SQANTI3 report

Unique Genes: 25797

Unique Isoforms: 558113

Transcript Classification

Category	Isoforms, count
FSM	130505
ISM	261895
NIC	83814
NNC	72591
Genic Genomic	724
Antisense	2241
Fusion	2407
Intergenic	3896
Genic Intron	0

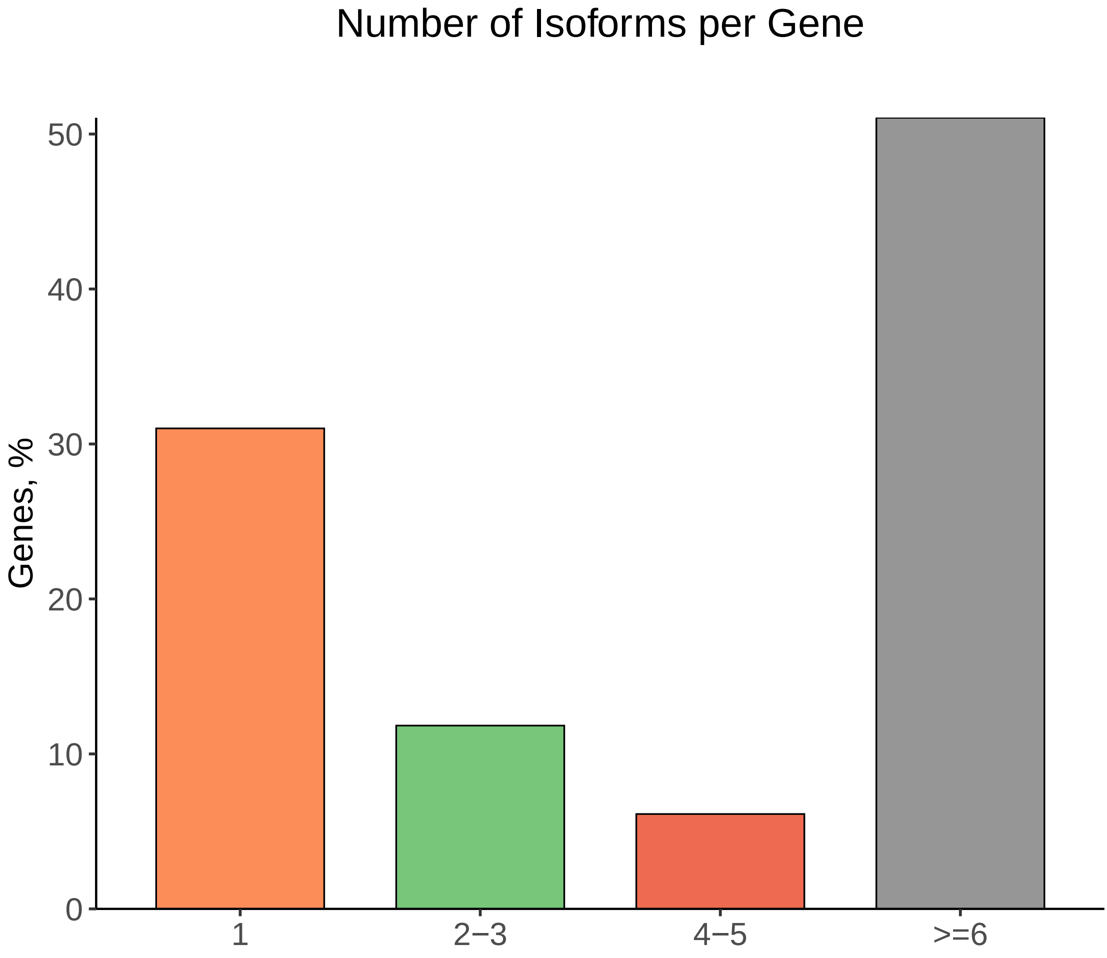
Gene Classification

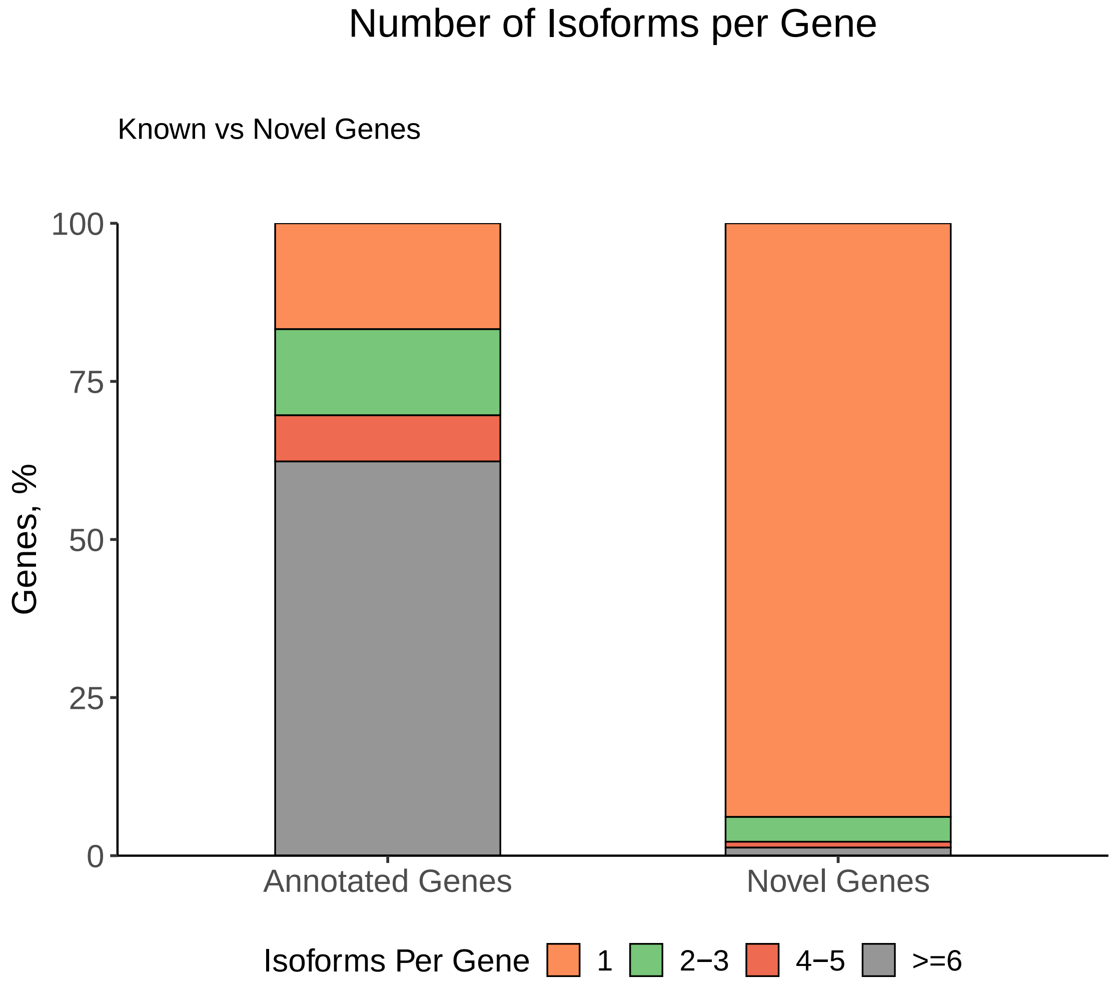
Category	Genes, count
Annotated Genes	21028
Novel Genes	4769

Splice Junction Classification

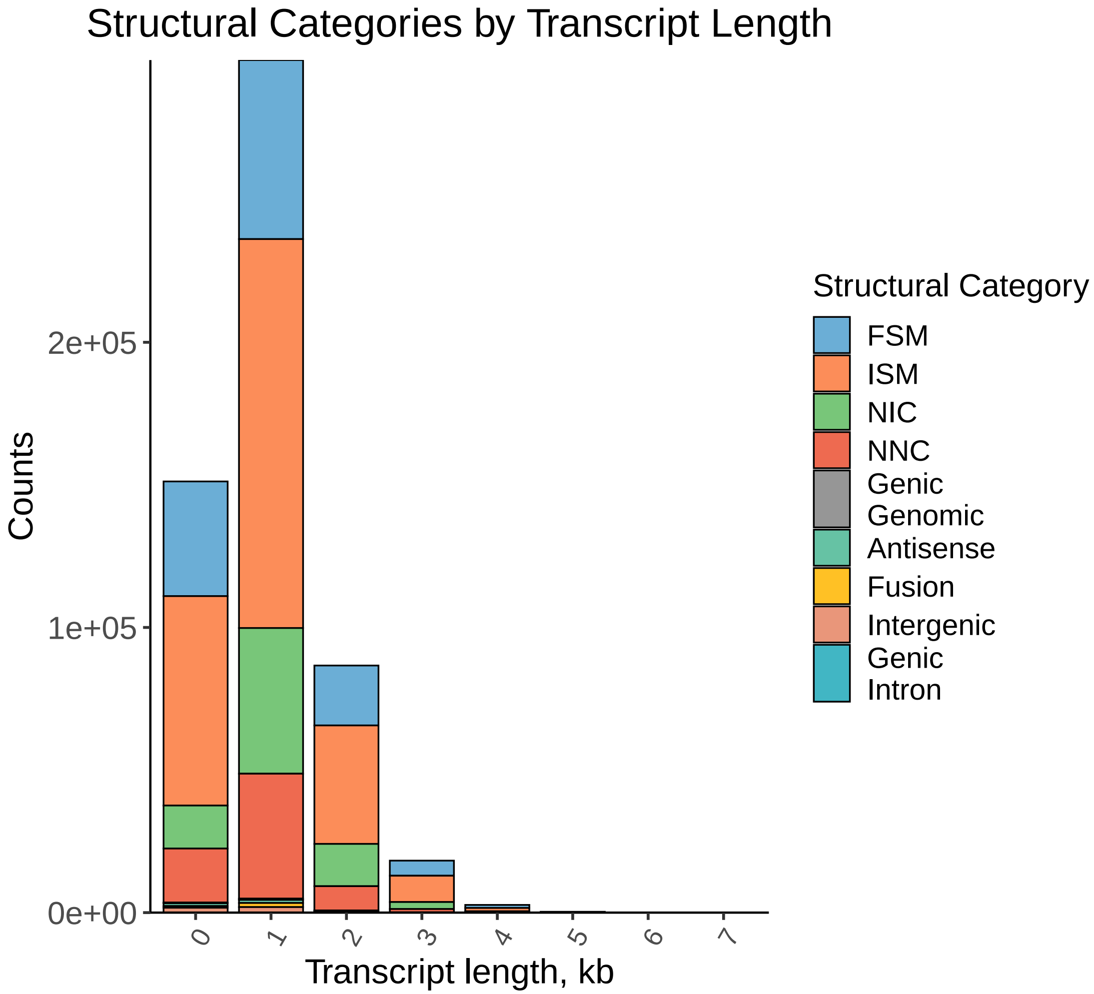
Category	SJs, count	Percent
Known canonical	171286	68.95
Known Non−canonical	61	0.02
Novel canonical	77063	31.02
Novel Non−canonical	0	0.00

Gene Characterization

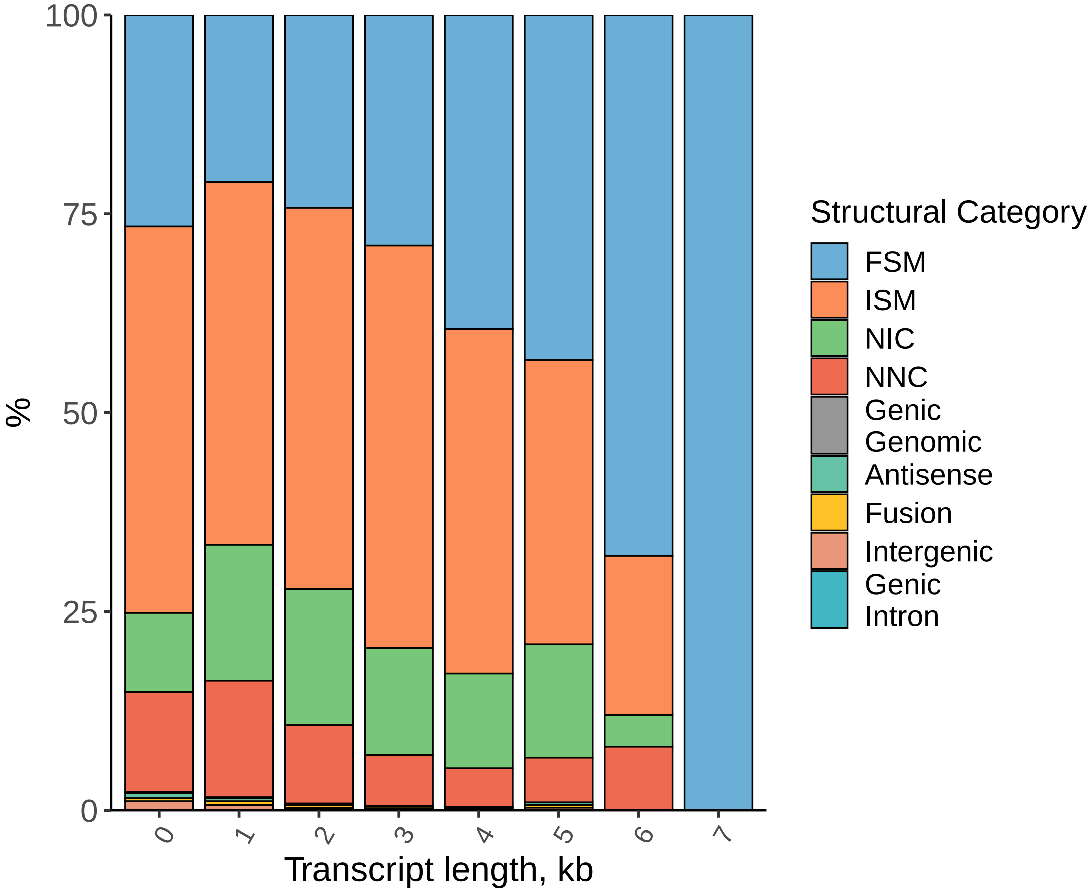

Number of Isoforms per Gene
Genes, %
0
10
20
30
40
50
1
2−3
4−5
>=6
Isoforms per gene

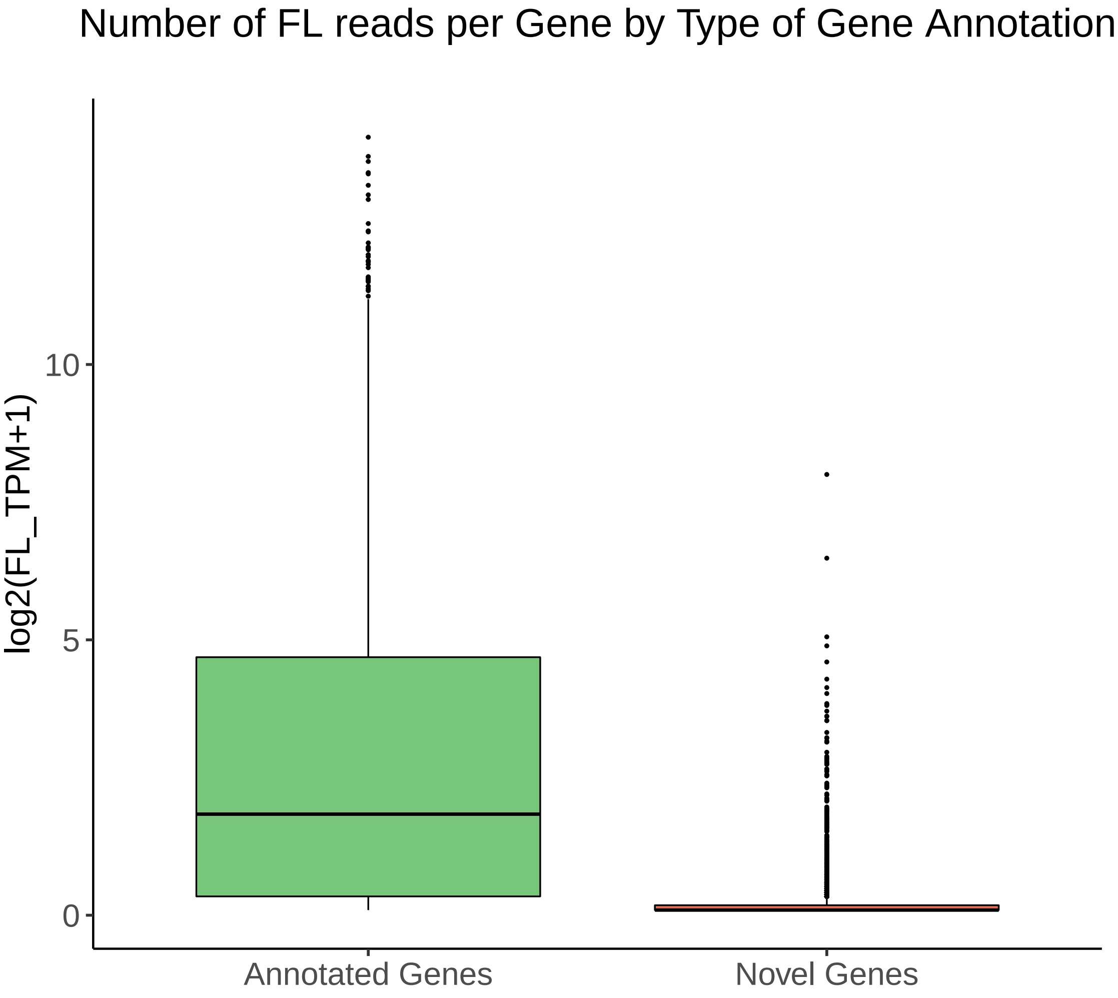

Number of Isoforms per Gene
Known vs Novel Genes
Genes, %
100
75
50
25
0
Annotated Genes
Novel Genes
Isoforms Per Gene
1
2−3
4−5
>=6

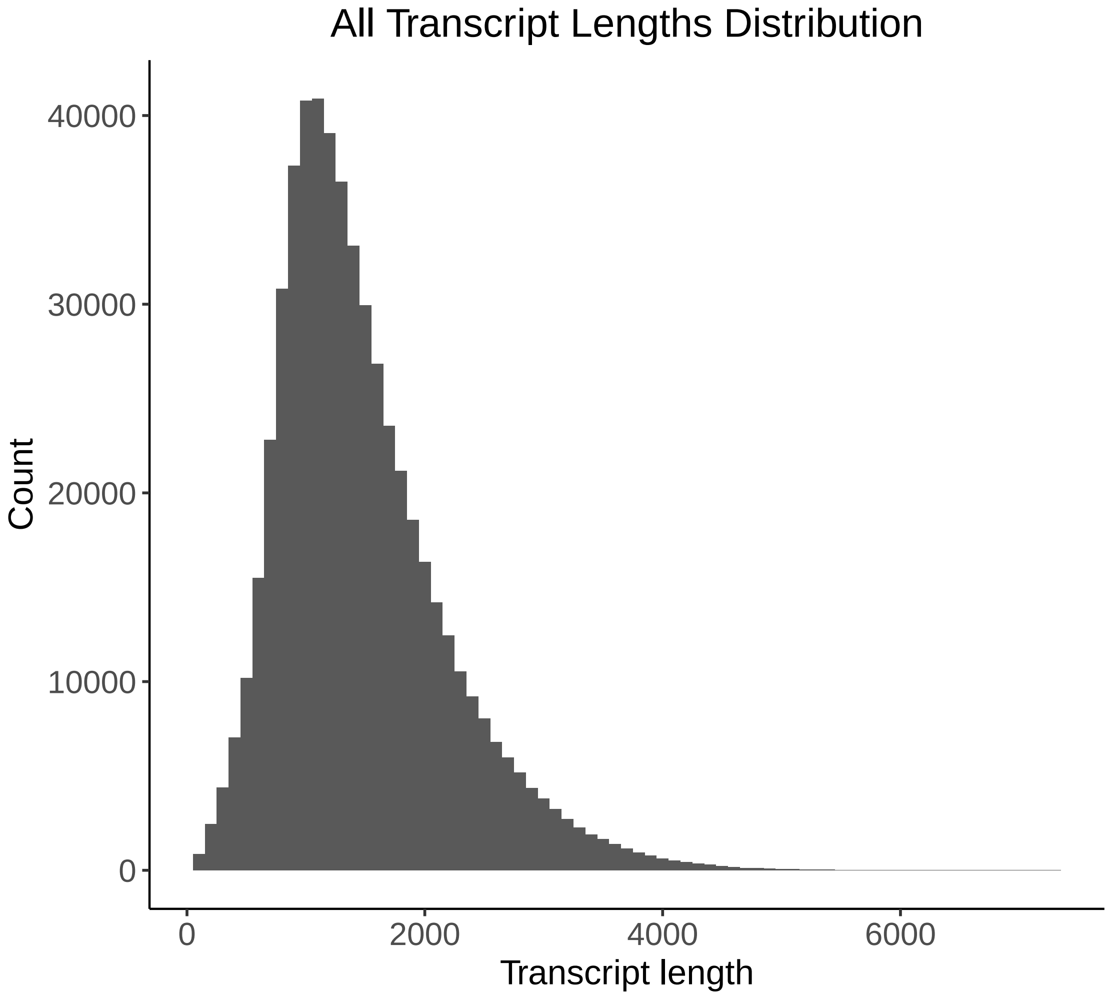

Distribution of Mono− vs Multi−Exon Transcripts
Transcripts, %
100
75
50
25
0
Novel Genes
Annotated Genes
Transcript type
Multi−Exon
Mono−Exon

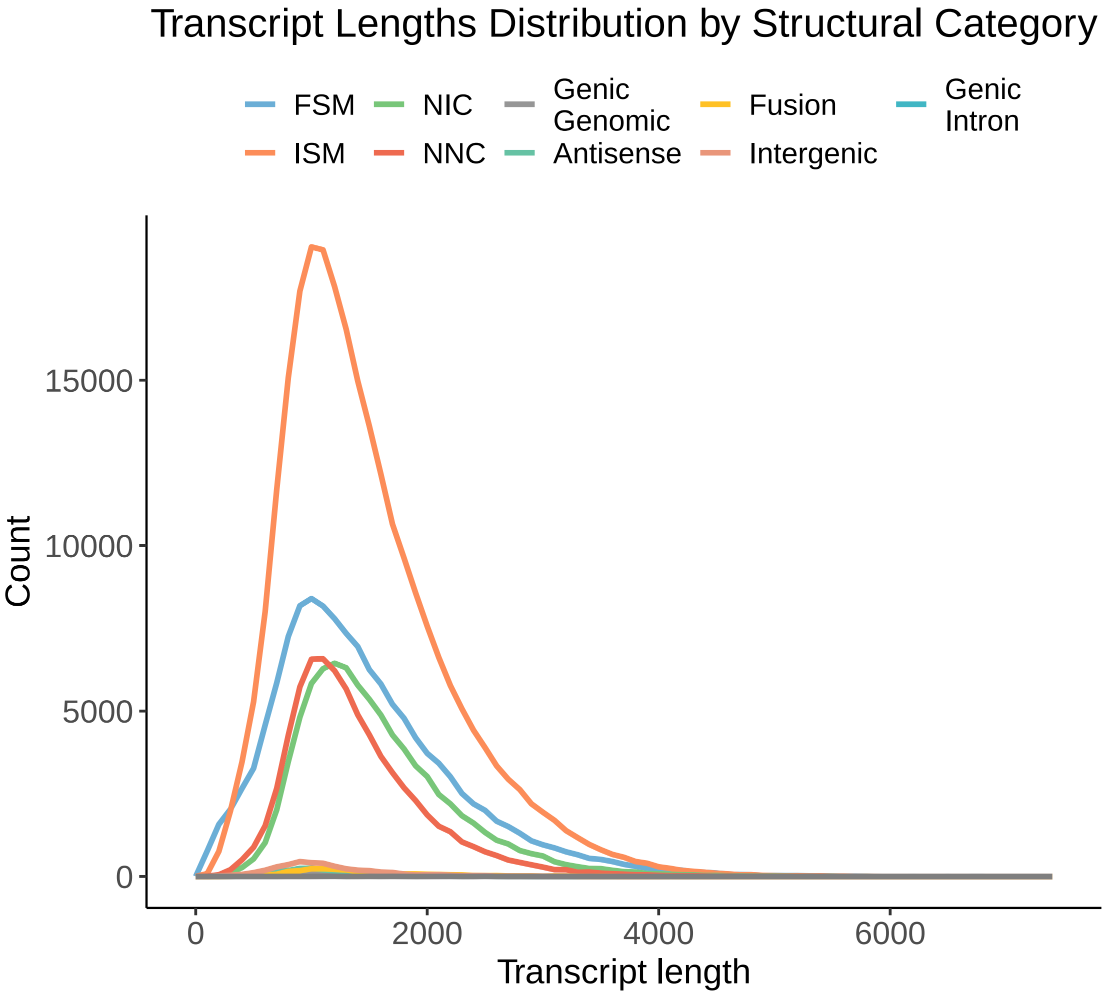

Structural Categories by Transcript Length
Counts
2e+05
1e+05
0e+00
0
1
2
3
4
5
6
7
Transcript length, kb
Structural Category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

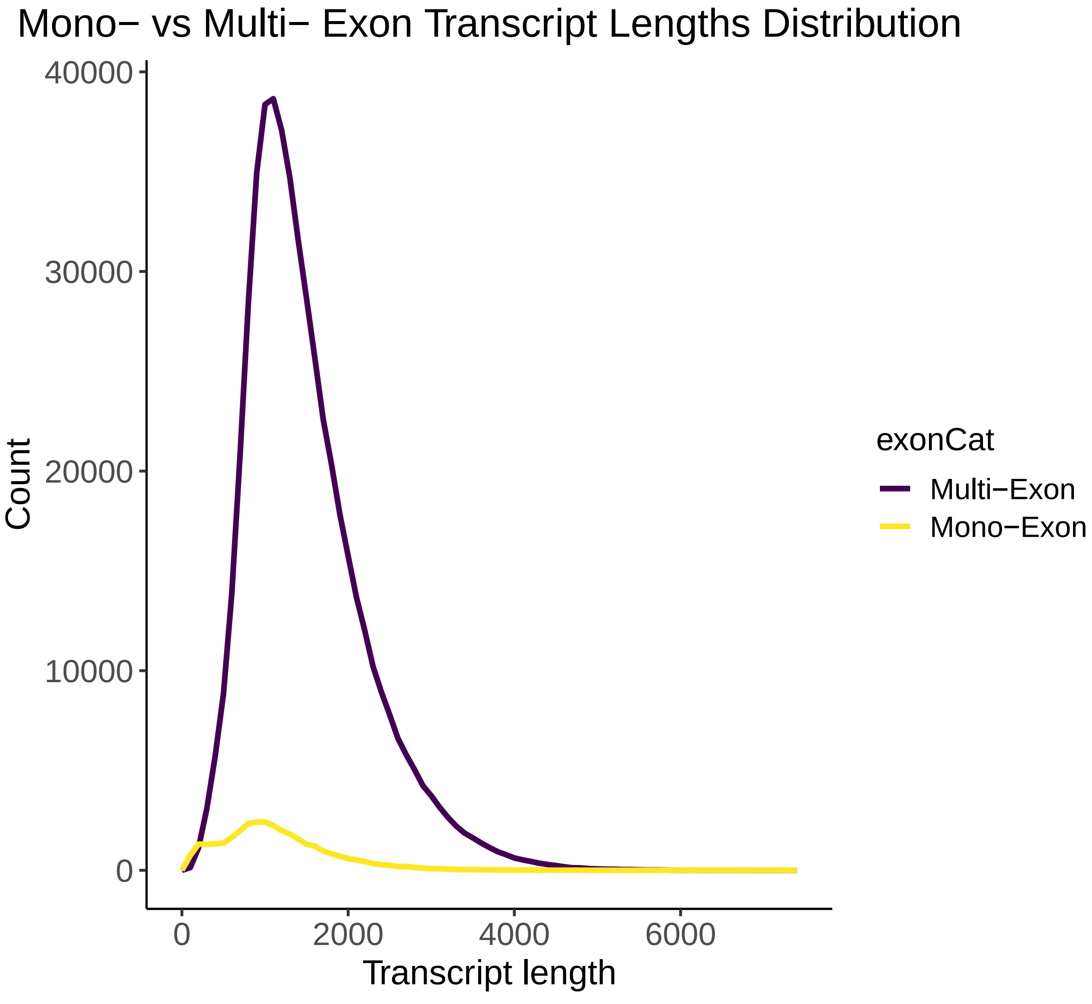
Structural Categories by Transcript Length

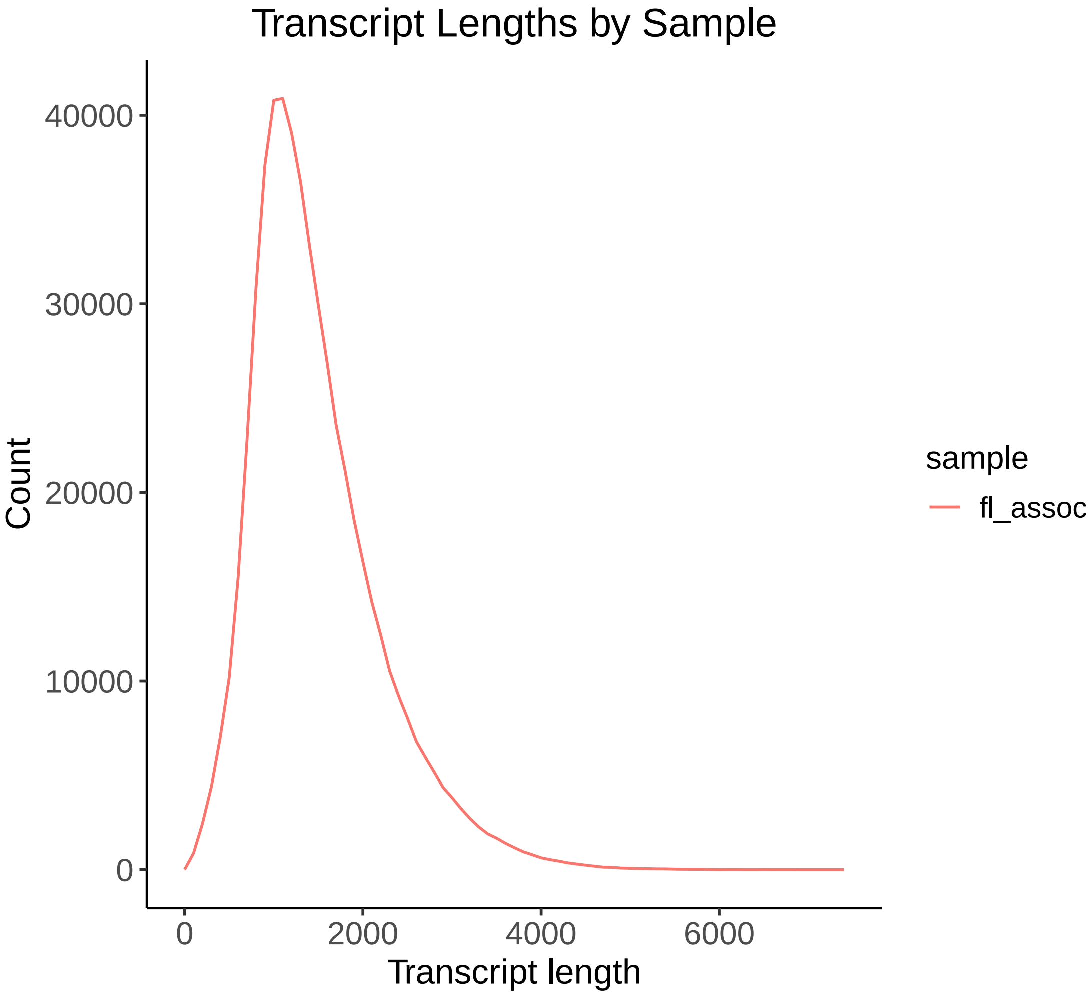


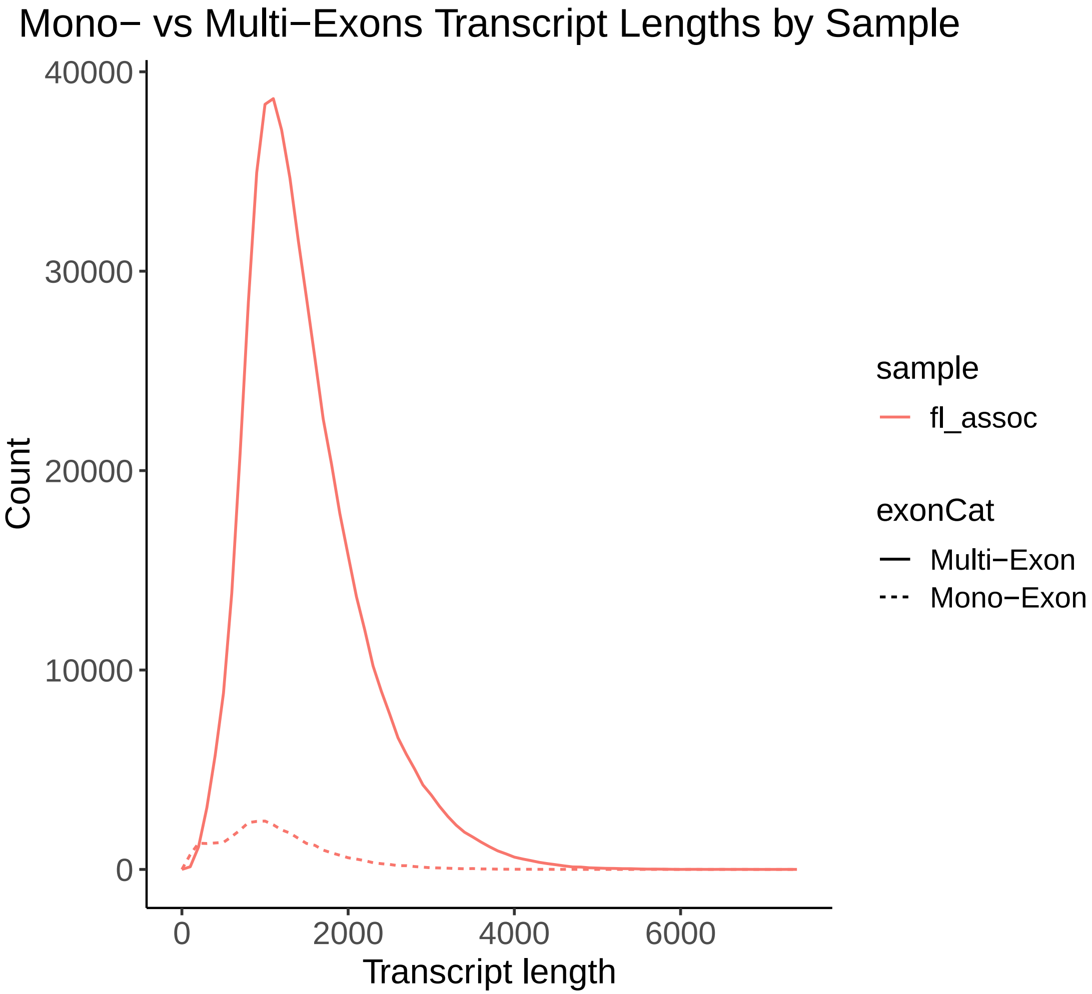

Number of FL reads per Gene by Type of Gene Annotation
log2(FL_TPM+1)
10
5
0
Annotated Genes
Novel Genes


All Transcript Lengths Distribution
Count
40000
30000
20000
10000
0
0
2000
4000
6000
Transcript length

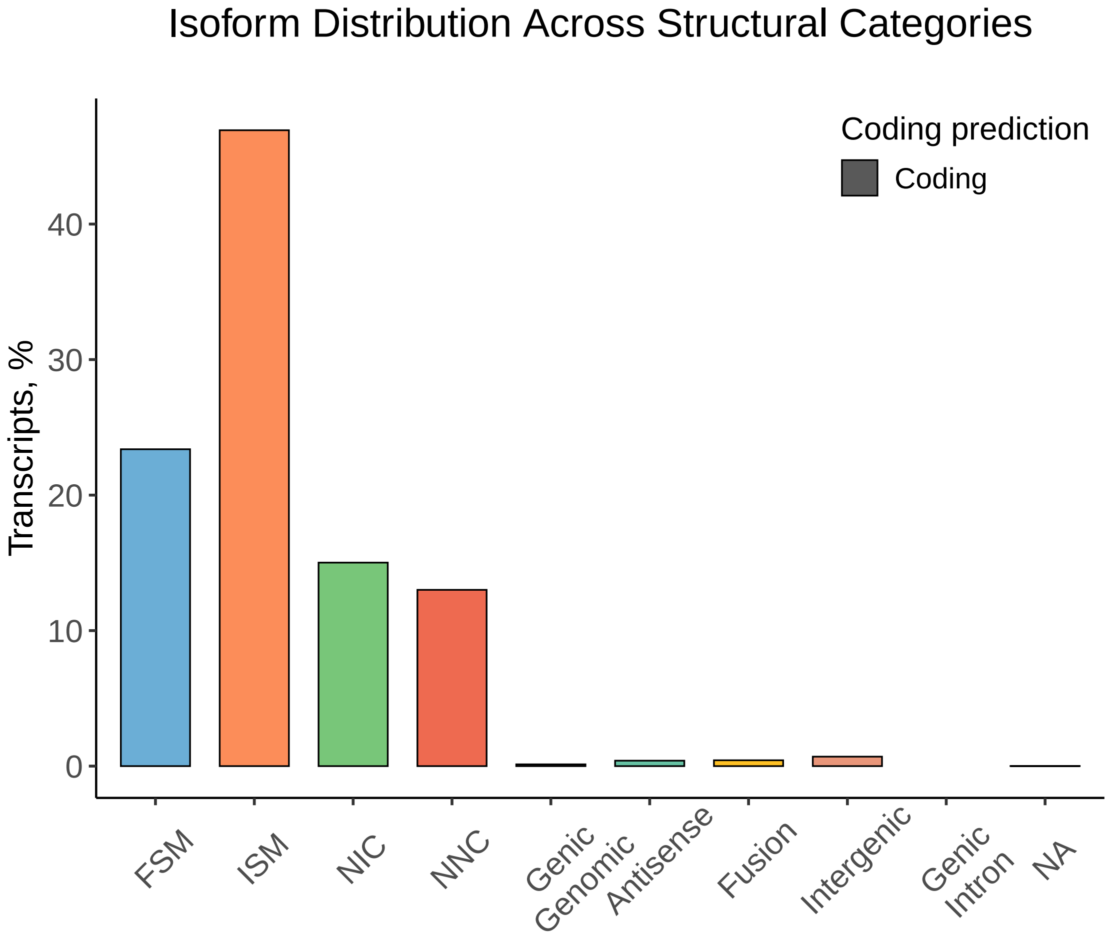

Transcript Lengths Distribution by Structural Category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron
Count
15000
10000
5000
0
0
2000
4000
6000
Transcript length

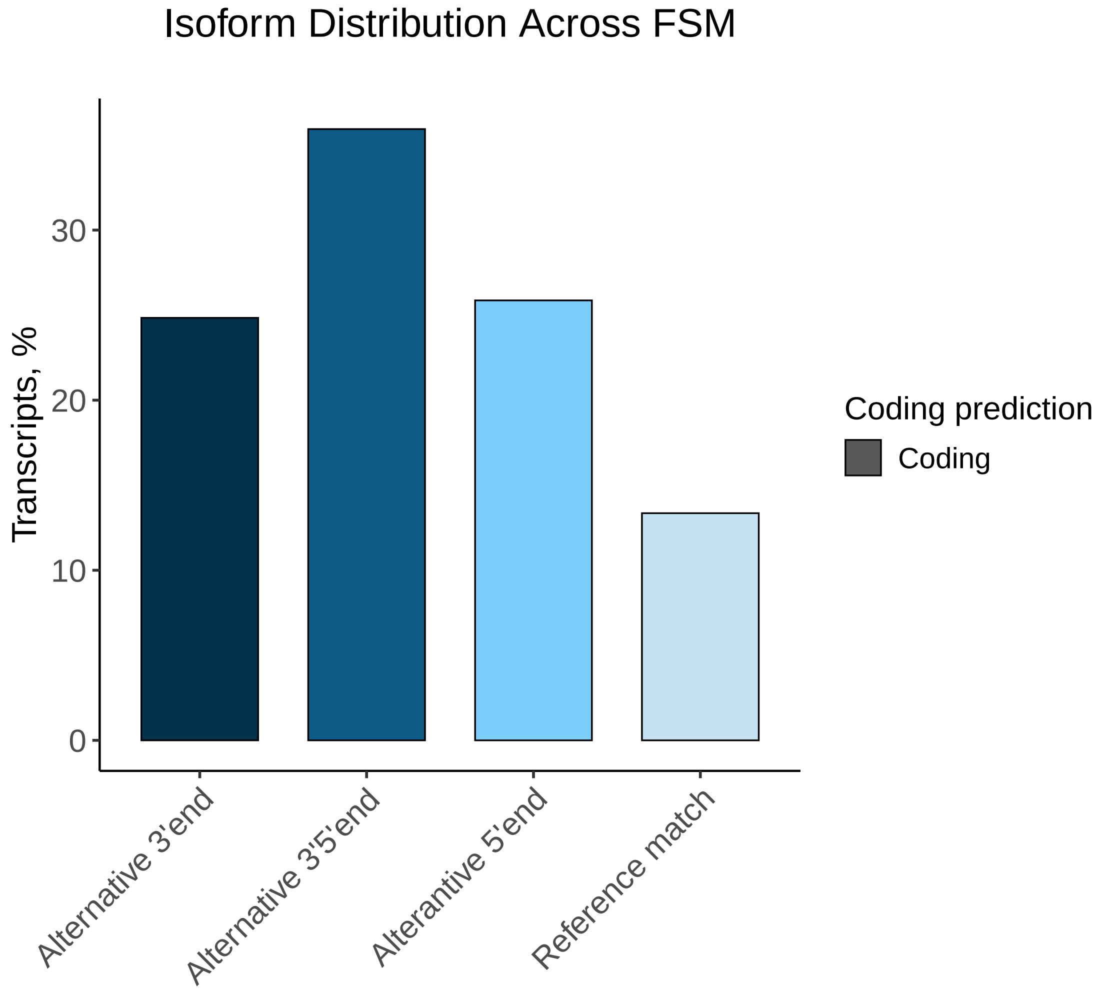

Mono− vs Multi− Exon Transcript Lengths Distribution
Count
40000
30000
20000
10000
0
0
2000
4000
6000
Transcript length
exonCat
Multi−Exon
Mono−Exon

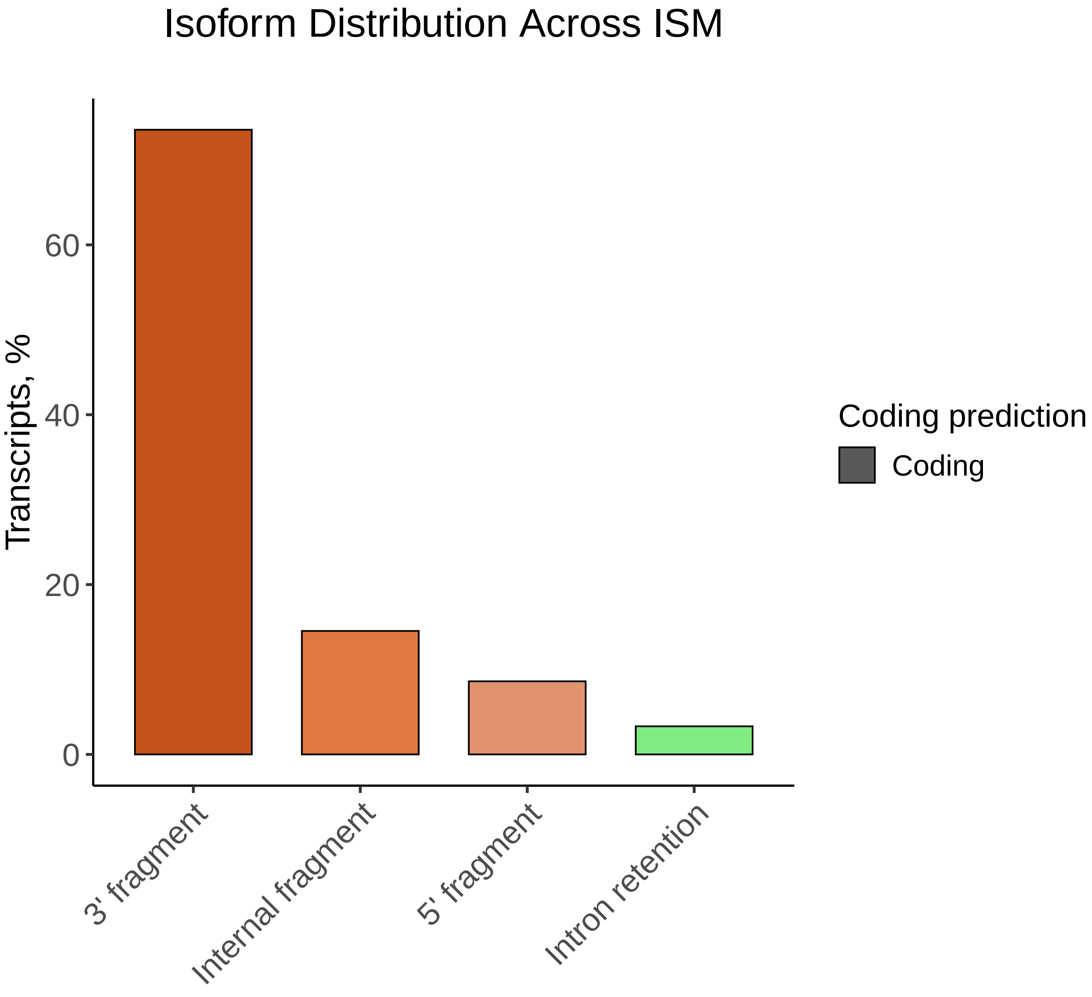

Transcript Lengths by Sample
Count
40000
30000
20000
10000
0
0
2000
4000
6000
Transcript length
sample
fl_assoc

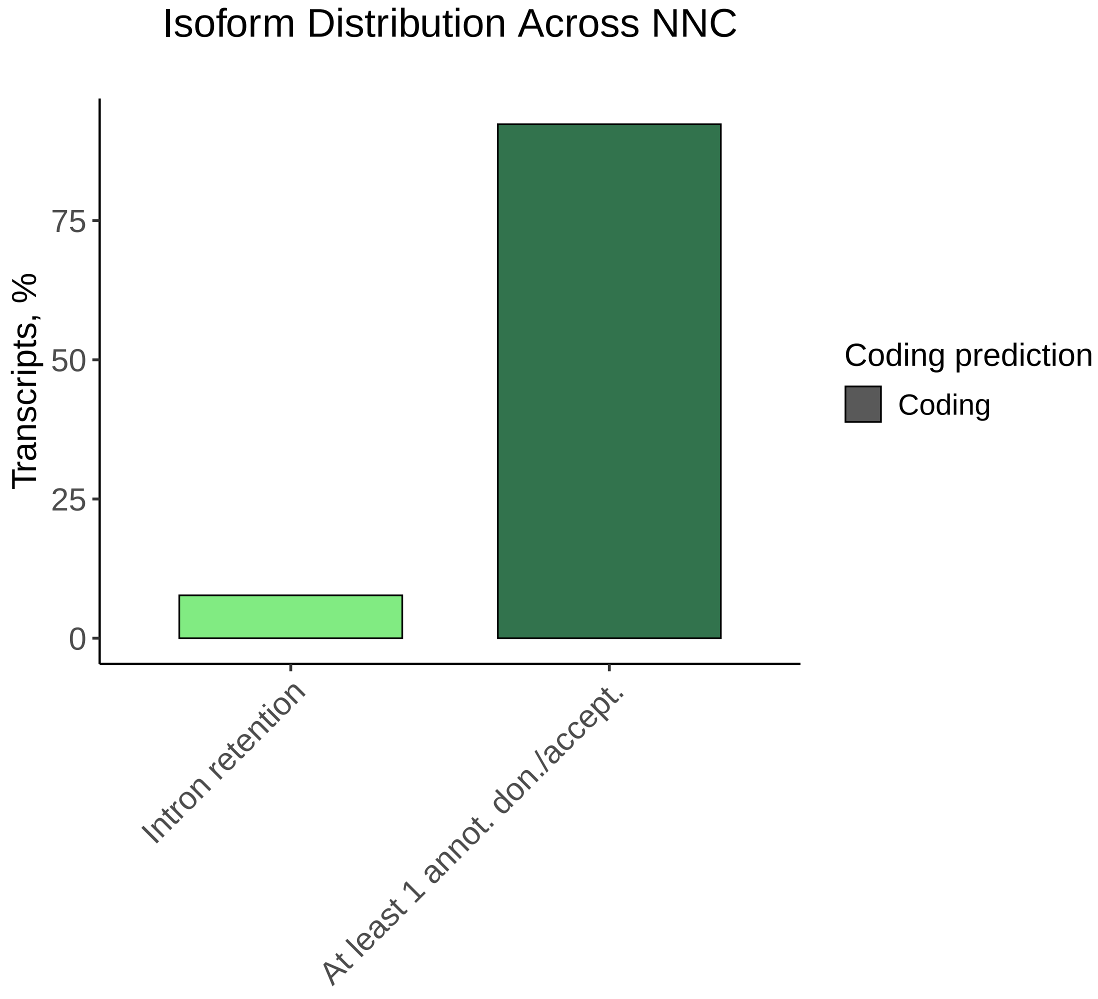

Mono− vs Multi−Exons Transcript Lengths by Sample
40000
30000
20000
10000
0
Count
0
2000
4000
6000
Transcript length
sample
fl_assoc
exonCat
Multi−Exon
Mono−Exon

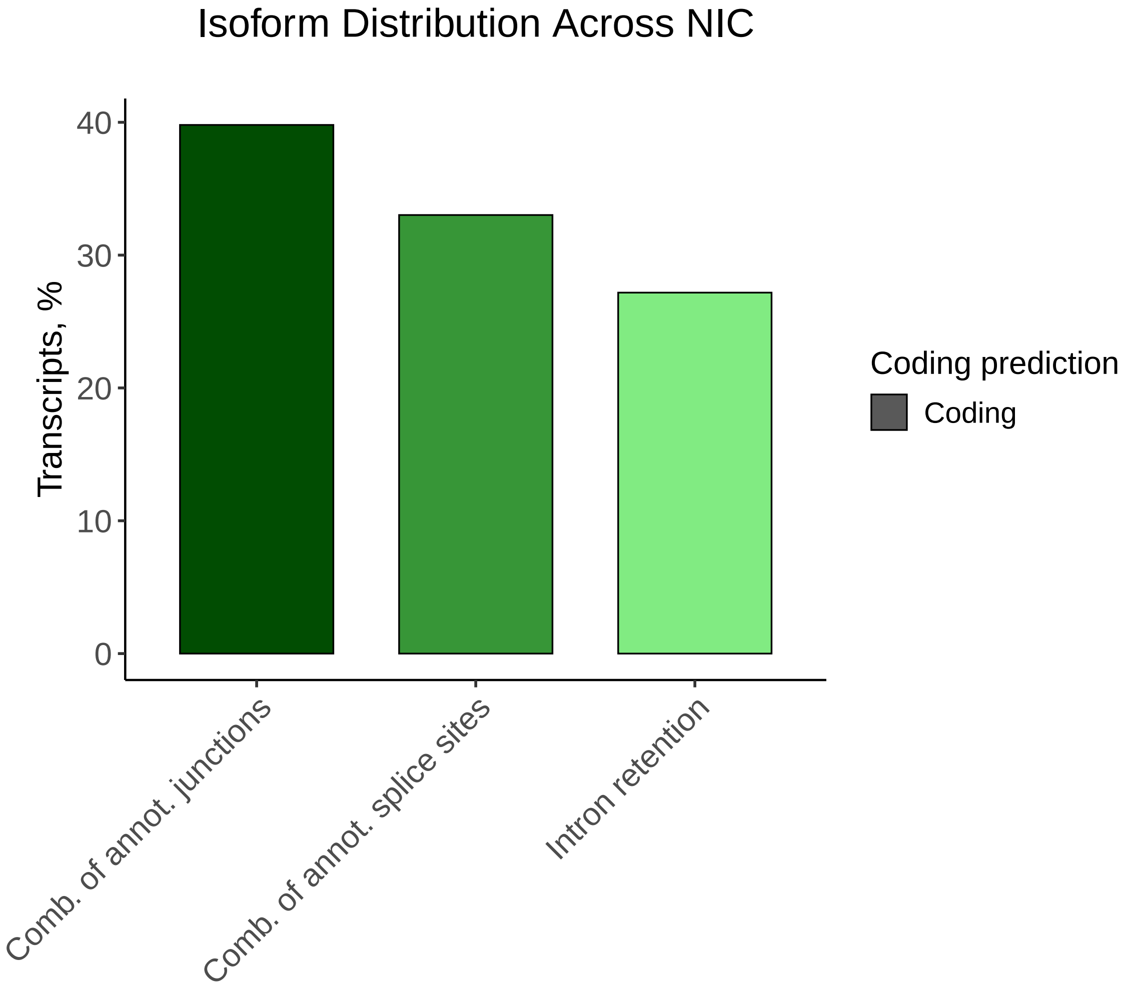
Structural Isoform Characterization

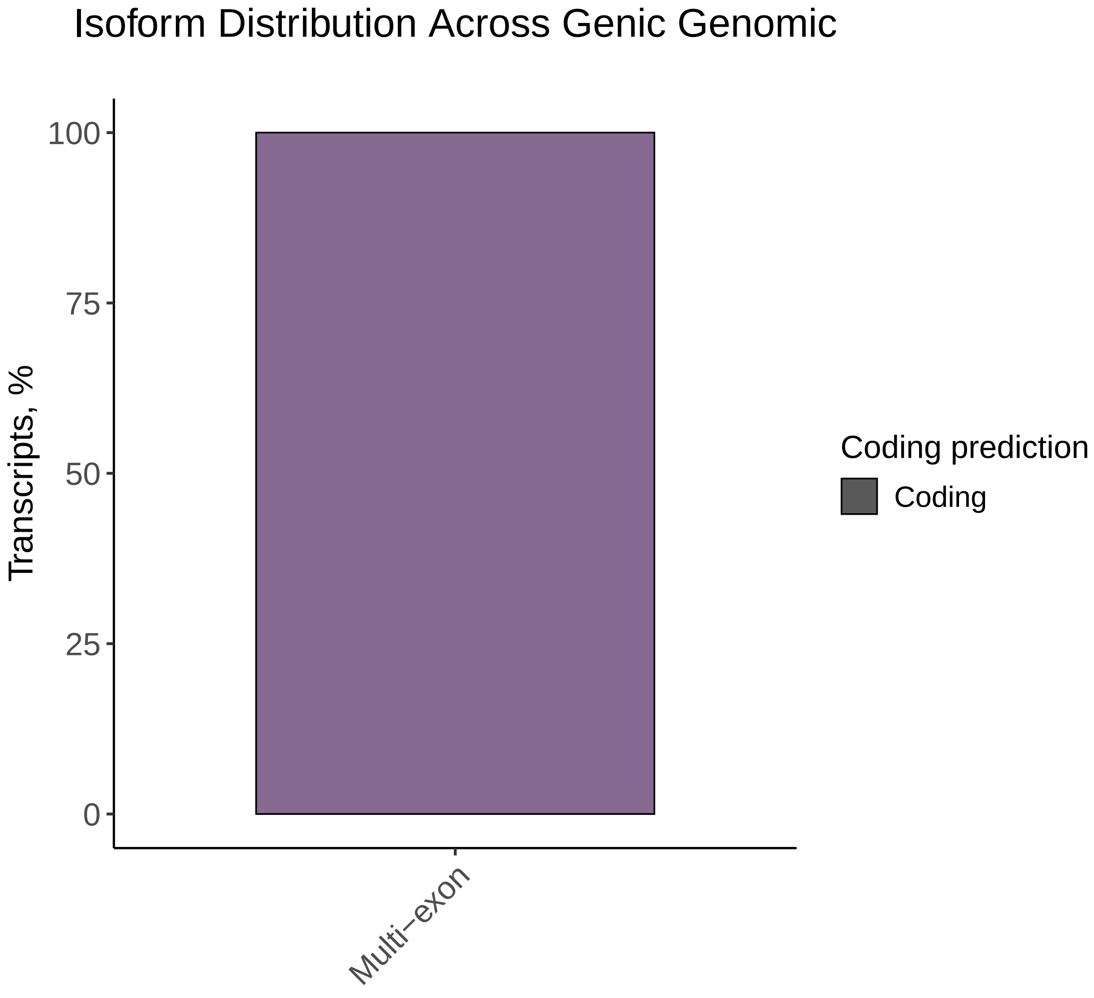

Isoform Distribution Across Structural Categories
Coding prediction
Coding
Transcripts, %
40
30
20
10
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron
NA

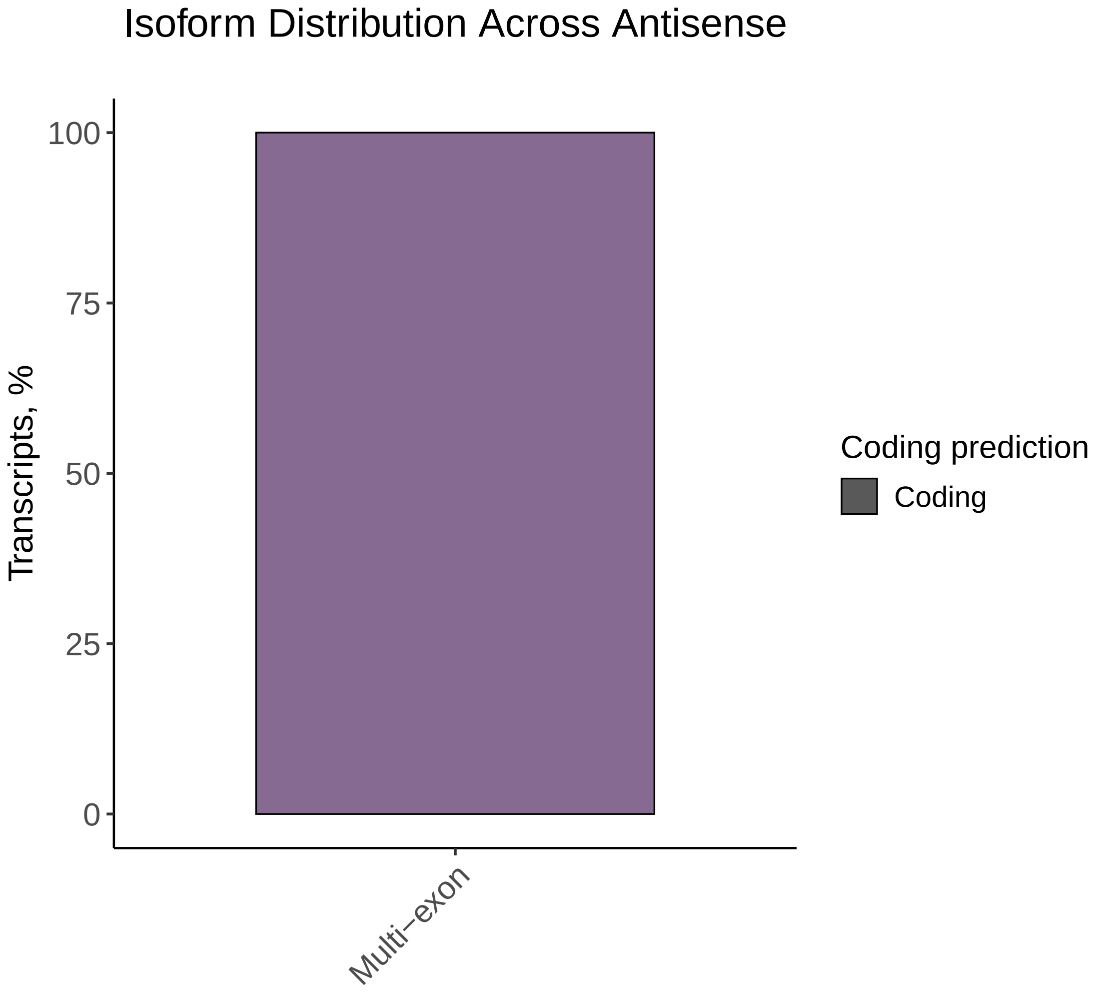

Isoform Distribution Across FSM
Transcripts, %
30
20
10
0
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
Coding prediction
Coding

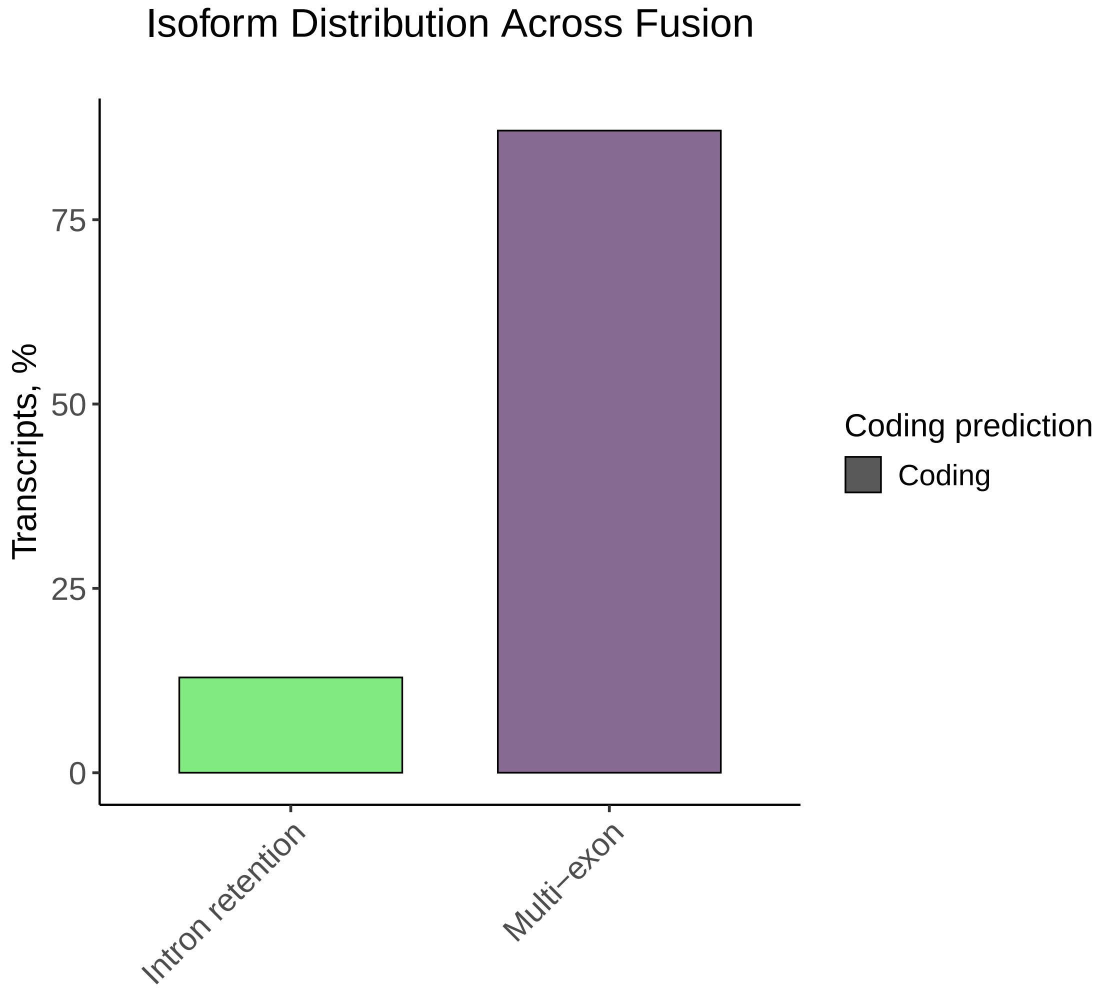

Isoform Distribution Across ISM
Transcripts, %
0
20
40
60
3' fragment
Internal fragment
5' fragment
Intron retention
Coding prediction
Coding

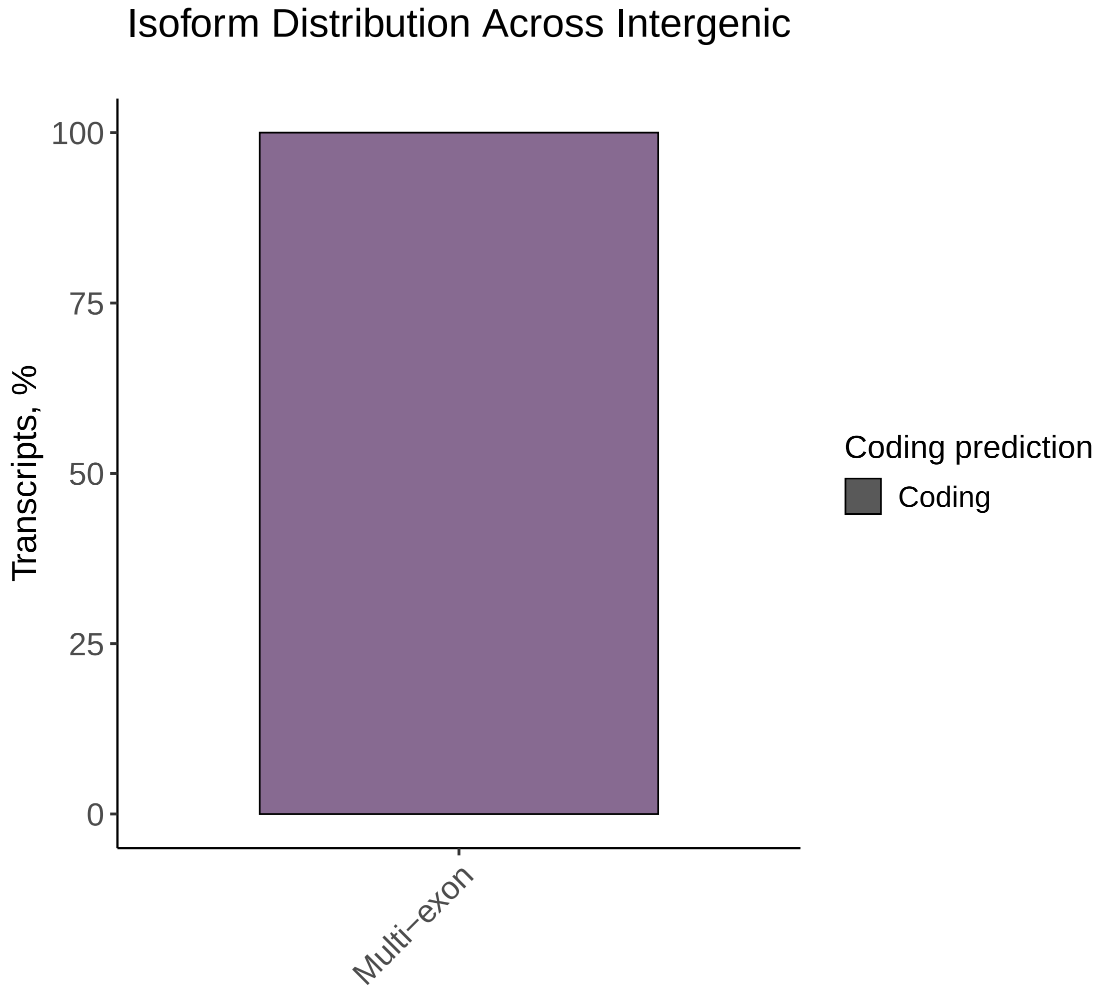

Isoform Distribution Across NNC
Transcripts, %
75
50
25
0
Intron retention
At least 1 annot. don./accept.
Coding prediction
Coding

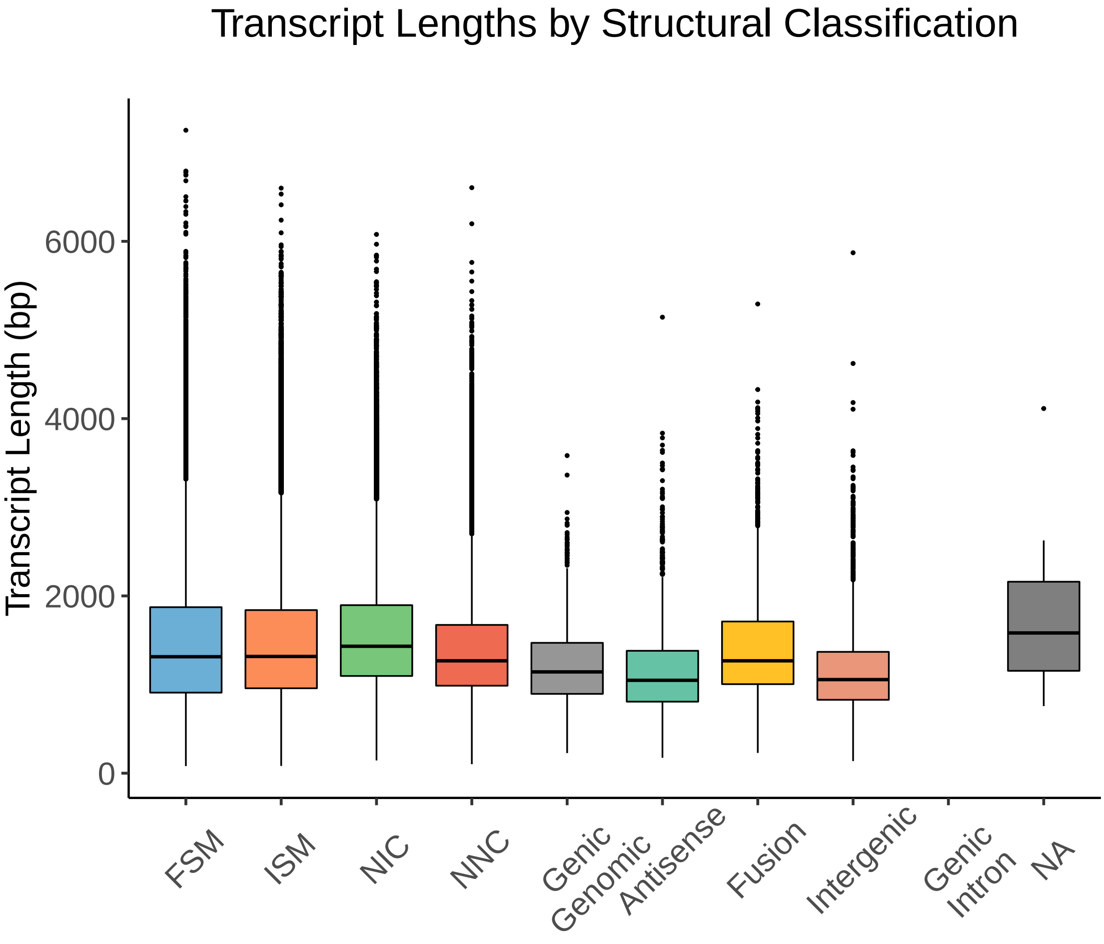

Isoform Distribution Across NIC
Transcripts, %
40
30
20
10
0
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Coding prediction
Coding


Isoform Distribution Across Genic Genomic
Transcripts, %
100
75
50
25
0
Multi–exon
Coding prediction
Coding

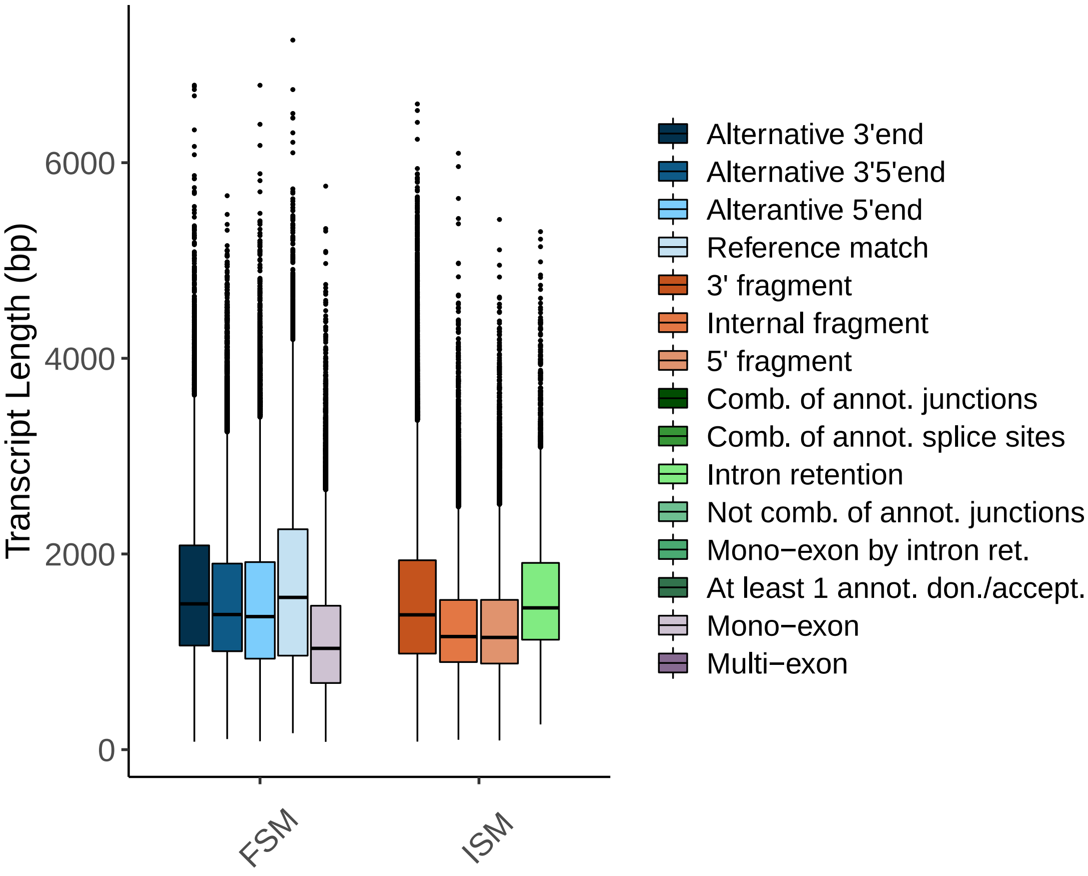

Isoform Distribution Across Antisense
Transcripts, %
100
75
50
25
0
Multi–exon
Coding prediction
Coding

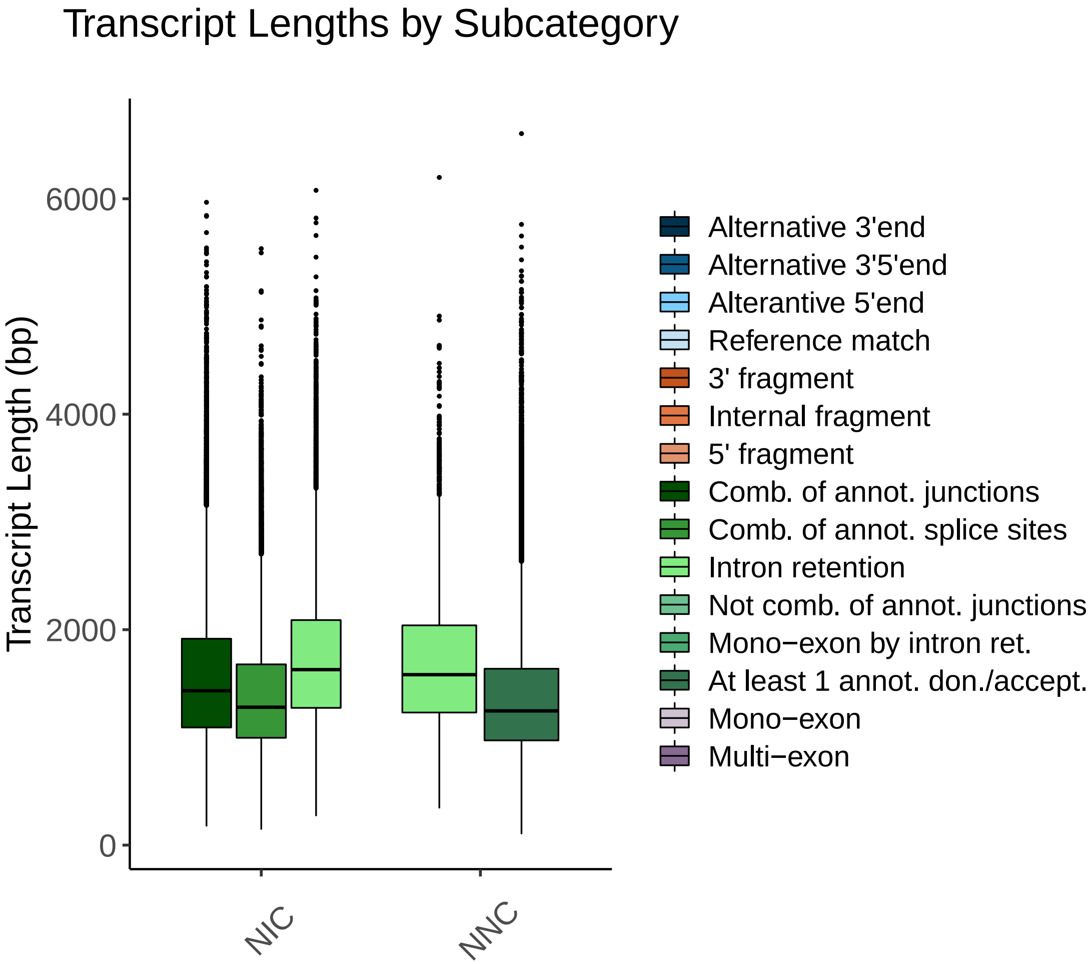

Isoform Distribution Across Fusion
Transcripts, %
0
25
50
75
Intron retention
Multi−exon
Coding prediction
Coding

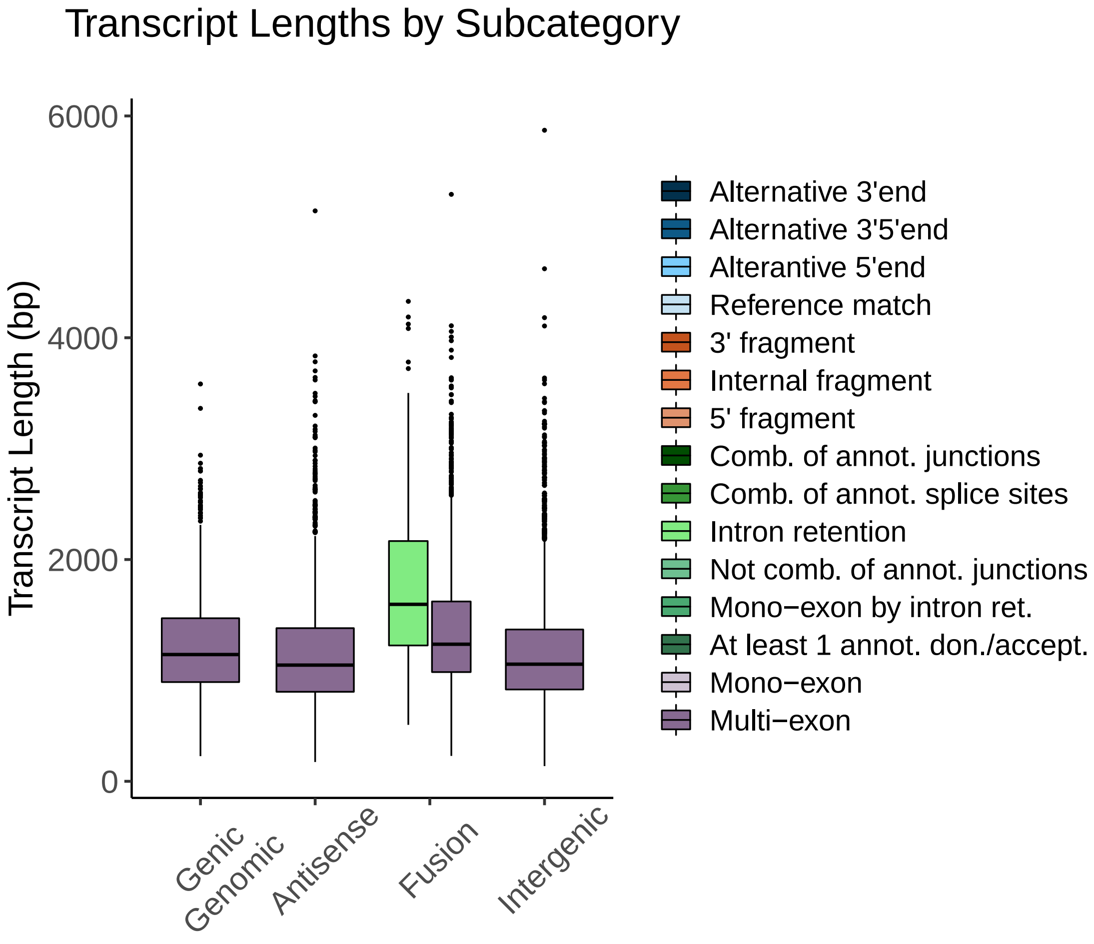

Isoform Distribution Across Intergenic
Transcripts, %
100
75
50
25
0
Multi–exon
Coding prediction
Coding


Transcript Lengths by Structural Classification
Transcript Length (bp)
0
2000
4000
6000
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron
NA

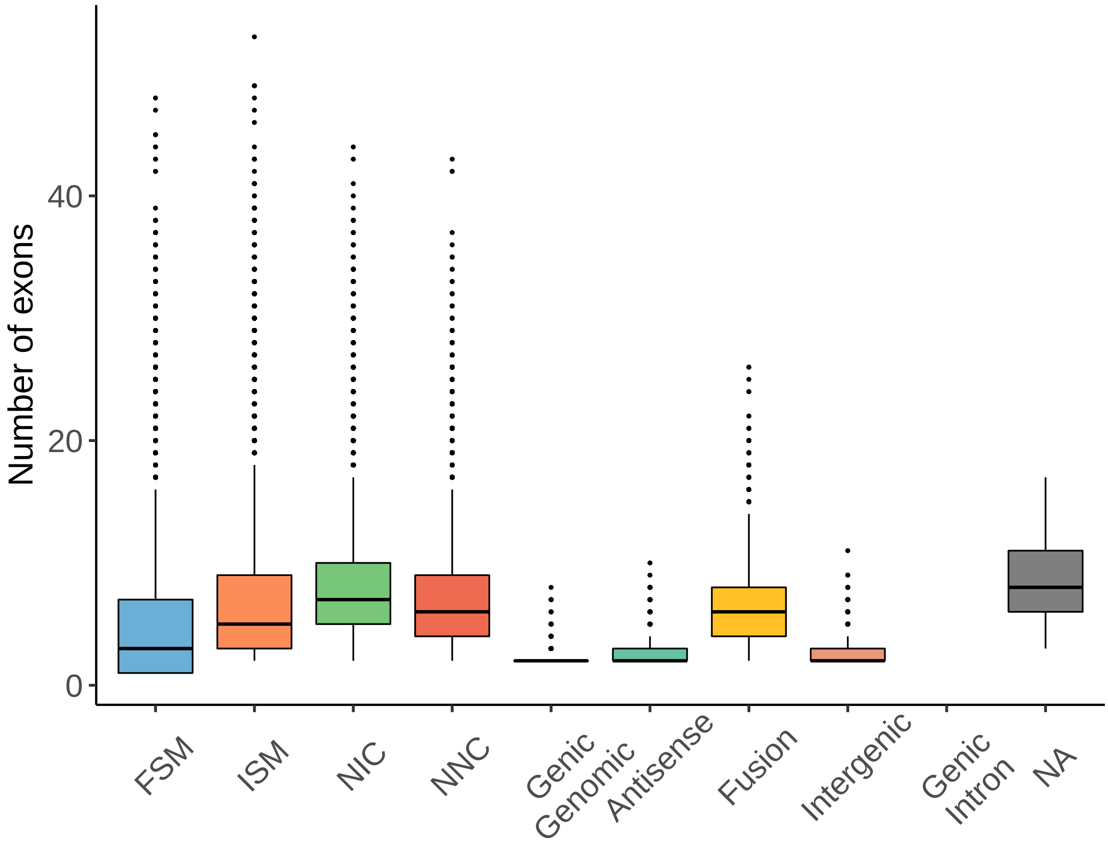
Transcript Lengths by Subcategory

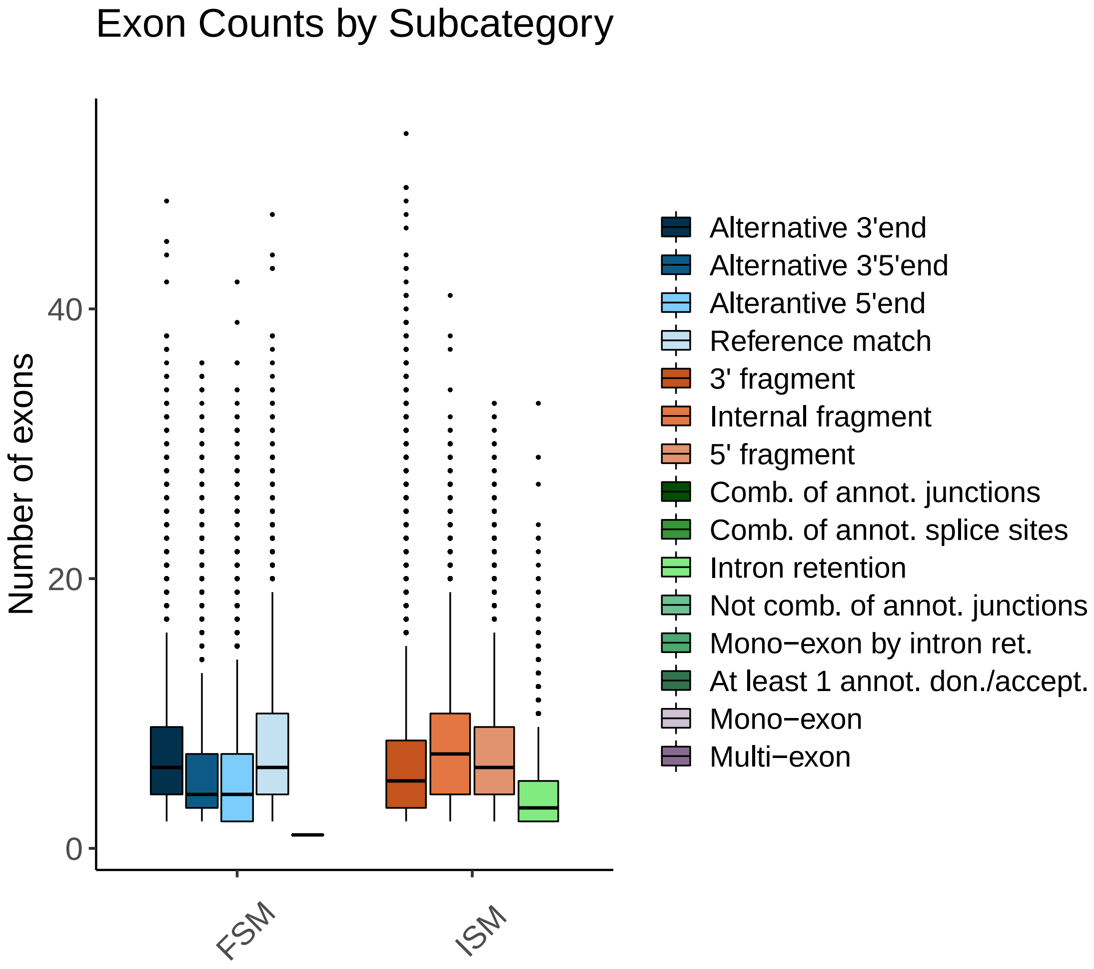


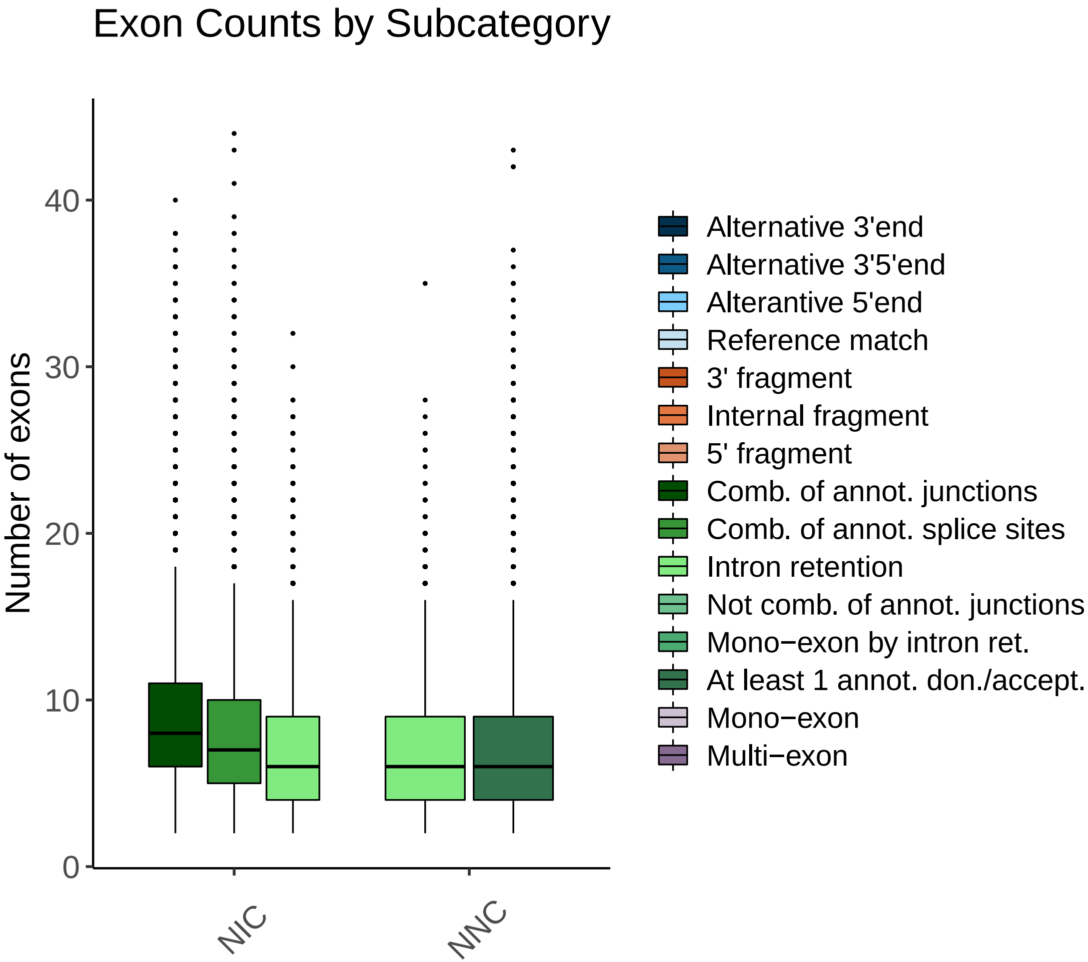

Transcript Lengths by Subcategory
Transcript Length (bp)
6000
4000
2000
0
NIC
NNC
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon

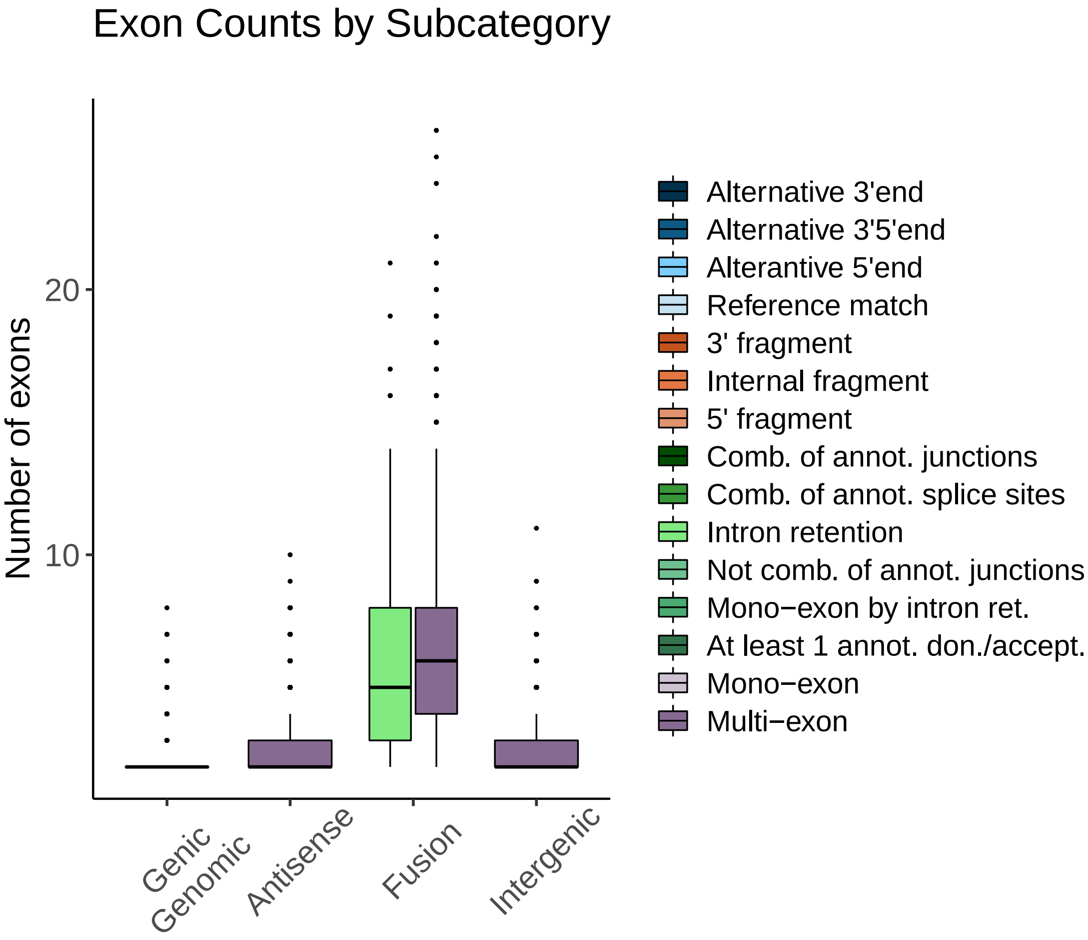

Transcript Lengths by Subcategory
Transcript Length (bp)
6000
4000
2000
0
Genic Genomic
Antisense
Fusion
Intergenic
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon

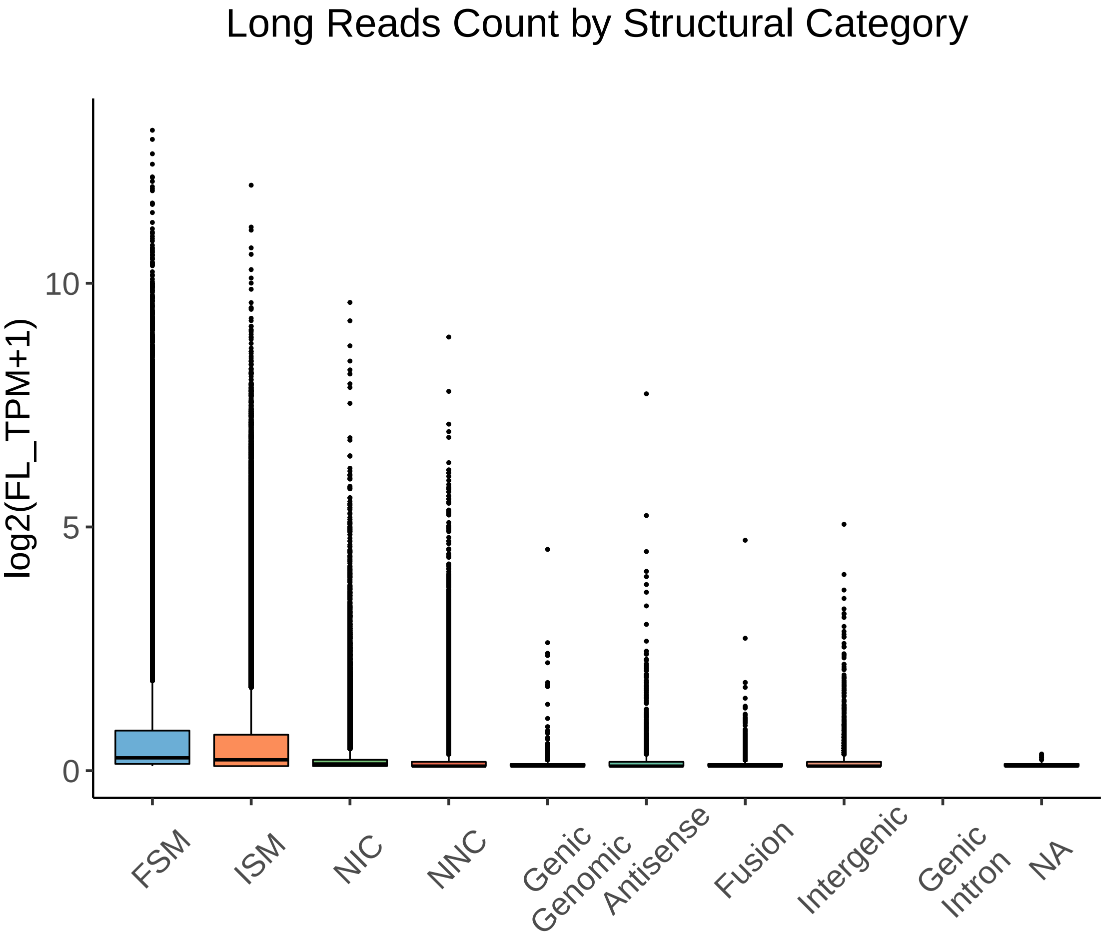
Exon Counts by Structural Classification




Exon Counts by Subcategory
Number of exons
0
20
40
FSM
ISM
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon

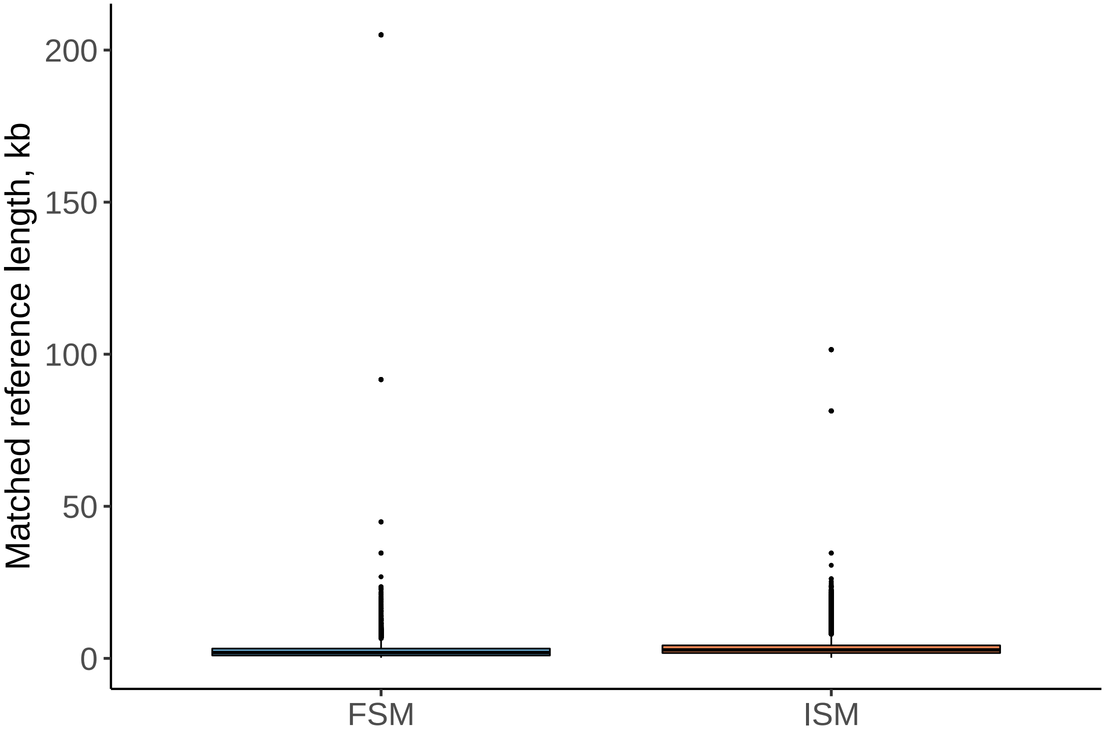

Exon Counts by Subcategory
Number of exons
0
10
20
30
40
NIC
NNC
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon


Exon Counts by Subcategory
Number of exons
20
10
Genic Genomic
Antisense
Fusion
Intergenic
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon


Long Reads Count by Structural Category
log2(FL_TPM+1)
10
5
0
FSM
ISM
NIC
NNC
Genic
Genomic
Antisense
Fusion
Intergenic
Genic
Intron
NA

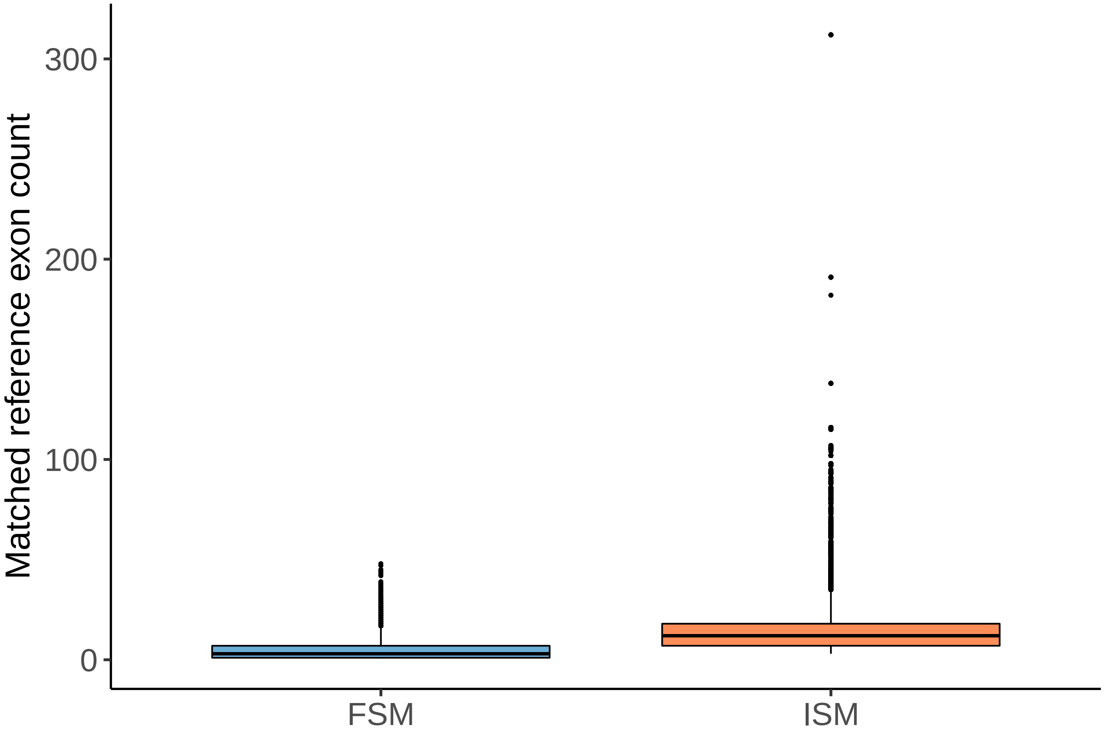
Length Distribution of Matched Reference Transcripts

Applicable Only to FSM and ISM Categories

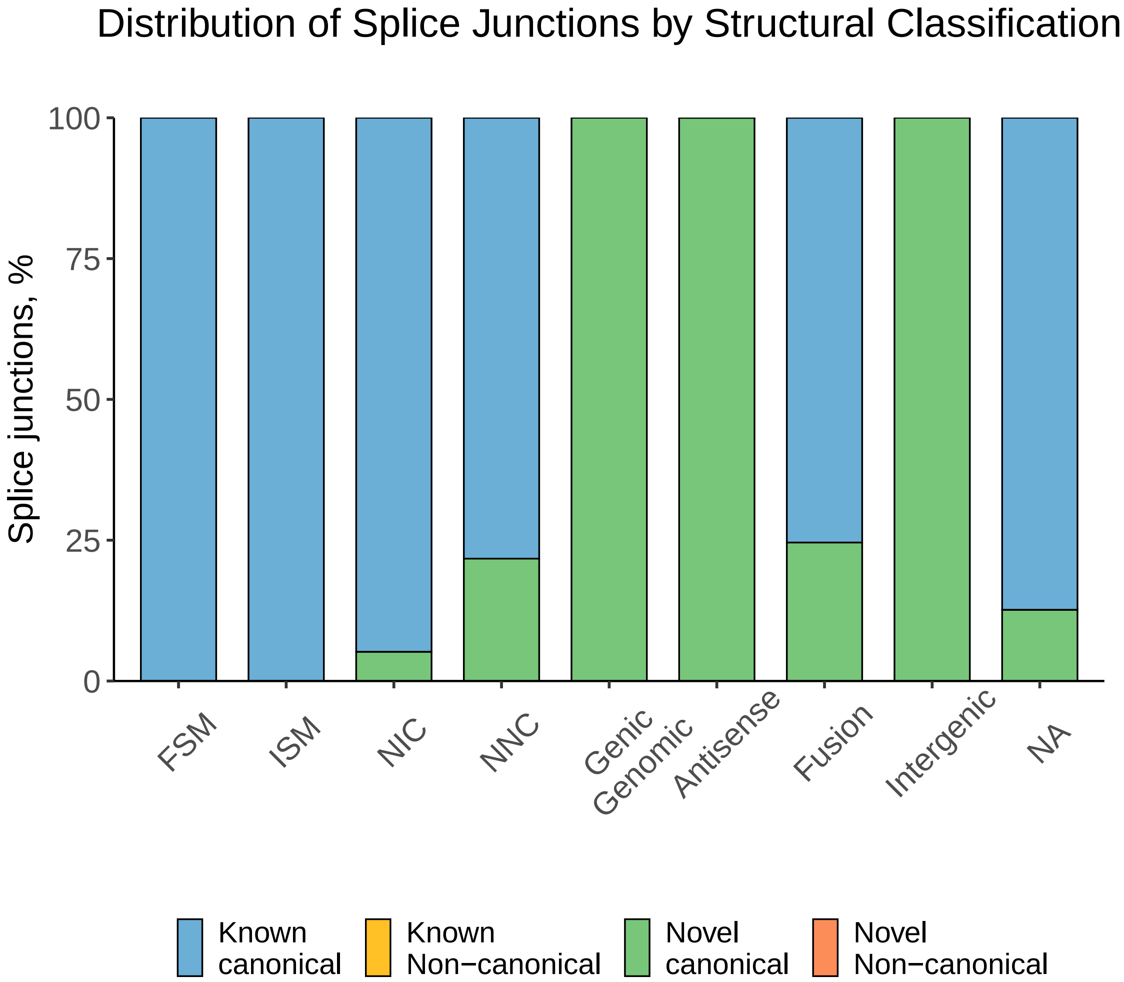


Exon Count Distribution of Matched Reference Transcripts

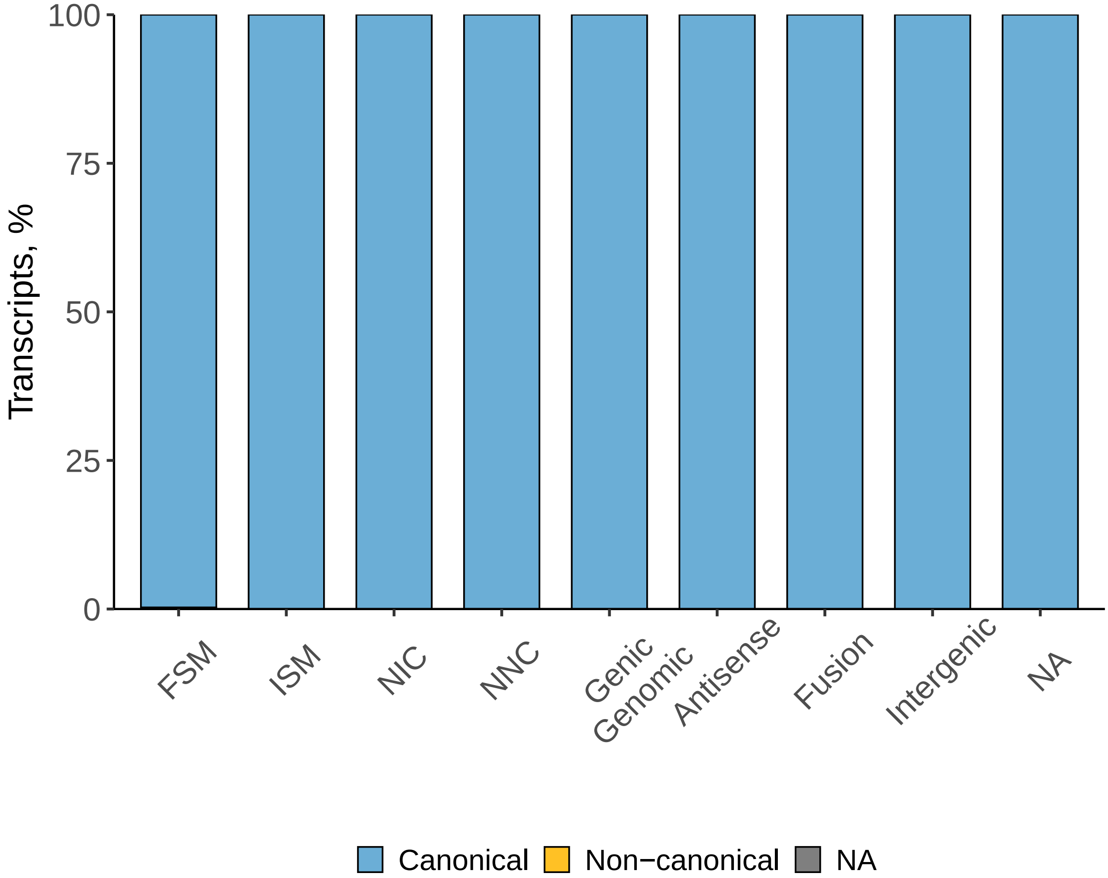
Applicable Only to FSM and ISM Categories

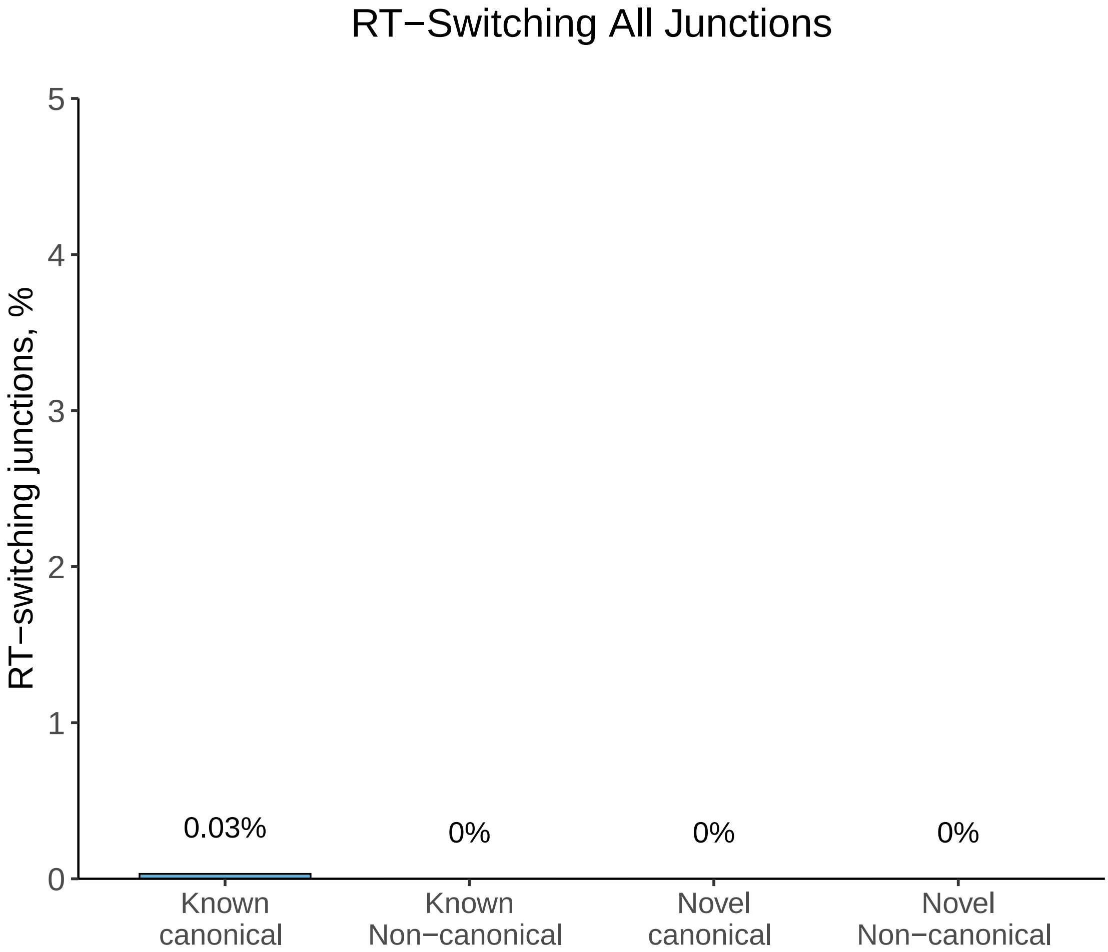


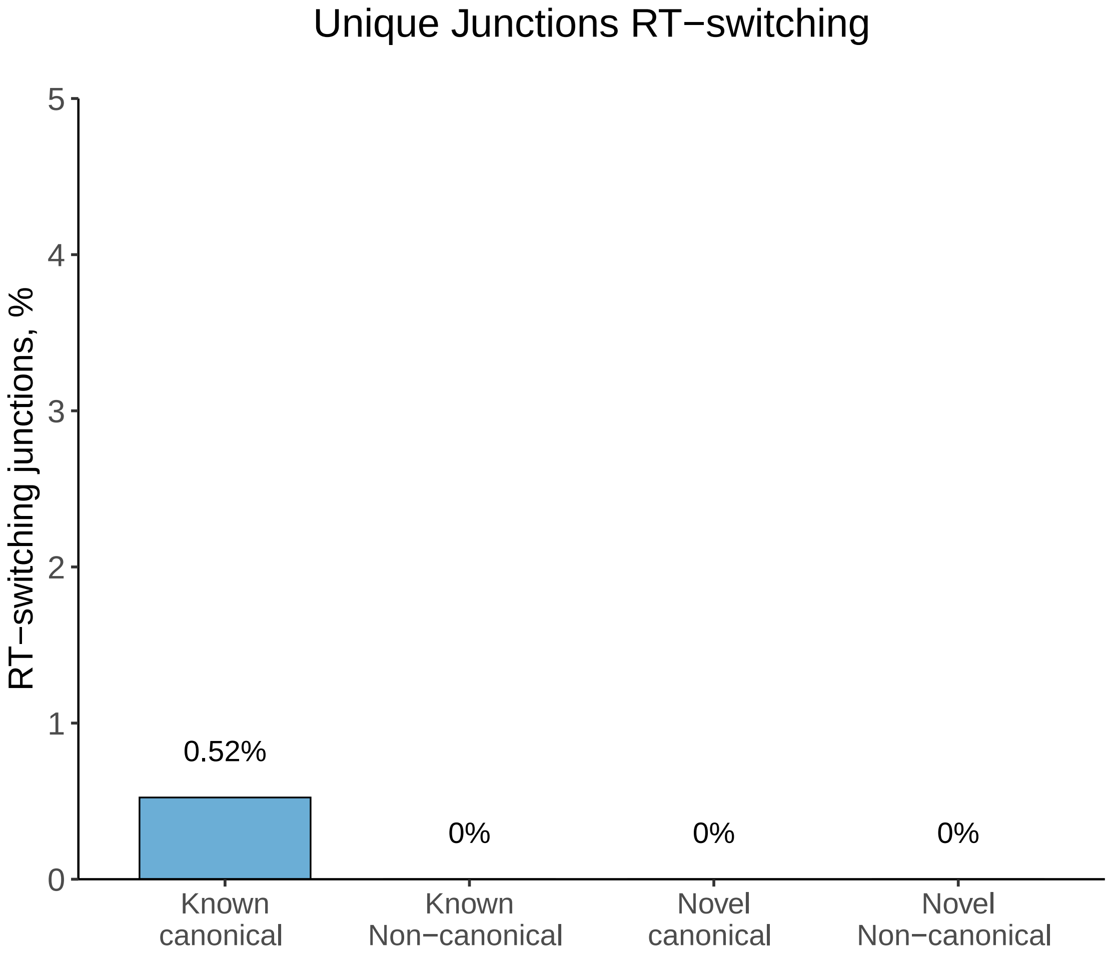
Splice Junction Characterization


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

Distribution of Transcripts by Splice Junctions



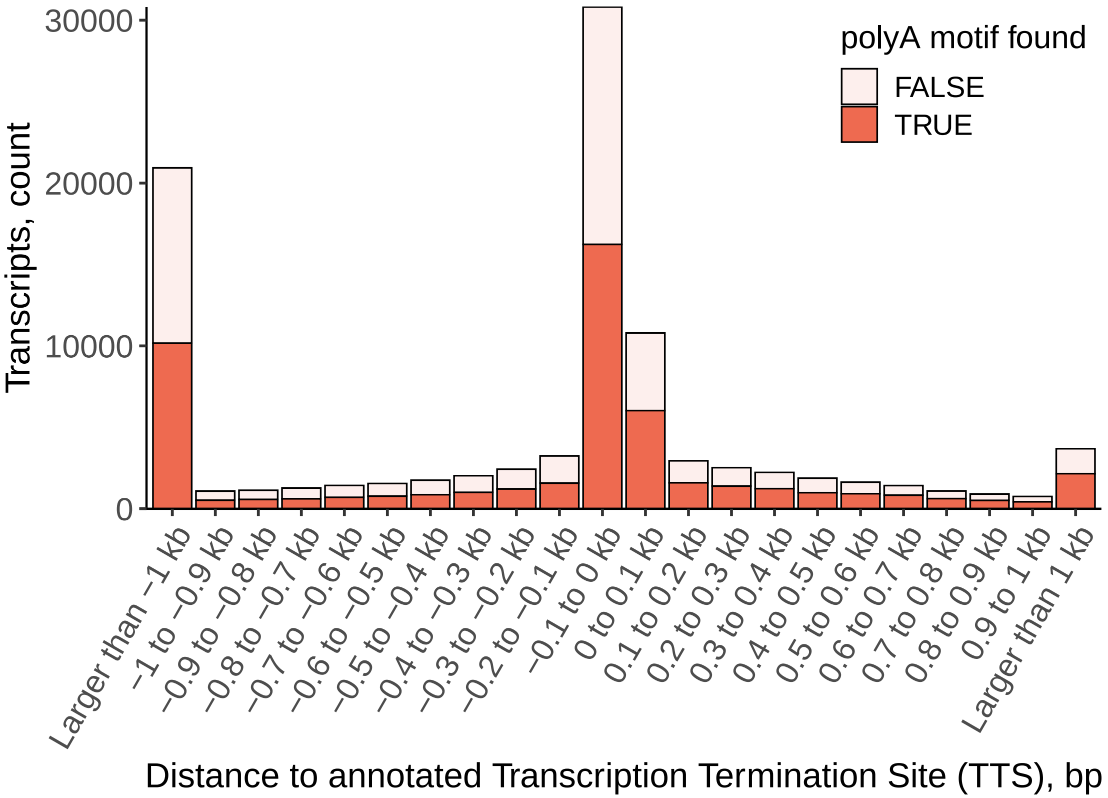

RT−Switching All Junctions
RT−switching junctions, %
5
4
3
2
1
0
0.03%
0%
0%
0%
Known
canonical
Known
Non−canonical
Novel
canonical
Novel
Non−canonical

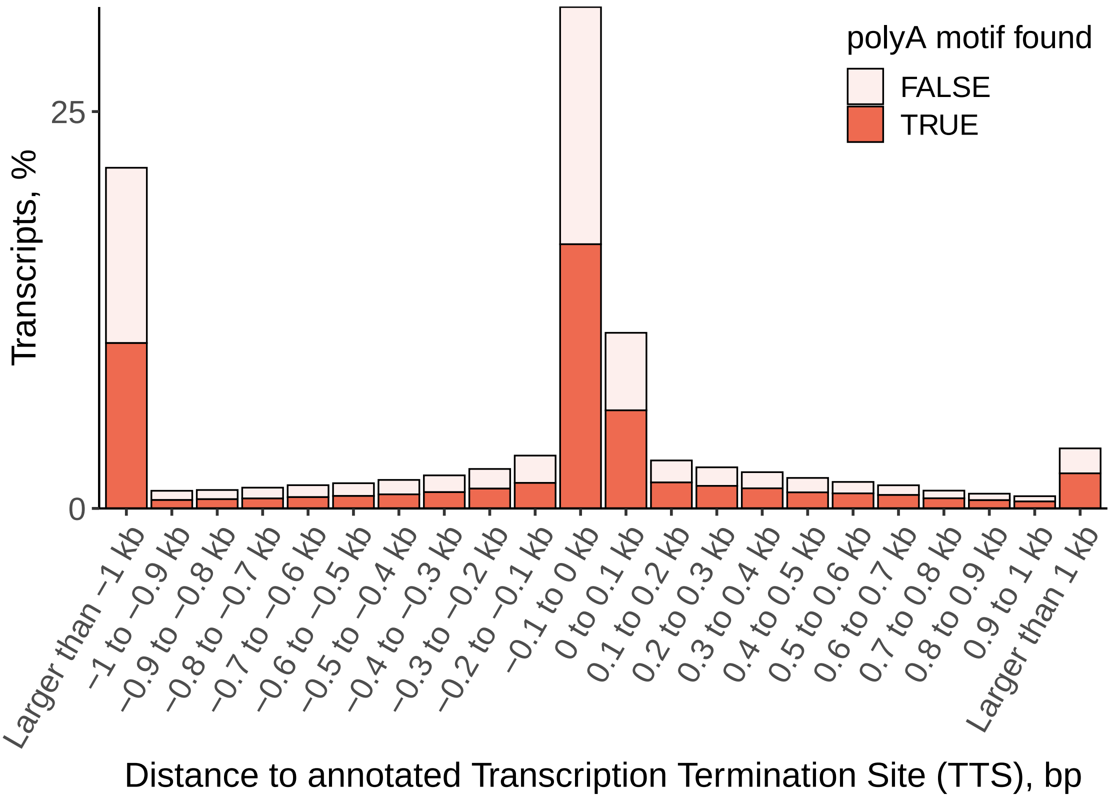

Unique Junctions RT−switching
RT−switching junctions, %
5
4
3
2
1
0
0.52%
0%
0%
0%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

Comparison With Annotated TSS and TTS

Distance to annotated Transcription Termination Site (TTS)
FSM

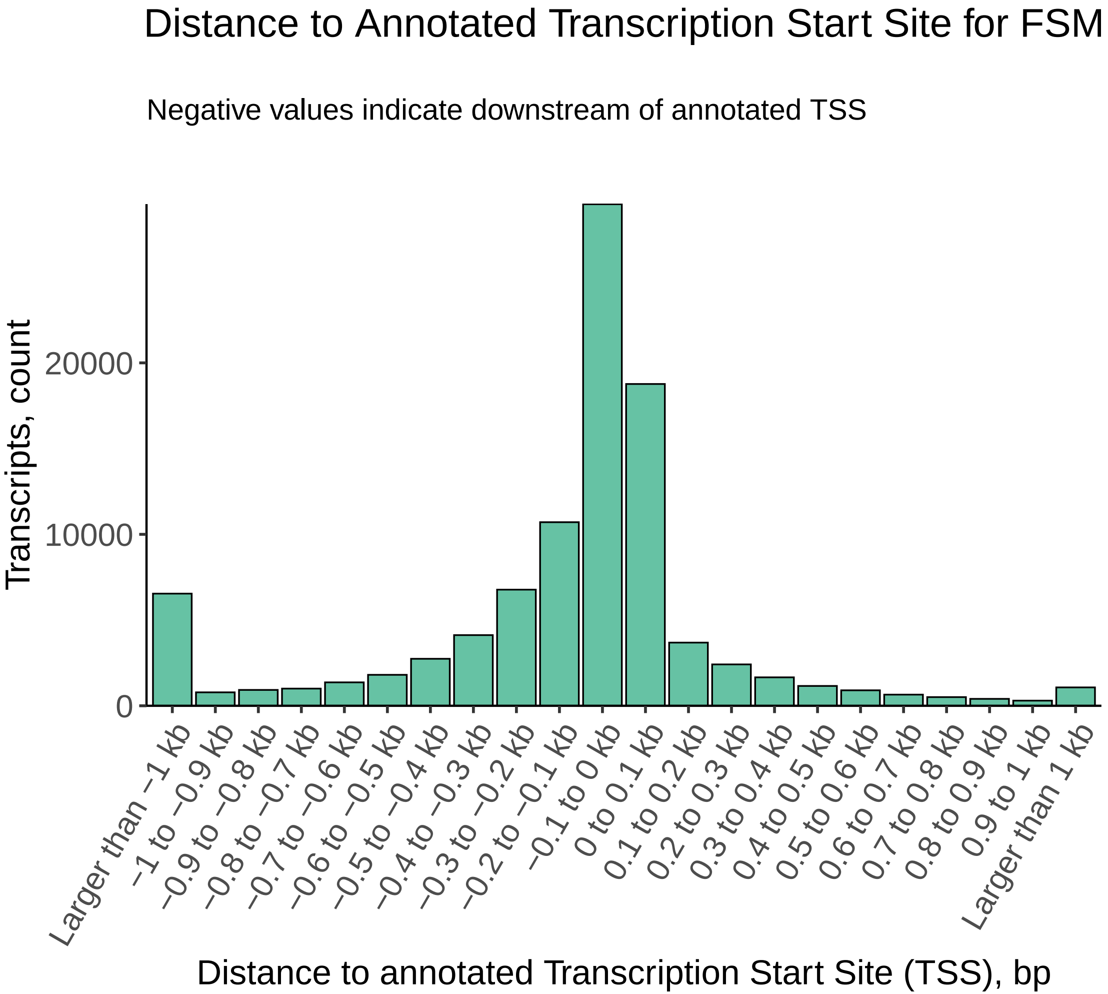
Negative values indicate upstream of annotated termination site





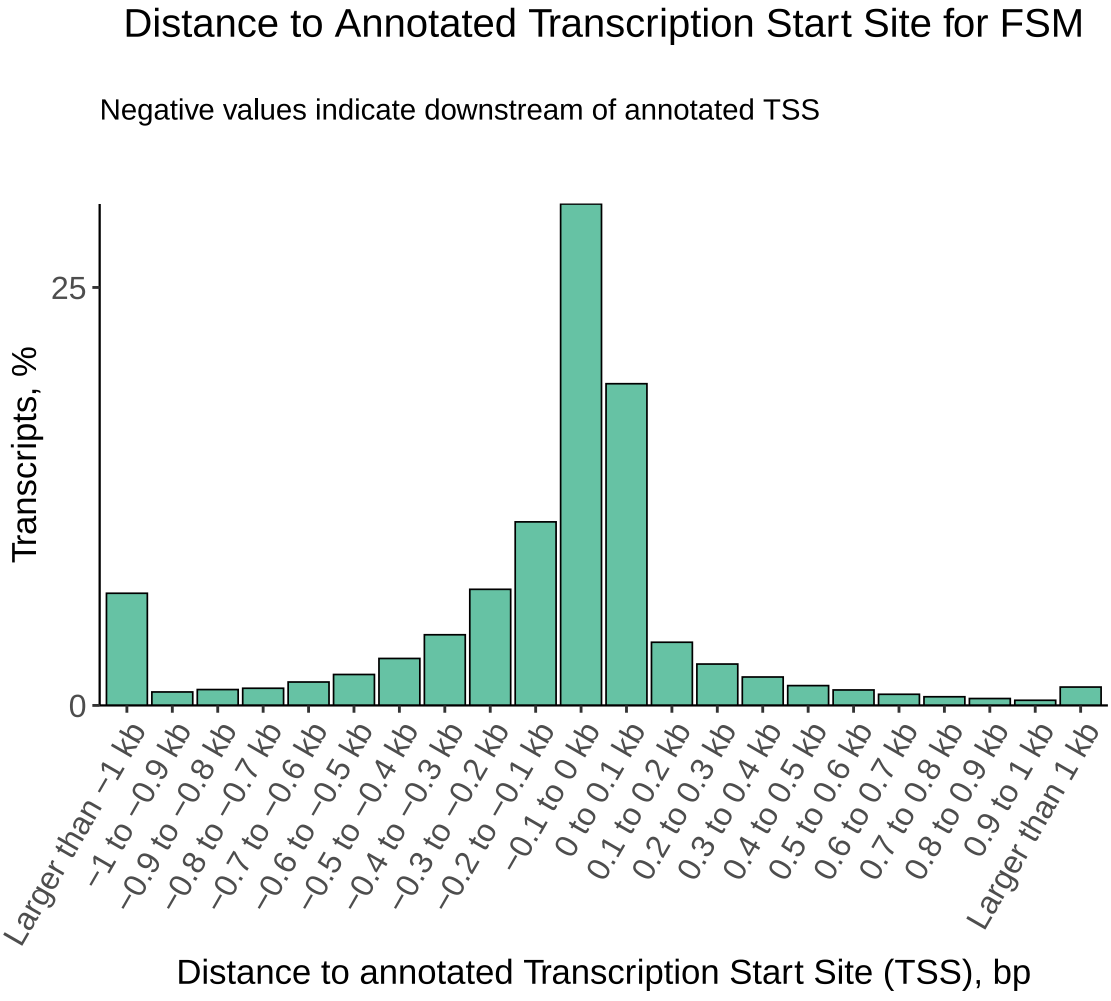
Distance to annotated Transcription Termination Site (TTS)
FSM

Negative values indicate upstream of annotated termination site



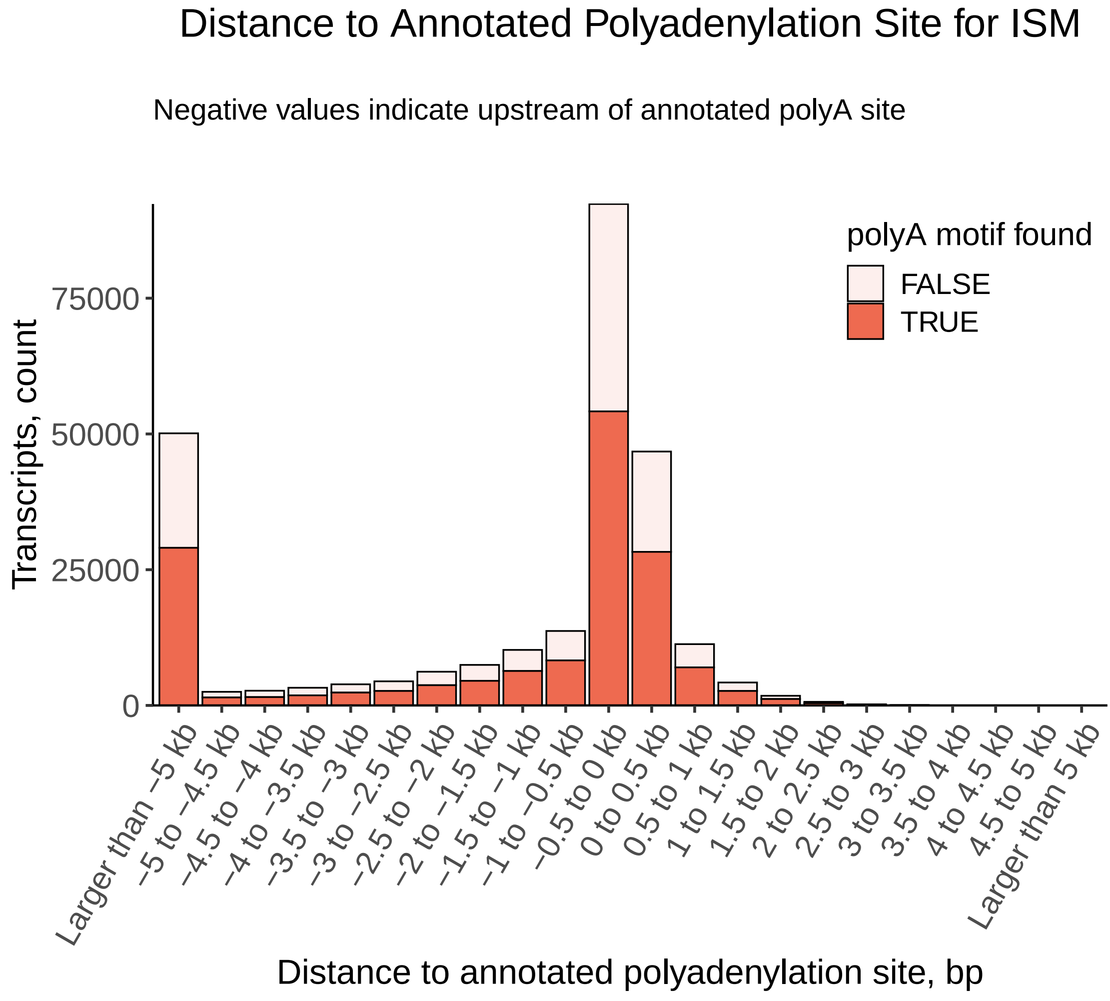
Distance to Annotated Transcription Start Site for FSM

Negative values indicate downstream of annotated TSS



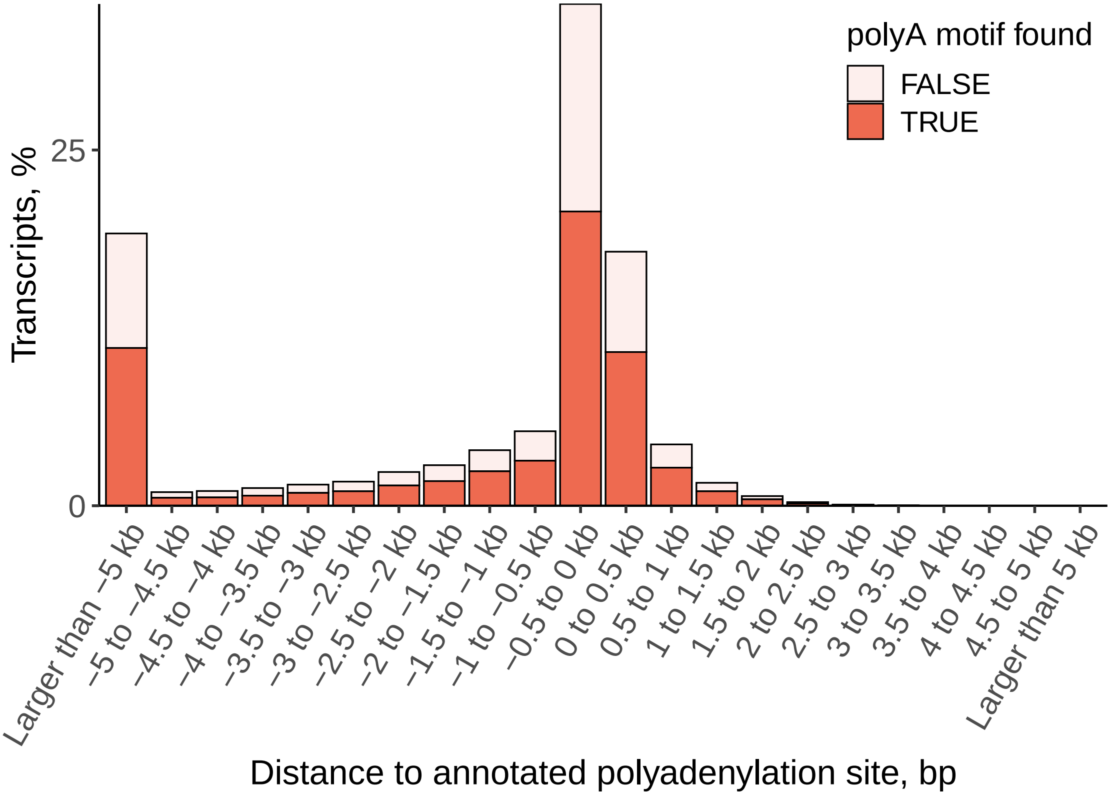
Distance to Annotated Transcription Start Site for FSM

Negative values indicate downstream of annotated TSS

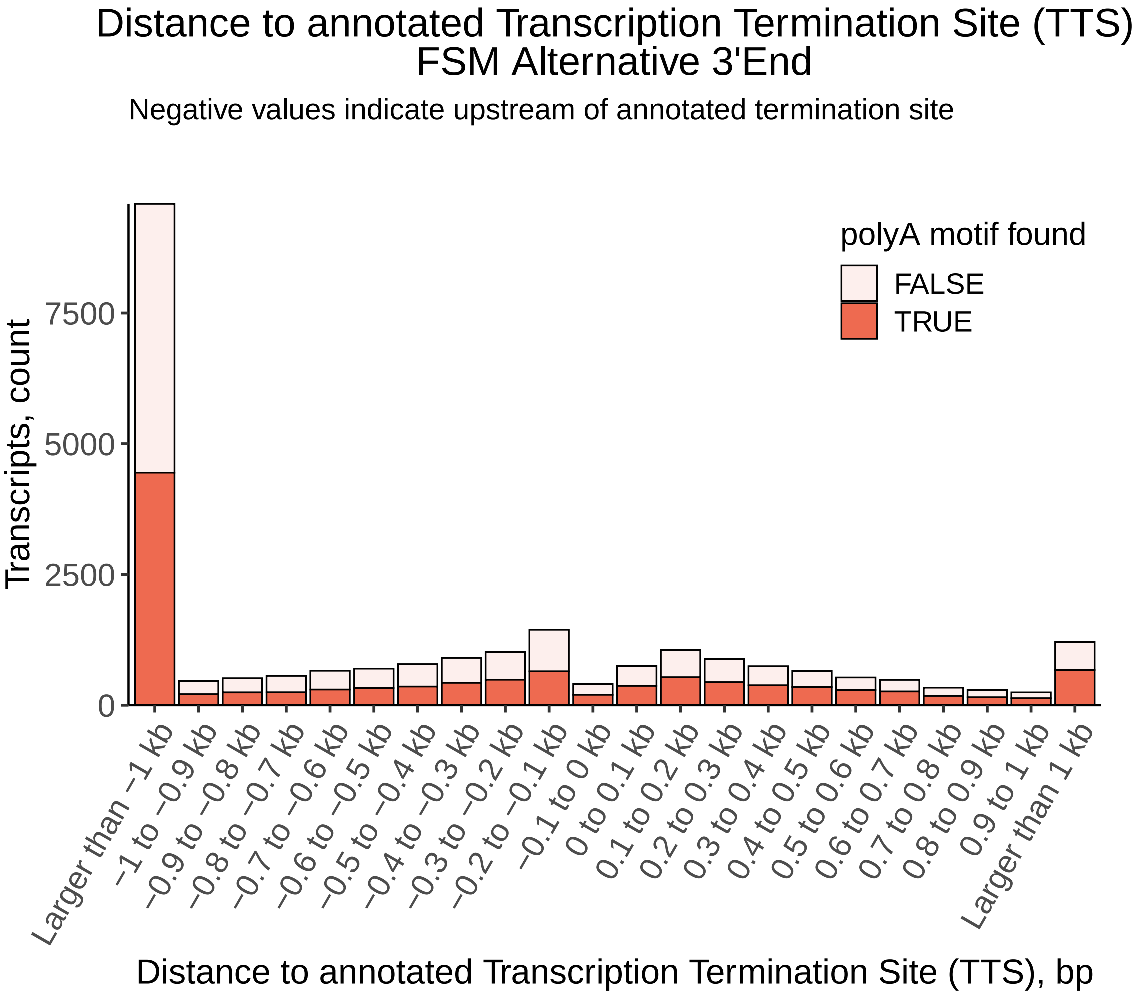

Distance to Annotated Polyadenylation Site for ISM
Negative values indicate upstream of annotated polyA site
polyA motif found
FALSE
TRUE
Transcripts, count
75000
50000
25000
0
Larger than −5 kb
−5 to −4.5 kb
−4.5 to −4 kb
−4 to −3.5 kb
−3.5 to −3 kb
−3 to −2.5 kb
−2.5 to −2 kb
−2 to −1.5 kb
−1.5 to −1 kb
−1 to −0.5 kb
−0.5 to 0 kb
0 to 0.5 kb
0.5 to 1 kb
1 to 1.5 kb
1.5 to 2 kb
2 to 2.5 kb
2.5 to 3 kb
3 to 3.5 kb
3.5 to 4 kb
4 to 4.5 kb
4.5 to 5 kb
Larger than 5 kb
Distance to annotated polyadenylation site, bp

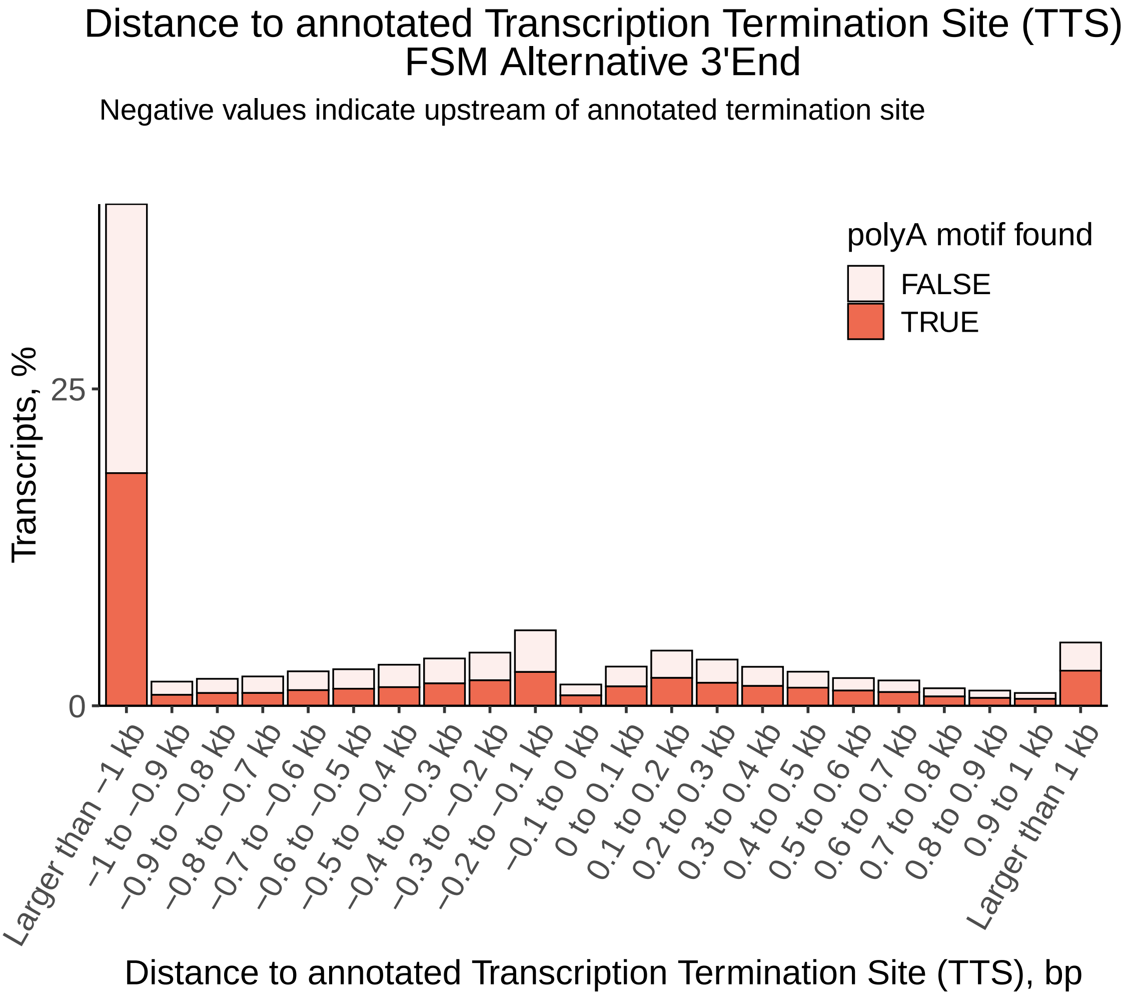
Distance to Annotated Polyadenylation Site for ISM

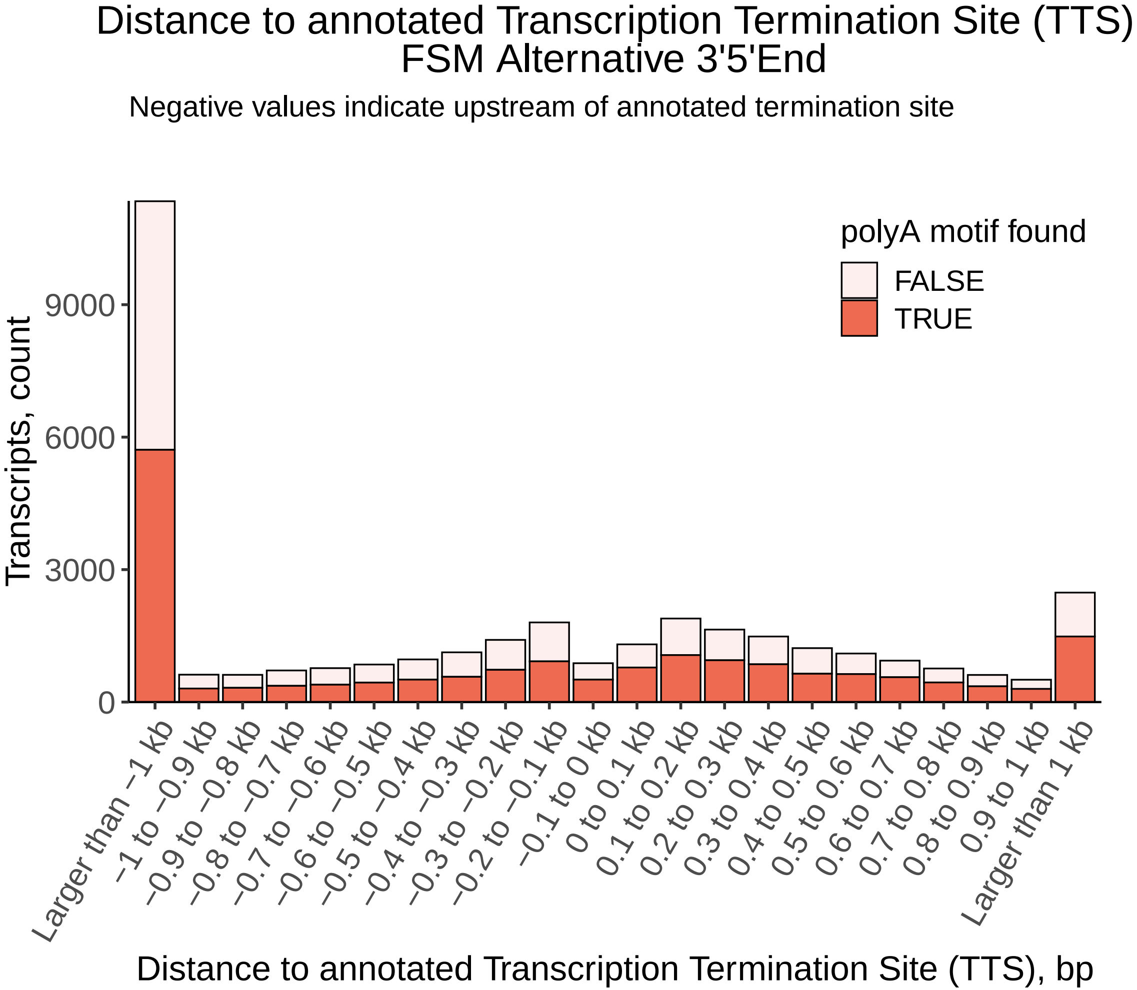
Negative values indicate upstream of annotated polyA site

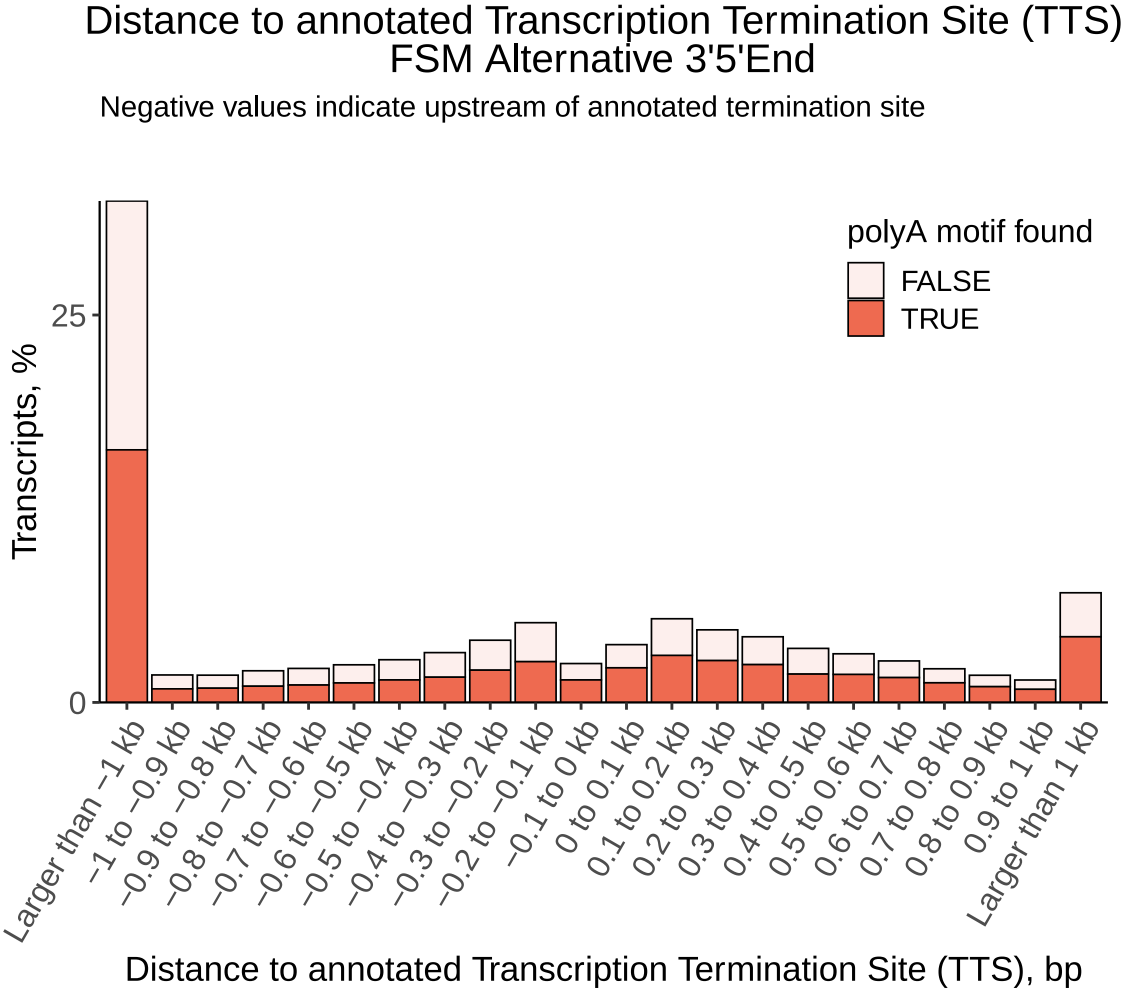


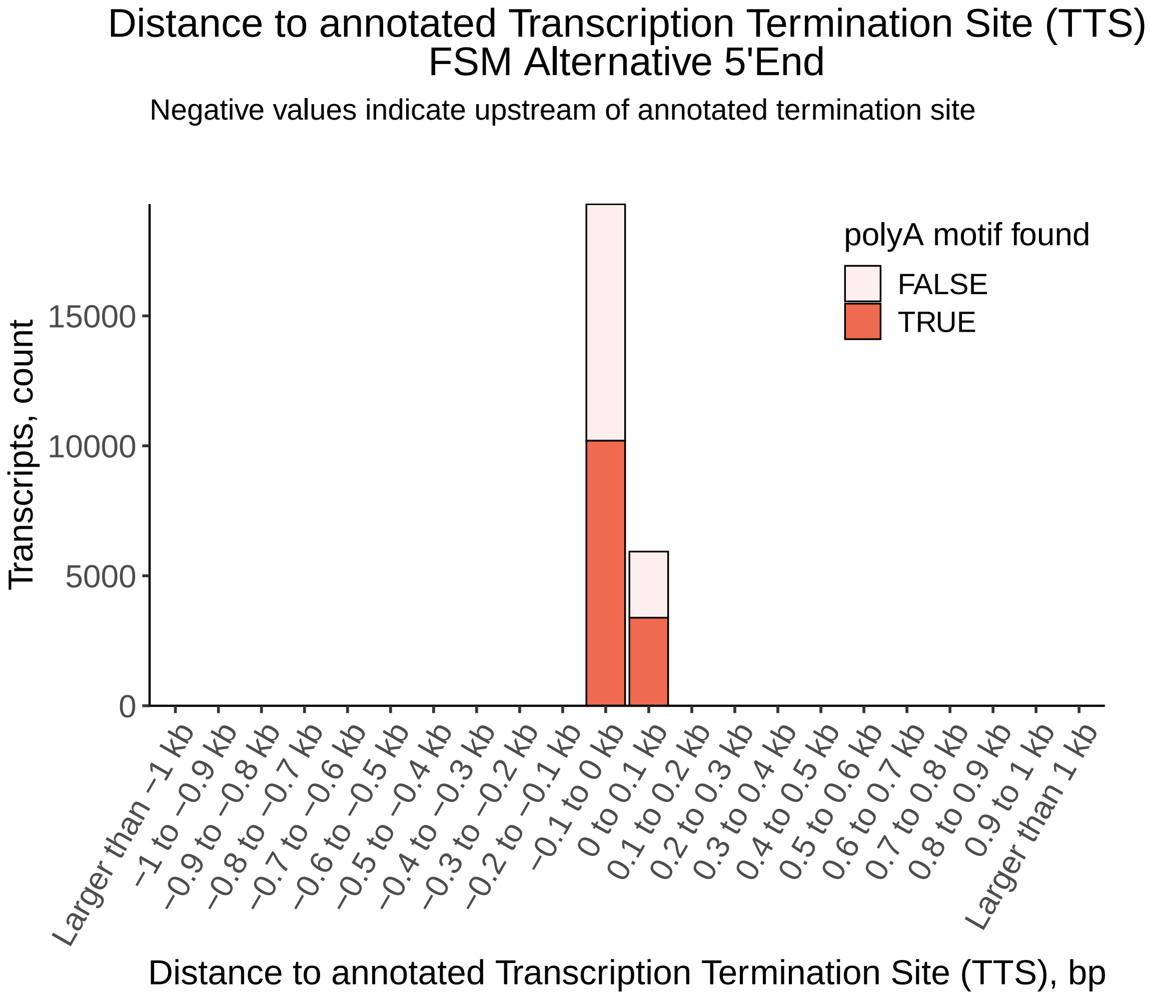
Comparison With Annotated TSS and TTS by Subcategories


Distance to annotated Transcription Termination Site (TTS)
FSM Alternative 3'End
Negative values indicate upstream of annotated termination site
Transcripts, count
0
2500
5000
7500
polyA motif found
FALSE
TRUE
Larger than −1 kb
−1 to −0.9 kb
−0.9 to −0.8 kb
−0.8 to −0.7 kb
−0.7 to −0.6 kb
−0.6 to −0.5 kb
−0.5 to −0.4 kb
−0.4 to −0.3 kb
−0.3 to −0.2 kb
−0.2 to −0.1 kb
−0.1 to 0 kb
0 to 0.1 kb
0.1 to 0.2 kb
0.2 to 0.3 kb
0.3 to 0.4 kb
0.4 to 0.5 kb
0.5 to 0.6 kb
0.6 to 0.7 kb
0.7 to 0.8 kb
0.8 to 0.9 kb
0.9 to 1 kb
Larger than 1 kb
Distance to annotated Transcription Termination Site (TTS), bp


Distance to annotated Transcription Termination Site (TTS)
FSM Alternative 3'End
Negative values indicate upstream of annotated termination site
polyA motif found
FALSE
TRUE
Transcripts, %
25
0
Larger than −1 kb
−1 to −0.9 kb
−0.9 to −0.8 kb
−0.8 to −0.7 kb
−0.7 to −0.6 kb
−0.6 to −0.5 kb
−0.5 to −0.4 kb
−0.4 to −0.3 kb
−0.3 to −0.2 kb
−0.2 to −0.1 kb
−0.1 to 0 kb
0 to 0.1 kb
0.1 to 0.2 kb
0.2 to 0.3 kb
0.3 to 0.4 kb
0.4 to 0.5 kb
0.5 to 0.6 kb
0.6 to 0.7 kb
0.7 to 0.8 kb
0.8 to 0.9 kb
0.9 to 1 kb
Larger than 1 kb
Distance to annotated Transcription Termination Site (TTS), bp

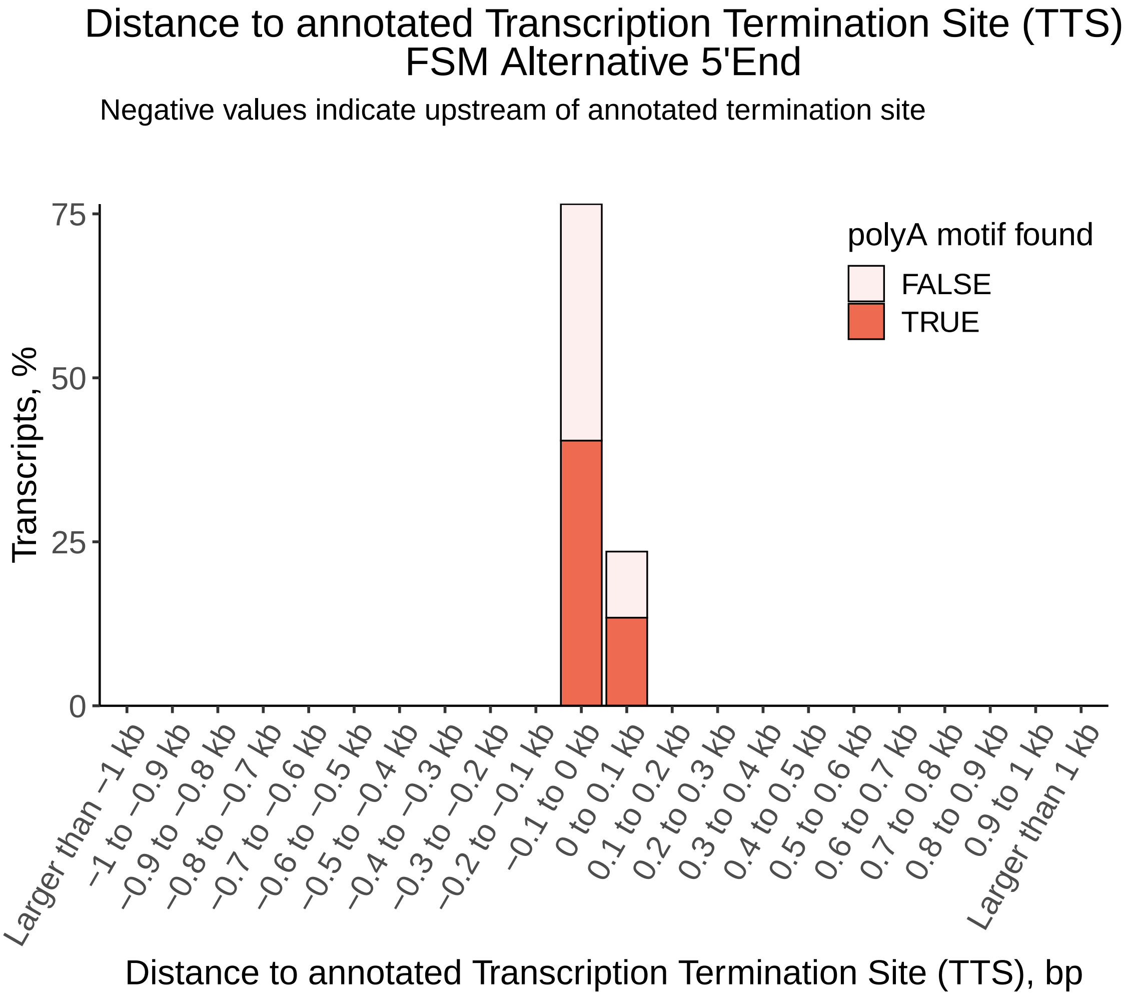

Distance to annotated Transcription Termination Site (TTS)
FSM Alternative 3'5'End
Negative values indicate upstream of annotated termination site
polyA motif found
FALSE
TRUE
Transcripts, count
0
3000
6000
9000
Larger than −1 kb
−1 to −0.9 kb
−0.9 to −0.8 kb
−0.8 to −0.7 kb
−0.7 to −0.6 kb
−0.6 to −0.5 kb
−0.5 to −0.4 kb
−0.4 to −0.3 kb
−0.3 to −0.2 kb
−0.2 to −0.1 kb
−0.1 to 0 kb
0 to 0.1 kb
0.1 to 0.2 kb
0.2 to 0.3 kb
0.3 to 0.4 kb
0.4 to 0.5 kb
0.5 to 0.6 kb
0.6 to 0.7 kb
0.7 to 0.8 kb
0.8 to 0.9 kb
0.9 to 1 kb
Larger than 1 kb
Distance to annotated Transcription Termination Site (TTS), bp

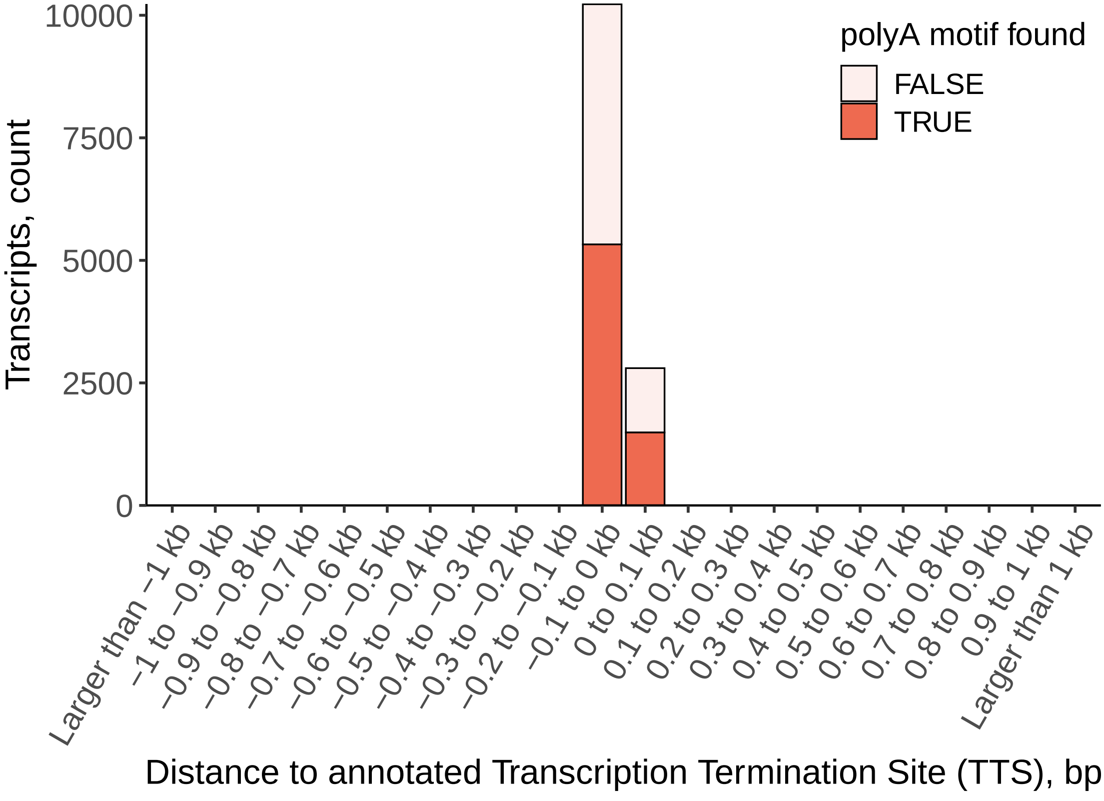

Distance to annotated Transcription Termination Site (TTS)
FSM Alternative 3'5'End
Negative values indicate upstream of annotated termination site
polyA motif found
FALSE
TRUE
Transcripts, %
25
0
Larger than −1 kb
−1 to −0.9 kb
−0.9 to −0.8 kb
−0.8 to −0.7 kb
−0.7 to −0.6 kb
−0.6 to −0.5 kb
−0.5 to −0.4 kb
−0.4 to −0.3 kb
−0.3 to −0.2 kb
−0.2 to −0.1 kb
−0.1 to 0 kb
0 to 0.1 kb
0.1 to 0.2 kb
0.2 to 0.3 kb
0.3 to 0.4 kb
0.4 to 0.5 kb
0.5 to 0.6 kb
0.6 to 0.7 kb
0.7 to 0.8 kb
0.8 to 0.9 kb
0.9 to 1 kb
Larger than 1 kb
Distance to annotated Transcription Termination Site (TTS), bp



Distance to annotated Transcription Termination Site (TTS)
FSM Alternative 5'End

Negative values indicate upstream of annotated termination site


Distance to annotated Transcription Termination Site (TTS)
FSM Alternative 5'End
Negative values indicate upstream of annotated termination site
Transcripts, %
75
50
25
0
polyA motif found
FALSE
TRUE
Larger than −1 kb
−1 to −0.9 kb
−0.9 to −0.8 kb
−0.8 to −0.7 kb
−0.7 to −0.6 kb
−0.6 to −0.5 kb
−0.5 to −0.4 kb
−0.4 to −0.3 kb
−0.3 to −0.2 kb
−0.2 to −0.1 kb
−0.1 to 0 kb
0 to 0.1 kb
0.1 to 0.2 kb
0.2 to 0.3 kb
0.3 to 0.4 kb
0.4 to 0.5 kb
0.5 to 0.6 kb
0.6 to 0.7 kb
0.7 to 0.8 kb
0.8 to 0.9 kb
0.9 to 1 kb
Larger than 1 kb
Distance to annotated Transcription Termination Site (TTS), bp

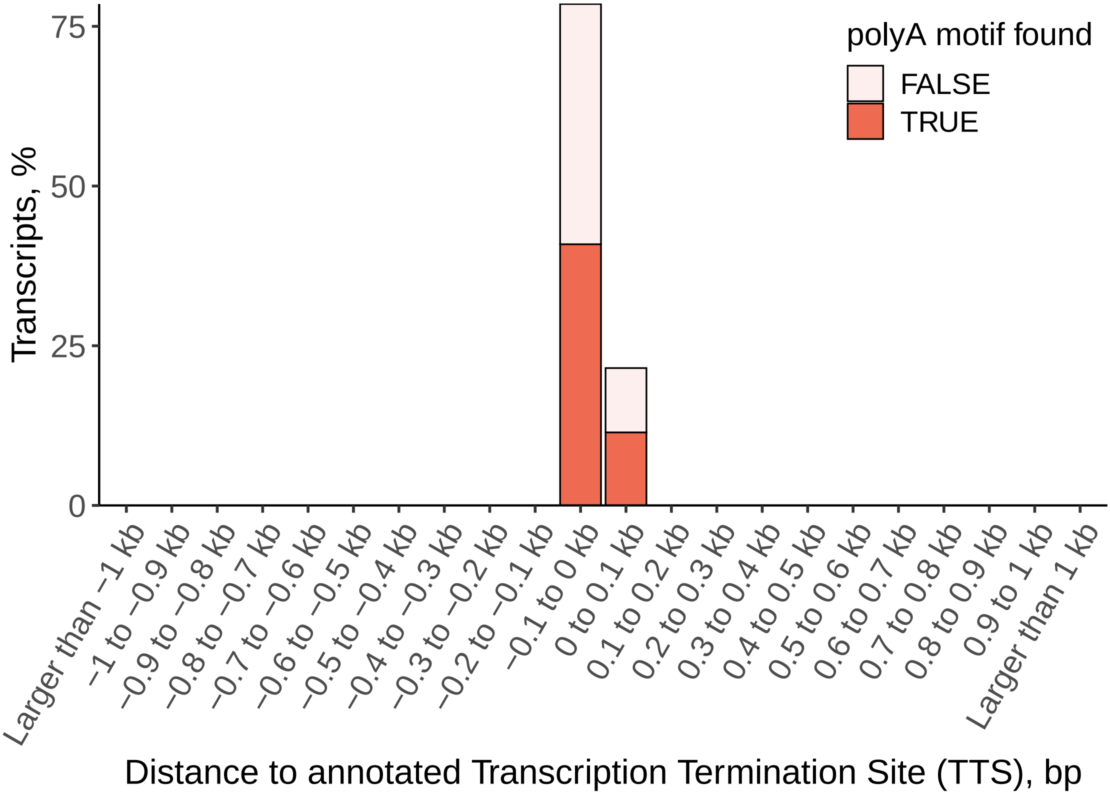


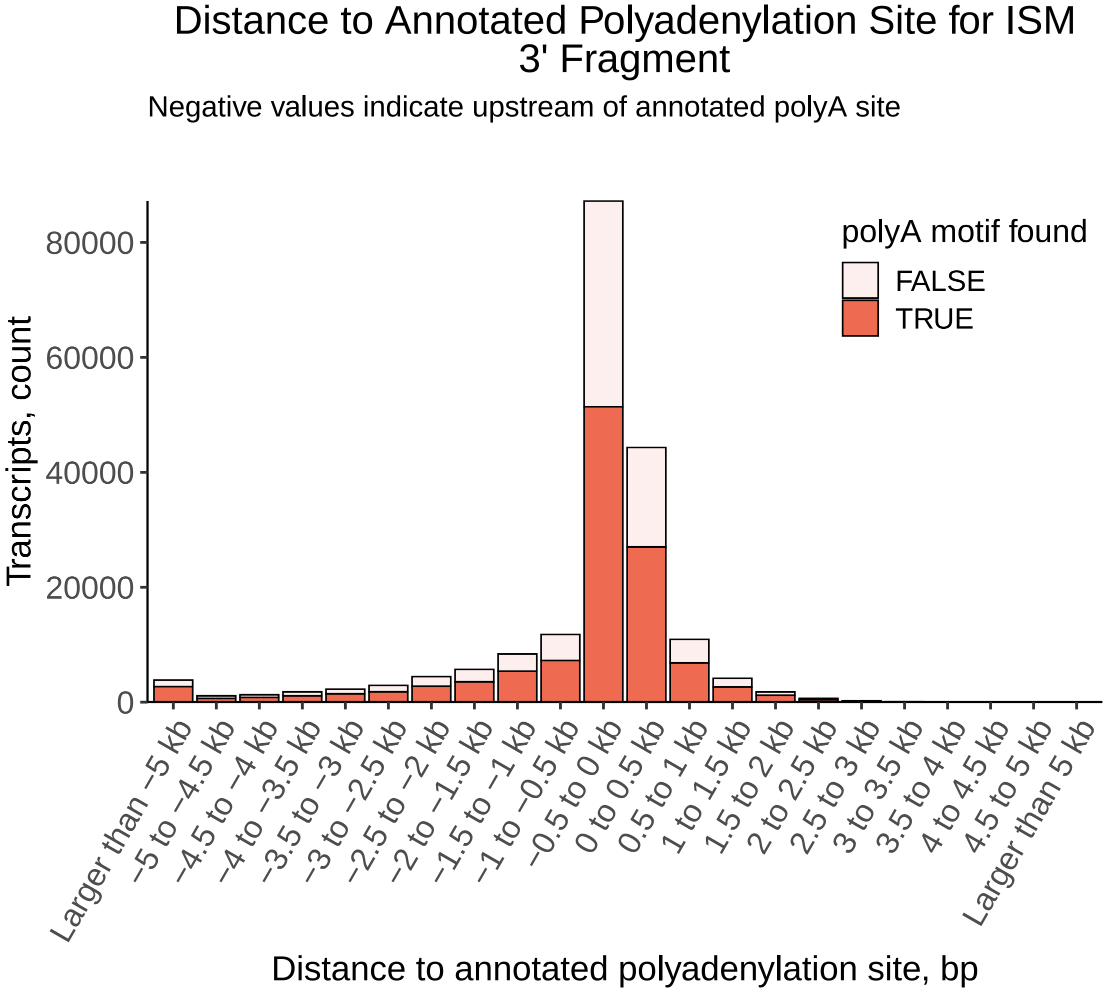
Distance to annotated Transcription Termination Site (TTS)
FSM Reference Match

Negative values indicate upstream of annotated termination site

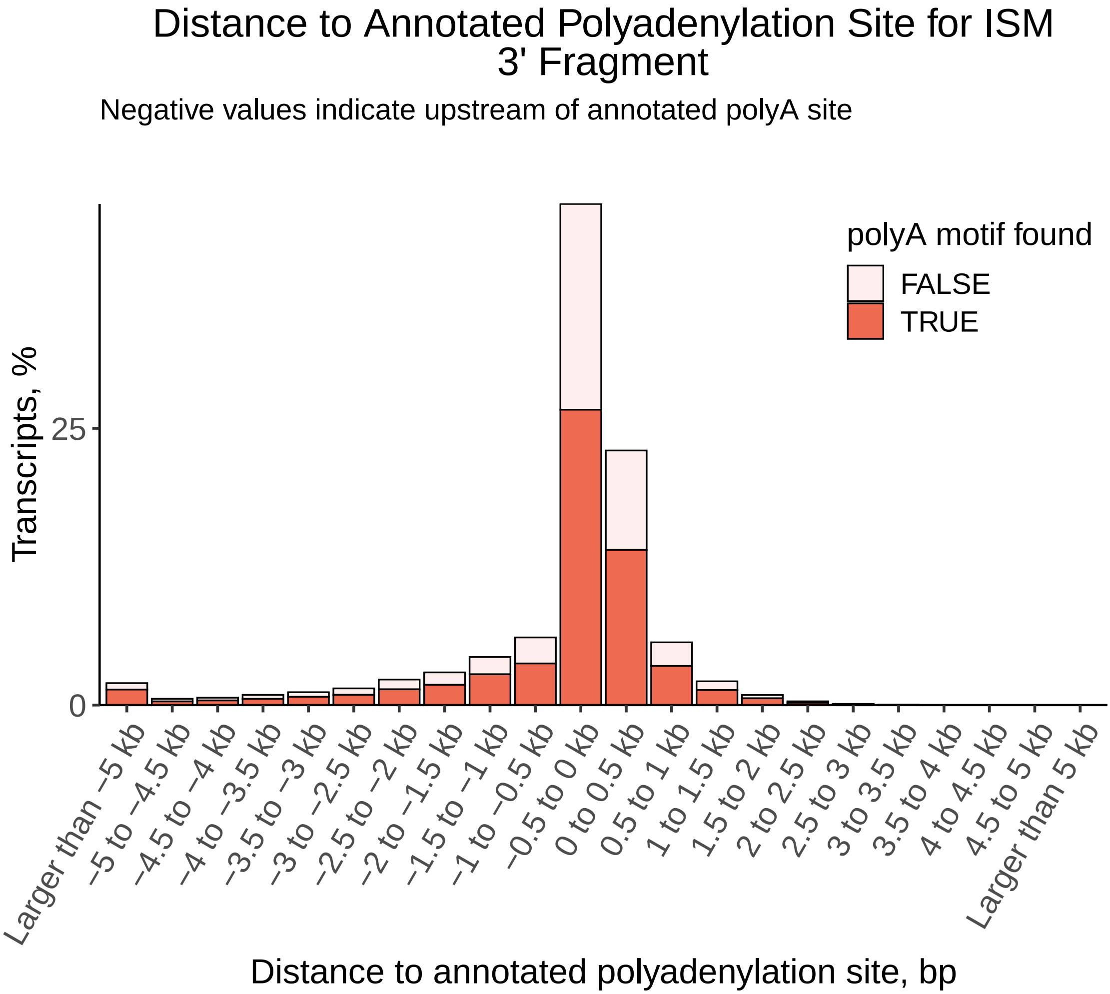
Distance to annotated Transcription Termination Site (TTS)
FSM Reference Match

Negative values indicate upstream of annotated termination site



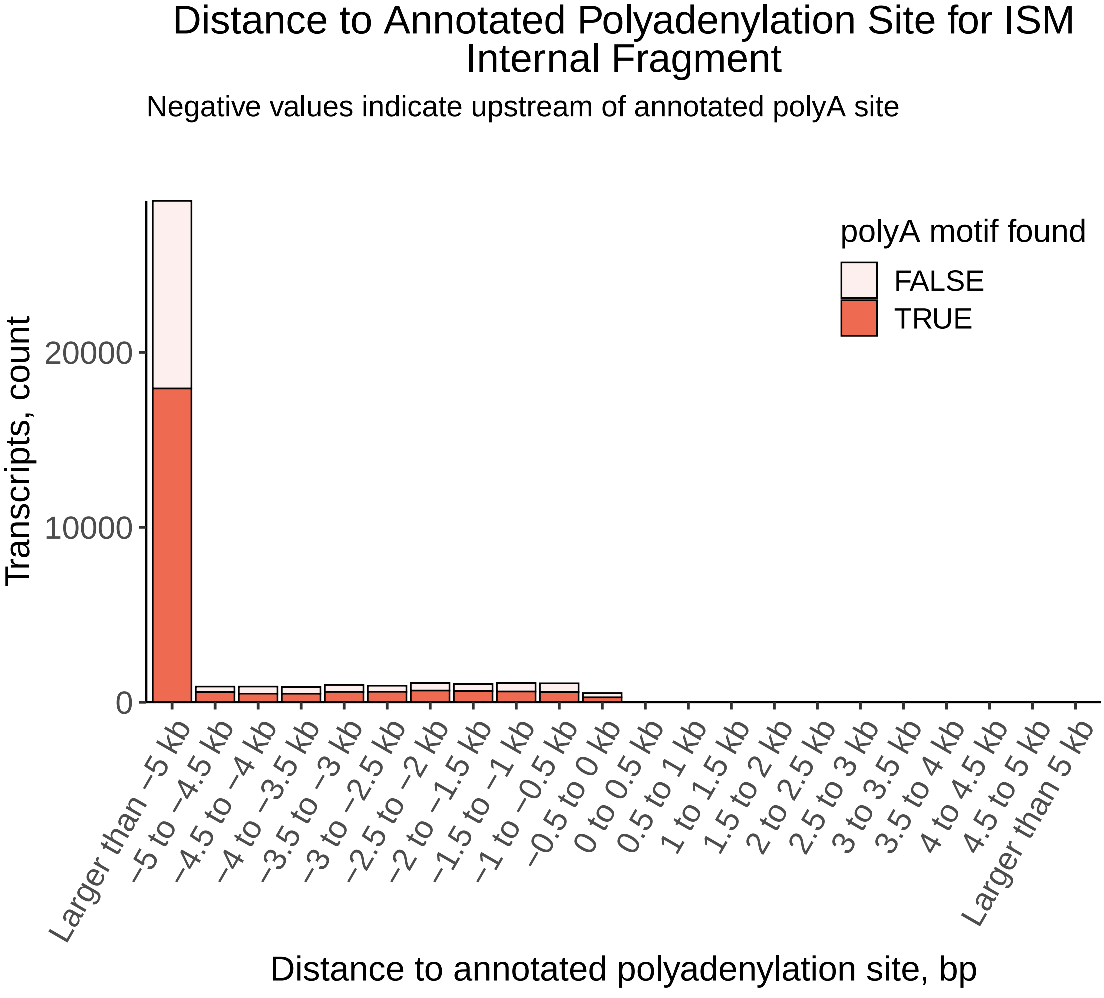

Distance to Annotated Polyadenylation Site for ISM 3' Fragment

Negative values indicate upstream of annotated polyA site



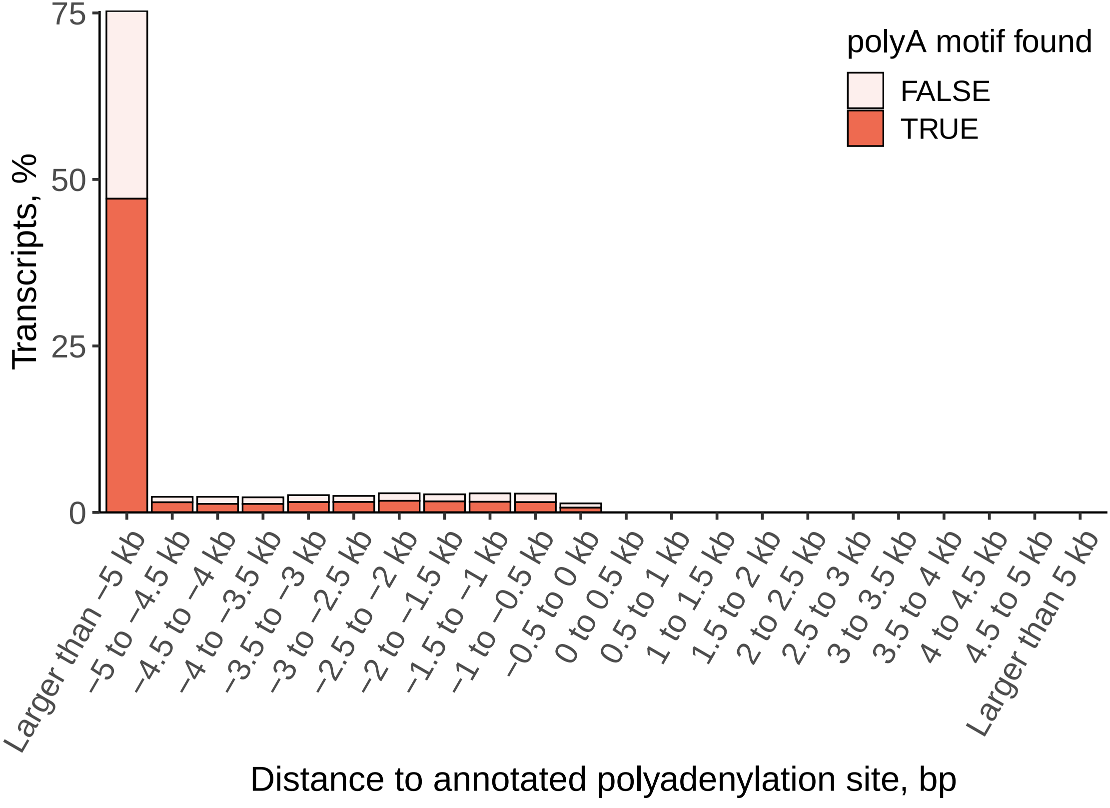
Distance to Annotated Polyadenylation Site for ISM 3' Fragment

Negative values indicate upstream of annotated polyA site



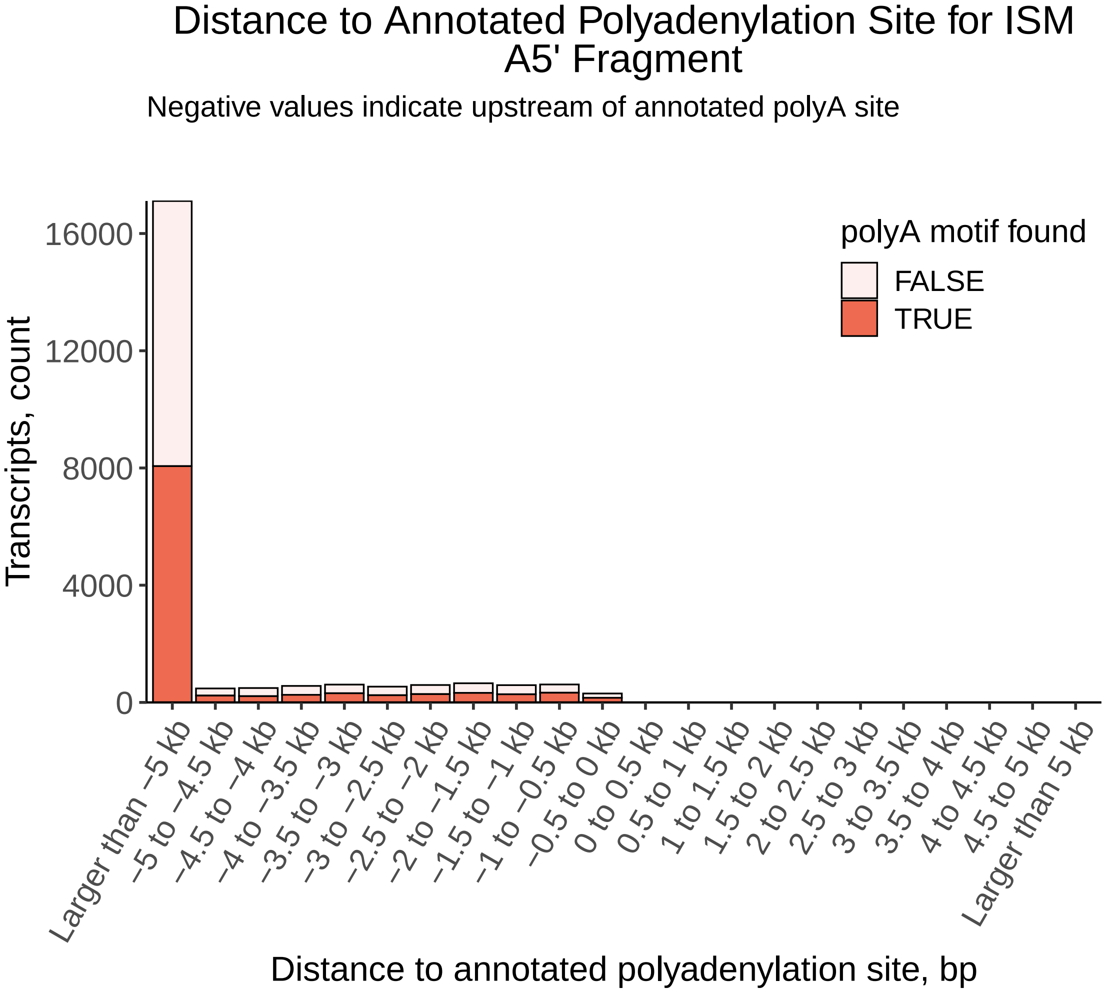
Distance to Annotated Polyadenylation Site for ISM Internal Fragment

Negative values indicate upstream of annotated polyA site



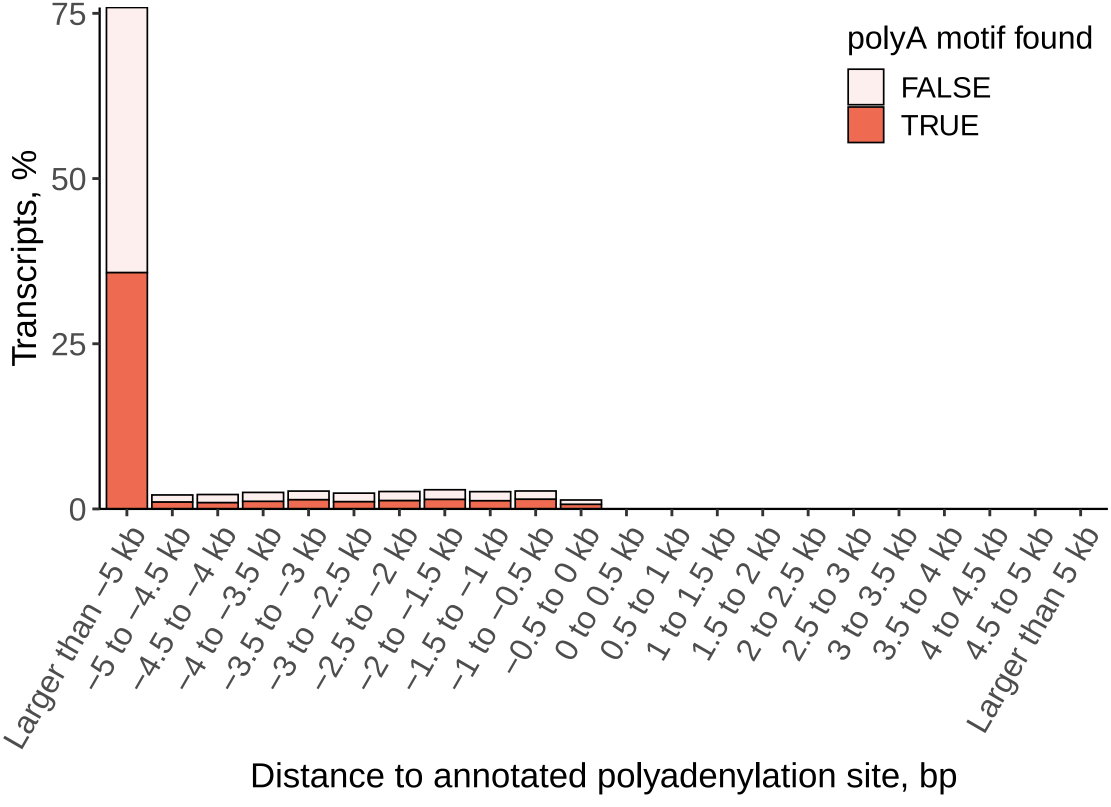
Distance to Annotated Polyadenylation Site for ISM Internal Fragment

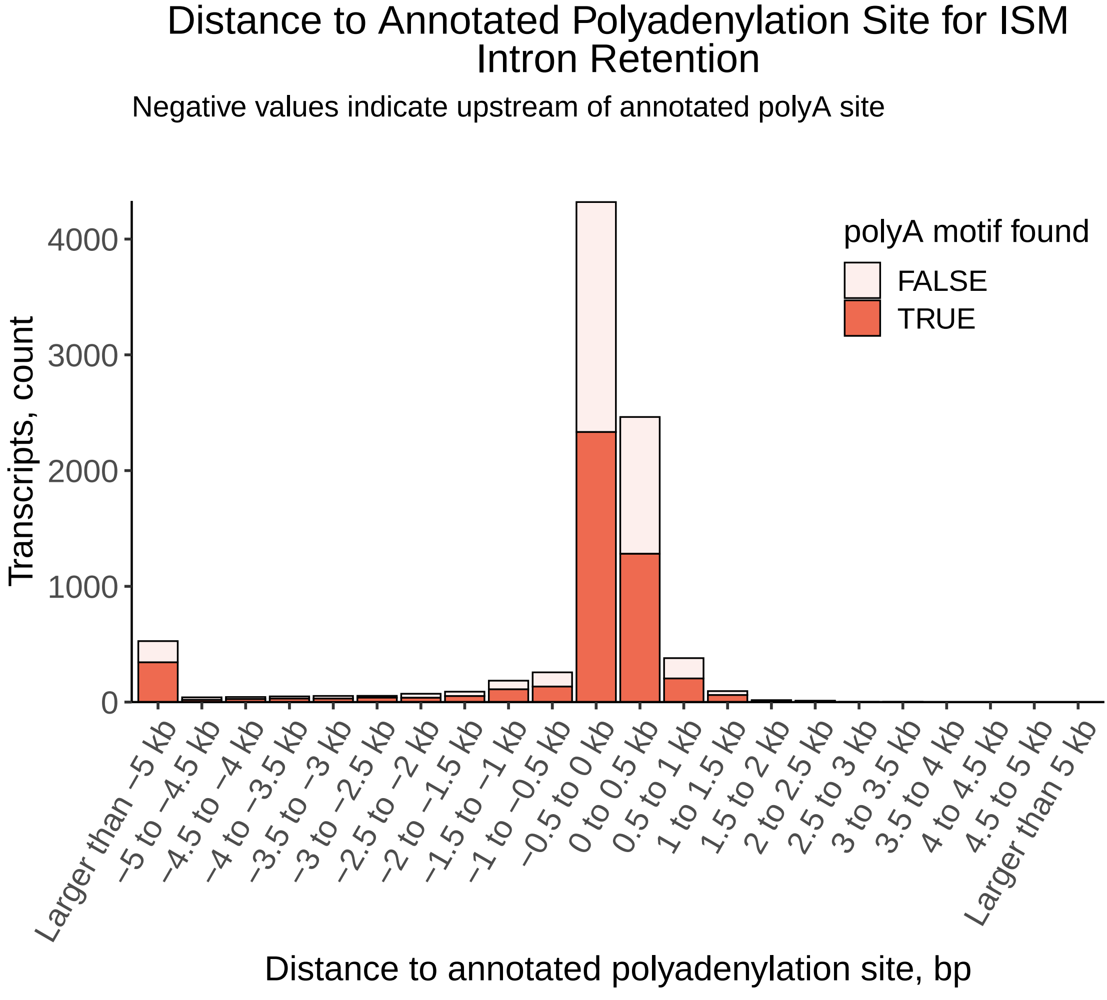
Negative values indicate upstream of annotated polyA site

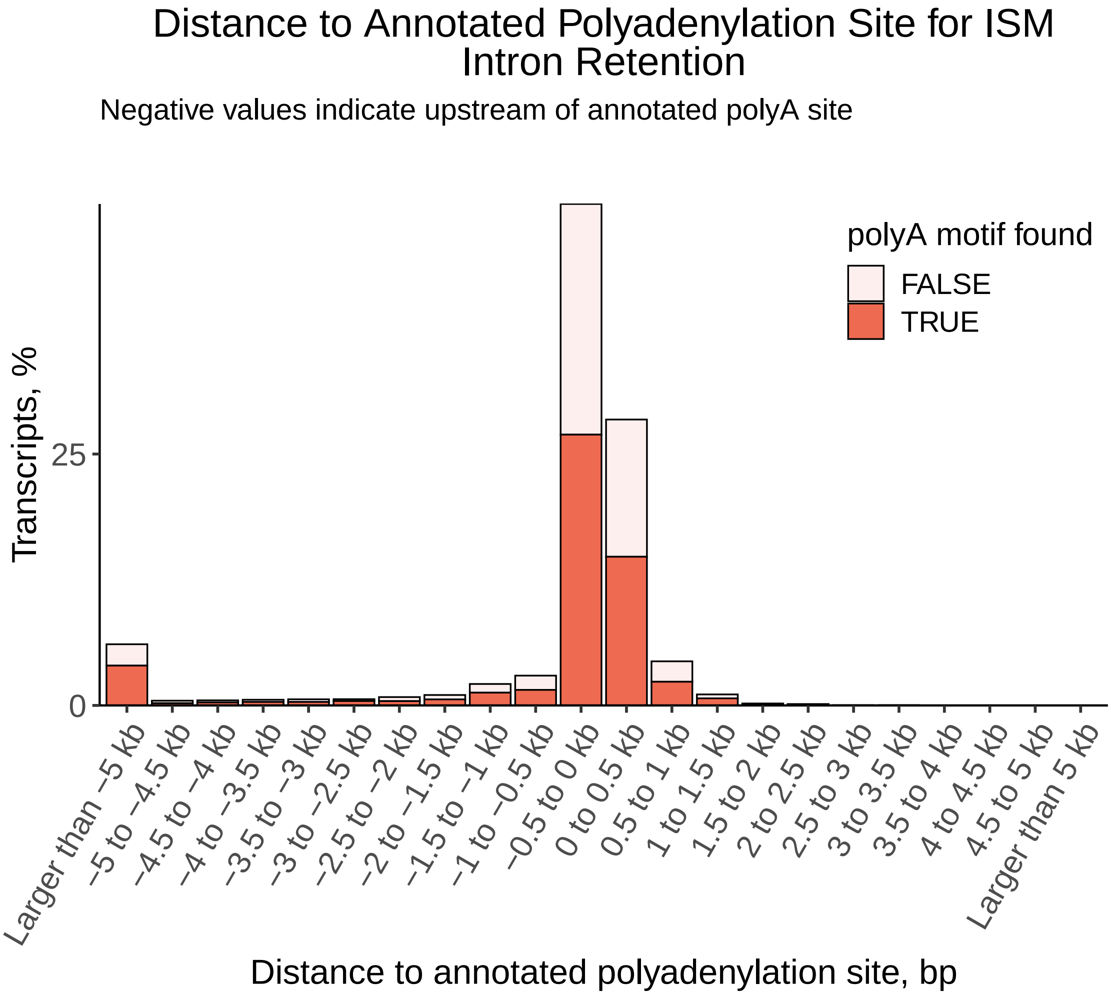

Distance to Annotated Polyadenylation Site for ISM A5' Fragment
Negative values indicate upstream of annotated polyA site
Transcripts, count
0
4000
8000
12000
16000
polyA motif found
FALSE
TRUE
Larger than −5 kb
−5 to −4.5 kb
−4.5 to −4 kb
−4 to −3.5 kb
−3.5 to −3 kb
−3 to −2.5 kb
−2.5 to −2 kb
−2 to −1.5 kb
−1.5 to −1 kb
−1 to −0.5 kb
−0.5 to 0 kb
0 to 0.5 kb
0.5 to 1 kb
1 to 1.5 kb
1.5 to 2 kb
2 to 2.5 kb
2.5 to 3 kb
3 to 3.5 kb
3.5 to 4 kb
4 to 4.5 kb
4.5 to 5 kb
Larger than 5 kb
Distance to annotated polyadenylation site, bp

Distance to Annotated Polyadenylation Site for ISM A5' Fragment

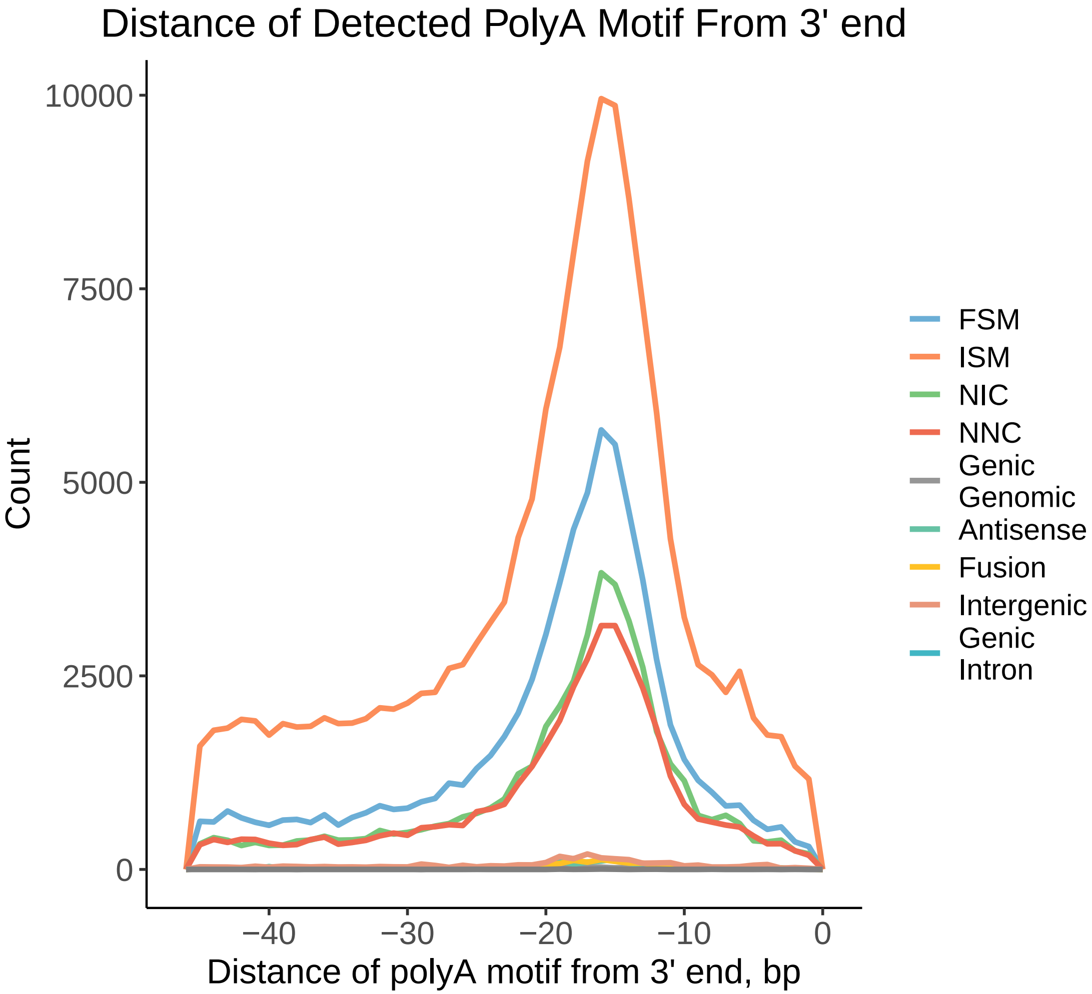
Negative values indicate upstream of annotated polyA site




Distance to Annotated Polyadenylation Site for ISM Intron Retention
Negative values indicate upstream of annotated polyA site
Transcripts, count
0
1000
2000
3000
4000
polyA motif found
FALSE
TRUE
Larger than −5 kb
−5 to −4.5 kb
−4.5 to −4 kb
−4 to −3.5 kb
−3.5 to −3 kb
−3 to −2.5 kb
−2.5 to −2 kb
−2 to −1.5 kb
−1.5 to −1 kb
−1 to −0.5 kb
−0.5 to 0 kb
0 to 0.5 kb
0.5 to 1 kb
1 to 1.5 kb
1.5 to 2 kb
2 to 2.5 kb
2.5 to 3 kb
3 to 3.5 kb
3.5 to 4 kb
4 to 4.5 kb
4.5 to 5 kb
Larger than 5 kb
Distance to annotated polyadenylation site, bp


Distance to Annotated Polyadenylation Site for ISM Intron Retention
Negative values indicate upstream of annotated polyA site
Transcripts, %
25
0
polyA motif found
FALSE
TRUE
Larger than −5 kb
−5 to −4.5 kb
−4.5 to −4 kb
−4 to −3.5 kb
−3.5 to −3 kb
−3 to −2.5 kb
−2.5 to −2 kb
−2 to −1.5 kb
−1.5 to −1 kb
−1 to −0.5 kb
−0.5 to 0 kb
0 to 0.5 kb
0.5 to 1 kb
1 to 1.5 kb
1.5 to 2 kb
2 to 2.5 kb
2.5 to 3 kb
3 to 3.5 kb
3.5 to 4 kb
4 to 4.5 kb
4.5 to 5 kb
Larger than 5 kb
Distance to annotated polyadenylation site, bp

PolyA Distance Analysis

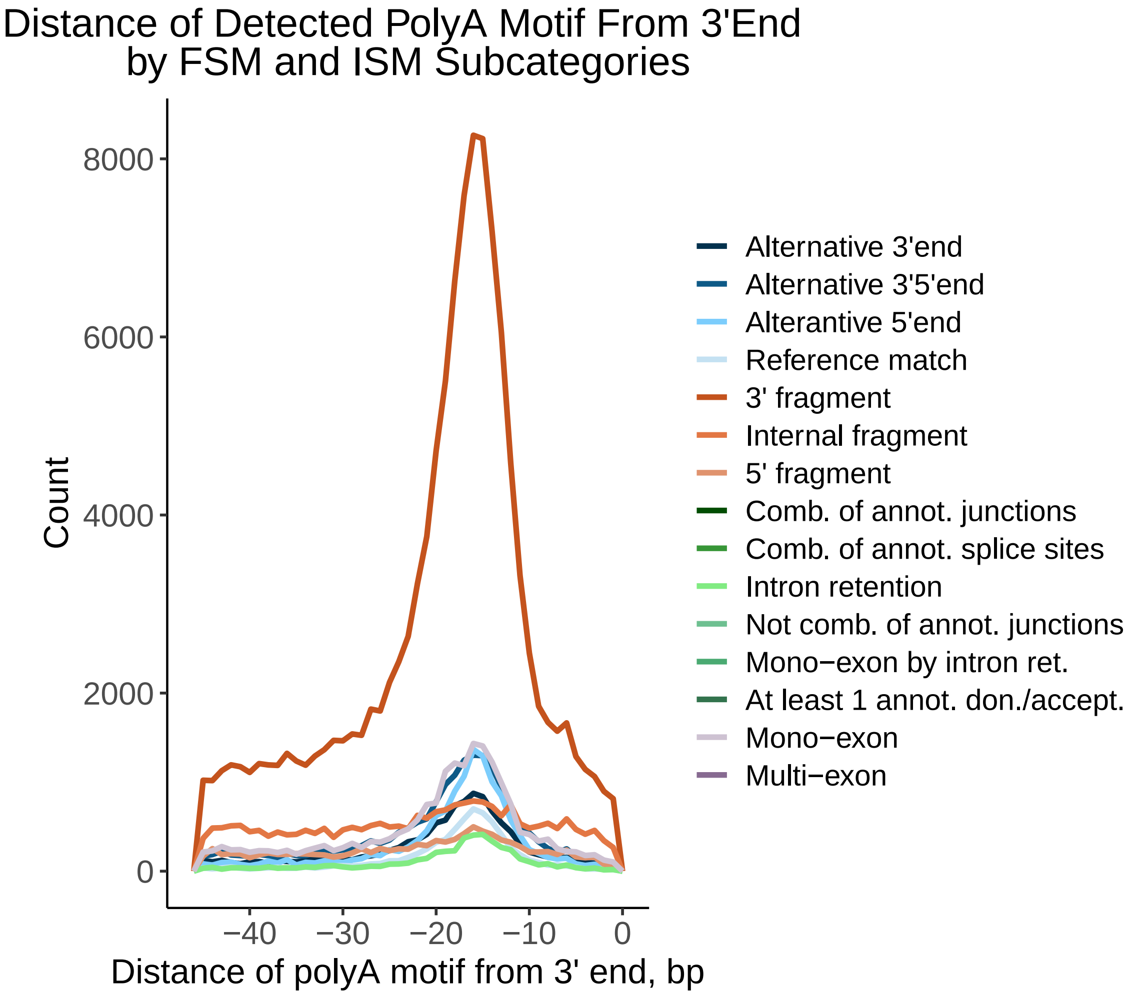

Distance of Detected PolyA Motif From 3' end
Count
10000
7500
5000
2500
0
−40
−30
−20
−10
0
Distance of polyA motif from 3' end, bp
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

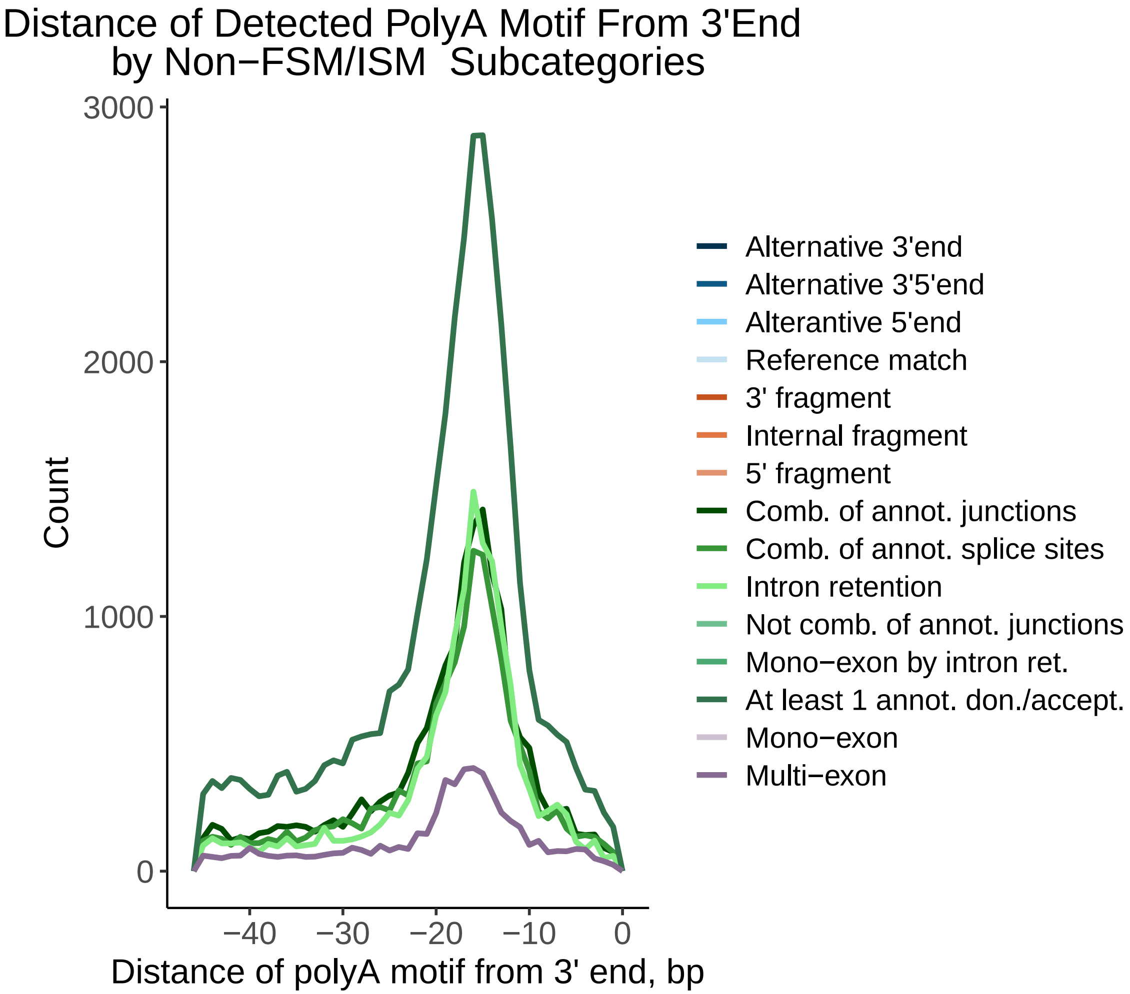
Number of polyA Motifs Detected

Category	Count	polyA Detected	%
FSM	130505	71418	55
ISM	261895	155713	59
NIC	83814	44729	53
NNC	72591	40828	56
Genic Genomic	724	427	59
Antisense	2241	1385	62
Fusion	2407	1436	60
Intergenic	3896	2596	67
NA	40	25	62

Frequency of PolyA Motifs

Motif	Count	%
AATAAA	139528	43.8
ATTAAA	43072	13.5
AAGAAA	15768	4.9
AAAAAG	15014	4.7
TATAAA	14380	4.5
AGTAAA	12920	4.1
TTTAAA	10705	3.4
CATAAA	8527	2.7
AATATA	8510	2.7
AAAACA	8476	2.7
GATAAA	8184	2.6
AATACA	8181	2.6
GGGGCT	8002	2.5
AATGAA	7938	2.5
AATAGA	4956	1.6
ACTAAA	4396	1.4


Distance of Detected PolyA Motif From 3'End
by FSM and ISM Subcategories
Count
8000
6000
4000
2000
0
−40
−30
−20
−10
0
Distance of polyA motif from 3' end, bp
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon


Distance of Detected PolyA Motif From 3'End
by Non−FSM/ISM Subcategories
Count
3000
2000
1000
0
−40
−30
−20
−10
0
Distance of polyA motif from 3' end, bp
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon

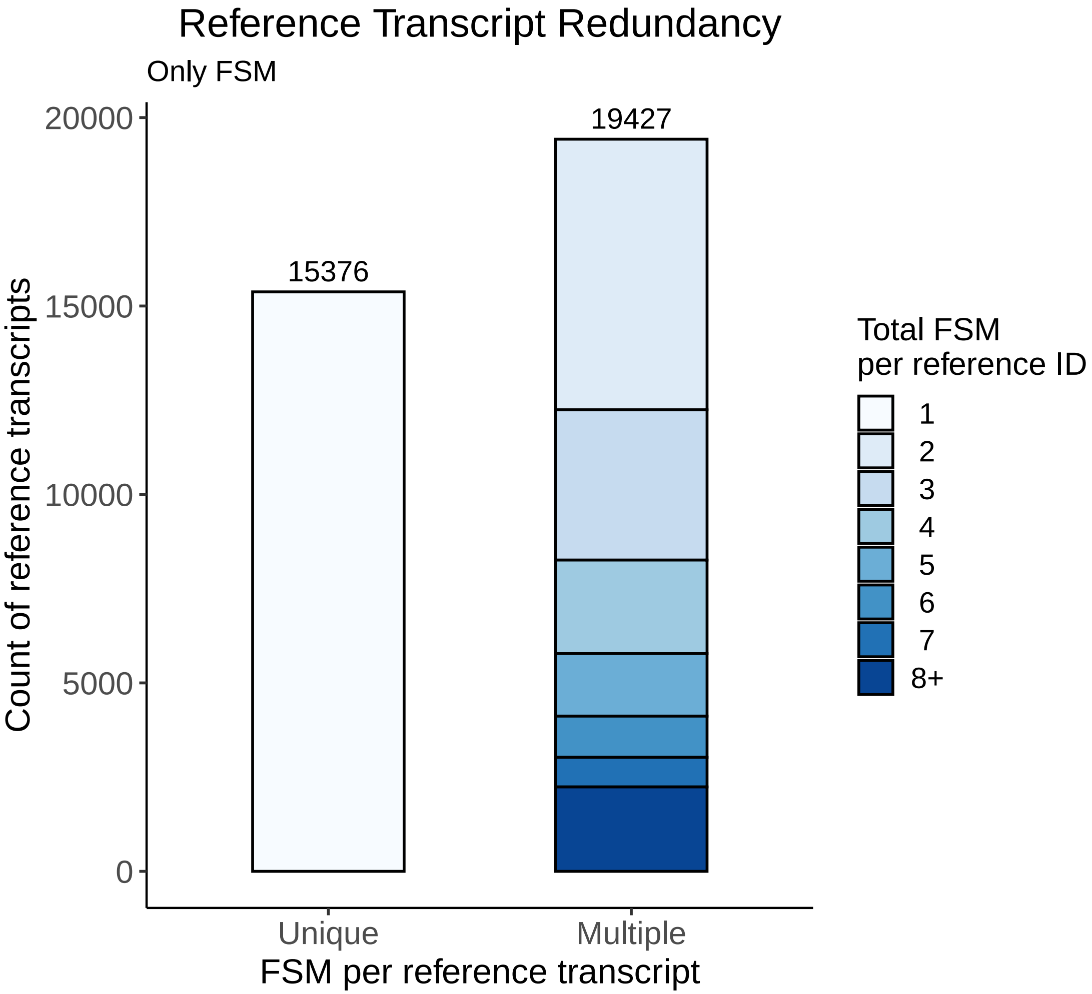
Number of polyA Motifs Detected

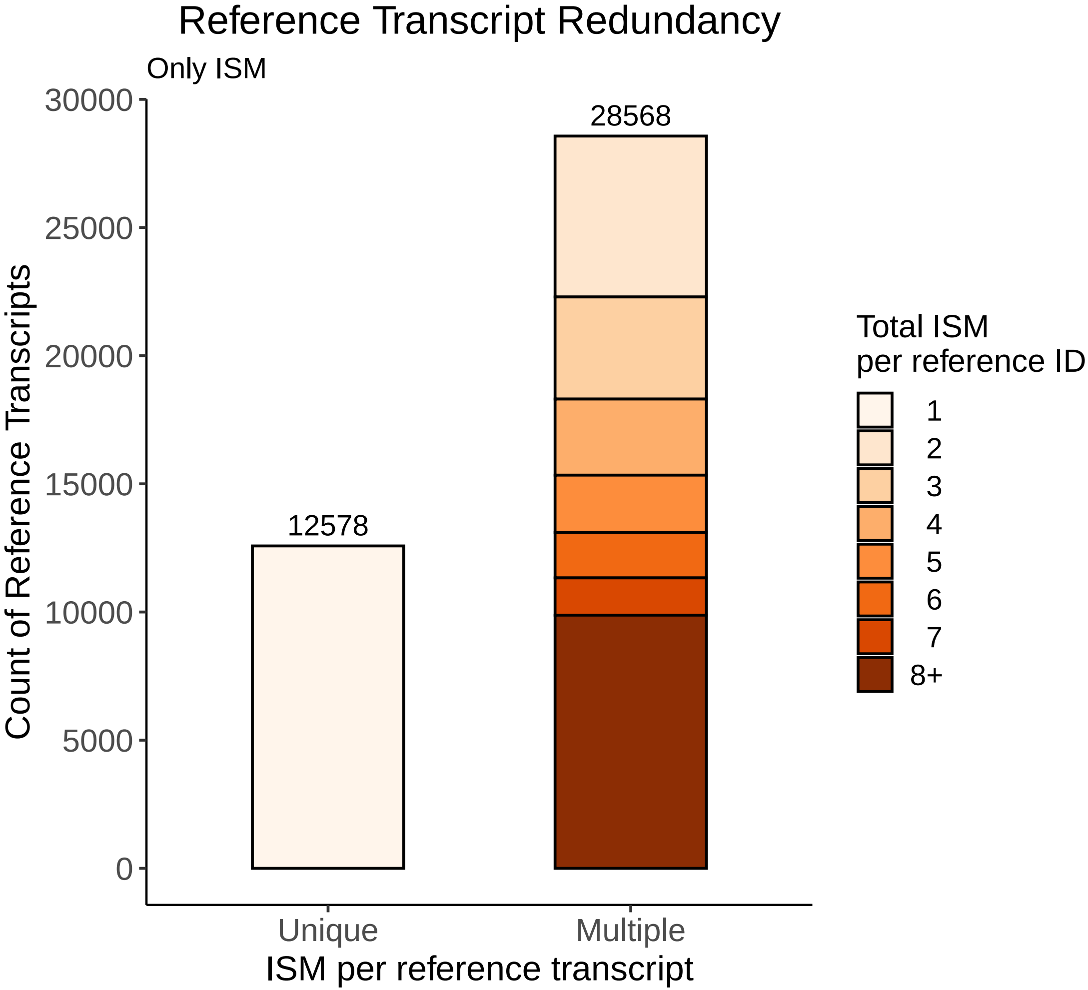
Subcategory	Count	polyA Detected	%
Alternative 3'end	24226	11637	48
Alternative 3'5'end	35052	18862	54
Alterantive 5'end	25228	13585	54
Reference match	13026	6813	52
3' fragment	192625	116859	61
Internal fragment	38062	23436	62
5' fragment	22545	10701	47
Comb. of annot. junctions	33360	17490	52
Comb. of annot. splice sites	27676	15219	55
Intron retention	37350	19812	53
At least 1 annot. don./accept.	66998	37931	57
Mono−exon	32973	20521	62
Multi−exon	8957	5670	63
NA	35	21	60

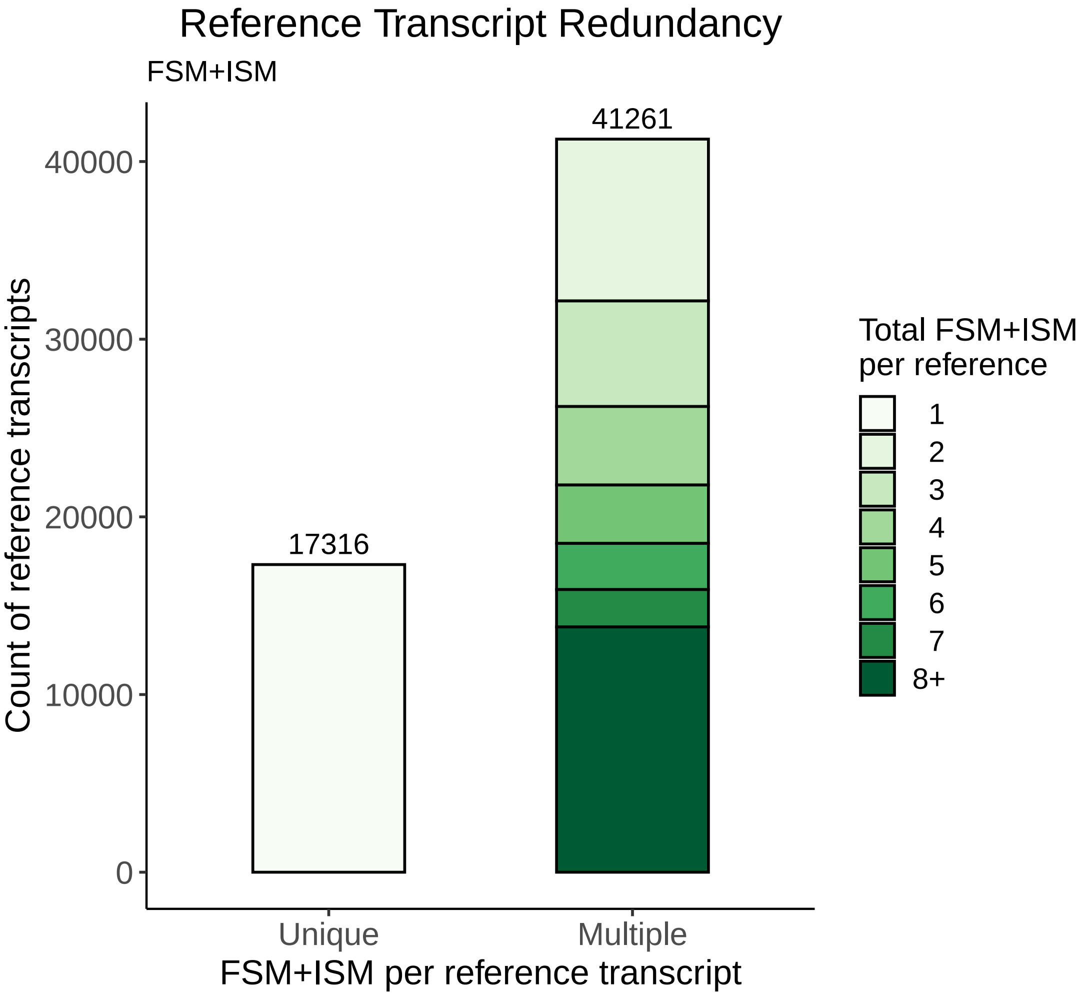
Frequency of PolyA Motifs

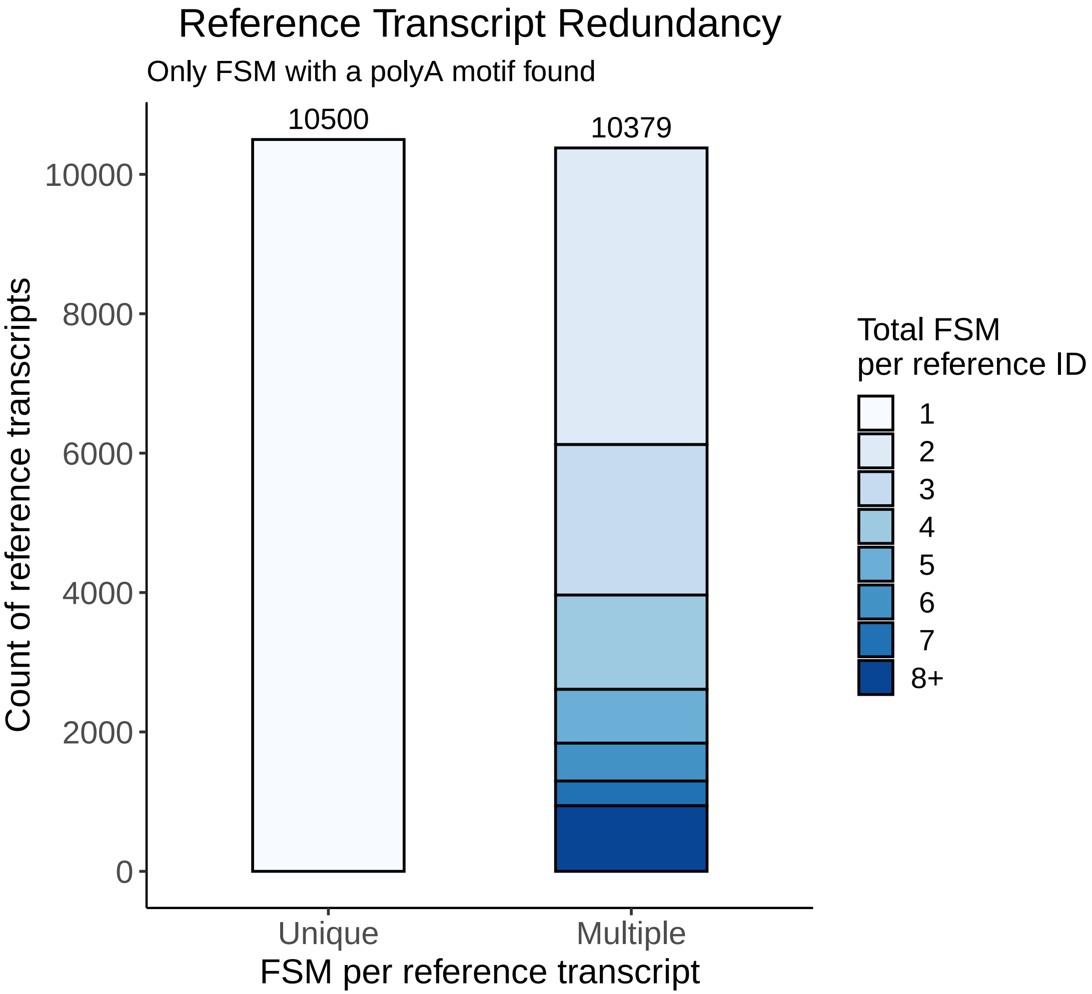
Motif	Count	%
AATAAA	139528	43.8
ATTAAA	43072	13.5
AAGAAA	15768	4.9
AAAAAG	15014	4.7
TATAAA	14380	4.5
AGTAAA	12920	4.1
TTTAAA	10705	3.4
CATAAA	8527	2.7
AATATA	8510	2.7
AAAACA	8476	2.7
GATAAA	8184	2.6
AATACA	8181	2.6
GGGGCT	8002	2.5
AATGAA	7938	2.5
AATAGA	4956	1.6
ACTAAA	4396	1.4

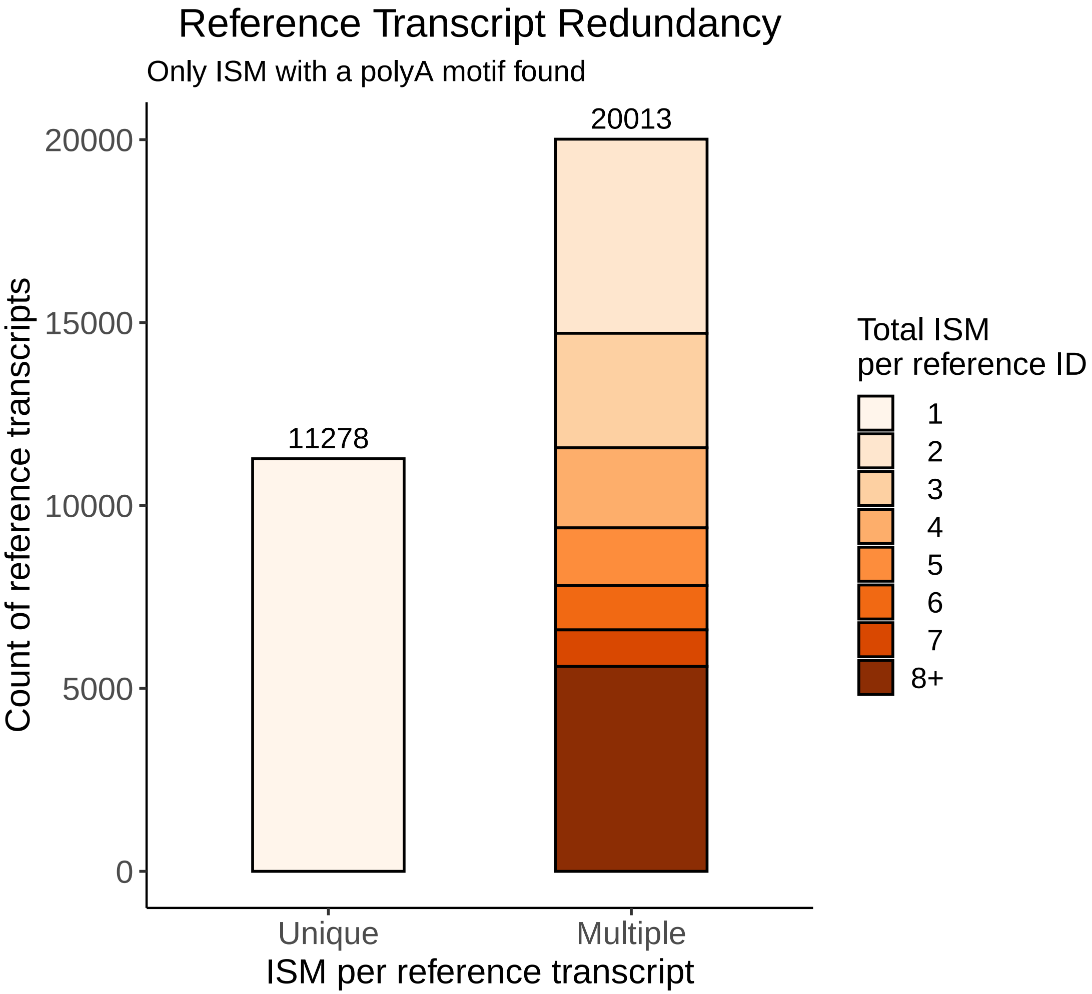
Redundancy Analysis

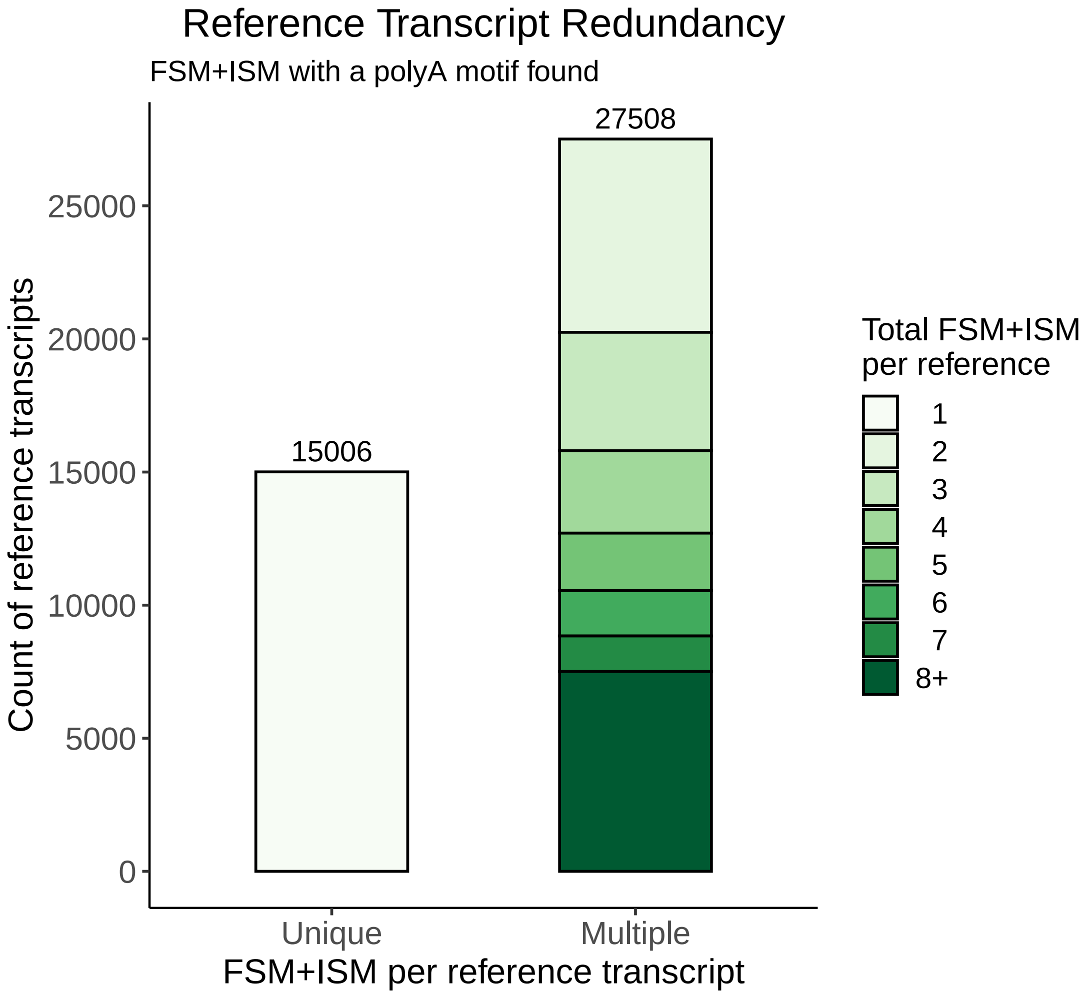

Reference Transcript Redundancy
Only FSM
20000
15000
10000
5000
0
15376
19427
Count of reference transcripts
Unique
Multiple
FSM per reference transcript
Total FSM
per reference ID
1
2
3
4
5
6
7
8+


Reference Transcript Redundancy
Only ISM
30000
25000
20000
15000
10000
5000
0
Count of Reference Transcripts
12578
28568
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
5
6
7
8+

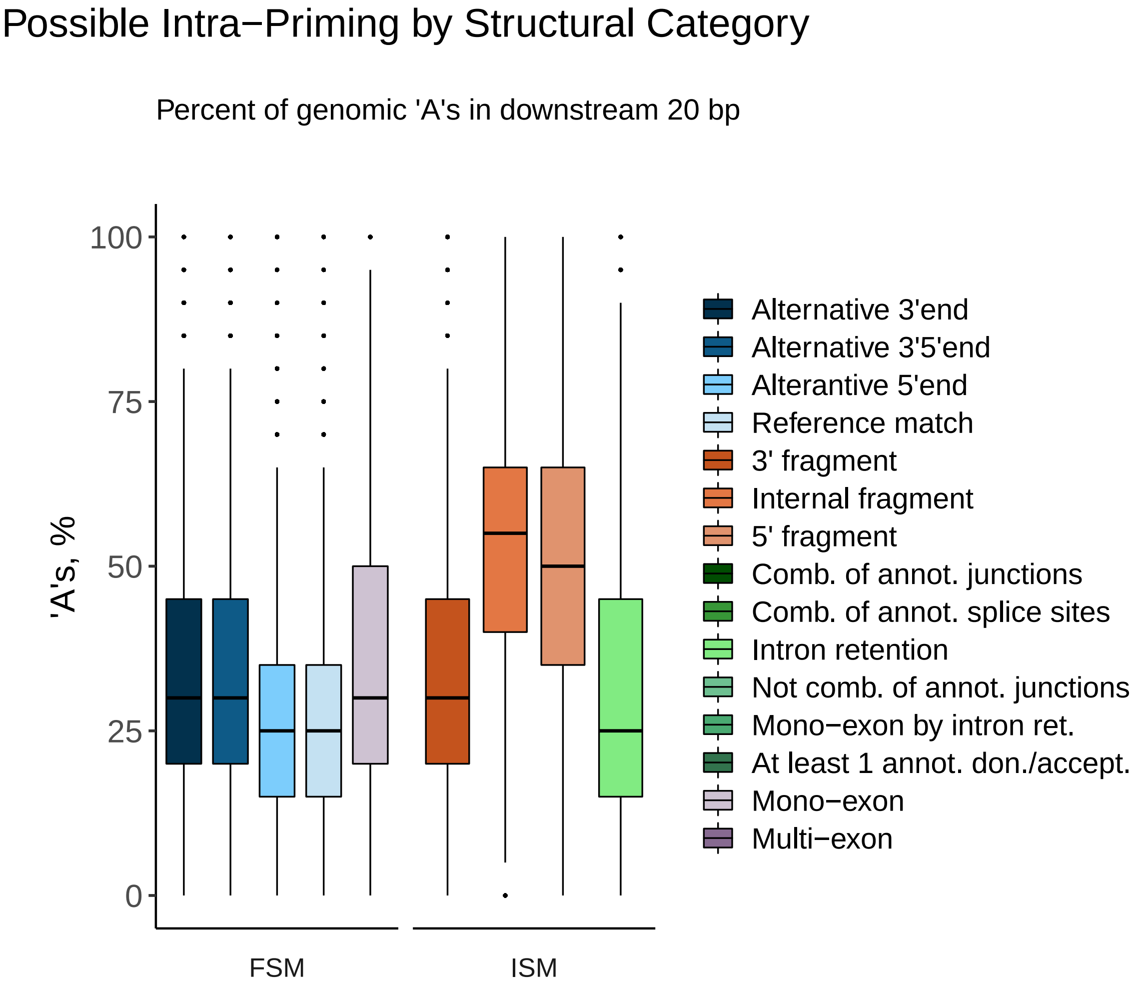

Reference Transcript Redundancy
FSM+ISM
Count of reference transcripts
40000
30000
20000
10000
0
17316
41261
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+


Reference Transcript Redundancy
Only FSM with a polyA motif found
10500
10379
Count of reference transcripts
0
2000
4000
6000
8000
10000
Unique
Multiple
FSM per reference transcript
Total FSM
per reference ID
1
2
3
4
5
6
7
8+


Reference Transcript Redundancy
Only ISM with a polyA motif found
Count of reference transcripts
0
5000
10000
15000
20000
11278
20013
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
5
6
7
8+

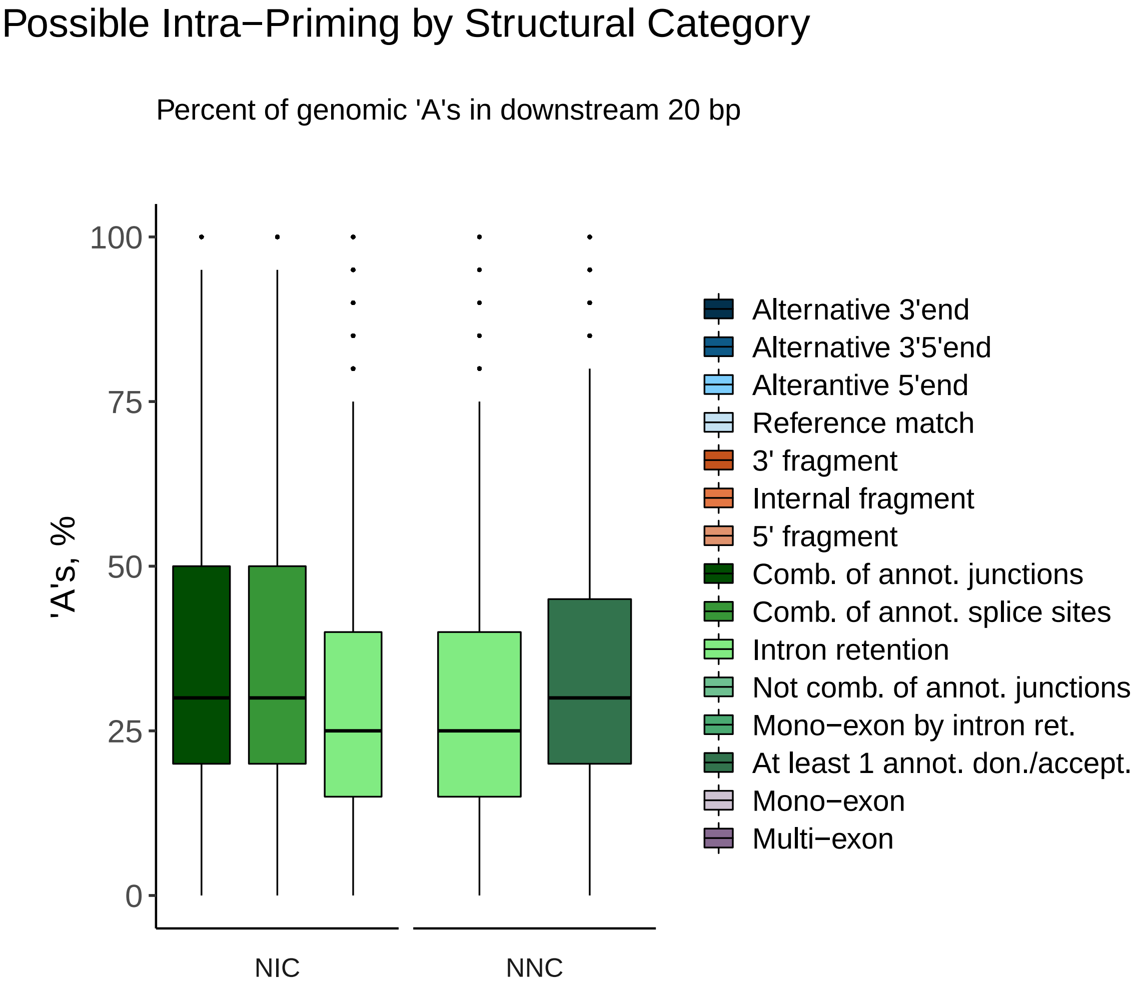

Reference Transcript Redundancy
FSM+ISM with a polyA motif found
Count of reference transcripts
25000
20000
15000
10000
5000
0
15006
27508
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+

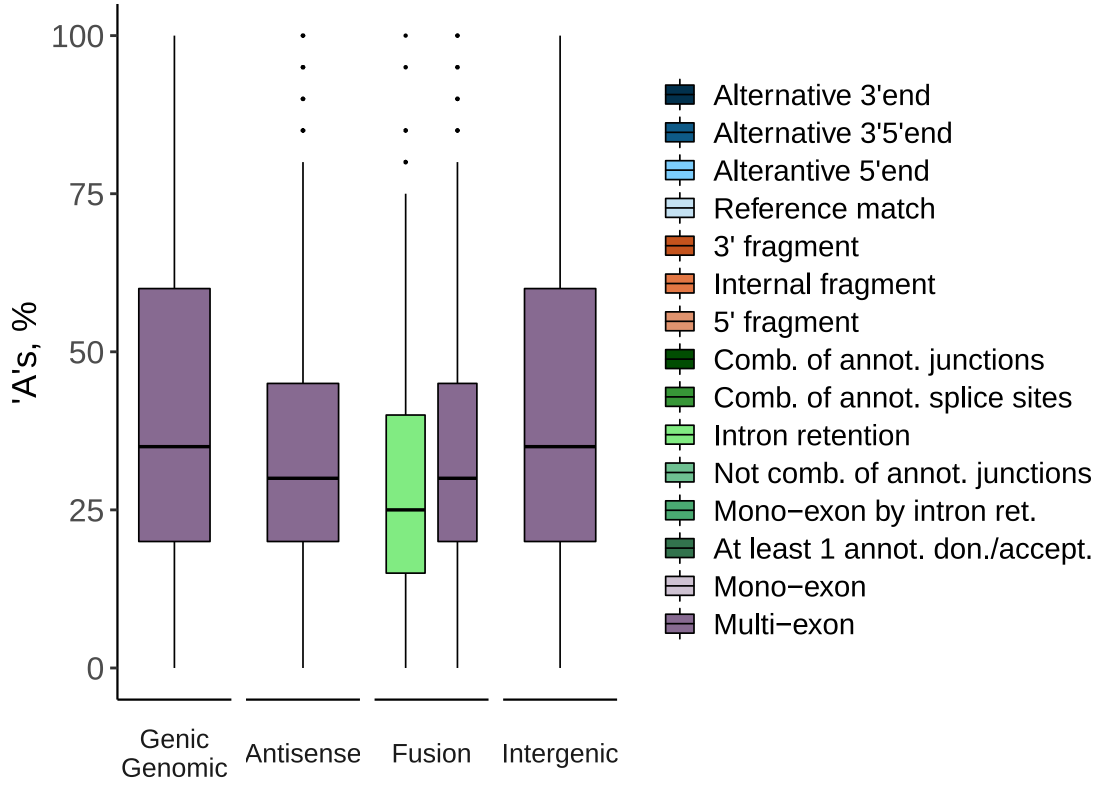
Intra−Priming Quality Check



Possible Intra−Priming by Structural Category

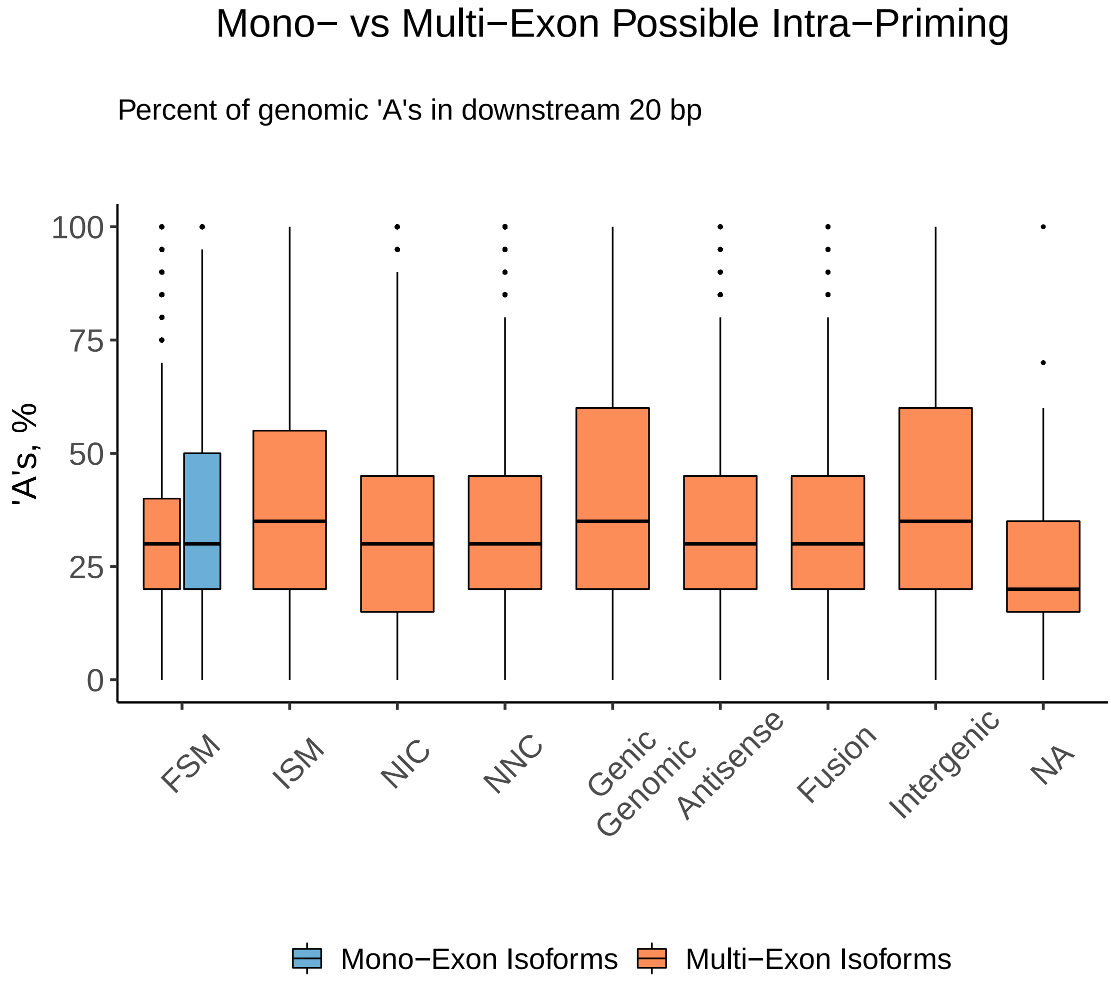
Percent of genomic 'A's in downstream 20 bp

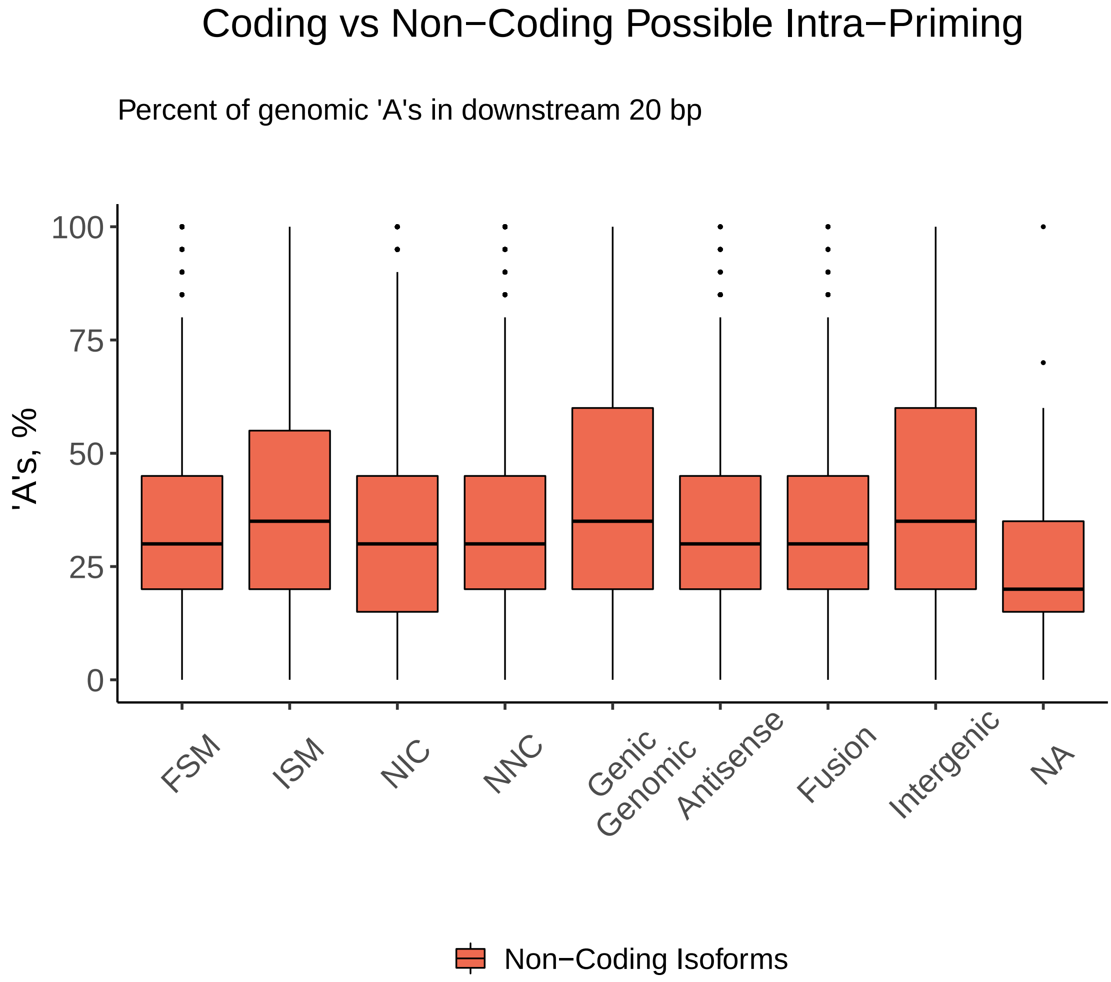

Possible Intra−Priming by Structural Category
Percent of genomic 'A's in downstream 20 bp
'A's, %
100
75
50
25
0
NIC
NNC
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon



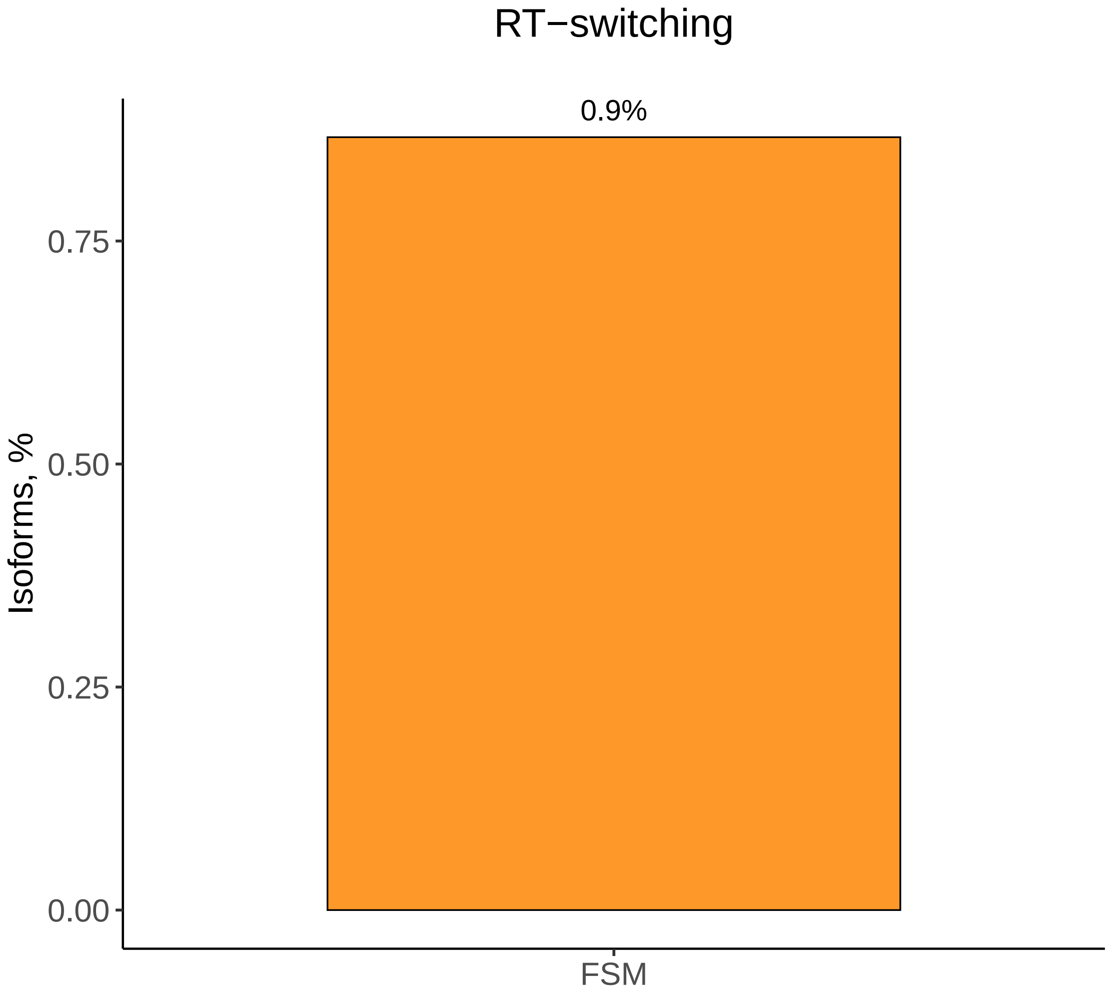
Possible Intra−Priming by Structural Category

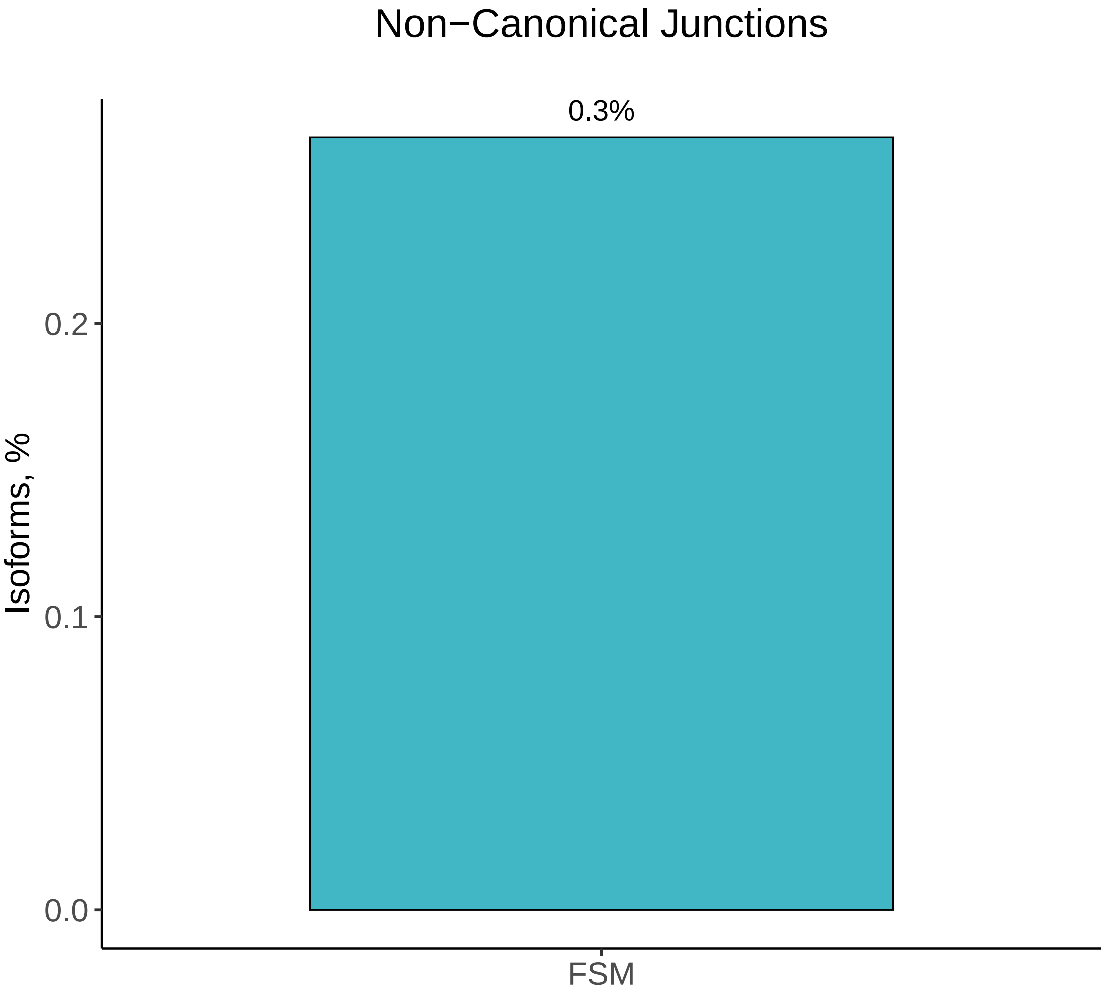
Percent of genomic 'A's in downstream 20 bp

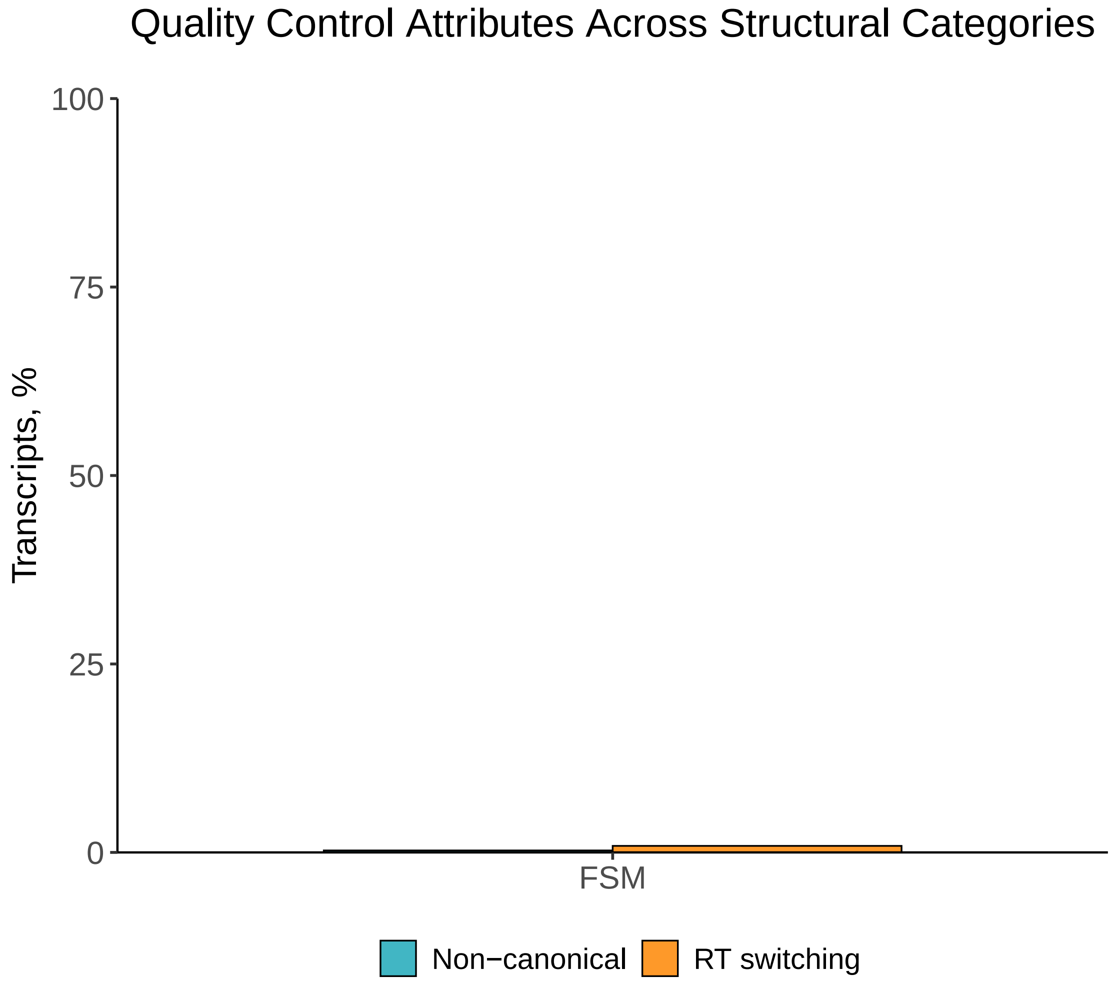

Mono− vs Multi−Exon Possible Intra−Priming
Percent of genomic 'A's in downstream 20 bp
'A's, %
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic
Genomic
Antisense
Fusion
Intergenic
NA
Mono−Exon Isoforms
Multi−Exon Isoforms


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

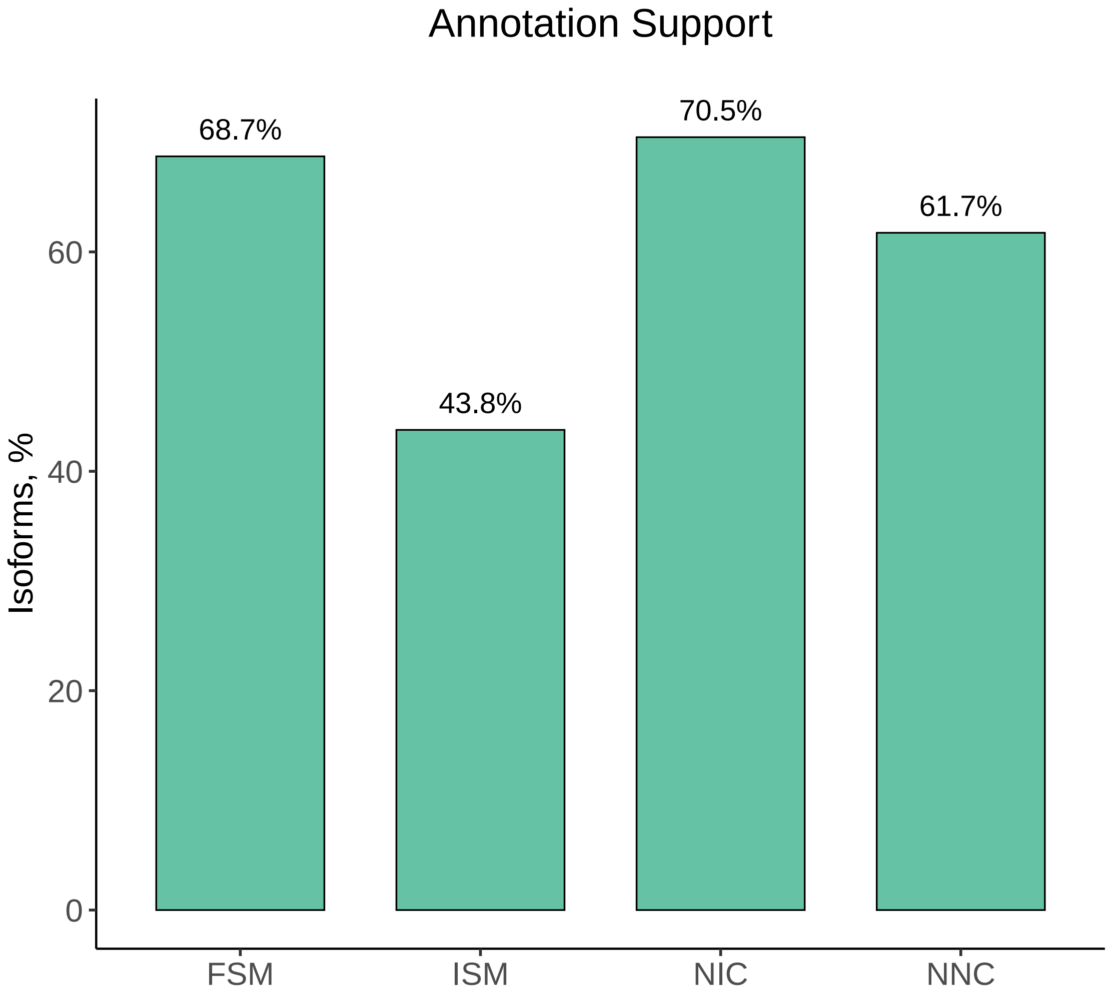
Features of Bad Quality

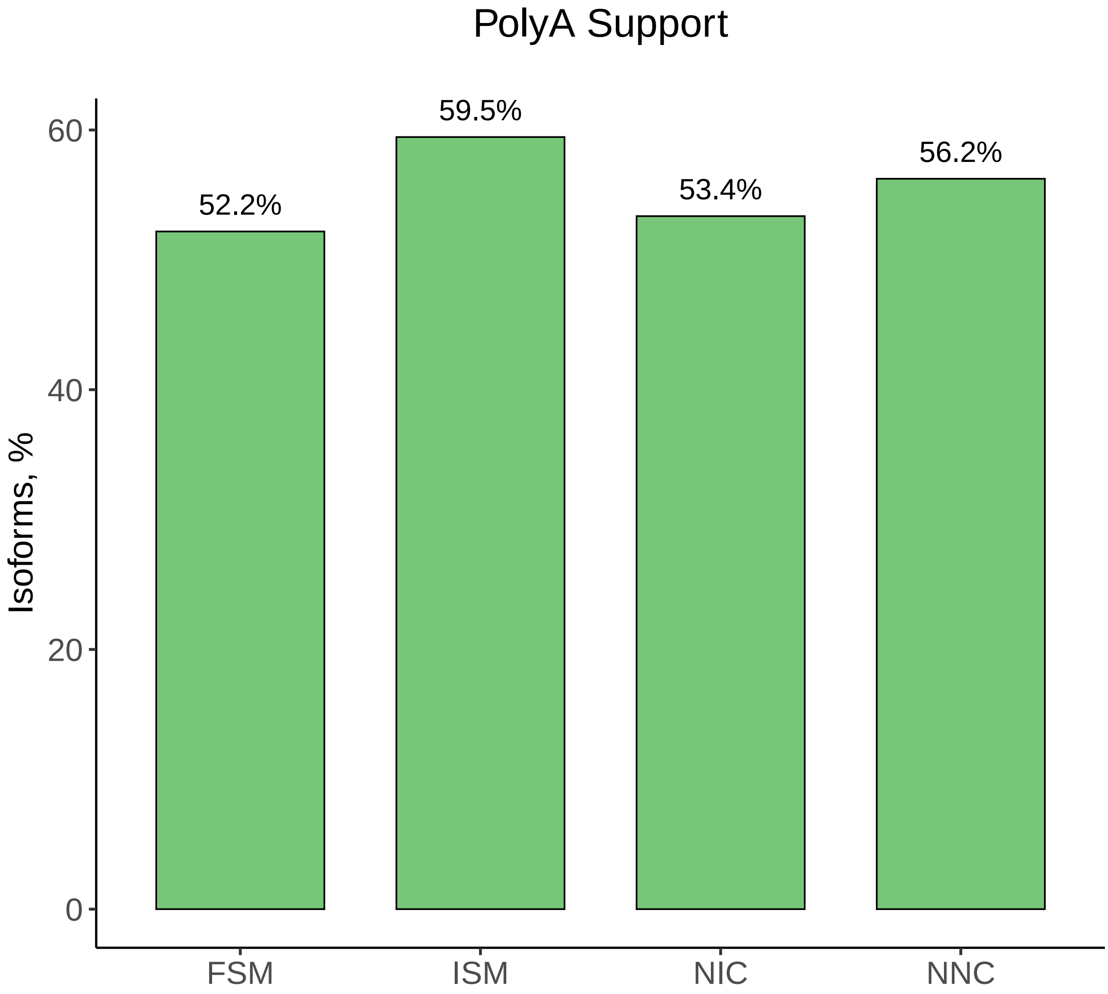

RT−switching
0.9%
Isoforms, %
0.75
0.50
0.25
0.00
FSM

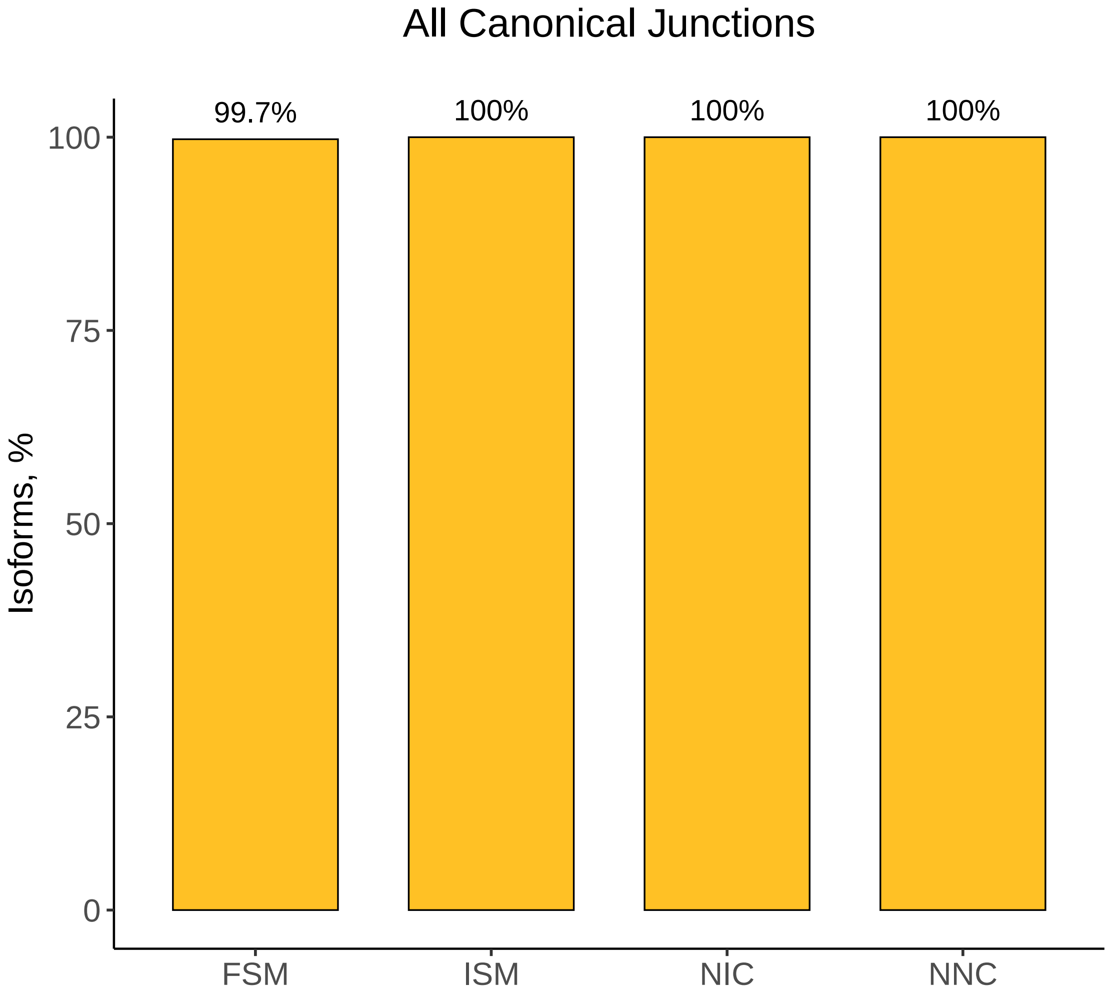

Non−Canonical Junctions
0.3%
Isoforms, %
0.2
0.1
0.0
FSM

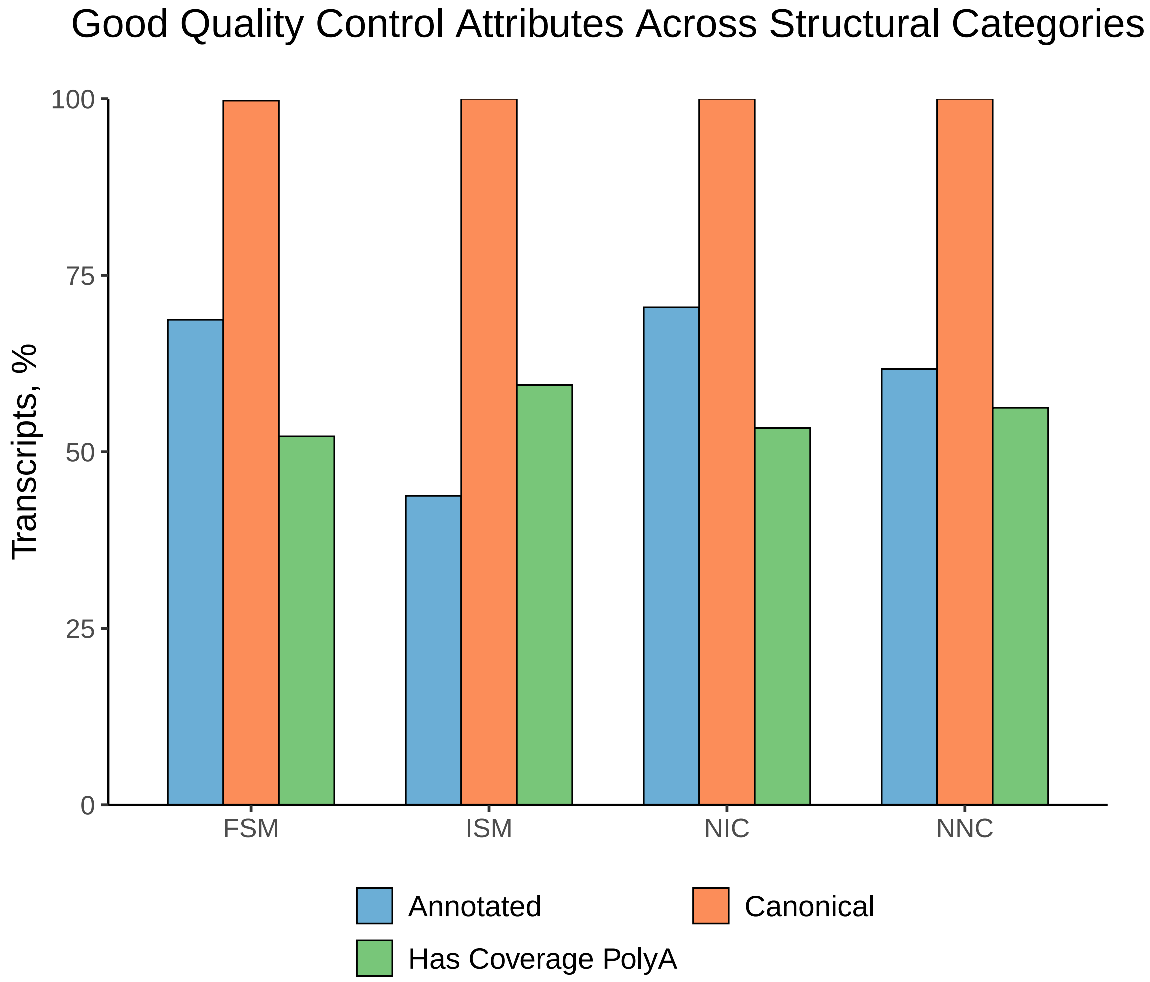

Quality Control Attributes Across Structural Categories
Transcripts, %
100
75
50
25
0
FSM
Non−canonical
RT switching

Features of Good Quality


Annotation Support
Isoforms, %
0
20
40
60
68.7%
43.8%
70.5%
61.7%
FSM
ISM
NIC
NNC


PolyA Support
Isoforms, %
60
40
20
0
52.2%
59.5%
53.4%
56.2%
FSM
ISM
NIC
NNC


All Canonical Junctions
Isoforms, %
100
75
50
25
0
99.7%
100%
100%
100%
FSM
ISM
NIC
NNC


Good Quality Control Attributes Across Structural Categories
Transcripts, %
100
75
50
25
0
FSM
ISM
NIC
NNC
Annotated
Canonical
Has Coverage PolyA