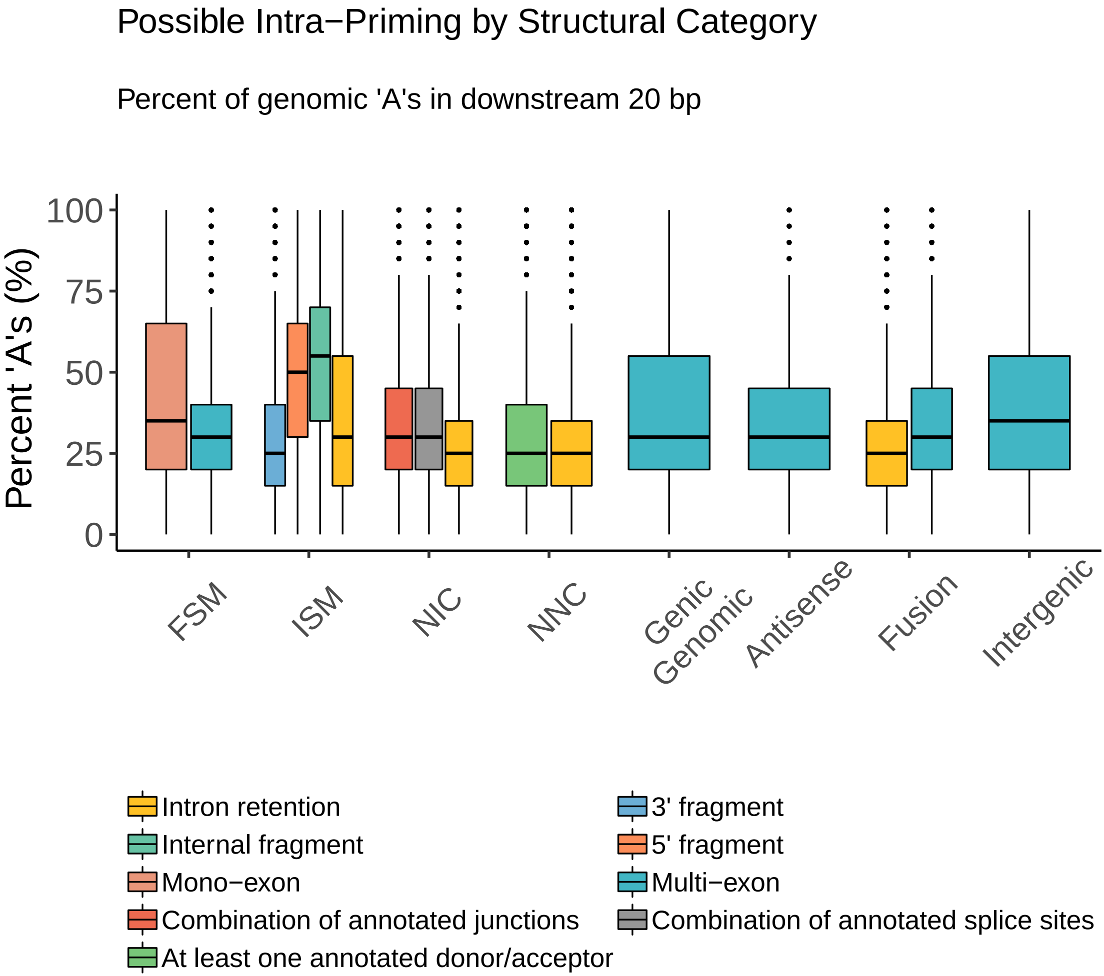

Reference transcript redundance
FSM+ISM with a polyA motif found
19020
15119
Count of Reference Transcripts
15000
10000
5000
0
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+


Reference transcript redundance
Only FSM with CAGE support and polyA motif
15130
8281
Count of Reference Transcripts
14000
12000
10000
8000
6000
4000
2000
0
Unique
Multiple
FSM per reference transcript
Total FSM
per reference ID
1
2
3
4
5
6
7
8+


Reference transcript redundance
Only ISM with CAGE support and polyA motif
5220
2710
5000
4000
3000
2000
1000
0
Count of Reference Transcripts
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
5
6
7
8+

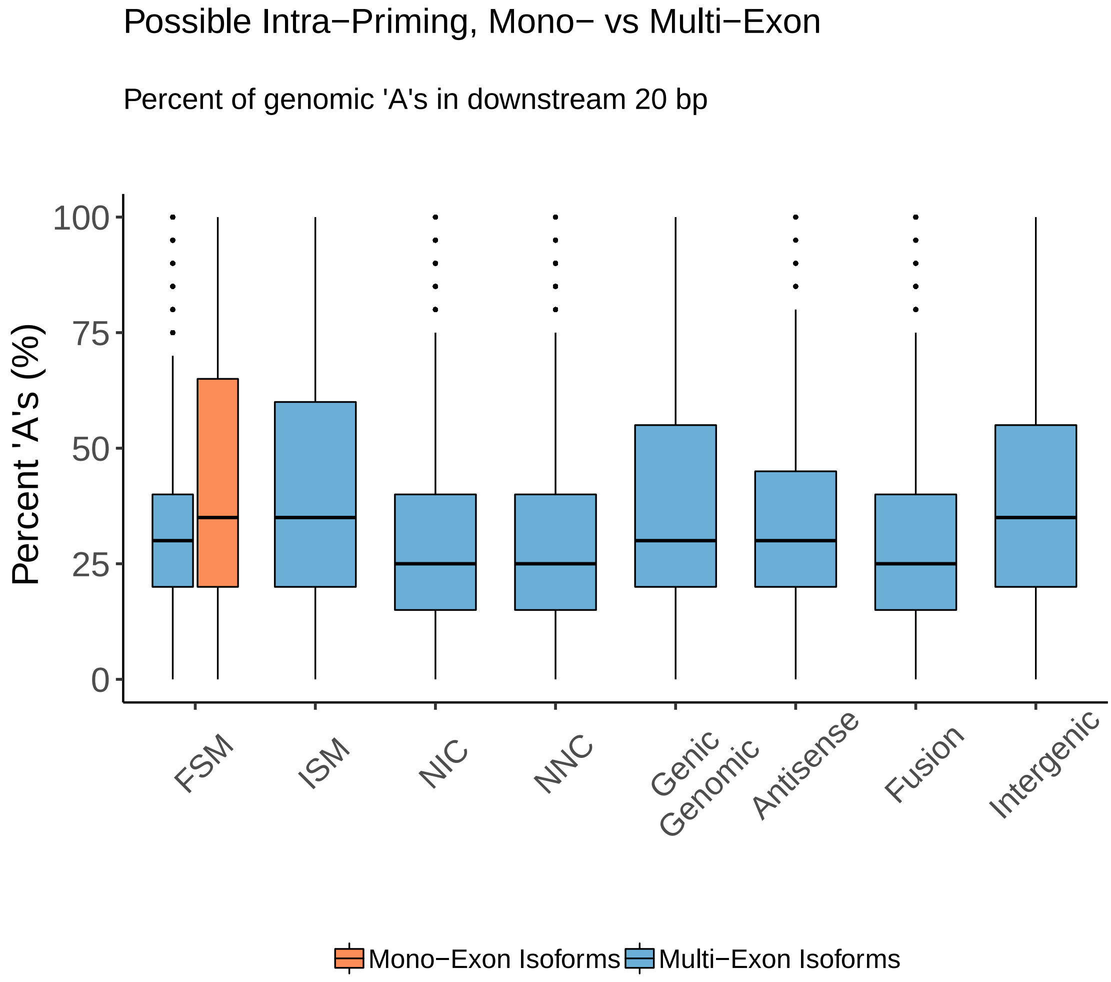

Reference transcript redundance
FSM+ISM with CAGE support and polyA motif
Count of Reference Transcripts
15000
10000
5000
0
15494
11343
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM
per reference
1
2
3
4
5
6
7
8+

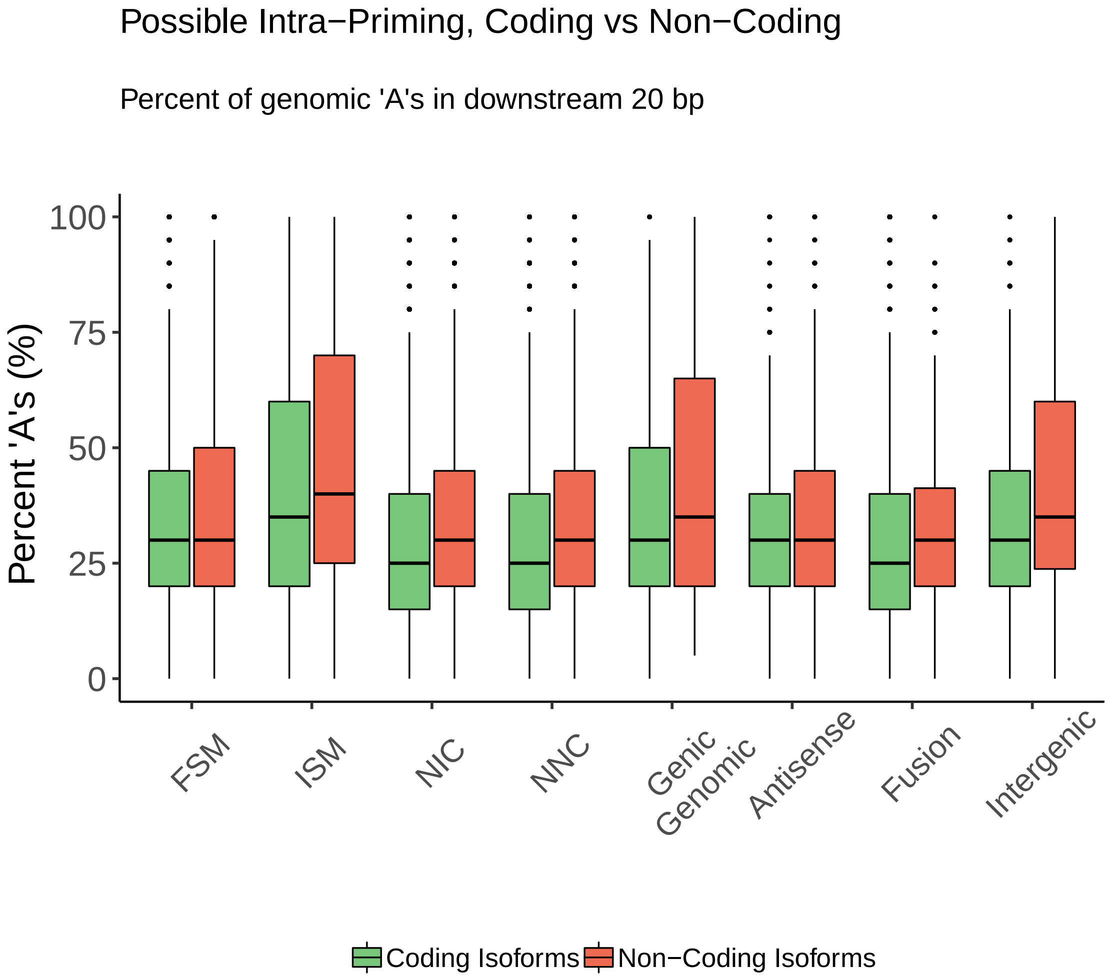
Intra−Priming Quality Check



Possible Intra−Priming by Structural Category

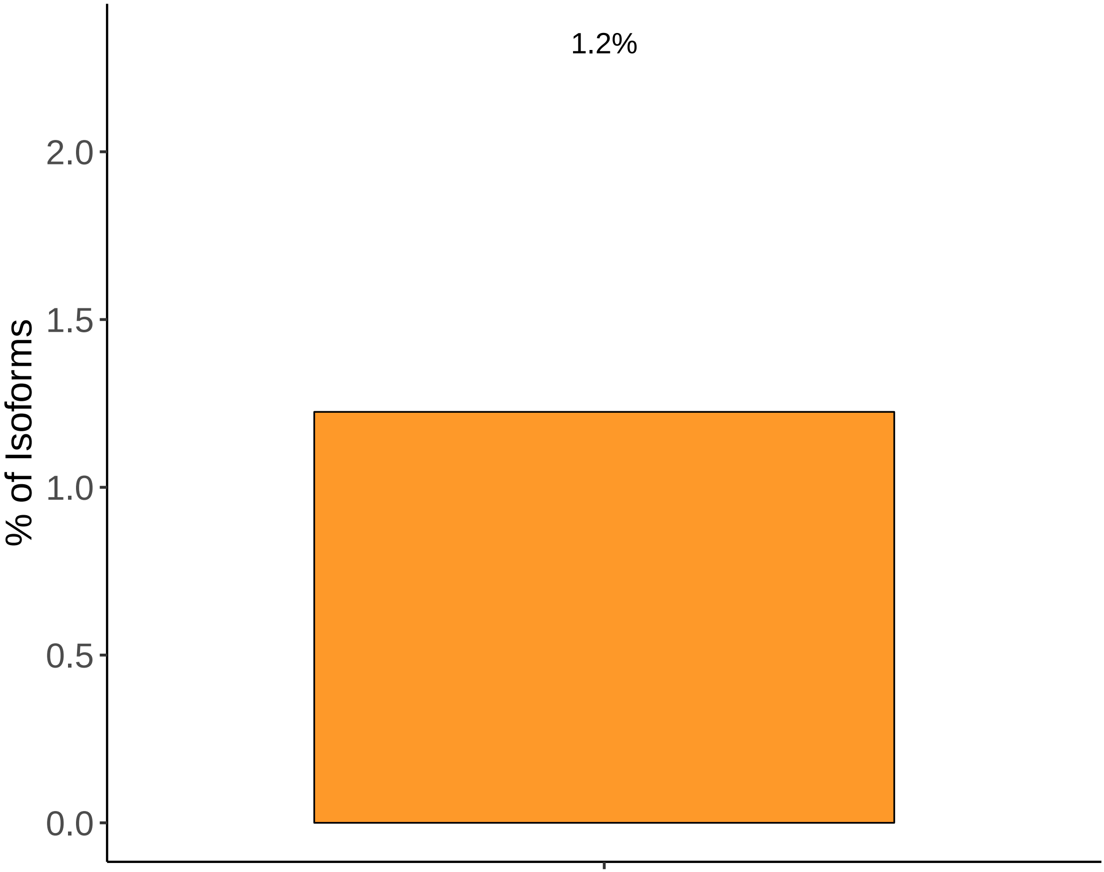
Percent of genomic 'A's in downstream 20 bp

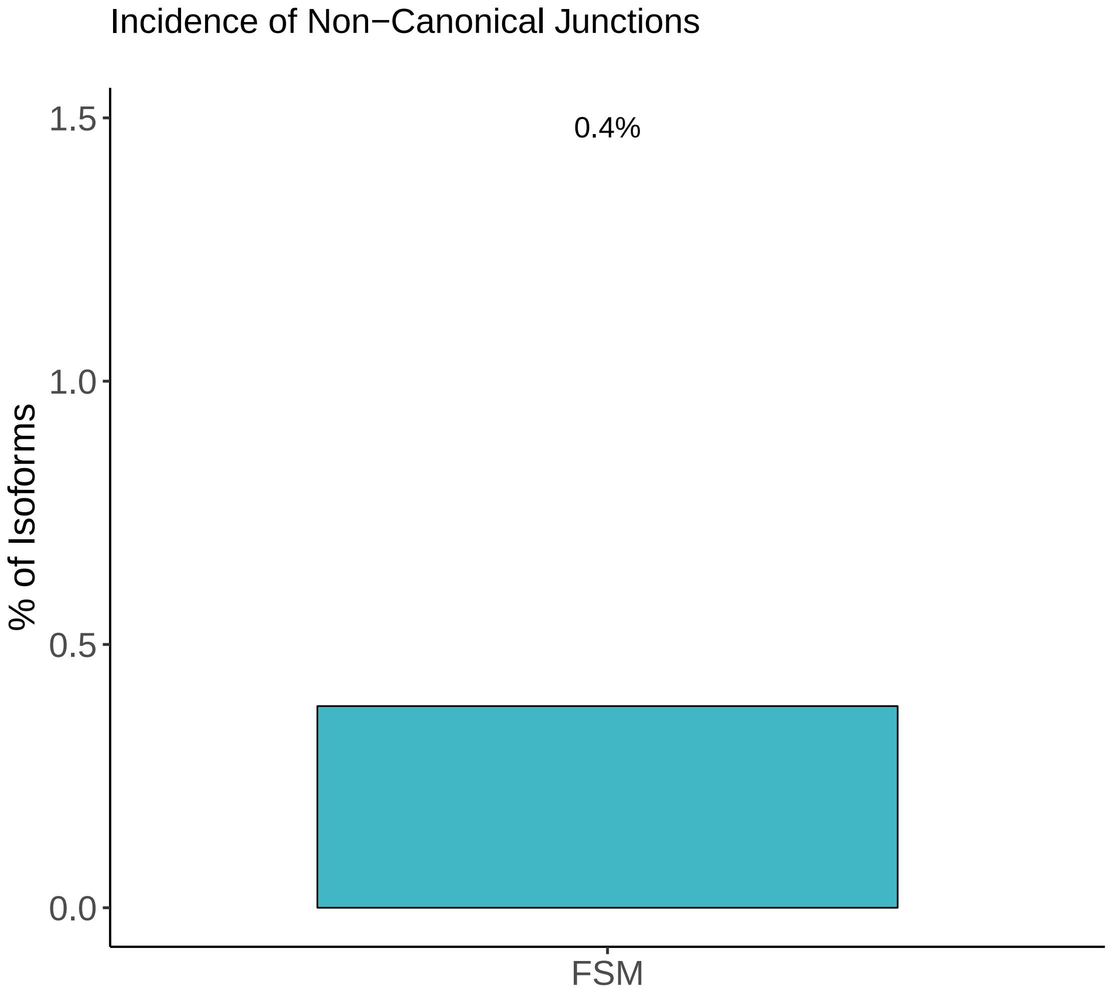

Possible Intra−Priming, Mono− vs Multi−Exon
Percent of genomic 'A's in downstream 20 bp
Percent 'A's (%)
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Mono−Exon Isoforms
Multi−Exon Isoforms

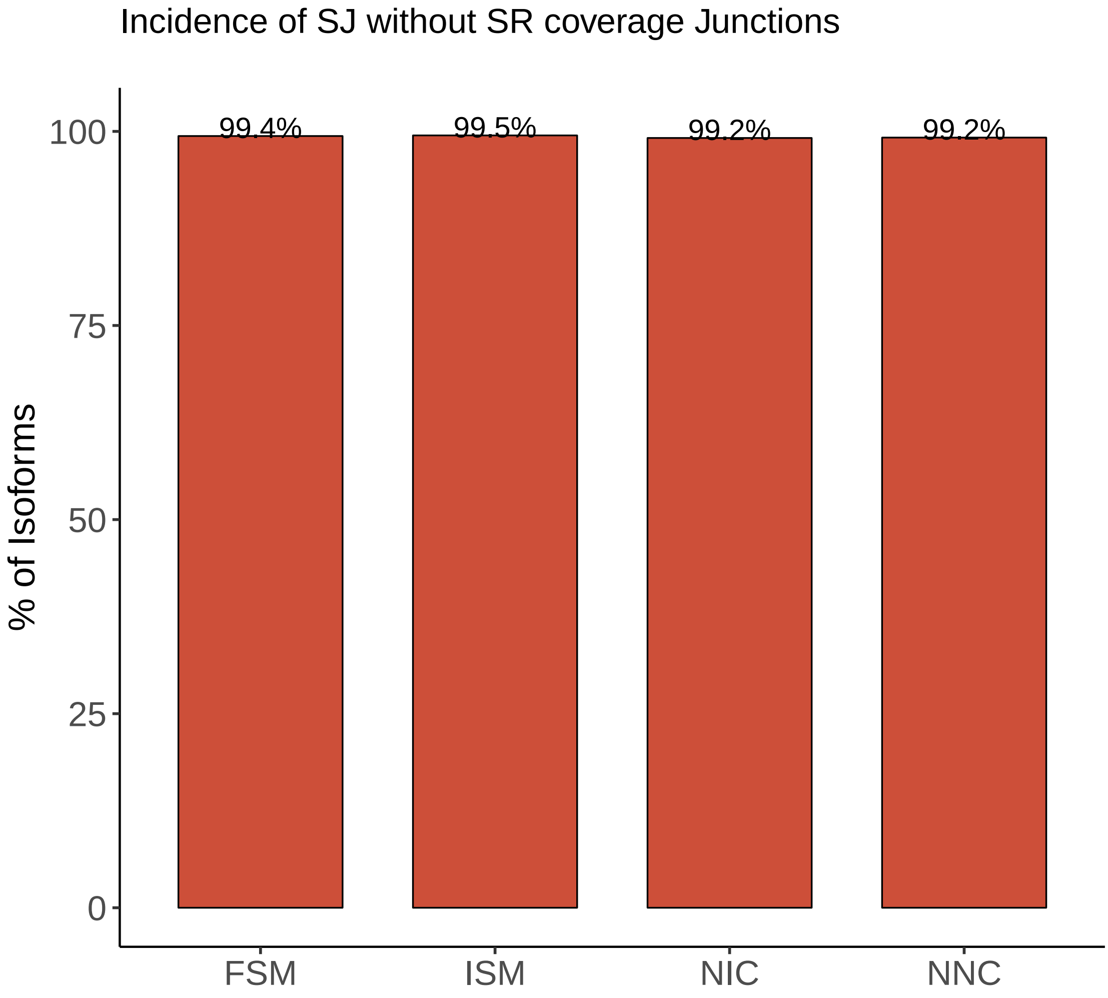

Possible Intra−Priming, Coding vs Non−Coding
Percent of genomic 'A's in downstream 20 bp
Percent 'A's (%)
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Coding Isoforms
Non−Coding Isoforms

Quality Controls

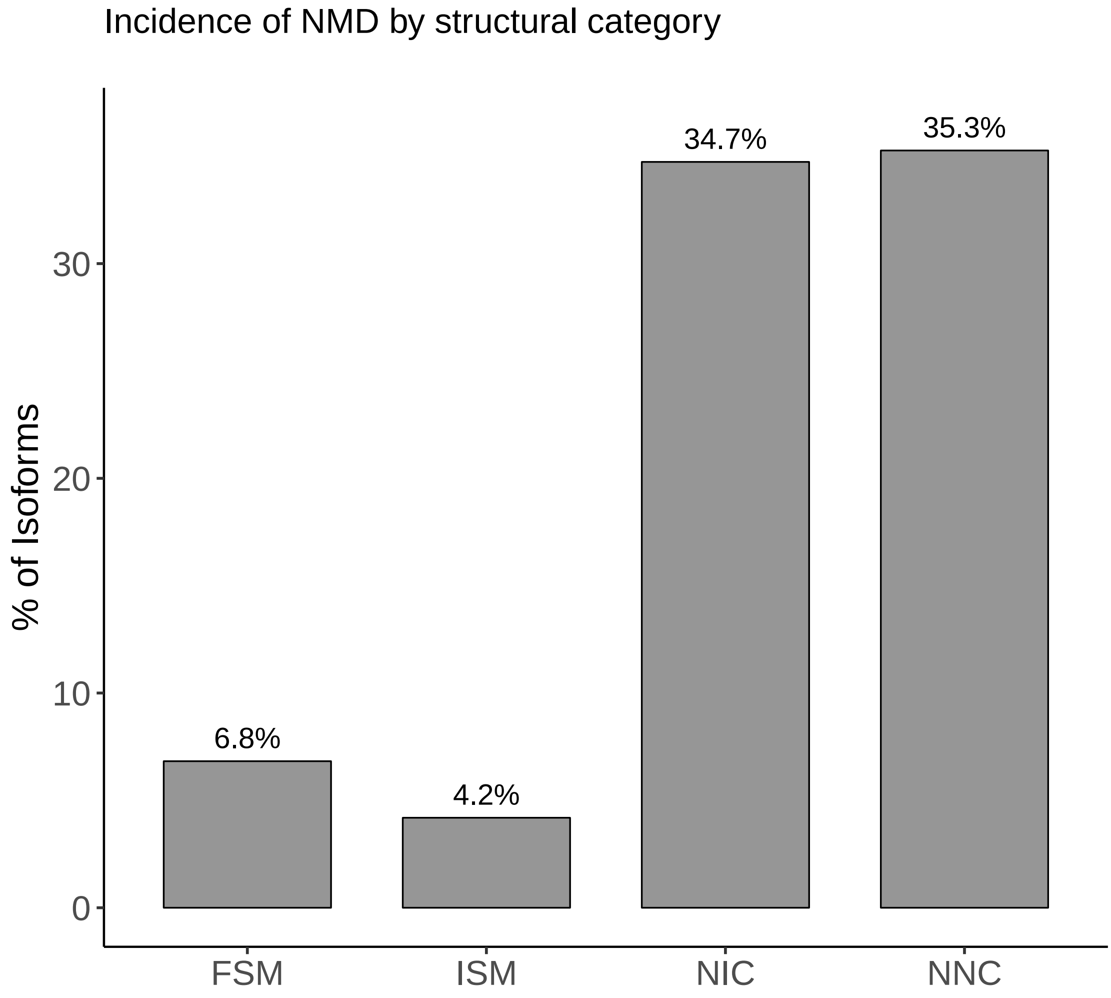
Incidence of RT−switching

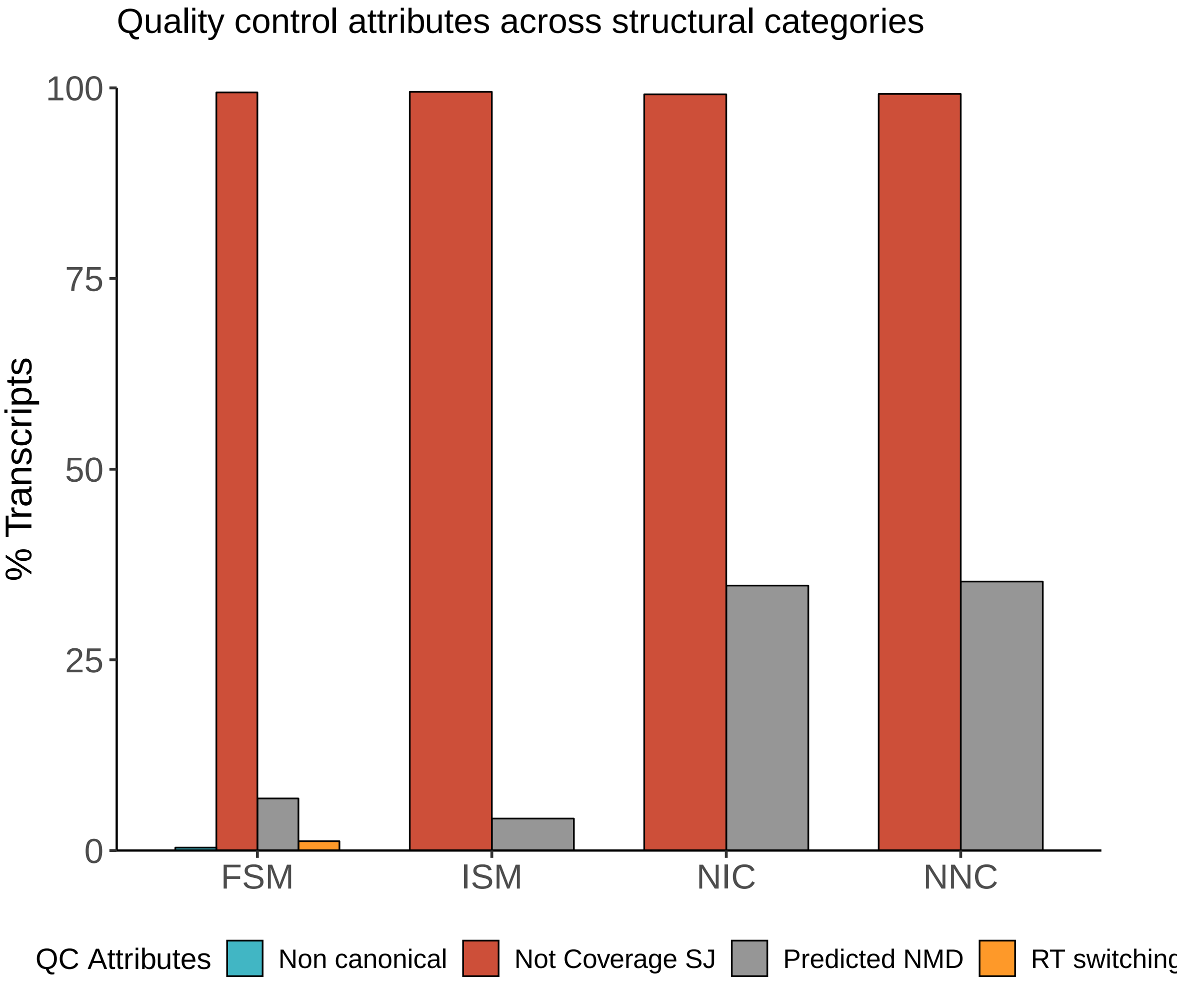



Incidence of Non−Canonical Junctions
% of Isoforms
1.5
1.0
0.5
0.0
0.4%
FSM



Incidence of SJ without SR coverage Junctions


Incidence of NMD by structural category
% of Isoforms
0
10
20
30
FSM
ISM
NIC
NNC
6.8%
4.2%
34.7%
35.3%


Quality control attributes across structural categories
% Transcripts
100
75
50
25
0
FSM
ISM
NIC
NNC
QC Attributes
Non canonical
Not Coverage SJ
Predicted NMD
RT switching