SQANTI3 report

Unique Genes: 248

Unique Isoforms: 1609

Gene classification

Category	# Genes
Annotated Genes	224
Novel Genes	24

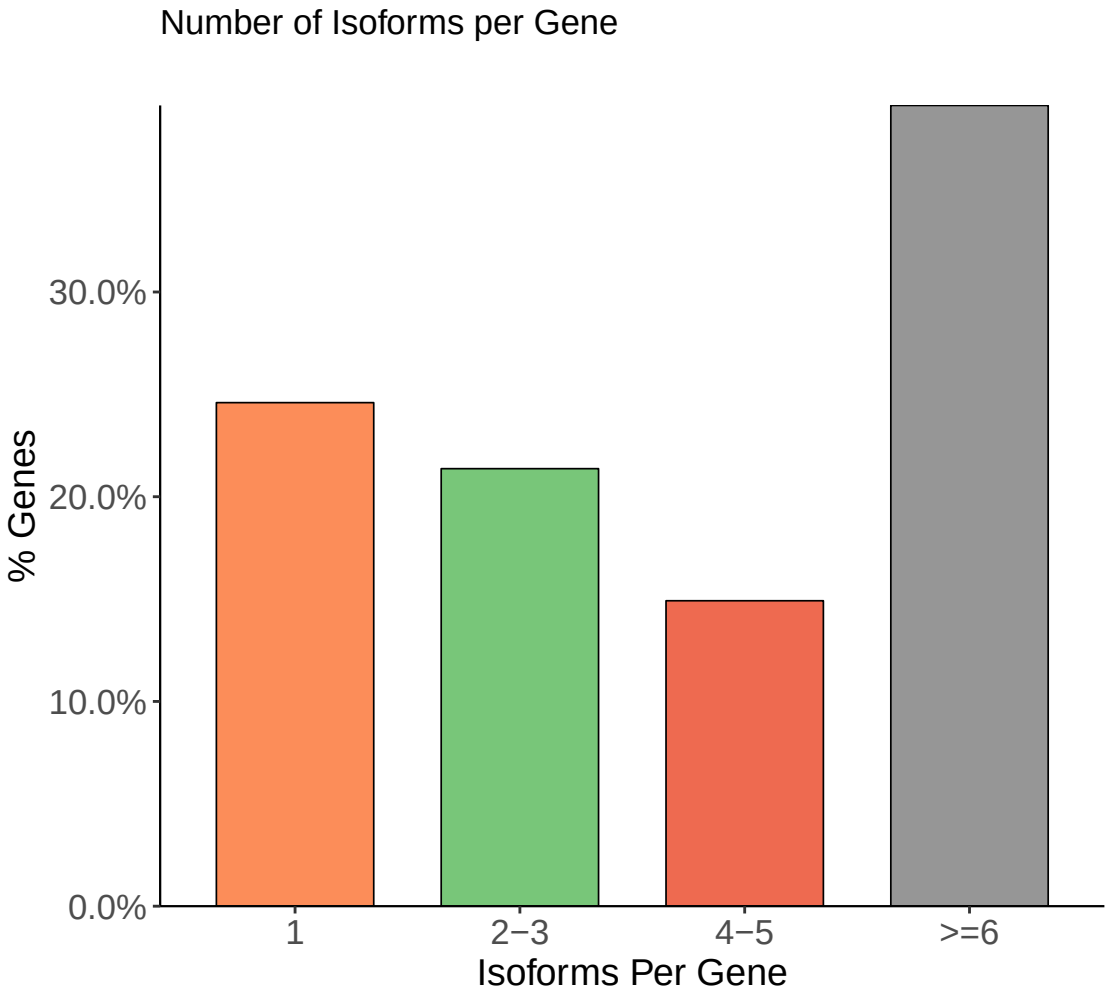
Splice Junction Classification

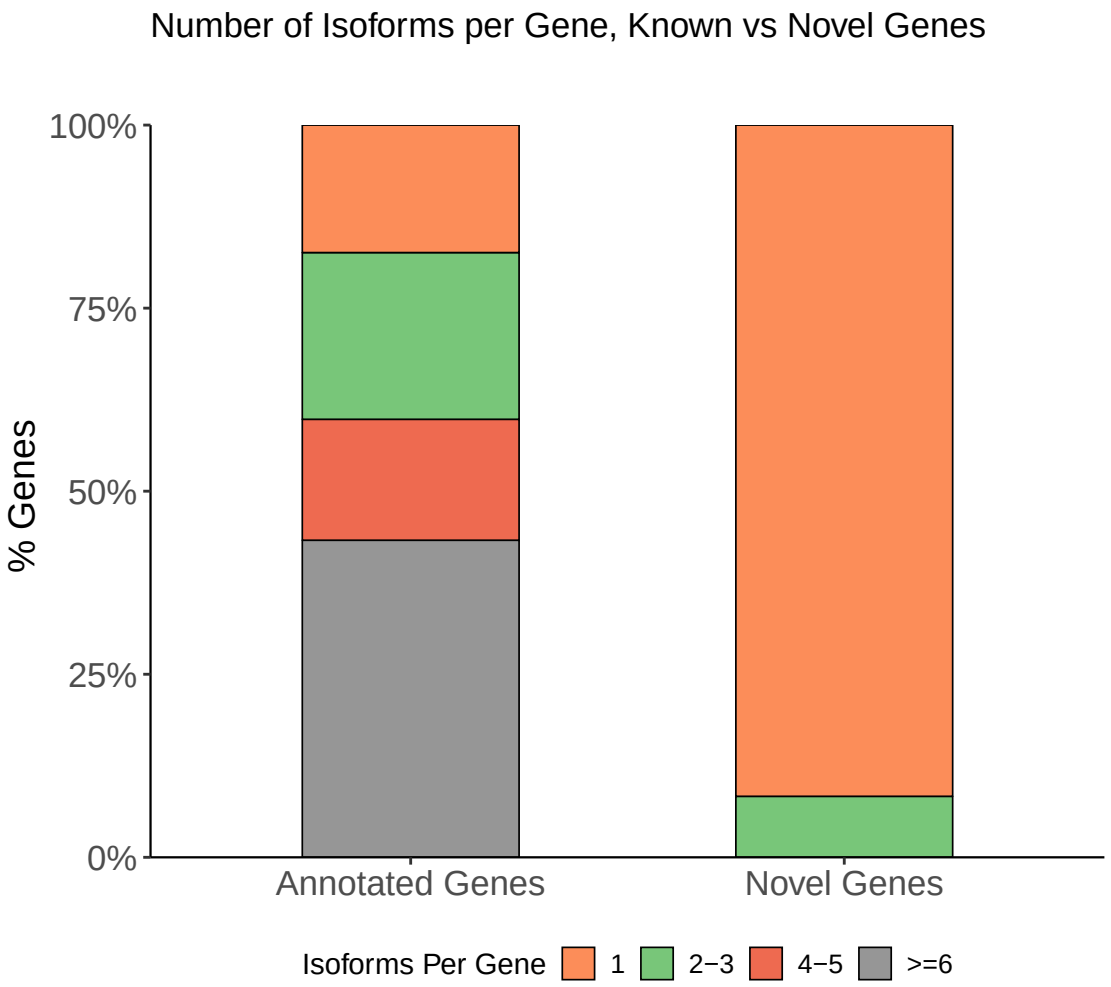
Category	# SJs	Percent
Known canonical	2465	82.86
Known Non−canonical	3	0.10
Novel canonical	443	14.89
Novel Non−canonical	64	2.15

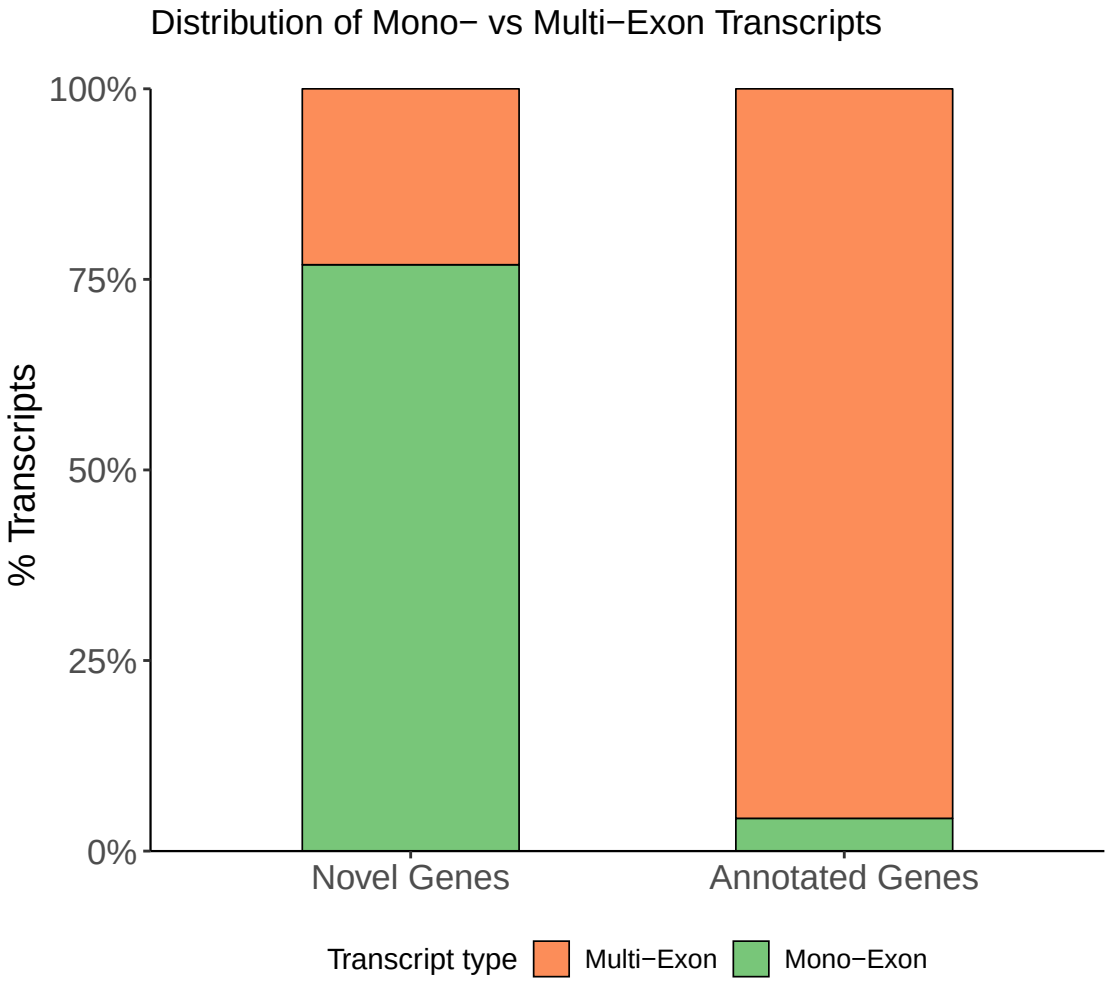
Characterization of transcripts based on splice junctions

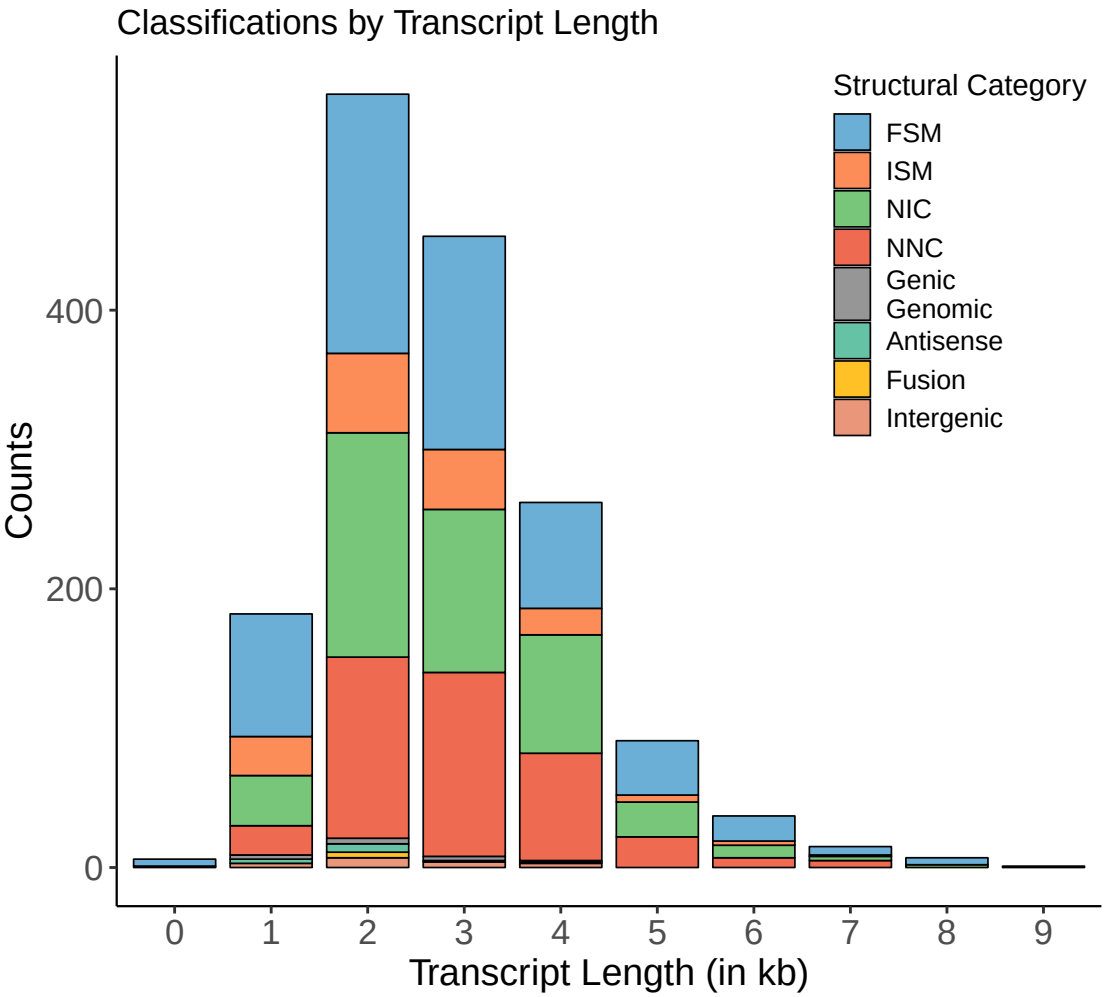
Category	# Isoforms
FSM	576
ISM	156
NIC	439
NNC	395
Genic Genomic	11
Antisense	9
Fusion	6
Intergenic	17
Genic Intron	0

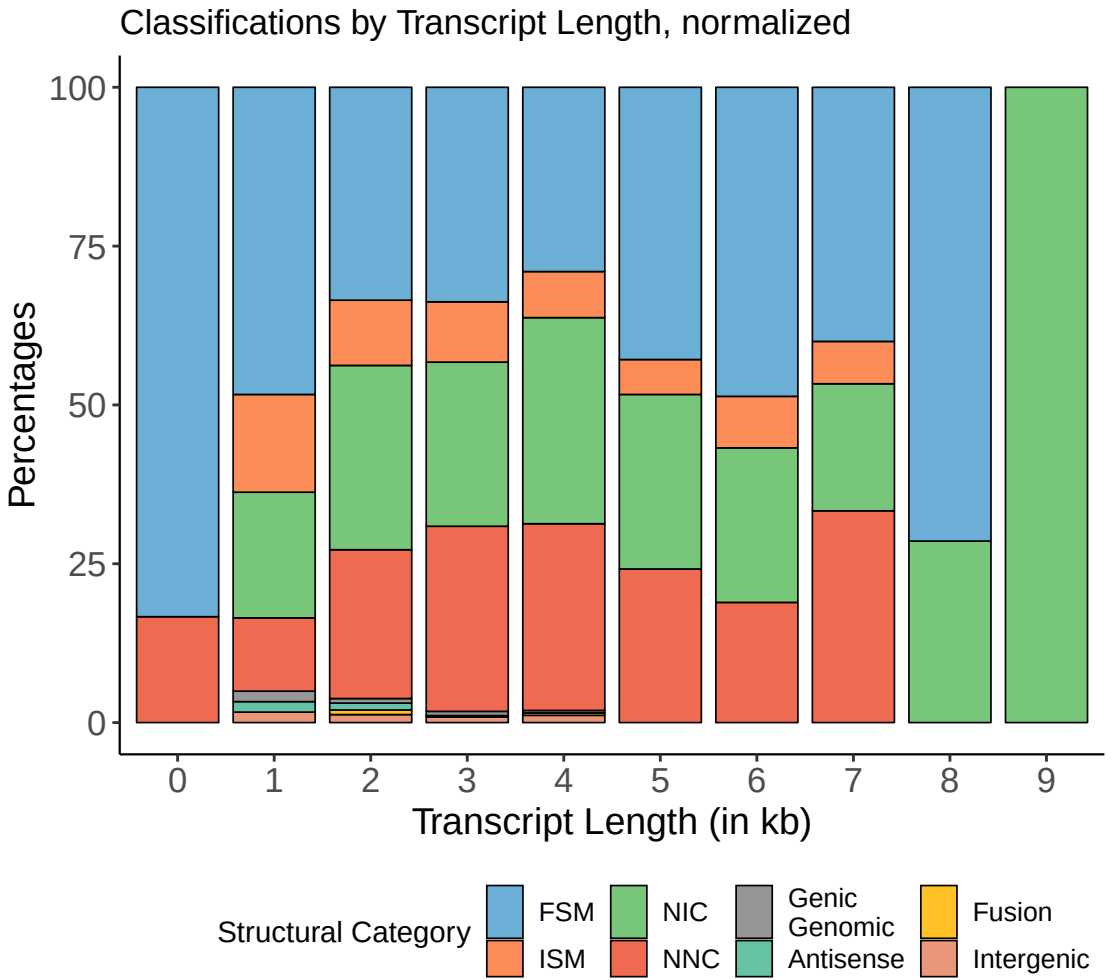
Gene Characterization

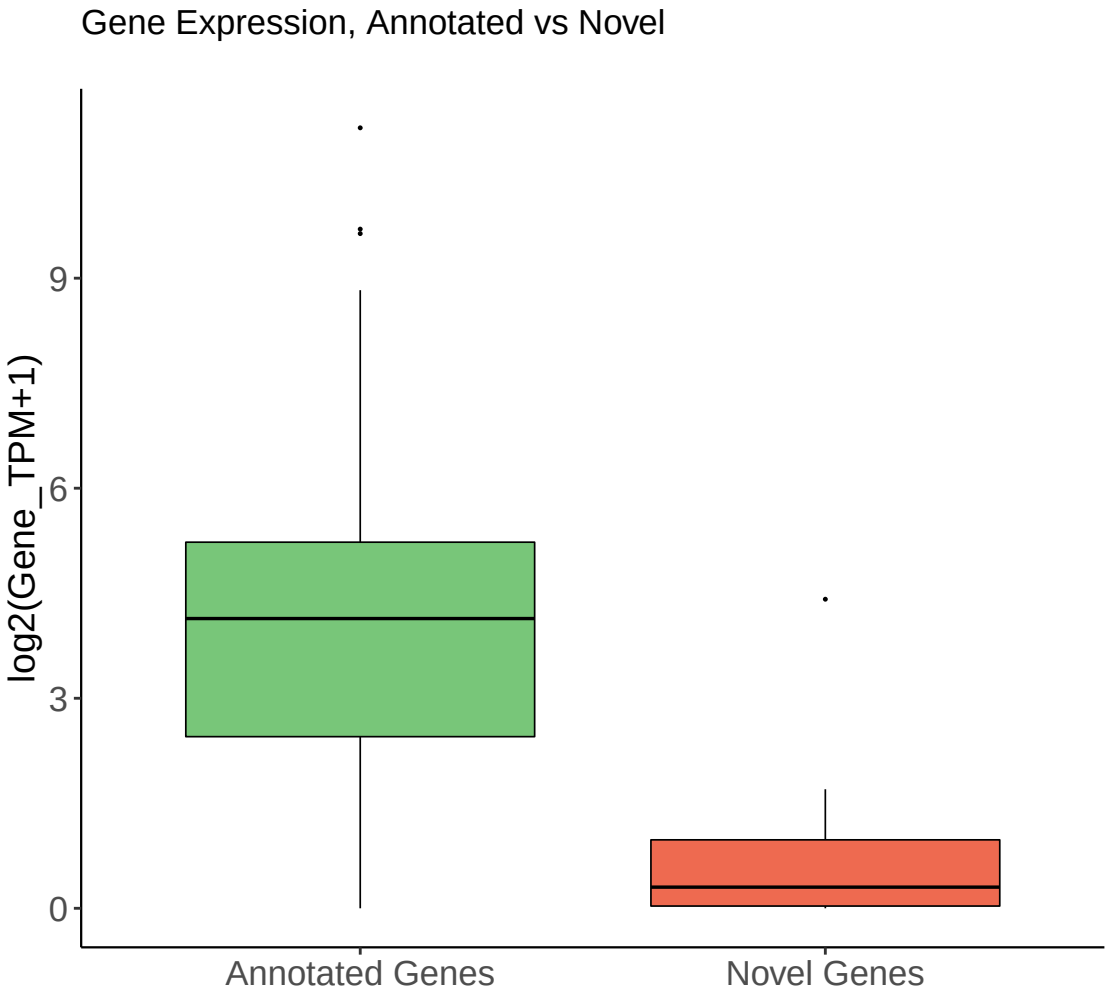

Number of Isoforms per Gene
% Genes
30.0%
20.0%
10.0%
0.0%
1
2−3
4−5
>=6
Isoforms Per Gene

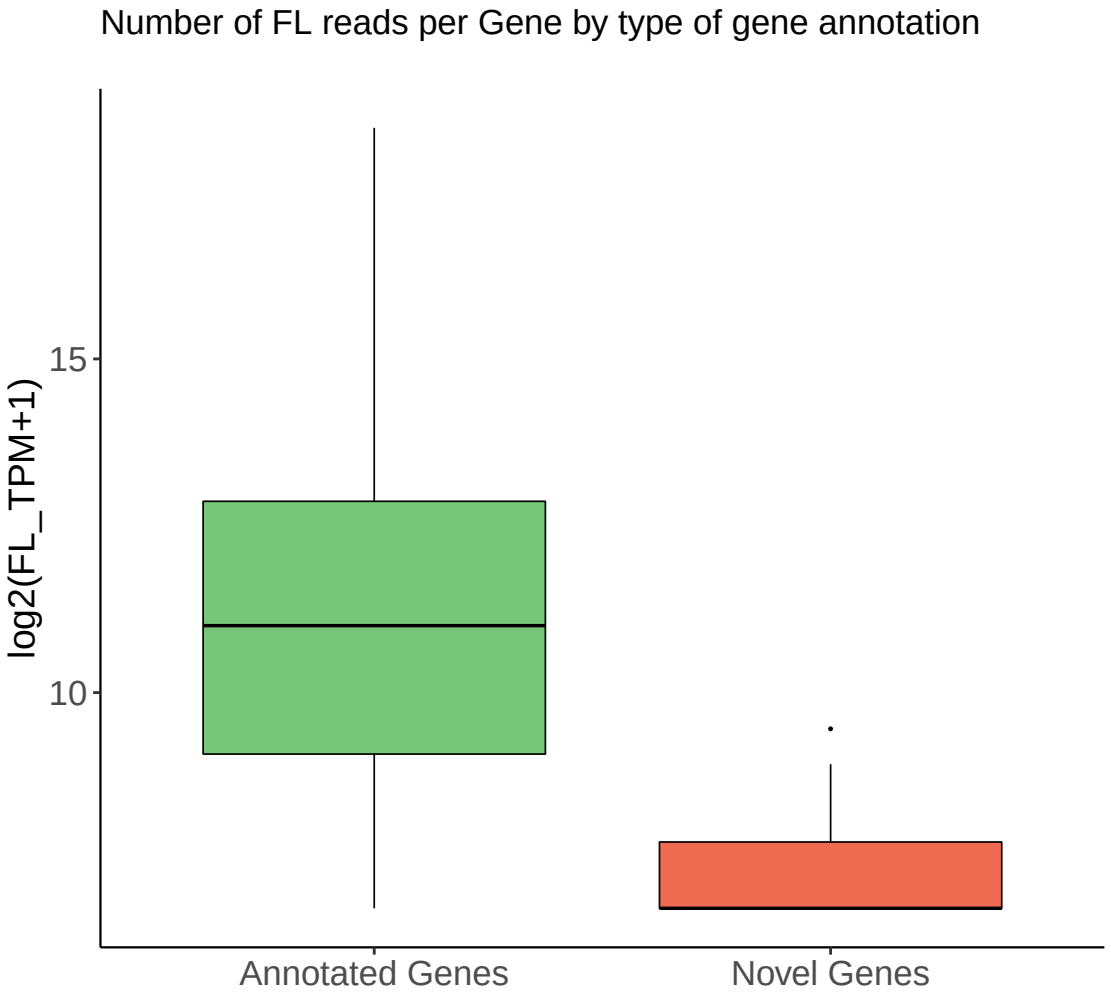

Number of Isoforms per Gene, Known vs Novel Genes
100%
75%
50%
25%
0%
% Genes
Annotated Genes
Novel Genes
Isoforms Per Gene
1
2−3
4−5
>=6

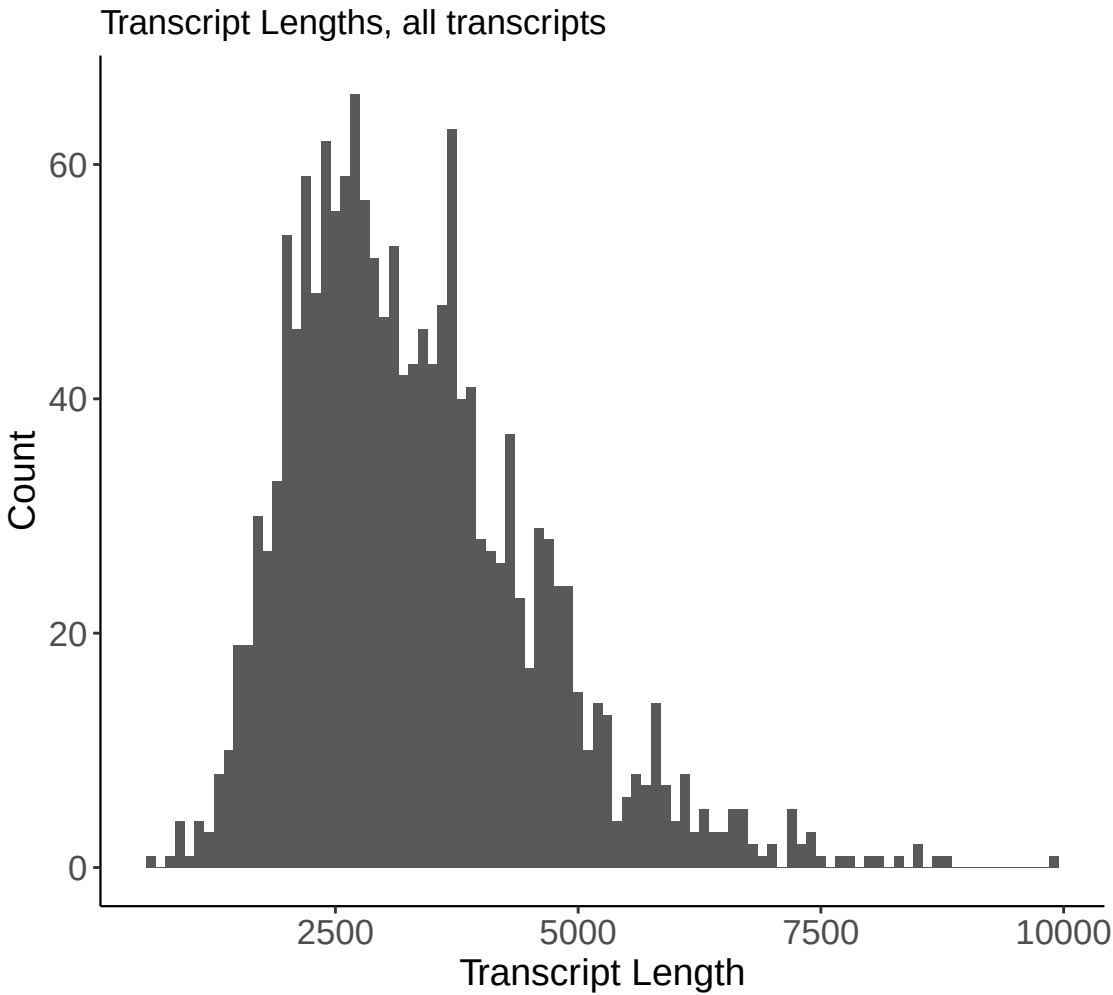

Distribution of Mono− vs Multi−Exon Transcripts
100%
75%
50%
25%
0%
% Transcripts
Novel Genes
Annotated Genes
Transcript type
Multi−Exon
Mono−Exon

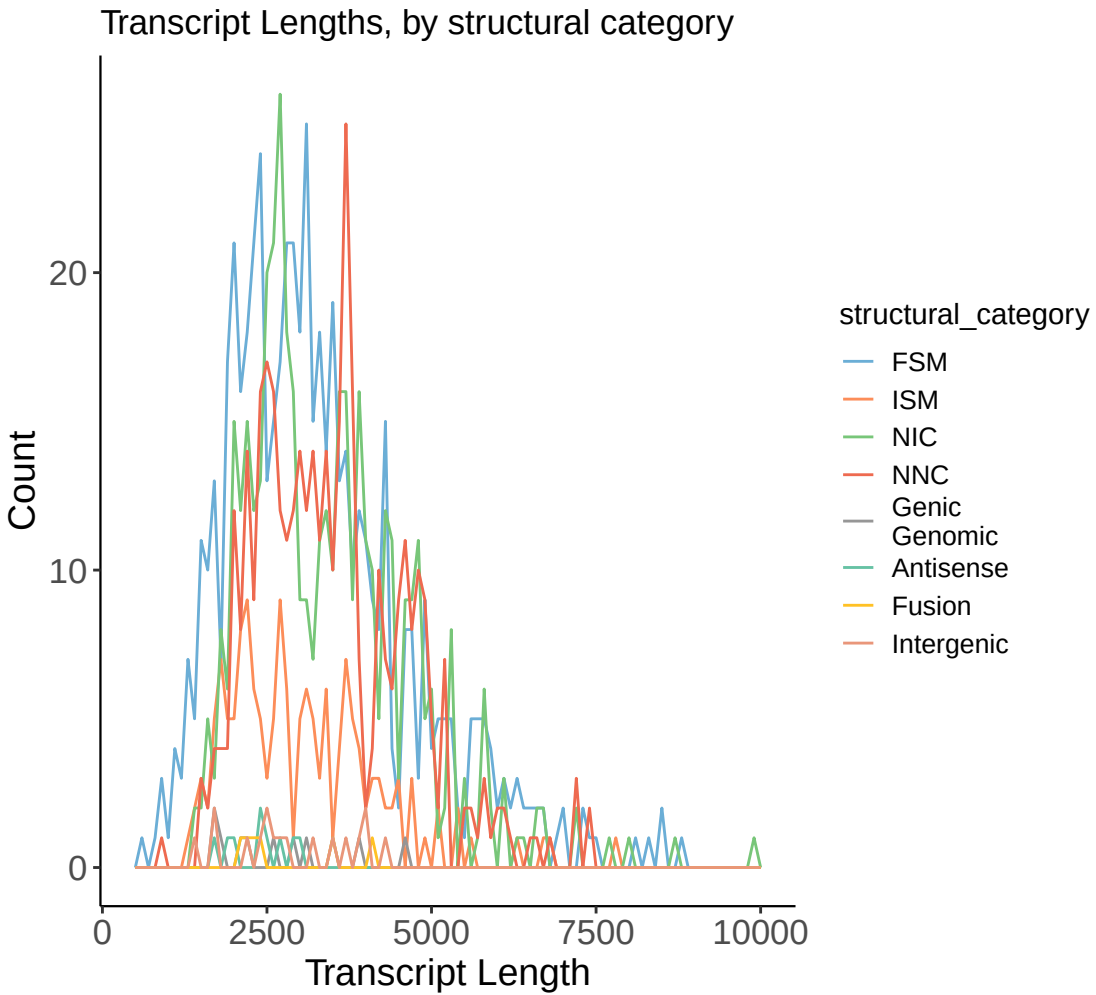

Classifications by Transcript Length
Structural Category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Counts
0
200
400
0
1
2
3
4
5
6
7
8
9
Transcript Length (in kb)

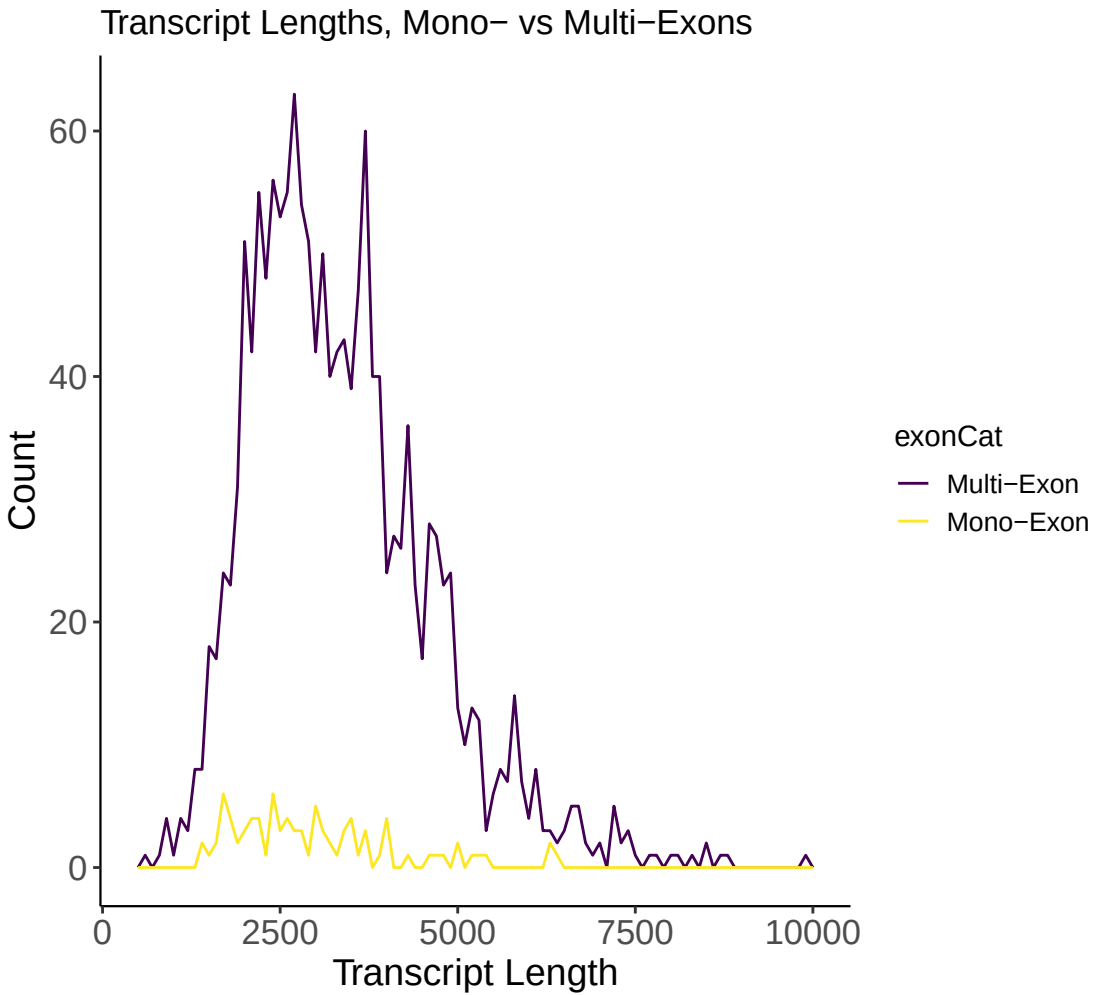

Classifications by Transcript Length, normalized
Percentages
100
75
50
25
0
0
1
2
3
4
5
6
7
8
9
Transcript Length (in kb)
Structural Category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic

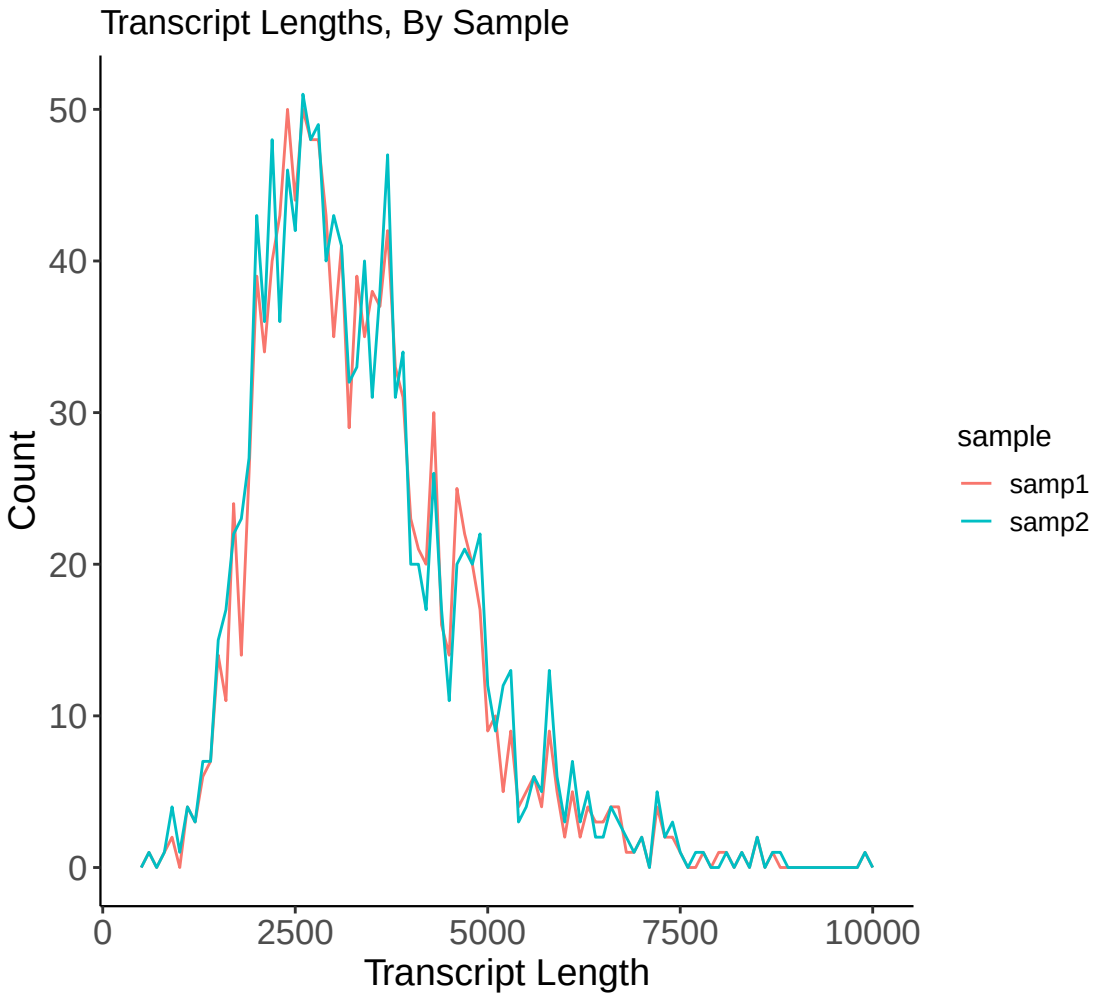

Gene Expression, Annotated vs Novel
log2(Gene_TPM+1)
9
6
3
0
Annotated Genes
Novel Genes

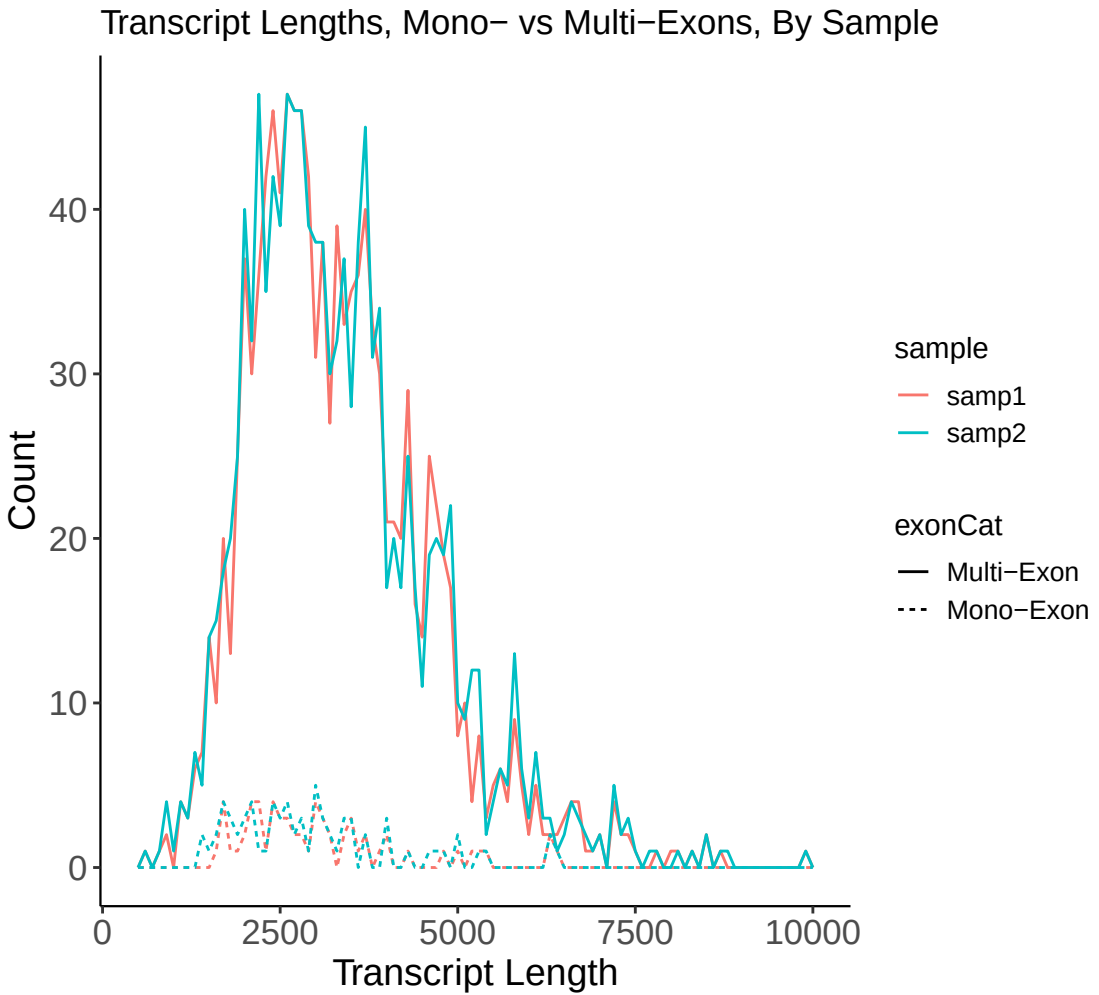

Number of FL reads per Gene by type of gene annotation
log2(FL_TPM+1)
15
10
Annotated Genes
Novel Genes


Transcript Lengths, all transcripts
Count
60
40
20
0
2500
5000
7500
10000
Transcript Length

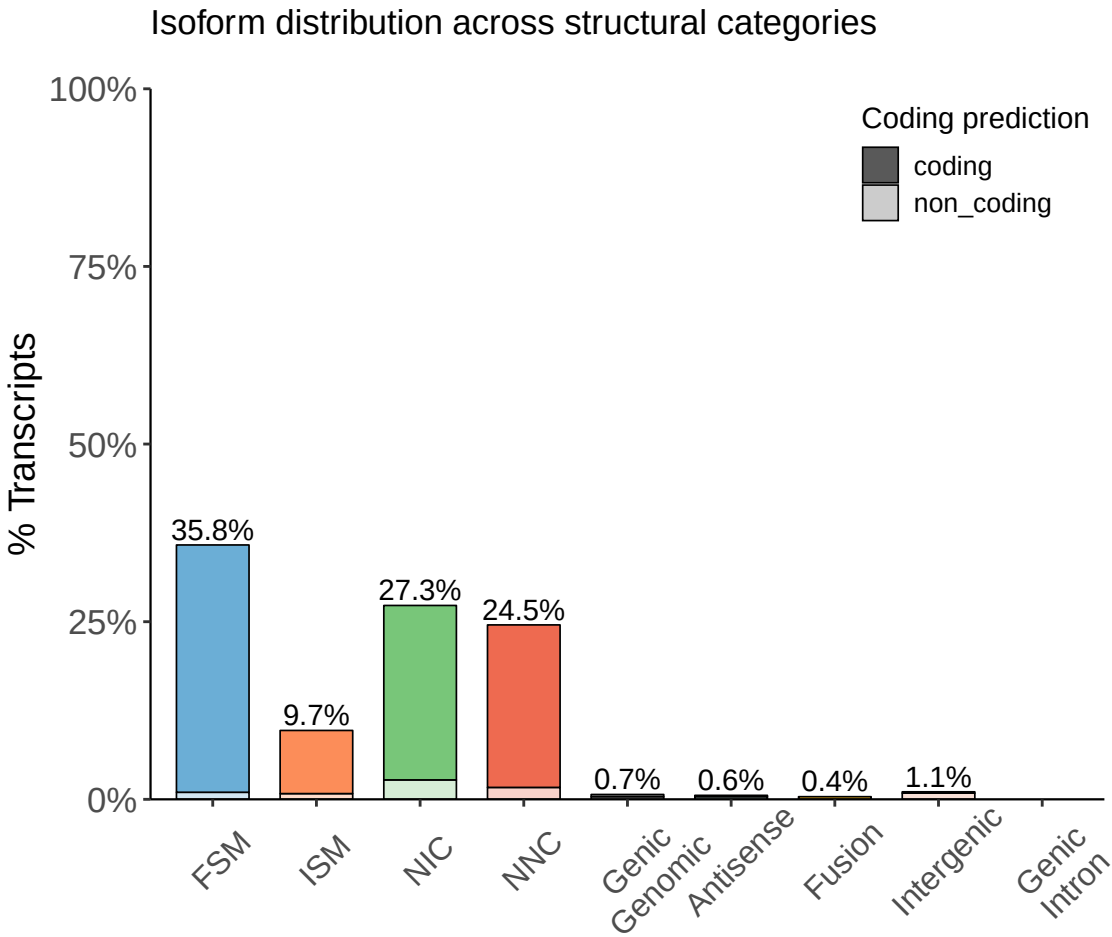

Transcript Lengths, by structural category
Count
20
10
0
0
2500
5000
7500
10000
Transcript Length
structural_category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic

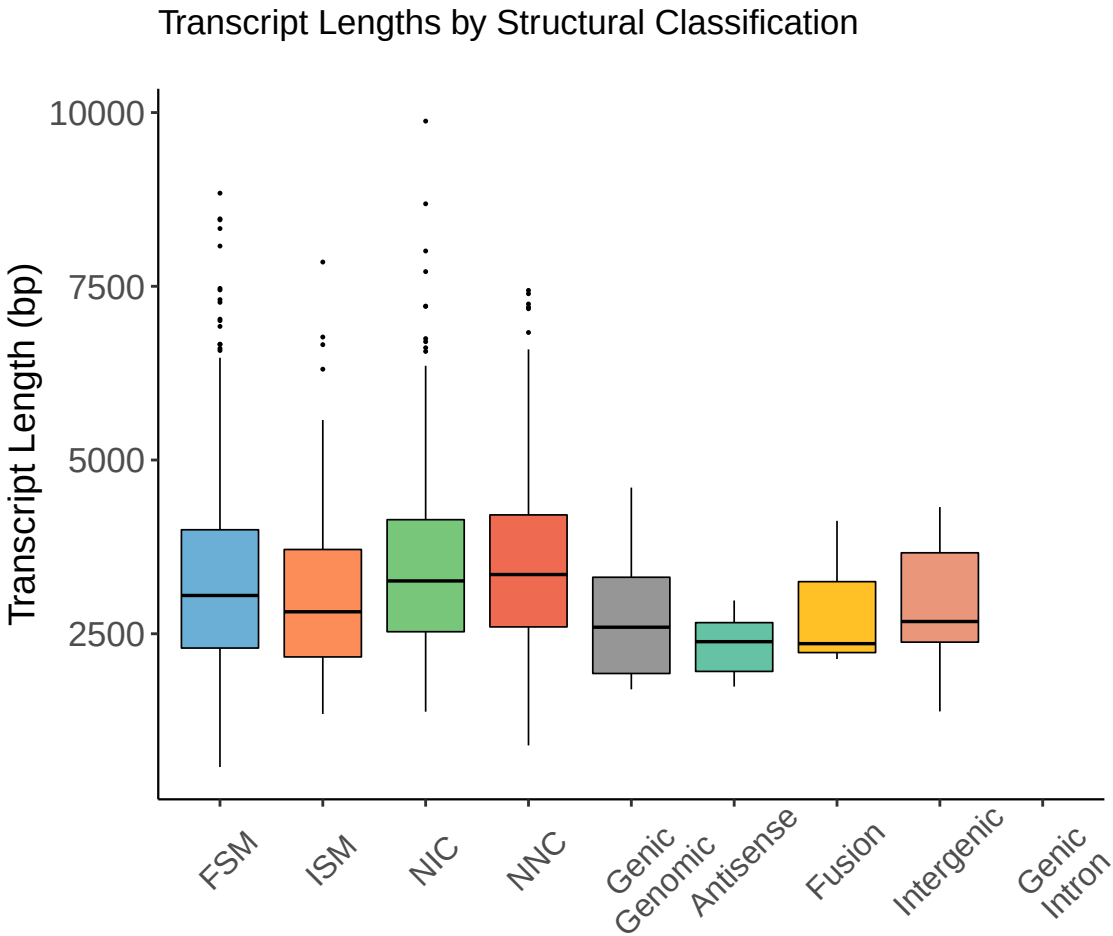

Transcript Lengths, Mono− vs Multi−Exons
Count
60
40
20
0
0
2500
5000
7500
10000
Transcript Length
exonCat
Multi−Exon
Mono−Exon

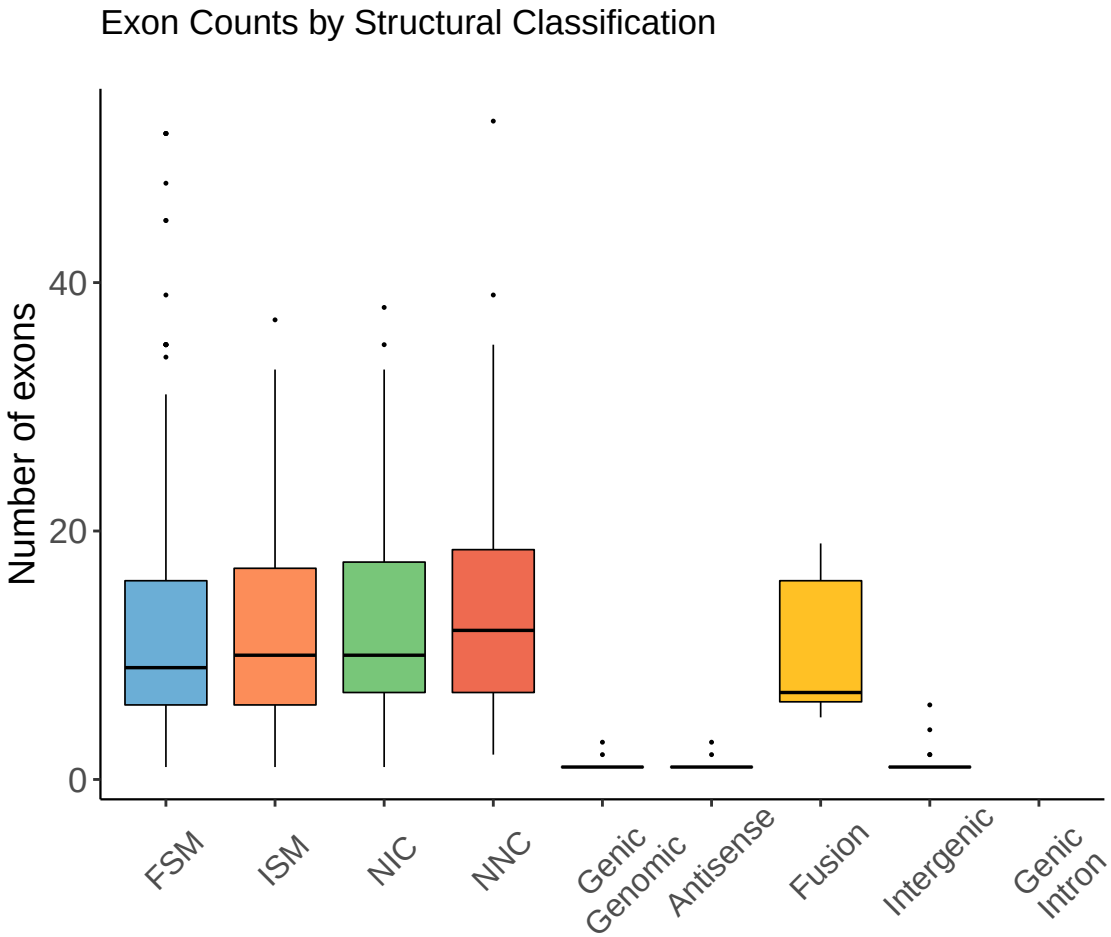

Transcript Lengths, By Sample
Count
50
40
30
20
10
0
0
2500
5000
7500
10000
Transcript Length
sample
samp1
samp2

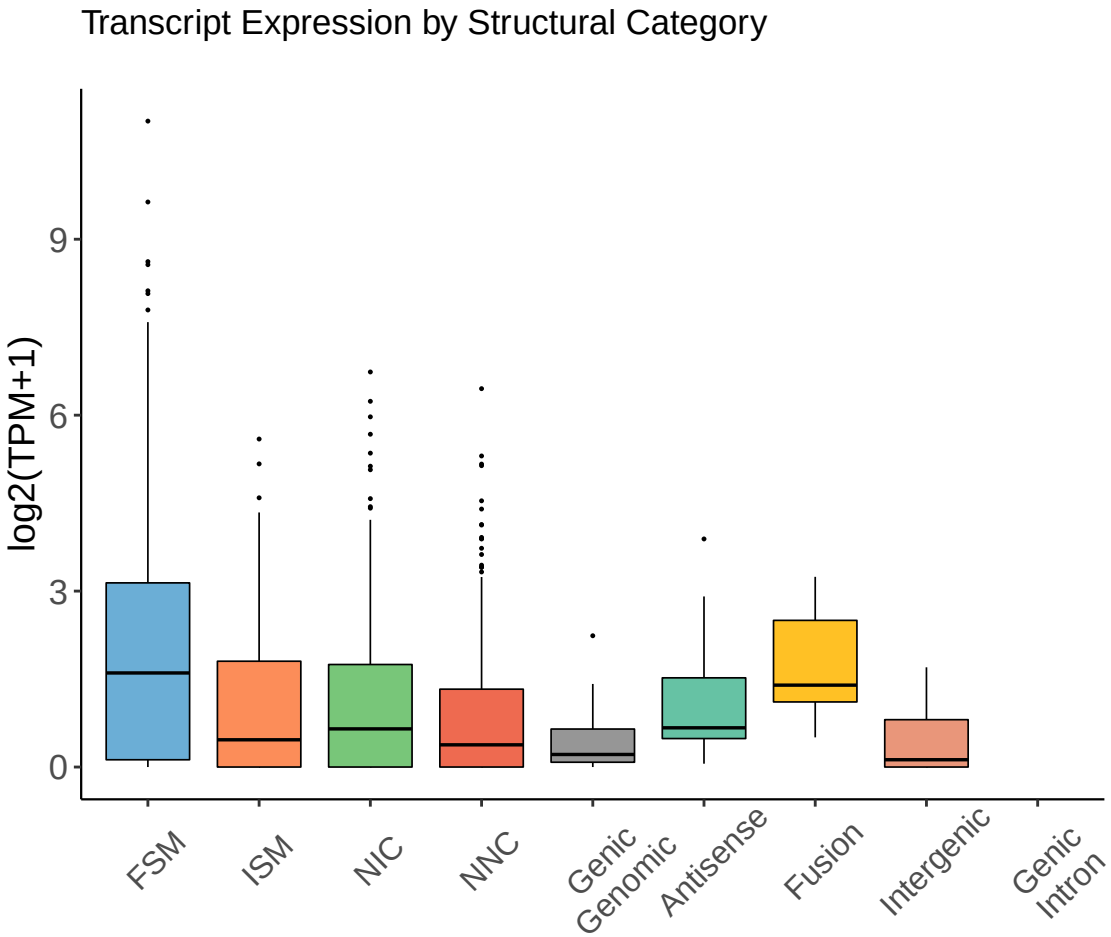

Transcript Lengths, Mono− vs Multi−Exons, By Sample
Count
0
10
20
30
40
0
2500
5000
7500
10000
Transcript Length
sample
samp1
samp2
exonCat
Multi−Exon
Mono−Exon

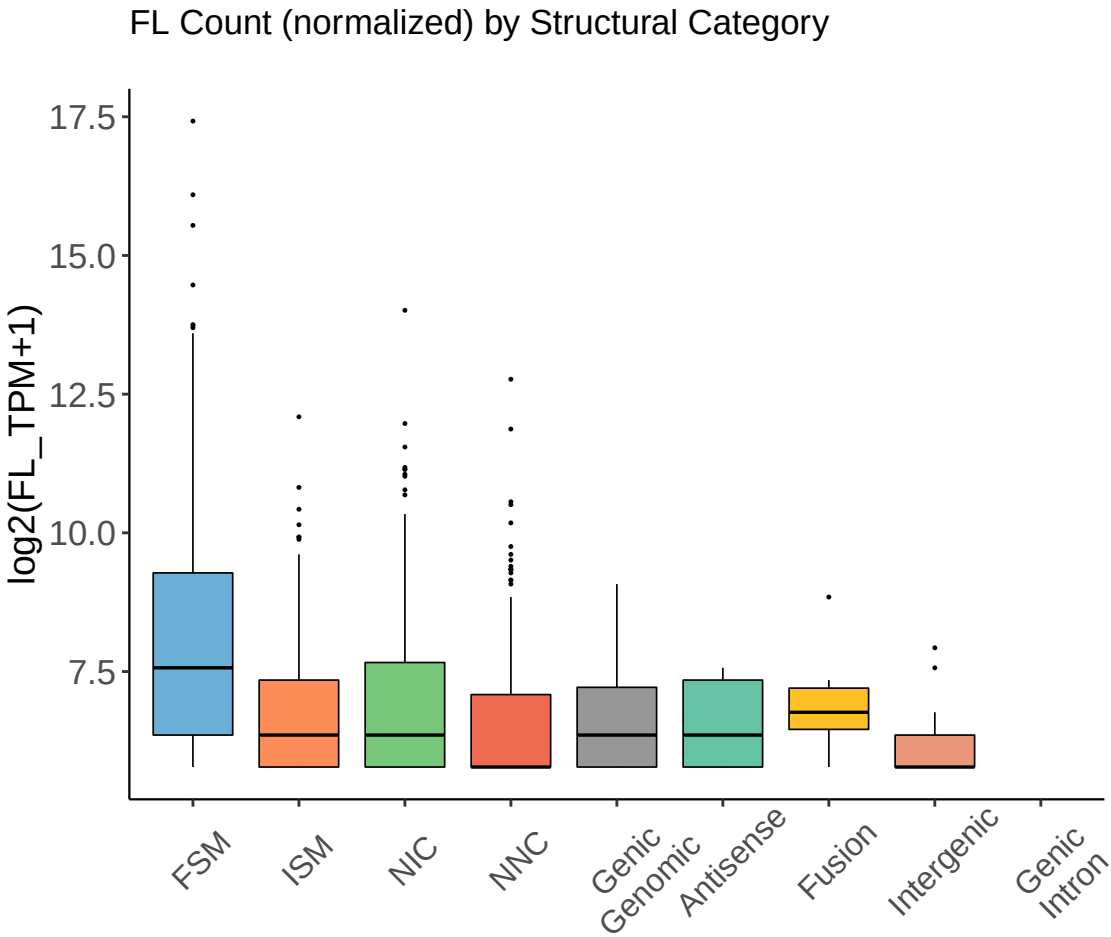
Structural Isoform Characterization by Splice Junctions


Isoform distribution across structural categories
% Transcripts
100%
75%
50%
25%
0%
Coding prediction
coding
non_coding
35.8%
9.7%
27.3%
24.5%
0.7%
0.6%
0.4%
1.1%
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron


Transcript Lengths by Structural Classification
Transcript Length (bp)
10000
7500
5000
2500
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

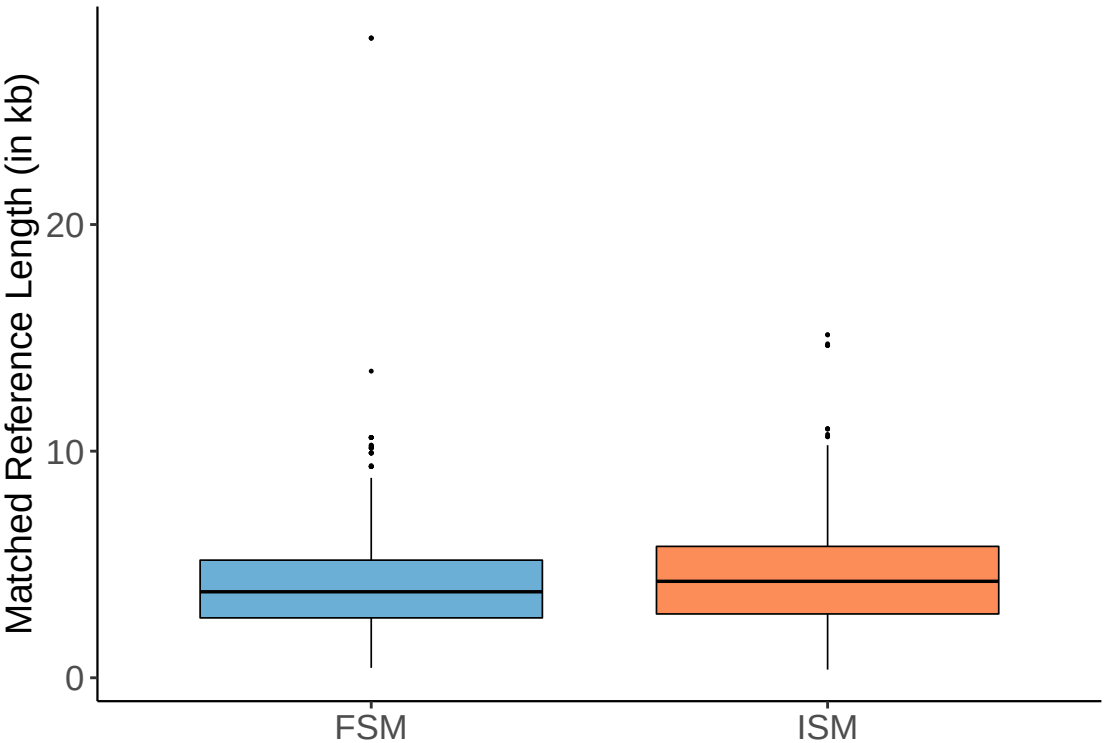

Exon Counts by Structural Classification
Number of exons
0
20
40
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron


Transcript Expression by Structural Category
log2(TPM+1)
0
3
6
9
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron


FL Count (normalized) by Structural Category
log2(FL_TPM+1)
17.5
15.0
12.5
10.0
7.5
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

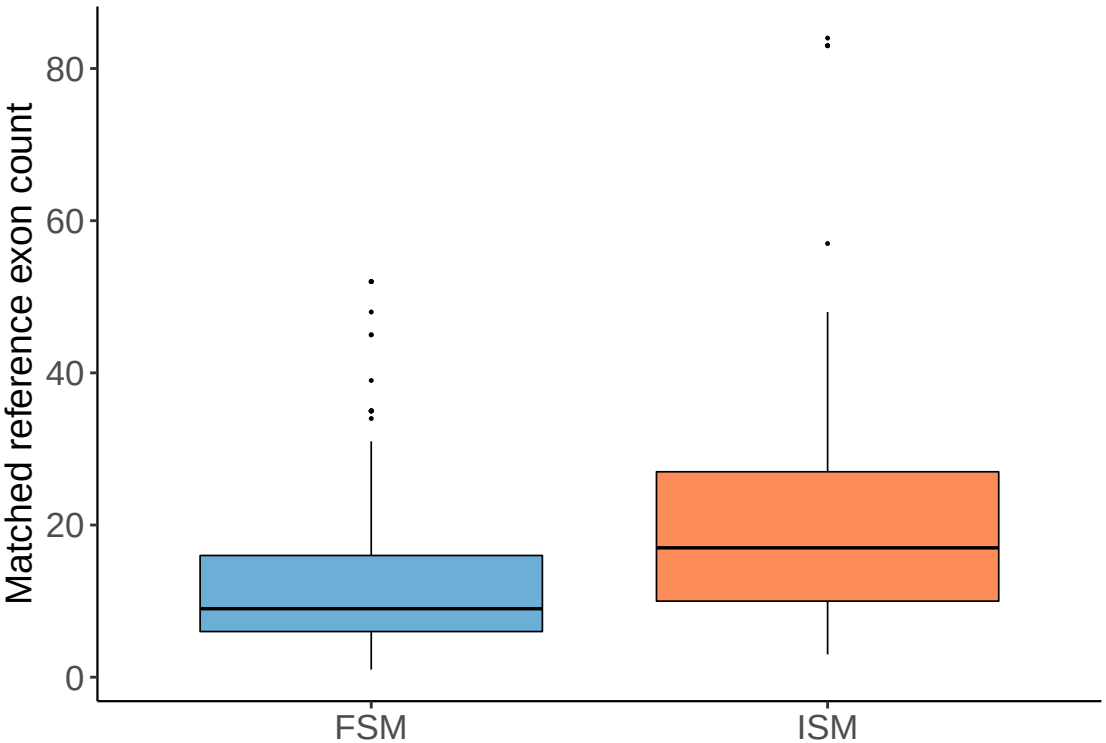
Length Distribution of Matched Reference Transcripts

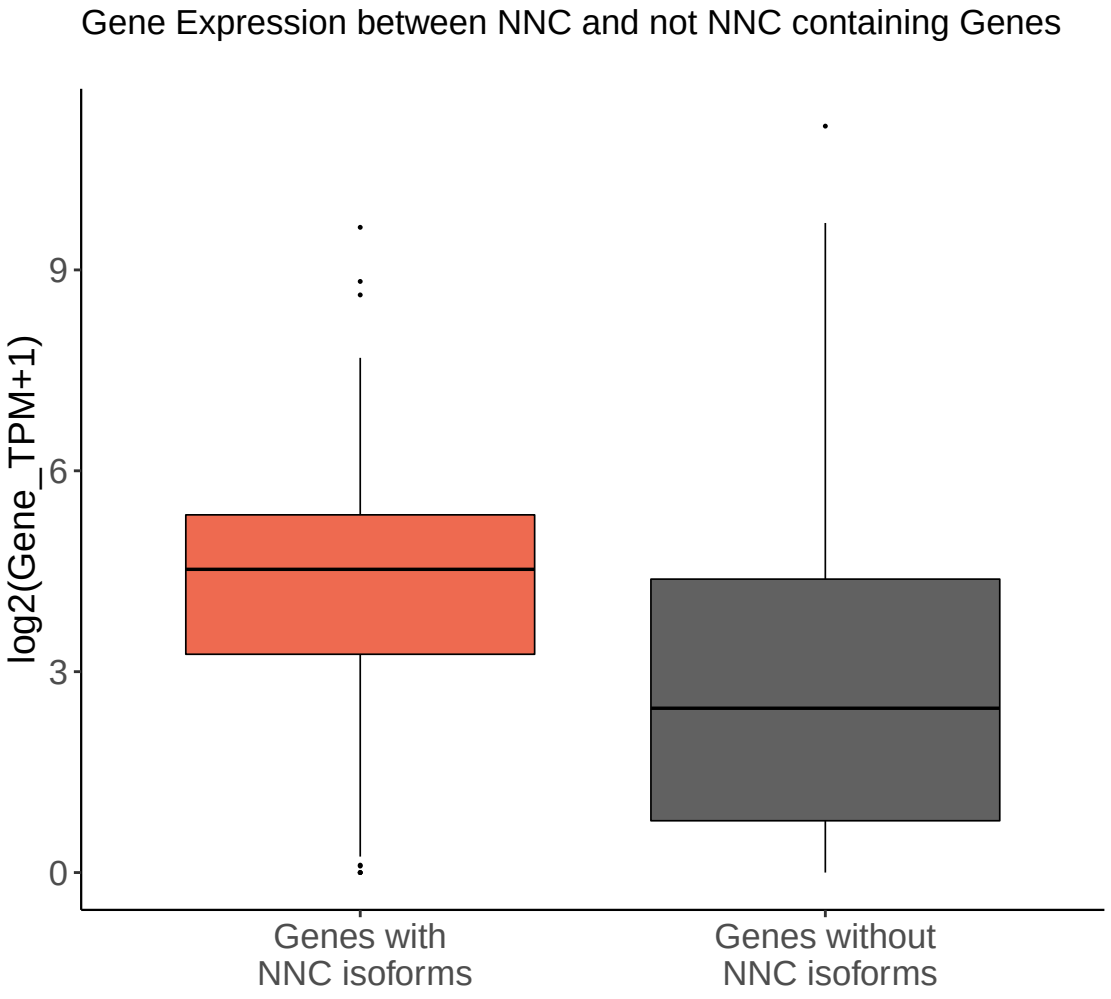
Applicable only to FSM and ISM categories



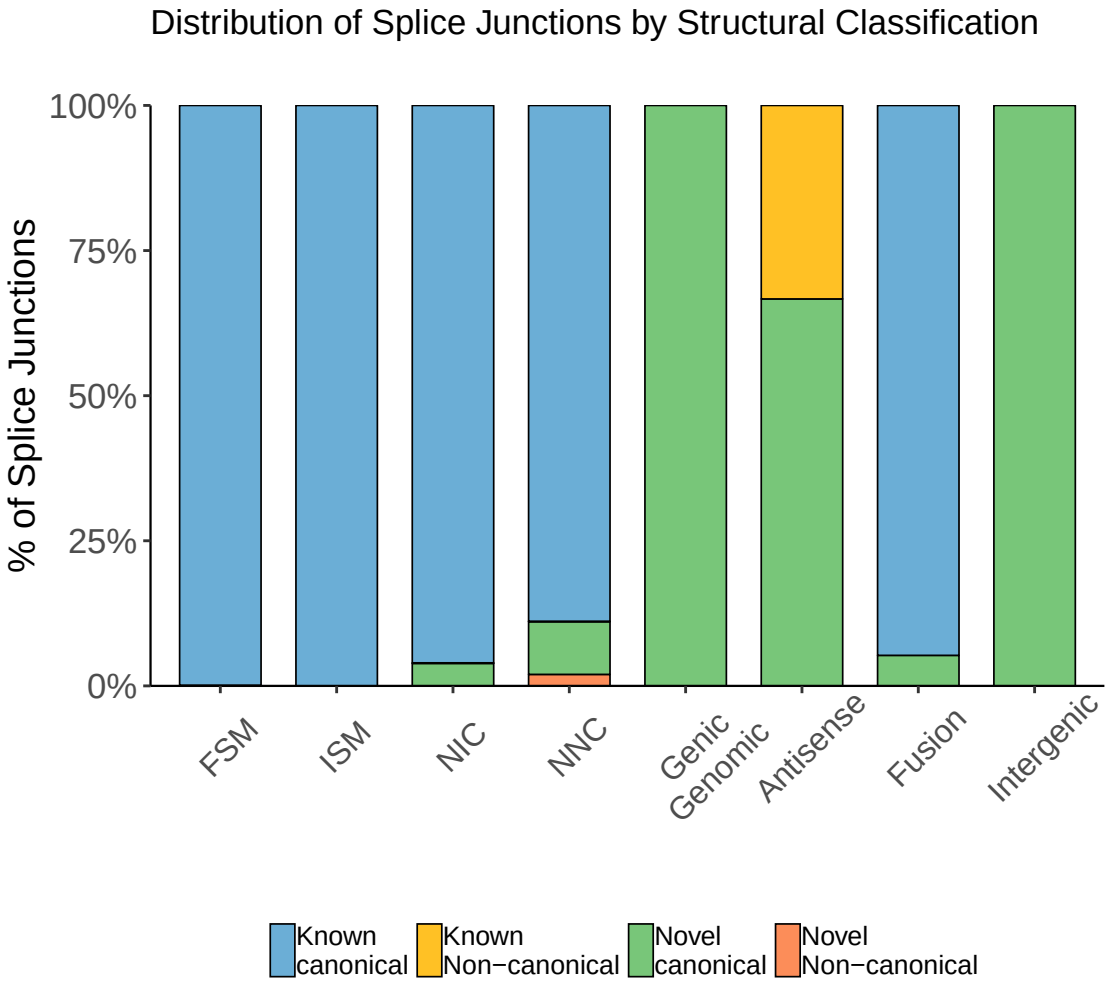
Exon Count Distribution of Matched Reference Transcripts

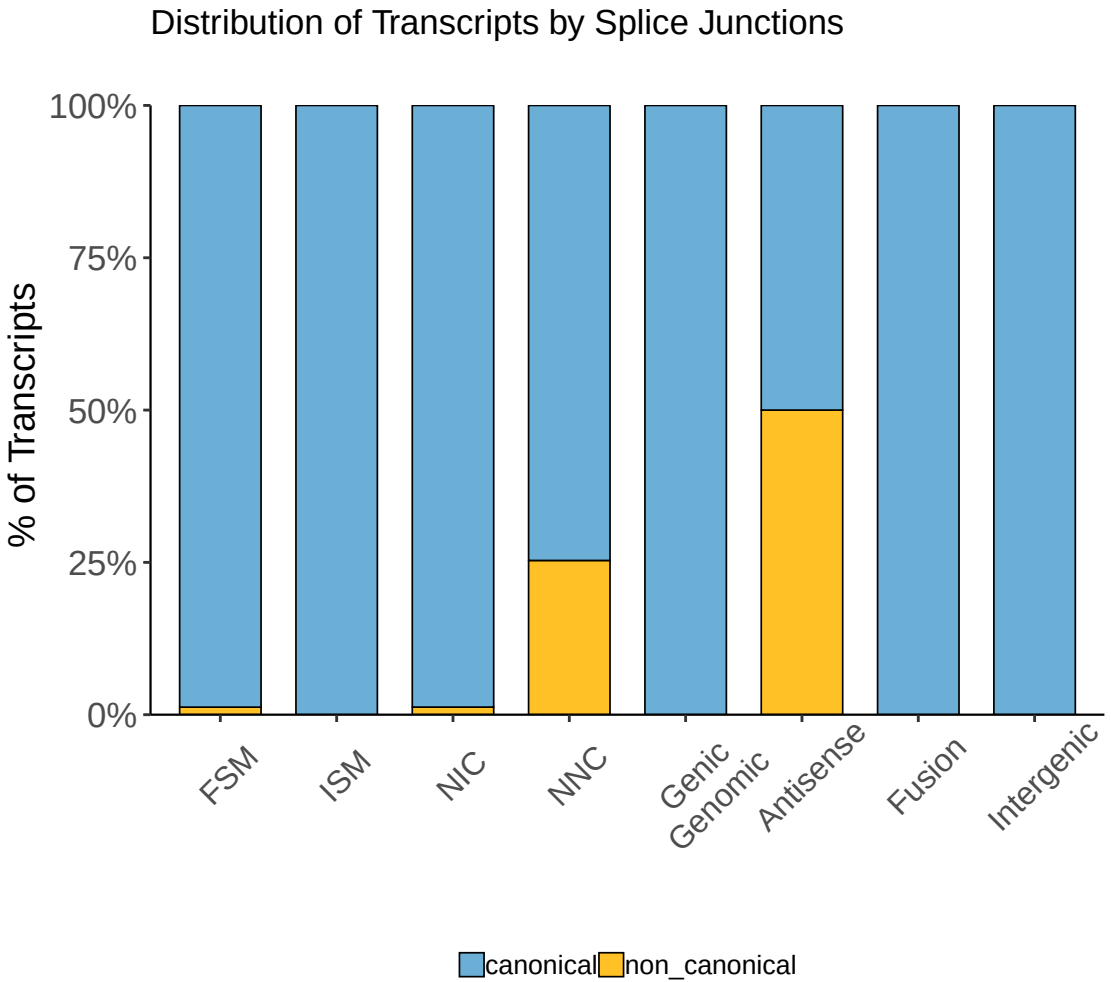
Applicable only to FSM and ISM categories

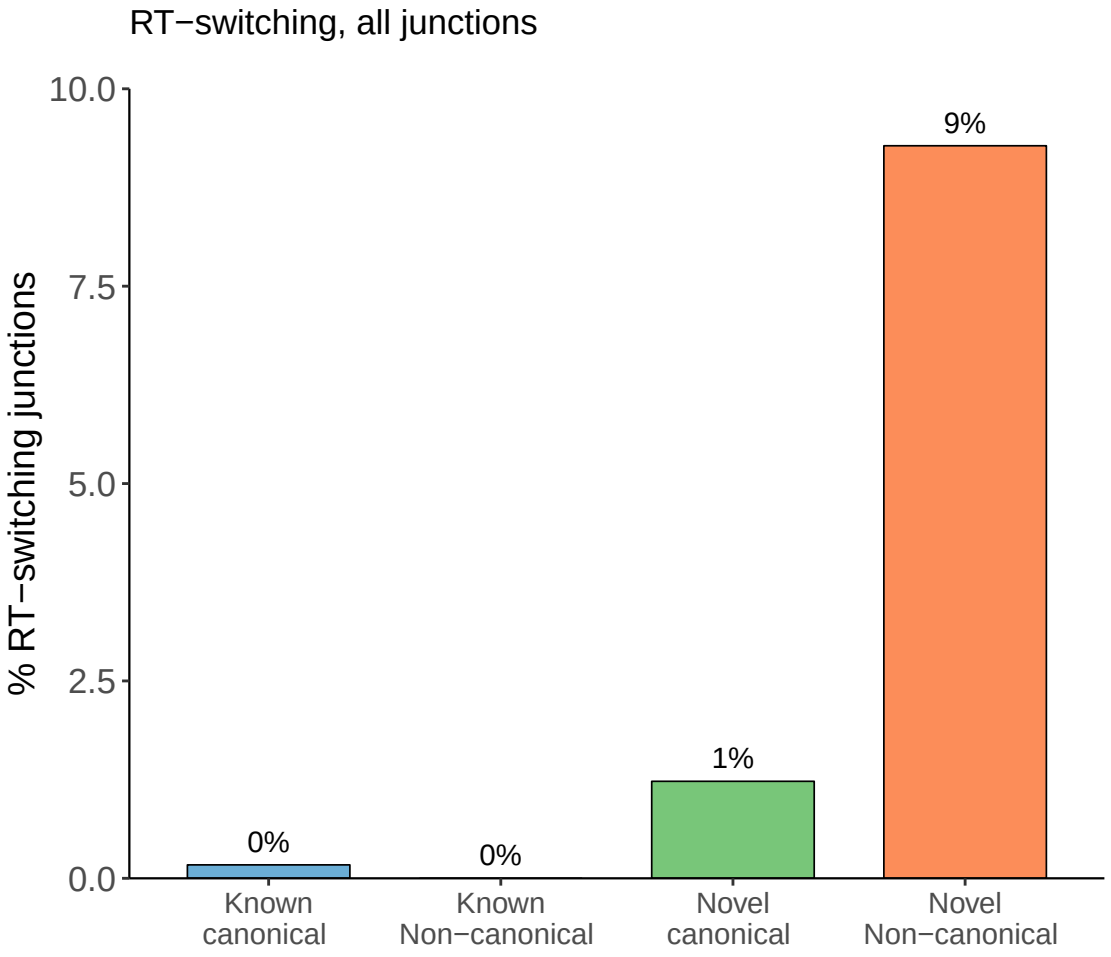


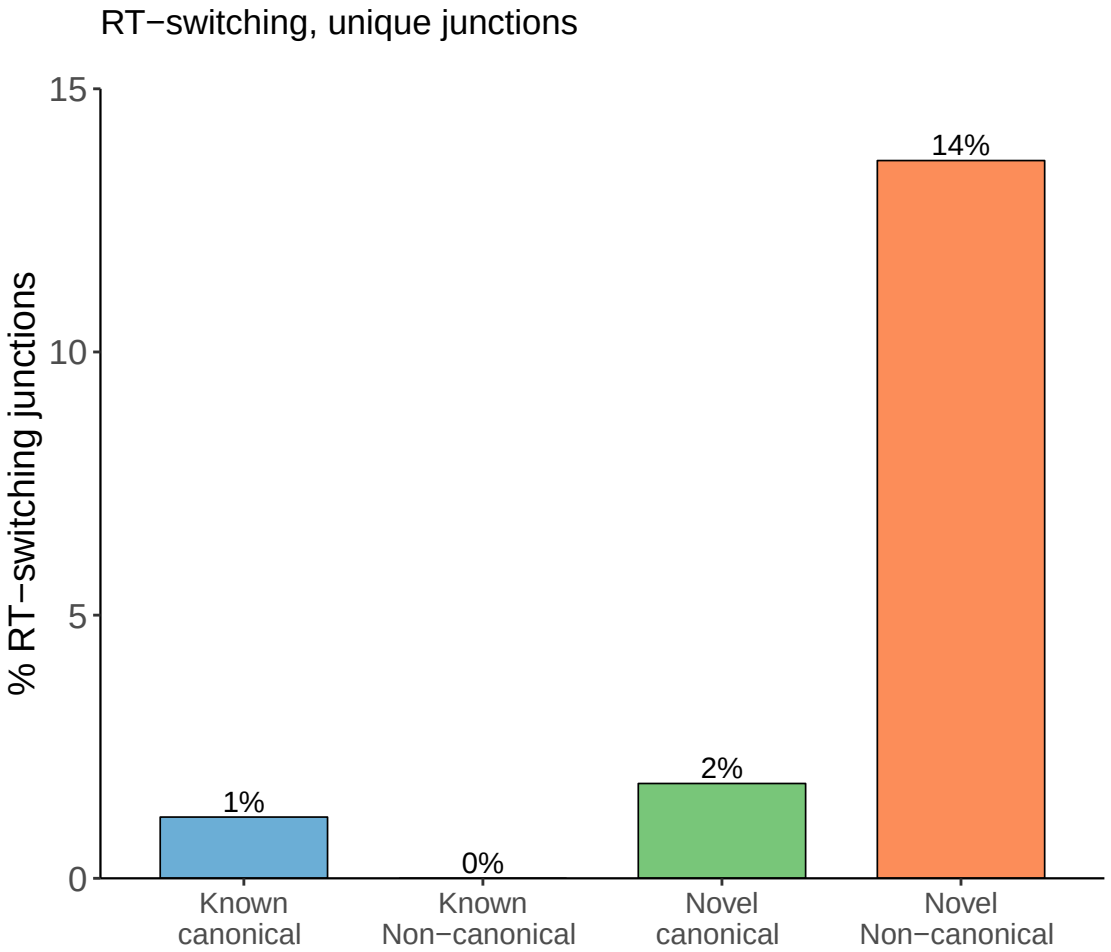

Gene Expression between NNC and not NNC containing Genes
log2(Gene_TPM+1)
9
6
3
0
Genes with
NNC isoforms
Genes without
NNC isoforms

Splice Junction Characterization

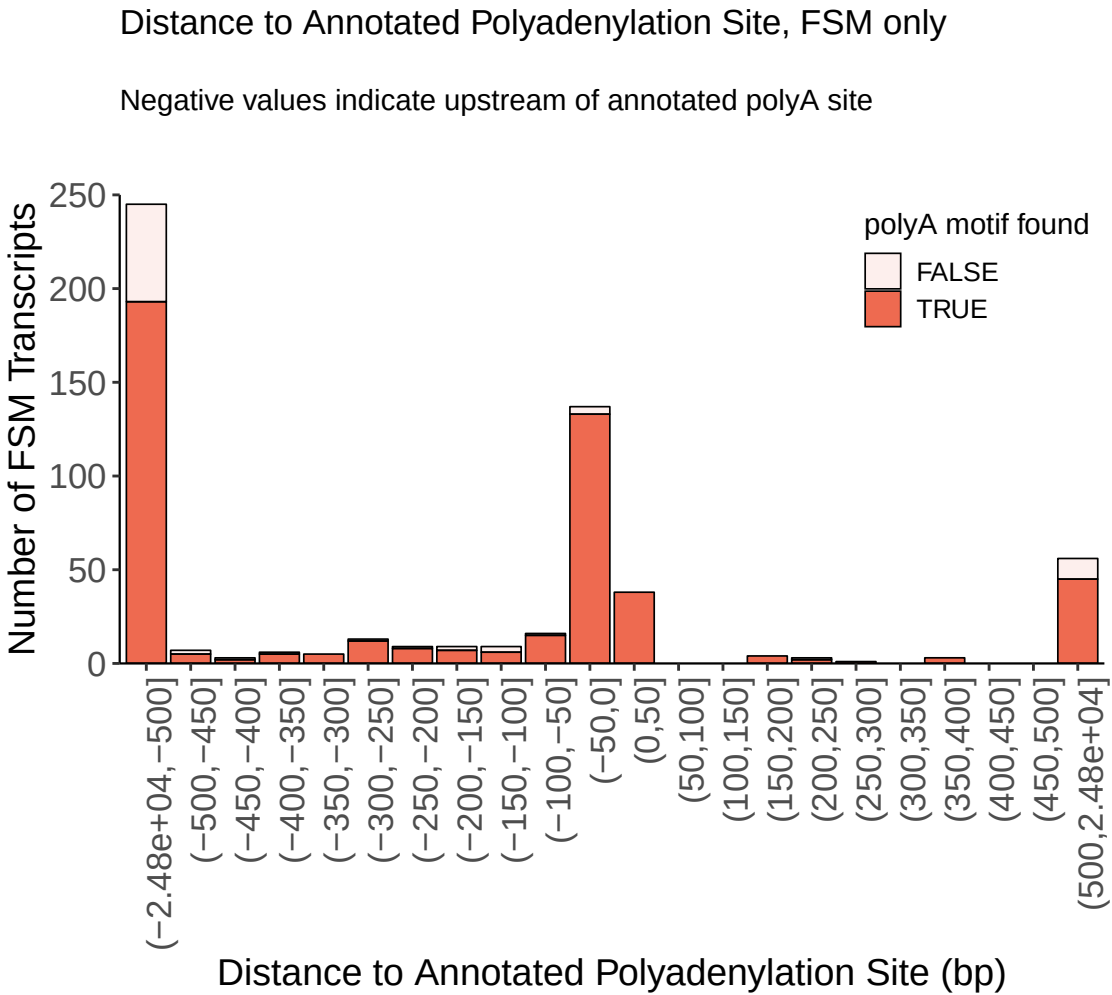

Distribution of Splice Junctions by Structural Classification
% of Splice Junctions
100%
75%
50%
25%
0%
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

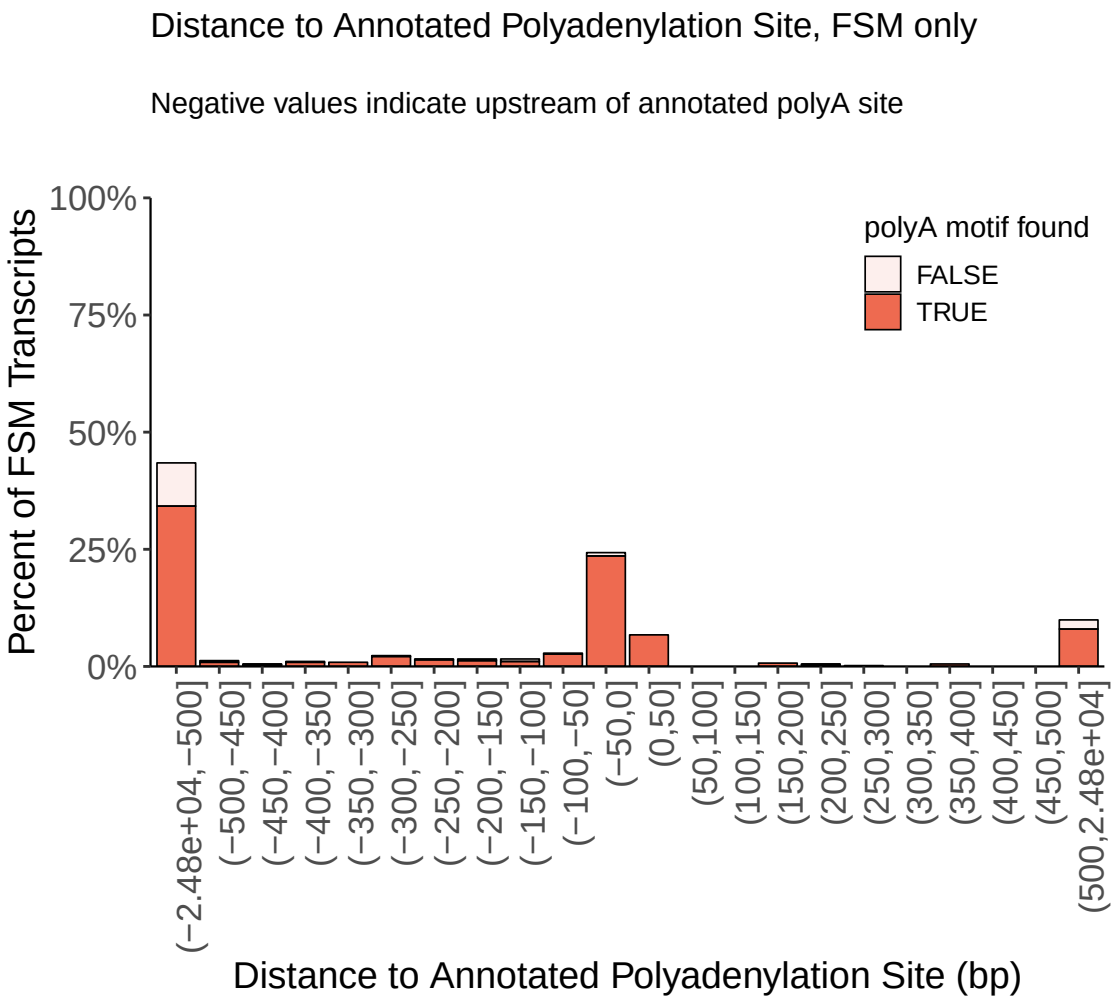

Distribution of Transcripts by Splice Junctions
% of Transcripts
100%
75%
50%
25%
0%
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
canonical
non_canonical

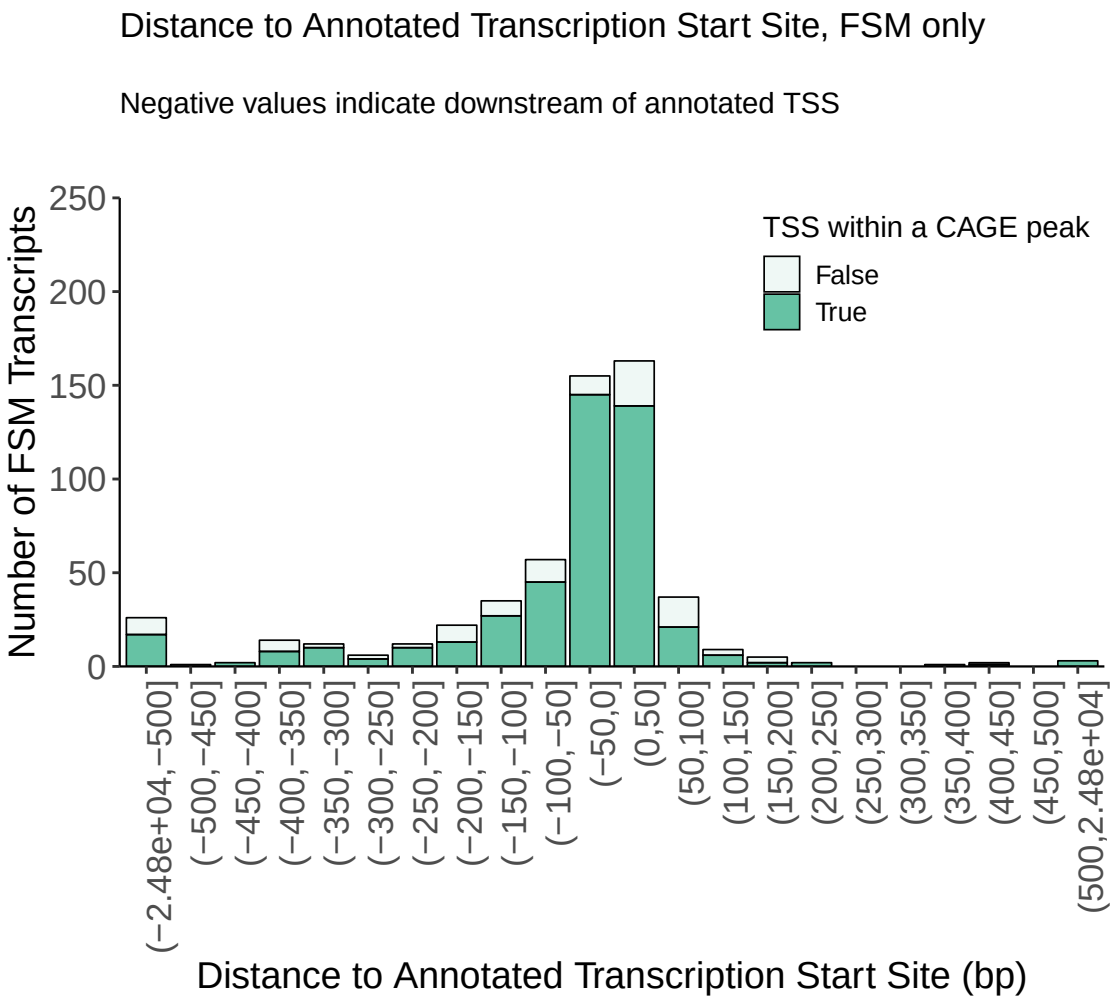

RT−switching, all junctions
% RT−switching junctions
10.0
7.5
5.0
2.5
0.0
0%
0%
1%
9%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

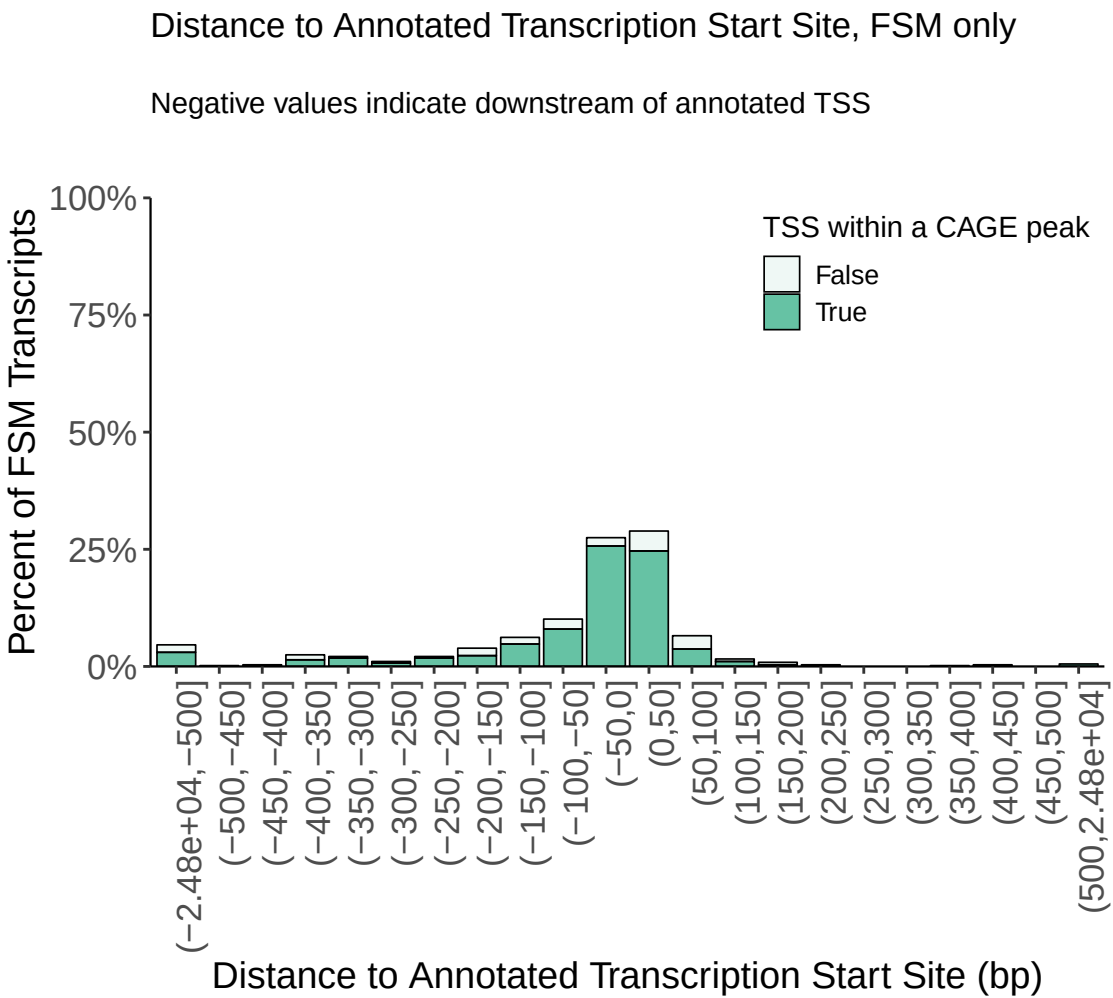

RT−switching, unique junctions
% RT−switching junctions
15
10
5
0
1%
0%
2%
14%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

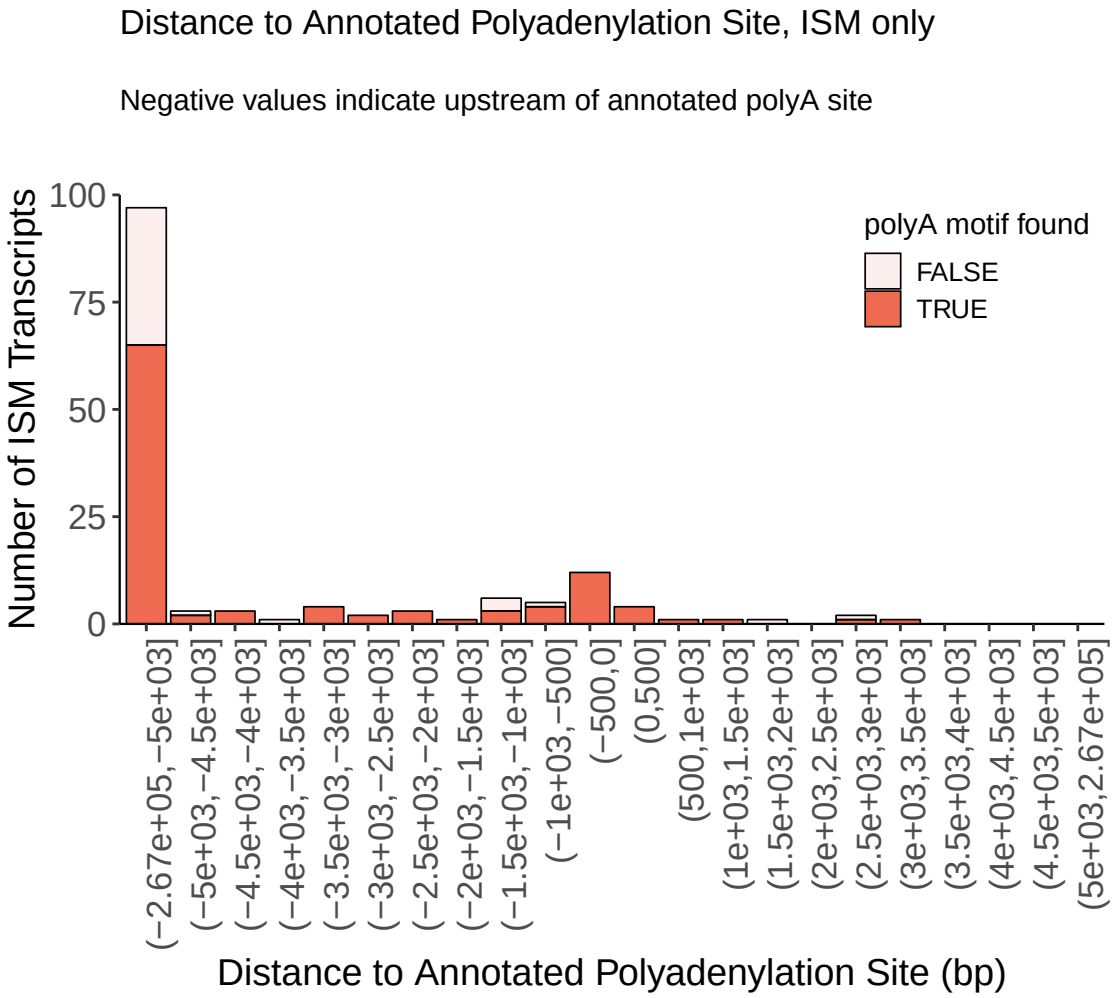
Comparison with Annotated TSS and PolyA Sites

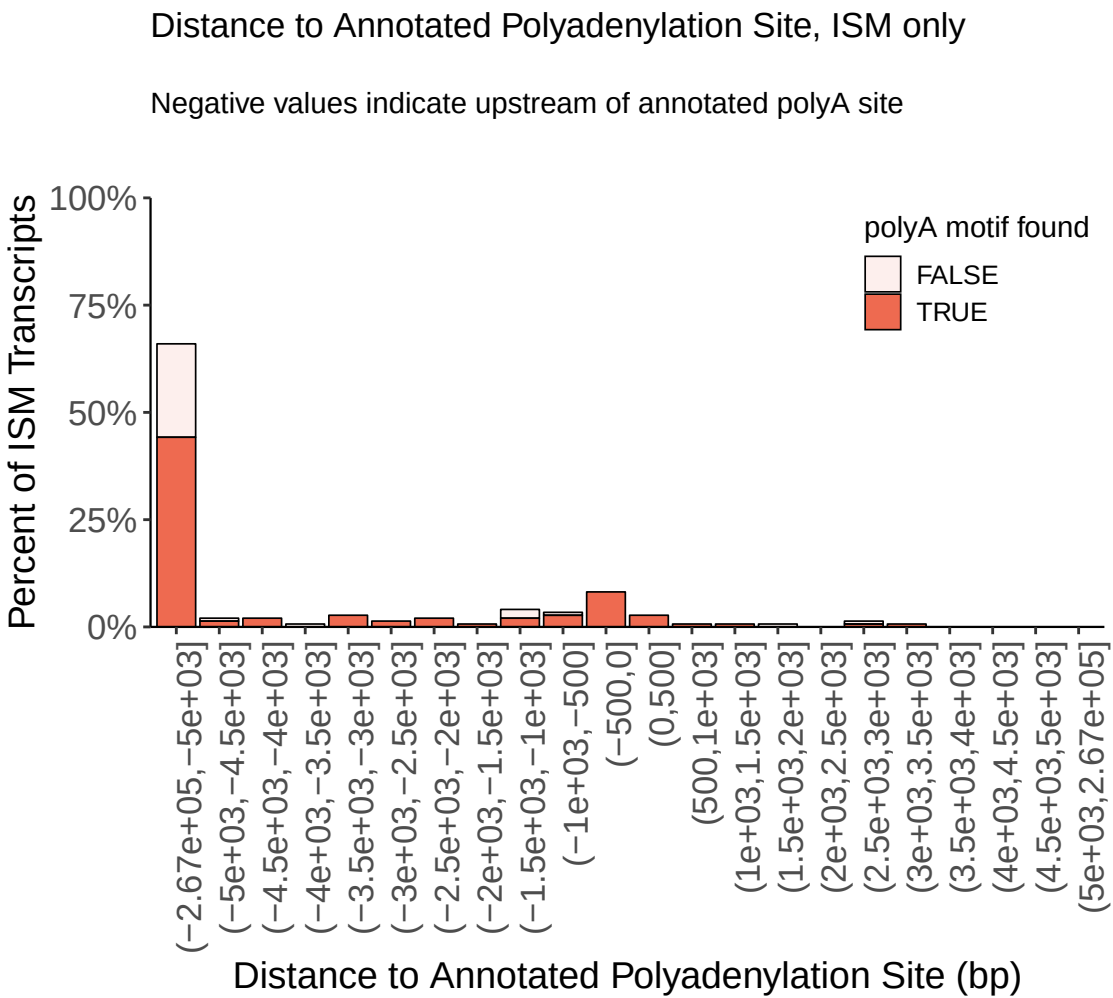

Distance to Annotated Polyadenylation Site, FSM only
Negative values indicate upstream of annotated polyA site
Number of FSM Transcripts
250
200
150
100
50
0
polyA motif found
FALSE
TRUE
(−2.48e+04,−500]
(−500,−450]
(−450,−400]
(−400,−350]
(−350,−300]
(−300,−250]
(−250,−200]
(−200,−150]
(−150,−100]
(−100,−50]
(−50,0]
(0,50]
(50,100]
(100,150]
(150,200]
(200,250]
(250,300]
(300,350]
(350,400]
(400,450]
(450,500]
(500,2.48e+04]
Distance to Annotated Polyadenylation Site (bp)

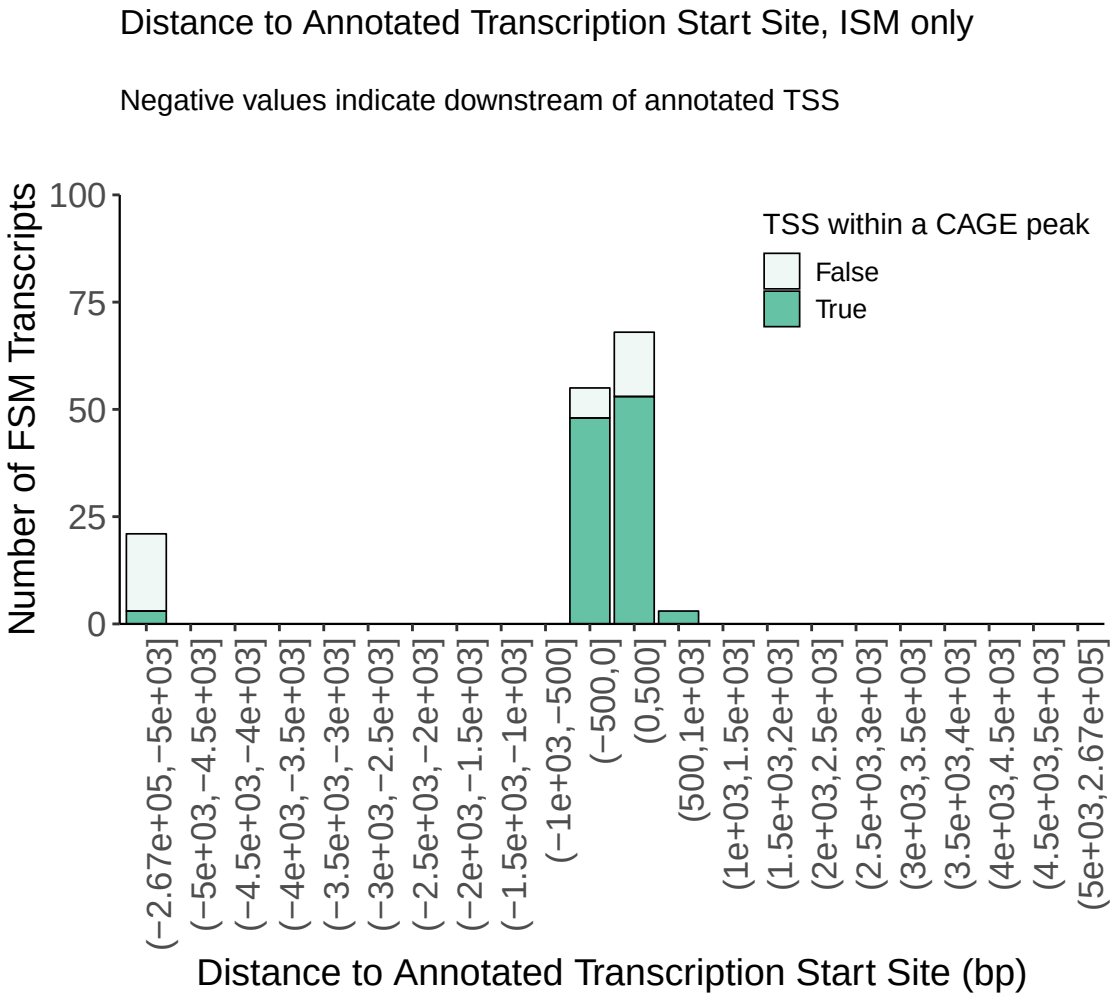

Distance to Annotated Polyadenylation Site, FSM only
Negative values indicate upstream of annotated polyA site
Percent of FSM Transcripts
100%
75%
50%
25%
0%
polyA motif found
FALSE
TRUE
(−2.48e+04,−500]
(−500,−450]
(−450,−400]
(−400,−350]
(−350,−300]
(−300,−250]
(−250,−200]
(−200,−150]
(−150,−100]
(−100,−50]
(−50,0]
(0,50]
(50,100]
(100,150]
(150,200]
(200,250]
(250,300]
(300,350]
(350,400]
(400,450]
(450,500]
(500,2.48e+04]
Distance to Annotated Polyadenylation Site (bp)

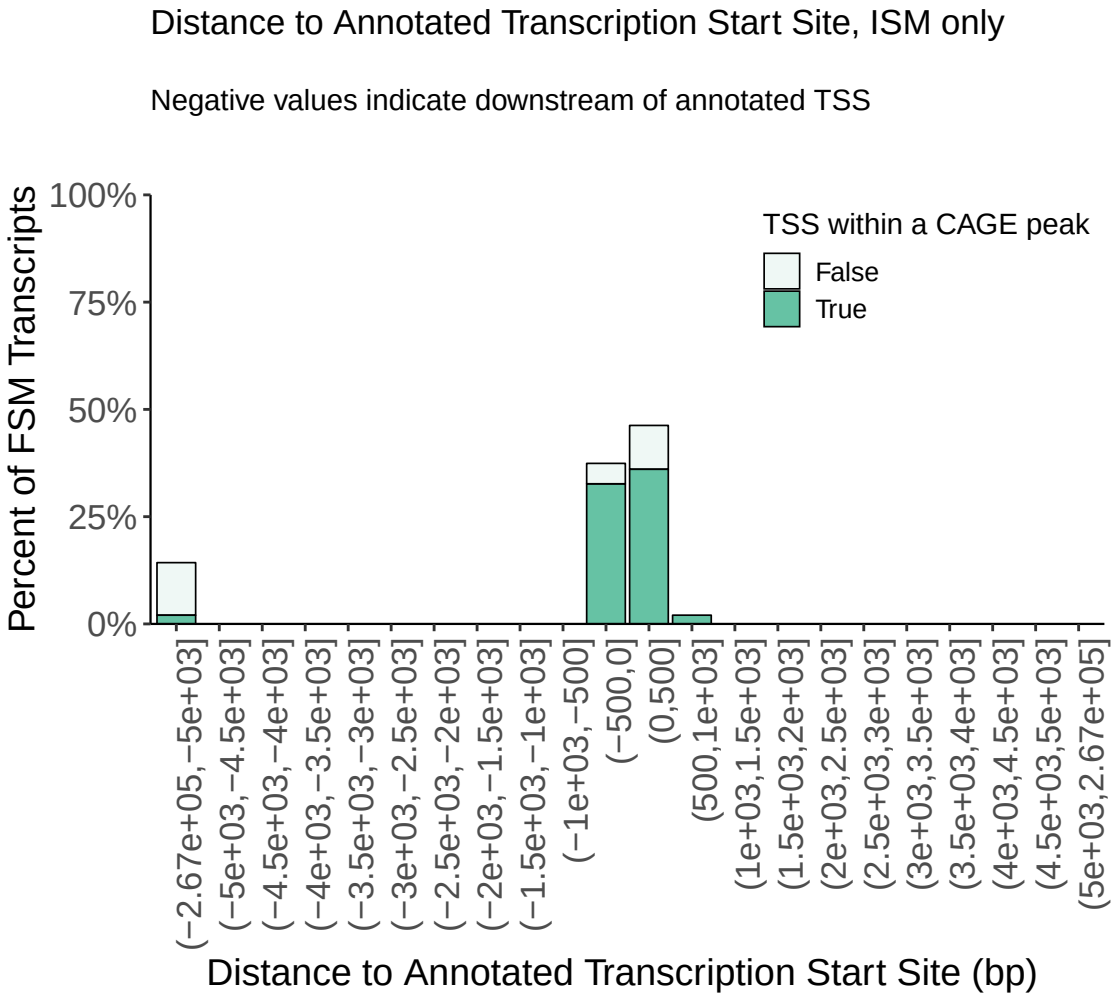

Distance to Annotated Transcription Start Site, FSM only
Negative values indicate downstream of annotated TSS
Number of FSM Transcripts
250
200
150
100
50
0
TSS within a CAGE peak
False
True
(−2.48e+04,−500]
(−500,−450]
(−450,−400]
(−400,−350]
(−350,−300]
(−300,−250]
(−250,−200]
(−200,−150]
(−150,−100]
(−100,−50]
(−50,0]
(0,50]
(50,100]
(100,150]
(150,200]
(200,250]
(250,300]
(300,350]
(350,400]
(400,450]
(450,500]
(500,2.48e+04]
Distance to Annotated Transcription Start Site (bp)

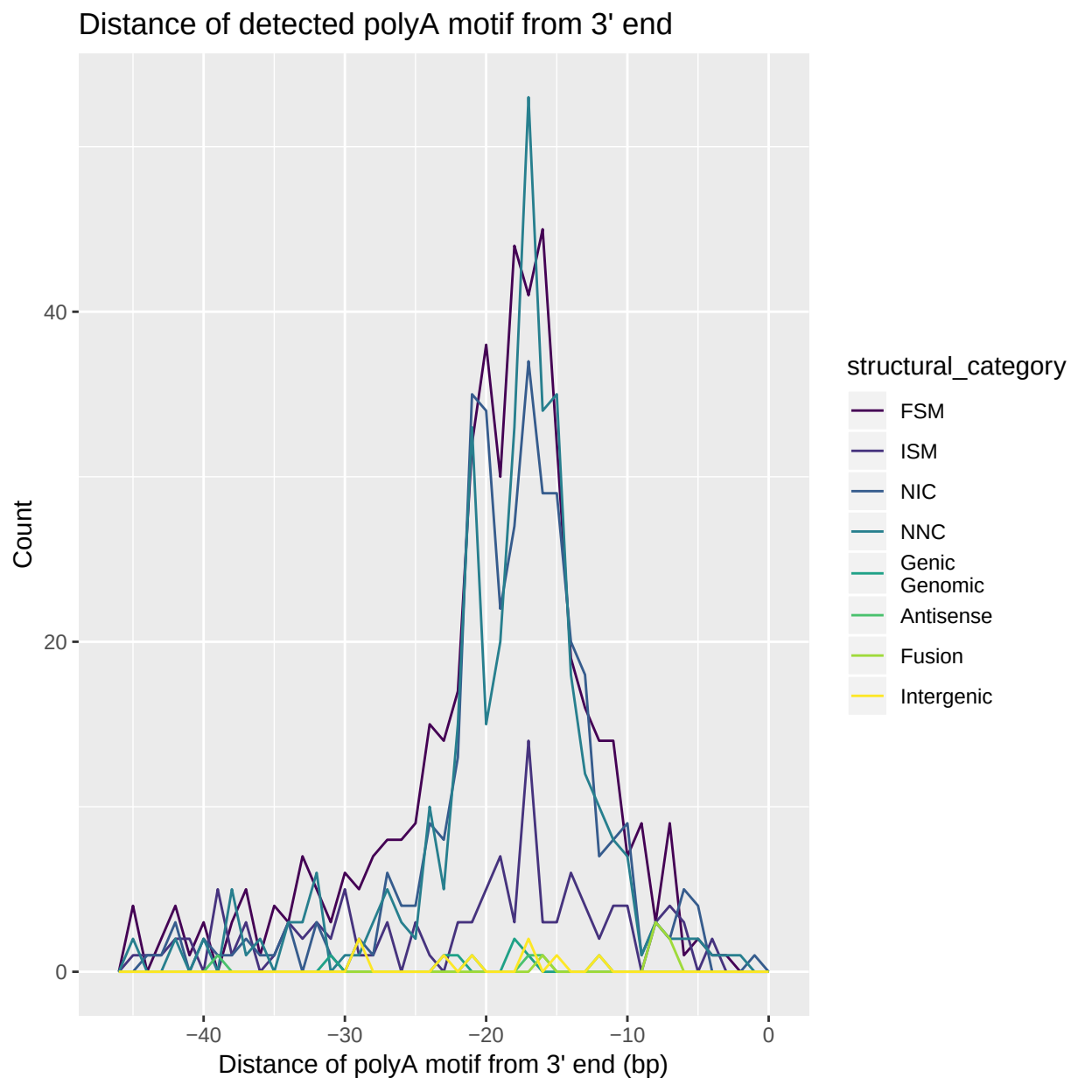

Distance to Annotated Transcription Start Site, FSM only
Negative values indicate downstream of annotated TSS
Percent of FSM Transcripts
100%
75%
50%
25%
0%
TSS within a CAGE peak
False
True
(−2.48e+04,−500]
(−500,−450]
(−450,−400]
(−400,−350]
(−350,−300]
(−300,−250]
(−250,−200]
(−200,−150]
(−150,−100]
(−100,−50]
(−50,0]
(0,50]
(50,100]
(100,150]
(150,200]
(200,250]
(250,300]
(300,350]
(350,400]
(400,450]
(450,500]
(500,2.48e+04]
Distance to Annotated Transcription Start Site (bp)


Distance to Annotated Polyadenylation Site, ISM only
Negative values indicate upstream of annotated polyA site
Number of ISM Transcripts
100
75
50
25
0
polyA motif found
FALSE
TRUE
(−2.67e+05,−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+03,2.67e+05]
Distance to Annotated Polyadenylation Site (bp)


Distance to Annotated Polyadenylation Site, ISM only
Negative values indicate upstream of annotated polyA site
Percent of ISM Transcripts
100%
75%
50%
25%
0%
polyA motif found
FALSE
TRUE
(−2.67e+05,−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+03,2.67e+05]
Distance to Annotated Polyadenylation Site (bp)


Distance to Annotated Transcription Start Site, ISM only
Negative values indicate downstream of annotated TSS
Number of FSM Transcripts
100
75
50
25
0
TSS within a CAGE peak
False
True
(−2.67e+05,−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+03,2.67e+05]
Distance to Annotated Transcription Start Site (bp)


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
(−2.67e+05,−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+03,2.67e+05]
Distance to Annotated Transcription Start Site (bp)


Distance of detected polyA motif from 3' end
Count
40
20
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

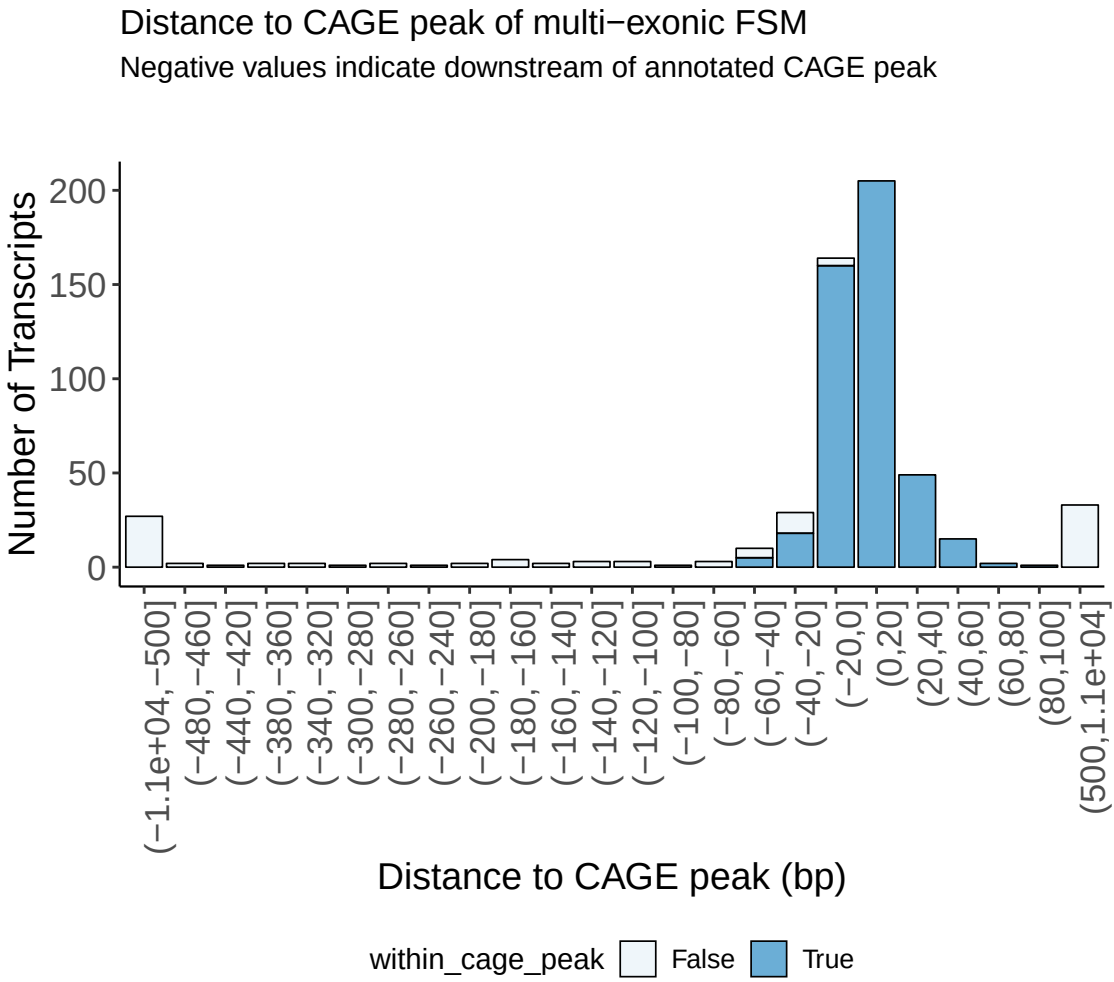
Number of polyA Motifs Detected

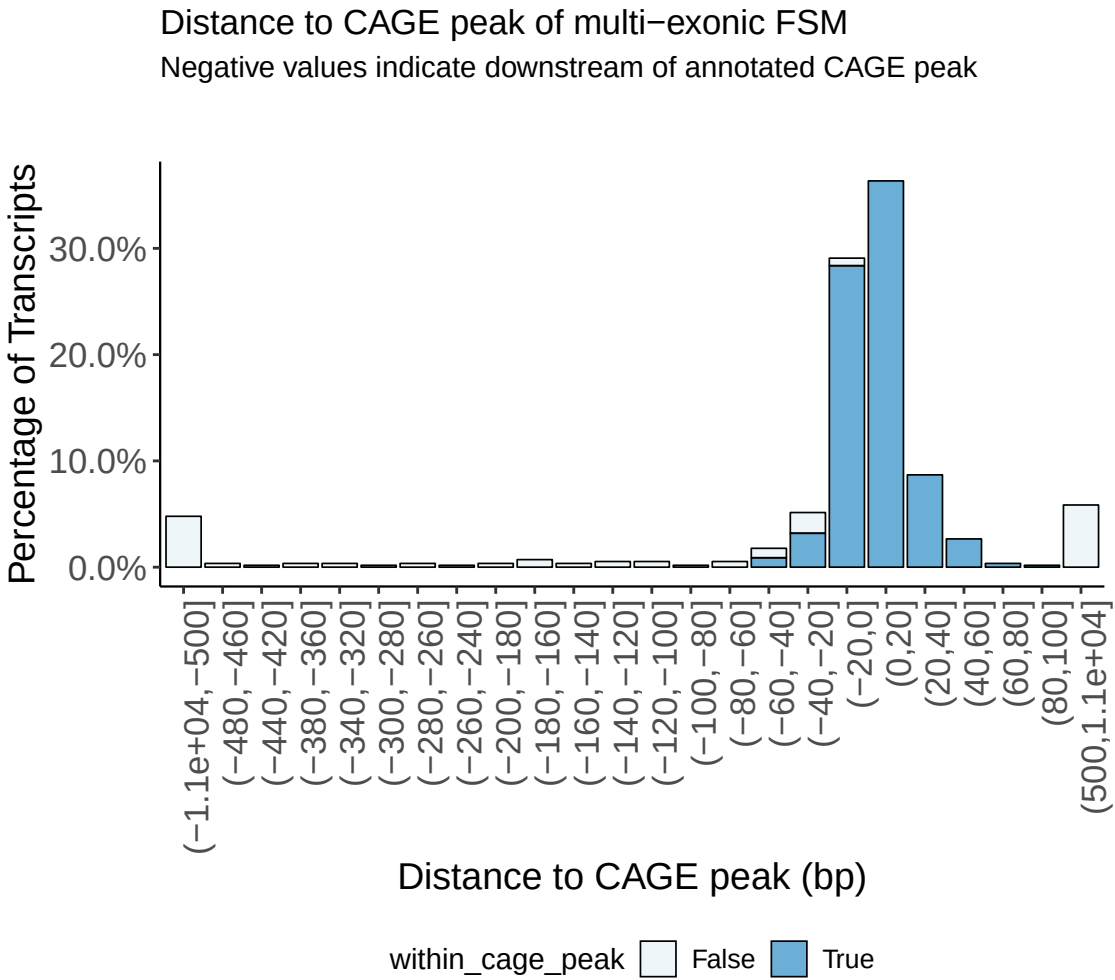
Category	Count	polyA Detected	%
FSM	576	492	85
ISM	156	114	73
NIC	439	358	82
NNC	395	362	92
Genic Genomic	11	6	55
Antisense	9	5	56
Fusion	6	6	100
Intergenic	17	8	47

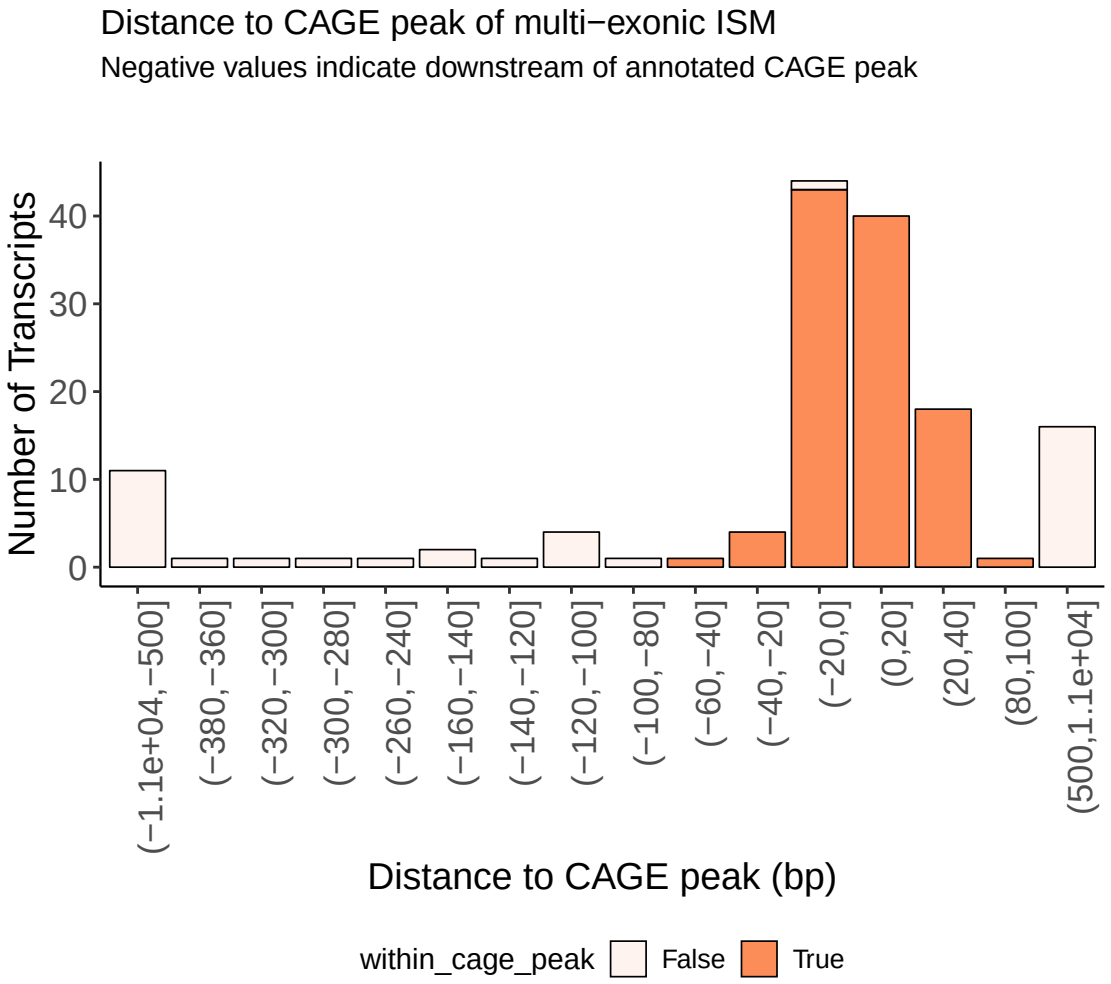
Frequency of polyA motifs

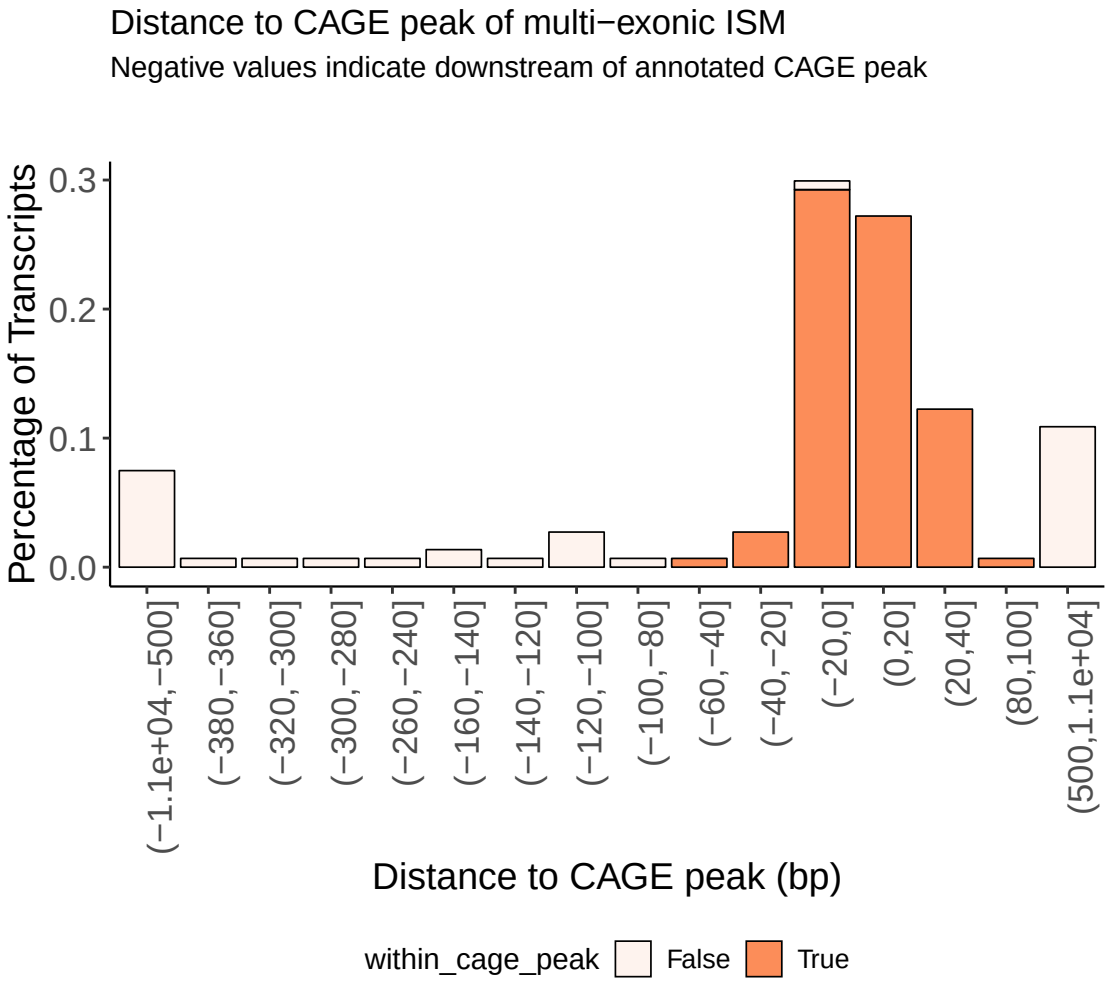
Motif	Count	%
AATAAA	844	62.5
ATTAAA	255	18.9
AATACA	38	2.8
TATAAA	34	2.5
AGTAAA	29	2.1
TTTAAA	28	2.1
AAGAAA	25	1.9
AAAAAG	20	1.5
AATATA	18	1.3
CATAAA	18	1.3
GATAAA	12	0.9
AATGAA	9	0.7
ACTAAA	8	0.6
AATAGA	7	0.5
AAAACA	6	0.4

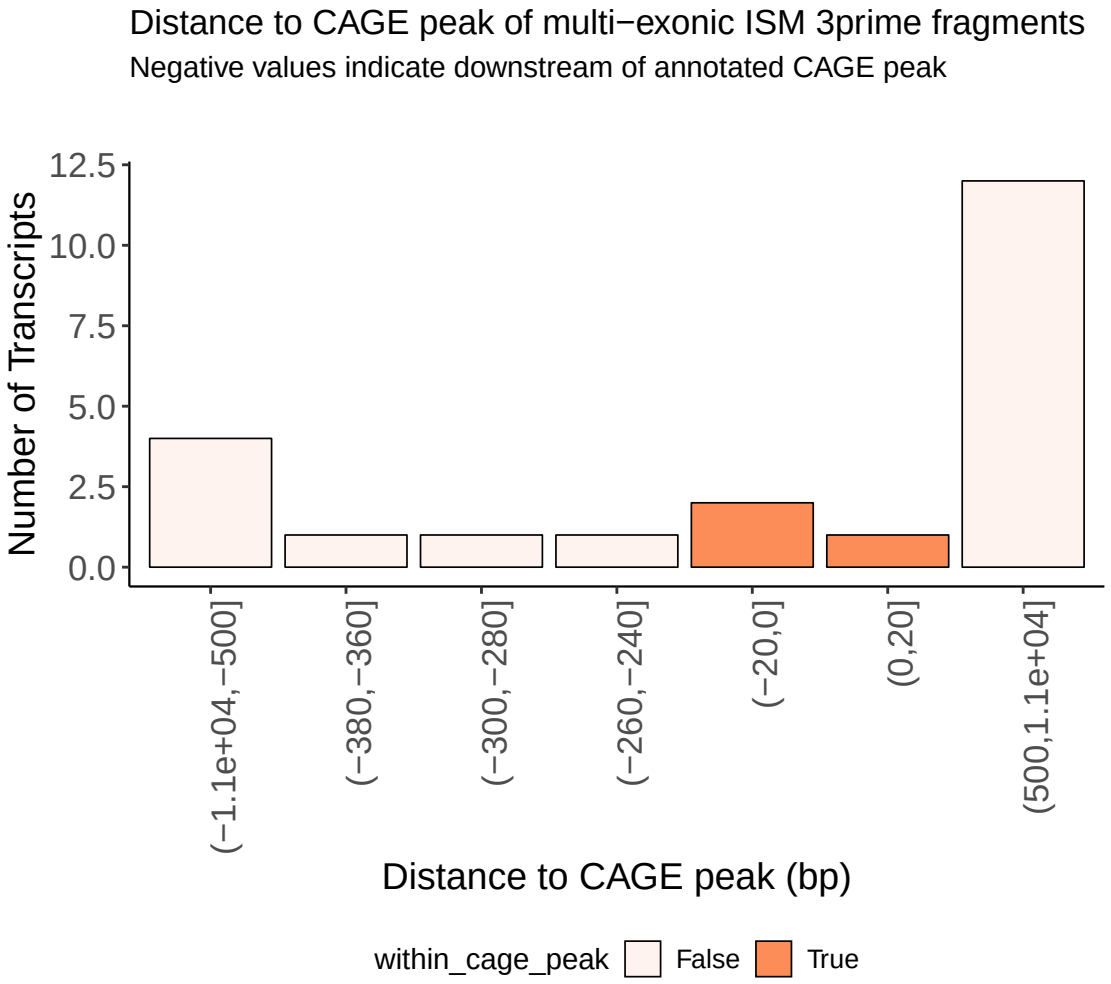
CAGE distances analysis

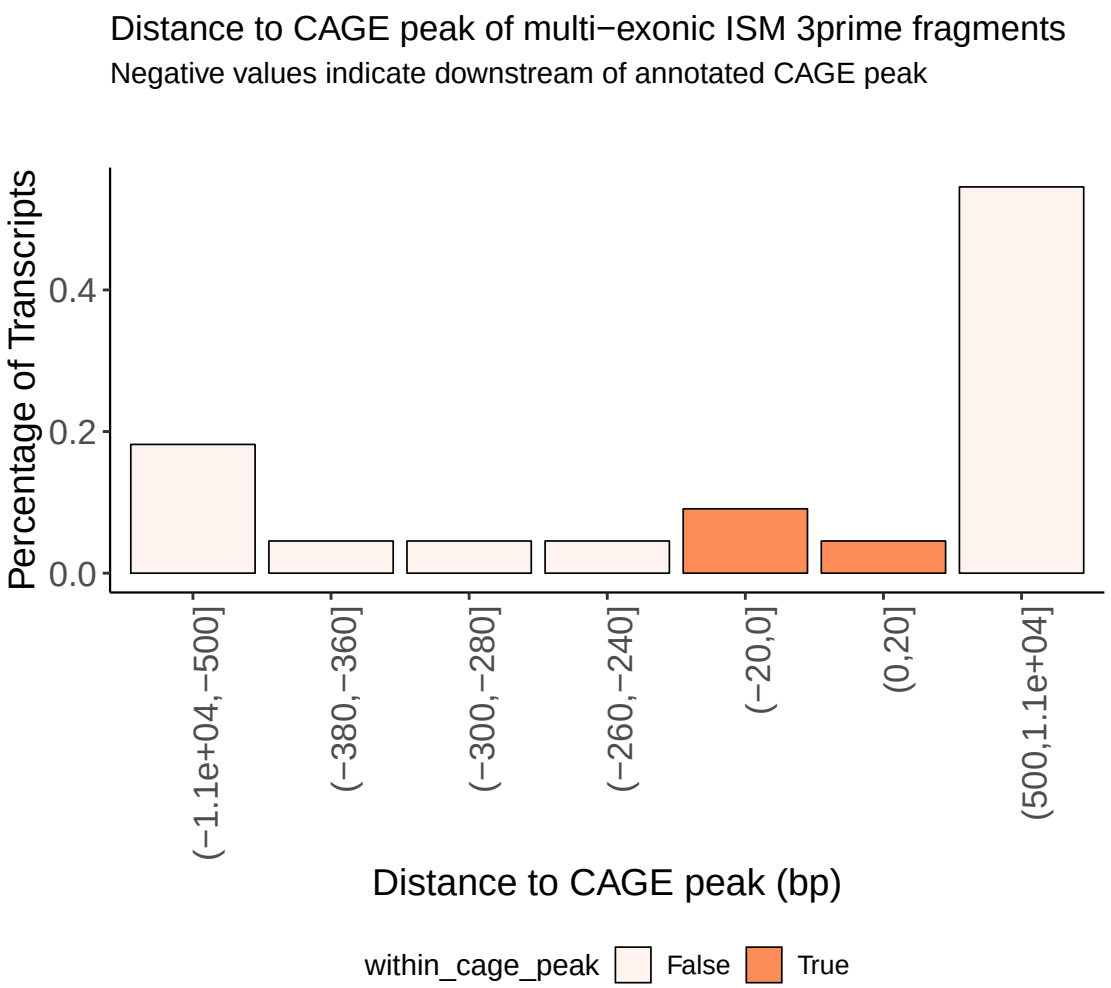

Distance to CAGE peak of multi−exonic FSM
Negative values indicate downstream of annotated CAGE peak
Number of Transcripts
200
150
100
50
0
(−1.1e+04,−500]
(−480,−460]
(−440,−420]
(−380,−360]
(−340,−320]
(−300,−280]
(−280,−260]
(−260,−240]
(−200,−180]
(−180,−160]
(−160,−140]
(−140,−120]
(−120,−100]
(−100,−80]
(−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

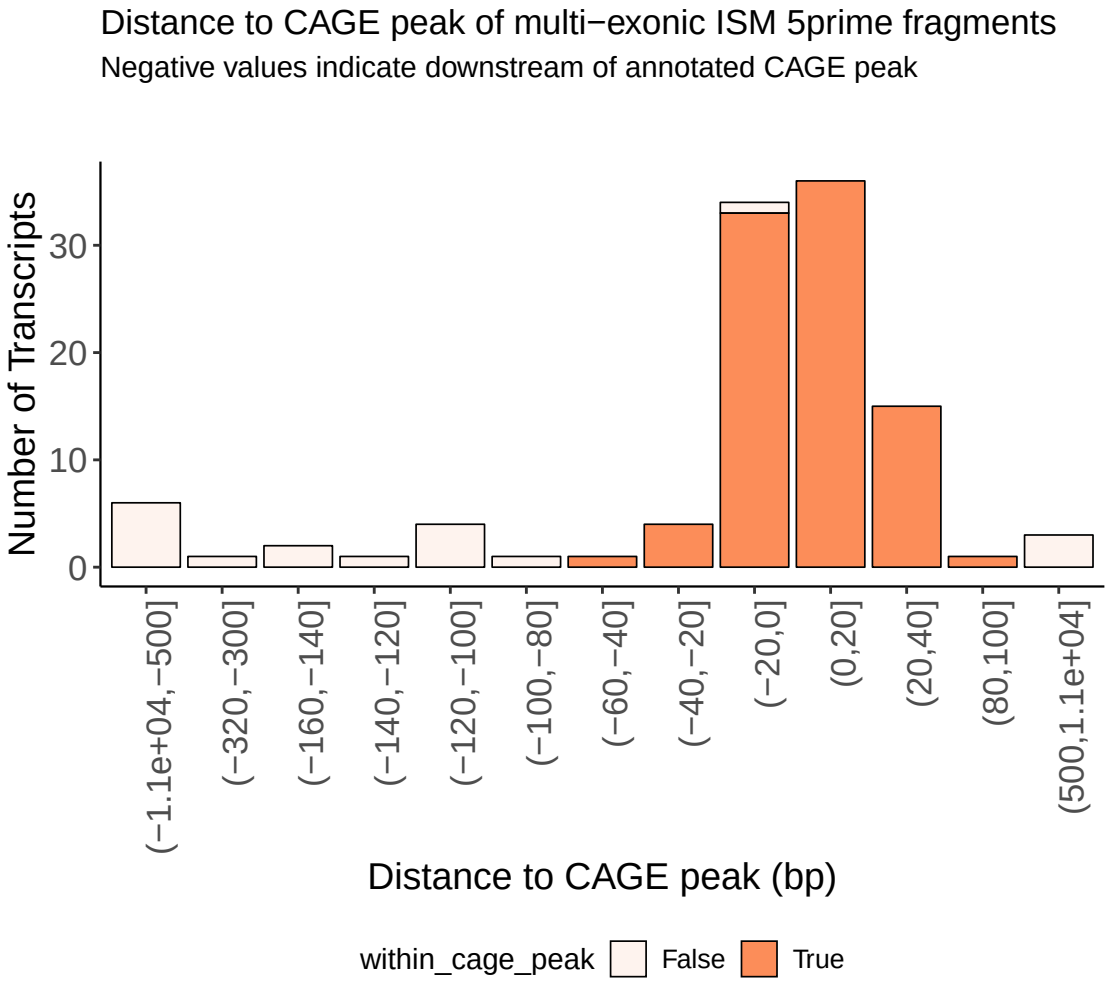

Distance to CAGE peak of multi−exonic FSM
Negative values indicate downstream of annotated CAGE peak
Percentage of Transcripts
30.0%
20.0%
10.0%
0.0%
(−1.1e+04,−500]
(−480,−460]
(−440,−420]
(−380,−360]
(−340,−320]
(−300,−280]
(−280,−260]
(−260,−240]
(−200,−180]
(−180,−160]
(−160,−140]
(−140,−120]
(−120,−100]
(−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

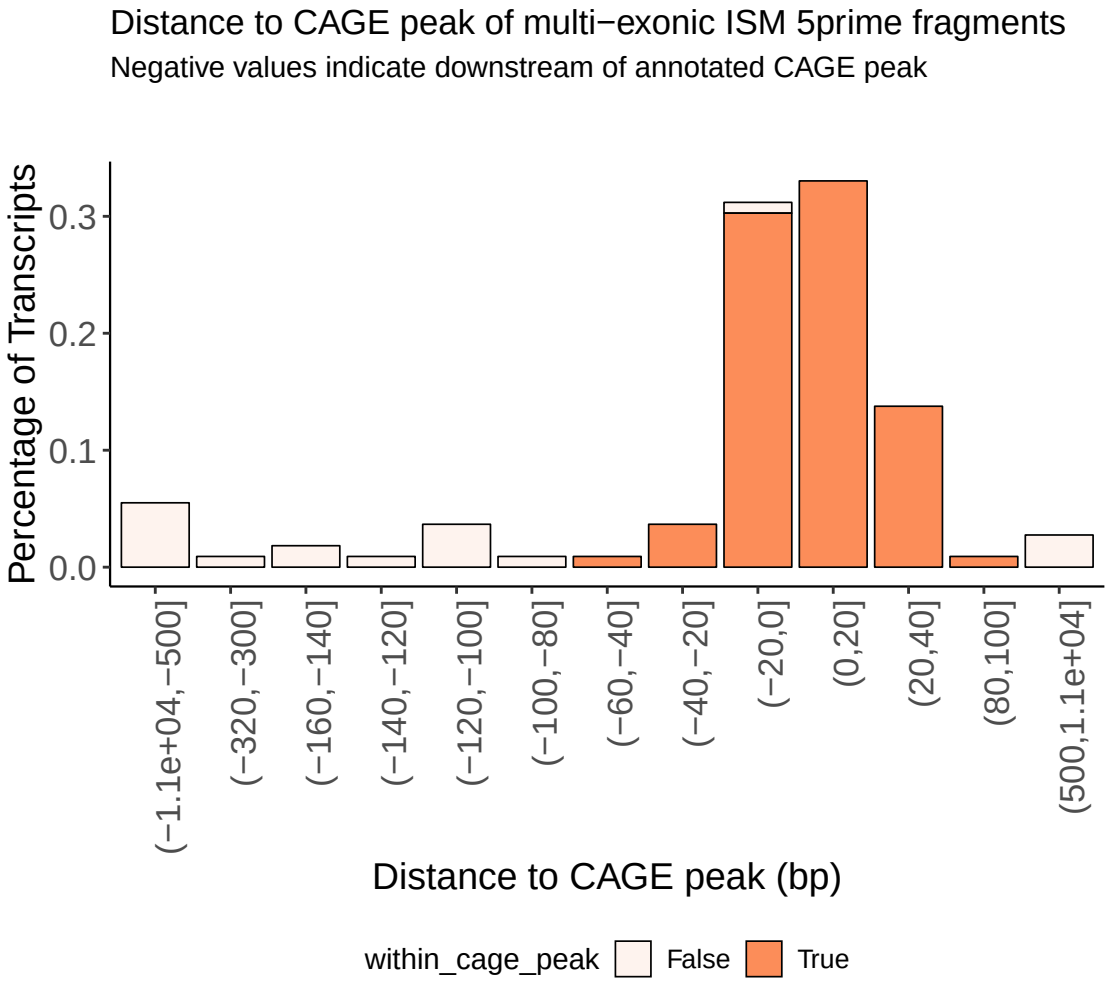

Distance to CAGE peak of multi−exonic ISM
Negative values indicate downstream of annotated CAGE peak
Number of Transcripts
0
10
20
30
40
(−1.1e+04,−500]
(−380,−360]
(−320,−300]
(−300,−280]
(−260,−240]
(−160,−140]
(−140,−120]
(−120,−100]
(−100,−80]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(80,100]
(500,1.1e+04]
Distance to CAGE peak (bp)
within_cage_peak
False
True

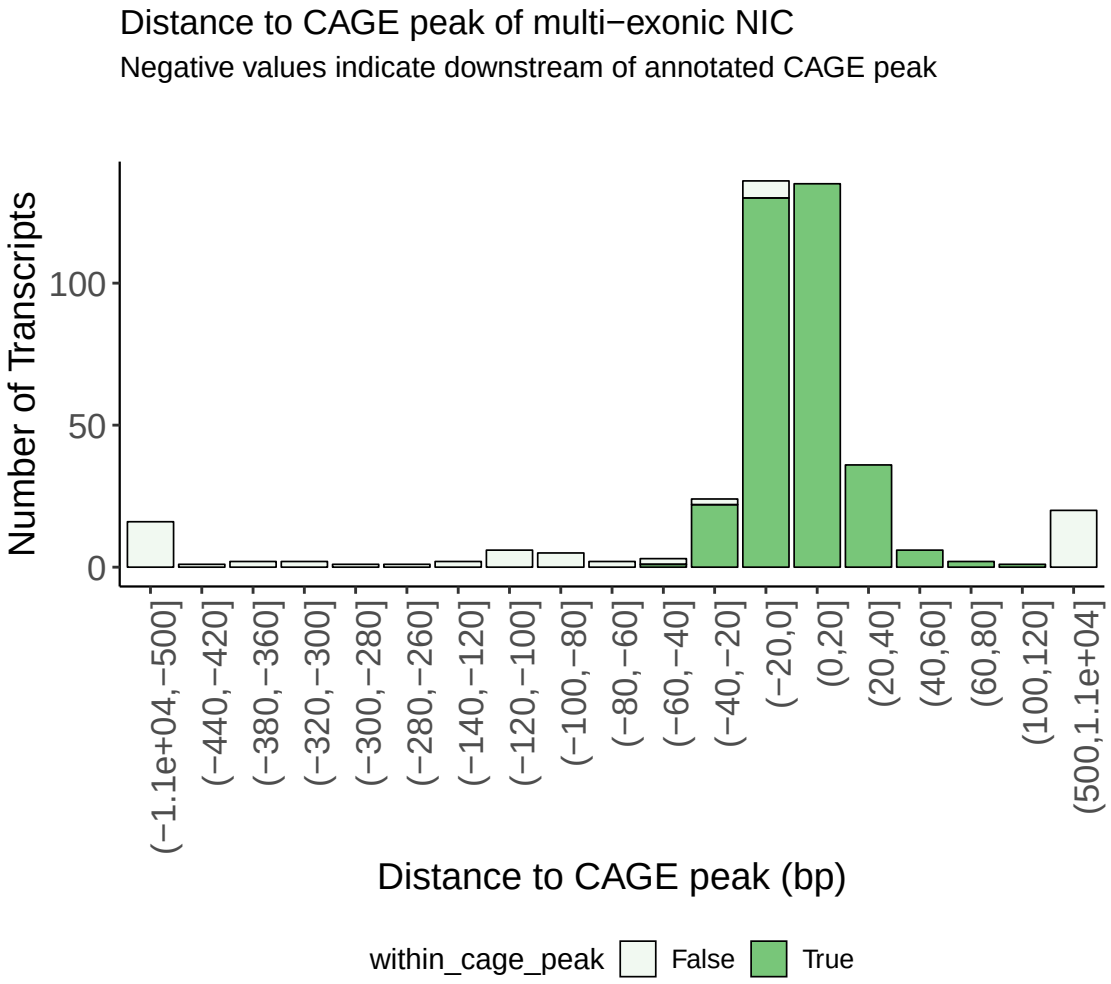

Distance to CAGE peak of multi−exonic ISM
Negative values indicate downstream of annotated CAGE peak
Percentage of Transcripts
0.3
0.2
0.1
0.0
(−1.1e+04,−500]
(−380,−360]
(−320,−300]
(−300,−280]
(−260,−240]
(−160,−140]
(−140,−120]
(−120,−100]
(−100,−80]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(80,100]
(500,1.1e+04]
Distance to CAGE peak (bp)
within_cage_peak
False
True

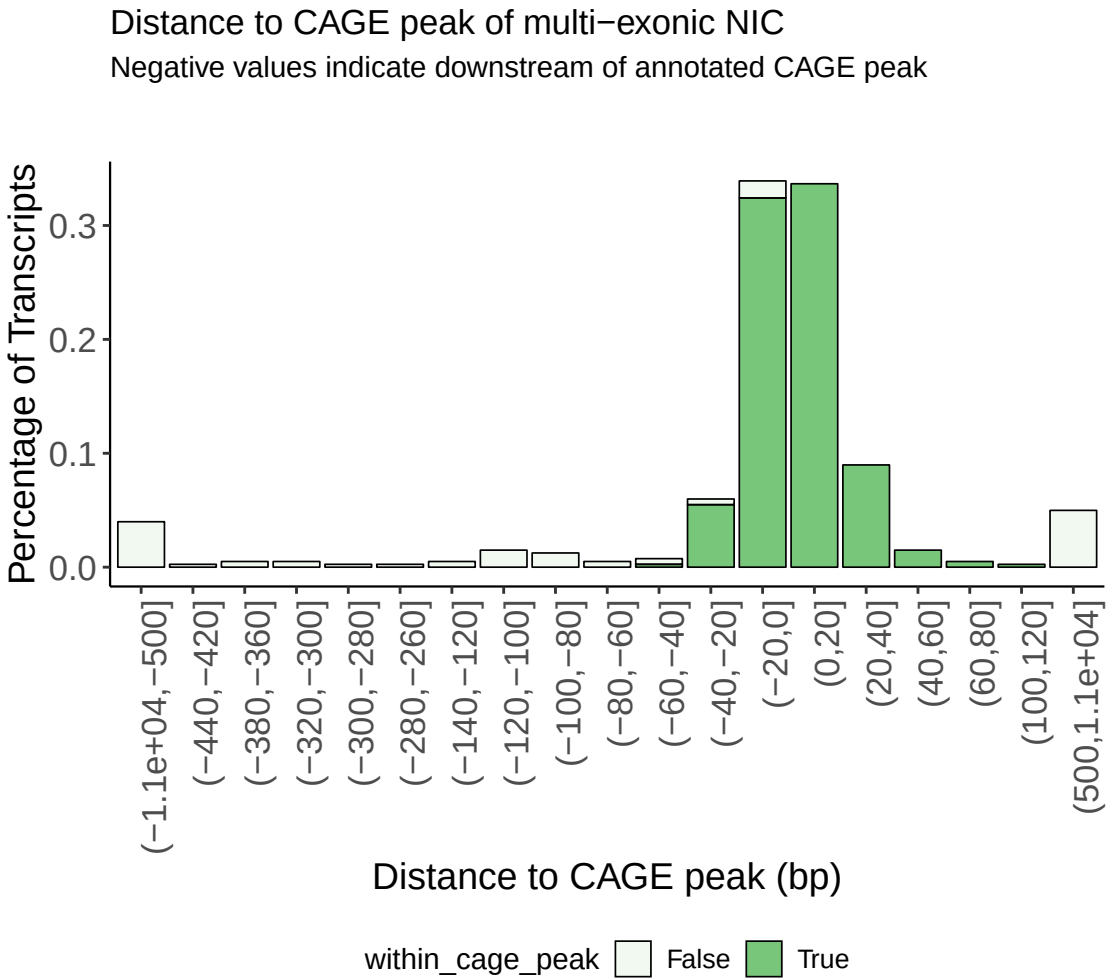

Distance to CAGE peak of multi−exonic ISM 3prime fragments
Negative values indicate downstream of annotated CAGE peak
Number of Transcripts
12.5
10.0
7.5
5.0
2.5
0.0
(−1.1e+04,−500]
(−380,−360]
(−300,−280]
(−260,−240]
(−20,0]
(0,20]
(500,1.1e+04]
Distance to CAGE peak (bp)
within_cage_peak
False
True

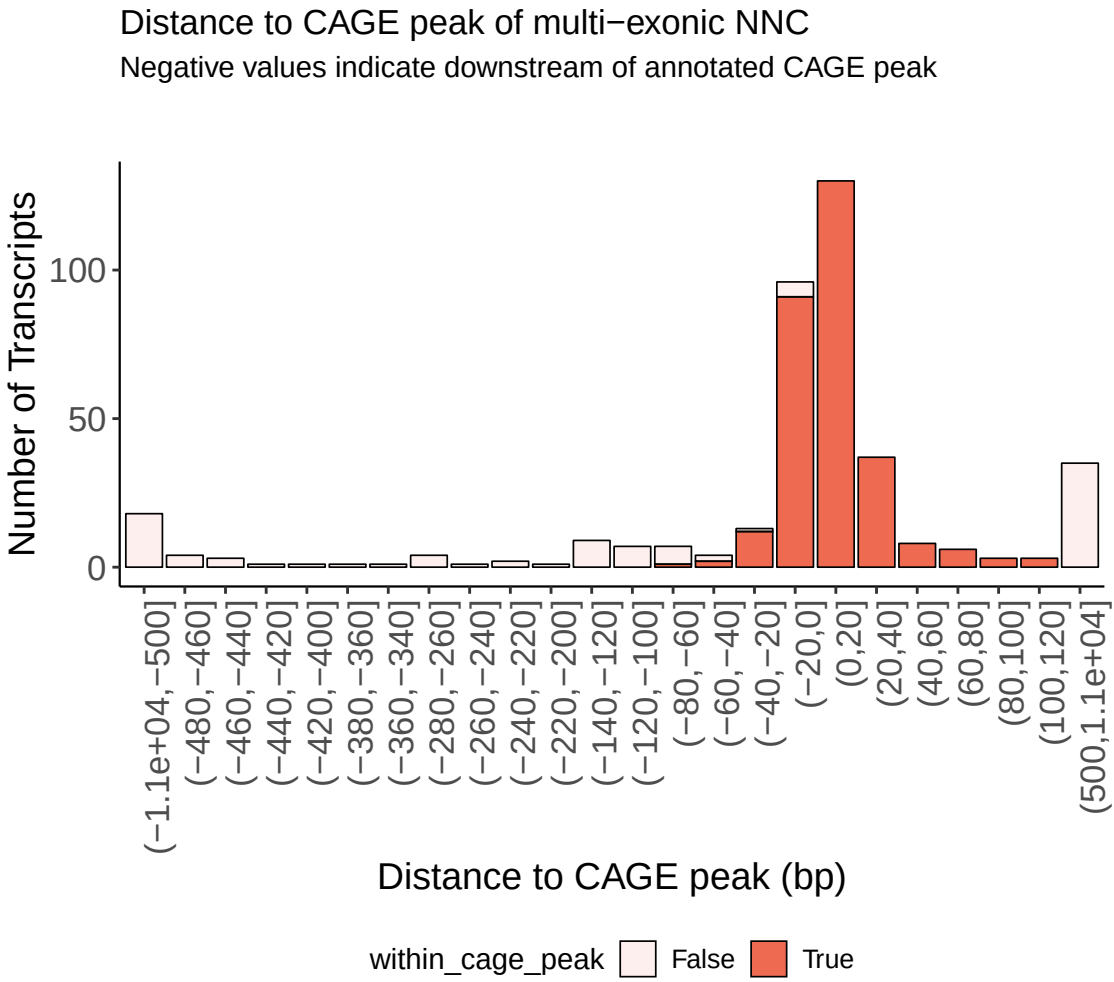

Distance to CAGE peak of multi−exonic ISM 3prime fragments
Negative values indicate downstream of annotated CAGE peak
Percentage of Transcripts
0.0
0.2
0.4
(−1.1e+04,−500]
(−380,−360]
(−300,−280]
(−260,−240]
(−20,0]
(0,20]
(500,1.1e+04]
Distance to CAGE peak (bp)
within_cage_peak
False
True

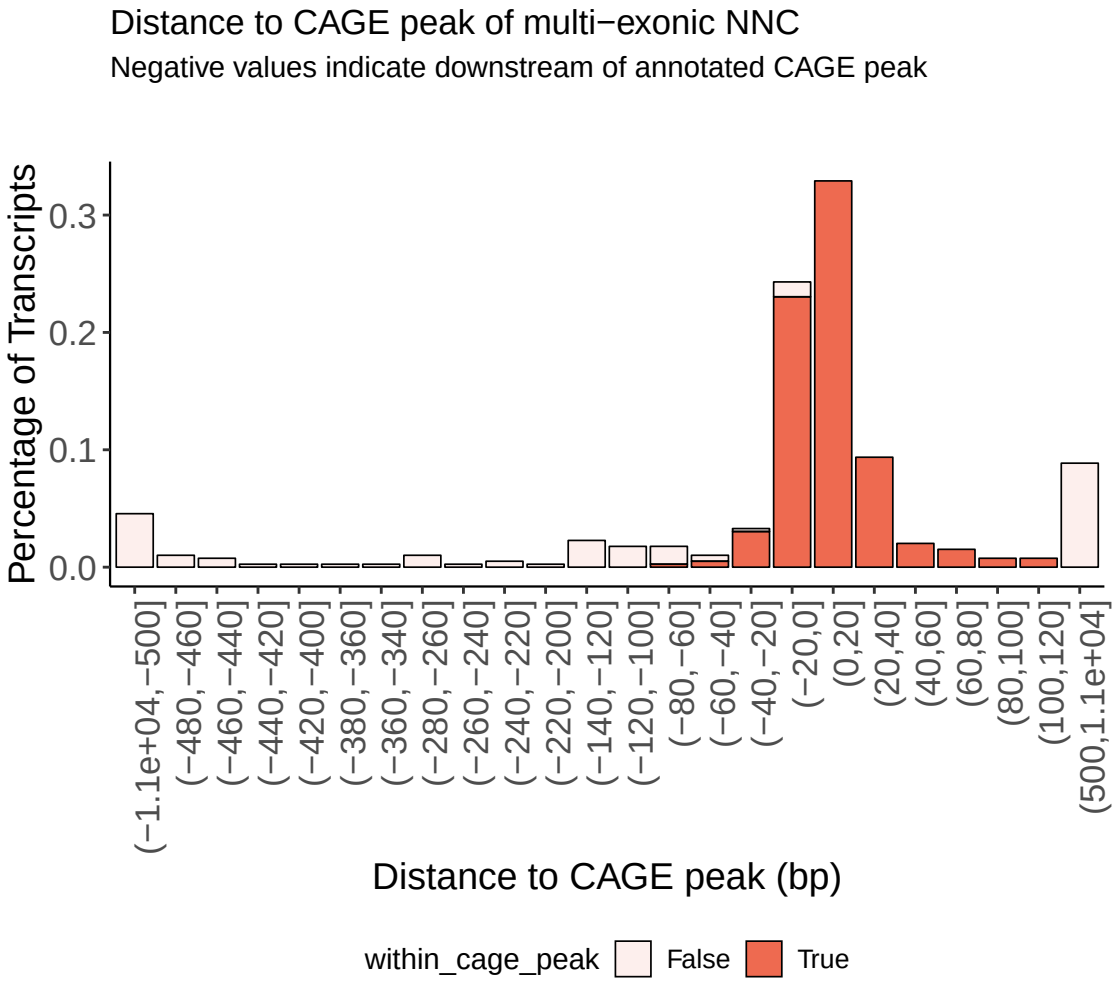

Distance to CAGE peak of multi−exonic ISM 5prime fragments
Negative values indicate downstream of annotated CAGE peak
Number of Transcripts
0
10
20
30
(−1.1e+04,−500]
(−320,−300]
(−160,−140]
(−140,−120]
(−120,−100]
(−100,−80]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(80,100]
(500,1.1e+04]
Distance 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
Percentage of Transcripts
0.0
0.1
0.2
0.3
(−1.1e+04,−500]
(−320,−300]
(−160,−140]
(−140,−120]
(−120,−100]
(−100,−80]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(80,100]
(500,1.1e+04]
Distance to CAGE peak (bp)
within_cage_peak
False
True

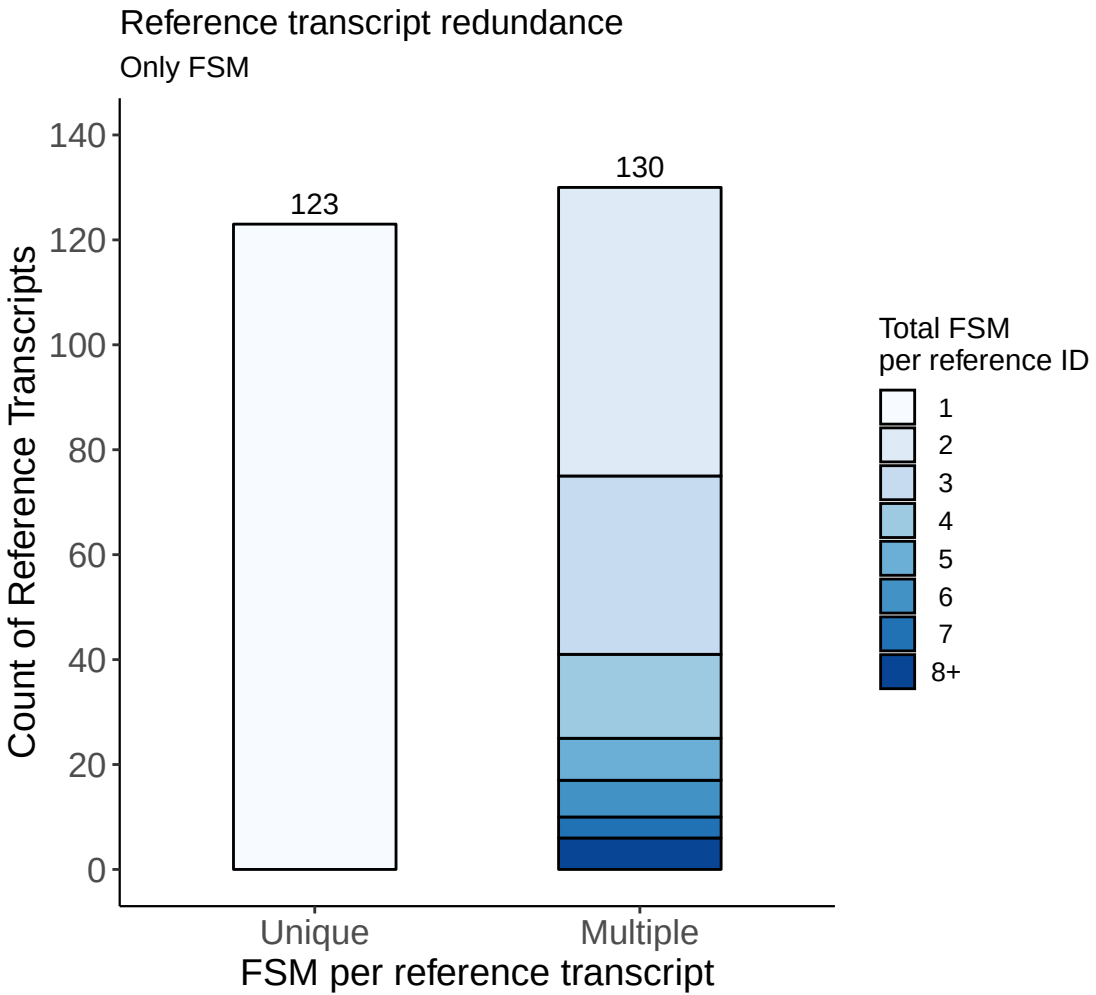

Distance to CAGE peak of multi−exonic NIC
Negative values indicate downstream of annotated CAGE peak
Number of Transcripts
100
50
0
(−1.1e+04,−500]
(−440,−420]
(−380,−360]
(−320,−300]
(−300,−280]
(−280,−260]
(−140,−120]
(−120,−100]
(−100,−80]
(−80,−60]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(40,60]
(60,80]
(100,120]
(500,1.1e+04]
Distance to CAGE peak (bp)
within_cage_peak
False
True

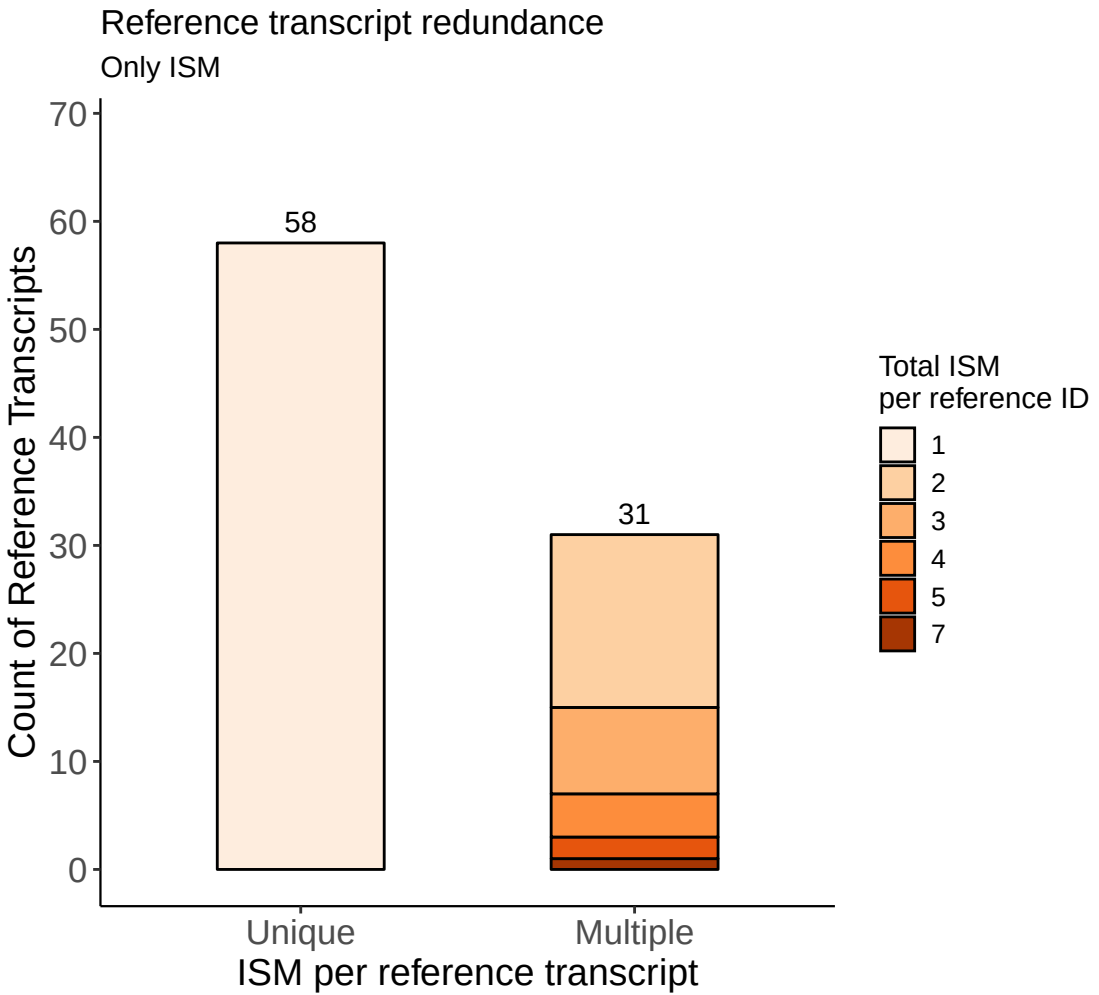

Distance to CAGE peak of multi−exonic NIC
Negative values indicate downstream of annotated CAGE peak
Percentage of Transcripts
0.0
0.1
0.2
0.3
(−1.1e+04,−500]
(−440,−420]
(−380,−360]
(−320,−300]
(−300,−280]
(−280,−260]
(−140,−120]
(−120,−100]
(−100,−80]
(−80,−60]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(40,60]
(60,80]
(100,120]
(500,1.1e+04]
Distance to CAGE peak (bp)
within_cage_peak
False
True

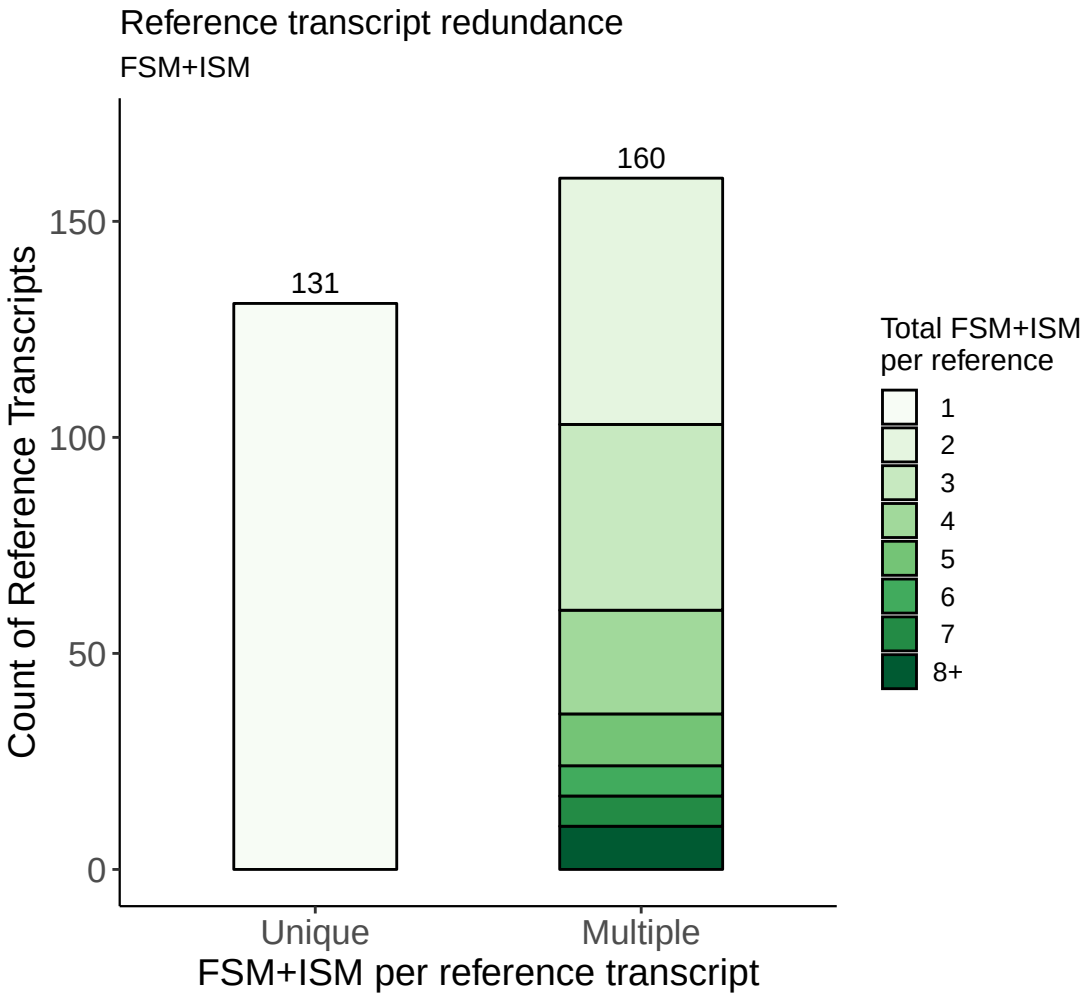

Distance to CAGE peak of multi−exonic NNC
Negative values indicate downstream of annotated CAGE peak
Number of Transcripts
100
50
0
(−1.1e+04,−500]
(−480,−460]
(−460,−440]
(−440,−420]
(−420,−400]
(−380,−360]
(−360,−340]
(−280,−260]
(−260,−240]
(−240,−220]
(−220,−200]
(−140,−120]
(−120,−100]
(−80,−60]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(40,60]
(60,80]
(80,100]
(100,120]
(500,1.1e+04]
Distance to CAGE peak (bp)
within_cage_peak
False
True

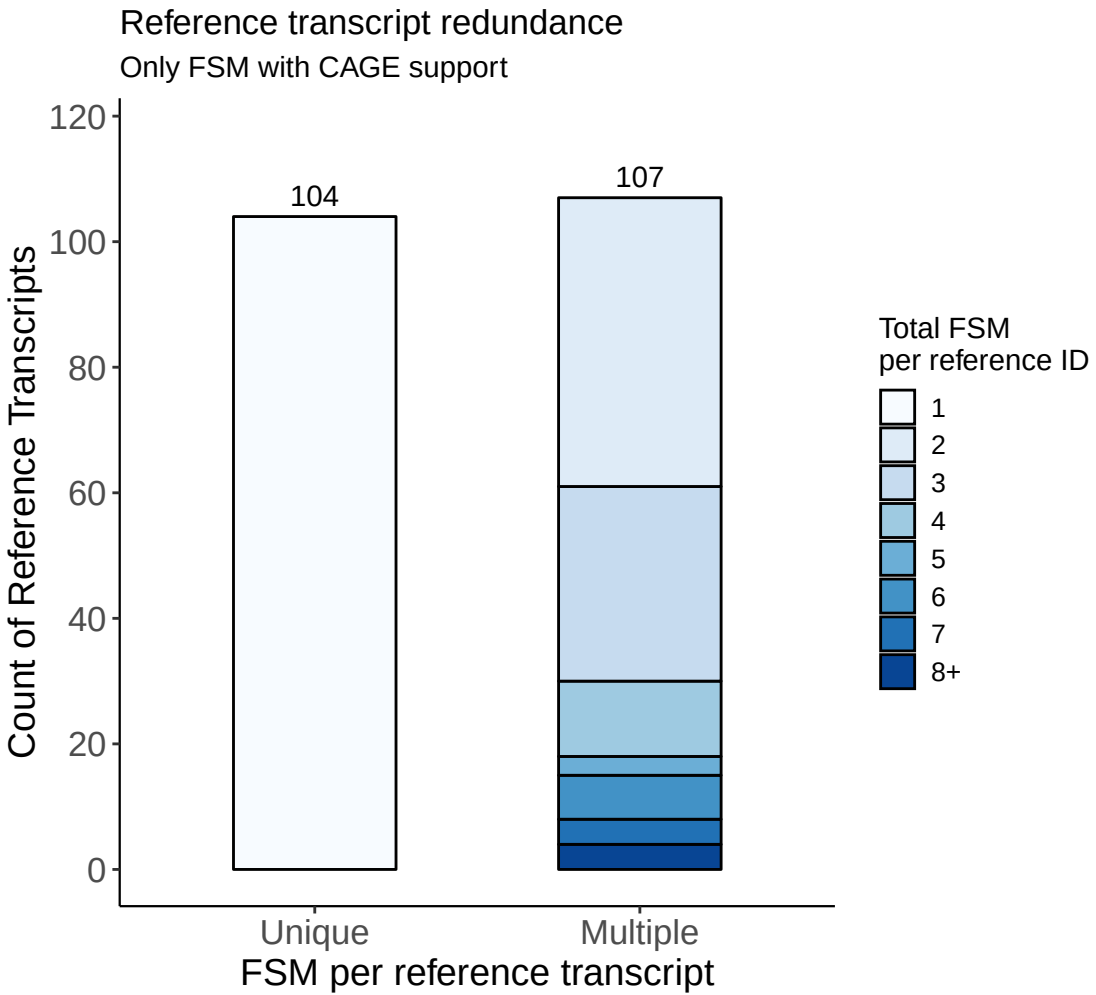

Distance to CAGE peak of multi−exonic NNC
Negative values indicate downstream of annotated CAGE peak
Percentage of Transcripts
0.0
0.1
0.2
0.3
(−1.1e+04,−500]
(−480,−460]
(−460,−440]
(−440,−420]
(−420,−400]
(−380,−360]
(−360,−340]
(−280,−260]
(−260,−240]
(−240,−220]
(−220,−200]
(−140,−120]
(−120,−100]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

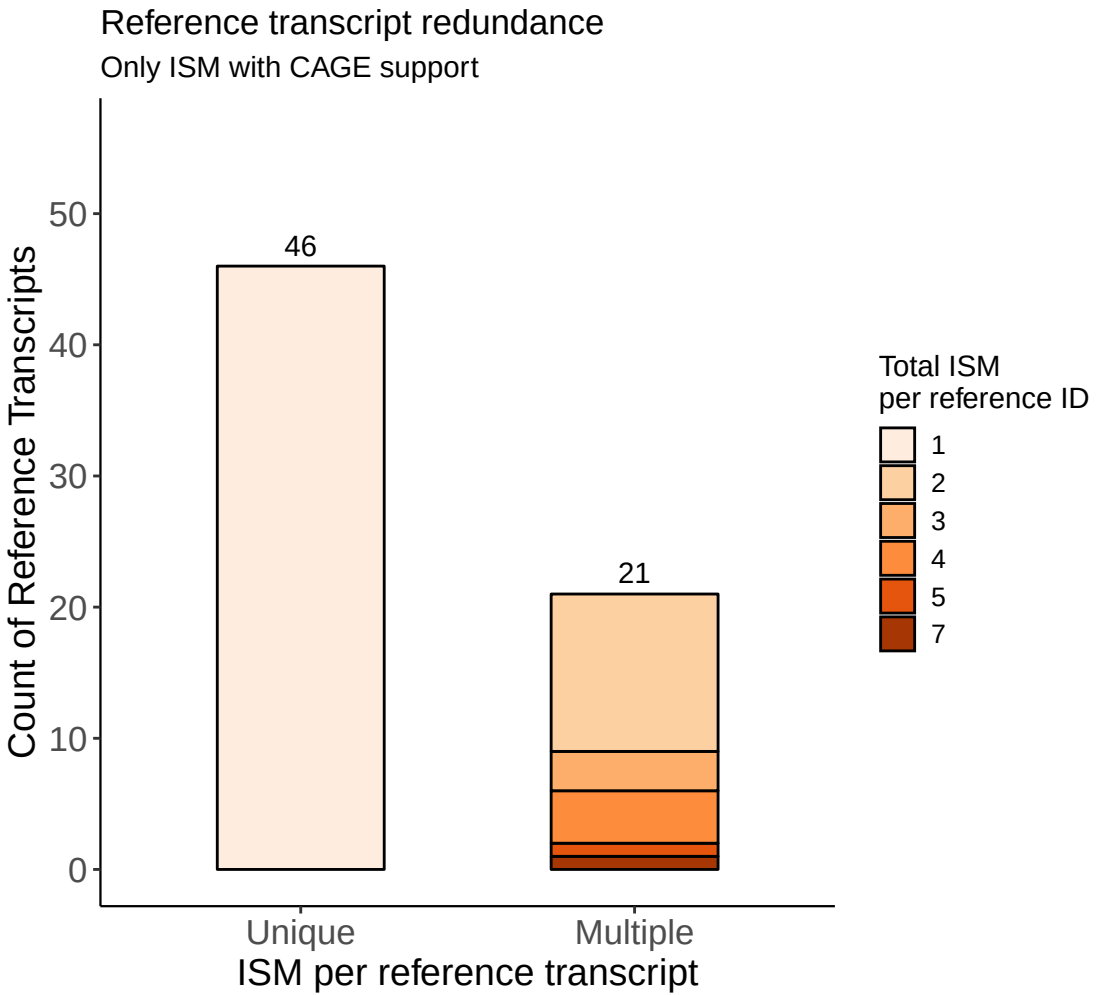
Accumulation of FSM and ISM
to the same reference transcript

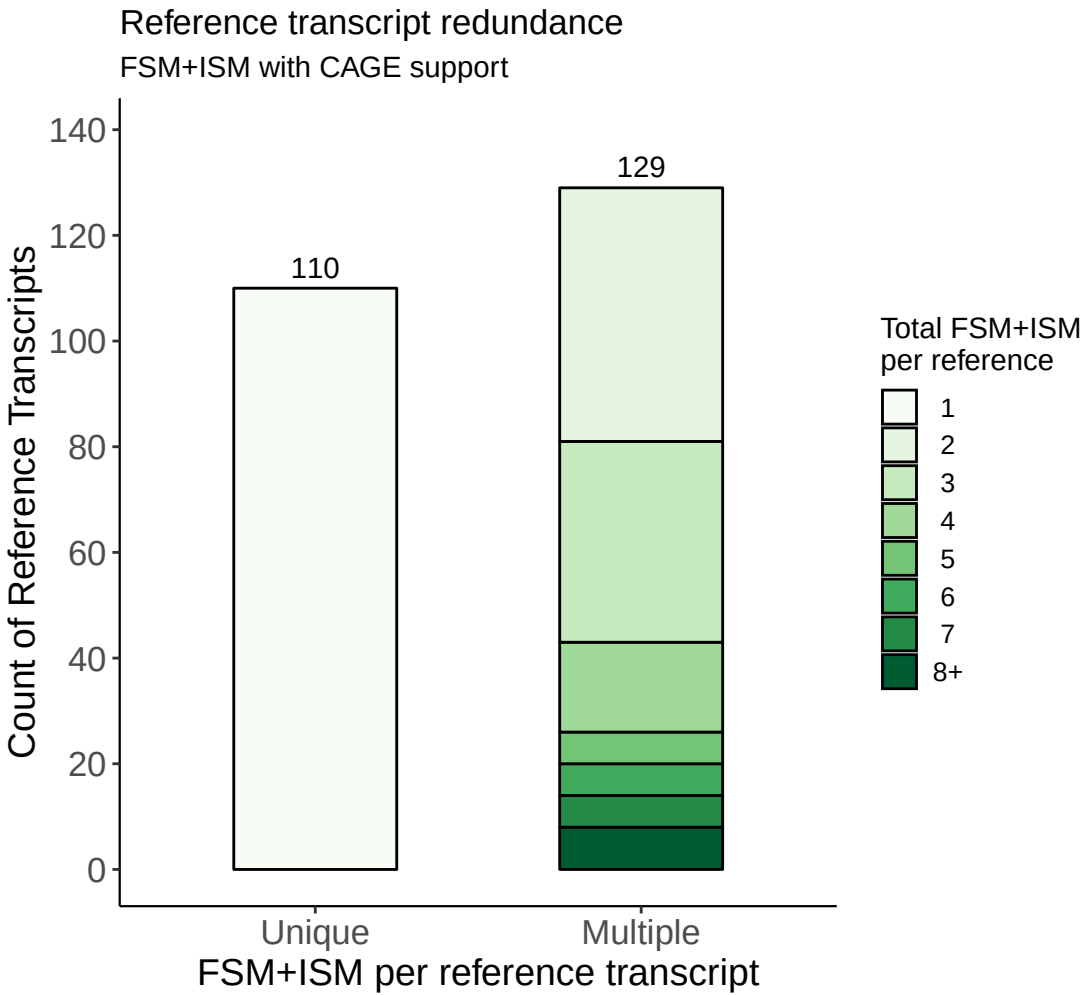

Reference transcript redundance
Only FSM
Count of Reference Transcripts
140
120
100
80
60
40
20
0
123
130
Unique
Multiple
FSM per reference transcript
Total FSM
per reference ID
1
2
3
4
5
6
7
8+

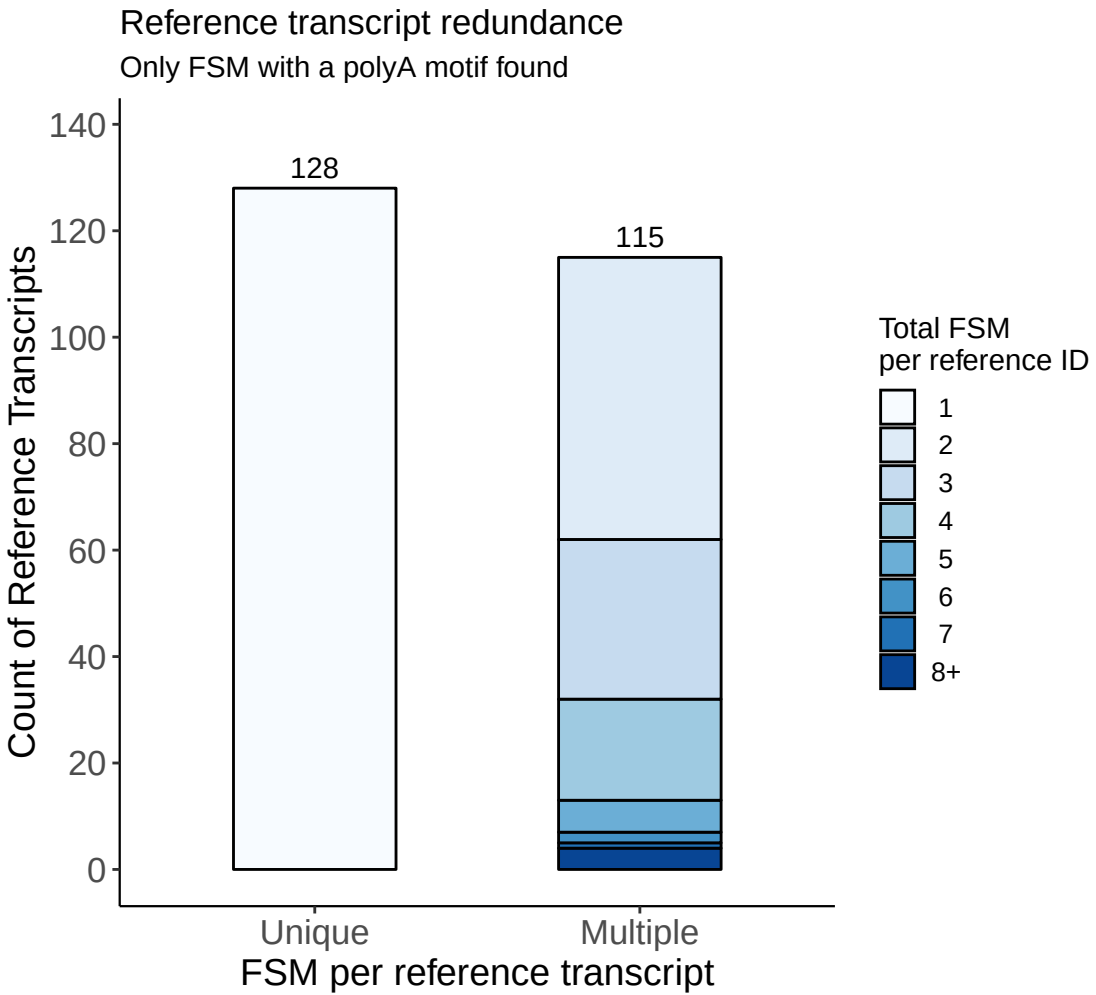

Reference transcript redundance
Only ISM
70
60
50
40
30
20
10
0
Count of Reference Transcripts
58
31
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
5
7

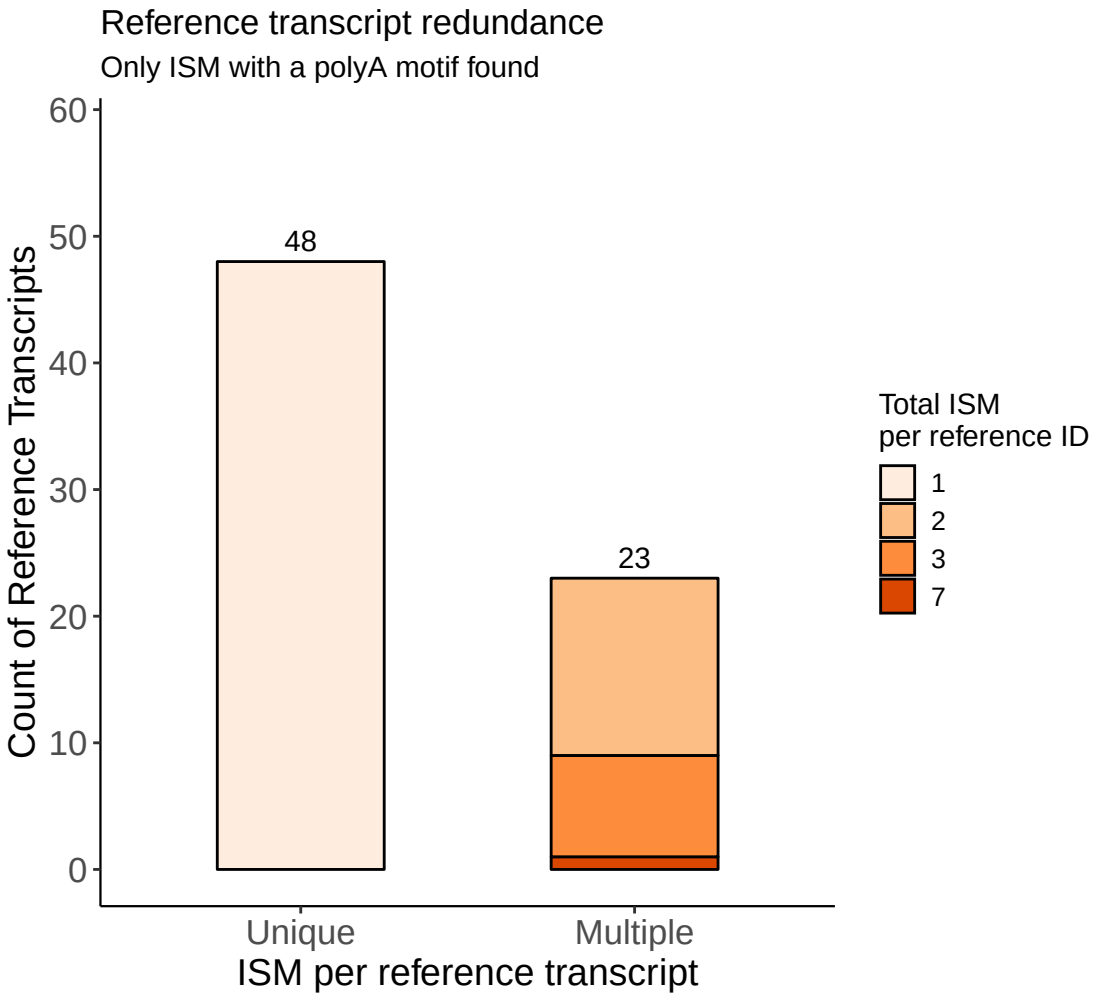

Reference transcript redundance
FSM+ISM
Count of Reference Transcripts
150
100
50
0
131
160
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+

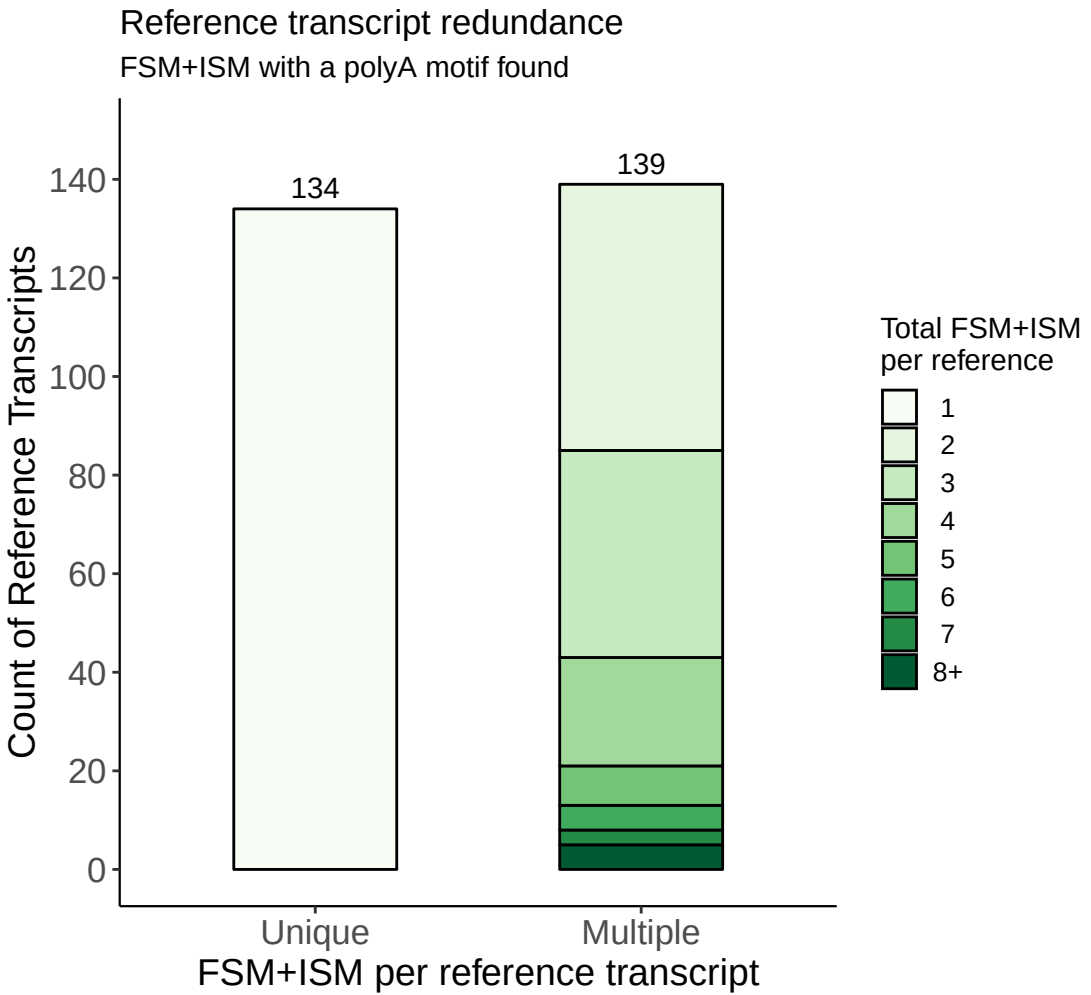

Reference transcript redundance
Only FSM with CAGE support
Count of Reference Transcripts
120
100
80
60
40
20
0
104
107
Unique
Multiple
FSM per reference transcript
Total FSM
per reference ID
1
2
3
4
5
6
7
8+

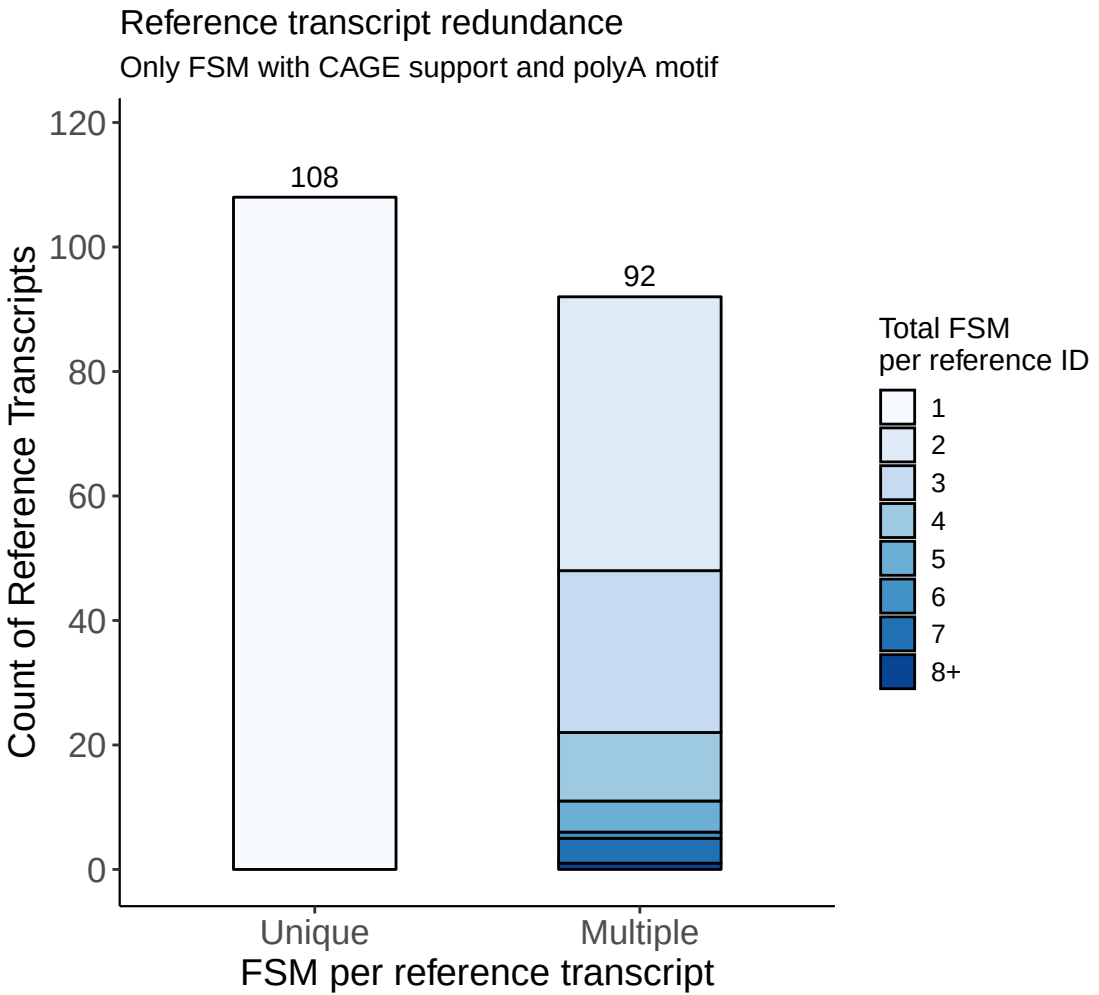

Reference transcript redundance
Only ISM with CAGE support
Count of Reference Transcripts
0
10
20
30
40
50
46
21
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
5
7

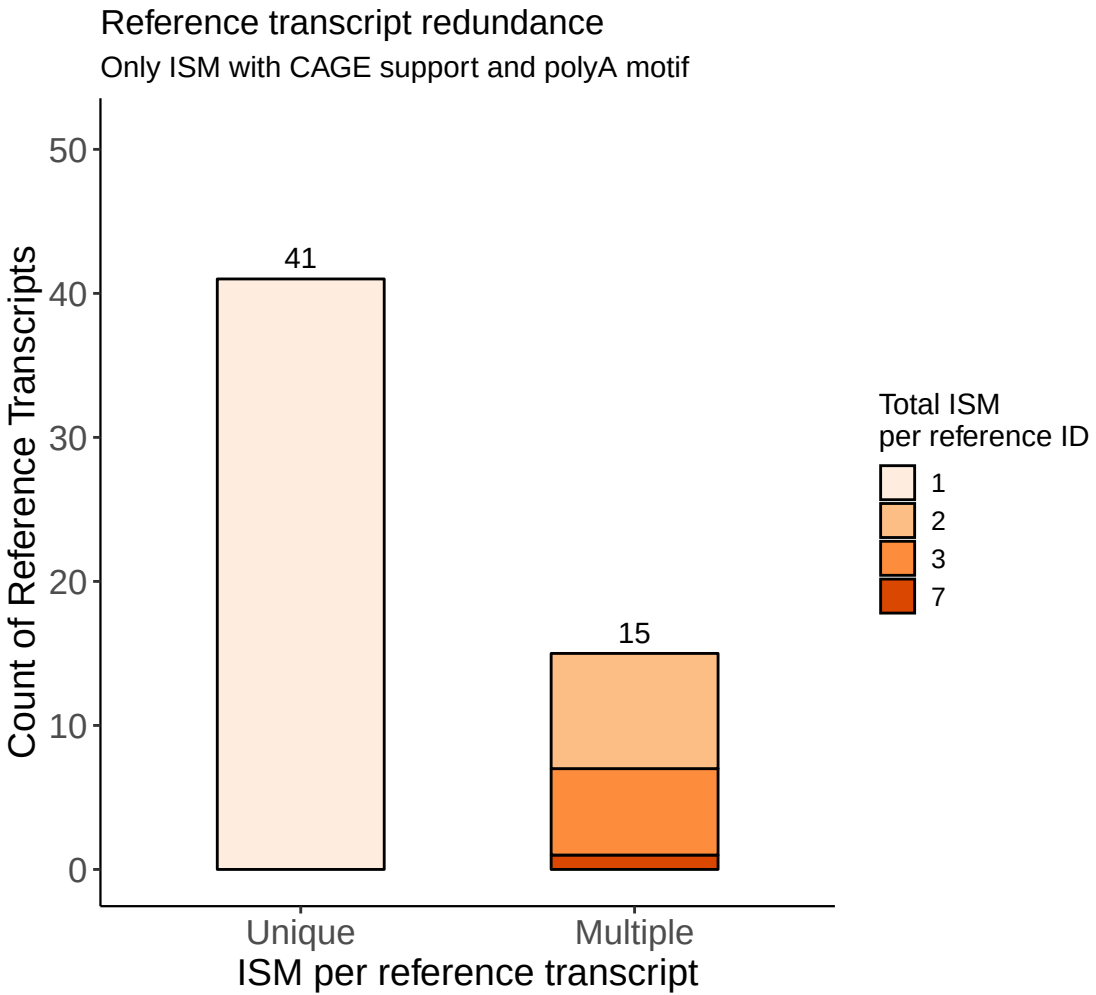

Reference transcript redundance
FSM+ISM with CAGE support
Count of Reference Transcripts
140
120
100
80
60
40
20
0
110
129
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+

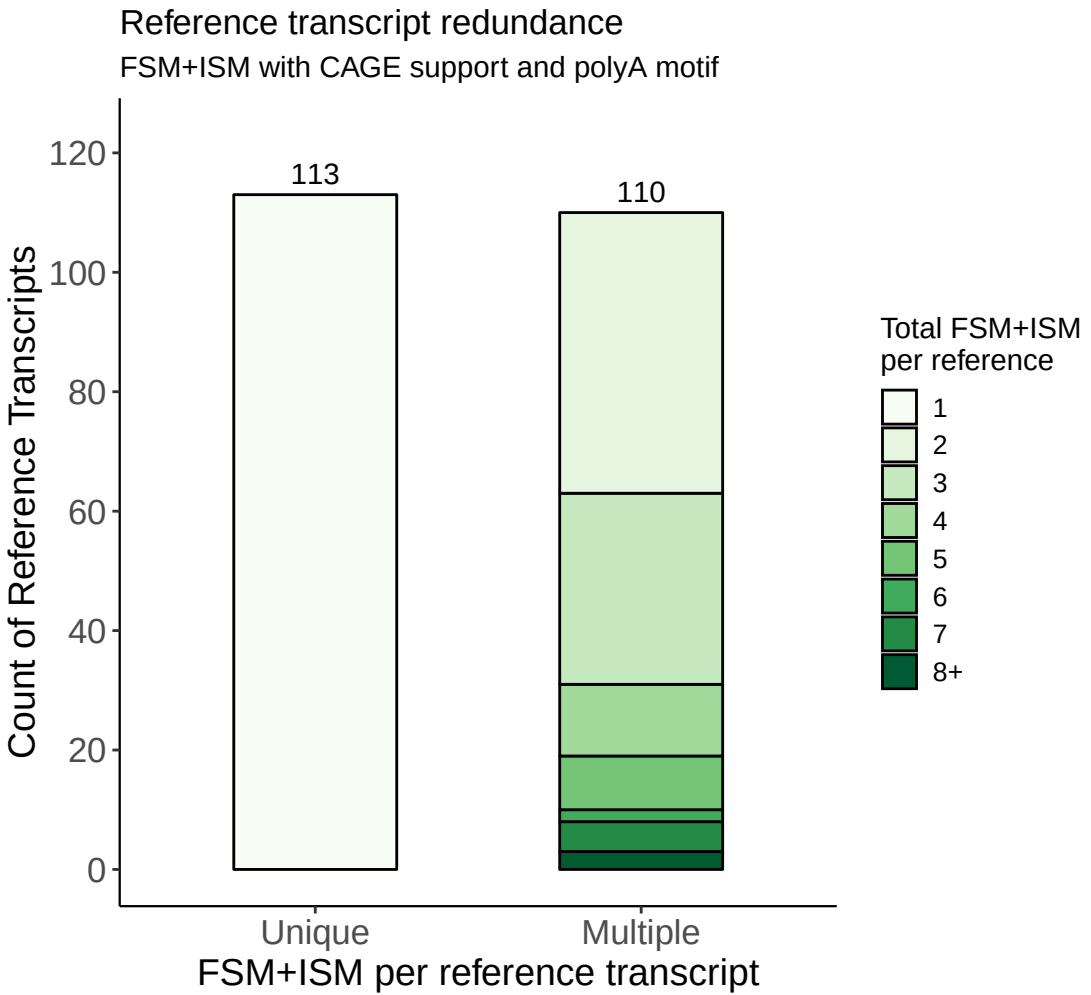

Reference transcript redundance
Only FSM with a polyA motif found
Count of Reference Transcripts
140
120
100
80
60
40
20
0
128
115
Unique
Multiple
FSM per reference transcript
Total FSM
per reference ID
1
2
3
4
5
6
7
8+


Reference transcript redundance
Only ISM with a polyA motif found
Count of Reference Transcripts
60
50
40
30
20
10
0
48
23
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
7

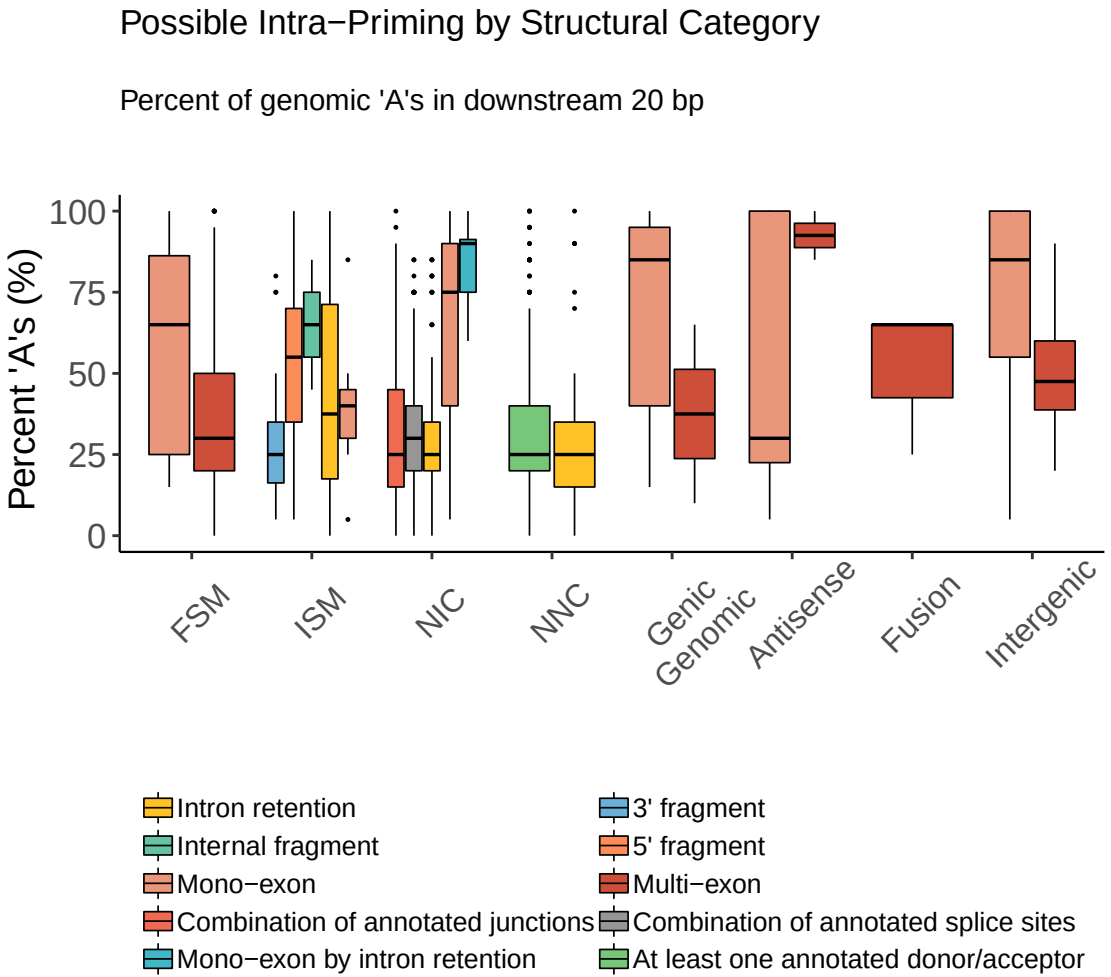

Reference transcript redundance
FSM+ISM with a polyA motif found
Count of Reference Transcripts
140
120
100
80
60
40
20
0
134
139
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+

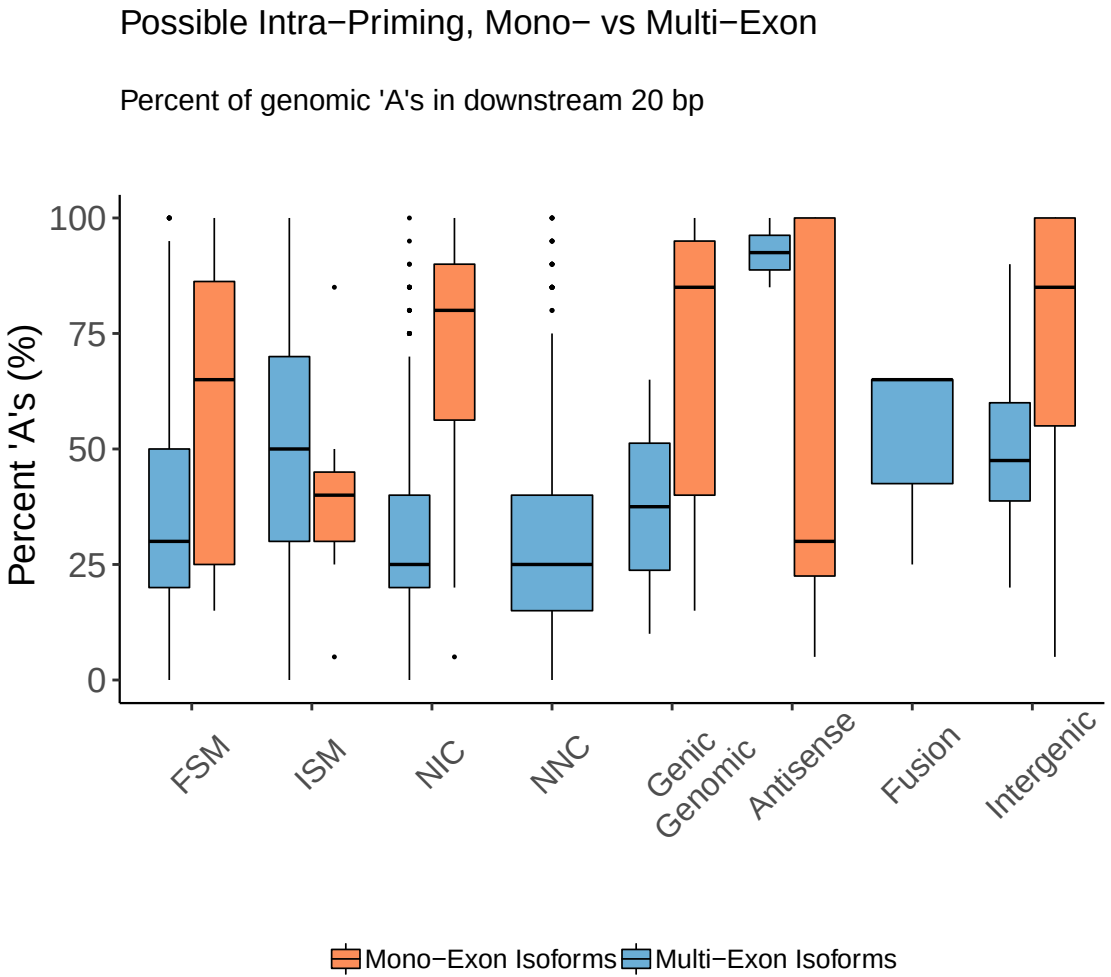

Reference transcript redundance
Only FSM with CAGE support and polyA motif
Count of Reference Transcripts
120
100
80
60
40
20
0
108
92
Unique
Multiple
FSM per reference transcript
Total FSM
per reference ID
1
2
3
4
5
6
7
8+

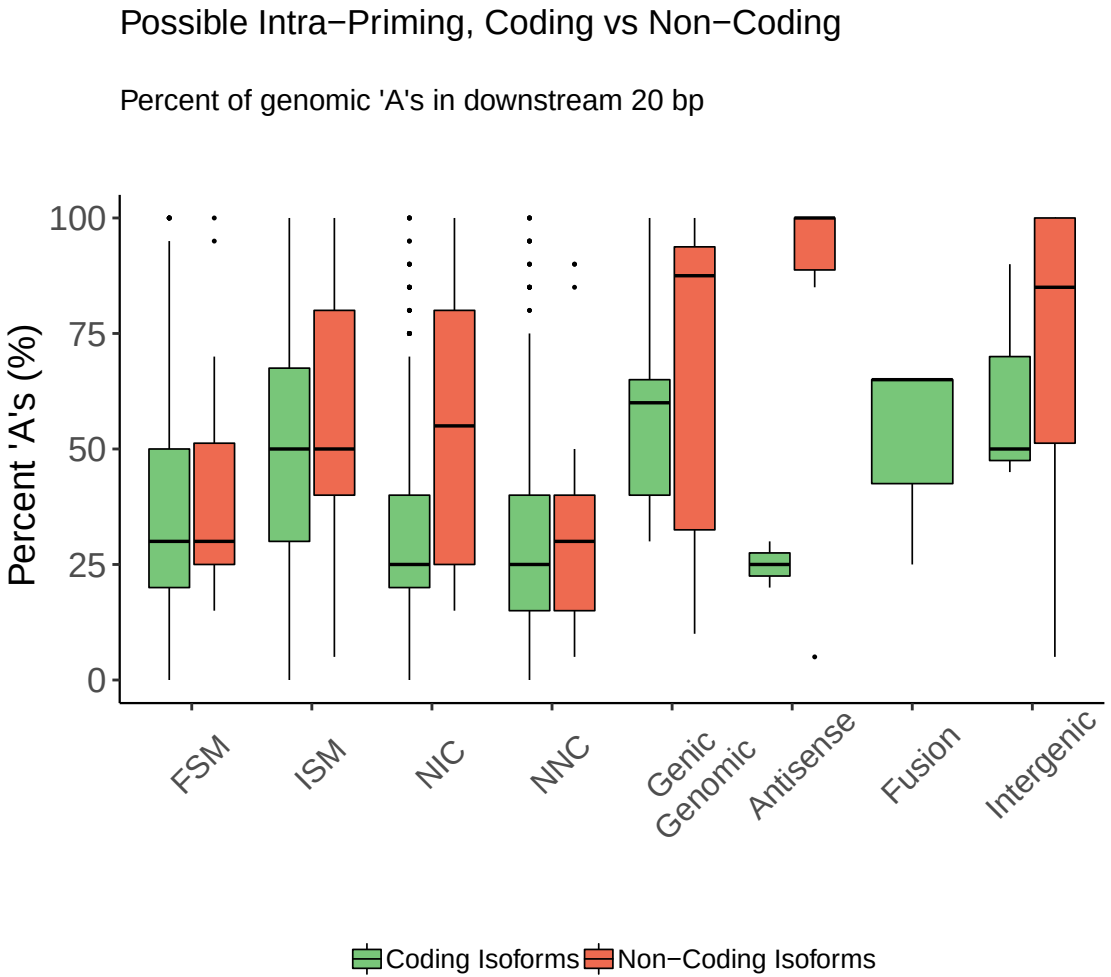

Reference transcript redundance
Only ISM with CAGE support and polyA motif
Count of Reference Transcripts
50
40
30
20
10
0
41
15
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
7


Reference transcript redundance
FSM+ISM with CAGE support and polyA motif
Count of Reference Transcripts
120
100
80
60
40
20
0
113
110
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+

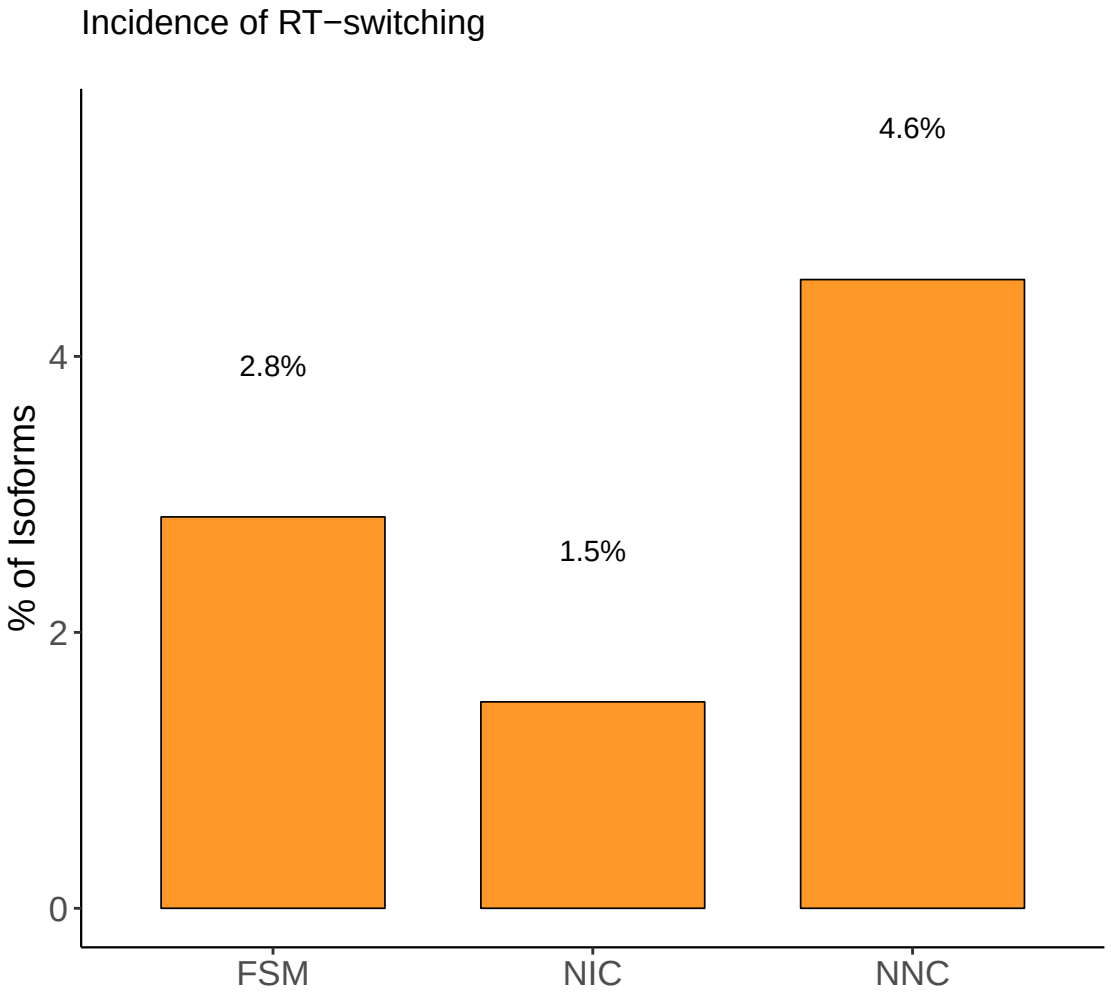
Intra−Priming Quality Check

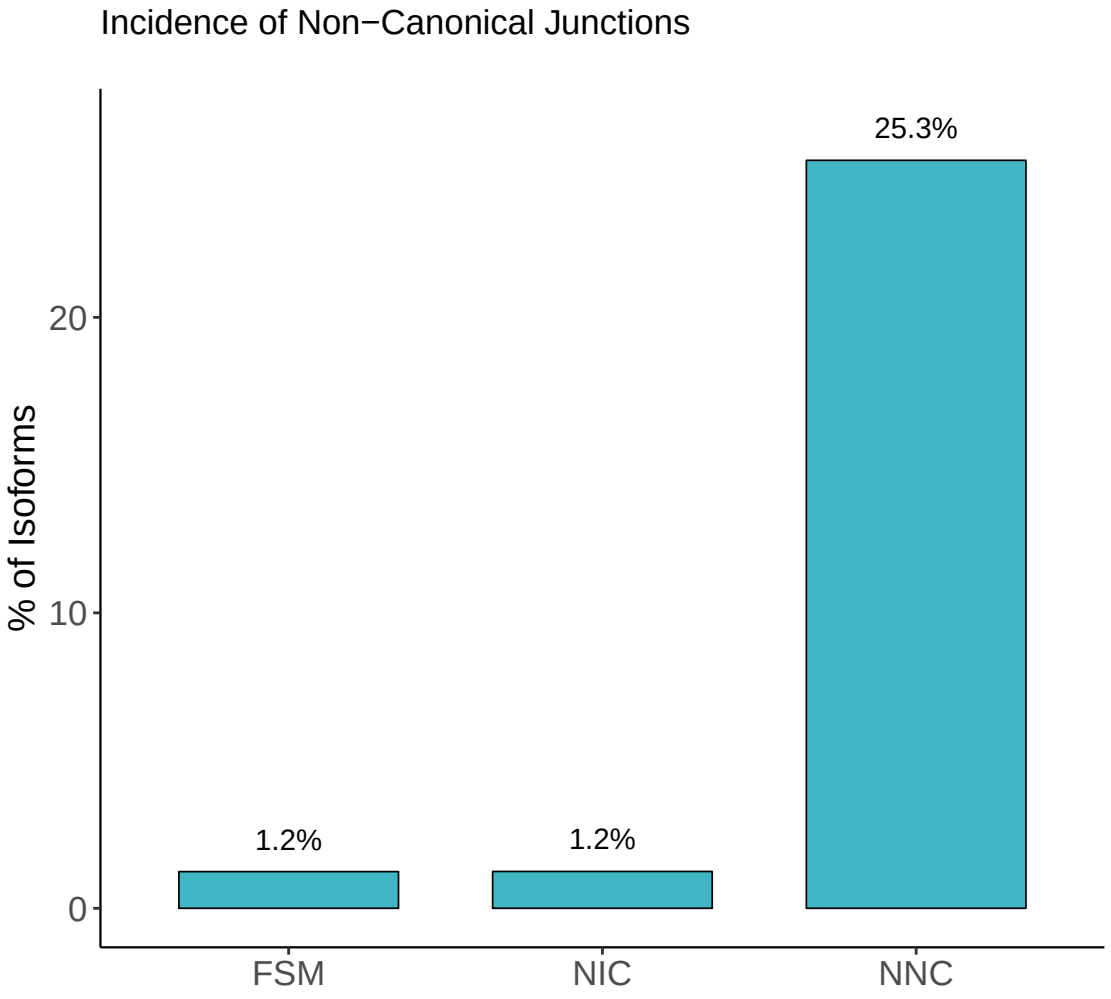

Possible Intra−Priming by Structural Category
Percent of genomic 'A's in downstream 20 bp
Percent 'A's (%)
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Intron retention
3' fragment
Internal fragment
5' fragment
Mono−exon
Multi−exon
Combination of annotated junctions
Combination of annotated splice sites
Mono−exon by intron retention
At least one annotated donor/acceptor

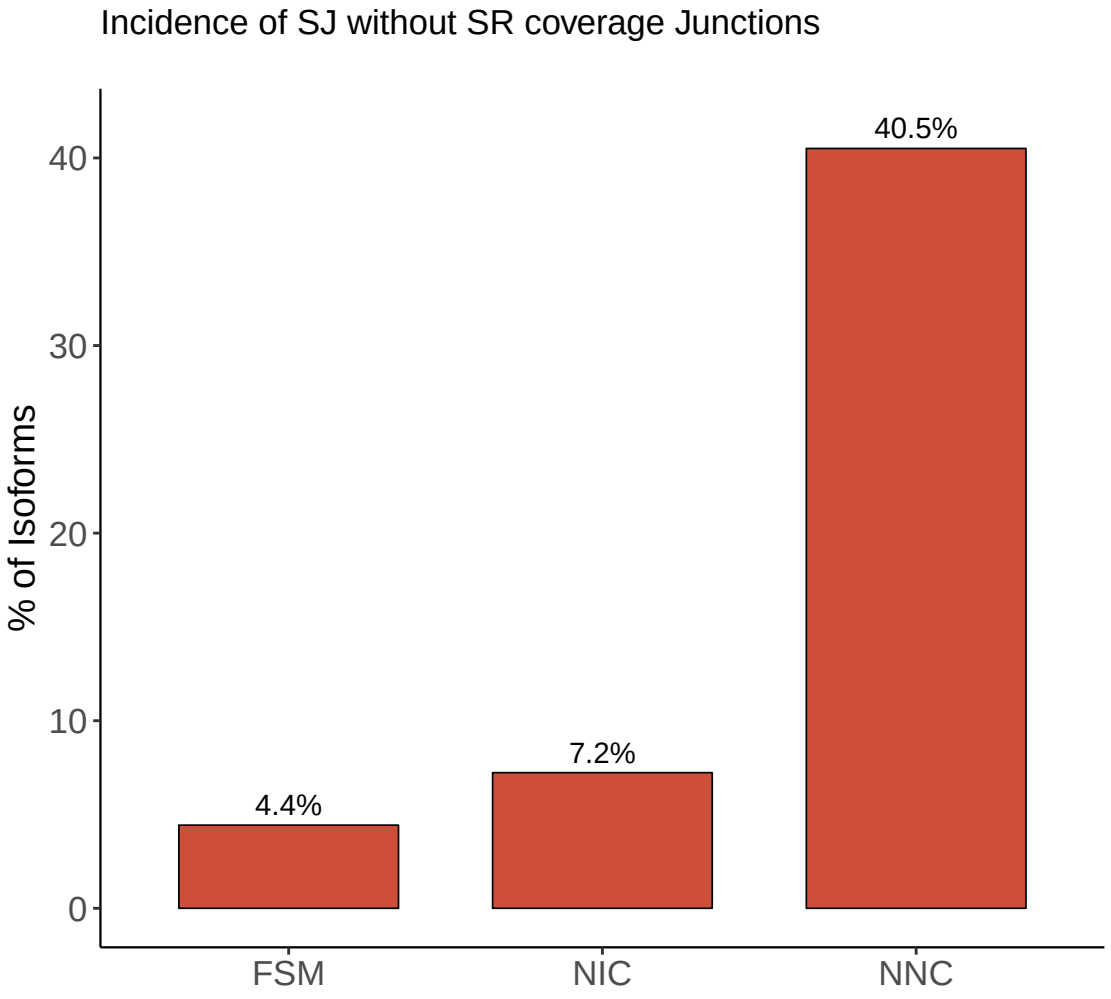

Possible Intra−Priming, Mono− vs Multi−Exon
Percent of genomic 'A's in downstream 20 bp
Percent 'A's (%)
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Mono−Exon Isoforms
Multi−Exon Isoforms

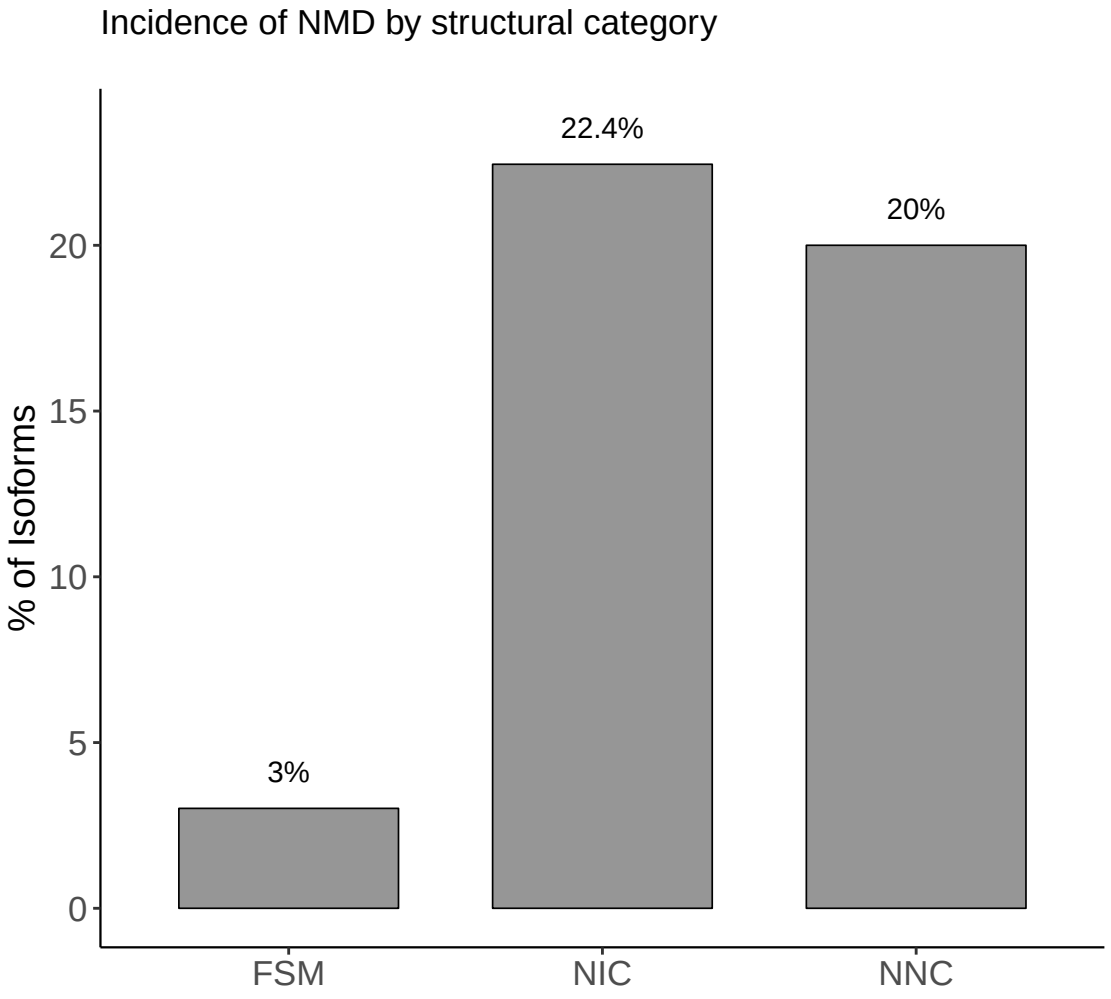

Possible Intra−Priming, Coding vs Non−Coding
Percent of genomic 'A's in downstream 20 bp
Percent 'A's (%)
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Coding Isoforms
Non−Coding Isoforms

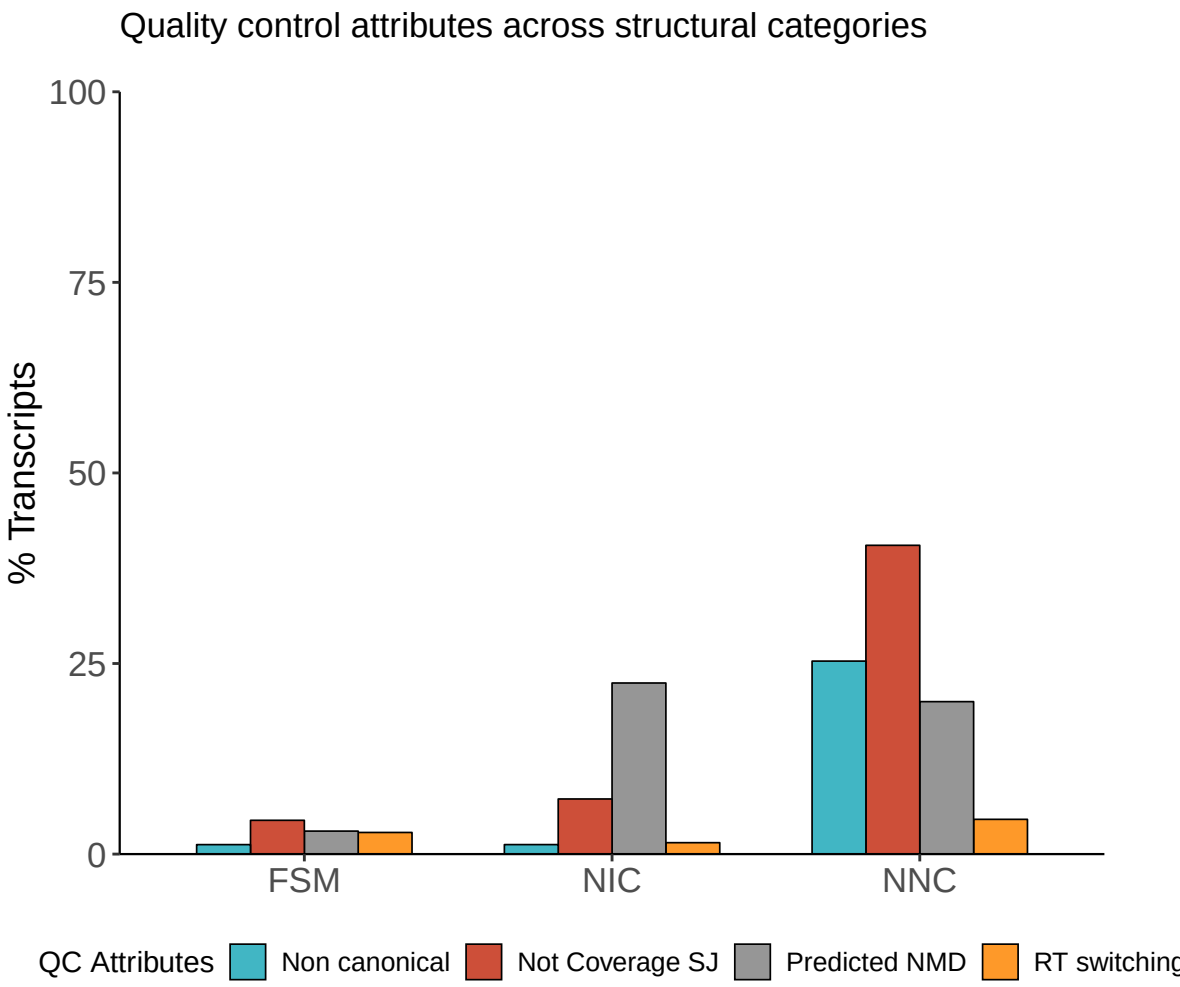
Quality Controls


Incidence of RT−switching
% of Isoforms
0
2
4
2.8%
1.5%
4.6%
FSM
NIC
NNC


Incidence of Non−Canonical Junctions
% of Isoforms
0
10
20
1.2%
1.2%
25.3%
FSM
NIC
NNC


Incidence of SJ without SR coverage Junctions
% of Isoforms
0
10
20
30
40
4.4%
7.2%
40.5%
FSM
NIC
NNC


Incidence of NMD by structural category
% of Isoforms
0
5
10
15
20
3%
22.4%
20%
FSM
NIC
NNC


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