SQANTI3 report

Unique Genes: 24253

Unique Isoforms: 639521

Transcript Classification

Category	Isoforms, count
FSM	149437
ISM	236718
NIC	138600
NNC	105766
Genic Genomic	752
Antisense	2146
Fusion	3044
Intergenic	2989
Genic Intron	0

Gene Classification

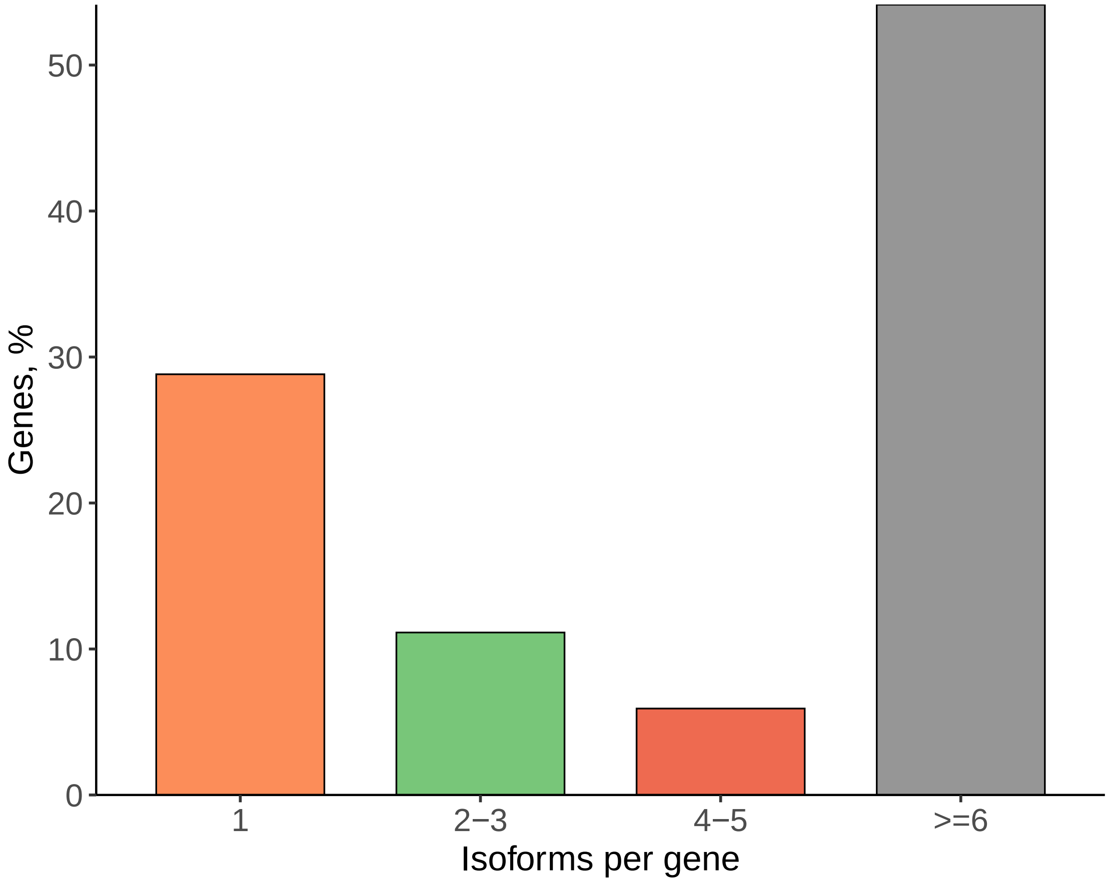
Category	Genes, count
Annotated Genes	20570
Novel Genes	3683

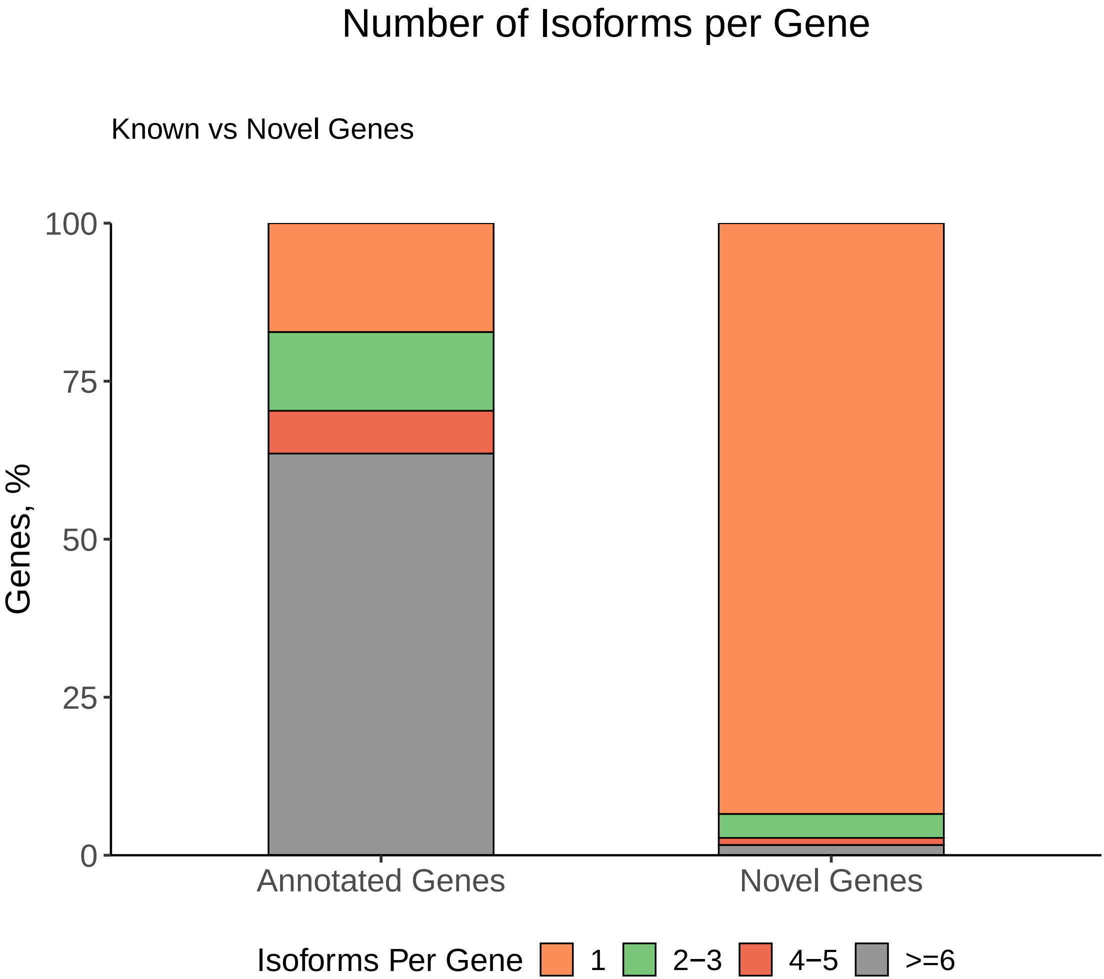
Splice Junction Classification

Category	SJs, count	Percent
Known canonical	189488	64.49
Known Non−canonical	87	0.03
Novel canonical	104228	35.48
Novel Non−canonical	0	0.00

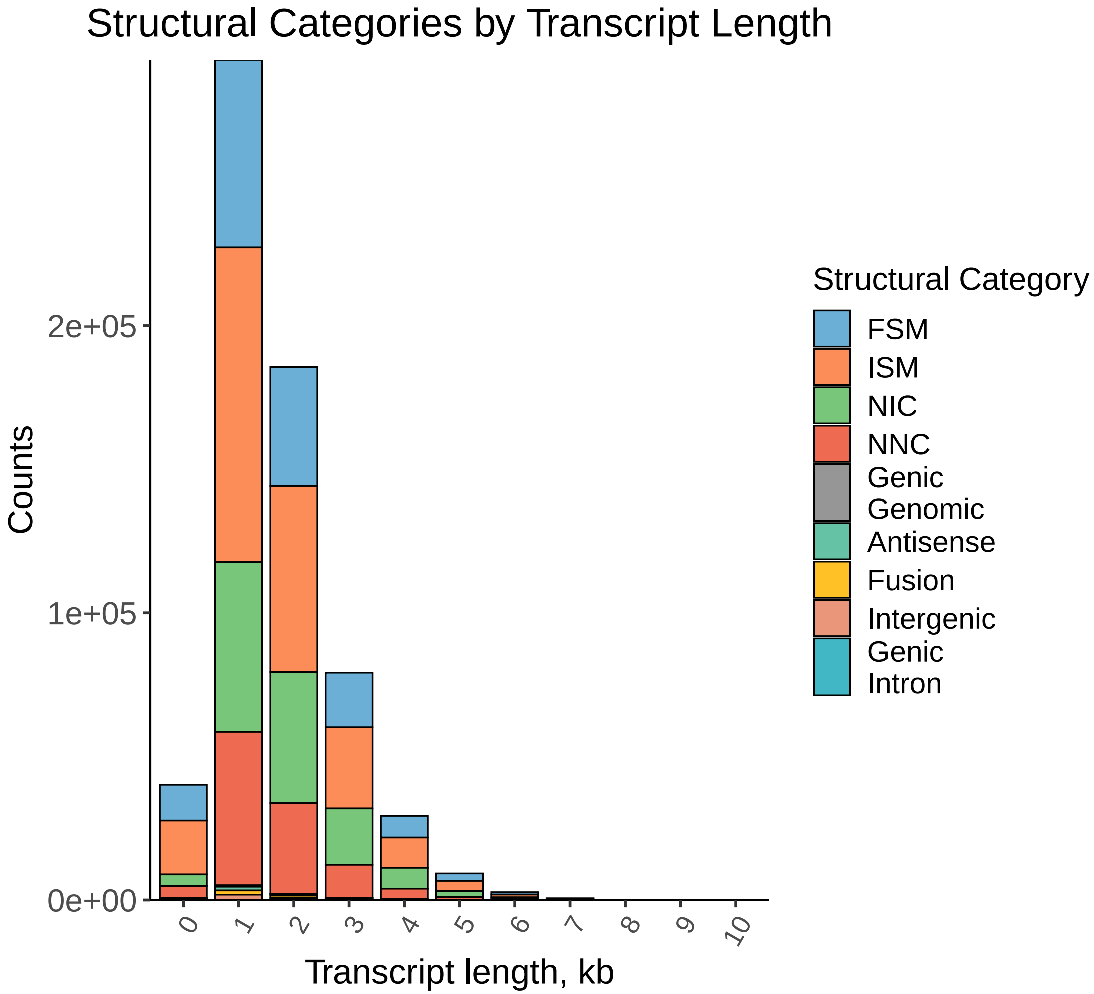
Gene Characterization

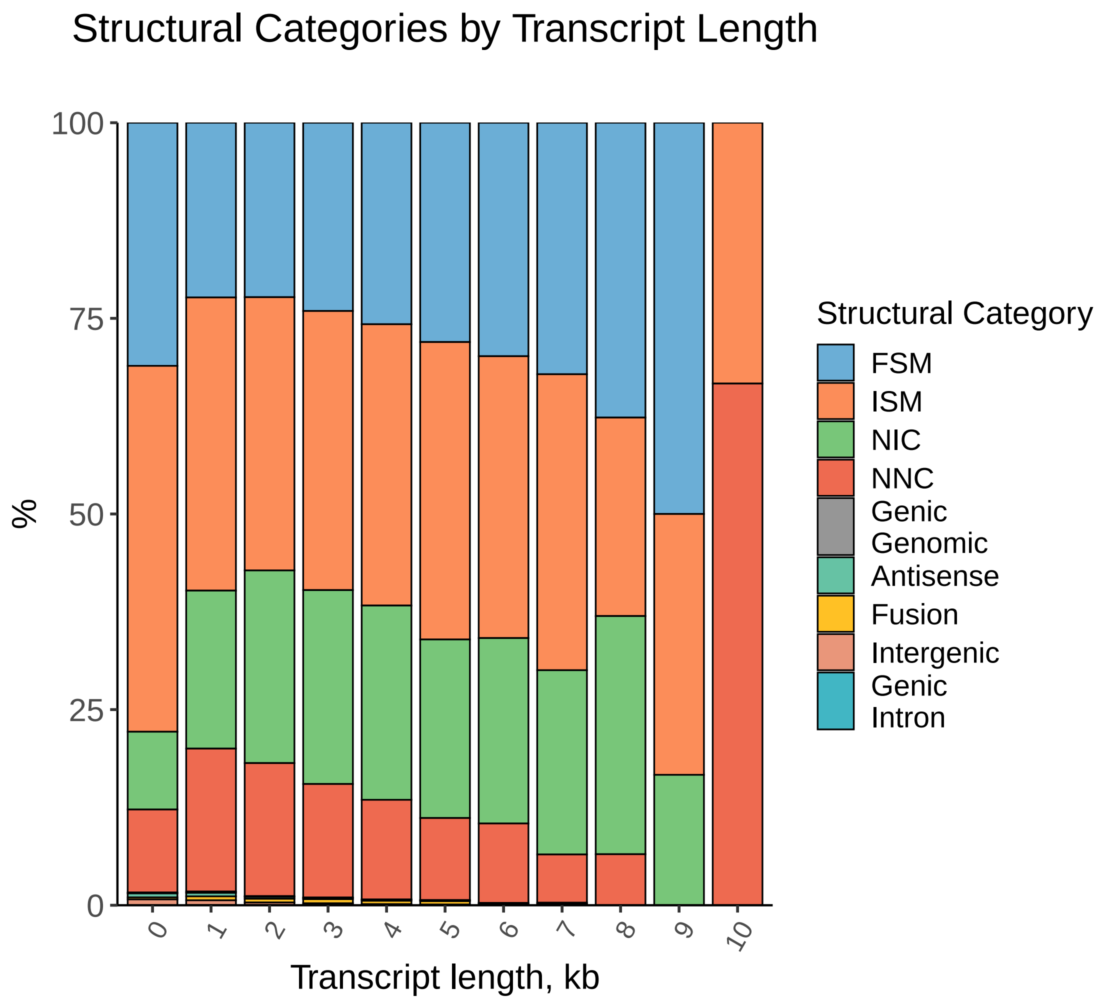
Number of Isoforms per Gene

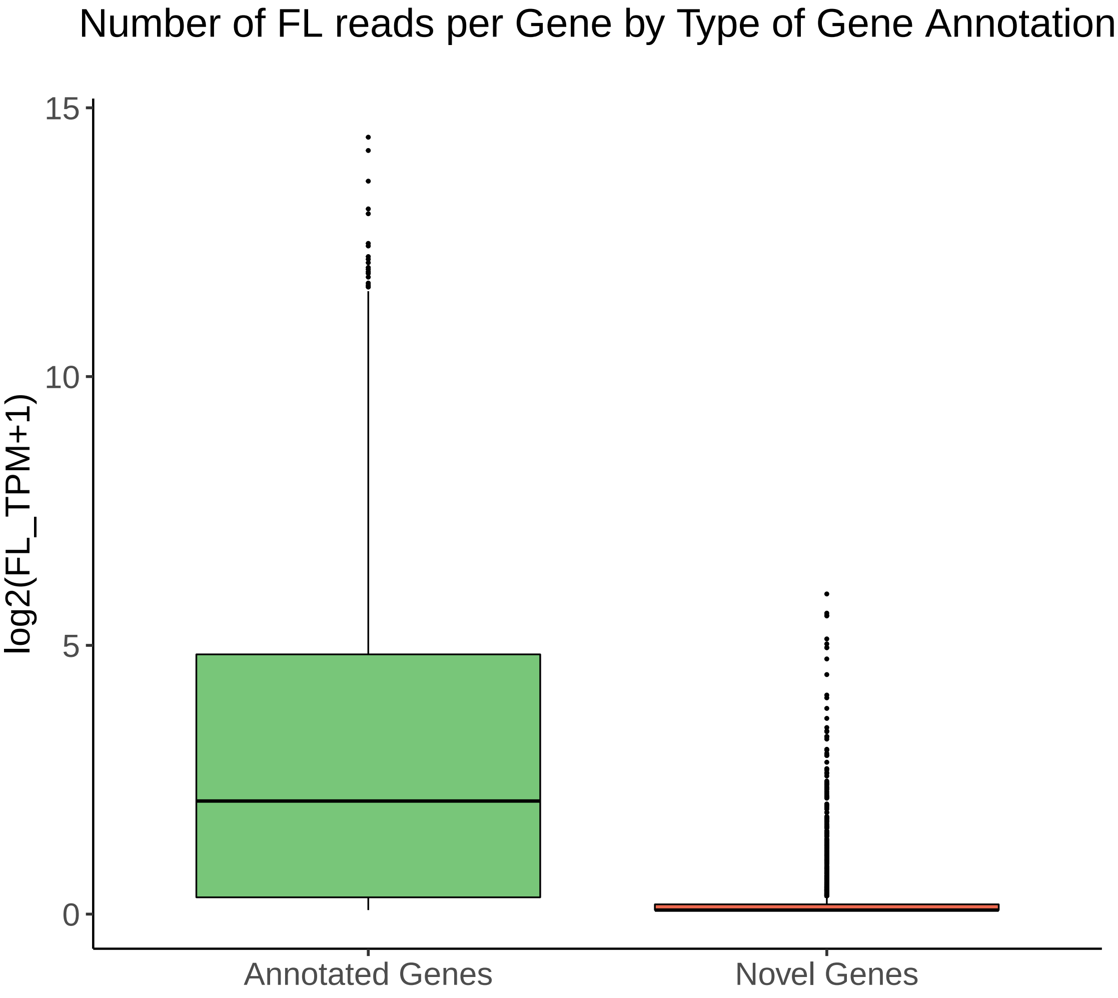


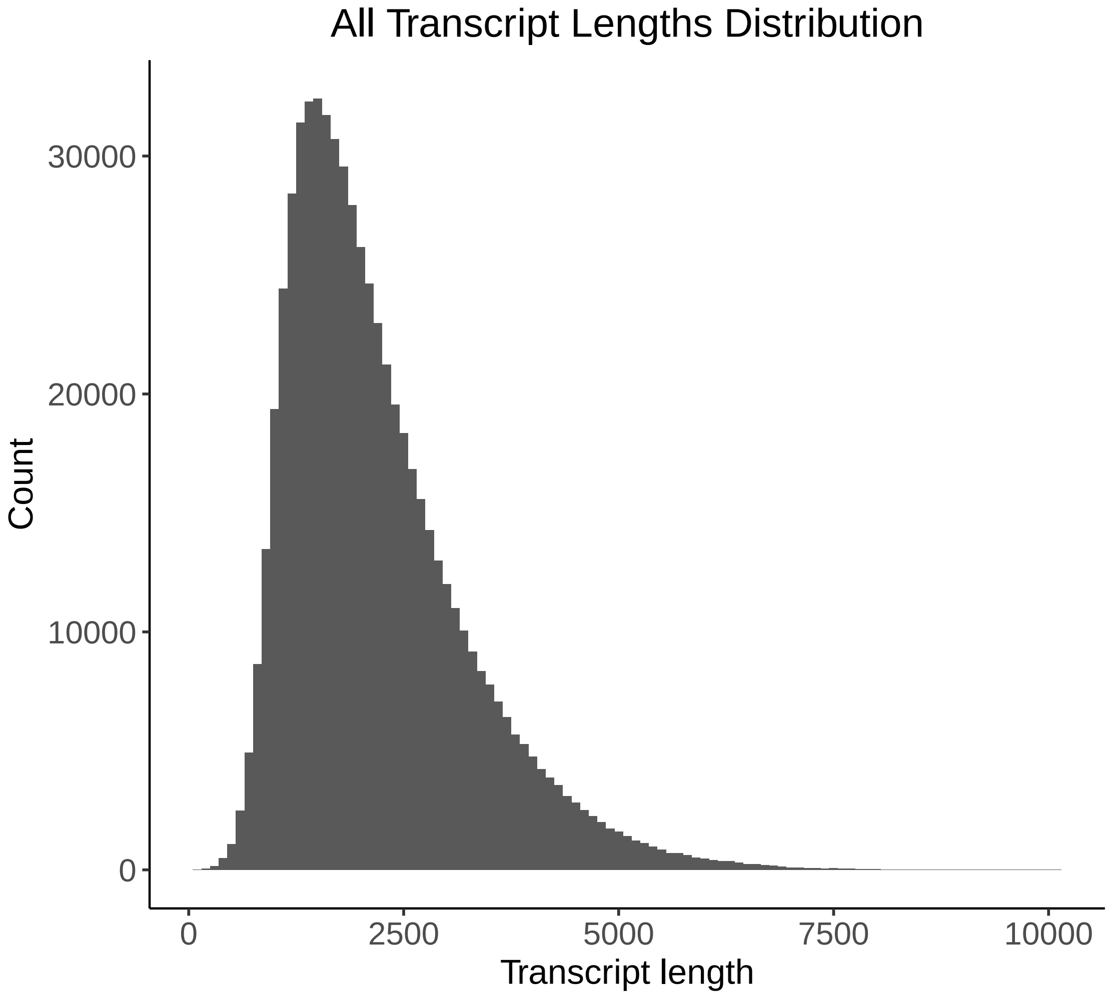

Number of Isoforms per Gene
Known vs Novel Genes
Genes, %
100
75
50
25
0
Annotated Genes
Novel Genes
Isoforms Per Gene
1
2−3
4−5
>=6

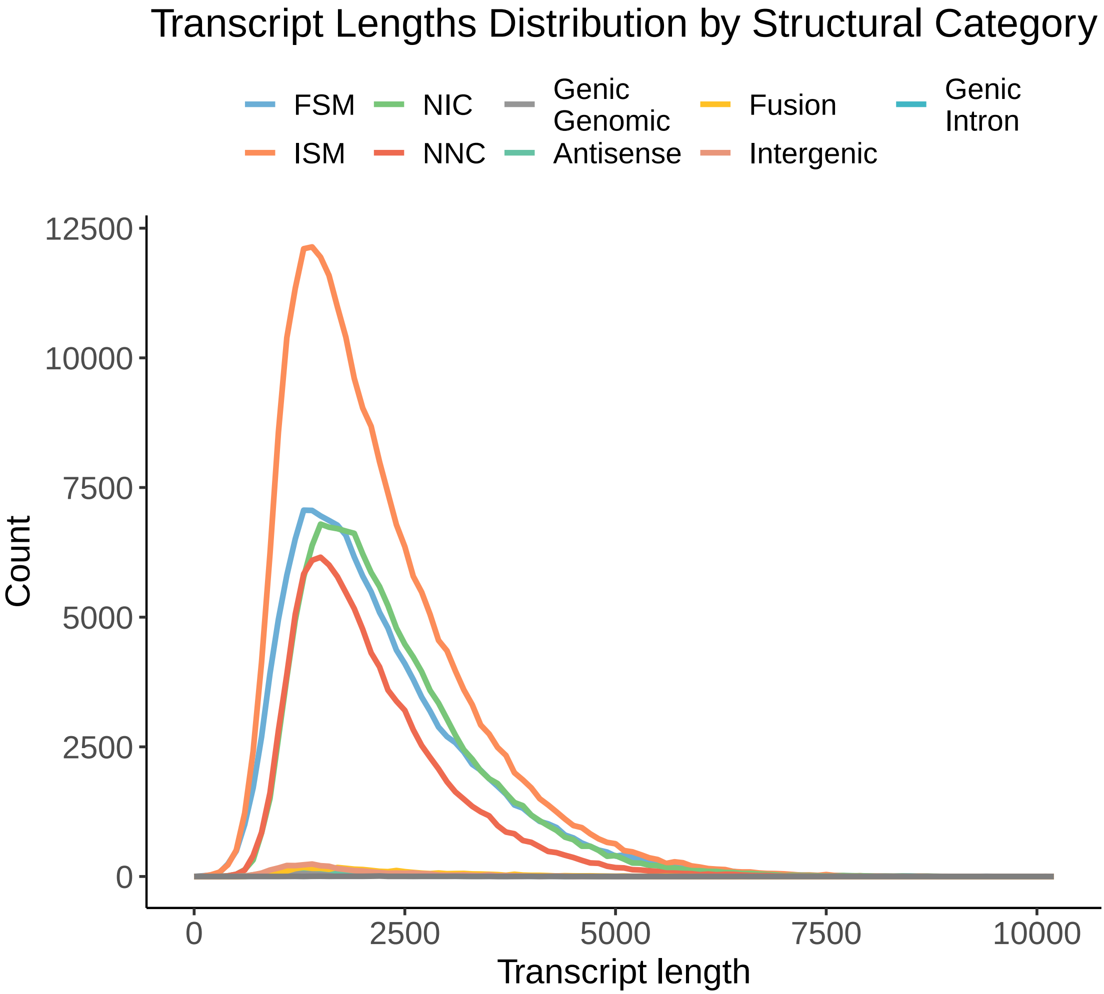

Distribution of Mono− vs Multi−Exon Transcripts
Transcripts, %
100
75
50
25
0
Novel Genes
Annotated Genes
Transcript type
Multi−Exon
Mono−Exon

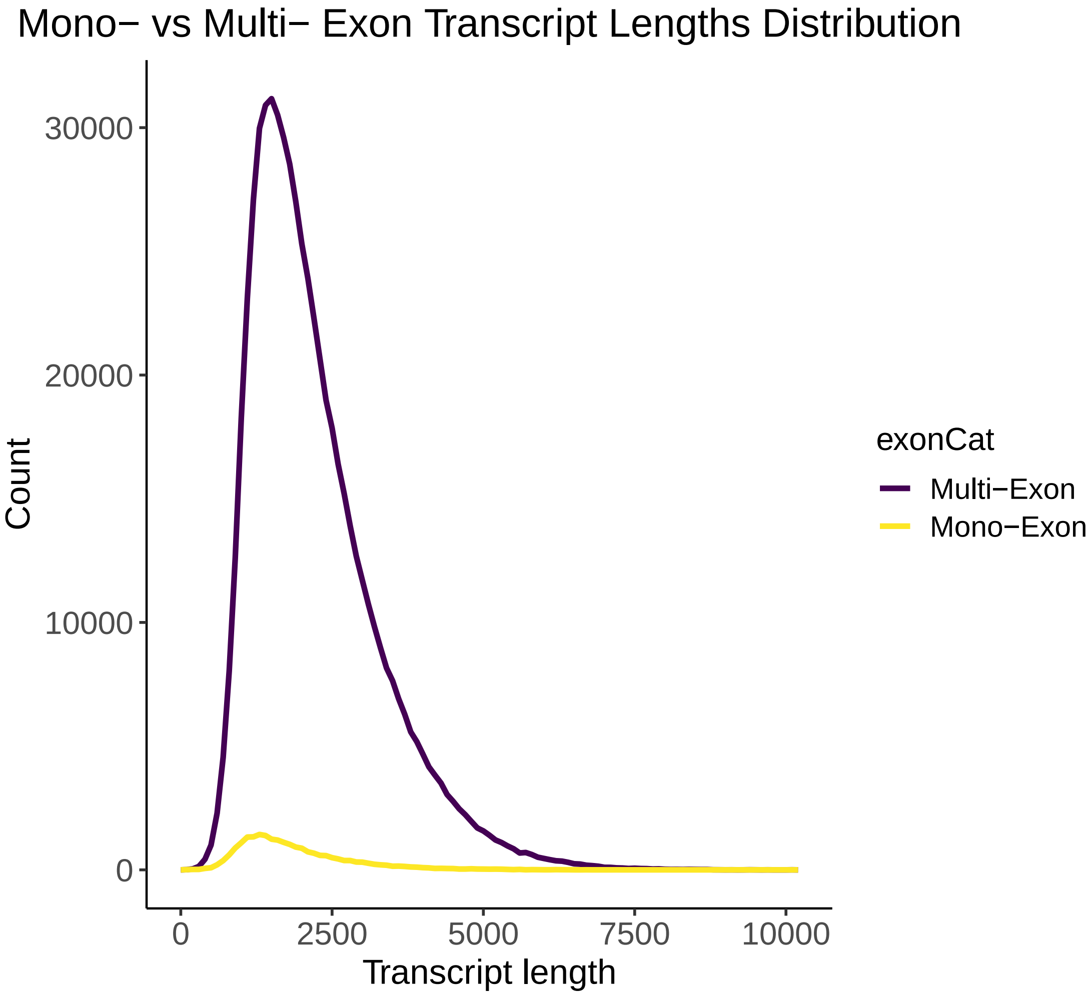

Structural Categories by Transcript Length
Counts
2e+05
1e+05
0e+00
0
1
2
3
4
5
6
7
8
9
10
Transcript length, kb
Structural Category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

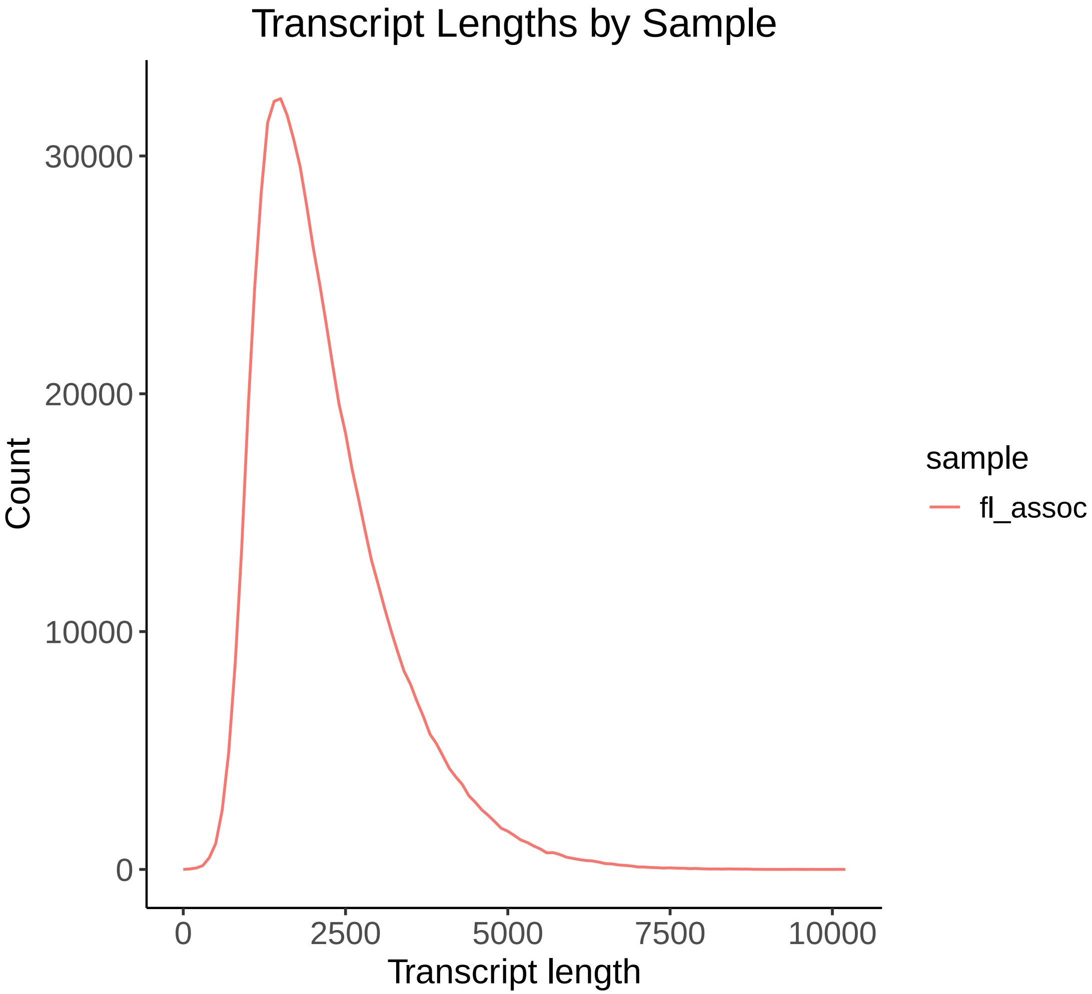

Structural Categories by Transcript Length
%
100
75
50
25
0
0
1
2
3
4
5
6
7
8
9
10
Transcript length, kb
Structural Category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

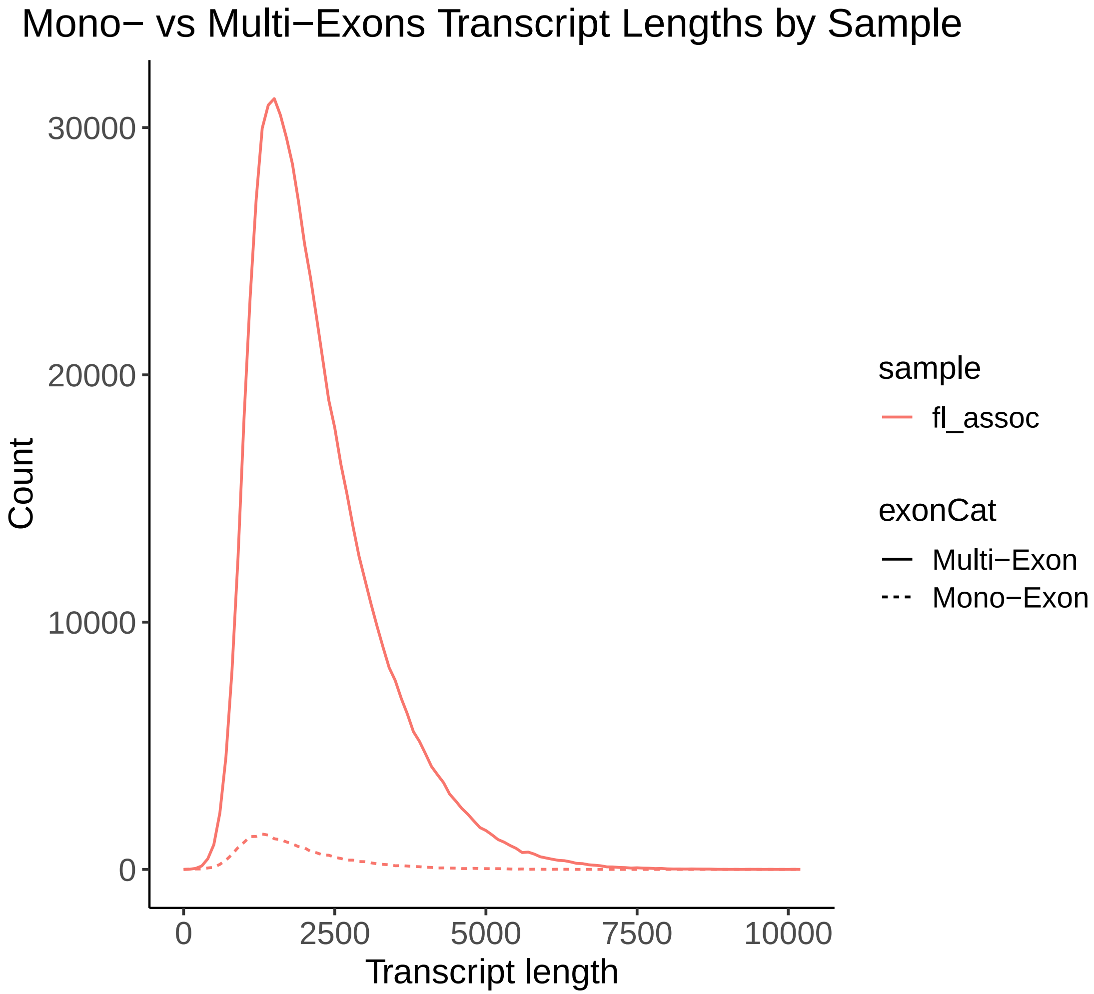

Number of FL reads per Gene by Type of Gene Annotation
log2(FL_TPM+1)
15
10
5
0
Annotated Genes
Novel Genes


All Transcript Lengths Distribution
Count
30000
20000
10000
0
0
2500
5000
7500
10000
Transcript length


Transcript Lengths Distribution by Structural Category
FSM
NIC
Genic Genomic
Fusion
Genic Intron
ISM
NNC
Antisense
Intergenic
Count
12500
10000
7500
5000
2500
0
0
2500
5000
7500
10000
Transcript length

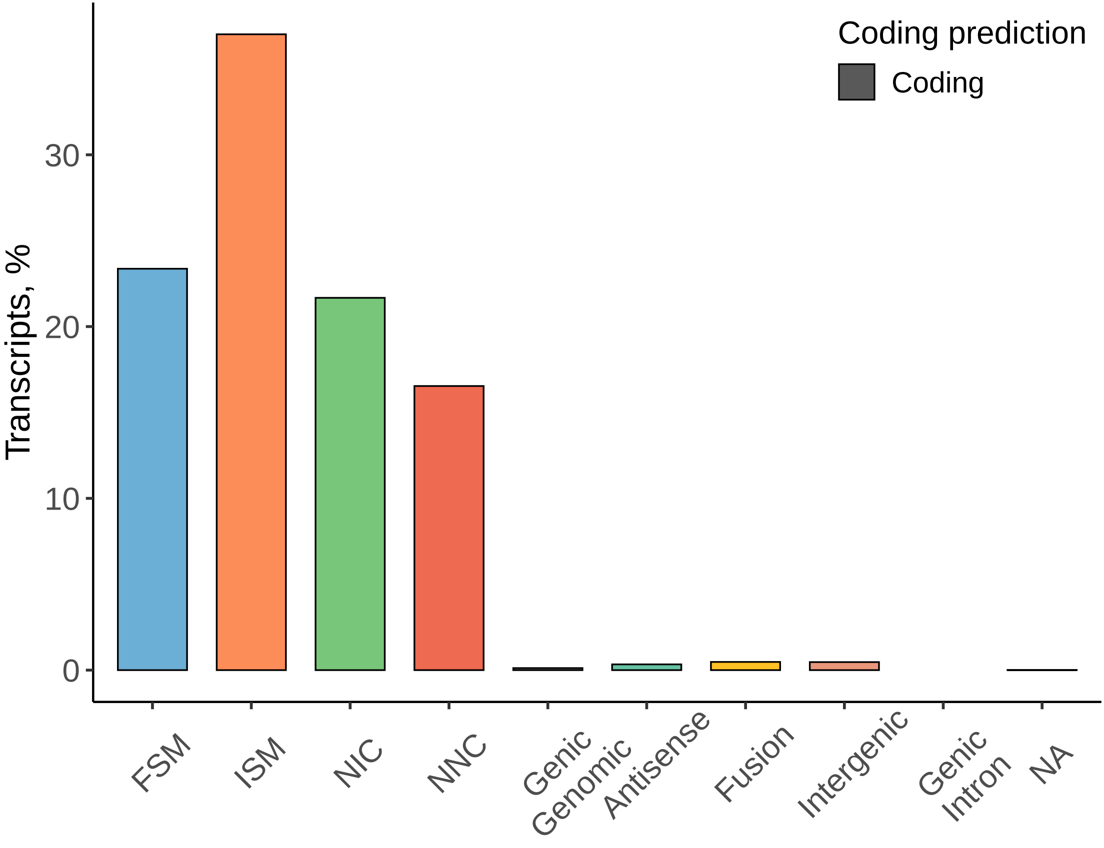

Mono− vs Multi− Exon Transcript Lengths Distribution
Count
30000
20000
10000
0
0
2500
5000
7500
10000
Transcript length
exonCat
Multi−Exon
Mono−Exon

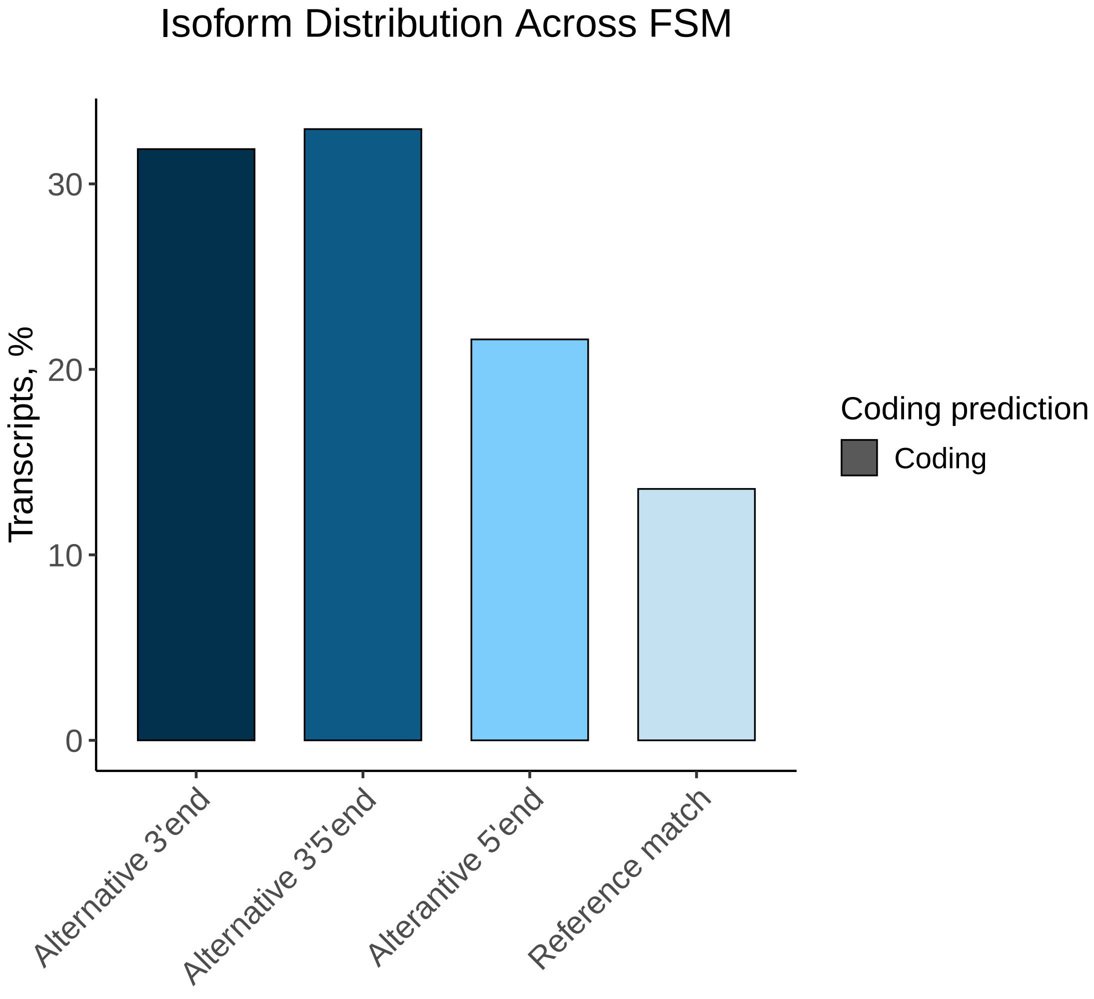

Transcript Lengths by Sample
Count
30000
20000
10000
0
0
2500
5000
7500
10000
Transcript length
sample
fl_assoc

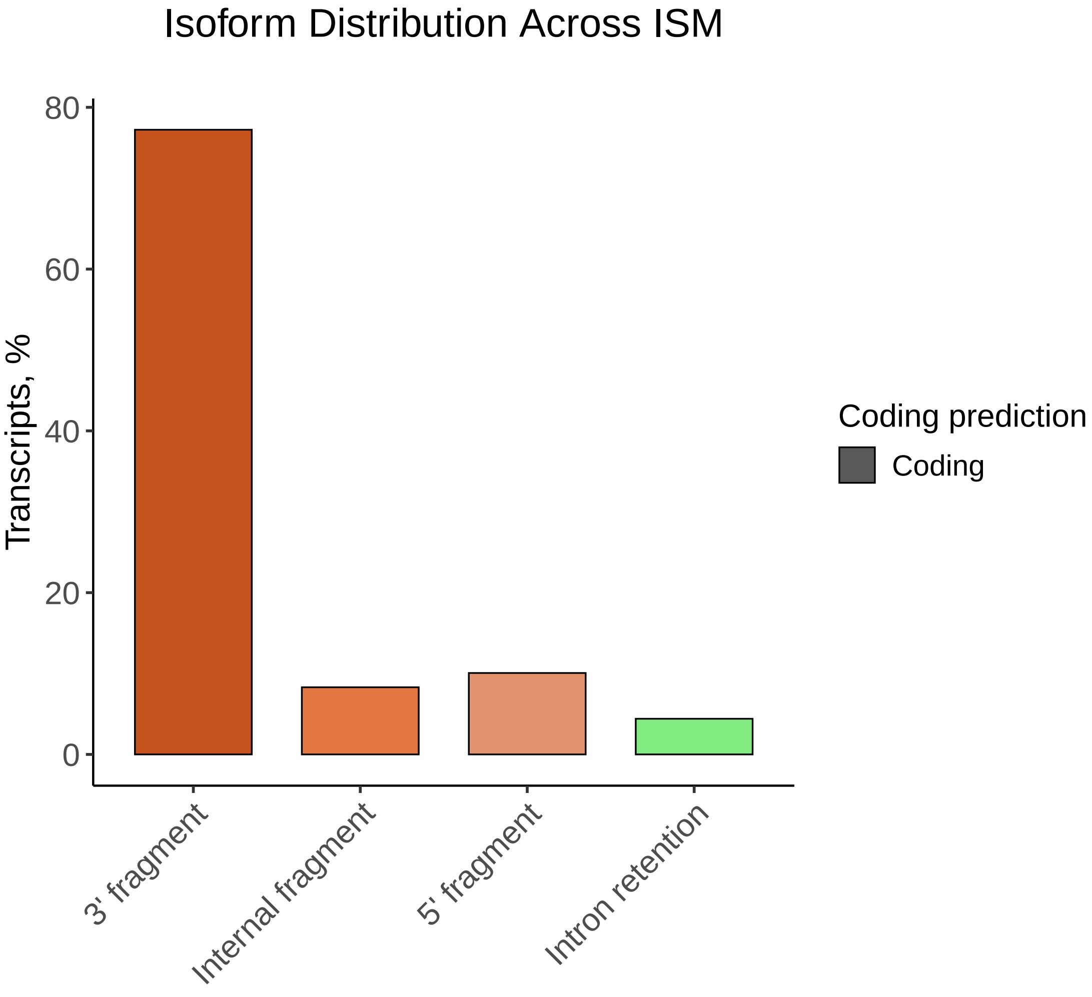

Mono− vs Multi−Exons Transcript Lengths by Sample
Count
30000
20000
10000
0
0
2500
5000
7500
10000
Transcript length
sample
fl_assoc
exonCat
Multi−Exon
Mono−Exon

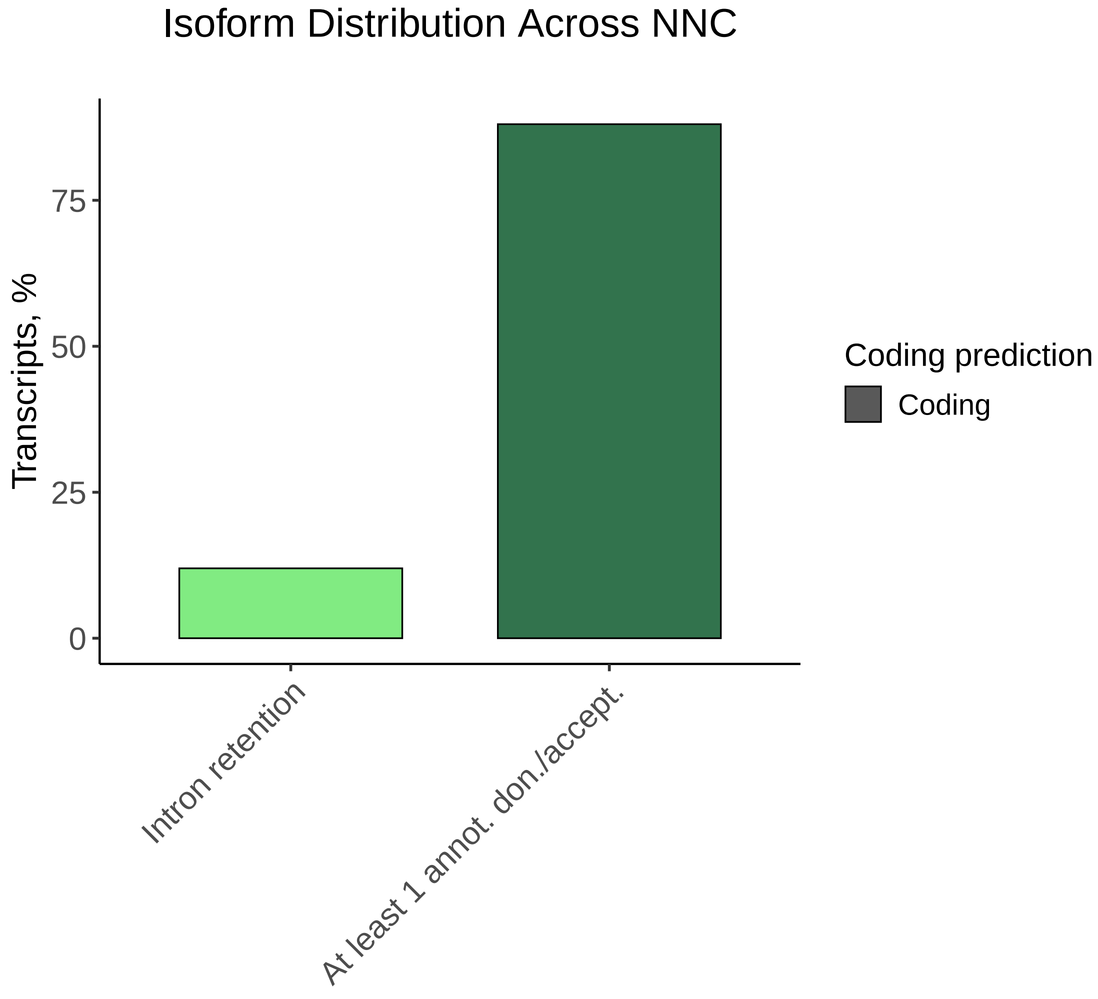
Structural Isoform Characterization

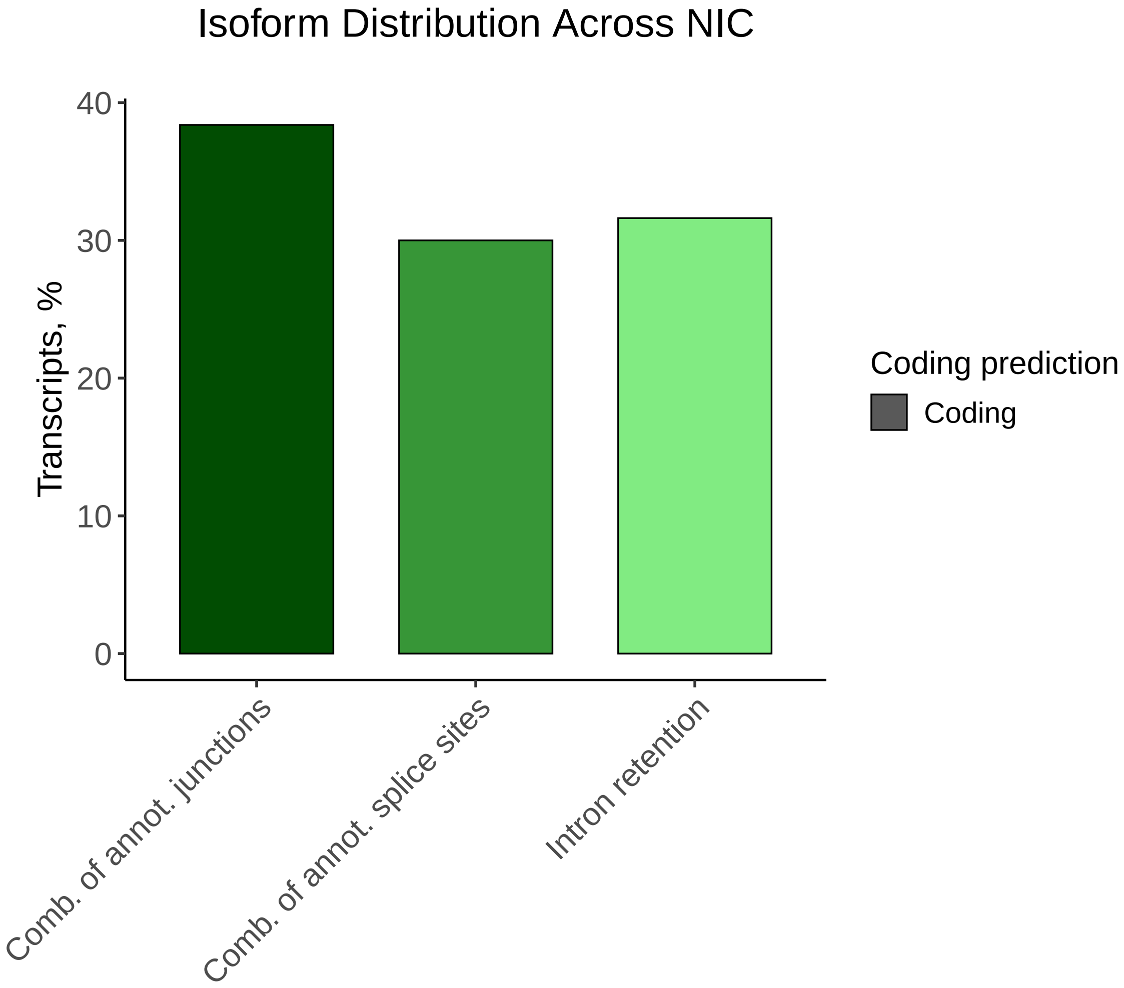
Isoform Distribution Across Structural Categories

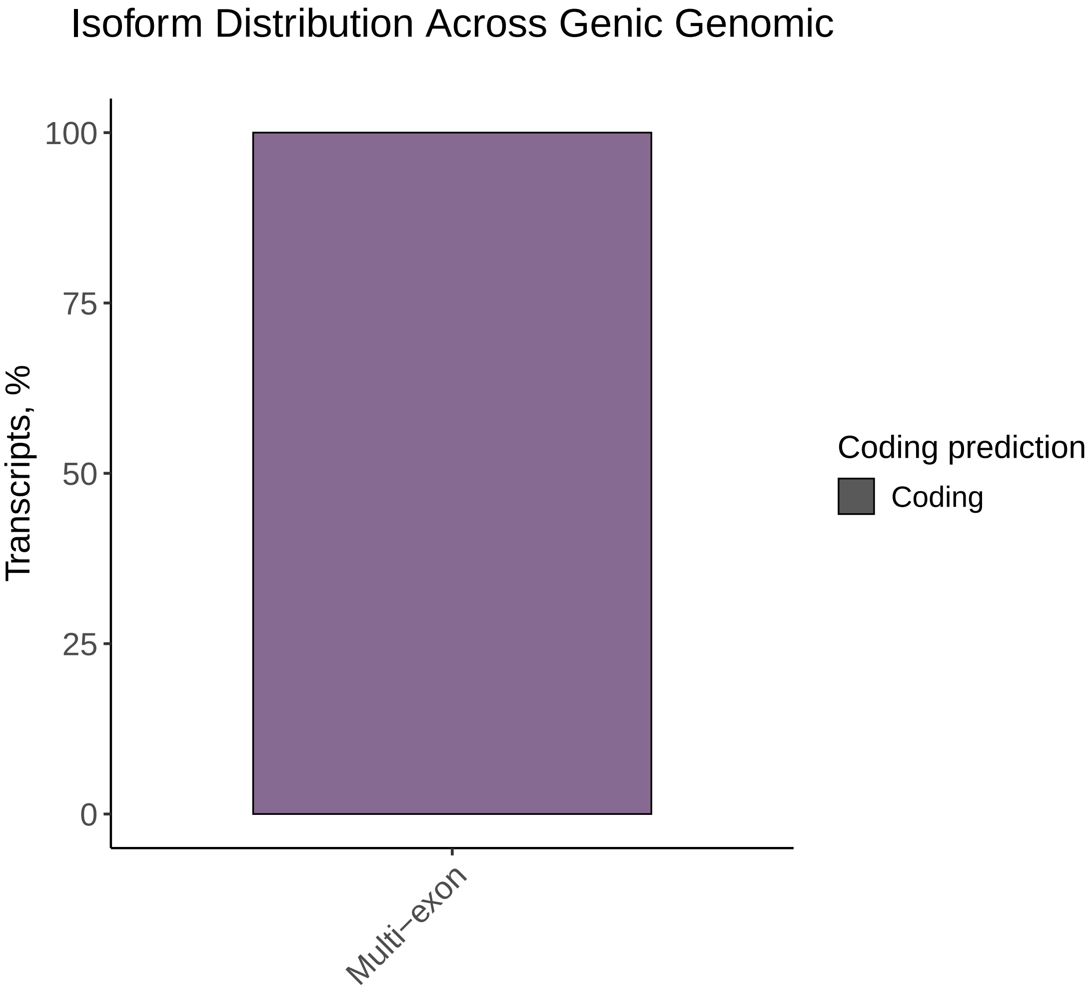


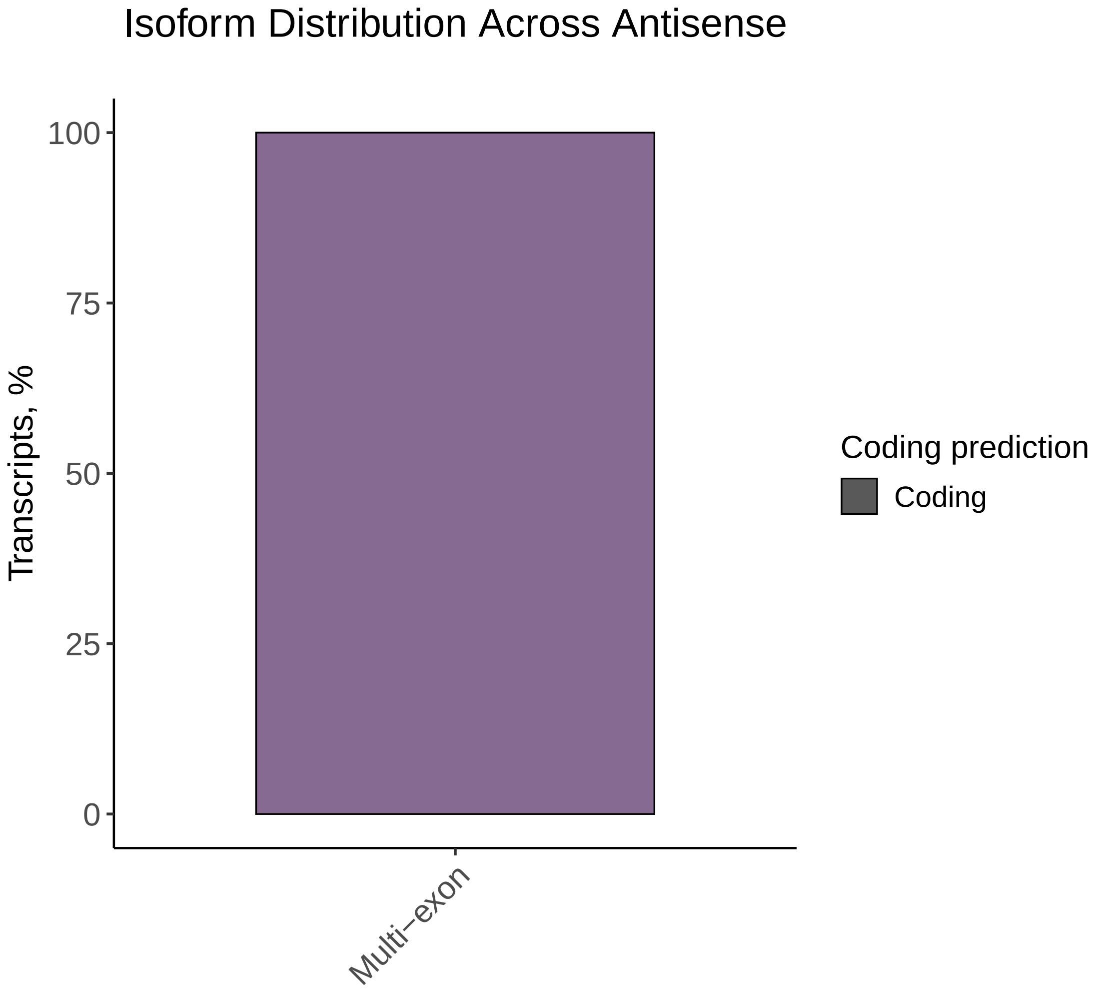

Isoform Distribution Across FSM
Transcripts, %
0
10
20
30
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
Coding prediction
Coding

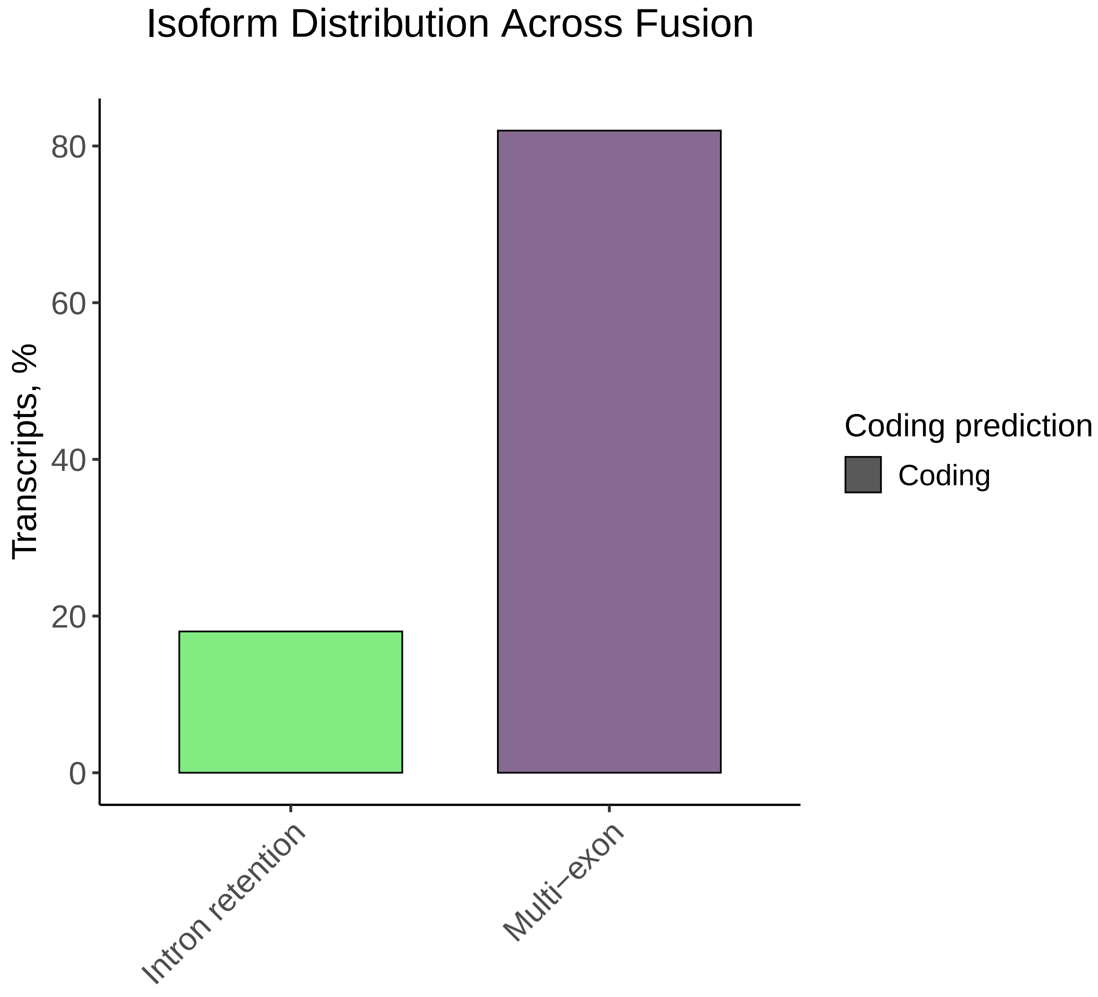

Isoform Distribution Across ISM
Transcripts, %
80
60
40
20
0
3' fragment
Internal fragment
5' fragment
Intron retention
Coding prediction
Coding

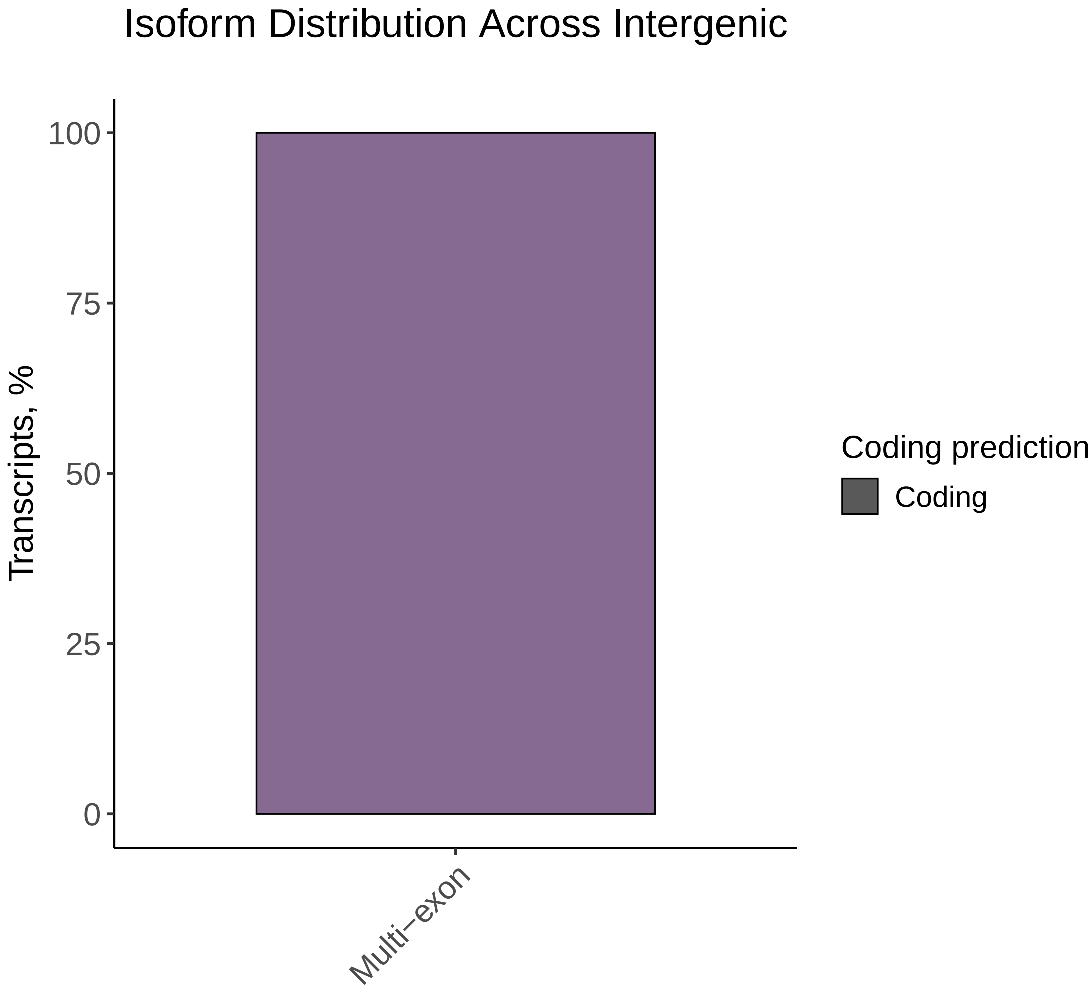

Isoform Distribution Across NNC
Transcripts, %
0
25
50
75
Intron retention
At least 1 annot. don./accept.
Coding prediction
Coding

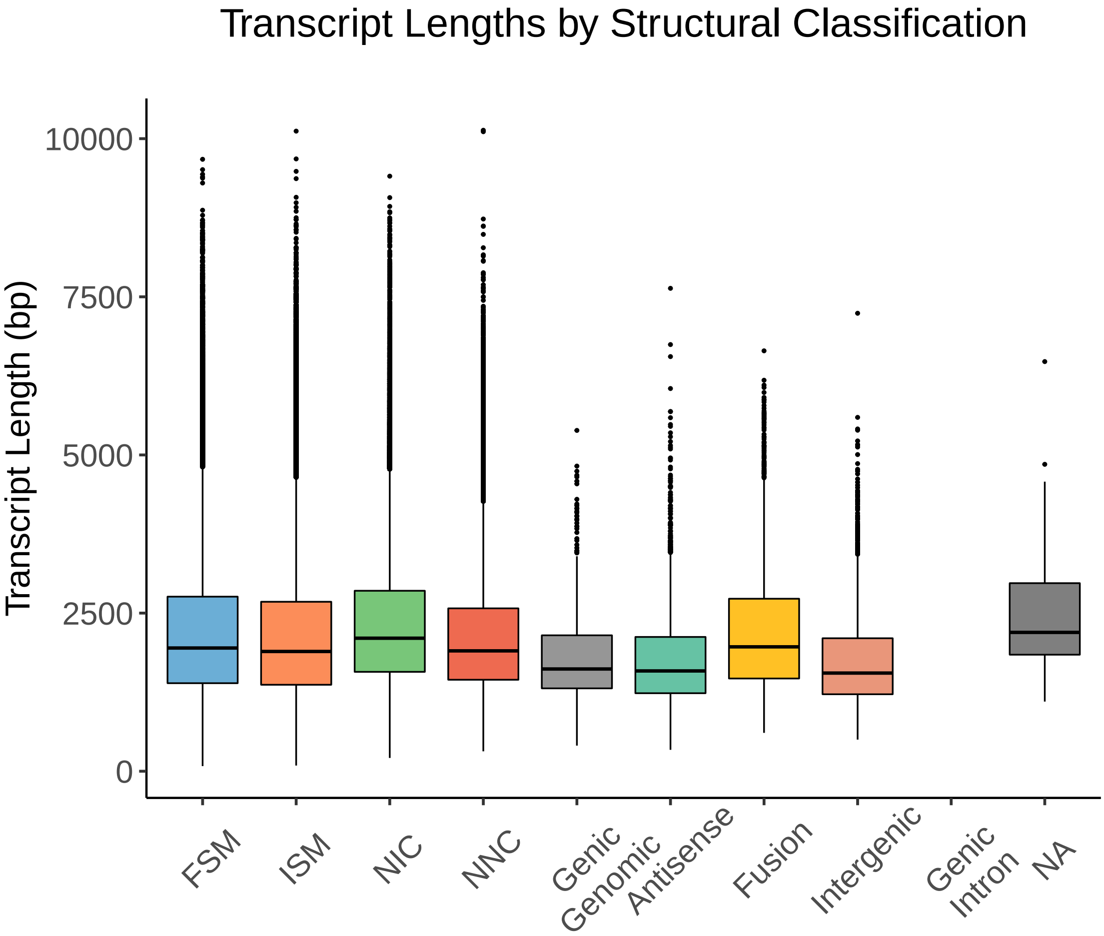

Isoform Distribution Across NIC
Transcripts, %
0
10
20
30
40
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Coding prediction
Coding

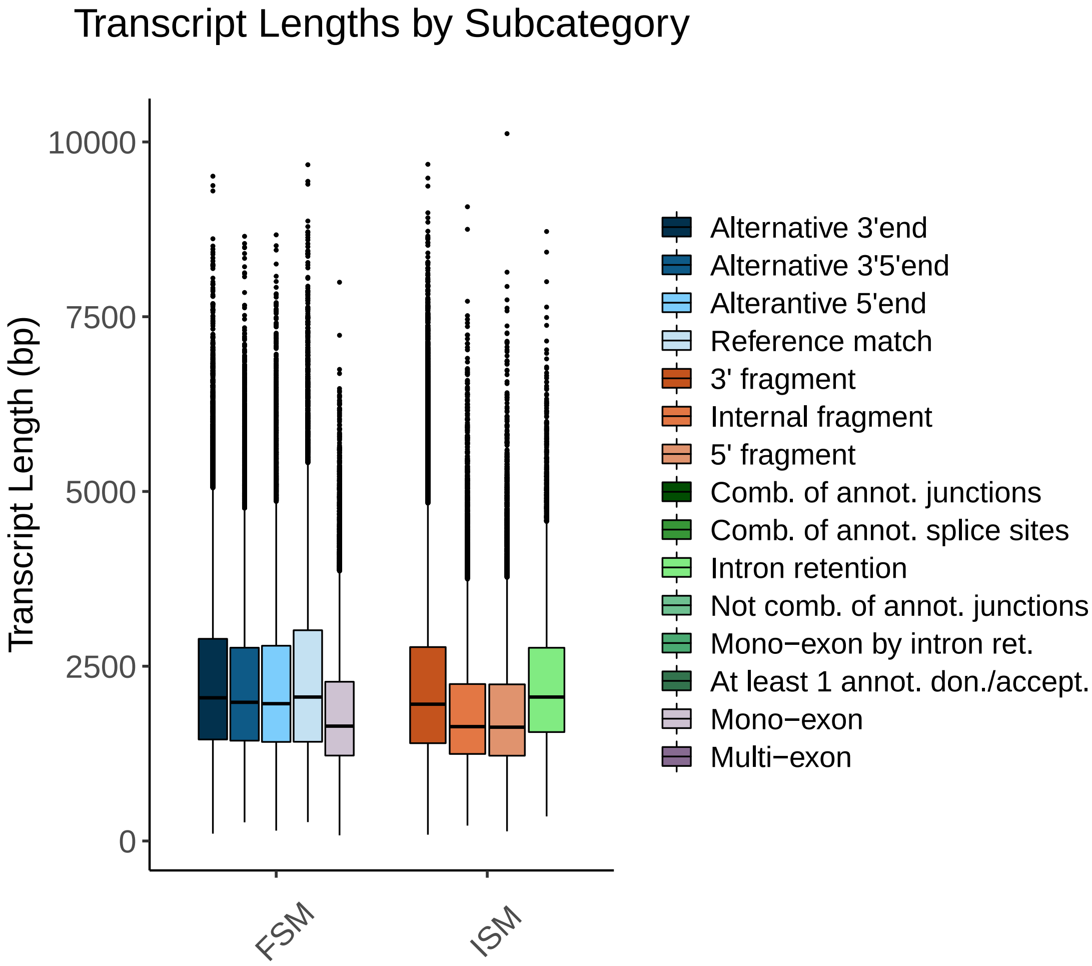

Isoform Distribution Across Genic Genomic
Transcripts, %
100
75
50
25
0
Multi–exon
Coding prediction
Coding

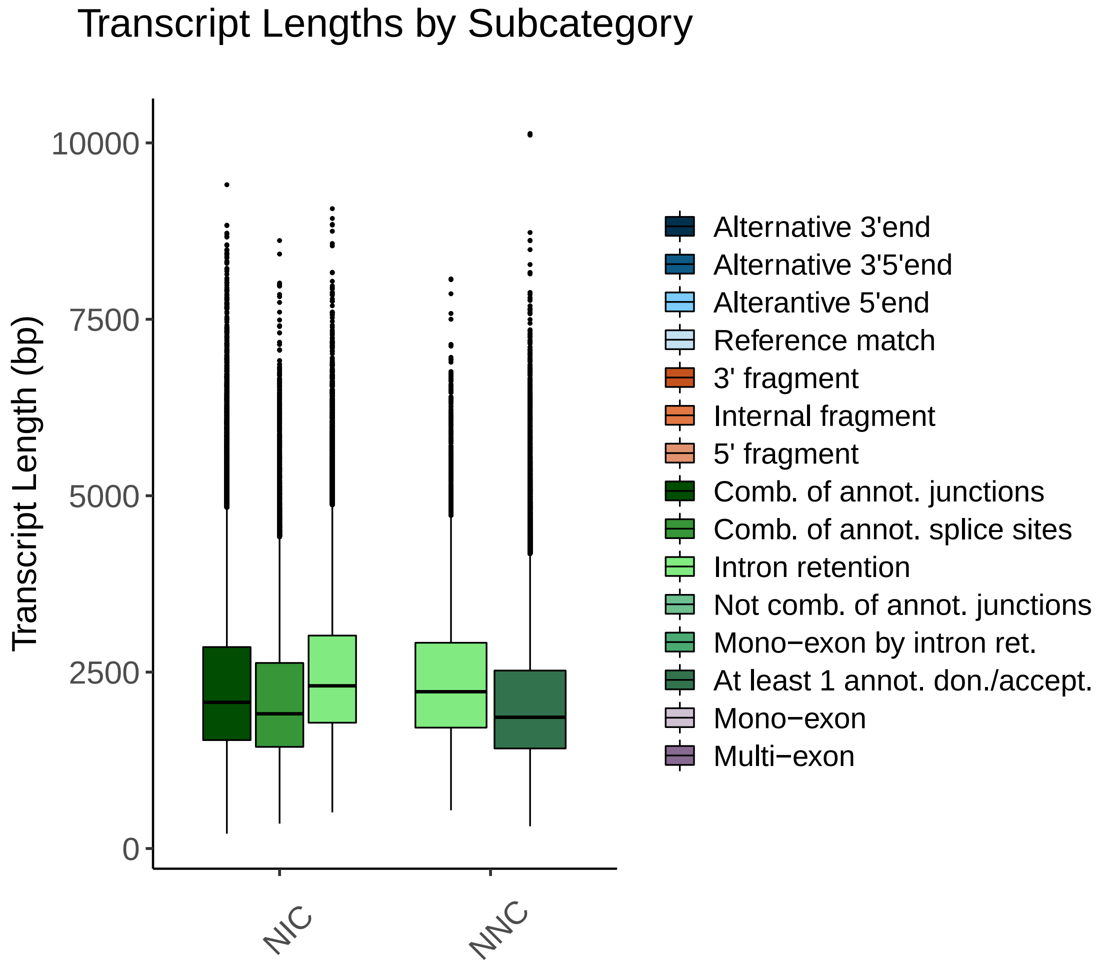

Isoform Distribution Across Antisense
Transcripts, %
100
75
50
25
0
Multi−exon
Coding prediction
Coding

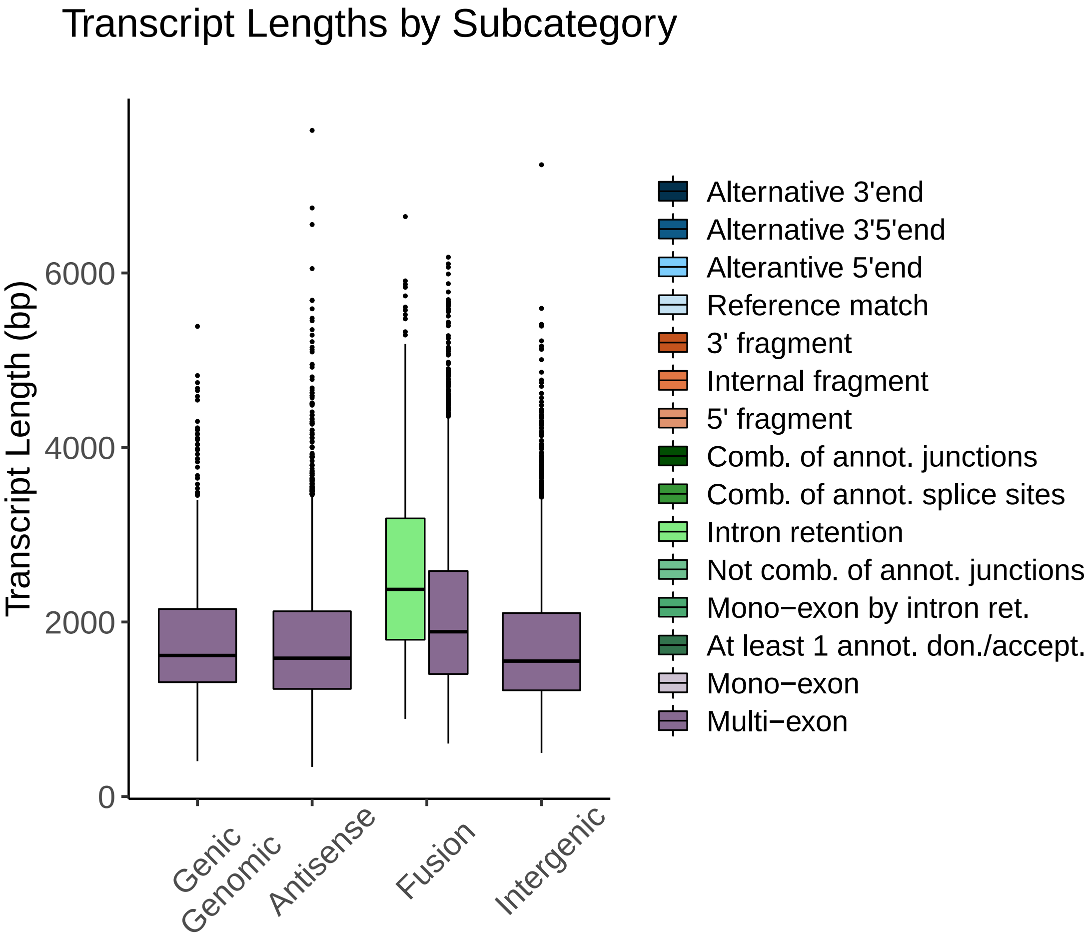

Isoform Distribution Across Fusion
Transcripts, %
80
60
40
20
0
Intron retention
Multi−exon
Coding prediction
Coding

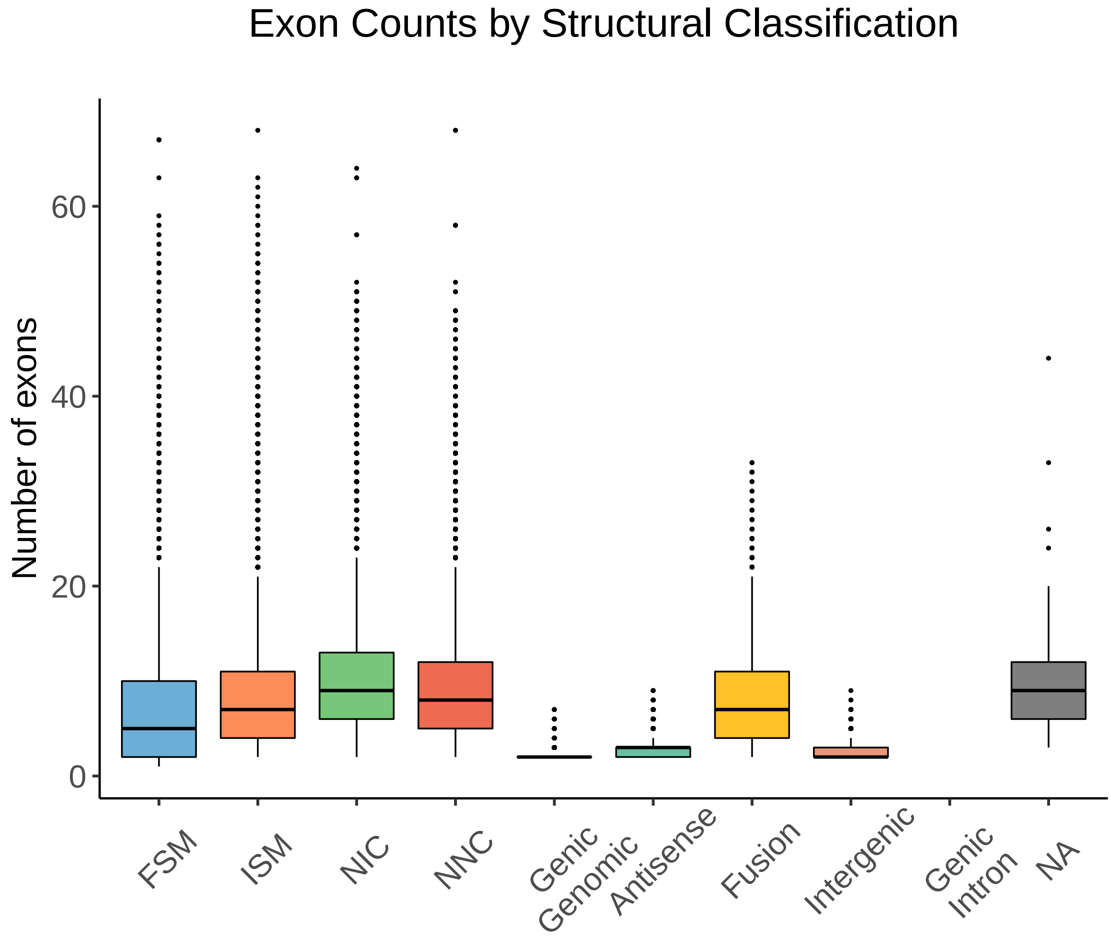

Isoform Distribution Across Intergenic
Transcripts, %
100
75
50
25
0
Multi–exon
Coding prediction
Coding

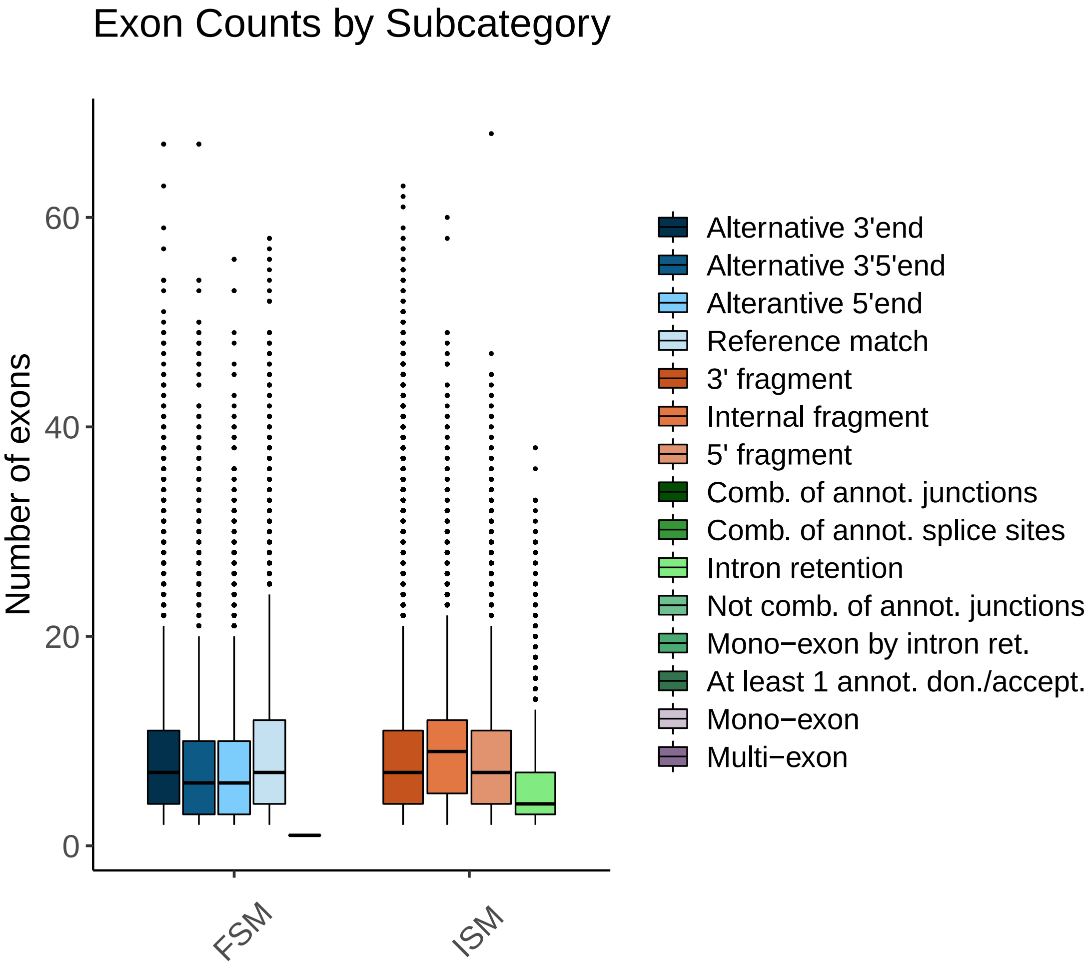

Transcript Lengths by Structural Classification
Transcript Length (bp)
10000
7500
5000
2500
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron
NA

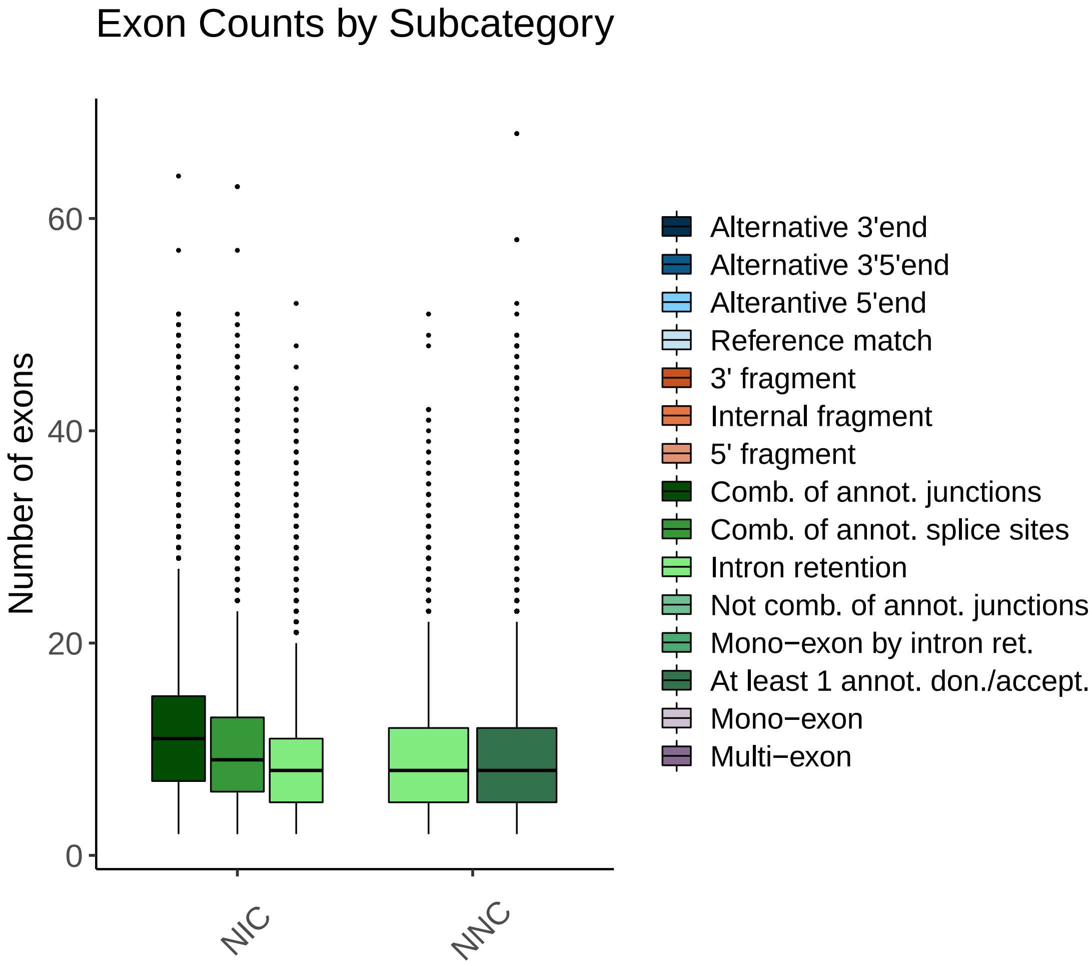

Transcript Lengths by Subcategory
Transcript Length (bp)
10000
7500
5000
2500
0
FSM
ISM
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon

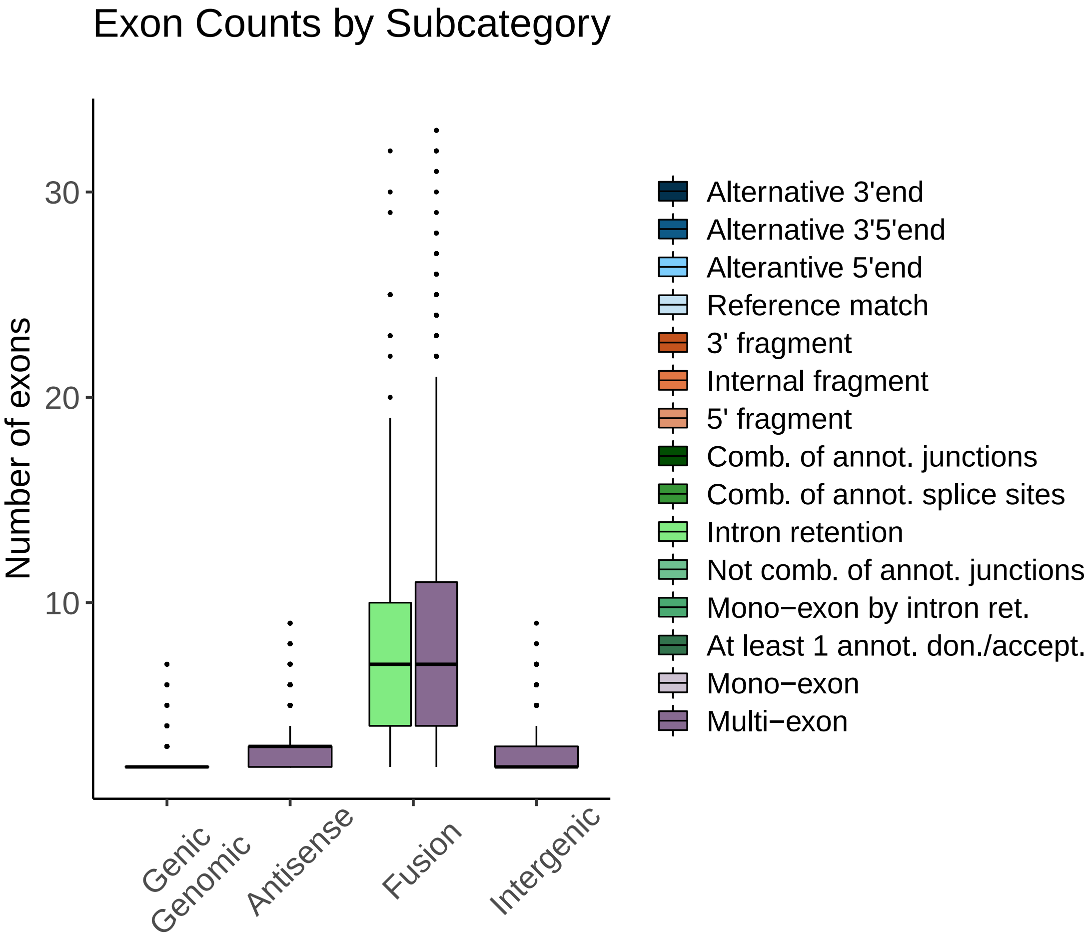

Transcript Lengths by Subcategory
Transcript Length (bp)
10000
7500
5000
2500
0
NIC
NNC
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon

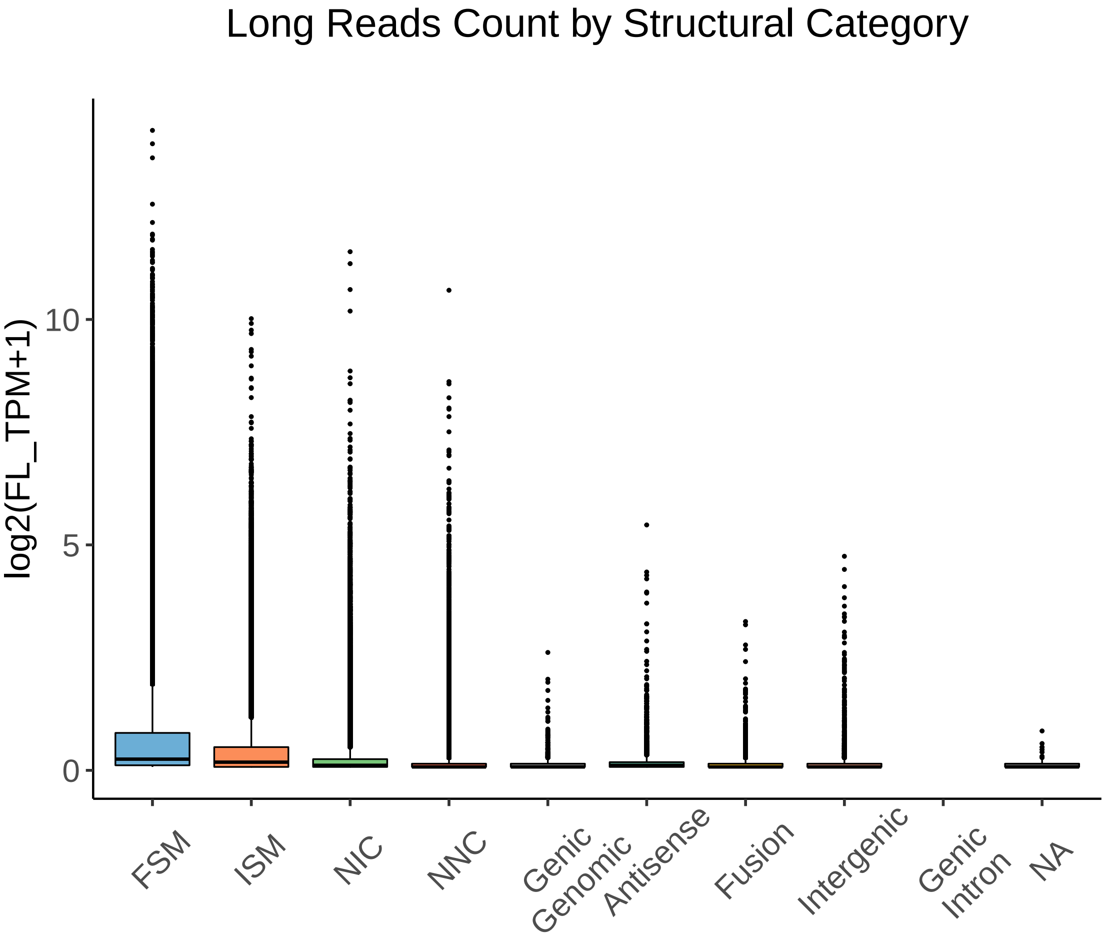

Transcript Lengths by Subcategory
Transcript Length (bp)
0
2000
4000
6000
Genic
Genomic
Antisense
Fusion
Intergenic
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
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
Number of exons
0
20
40
60
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron
NA


Exon Counts by Subcategory
Number of exons
0
20
40
60
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

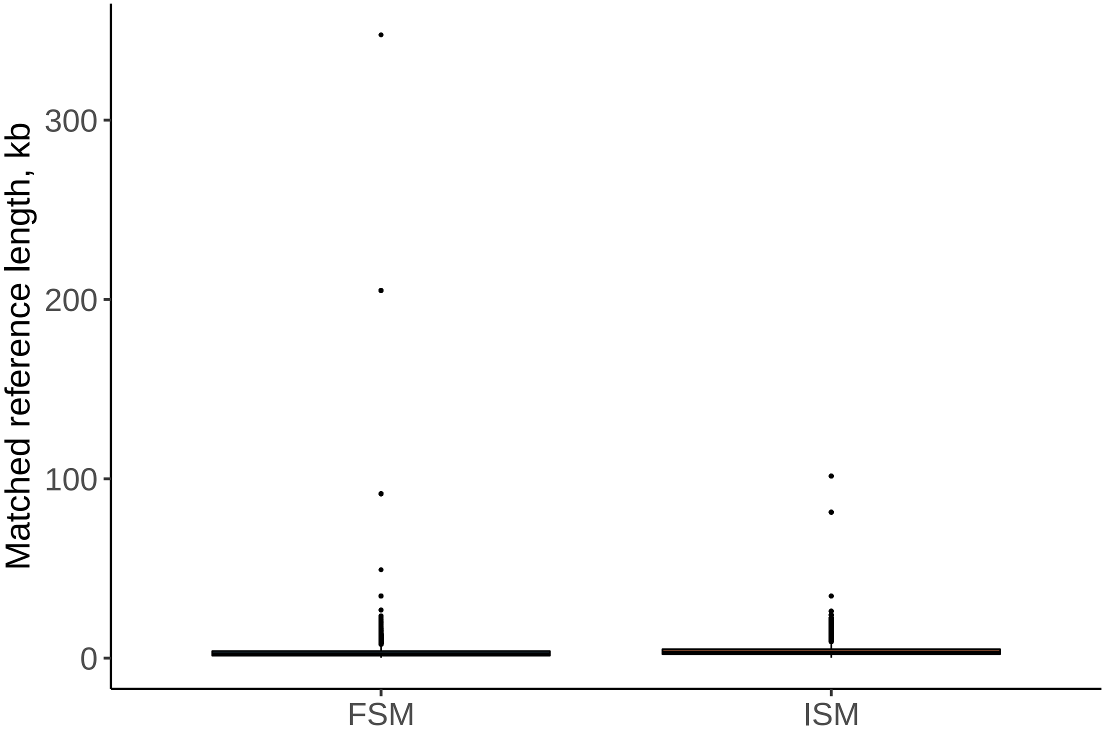

Exon Counts by Subcategory
Number of exons
0
20
40
60
NIC
NNC
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon

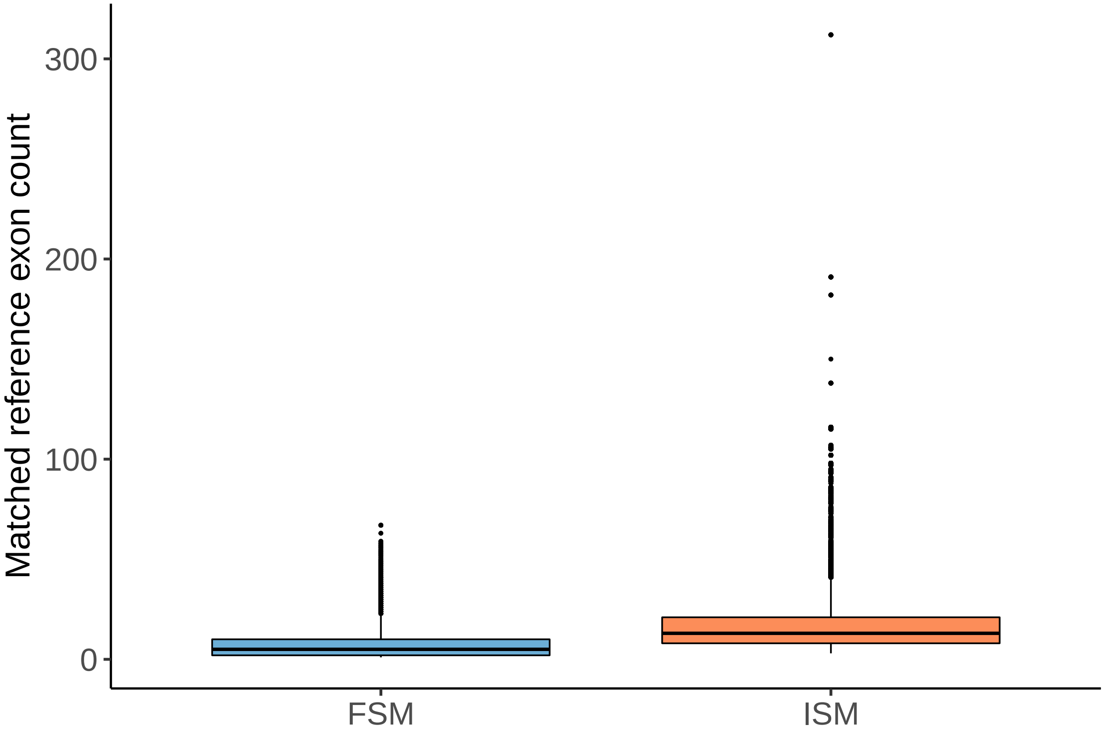

Exon Counts by Subcategory
Number of exons
30
20
10
Genic Genomic
Antisense
Fusion
Intergenic
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
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

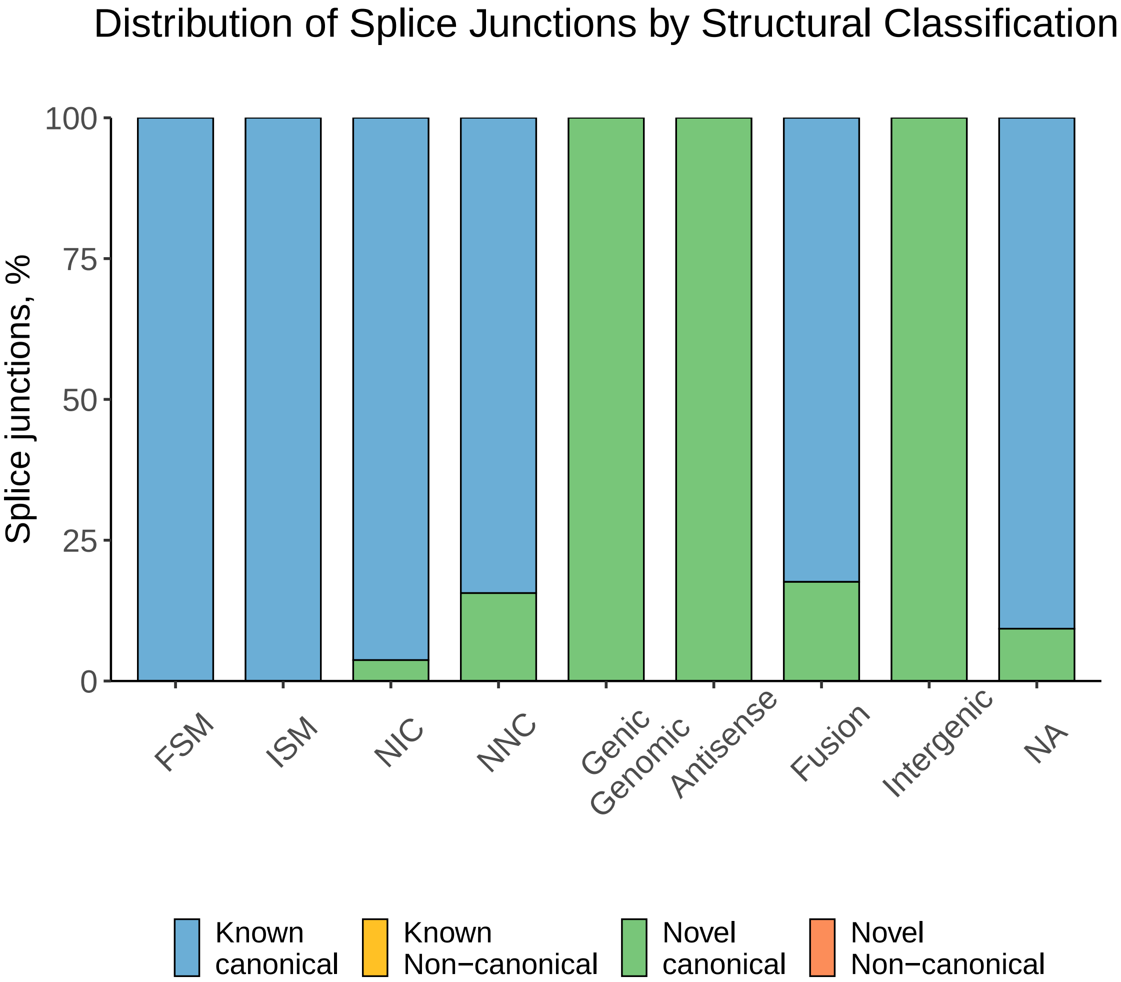


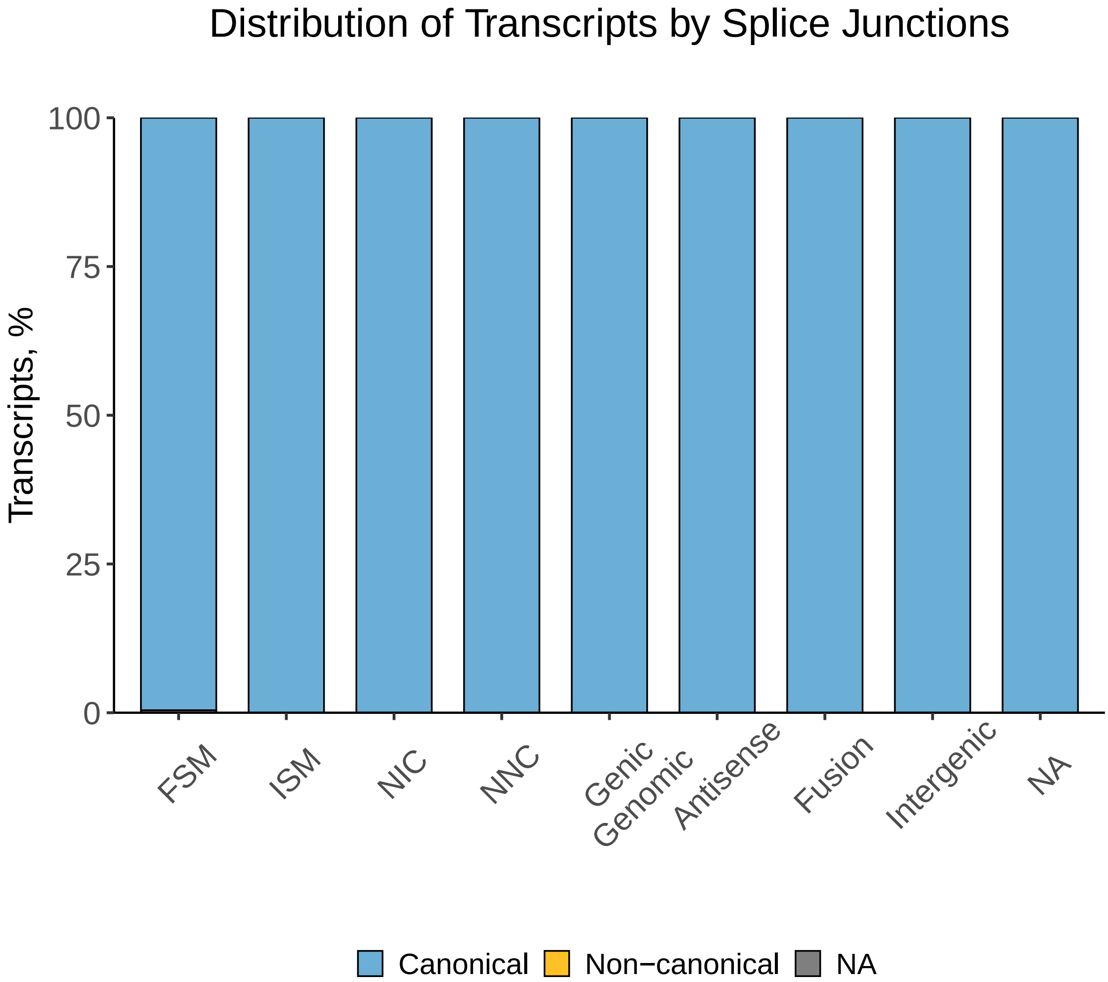


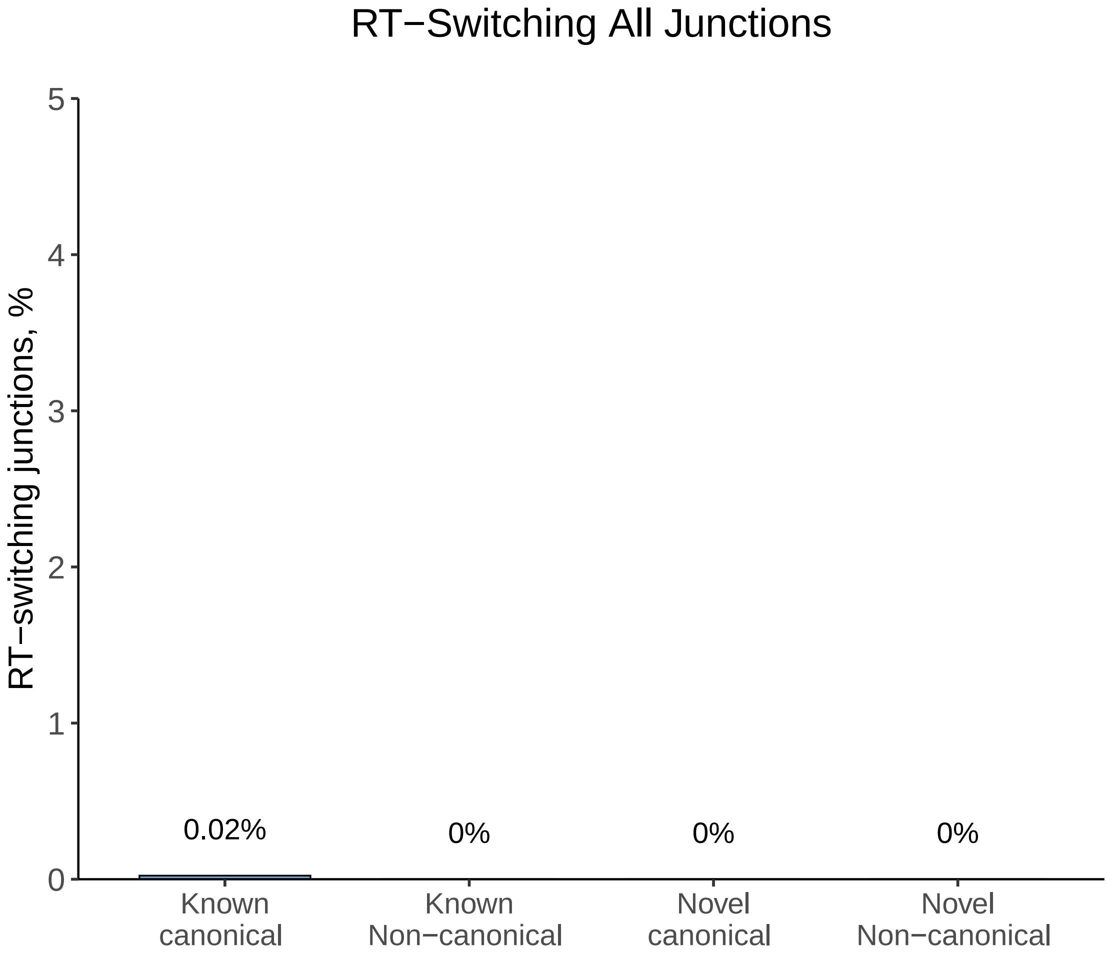
Exon Count Distribution of Matched Reference Transcripts

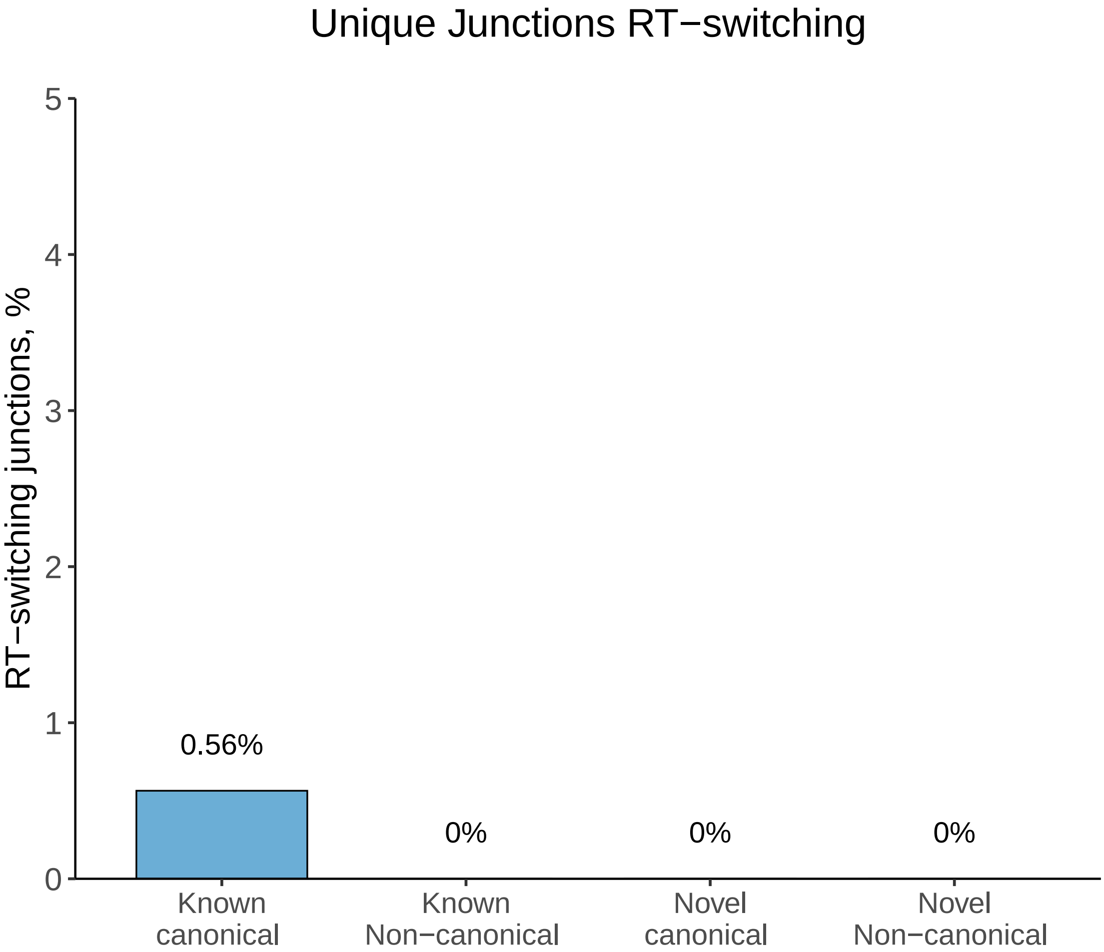
Applicable Only to FSM and ISM Categories

Splice Junction Characterization

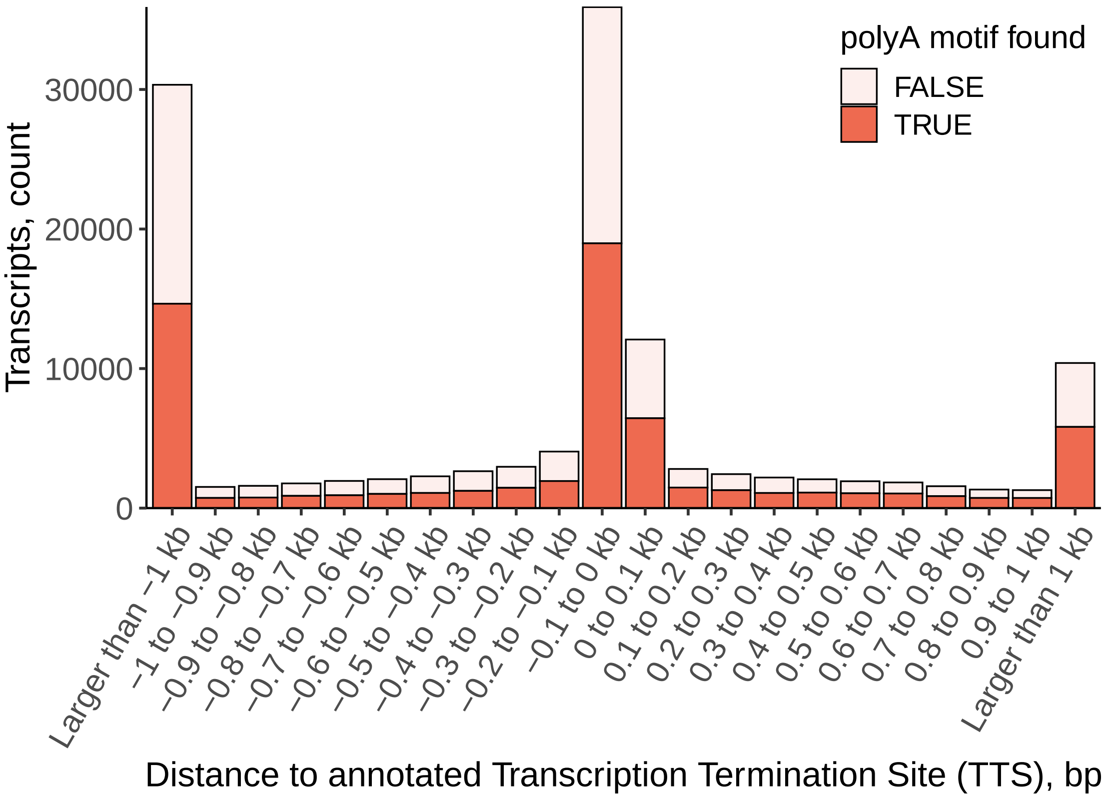

Distribution of Splice Junctions by Structural Classification
Splice junctions, %
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
NA
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical


Distribution of Transcripts by Splice Junctions
Transcripts, %
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic
Genomic
Antisense
Fusion
Intergenic
NA
Canonical
Non−canonical
NA


RT−Switching All Junctions
RT−switching junctions, %
5
4
3
2
1
0
0.02%
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
0.56%
0%
0%
0%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

Comparison With Annotated TSS and TTS

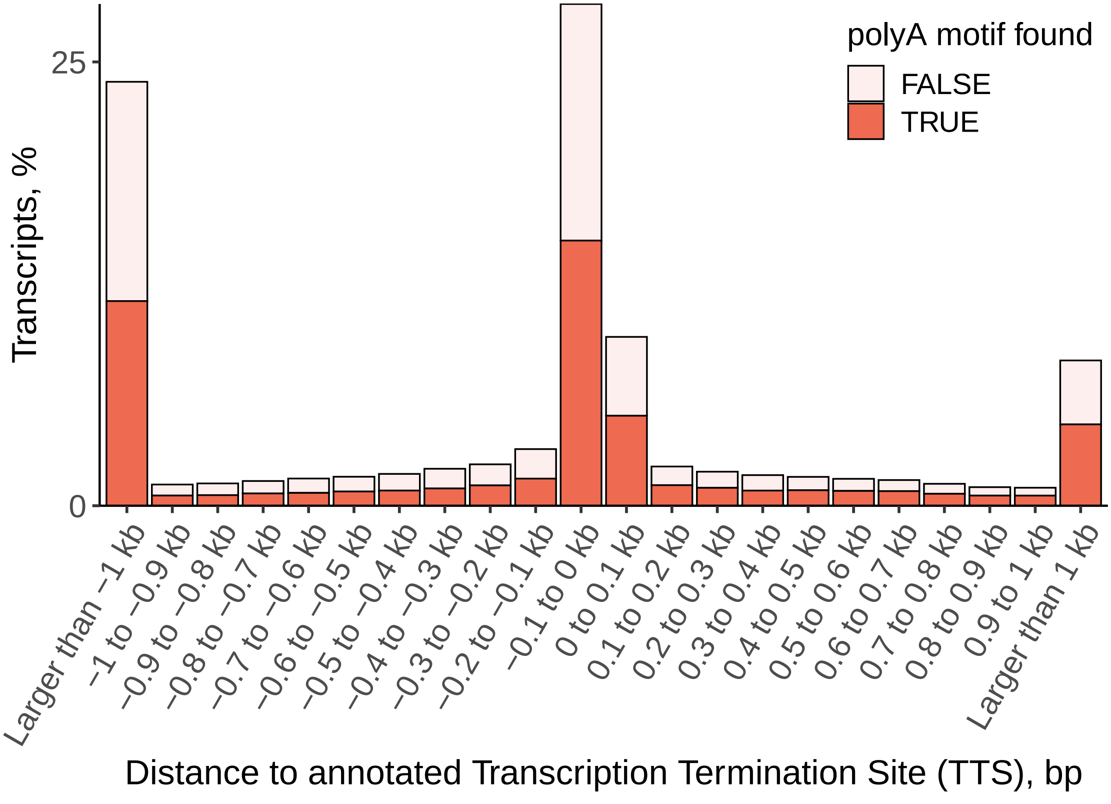


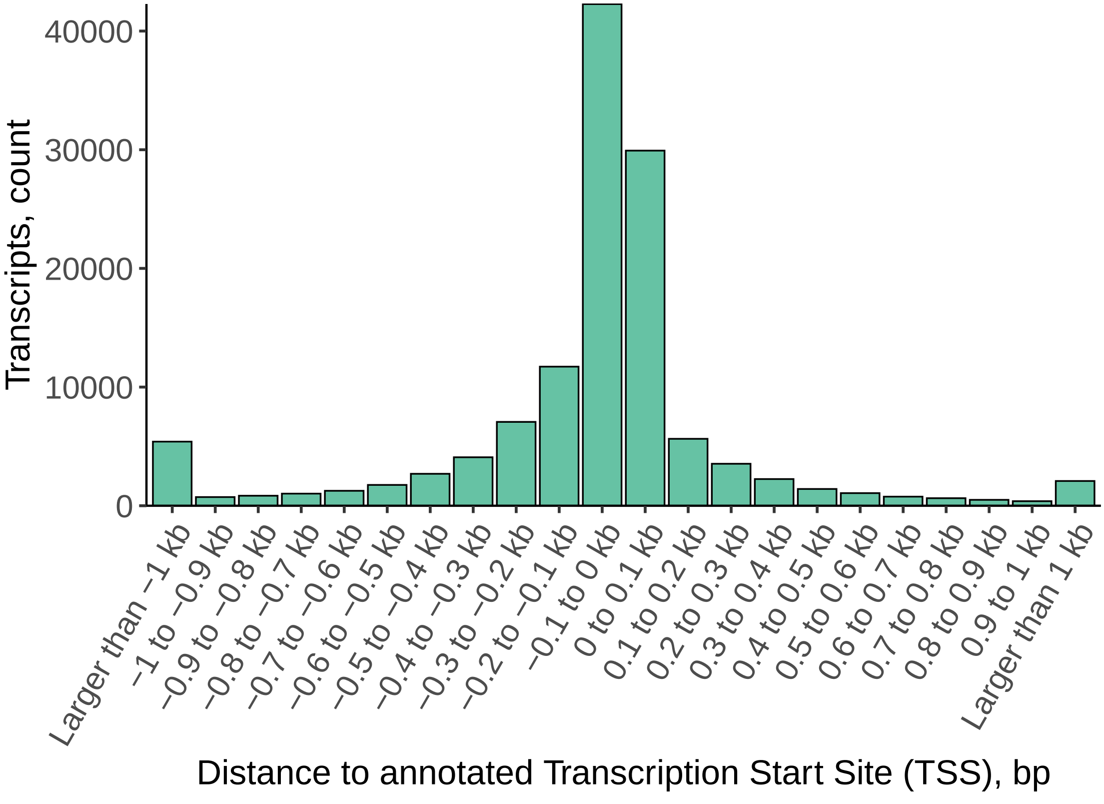
Distance to annotated Transcription Termination Site (TTS)
FSM

Negative values indicate upstream of annotated termination site

Distance to annotated Transcription Termination Site (TTS)
FSM

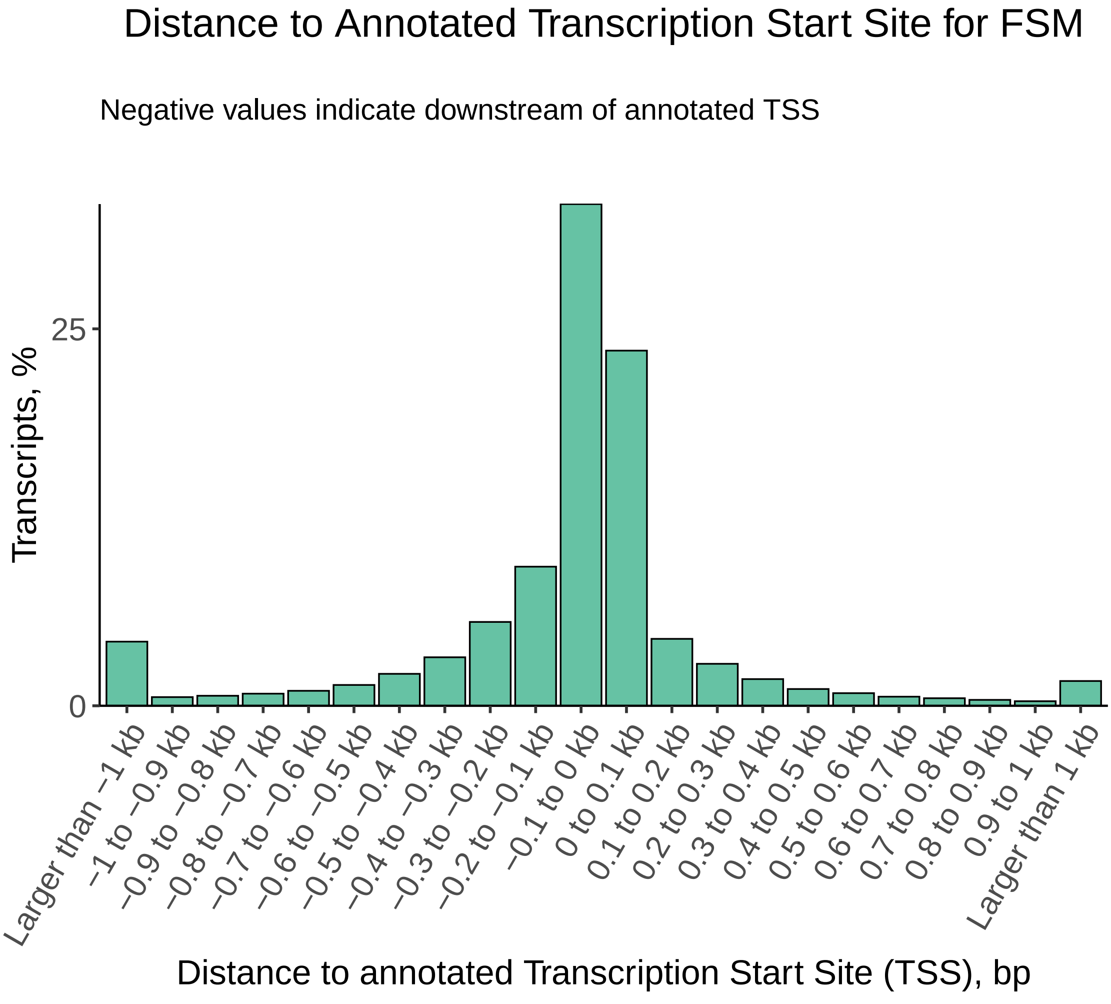
Negative values indicate upstream of annotated termination site

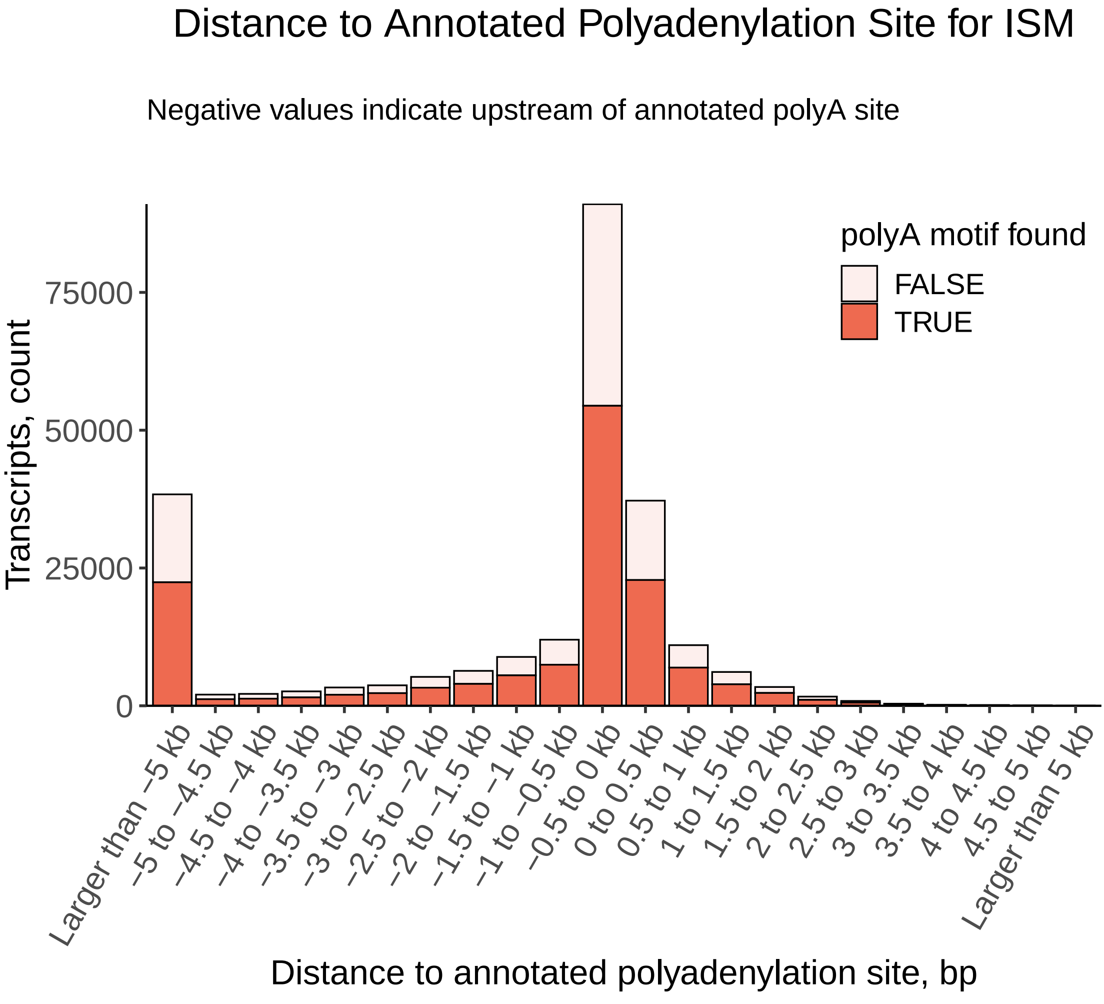


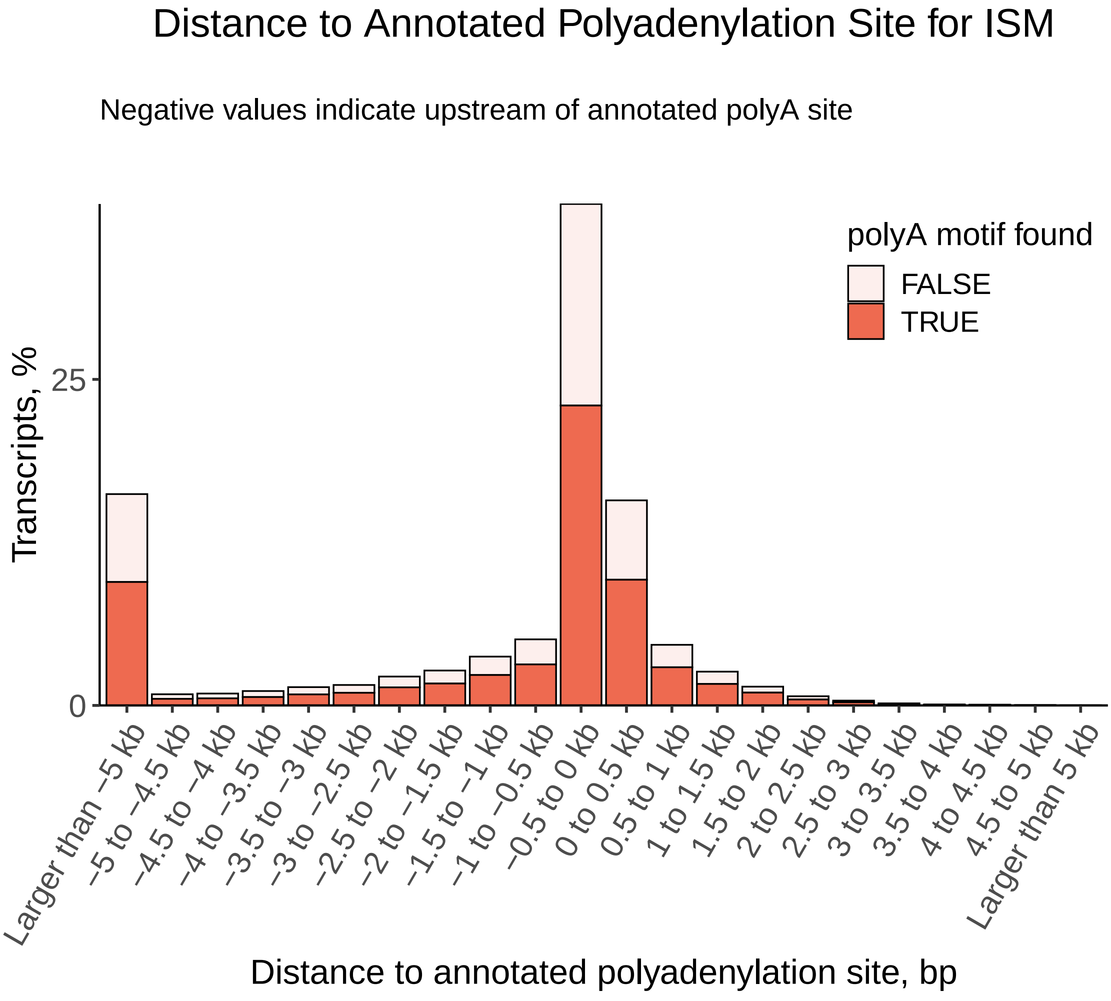


Distance to Annotated Transcription Start Site for FSM

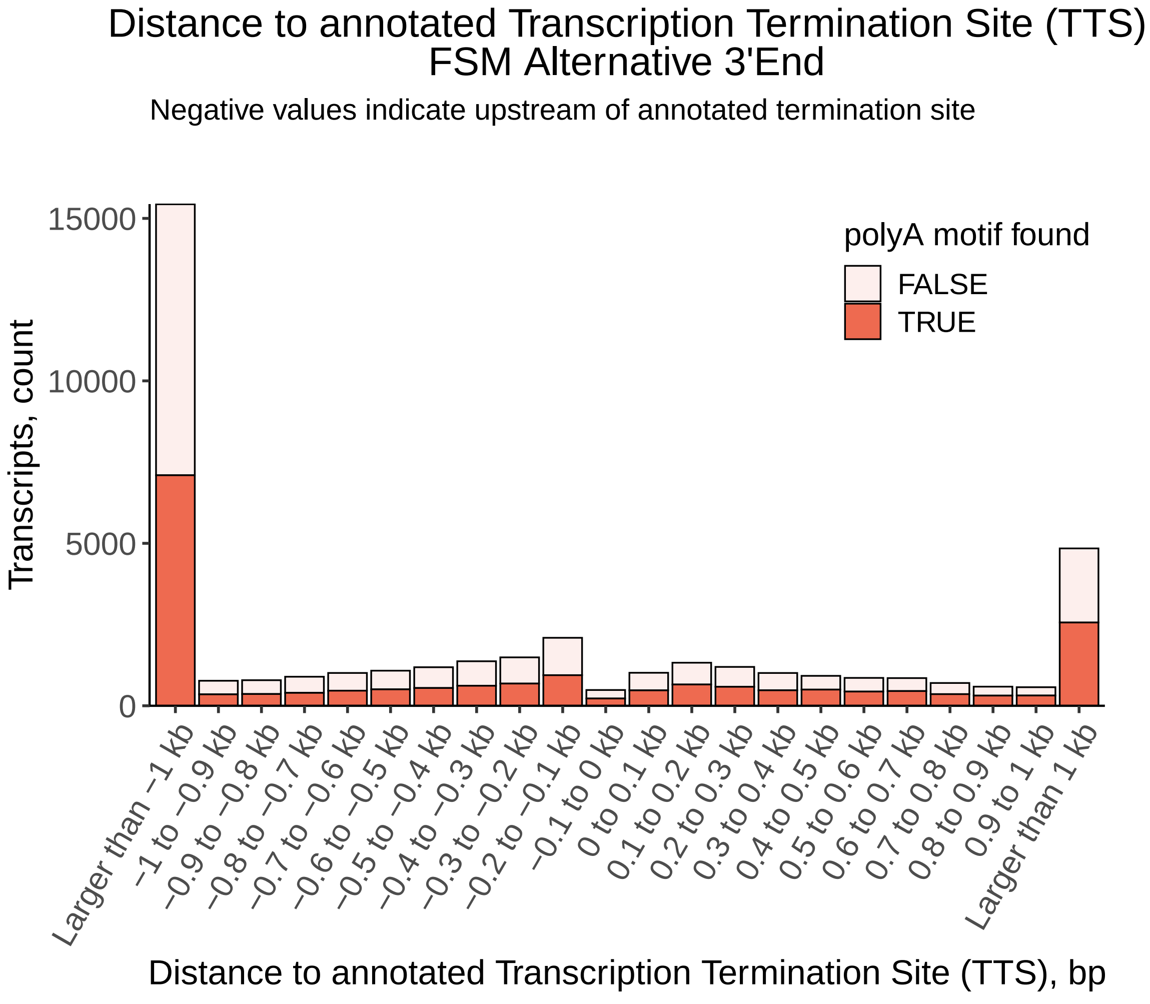
Negative values indicate downstream of annotated TSS


Distance to Annotated Transcription Start Site for FSM
Negative values indicate downstream of annotated TSS
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 Start Site (TSS), bp


Distance to Annotated Polyadenylation Site for ISM
Negative values indicate upstream of annotated polyA site
Transcripts, count
75000
50000
25000
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

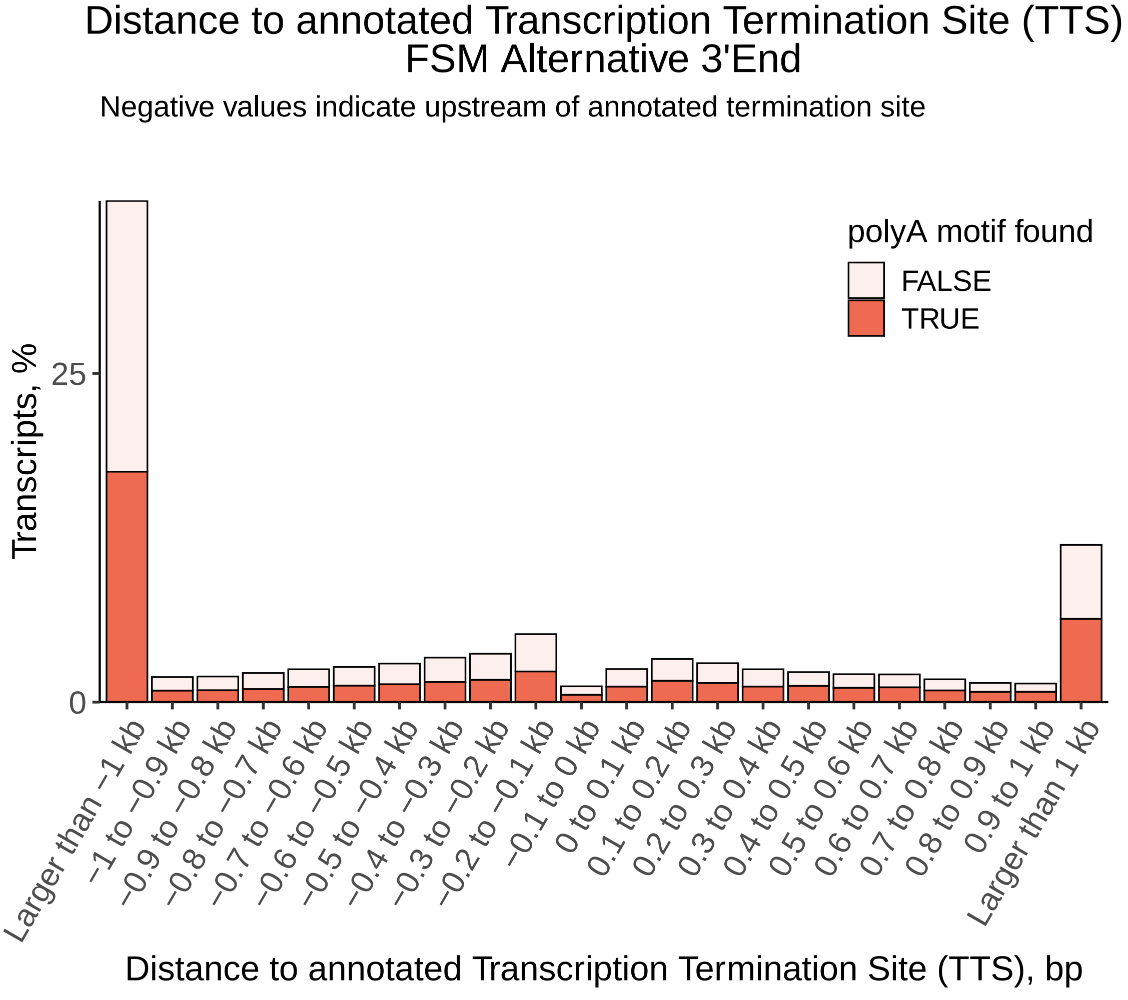

Distance to Annotated Polyadenylation Site for ISM
Negative values indicate upstream of annotated polyA site
Transcripts, %
25
0
polyA motif found
FALSE
TRUE
Larger than −5 kb
−5 to −4.5 kb
−4.5 to −4 kb
−4 to −3.5 kb
−3.5 to −3 kb
−3 to −2.5 kb
−2.5 to −2 kb
−2 to −1.5 kb
−1.5 to −1 kb
−1 to −0.5 kb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

Comparison With Annotated TSS and TTS by Subcategories



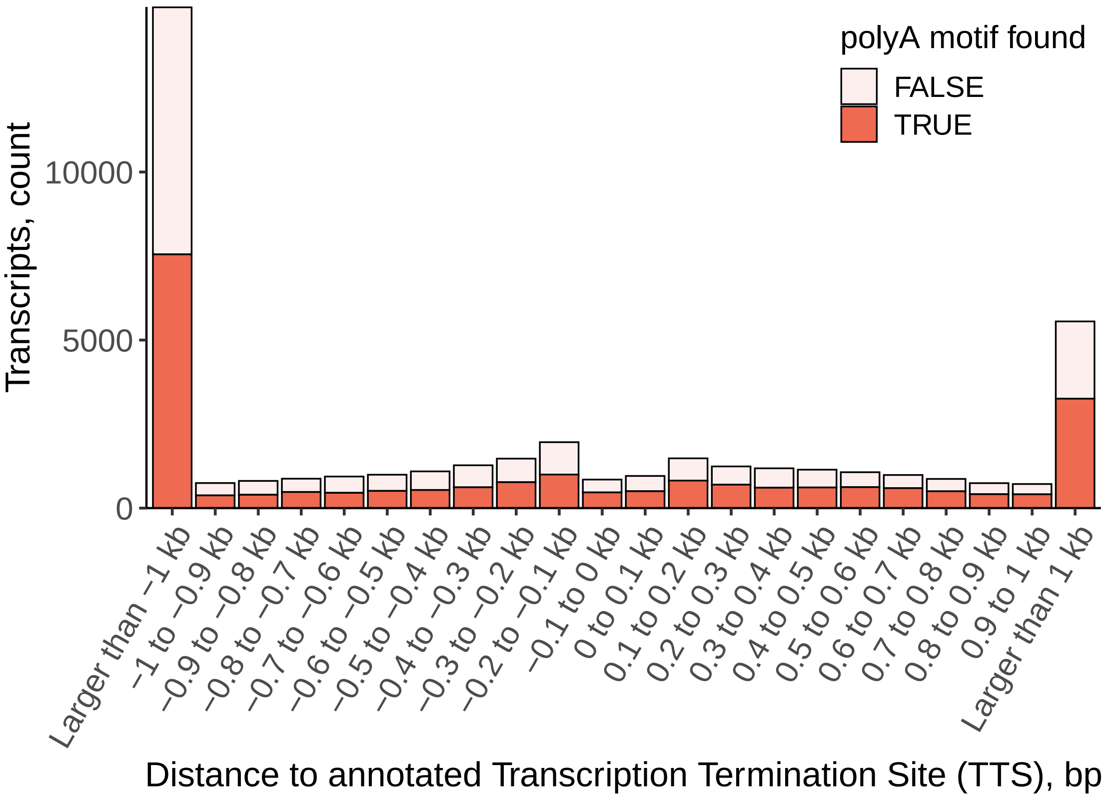
Distance to annotated Transcription Termination Site (TTS)
FSM Alternative 3'End

Negative values indicate upstream of annotated termination site



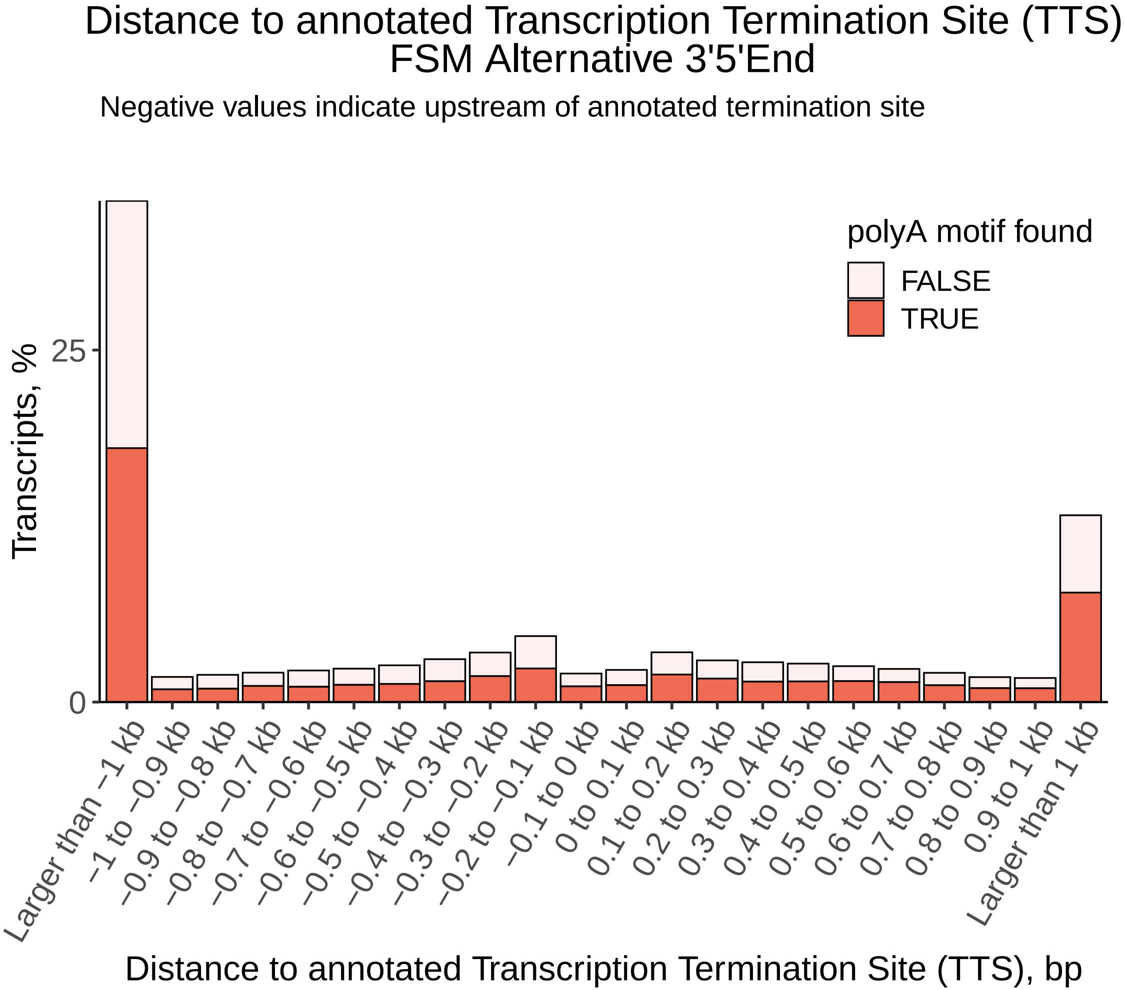
Distance to annotated Transcription Termination Site (TTS)
FSM Alternative 3'End

Negative values indicate upstream of annotated termination site



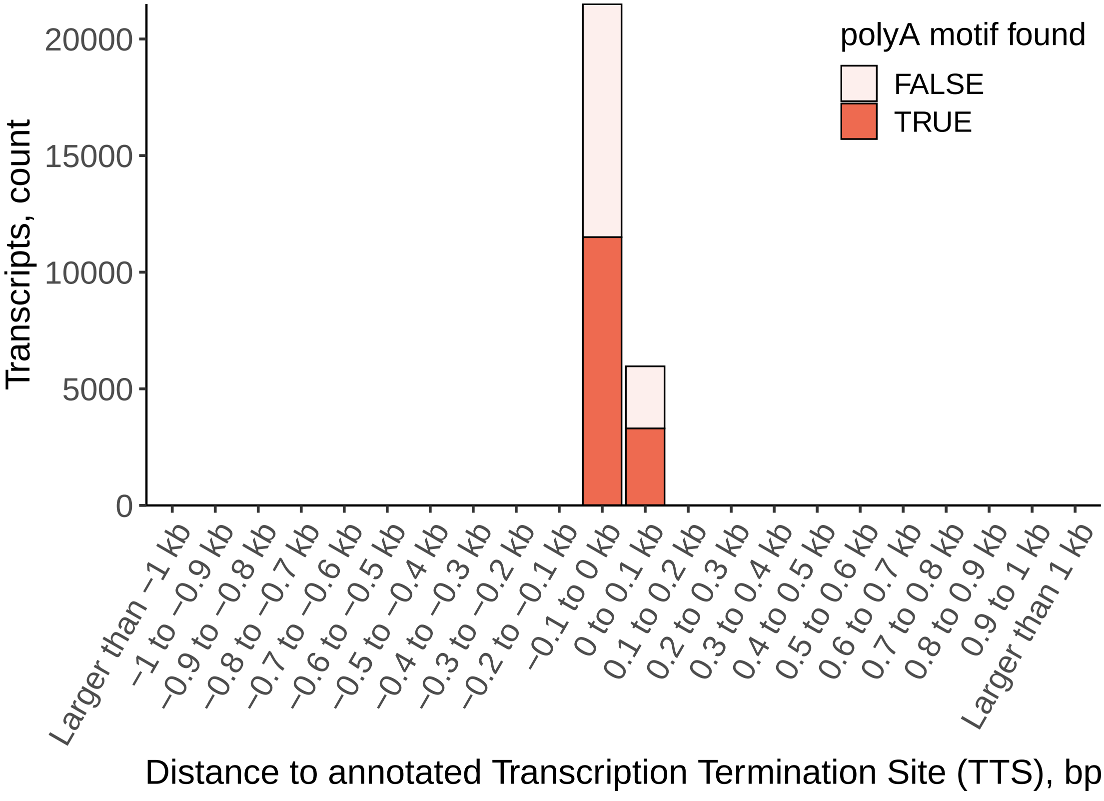
Distance to annotated Transcription Termination Site (TTS)
FSM Alternative 3'5'End

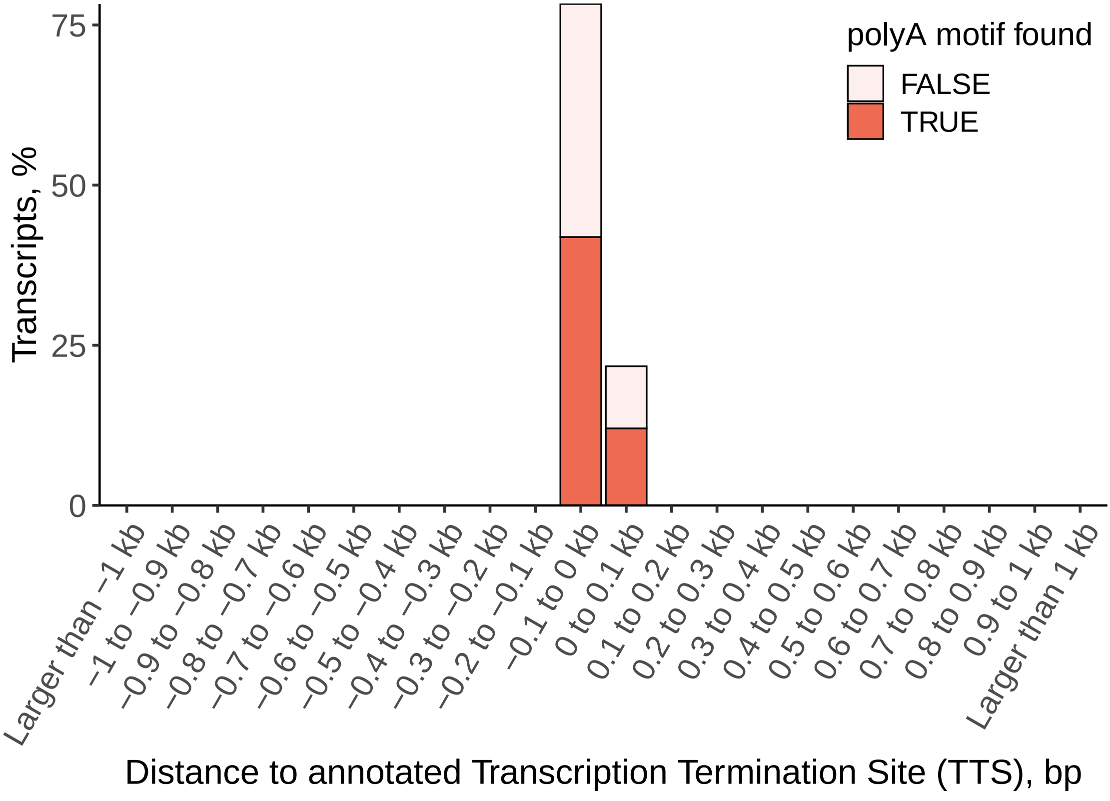
Negative values indicate upstream of annotated termination site


Distance to annotated Transcription Termination Site (TTS)
FSM Alternative 3'5'End
Negative values indicate upstream of annotated termination site
Transcripts, %
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 Alternative 5'End

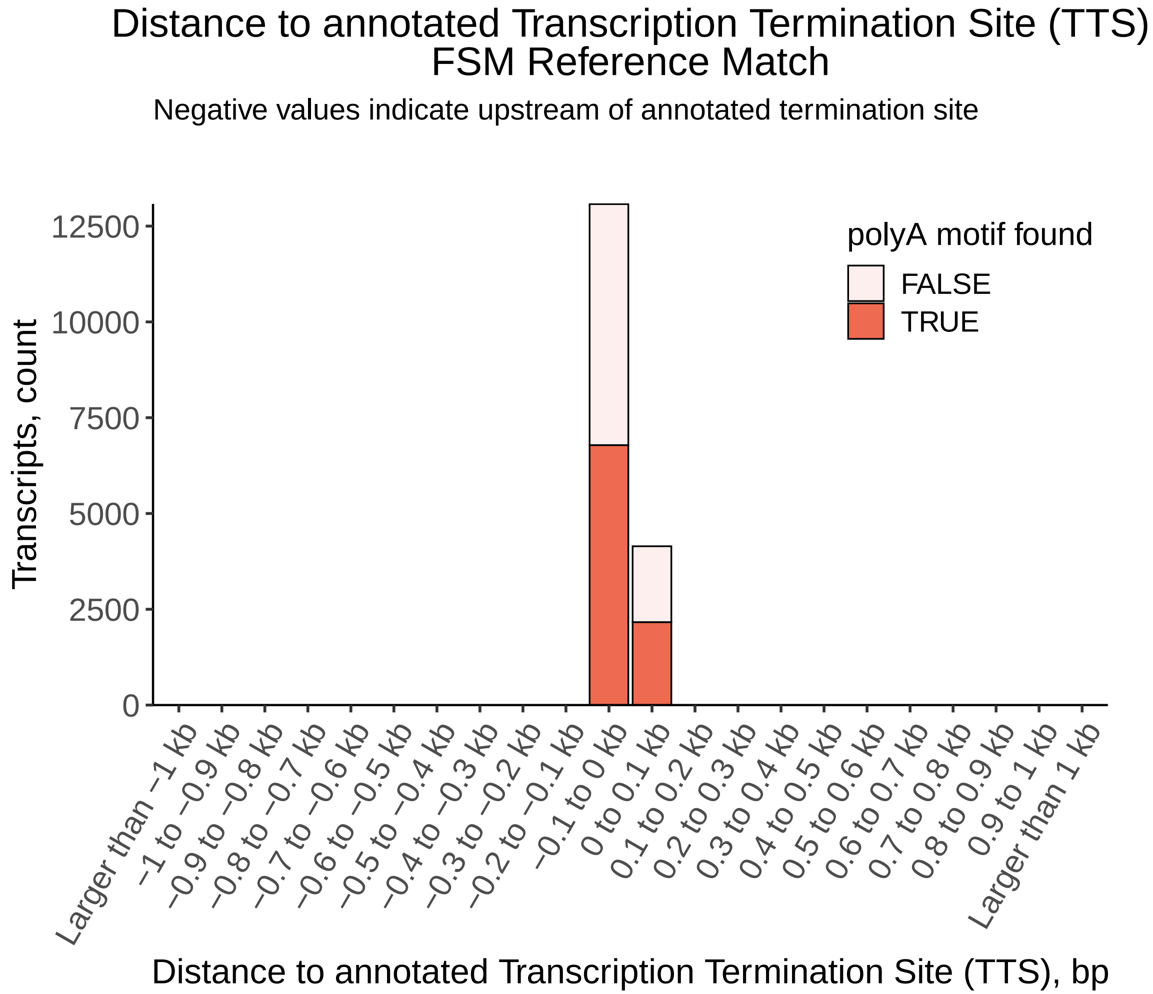
Negative values indicate upstream of annotated termination site





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

Negative values indicate upstream of annotated termination site

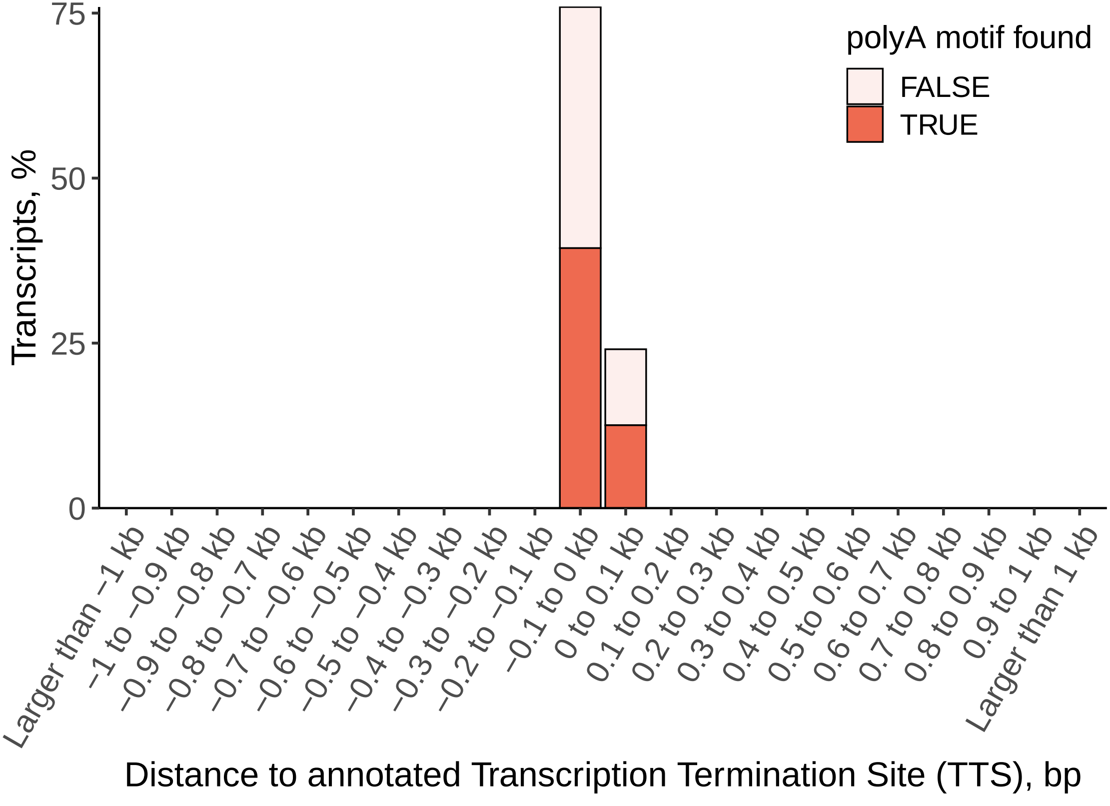


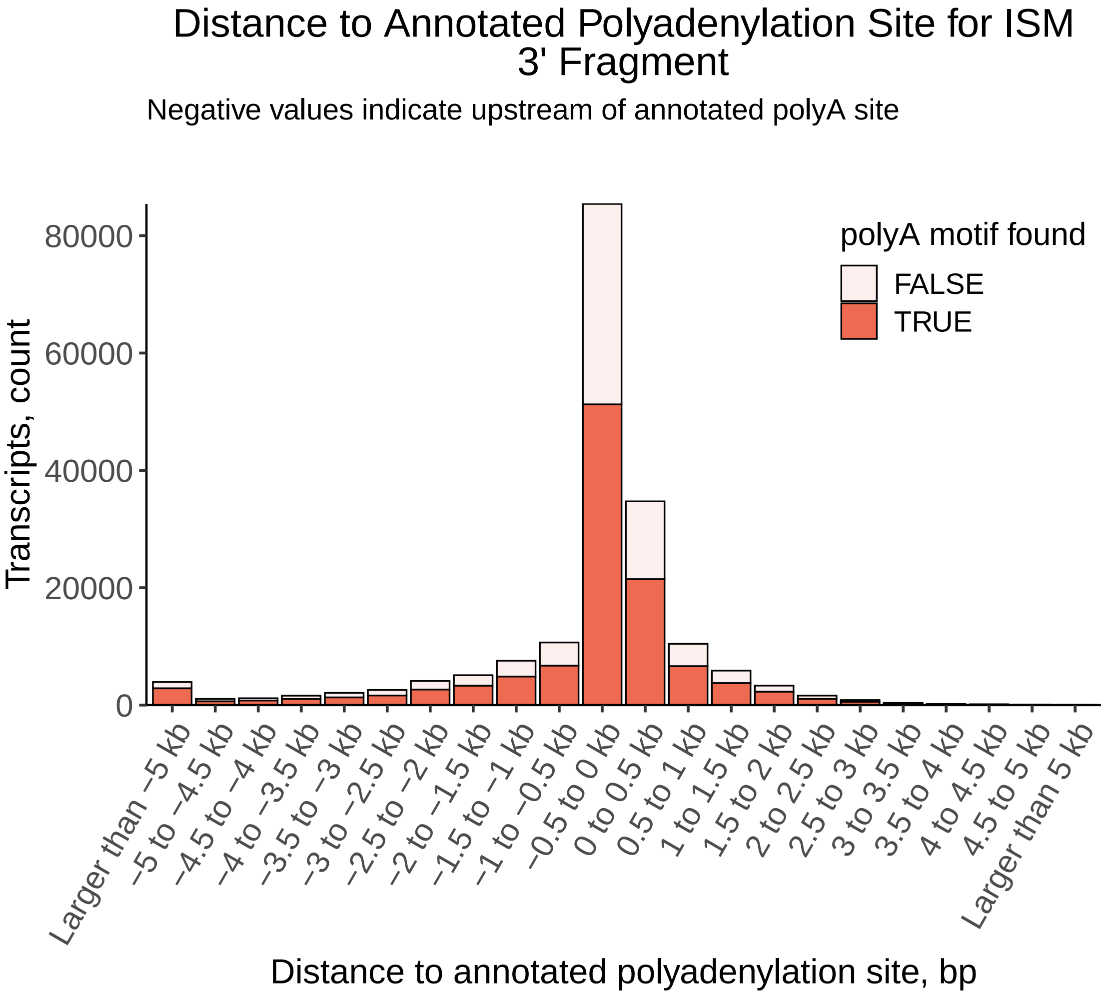
Distance to annotated Transcription Termination Site (TTS)
FSM Reference Match

Negative values indicate upstream of annotated termination site

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

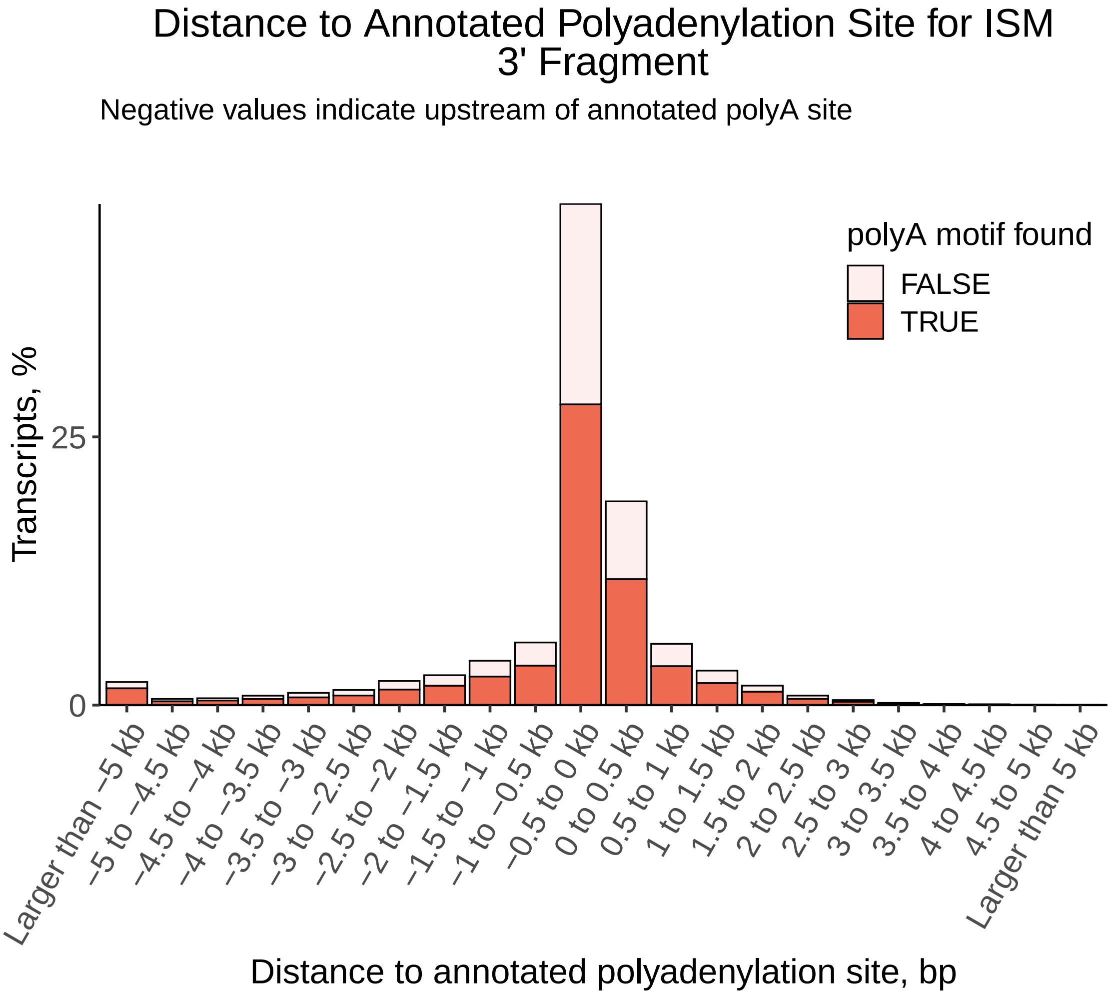
Negative values indicate upstream of annotated termination site





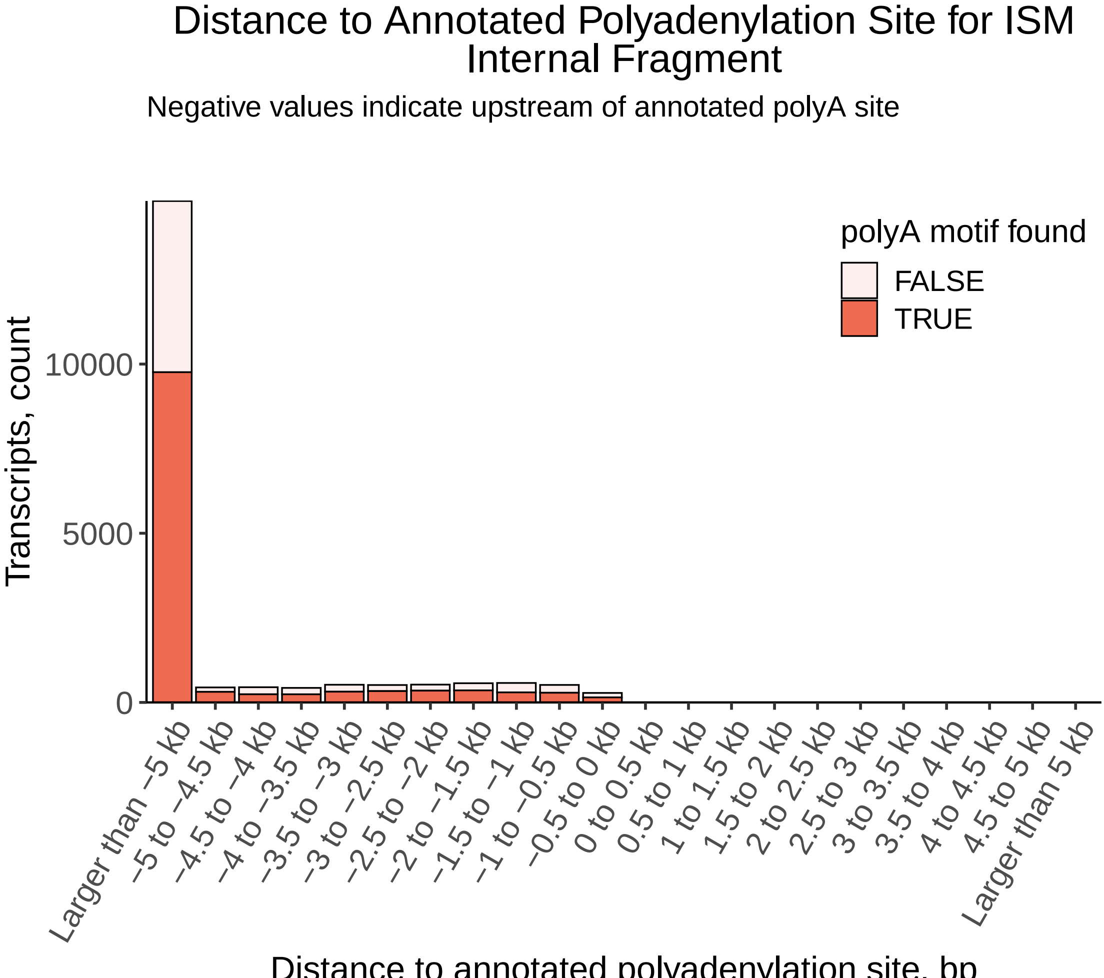
Distance to Annotated Polyadenylation Site for ISM 3' Fragment

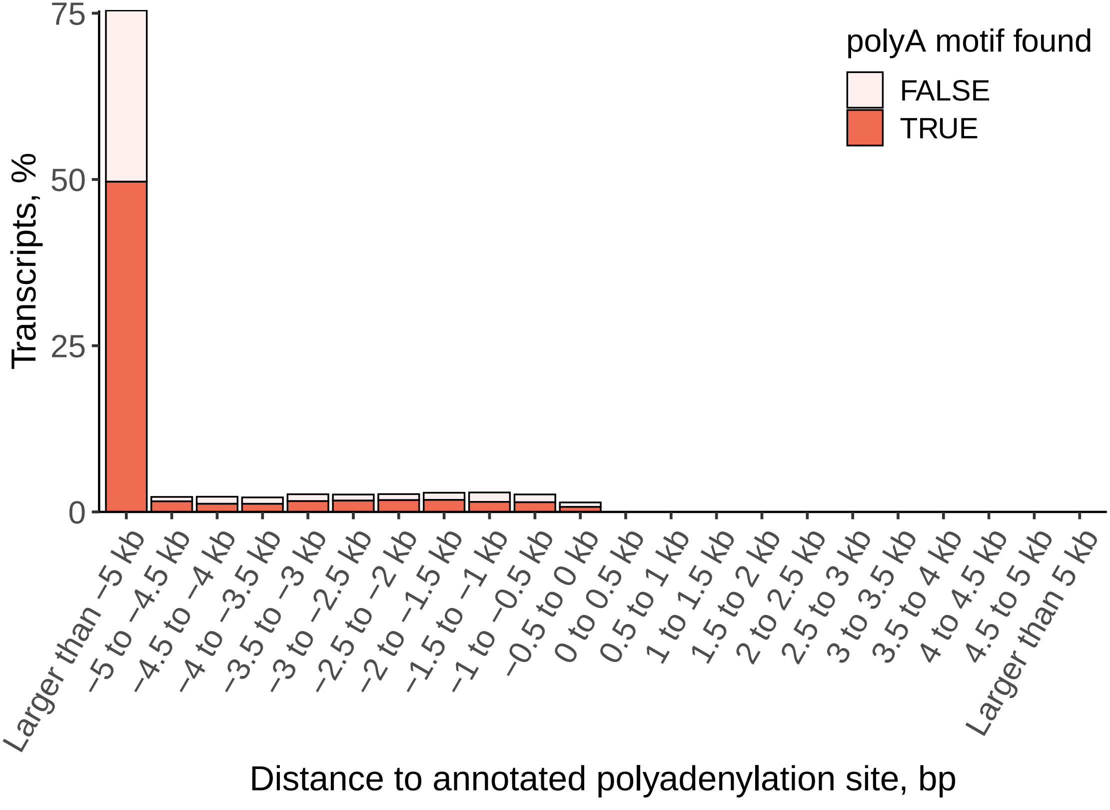
Negative values indicate upstream of annotated polyA site



Distance to Annotated Polyadenylation Site for ISM 3' Fragment

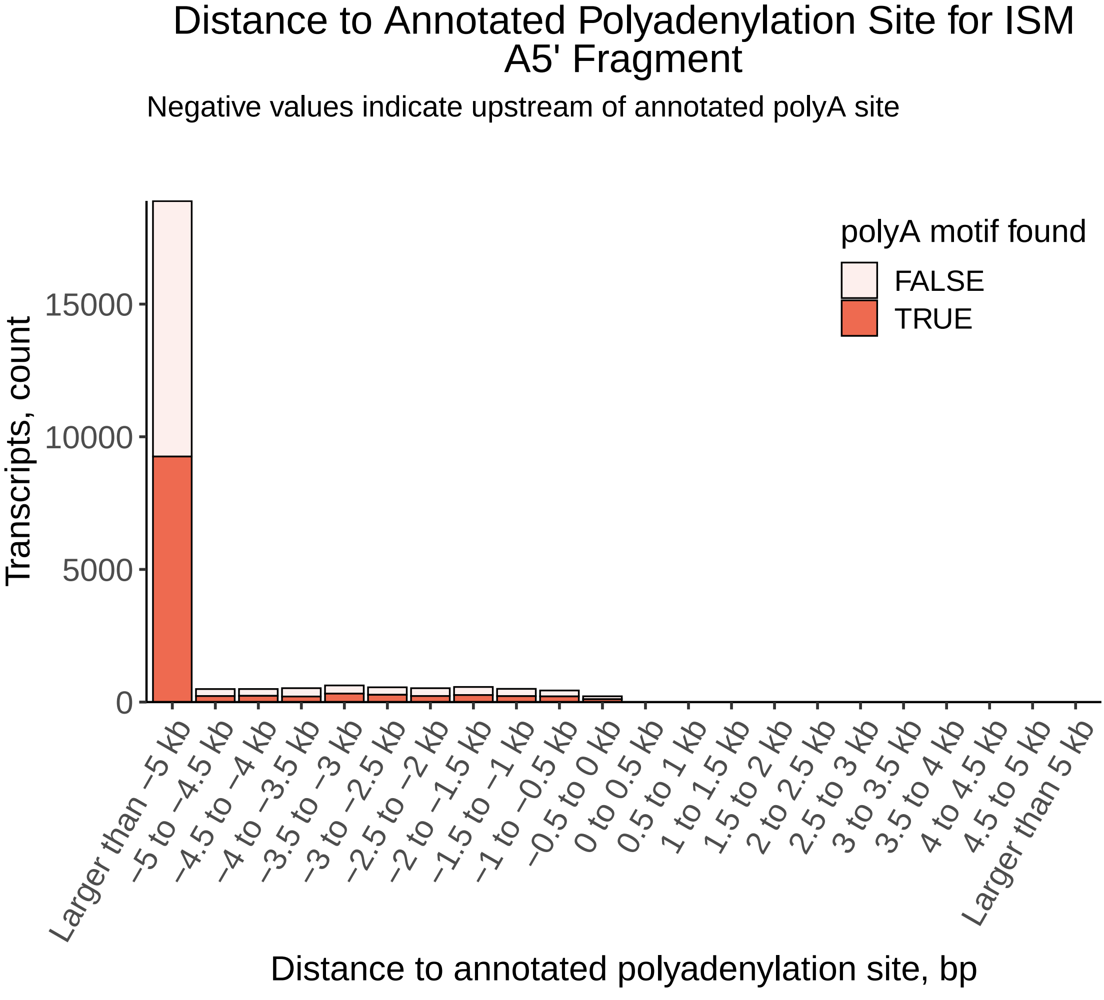
Negative values indicate upstream of annotated polyA site

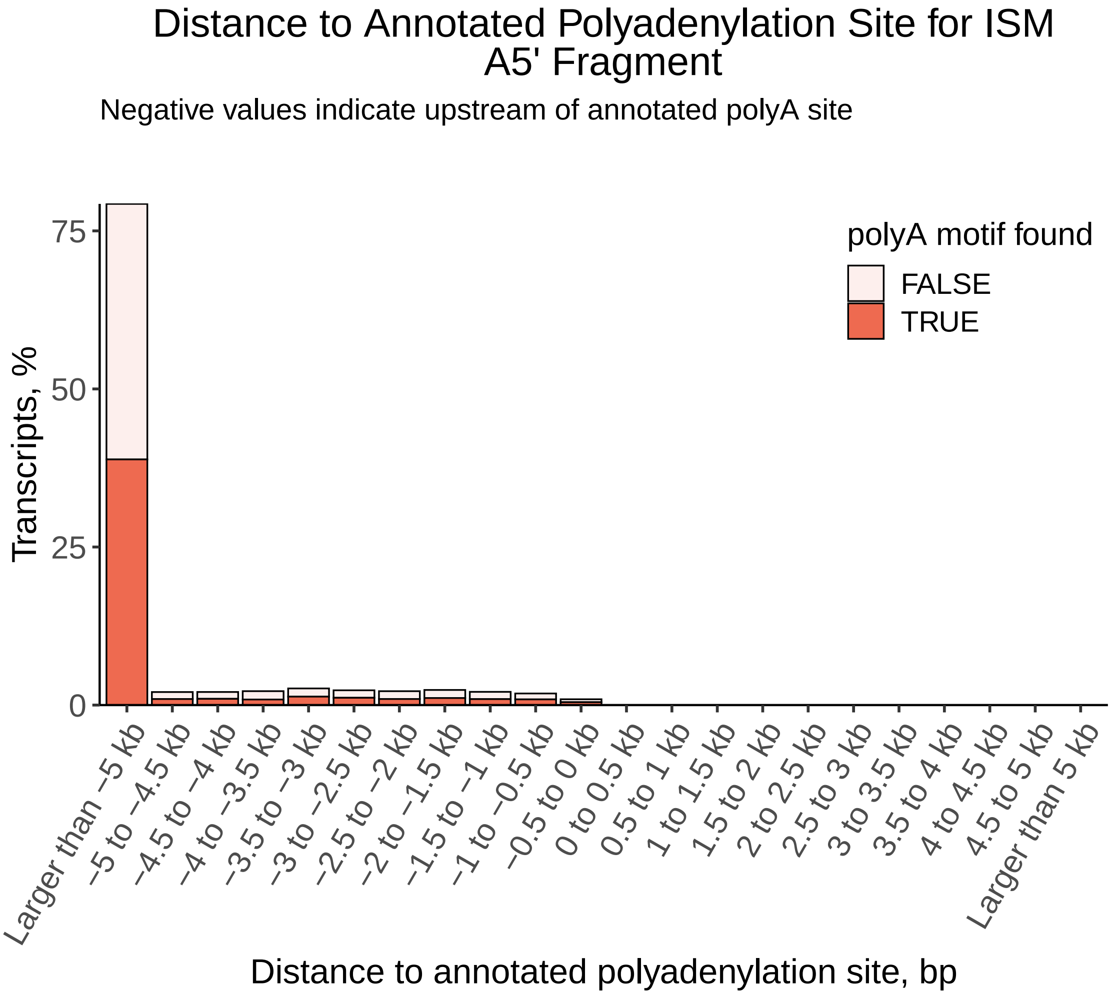

Distance to Annotated Polyadenylation Site for ISM Internal Fragment
Negative values indicate upstream of annotated polyA site
Transcripts, count
0
5000
10000
polyA motif found
FALSE
TRUE
Larger than −5 kb
−5 to −4.5 kb
−4.5 to −4 kb
−4 to −3.5 kb
−3.5 to −3 kb
−3 to −2.5 kb
−2.5 to −2 kb
−2 to −1.5 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 Internal Fragment

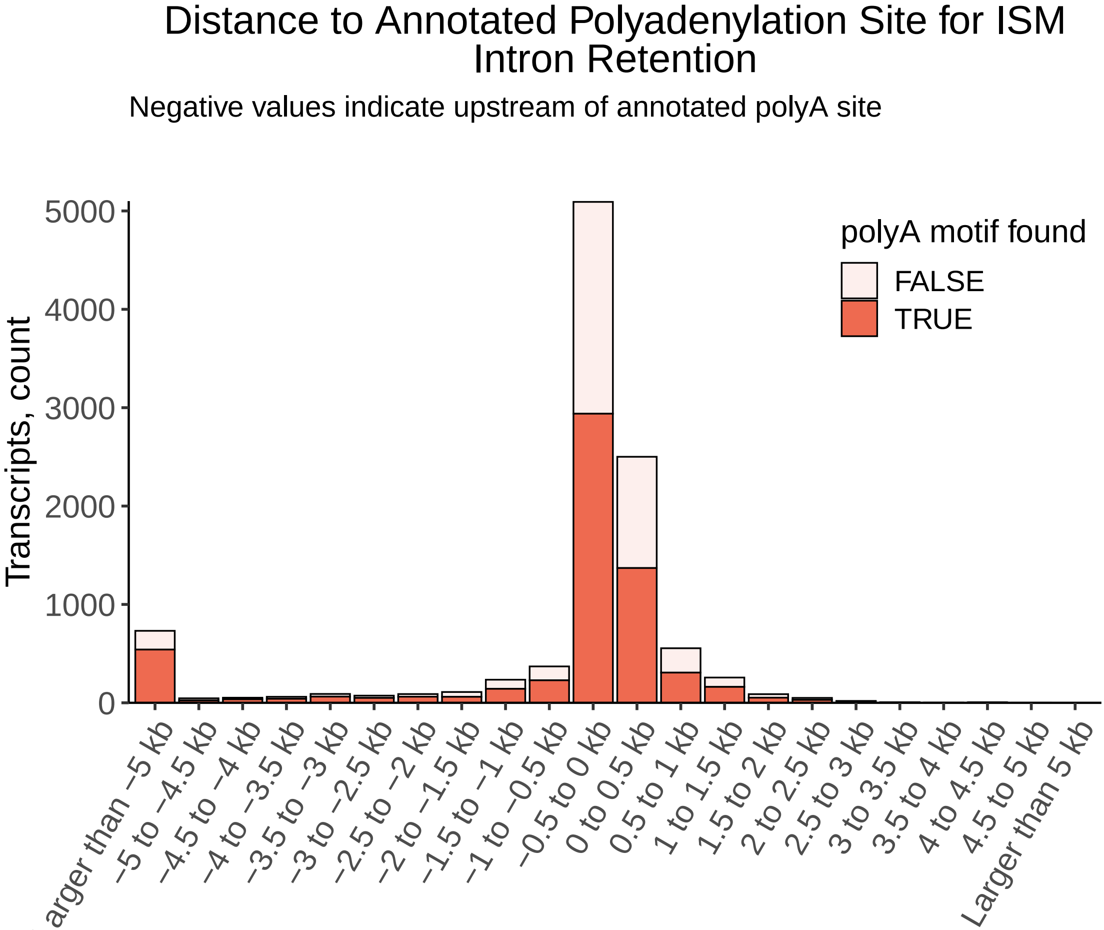
Negative values indicate upstream of annotated polyA site


Distance to Annotated Polyadenylation Site for ISM A5' Fragment
Negative values indicate upstream of annotated polyA site
polyA motif found
FALSE
TRUE
Transcripts, count
0
5000
10000
15000
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



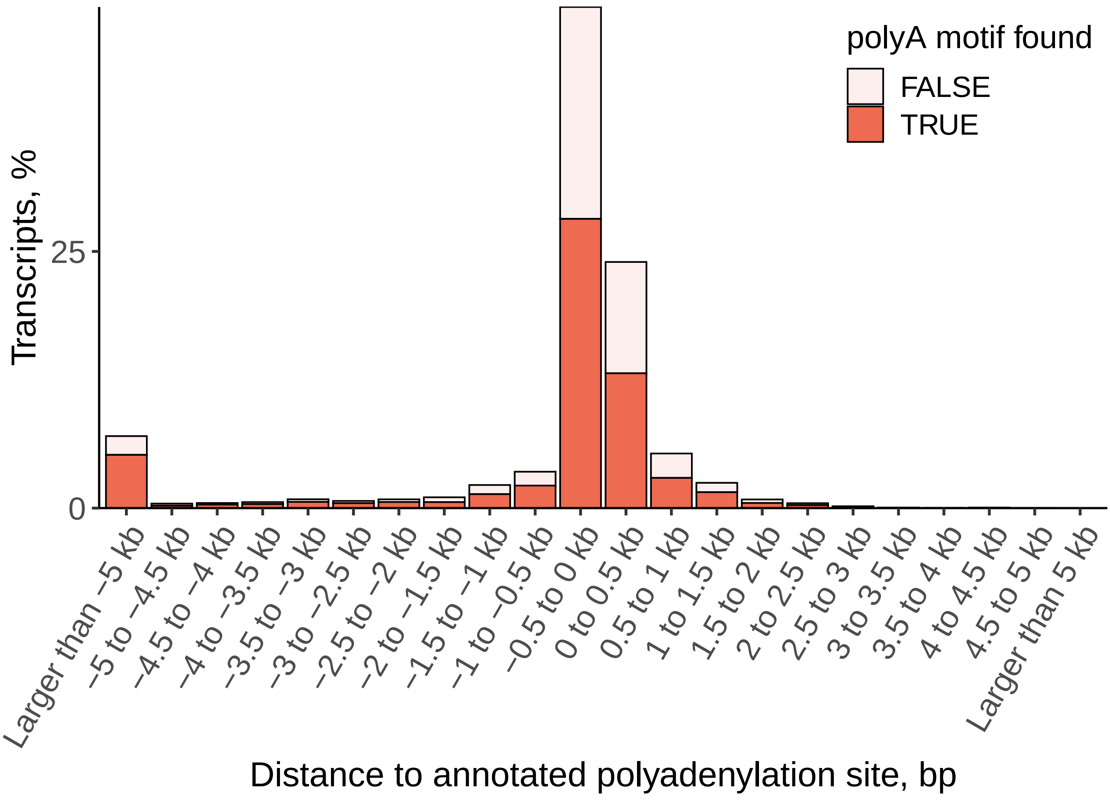
Distance to Annotated Polyadenylation Site for ISM A5' Fragment

Negative values indicate upstream of annotated polyA site

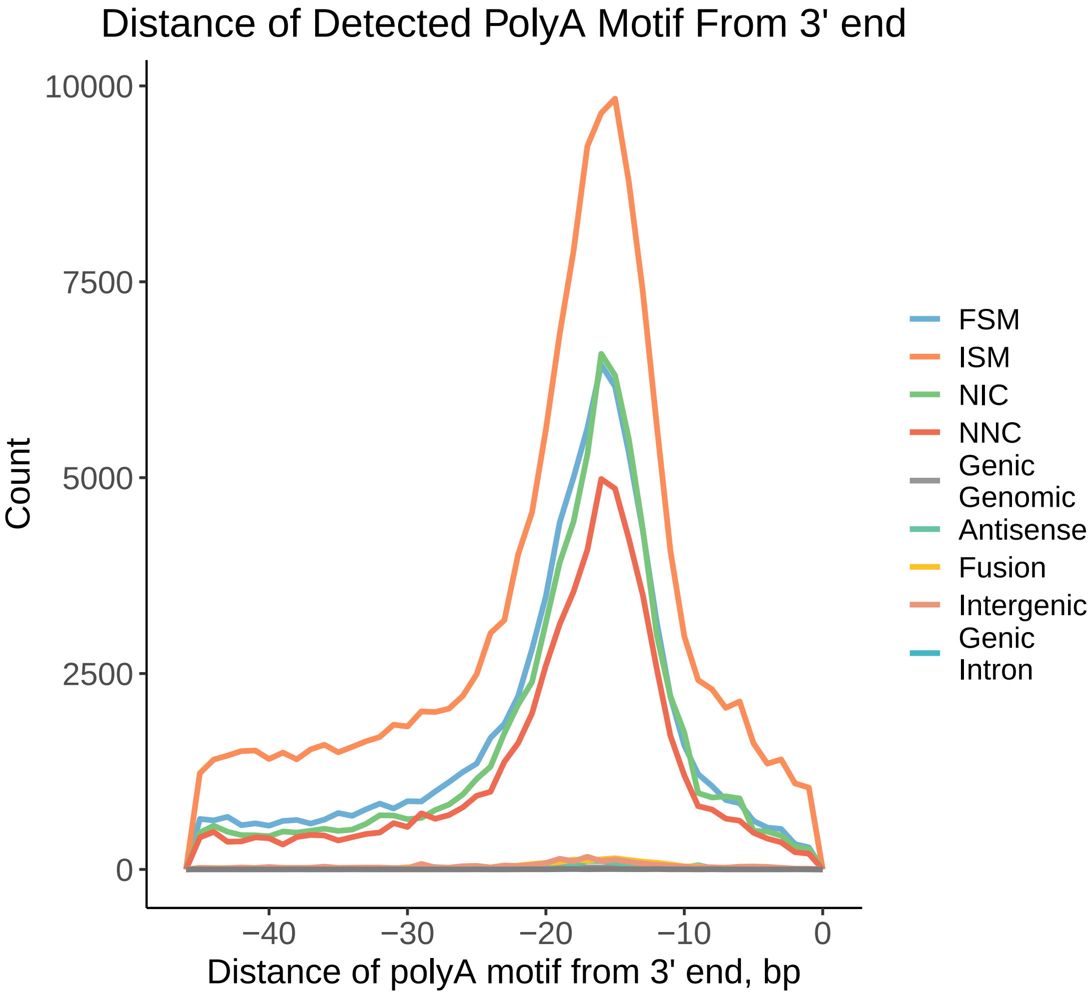

Distance to Annotated Polyadenylation Site for ISM Intron Retention
Negative values indicate upstream of annotated polyA site
polyA motif found
FALSE
TRUE
Transcripts, count
0
1000
2000
3000
4000
5000
Larger than −5 kb
−5 to −4.5 kb
−4.5 to −4 kb
−4 to −3.5 kb
−3.5 to −3 kb
−3 to −2.5 kb
−2.5 to −2 kb
−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



PolyA Distance Analysis

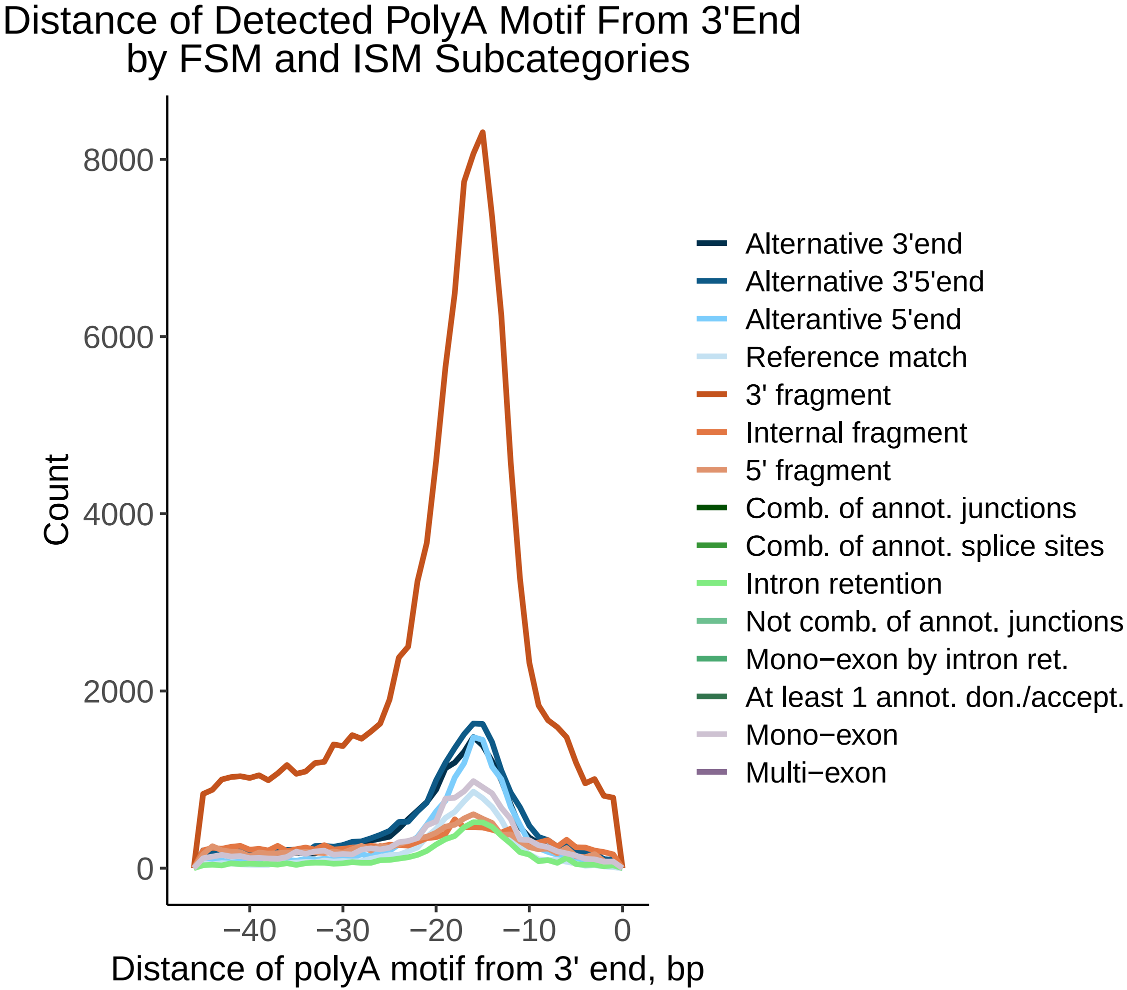

Distance of Detected PolyA Motif From 3' end
Count
10000
7500
5000
2500
0
−40
−30
−20
−10
0
Distance of polyA motif from 3' end, bp
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

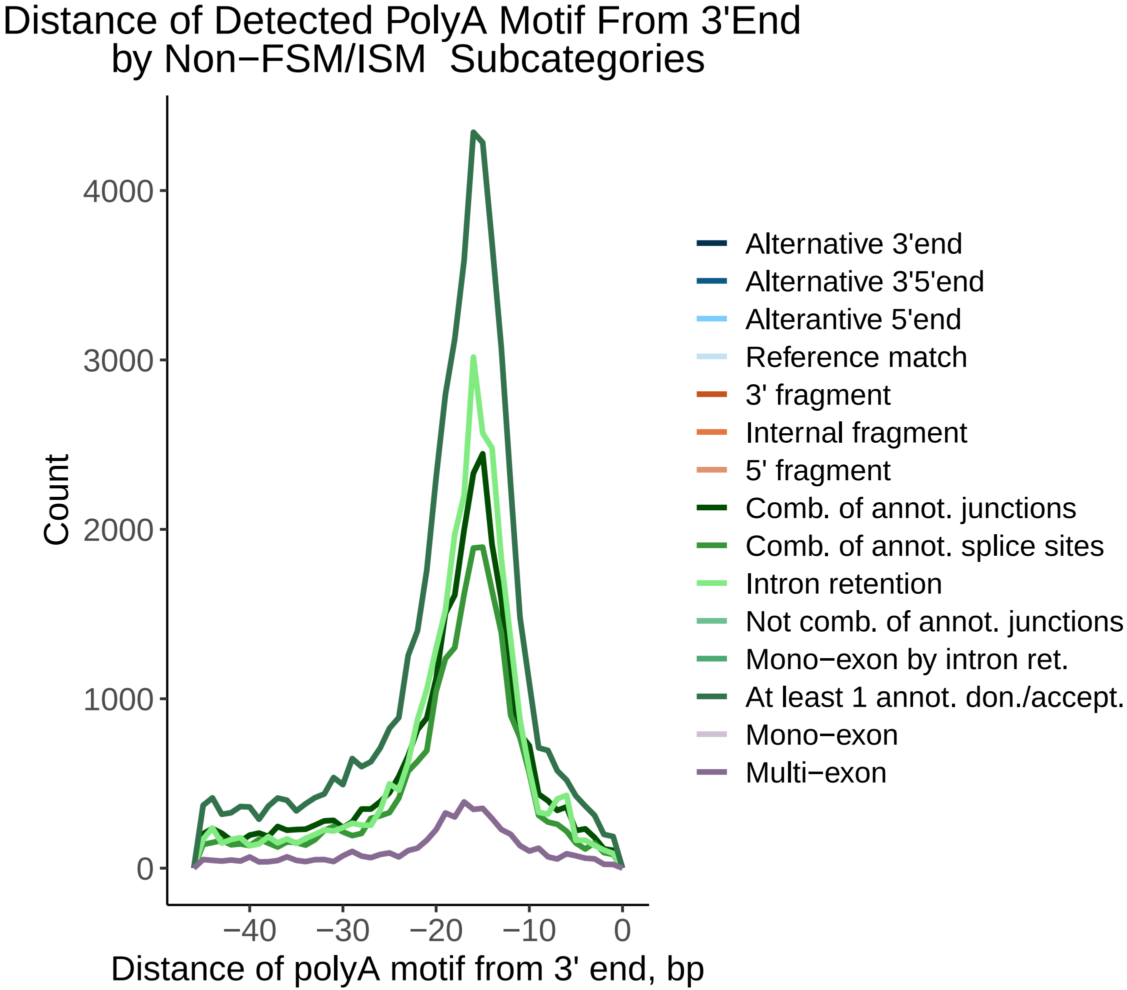
Number of polyA Motifs Detected

Category	Count	polyA Detected	%
FSM	149437	78949	53
ISM	236718	143526	61
NIC	138600	72445	52
NNC	105766	57474	54
Genic Genomic	752	412	55
Antisense	2146	1215	57
Fusion	3044	1764	58
Intergenic	2989	1978	66
NA	69	40	58

Frequency of PolyA Motifs

Motif	Count	%
AATAAA	175837	49.1
ATTAAA	51781	14.5
TATAAA	14342	4.0
AGTAAA	13575	3.8
AAAAAG	13444	3.8
AAGAAA	13382	3.7
GGGGCT	10508	2.9
TTTAAA	9928	2.8
AATATA	8374	2.3
CATAAA	8172	2.3
AATACA	7760	2.2
GATAAA	7755	2.2
AAAACA	7585	2.1
AATGAA	6999	2.0
AATAGA	4314	1.2
ACTAAA	4047	1.1


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
4000
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

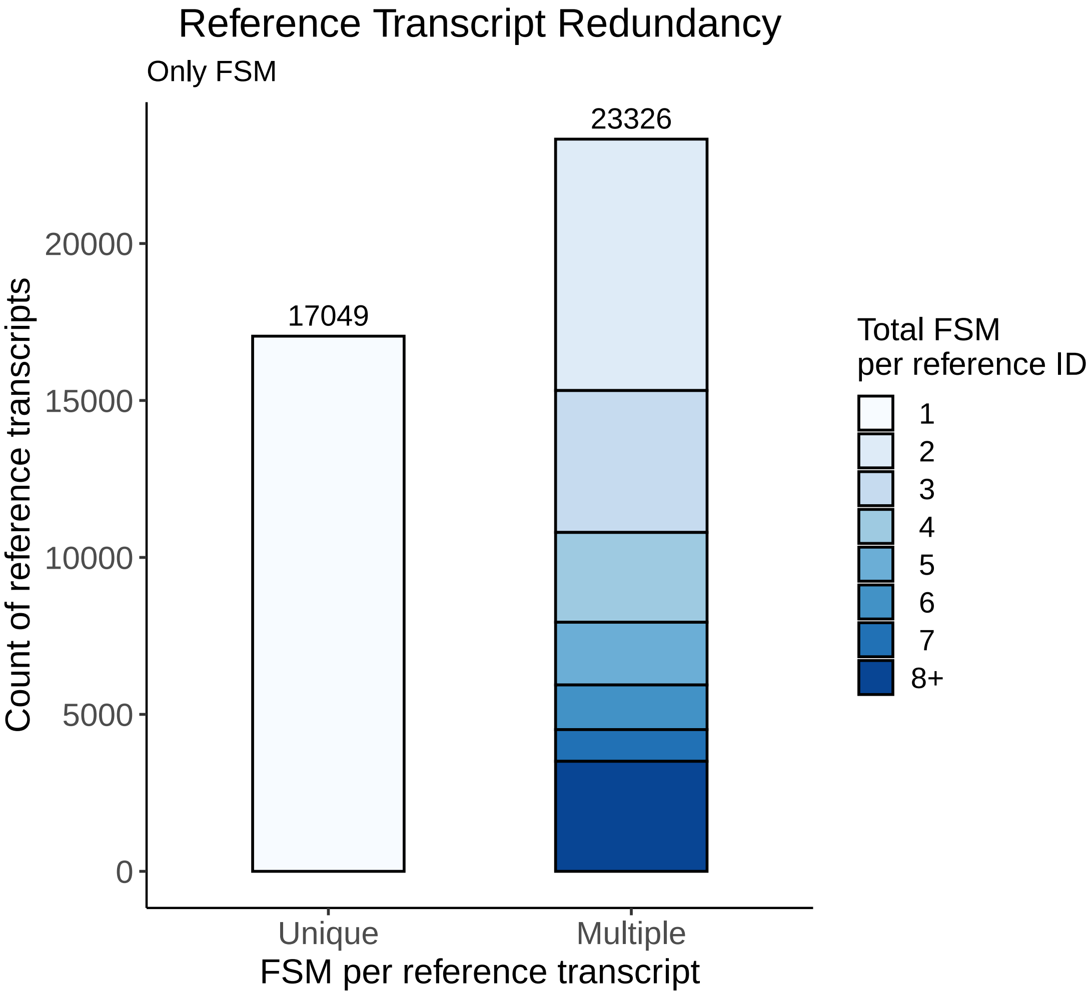
Number of polyA Motifs Detected

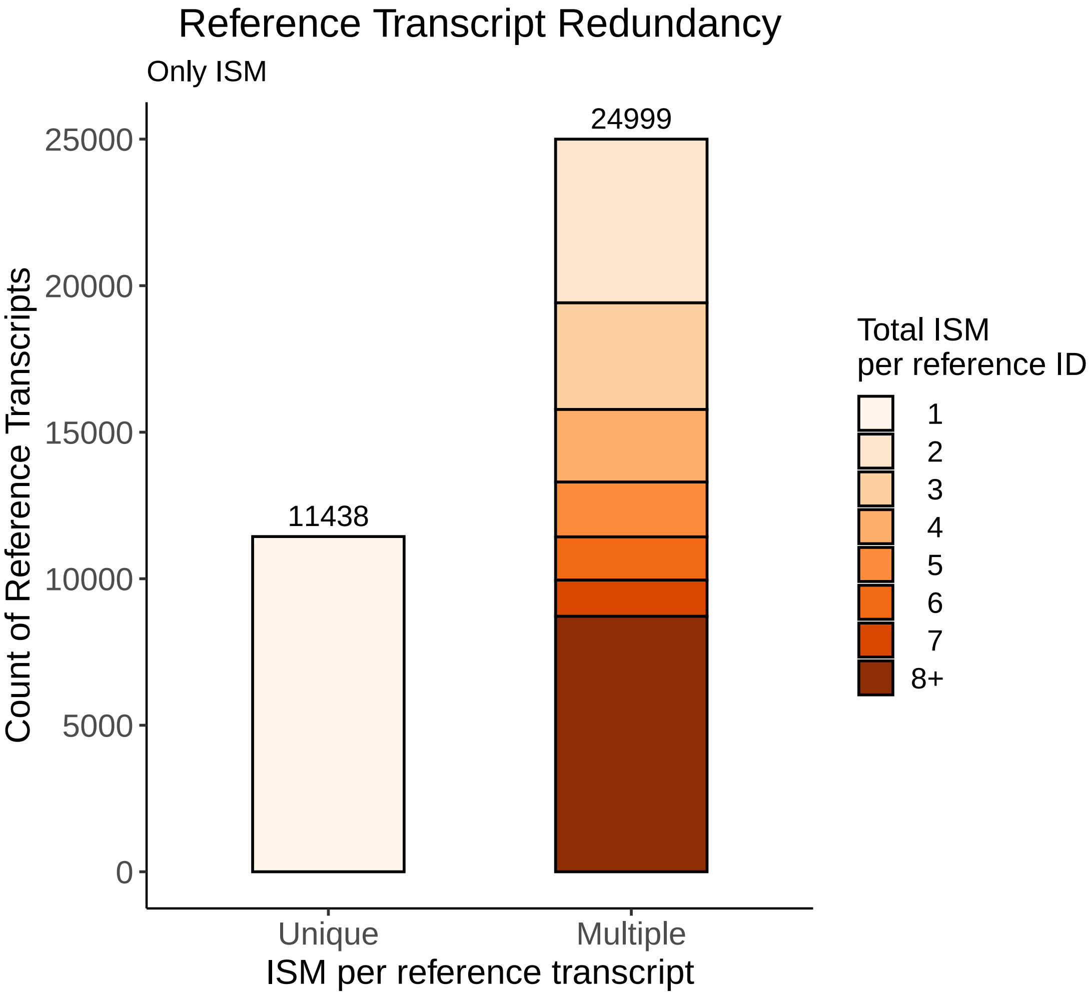
Subcategory	**Count**	**polyA Detected**	**%**
Alternative 3'end	40493	19324	48
Alternative 3'5'end	41864	22206	53
Alterantive 5'end	27458	14801	54
Reference match	17219	8945	52
3' fragment	182812	113187	62
Internal fragment	19651	12643	64
5' fragment	23822	11563	49
Comb. of annot. junctions	53196	27740	52
Comb. of annot. splice sites	41582	22131	53
Intron retention	67476	35502	53
At least 1 annot. don./accept.	93105	50965	55
Mono−exon	22403	13673	61
Multi−exon	8382	5093	61
NA	58	30	52

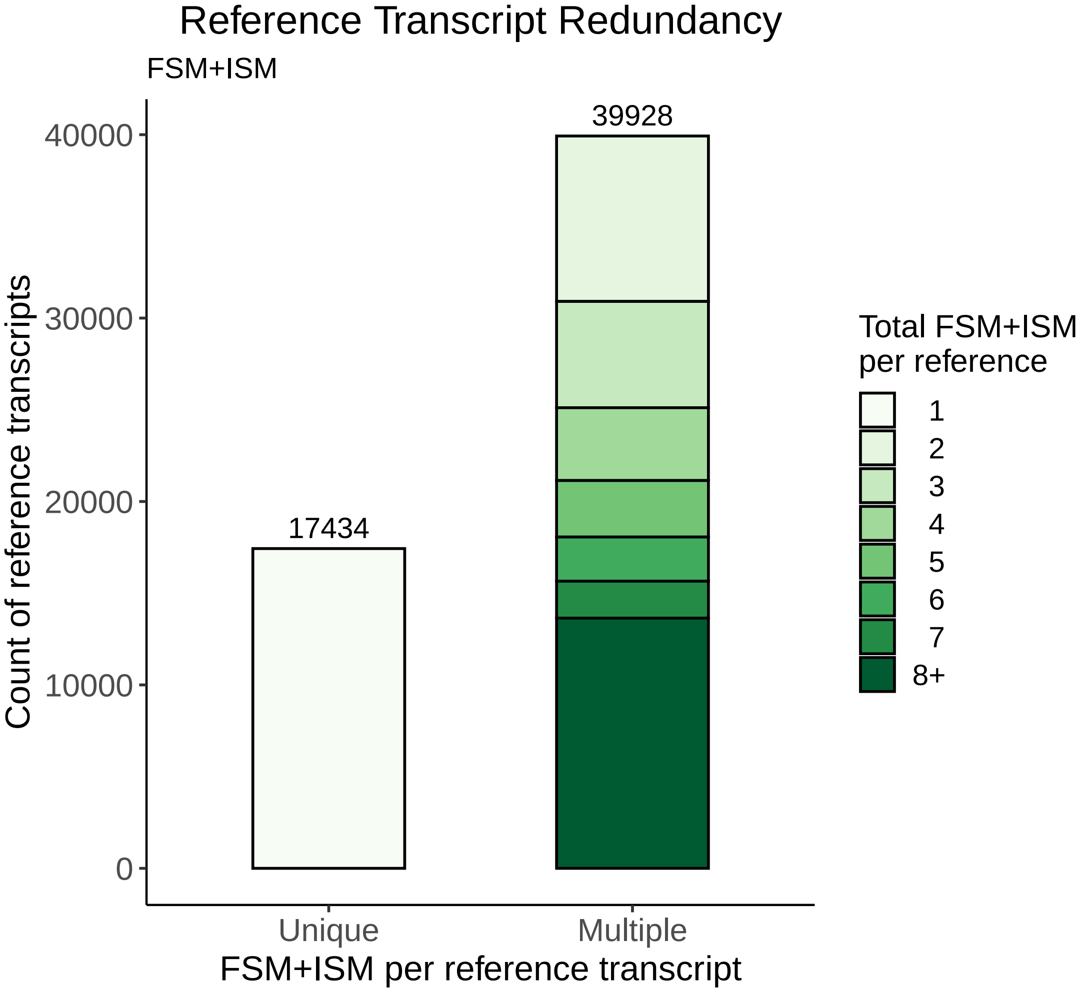
Frequency of PolyA Motifs

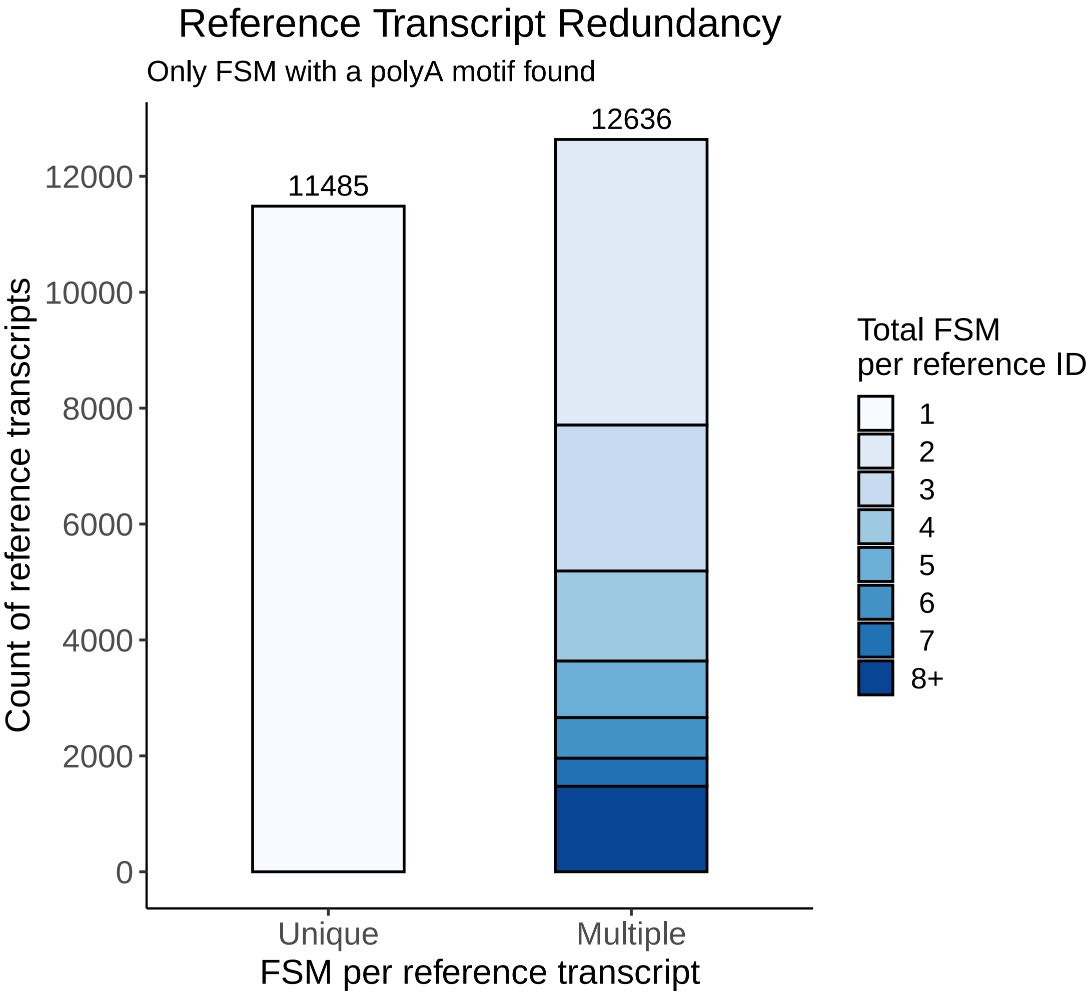
Motif	Count	%
AATAAA	175837	49.1
ATTAAA	51781	14.5
TATAAA	14342	4.0
AGTAAA	13575	3.8
AAAAAG	13444	3.8
AAGAAA	13382	3.7
GGGGCT	10508	2.9
TTTAAA	9928	2.8
AATATA	8374	2.3
CATAAA	8172	2.3
AATACA	7760	2.2
GATAAA	7755	2.2
AAAACA	7585	2.1
AATGAA	6999	2.0
AATAGA	4314	1.2
ACTAAA	4047	1.1

Redundancy Analysis


Reference Transcript Redundancy
Only FSM
23326
17049
20000
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+

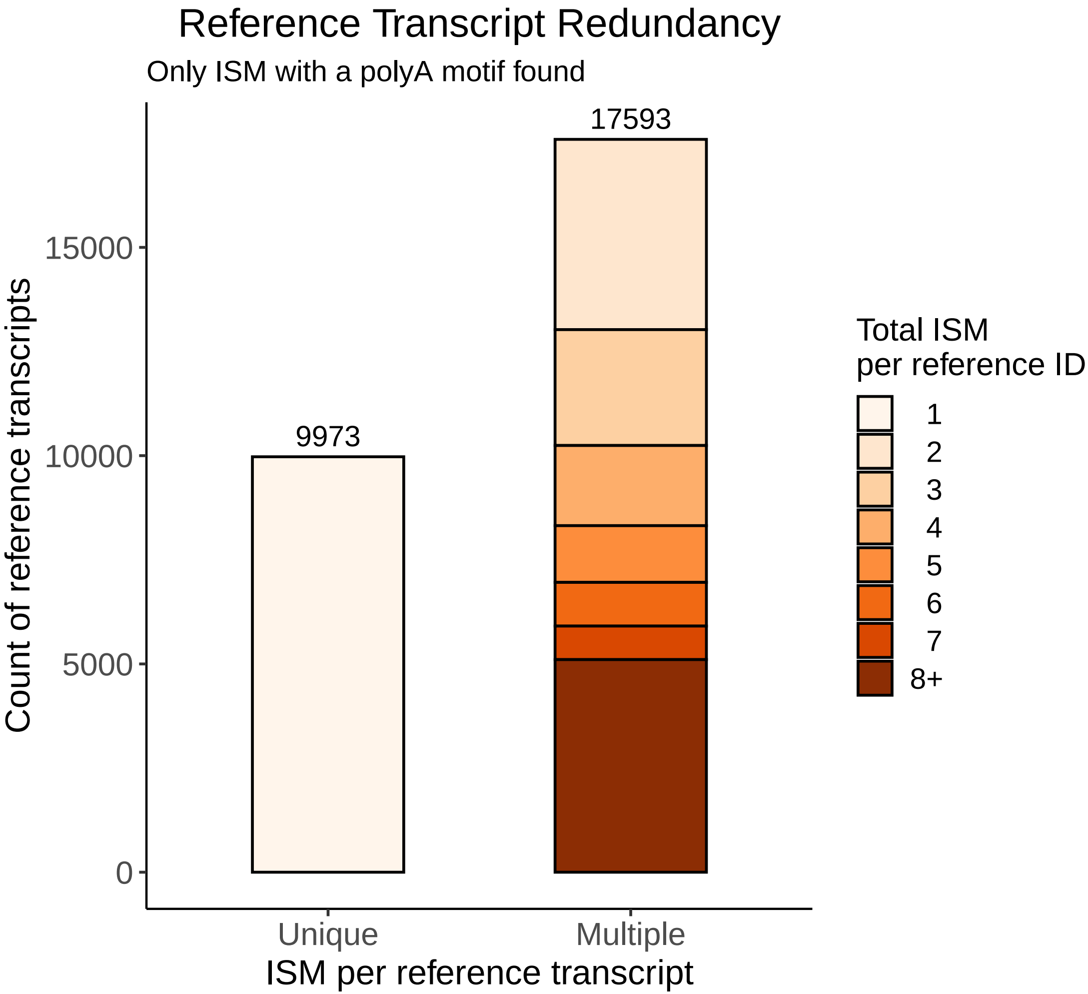

Reference Transcript Redundancy
Only ISM
24999
11438
25000
20000
15000
10000
5000
0
Count of Reference Transcripts
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
5
6
7
8+

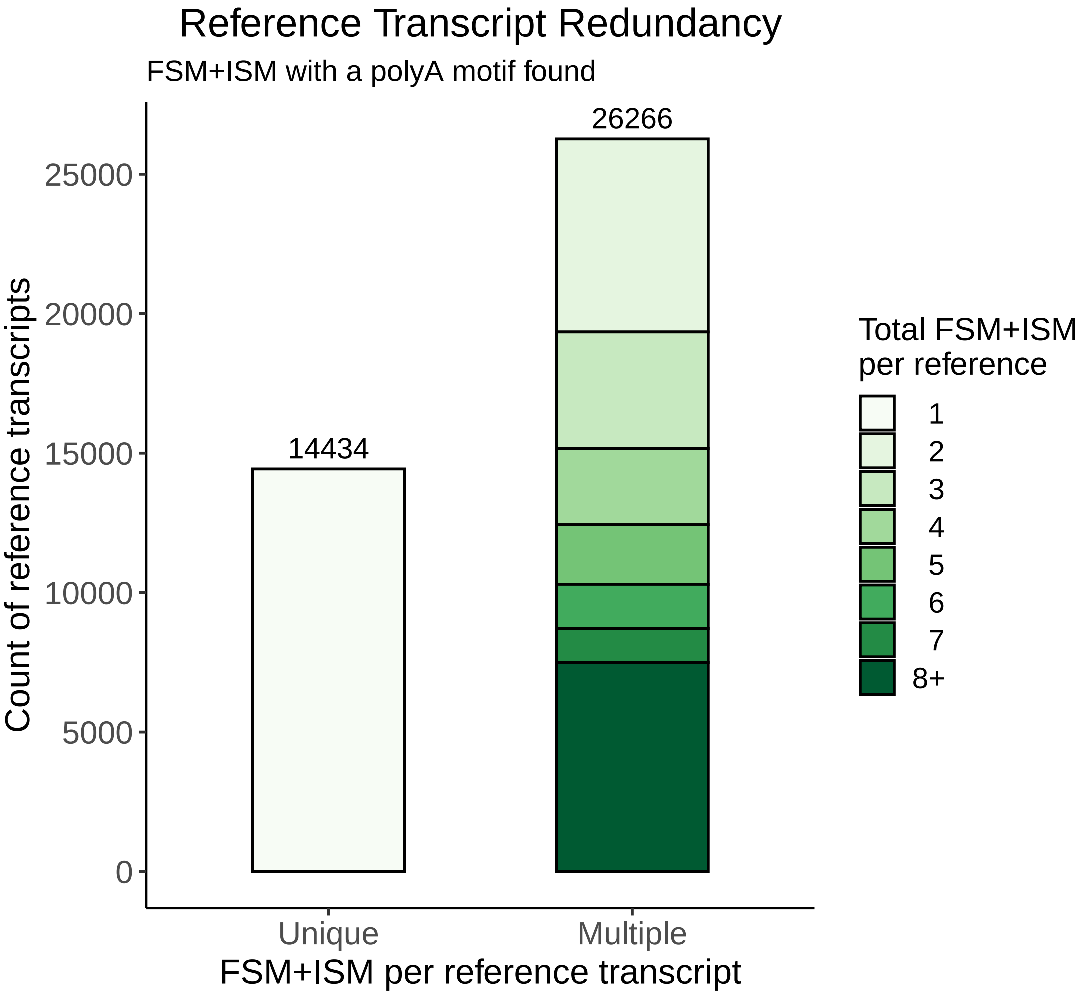

Reference Transcript Redundancy
FSM+ISM
Count of reference transcripts
40000
30000
20000
10000
0
17434
39928
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+



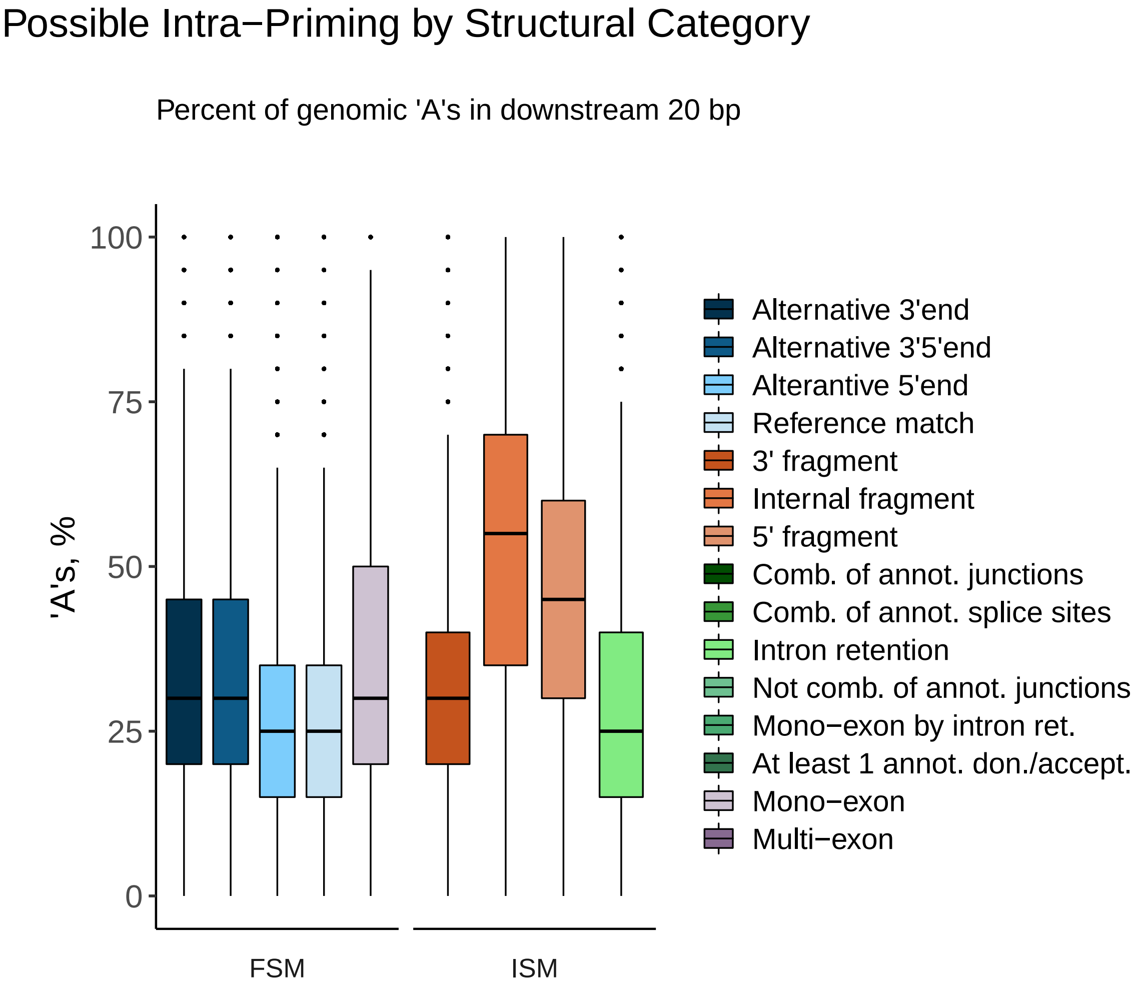
Reference Transcript Redundancy

Only FSM with a polyA motif found


Reference Transcript Redundancy
Only ISM with a polyA motif found
Count of reference transcripts
0
5000
10000
15000
9973
17593
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
5
6
7
8+

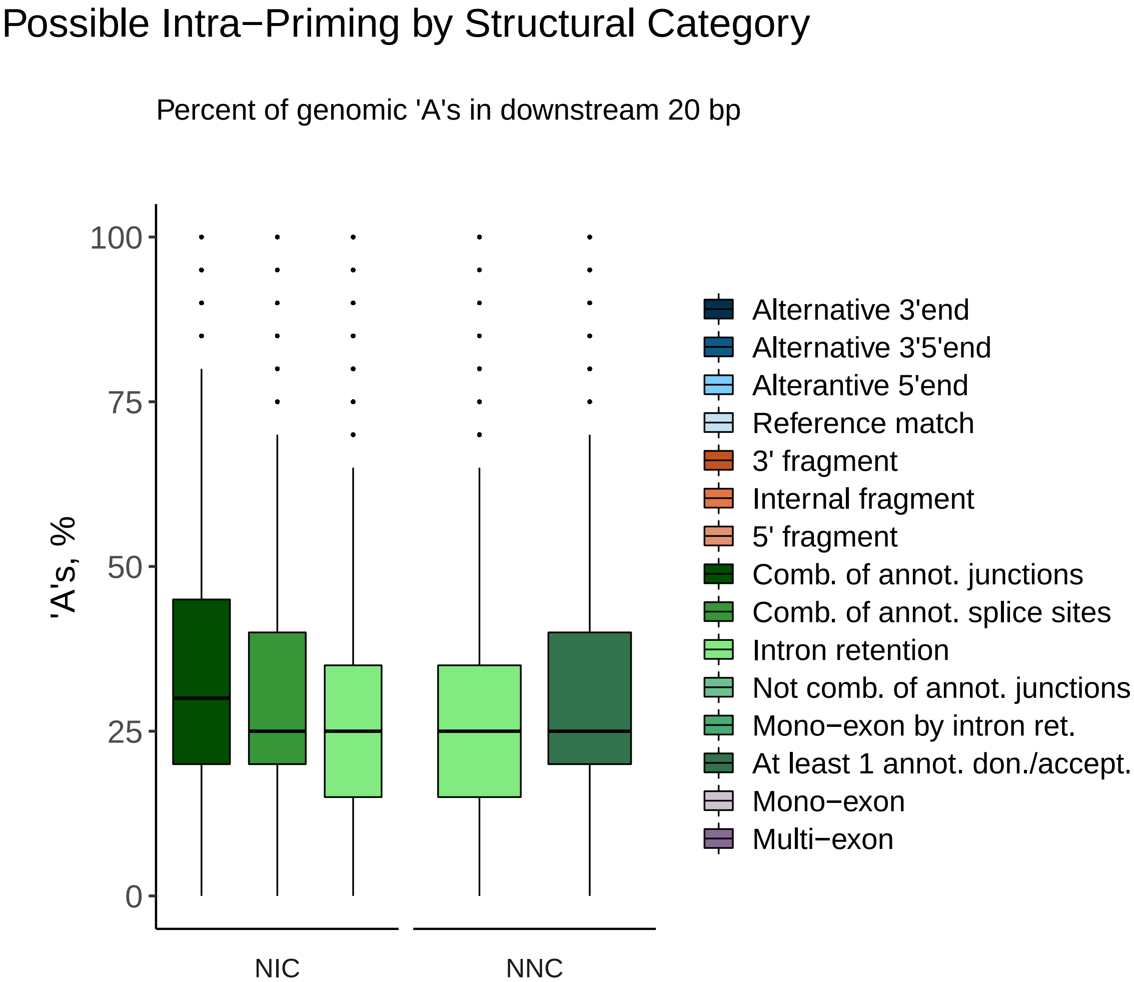

Reference Transcript Redundancy
FSM+ISM with a polyA motif found
26266
14434
25000
20000
15000
10000
5000
0
Count of reference transcripts
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+

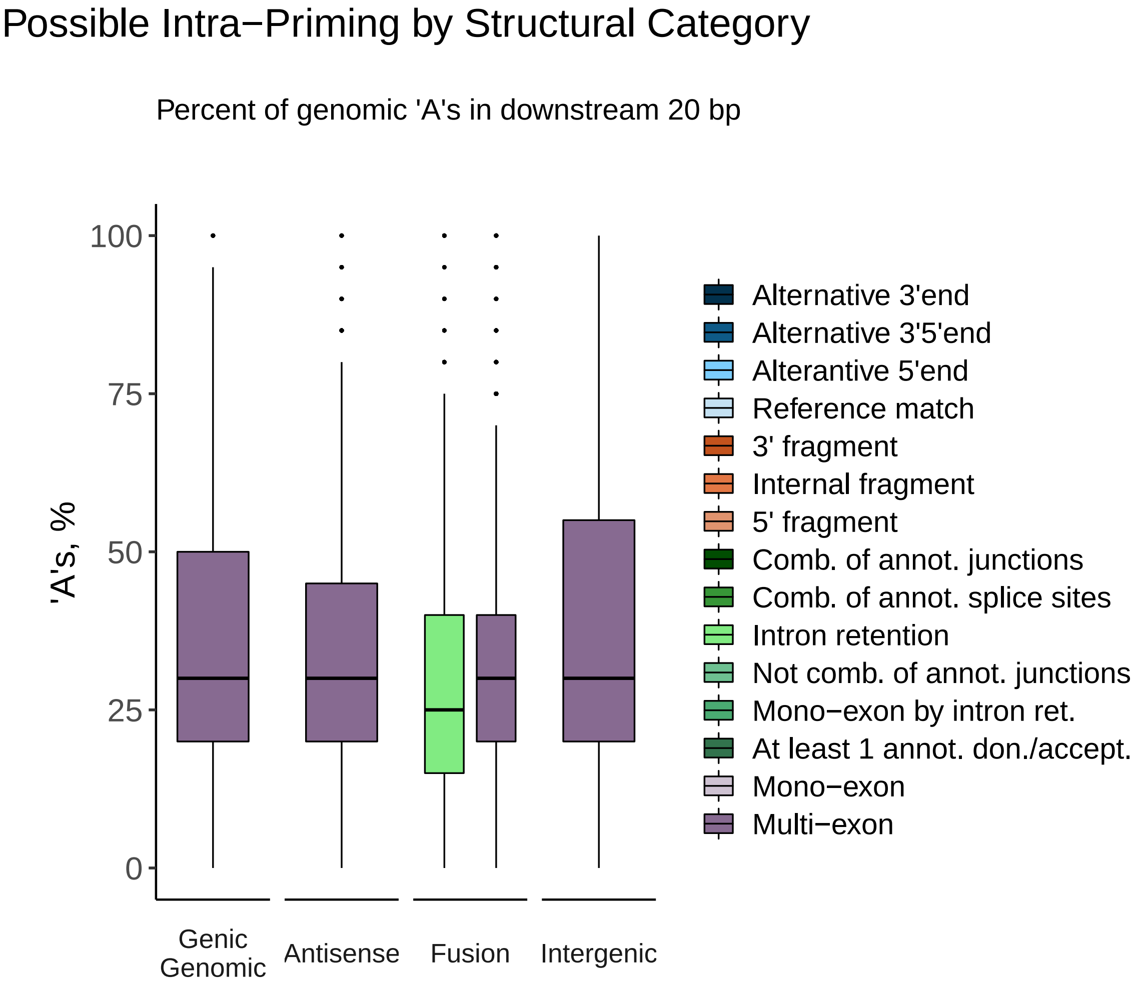
Intra−Priming Quality Check



Possible Intra−Priming by Structural Category

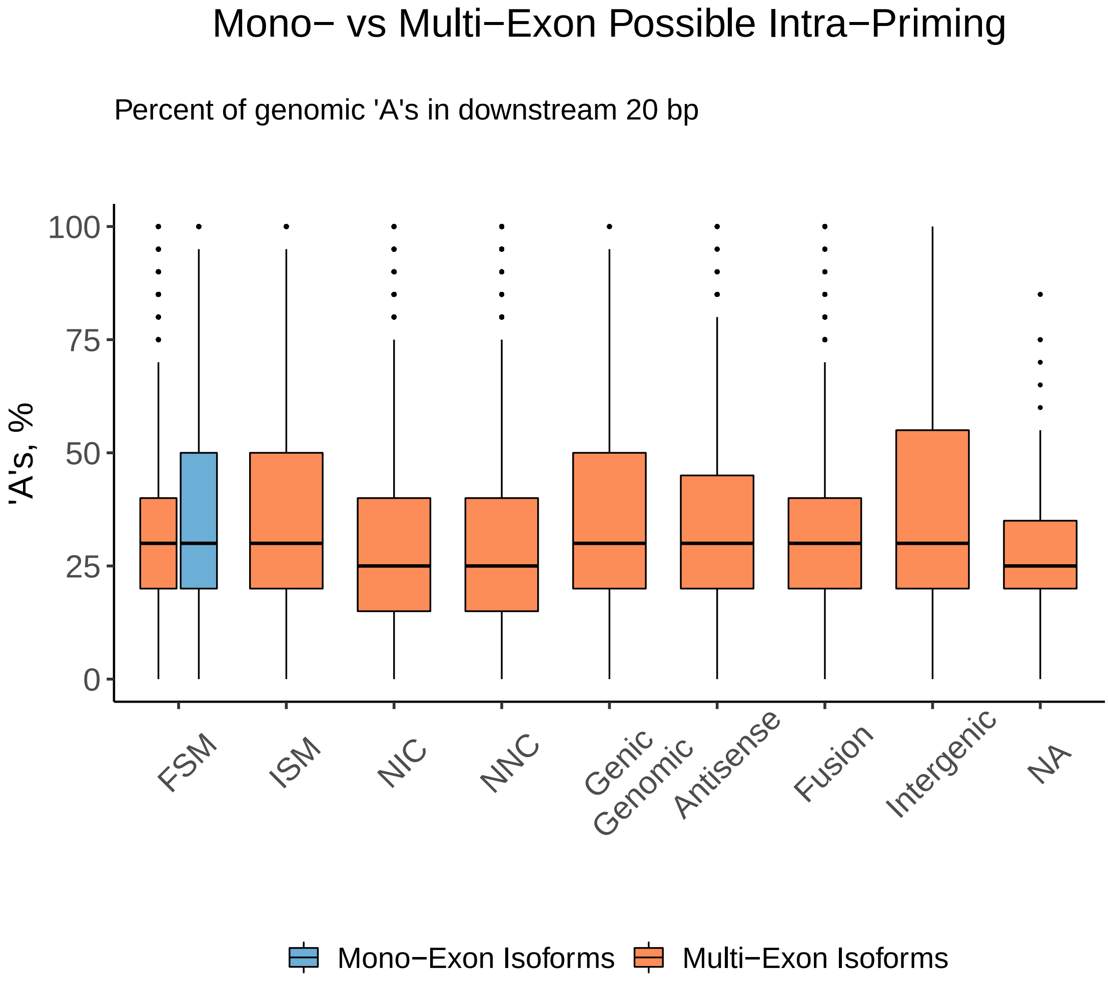
Percent of genomic 'A's in downstream 20 bp

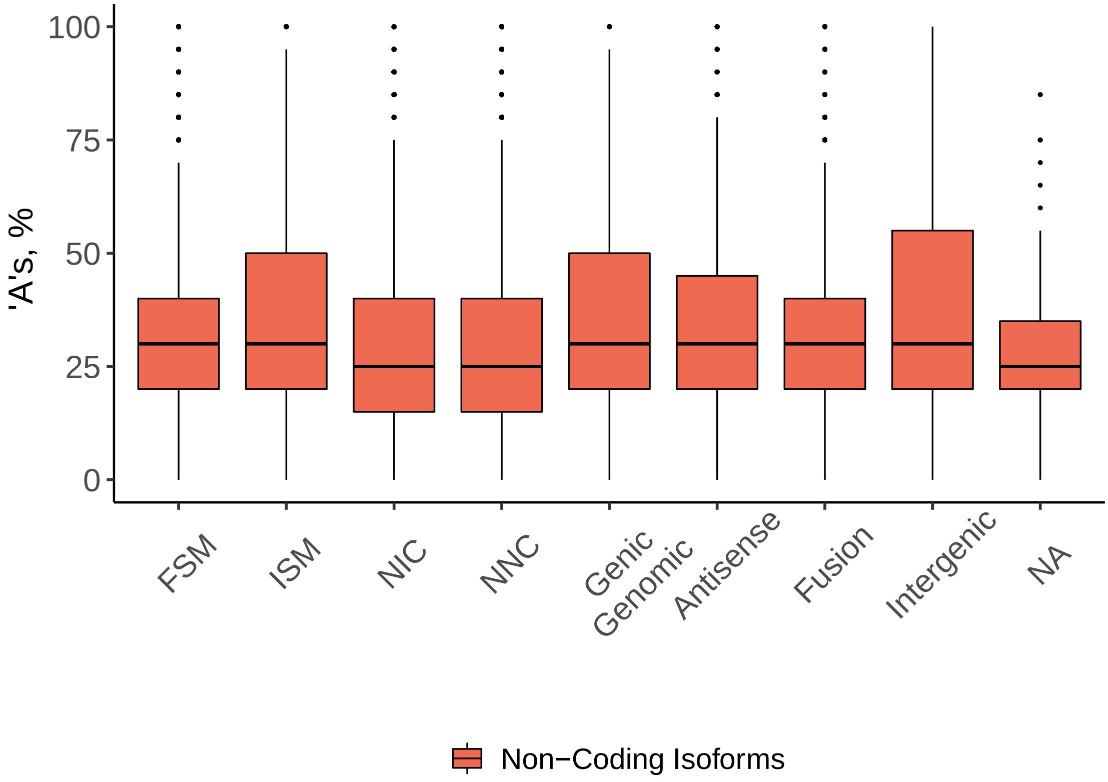

Possible Intra−Priming by Structural Category
Percent of genomic 'A's in downstream 20 bp
'A's, %
100
75
50
25
0
NIC
NNC
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
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

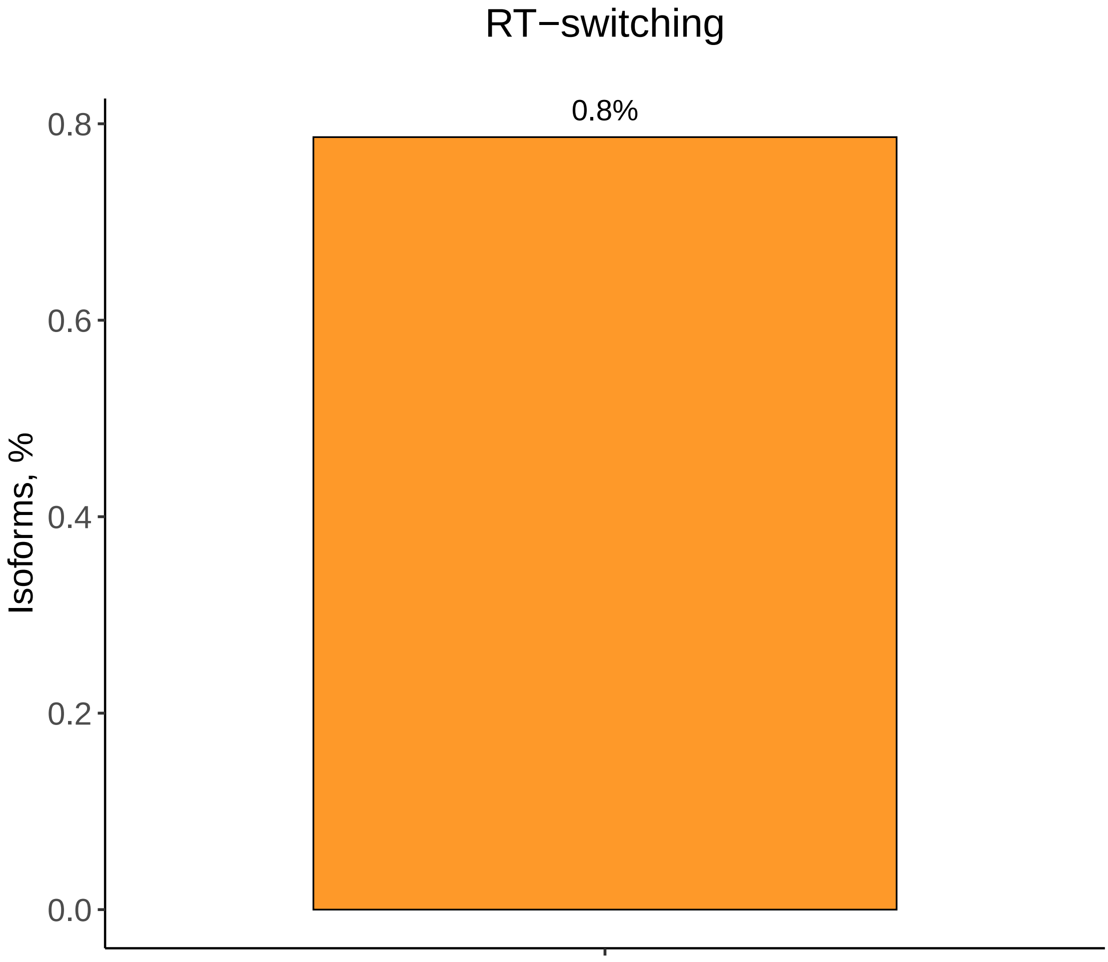

Mono− vs Multi−Exon Possible Intra−Priming
Percent of genomic 'A's in downstream 20 bp
'A's, %
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic
Genomic
Antisense
Fusion
Intergenic
NA
Mono−Exon Isoforms
Multi−Exon Isoforms

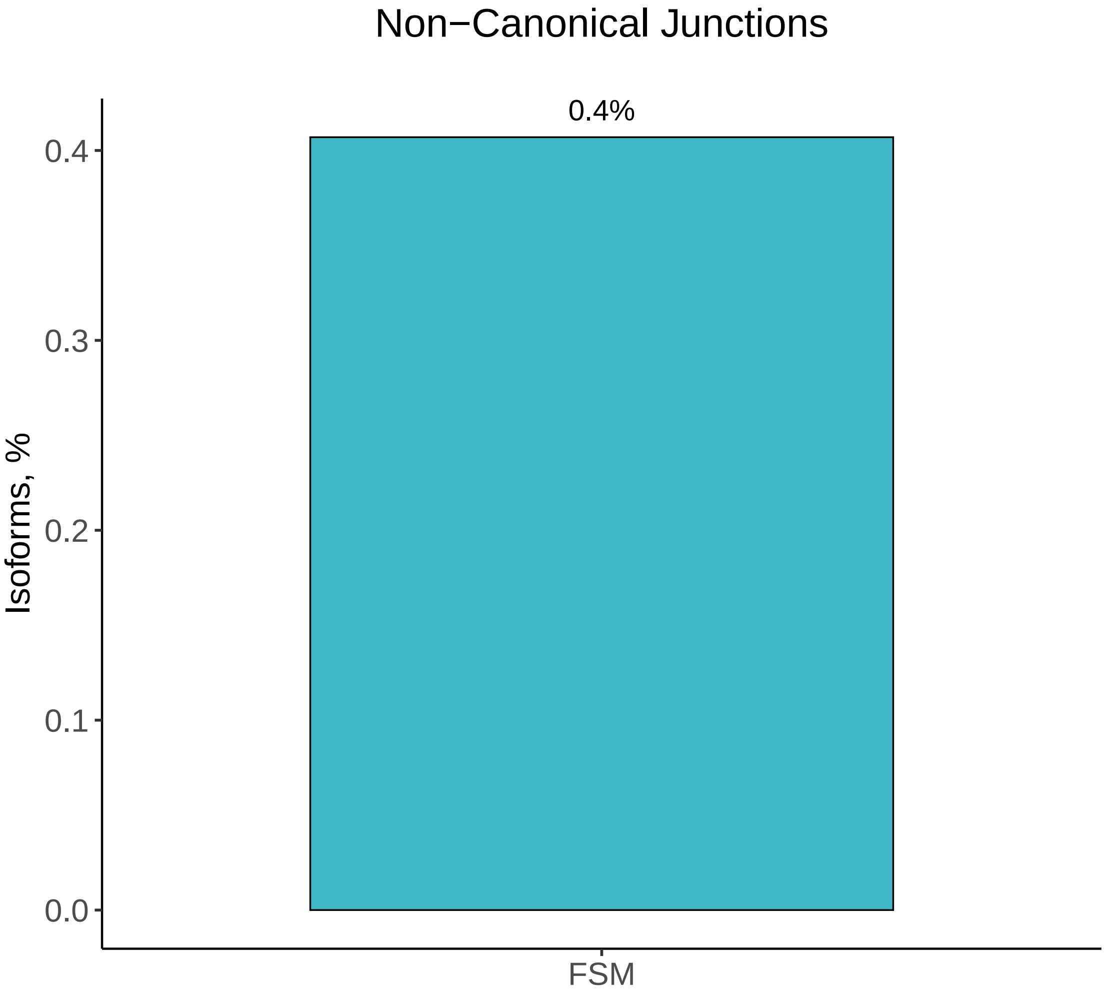


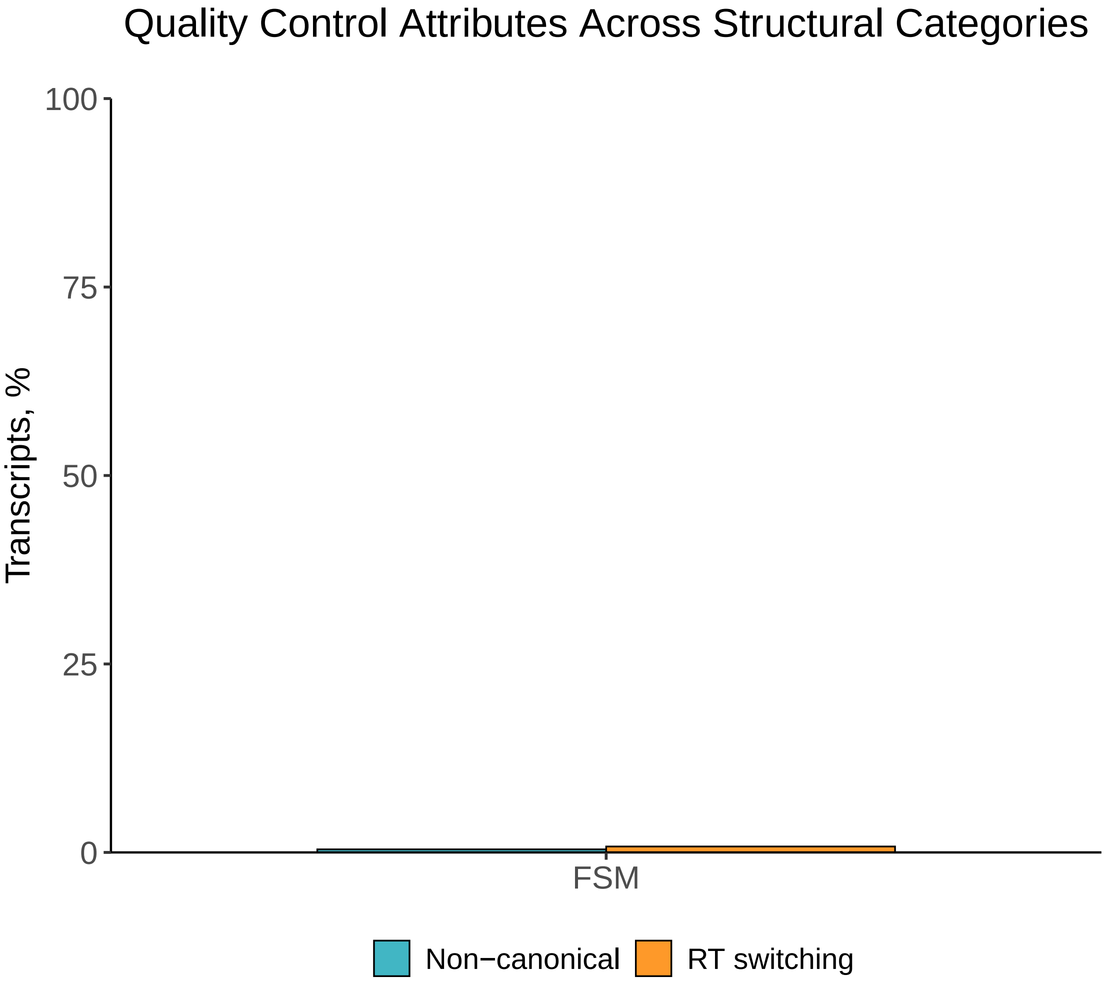
Coding vs Non−Coding Possible Intra−Priming

Percent of genomic 'A's in downstream 20 bp

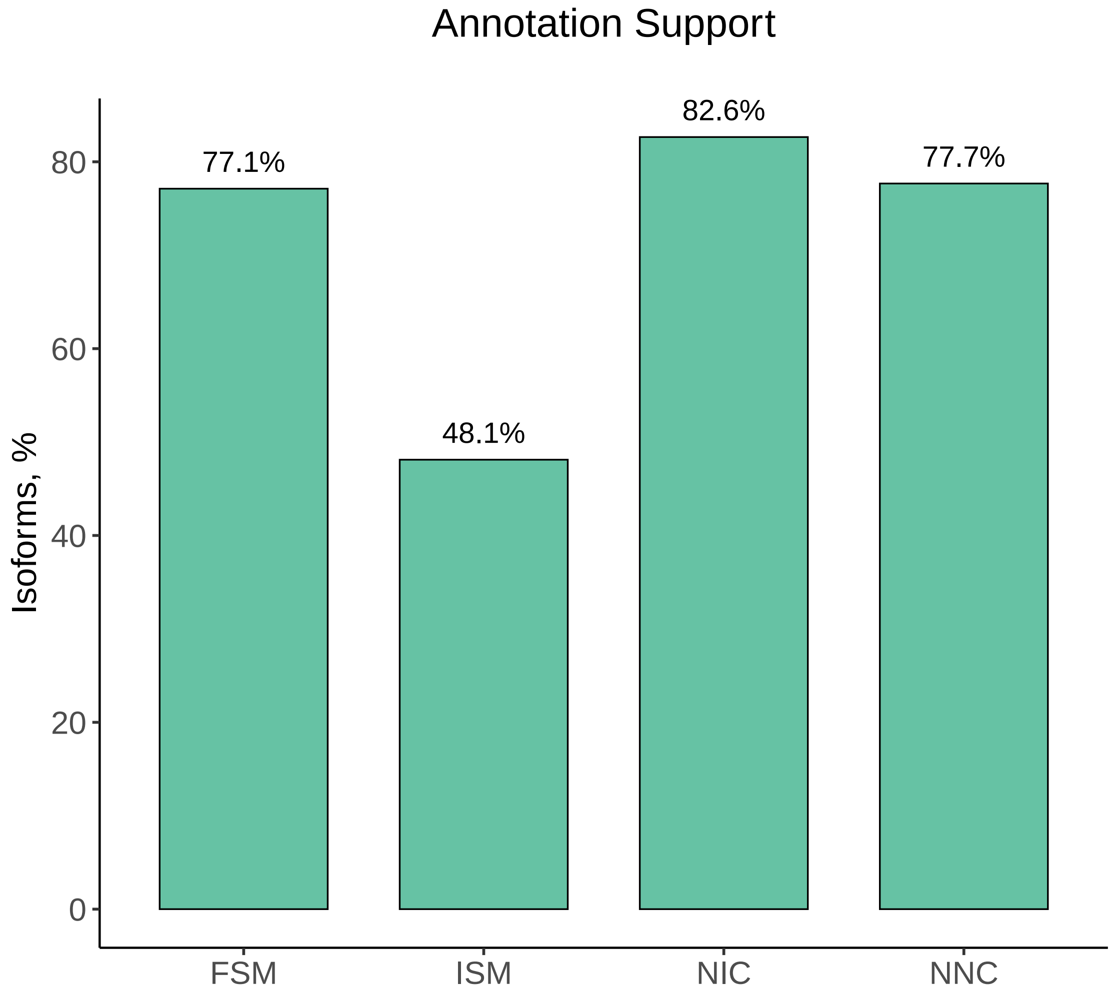
Features of Bad Quality

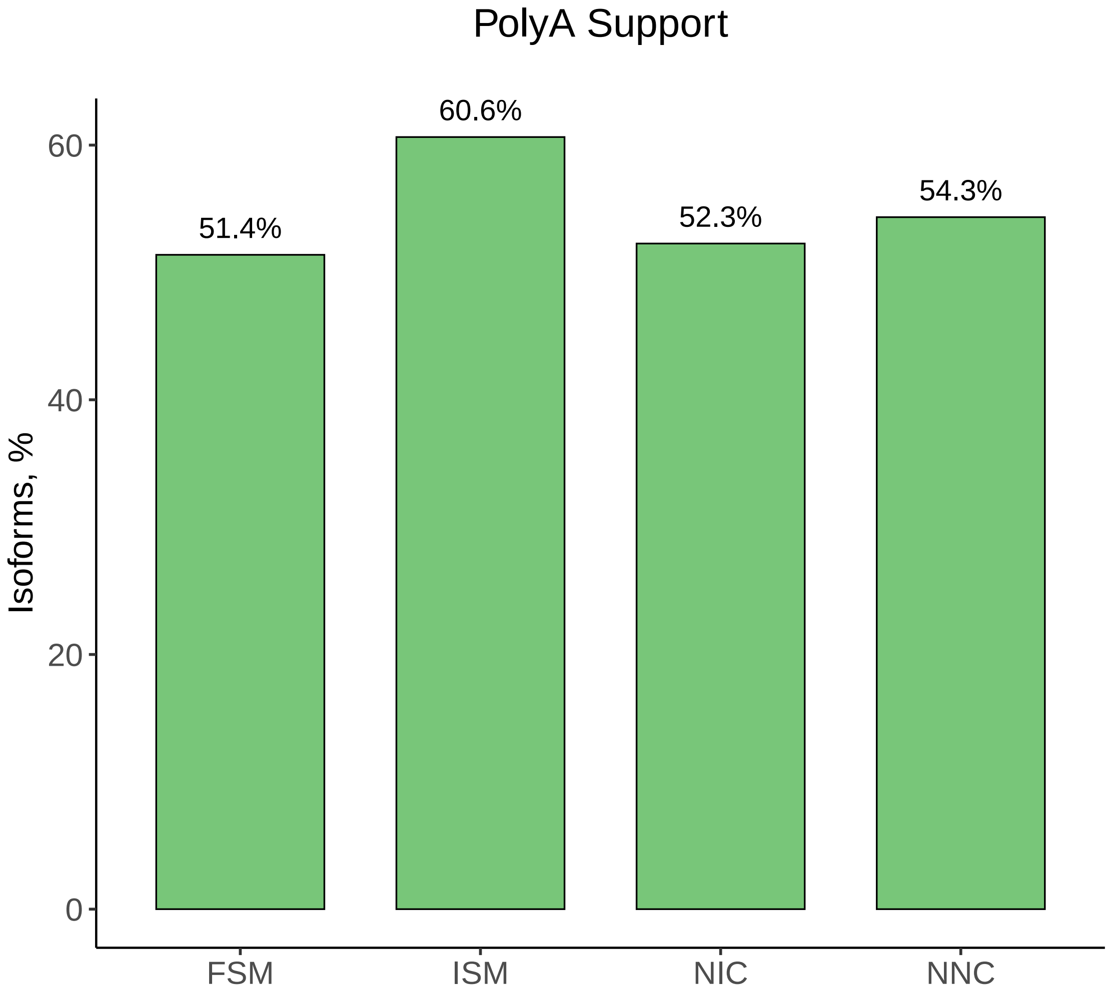

RT−switching
0.8%
Isoforms, %
0.8
0.6
0.4
0.2
0.0
FSM

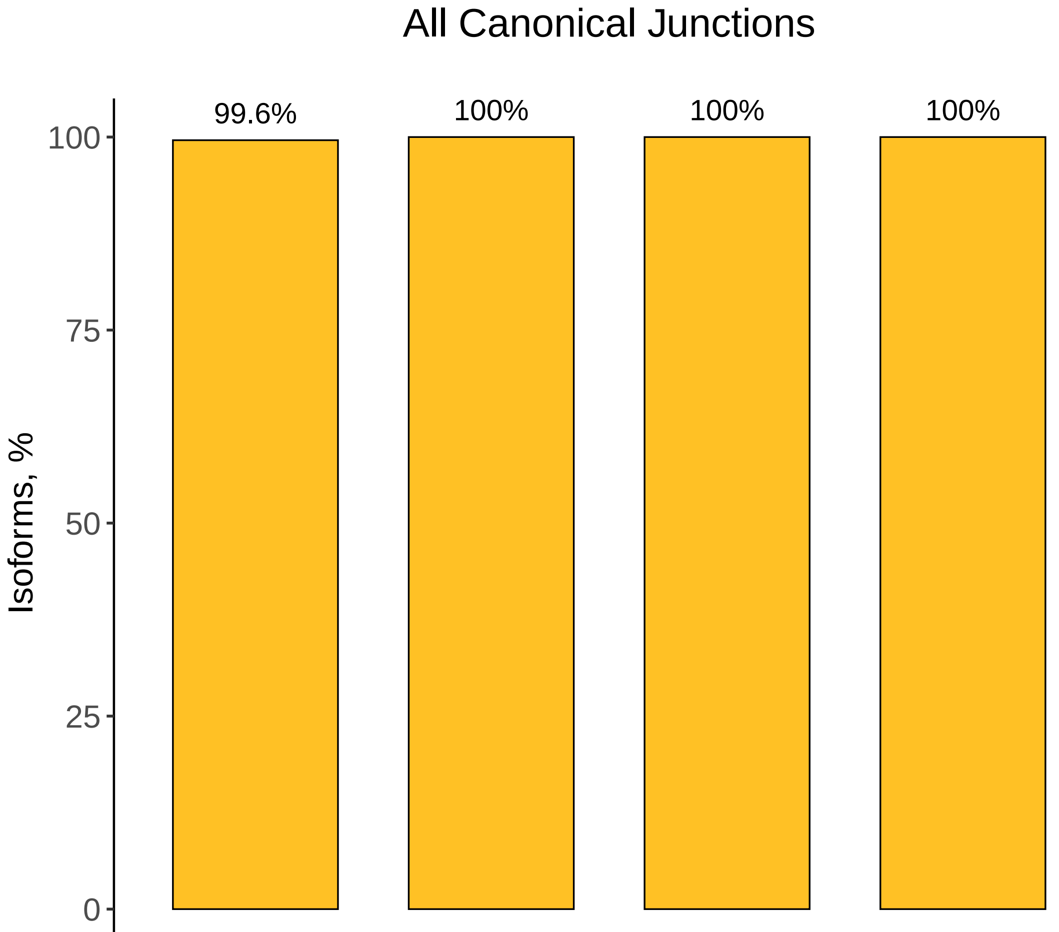

Non−Canonical Junctions
0.4%
0.4
0.3
0.2
0.1
0.0
Isoforms, %
FSM

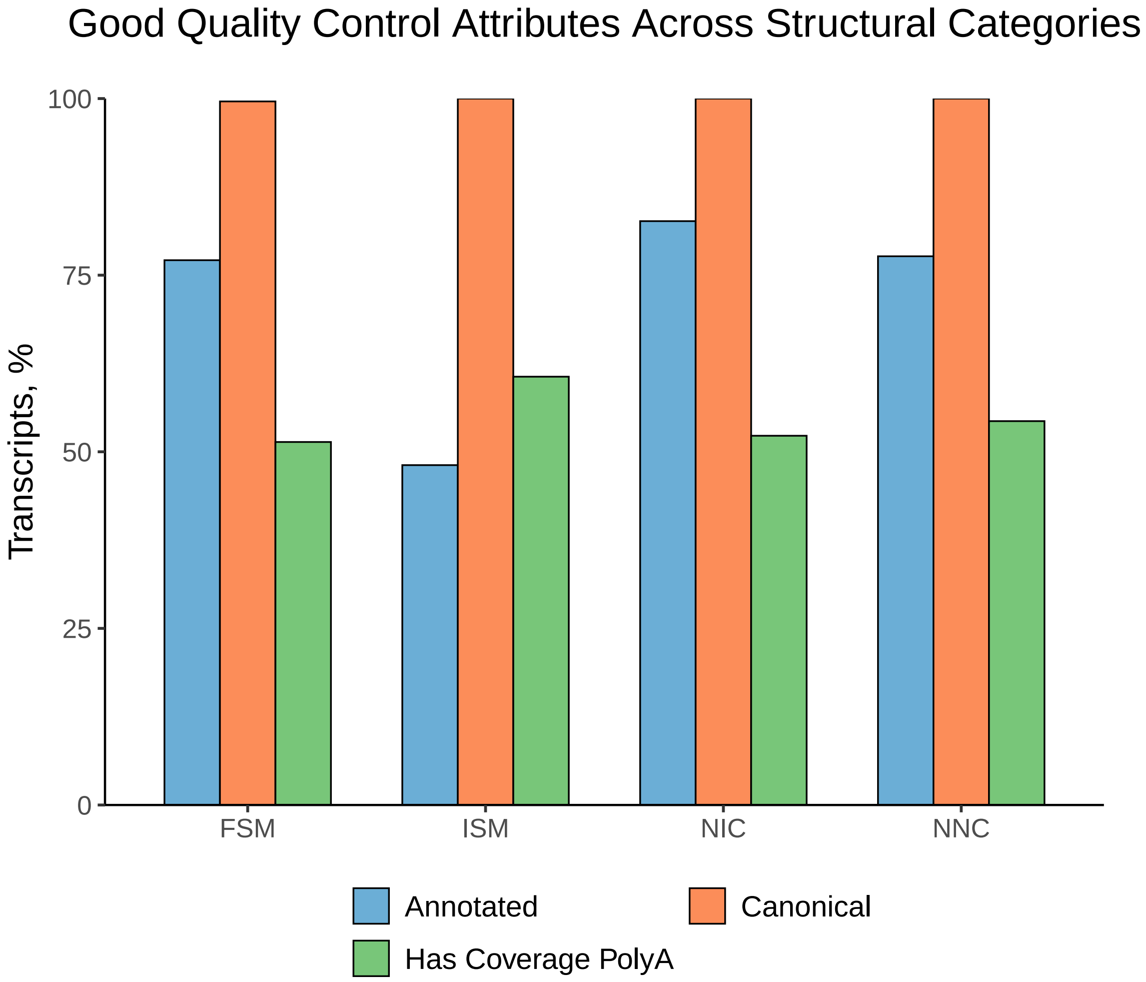

Quality Control Attributes Across Structural Categories
Transcripts, %
100
75
50
25
0
FSM
Non−canonical
RT switching

Features of Good Quality


Annotation Support
Isoforms, %
0
20
40
60
80
77.1%
48.1%
82.6%
77.7%
FSM
ISM
NIC
NNC


PolyA Support
Isoforms, %
60
40
20
0
51.4%
60.6%
52.3%
54.3%
FSM
ISM
NIC
NNC


All Canonical Junctions
Isoforms, %
100
75
50
25
0
99.6%
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