SQANTI3 report

Unique Genes: 8938

Unique Isoforms: 23894

Gene Classification

Category	Genes, count
Annotated Genes	8906
Novel Genes	32

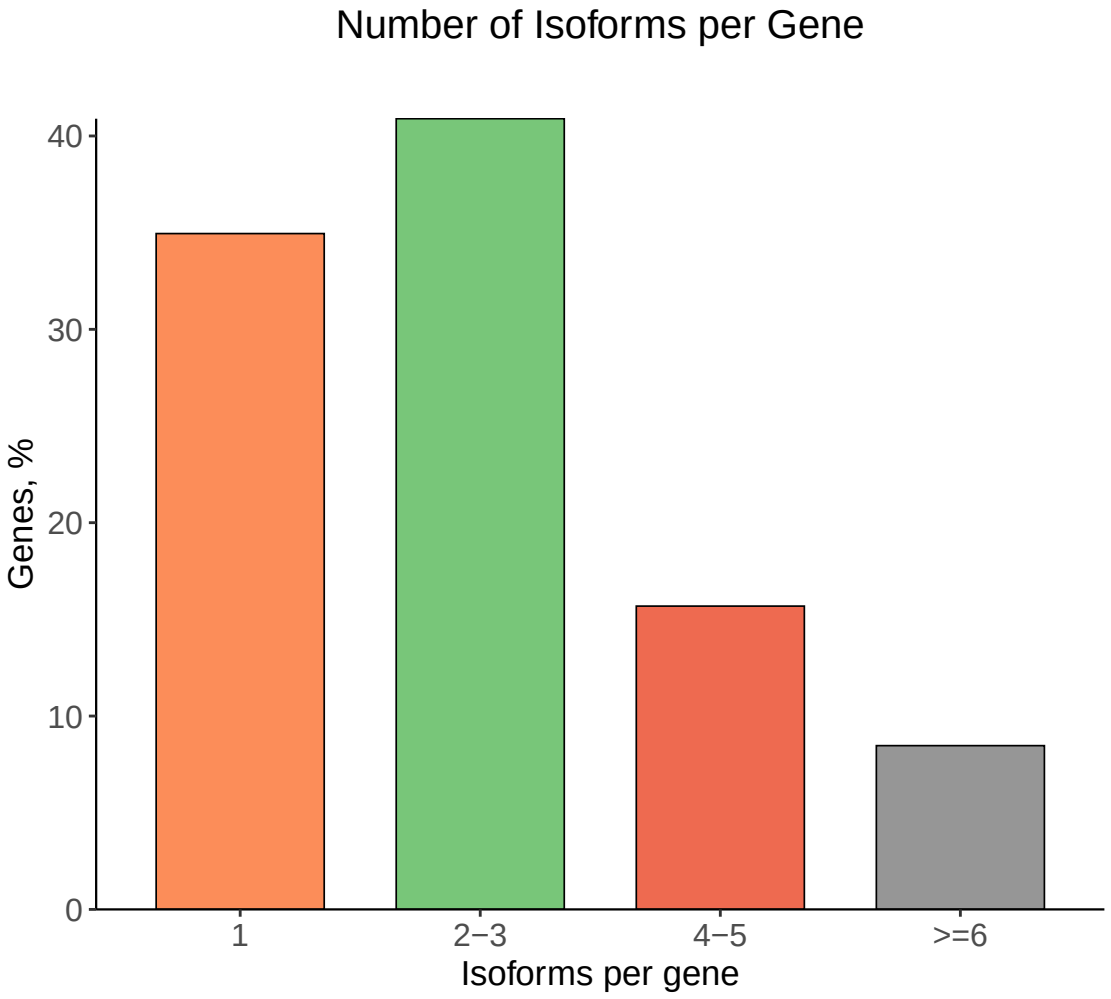
Transcript Classification

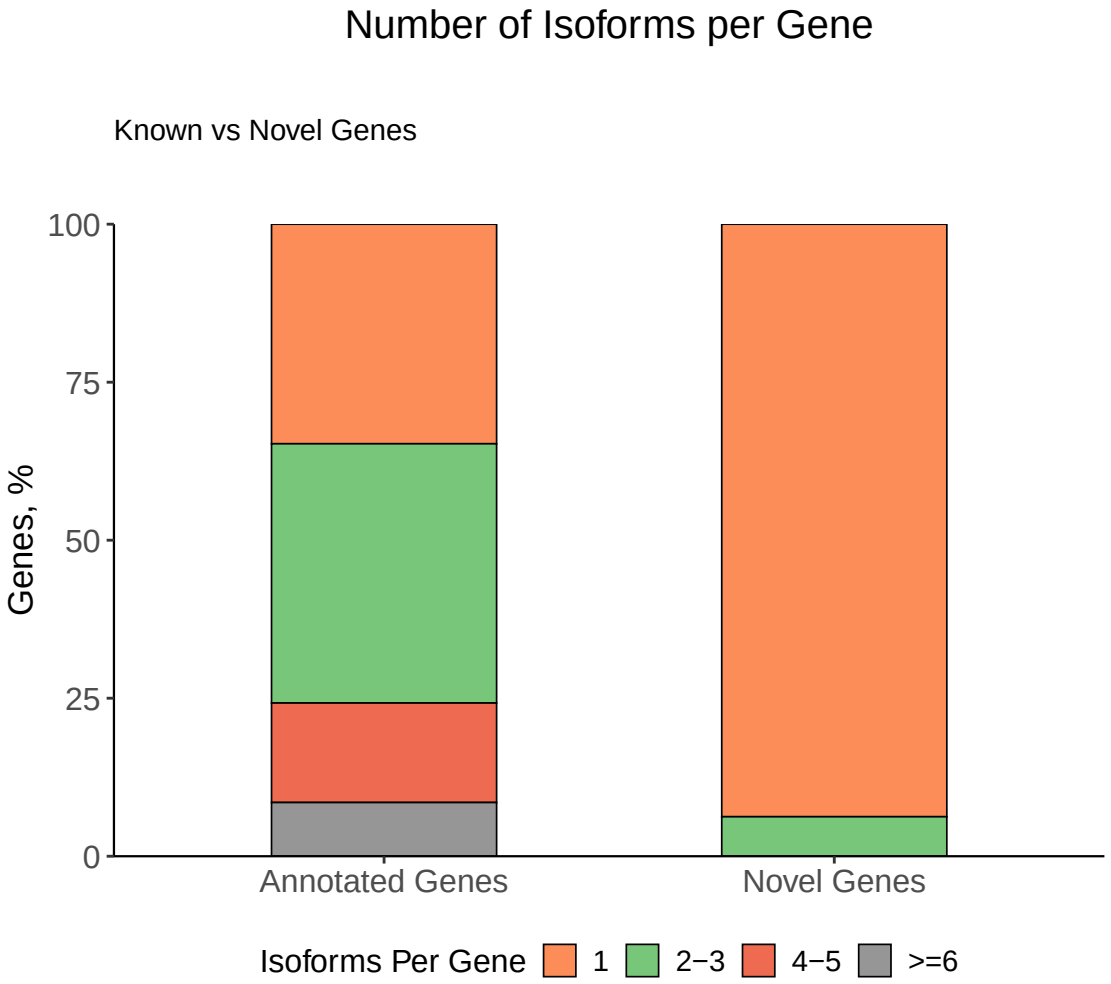
Category	Isoforms, count
FSM	18039
ISM	3458
NIC	2037
NNC	303
Genic Genomic	3
Antisense	13
Fusion	19
Intergenic	22
Genic Intron	0

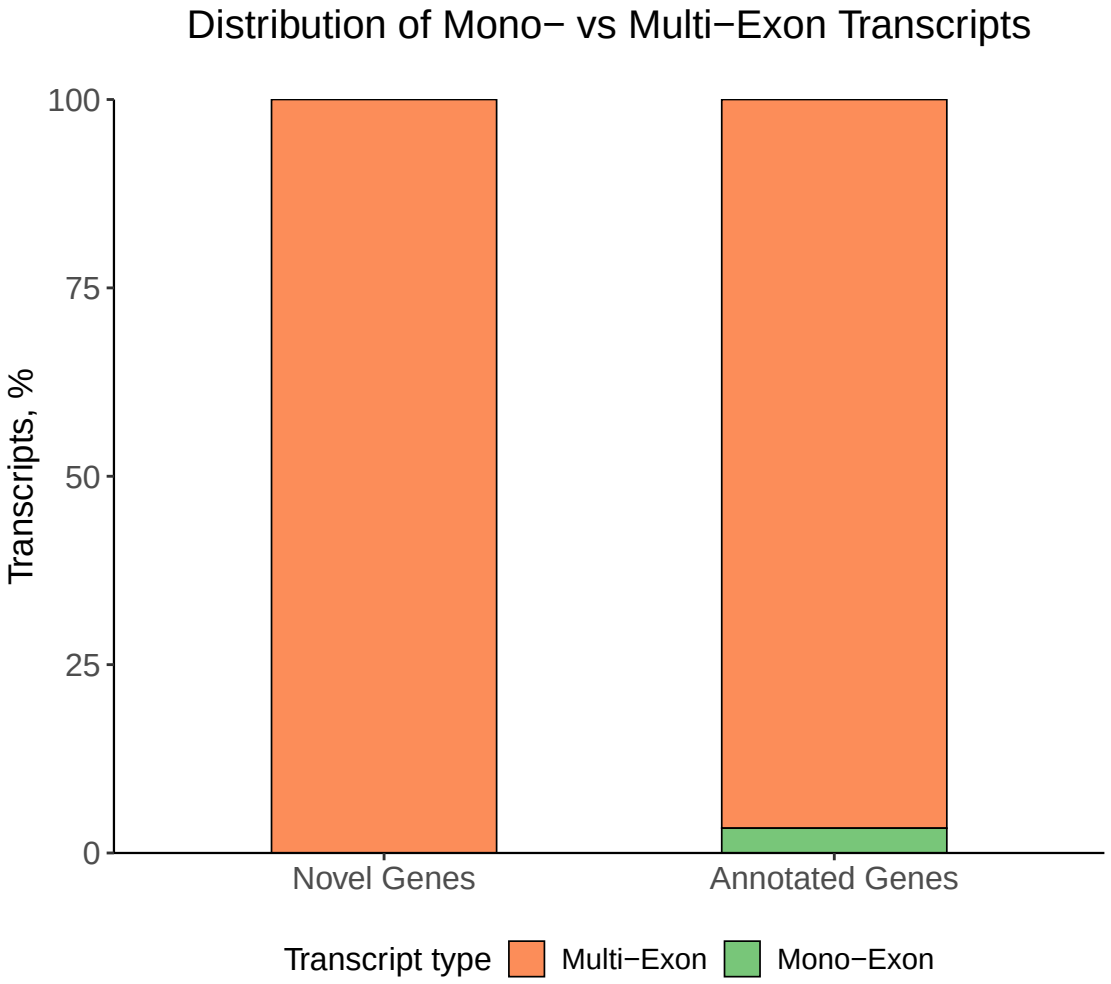
Splice Junction Classification

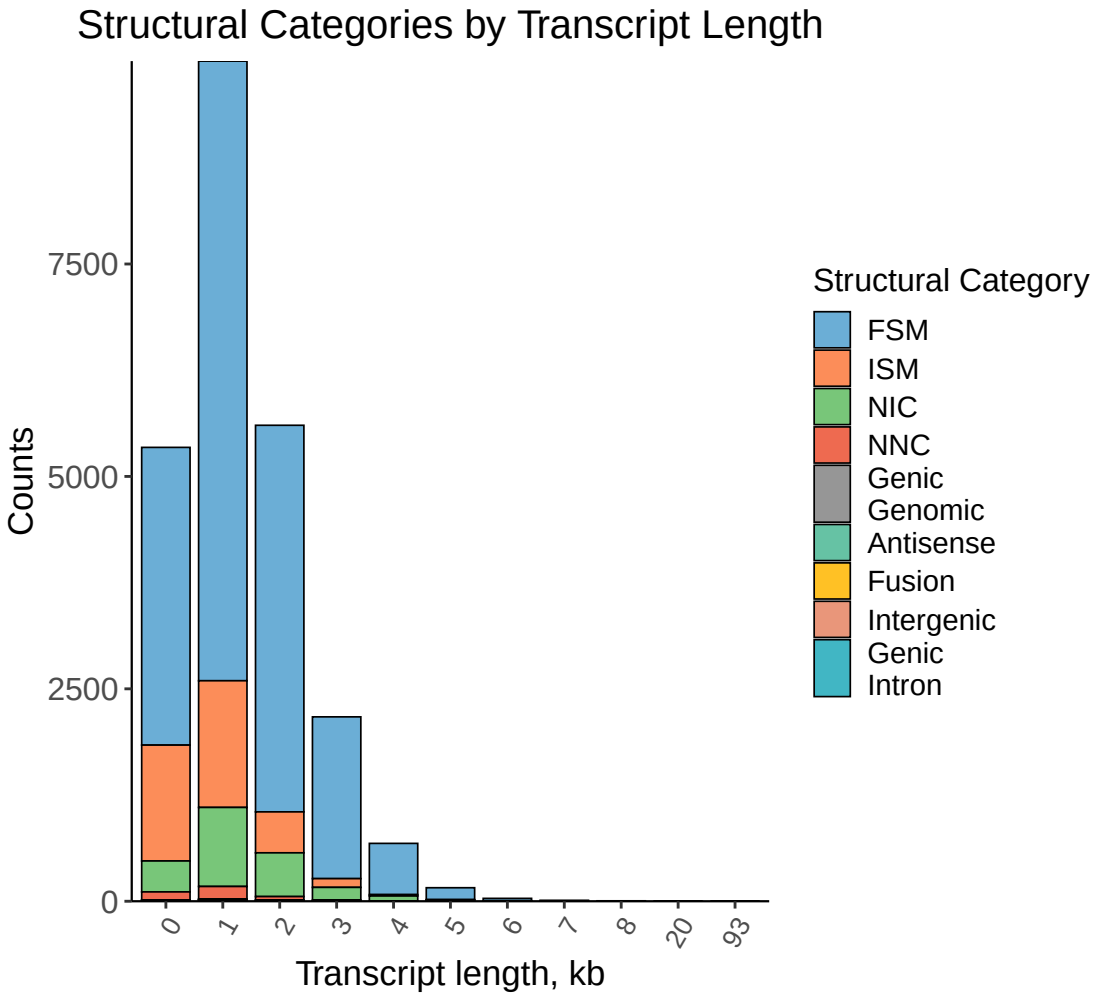
Category	SJs, count	Percent
Known canonical	72049	99.15
Known Non−canonical	32	0.04
Novel canonical	583	0.80
Novel Non−canonical	0	0.00

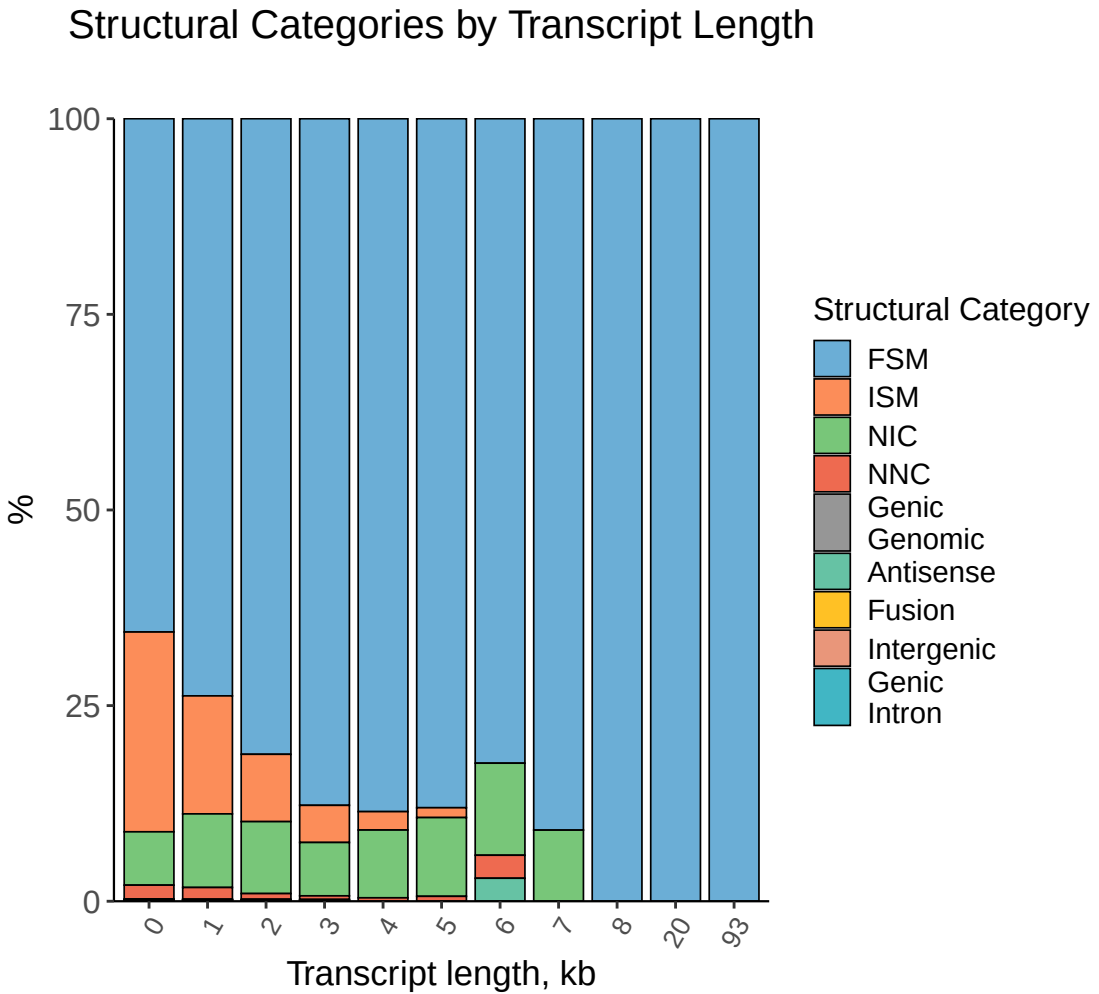
Gene Characterization

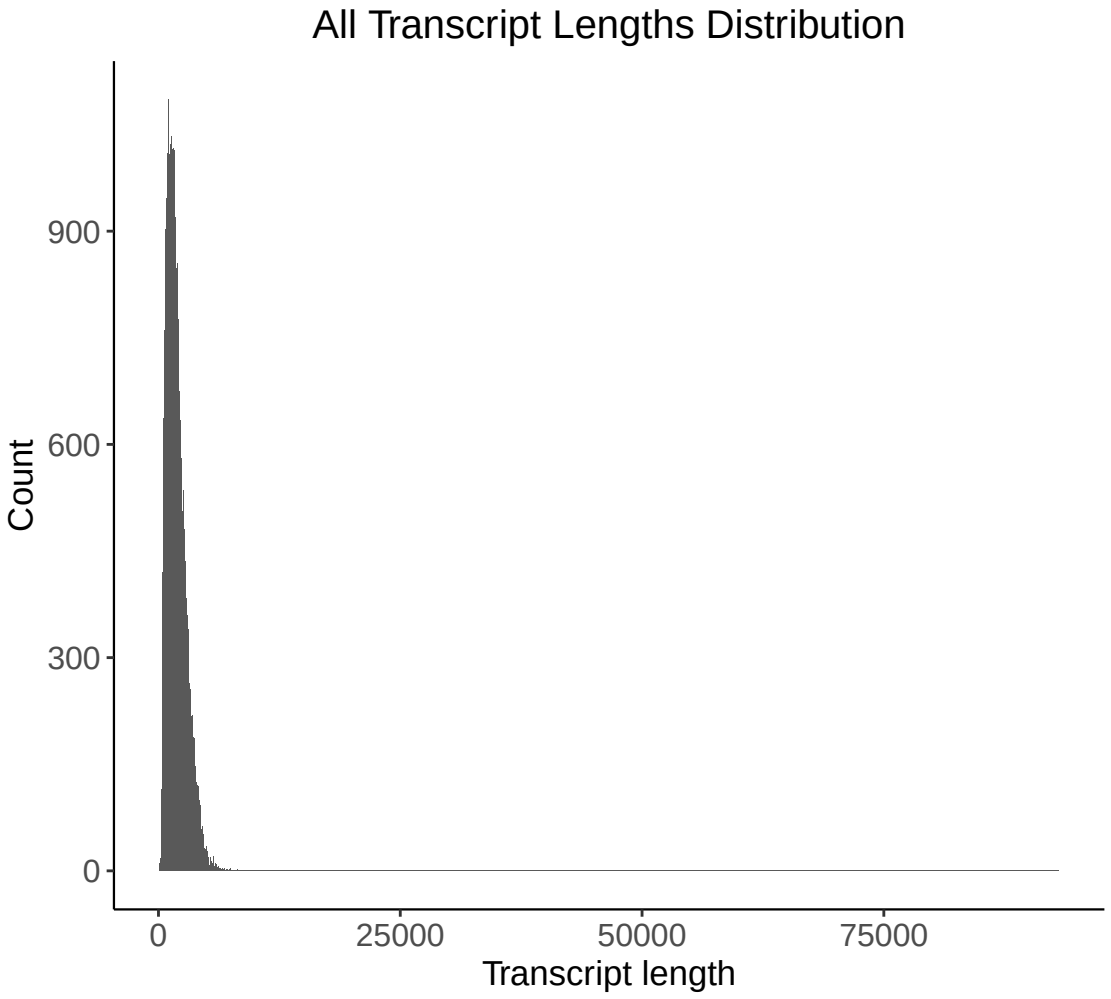

Number of Isoforms per Gene
Genes, %
40
30
20
10
0
1
2−3
4−5
>=6
Isoforms per gene

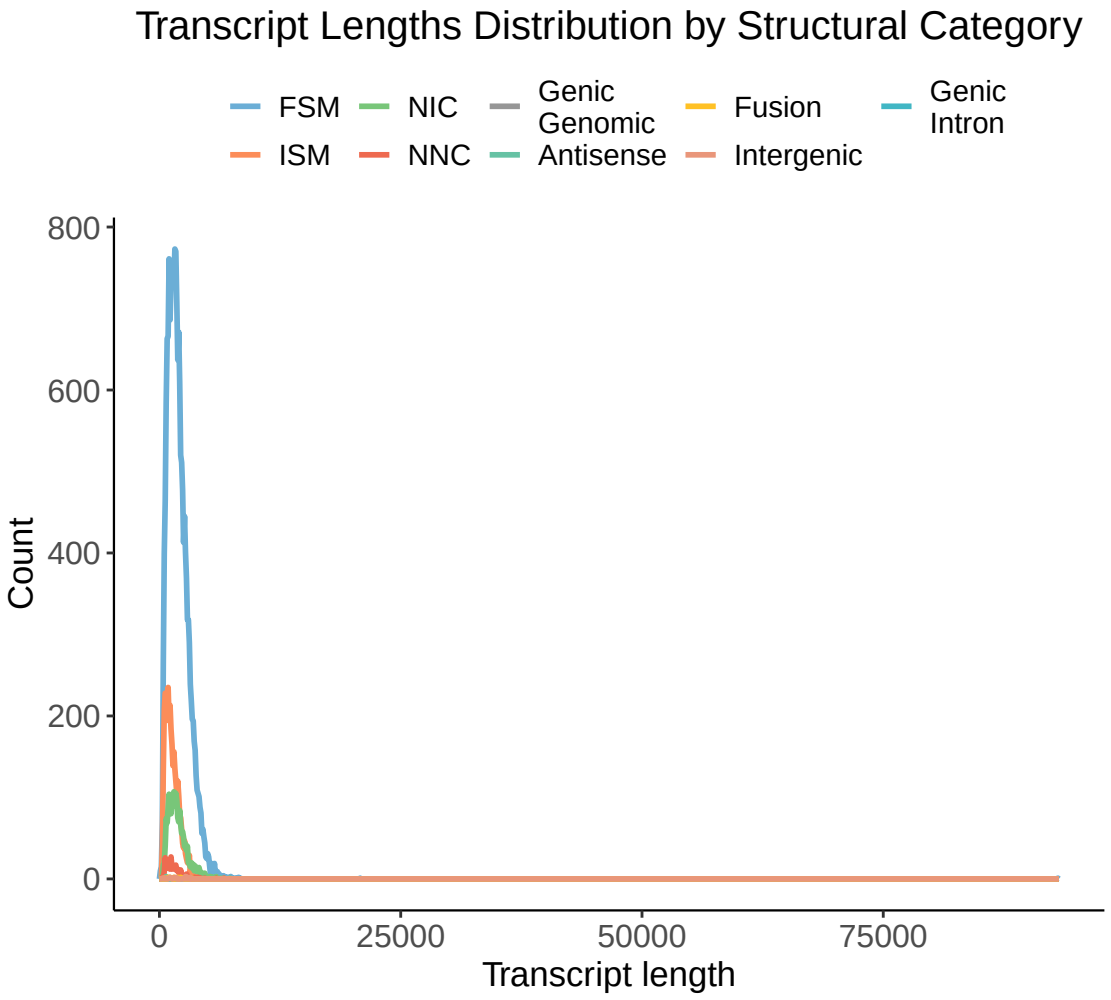

Number of Isoforms per Gene
Known vs Novel Genes
Genes, %
100
75
50
25
0
Annotated Genes
Novel Genes
Isoforms Per Gene
1
2−3
4−5
>=6

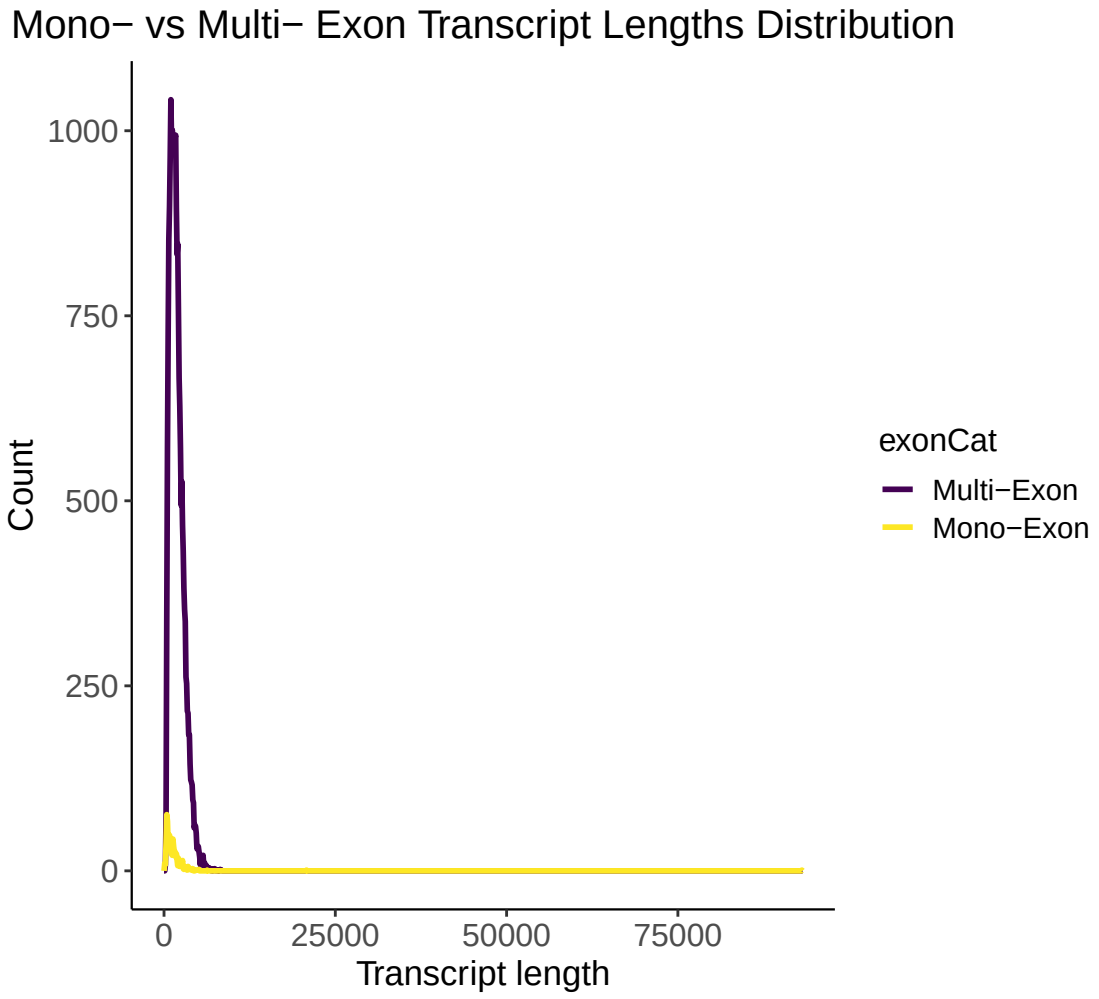

Distribution of Mono− vs Multi−Exon Transcripts
100
75
50
25
0
Transcripts, %
Novel Genes
Annotated Genes
Transcript type
Multi−Exon
Mono−Exon


Structural Categories by Transcript Length
Counts
7500
5000
2500
0
0
1
2
3
4
5
6
7
8
20
93
Transcript length, kb
Structural Category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

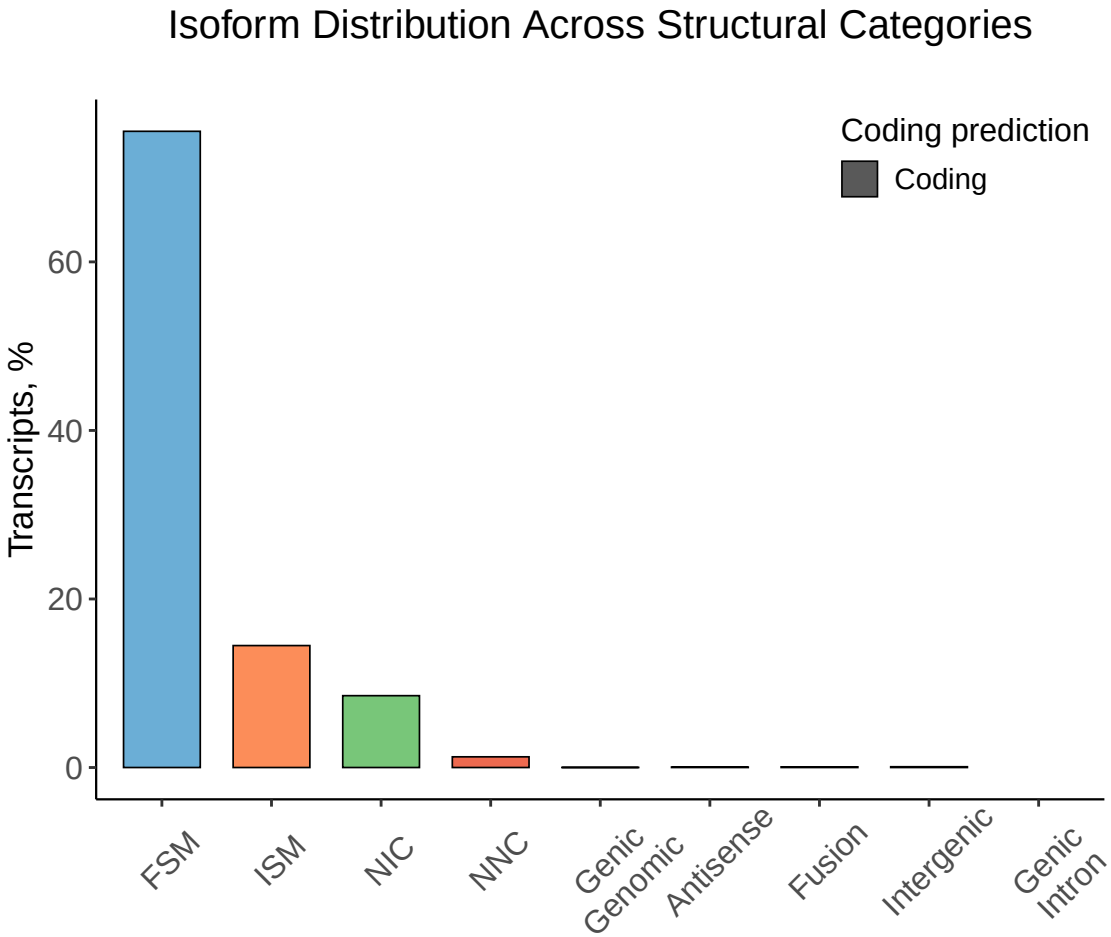

Structural Categories by Transcript Length
100
75
50
25
0
%
0
1
2
3
4
5
6
7
8
20
93
Transcript length, kb
Structural Category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

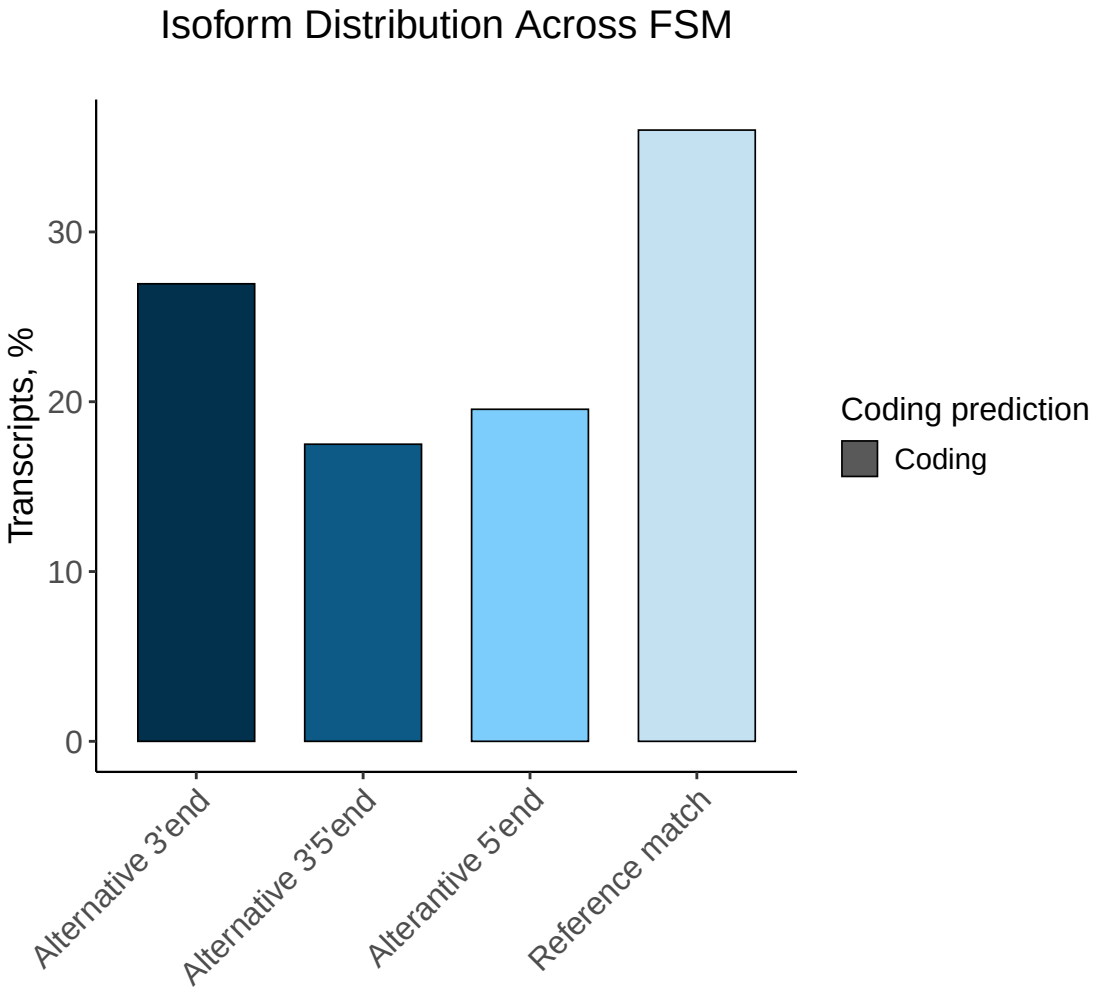

All Transcript Lengths Distribution
Count
900
600
300
0
0
25000
50000
75000
Transcript length

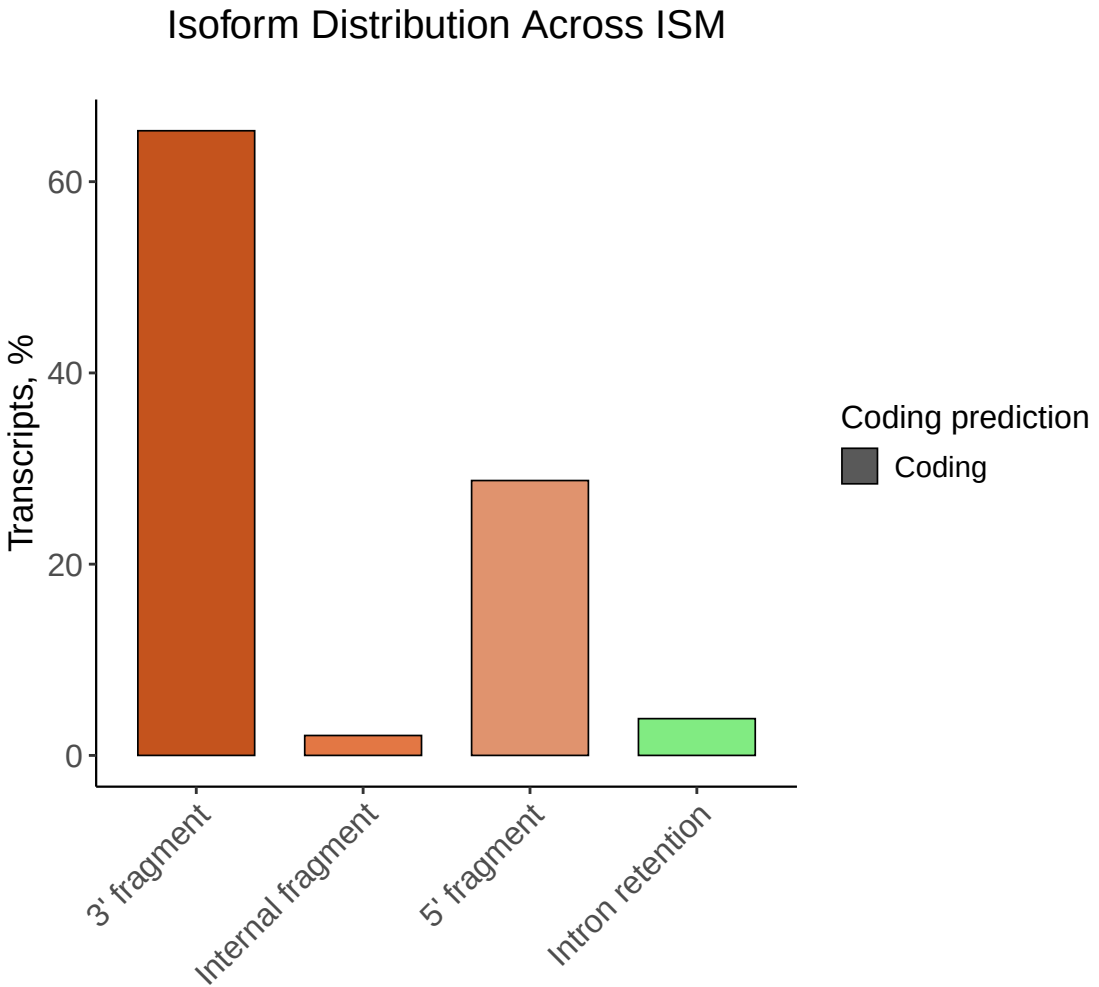

Transcript Lengths Distribution by Structural Category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron
800
600
400
200
0
Count
0
25000
50000
75000
Transcript length

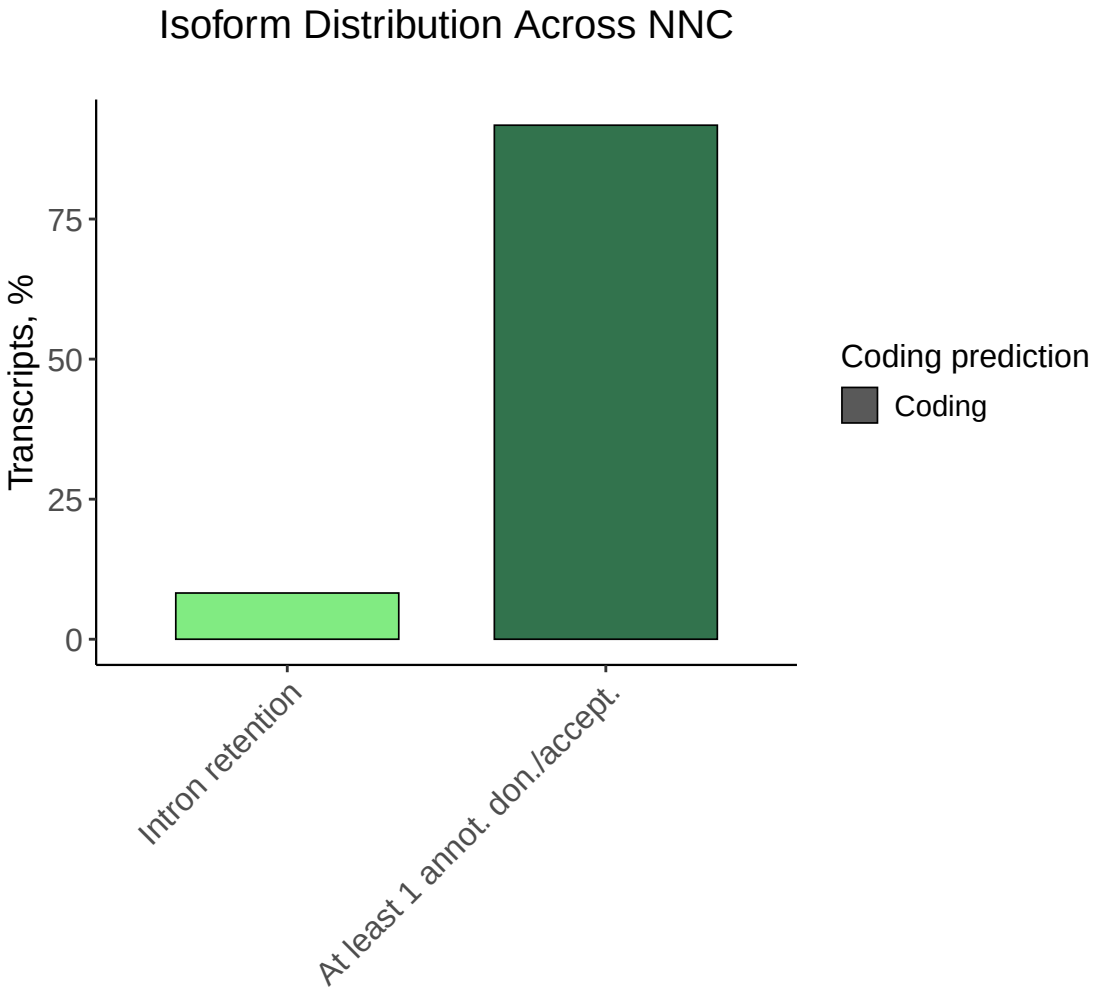

Mono− vs Multi− Exon Transcript Lengths Distribution
Count
1000
750
500
250
0
0
25000
50000
75000
Transcript length
exonCat
Multi−Exon
Mono−Exon

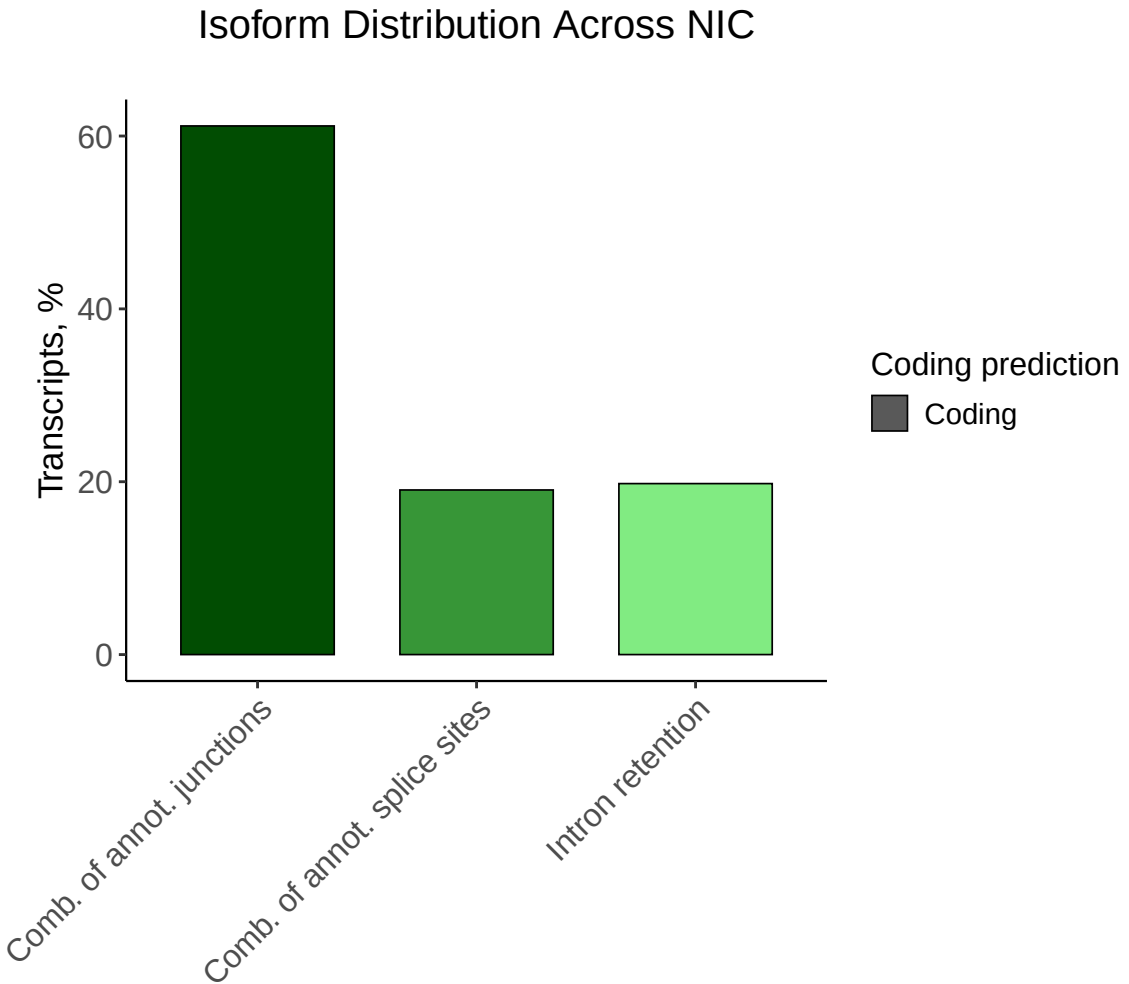
Structural Isoform Characterization

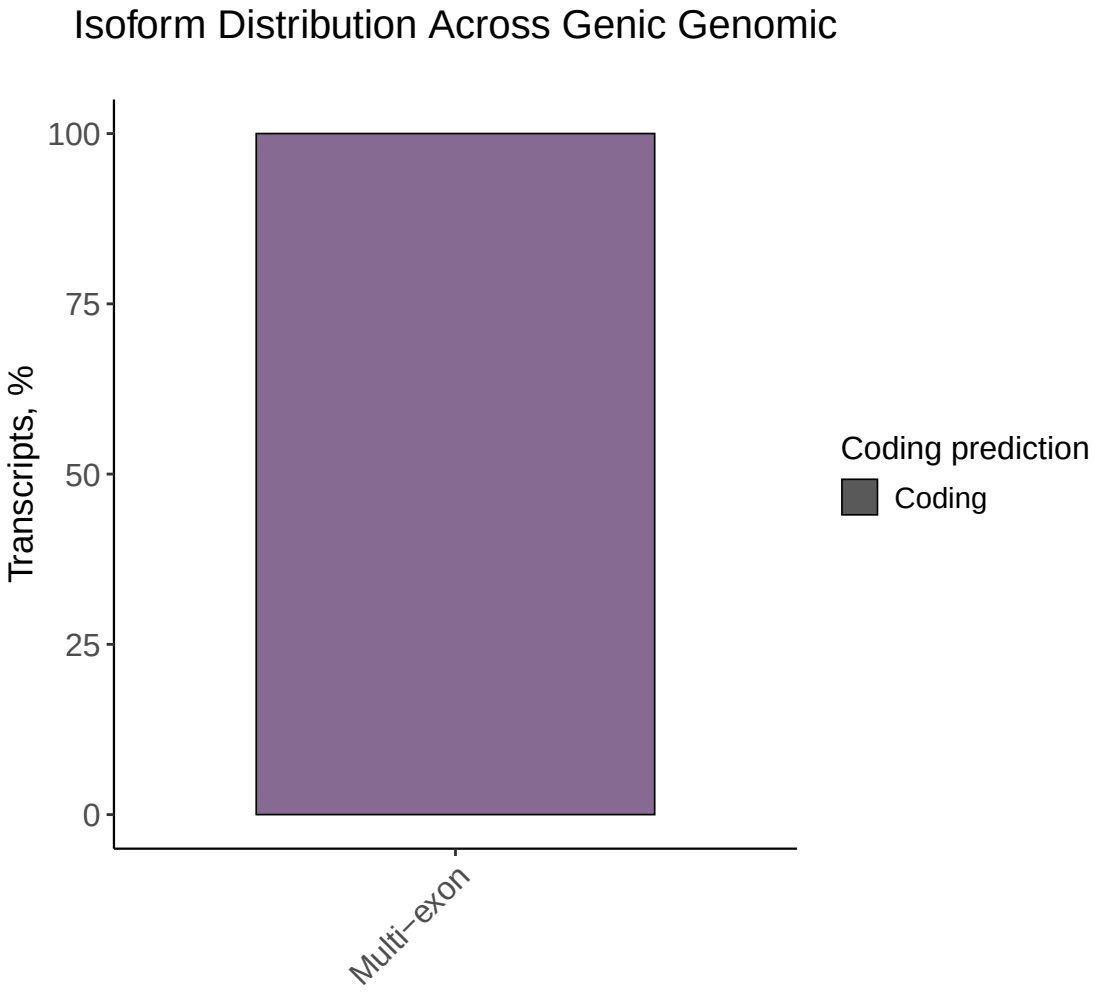

Isoform Distribution Across Structural Categories
Coding prediction
Coding
Transcripts, %
60
40
20
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

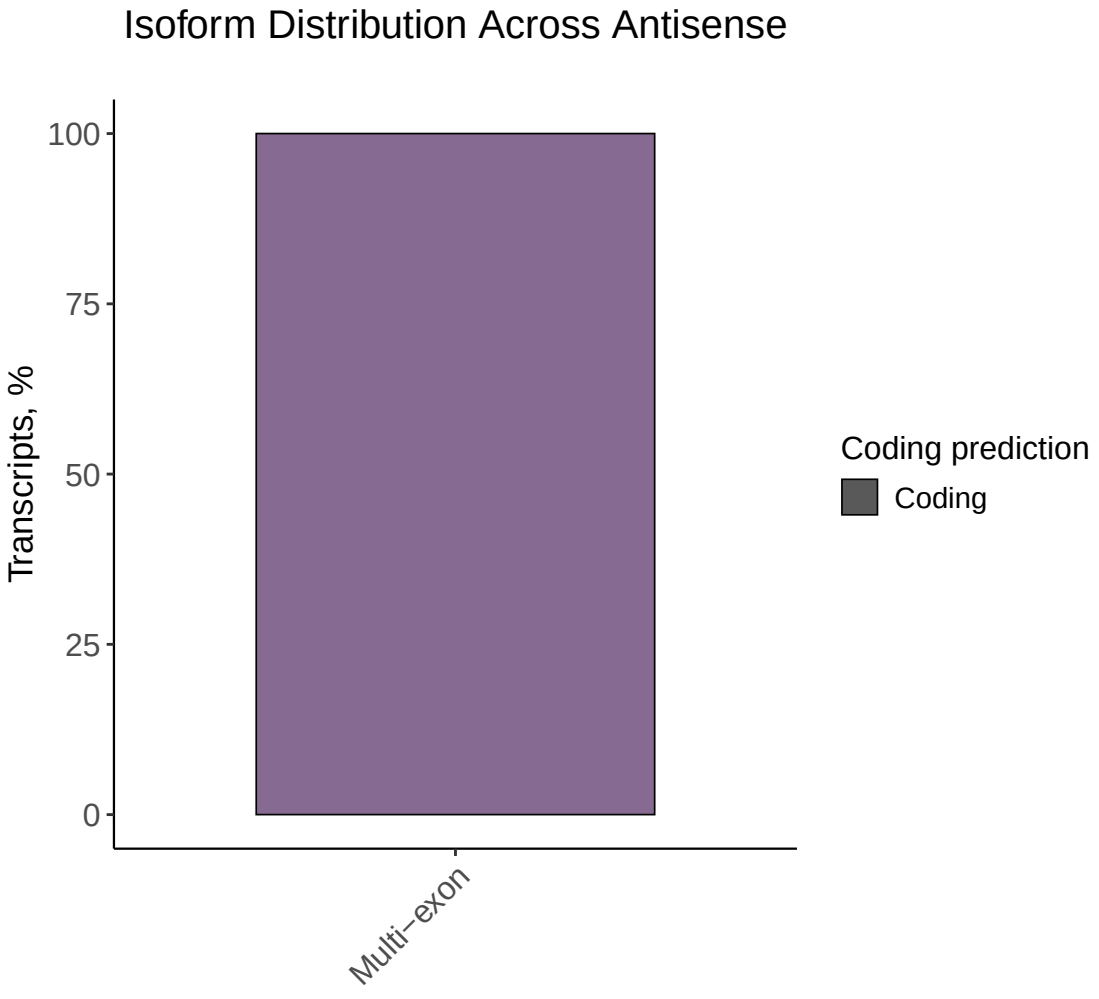

Isoform Distribution Across FSM
Transcripts, %
30
20
10
0
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
Coding prediction
Coding

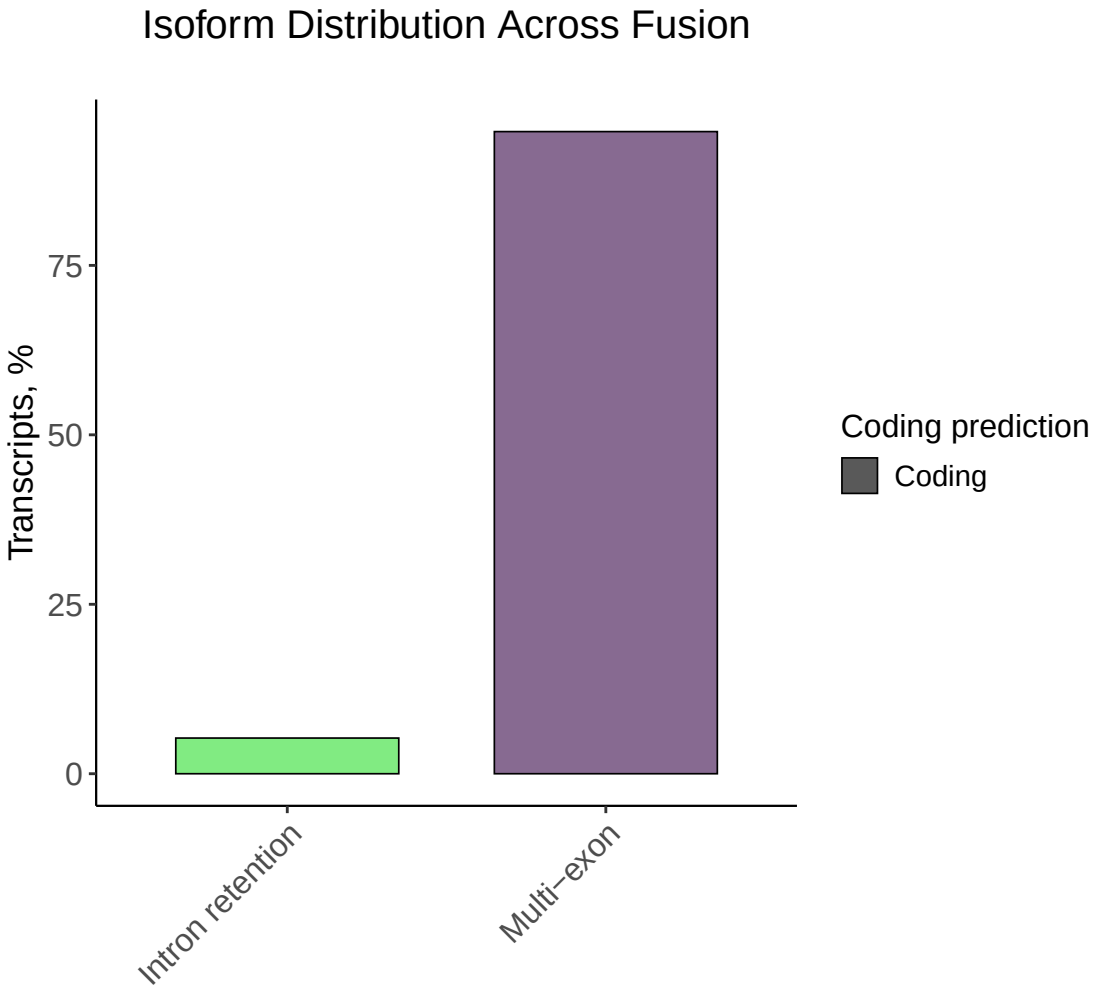

Isoform Distribution Across ISM
Transcripts, %
60
40
20
0
3' fragment
Internal fragment
5' fragment
Intron retention
Coding prediction
Coding

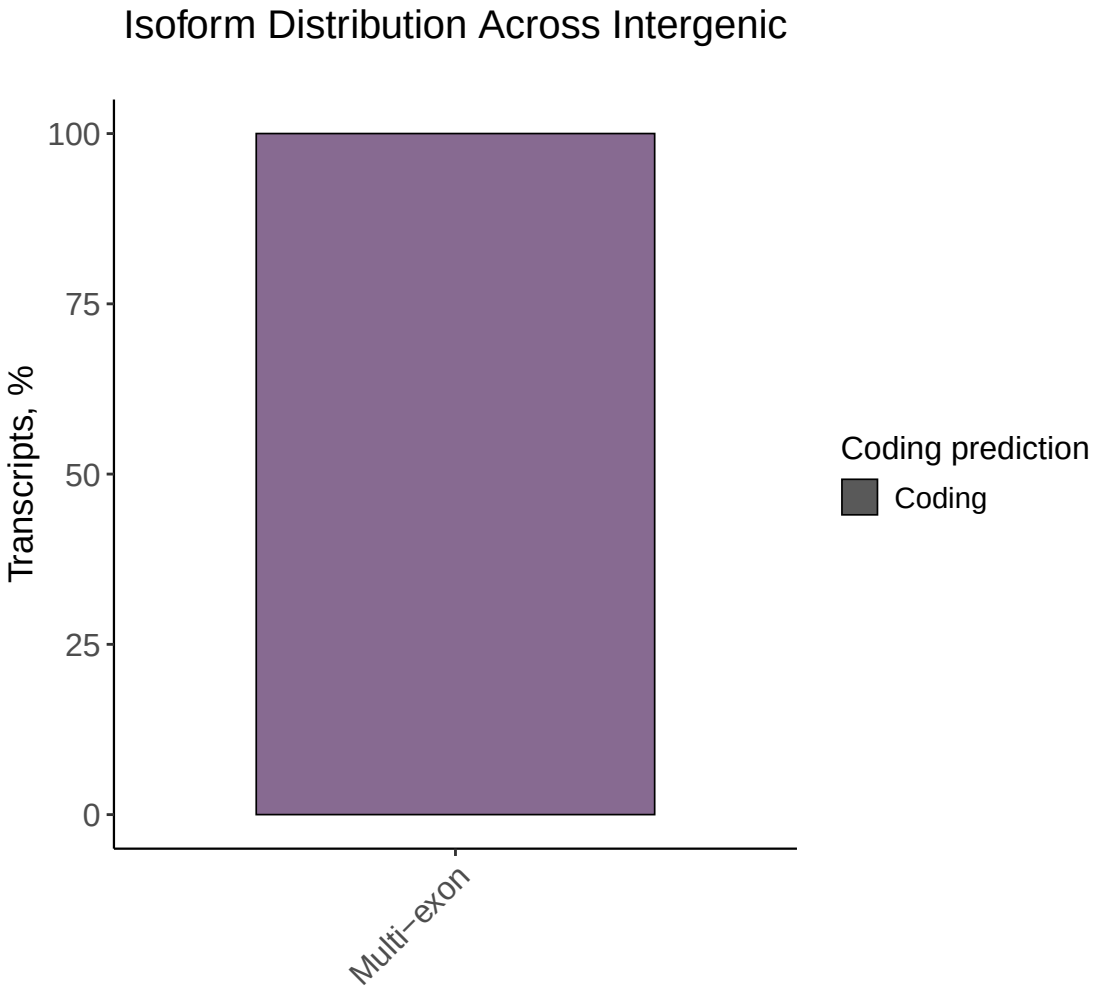

Isoform Distribution Across NNC
Transcripts, %
75
50
25
0
Intron retention
At least 1 annot. don./accept.
Coding prediction
Coding

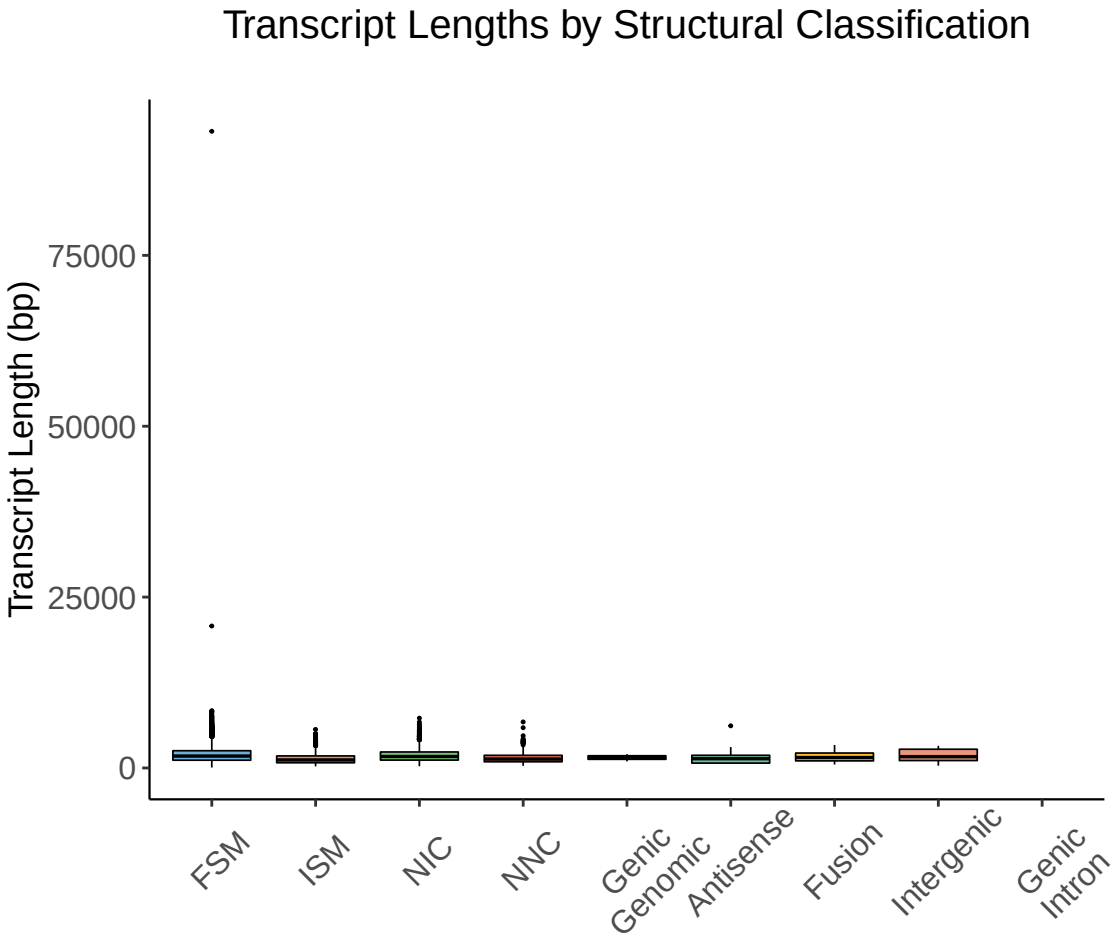

Isoform Distribution Across NIC
Transcripts, %
60
40
20
0
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Coding prediction
Coding

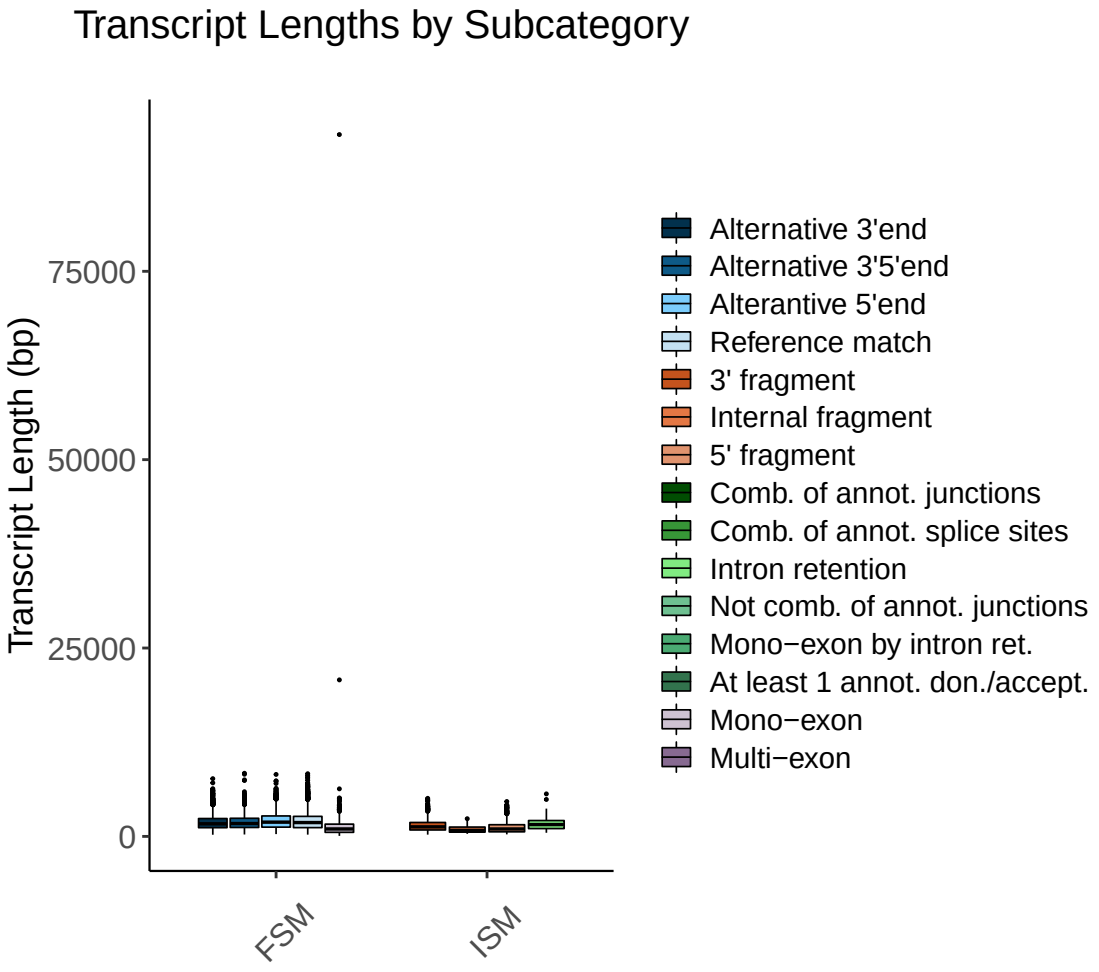

Isoform Distribution Across Genic Genomic
Transcripts, %
100
75
50
25
0
Multi−exon
Coding prediction
Coding

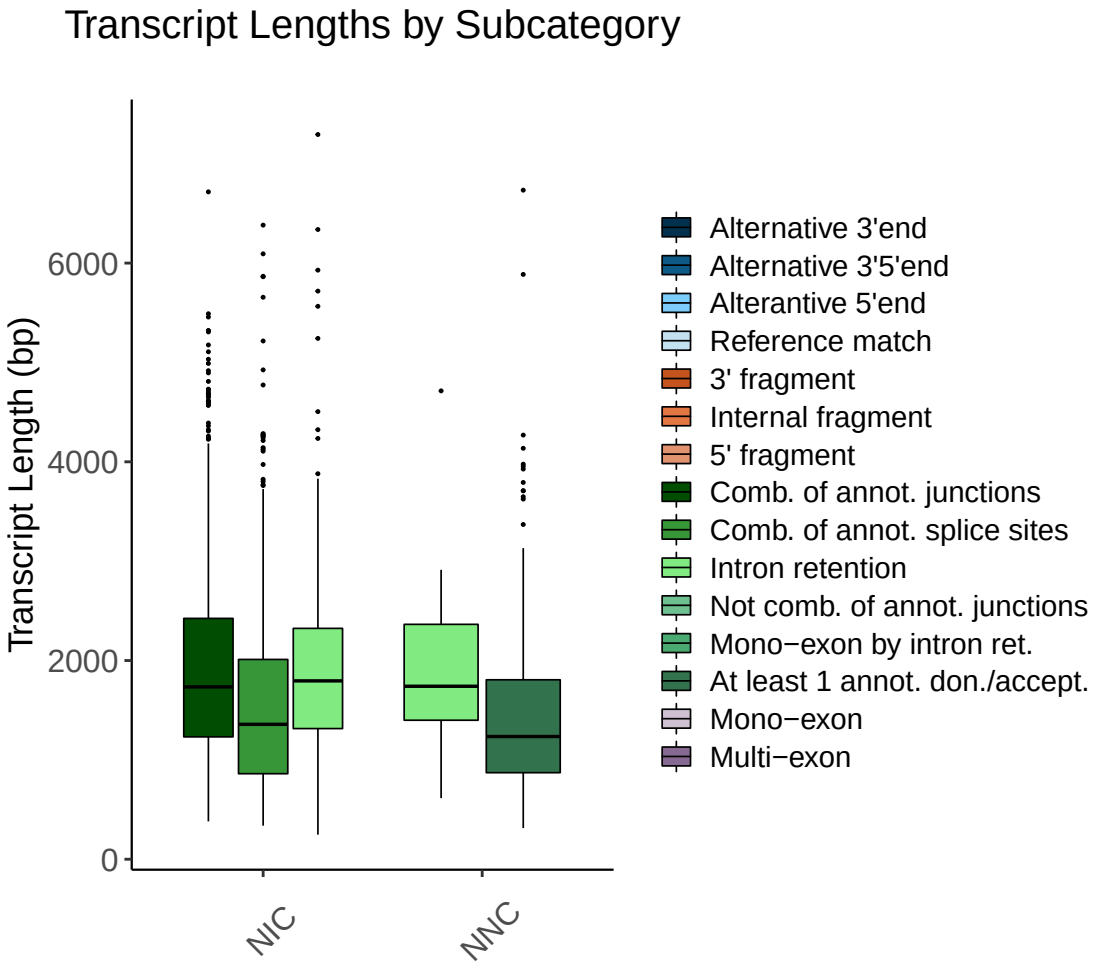

Isoform Distribution Across Antisense
Transcripts, %
100
75
50
25
0
Multi−exon
Coding prediction
Coding

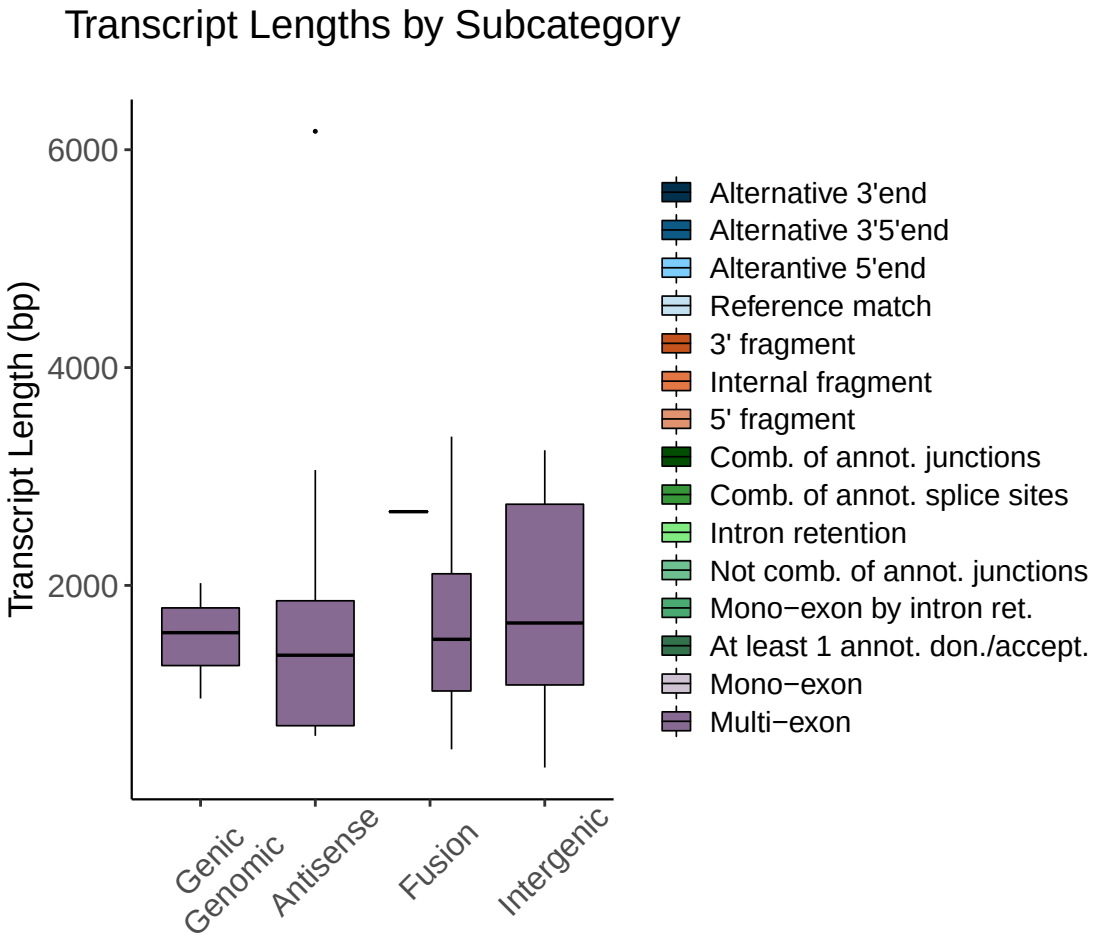

Isoform Distribution Across Fusion
Transcripts, %
75
50
25
0
Intron retention
Multi−exon
Coding prediction
Coding

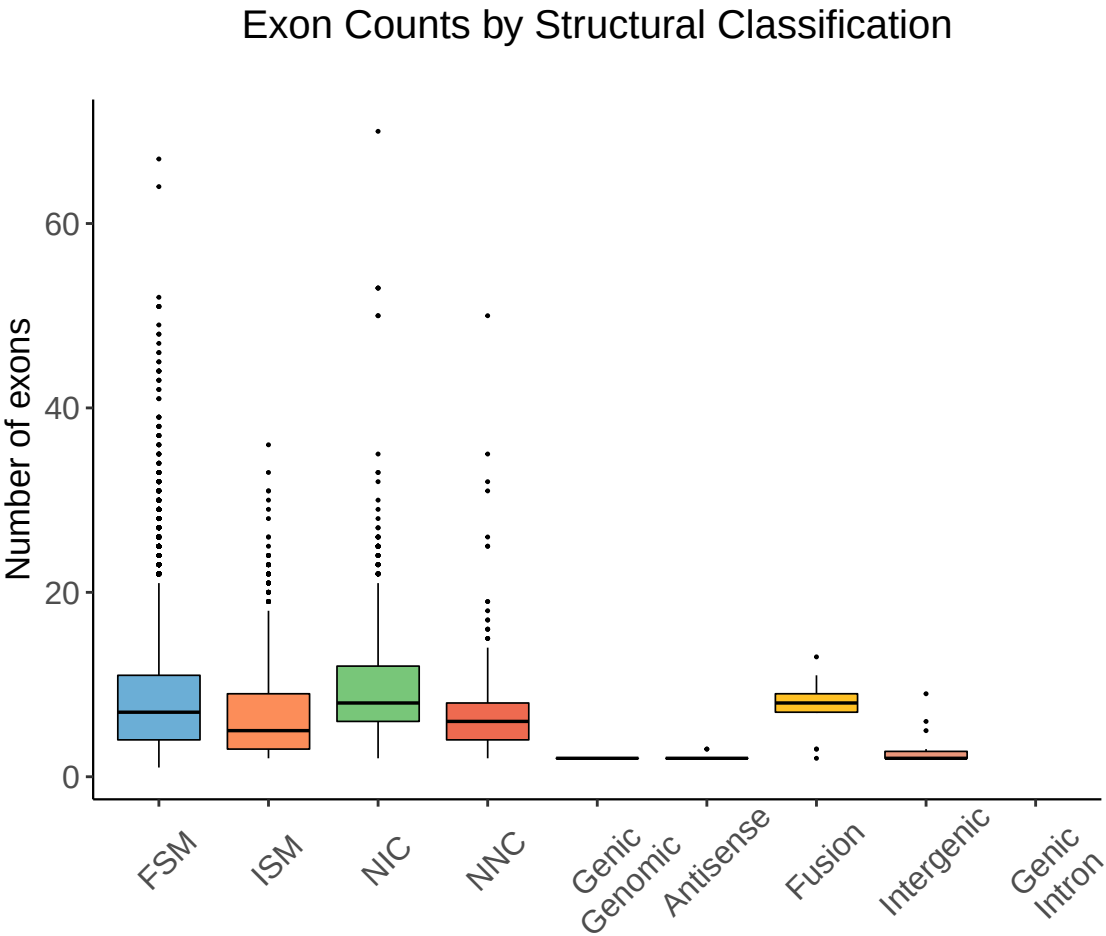

Isoform Distribution Across Intergenic
Transcripts, %
100
75
50
25
0
Multi−exon
Coding prediction
Coding

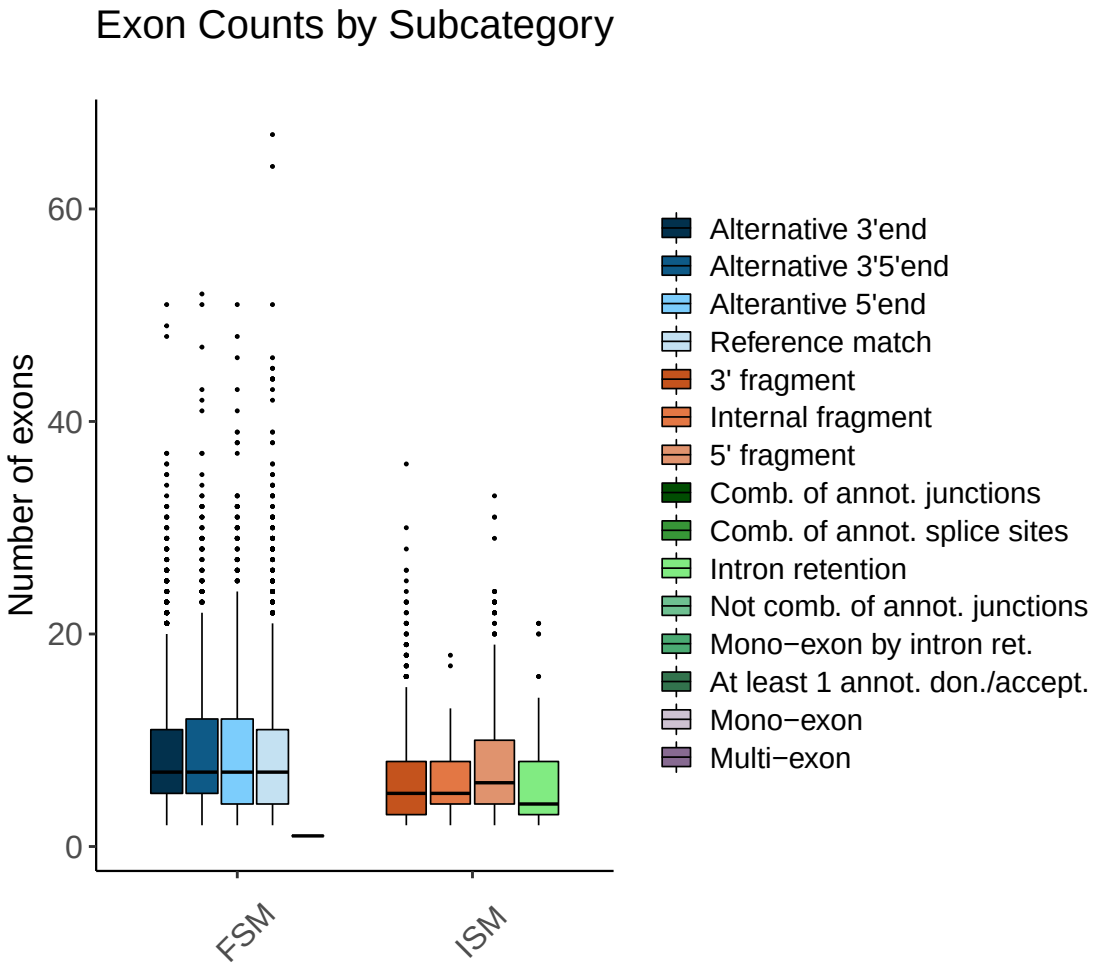

Transcript Lengths by Structural Classification
Transcript Length (bp)
75000
50000
25000
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

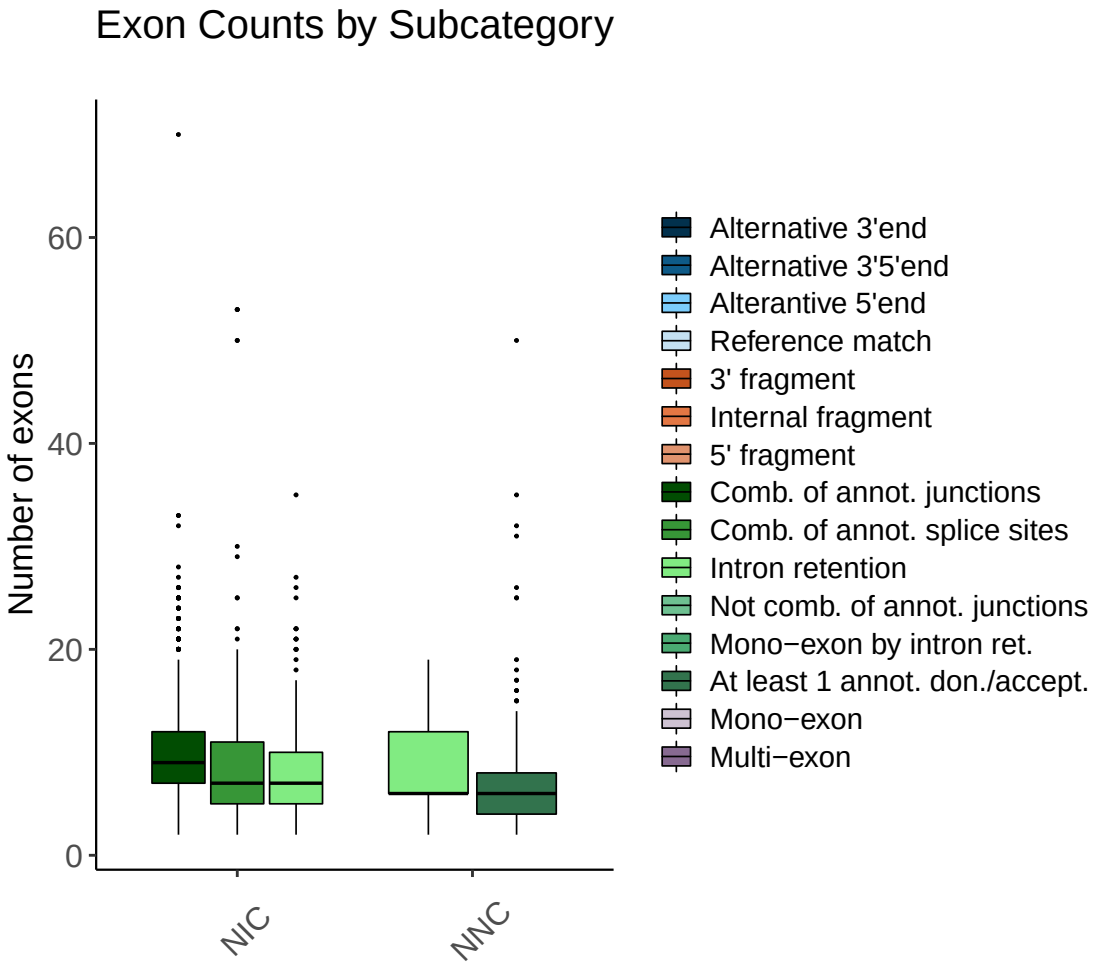

Transcript Lengths by Subcategory
Transcript Length (bp)
75000
50000
25000
0
FSM
ISM
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.
At least 1 annot. don./accept.
Mono−exon
Multi−exon

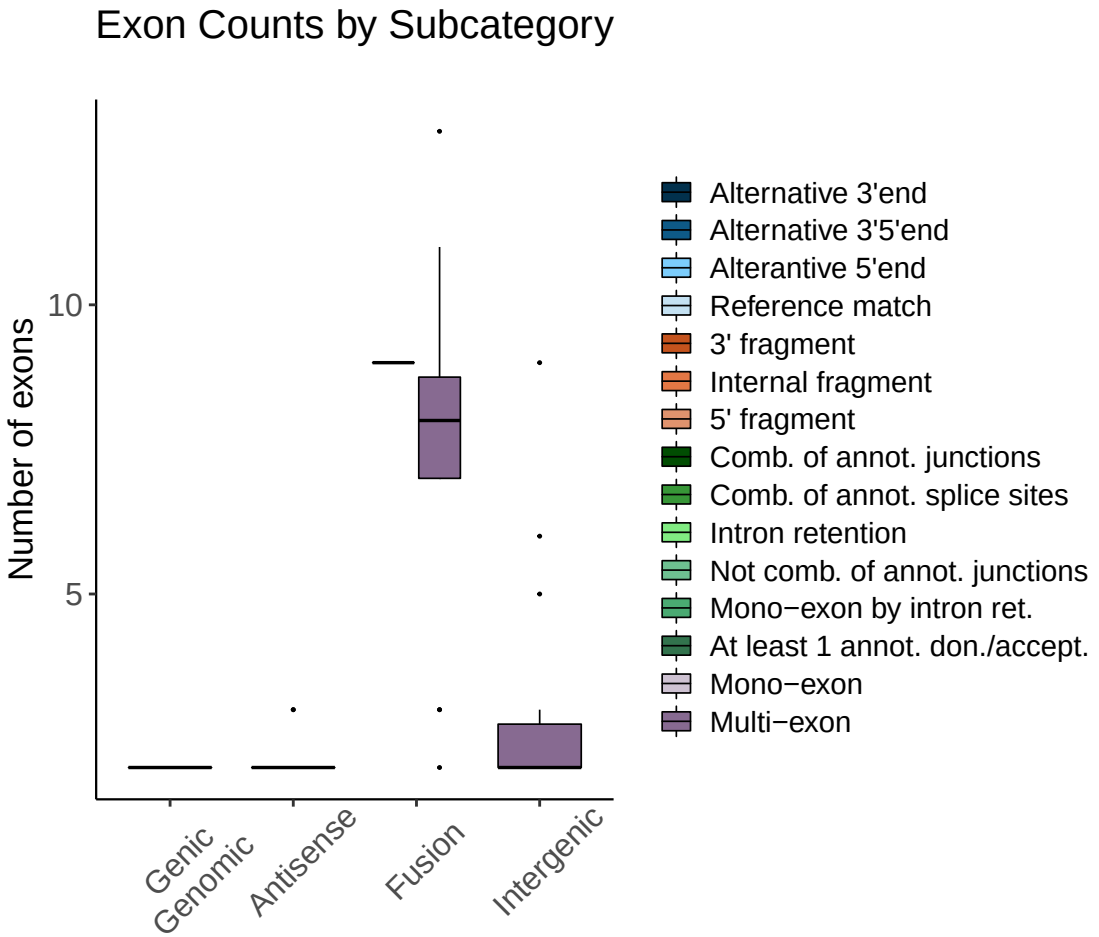

Transcript Lengths by Subcategory
Transcript Length (bp)
6000
4000
2000
0
NIC
NNC
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon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

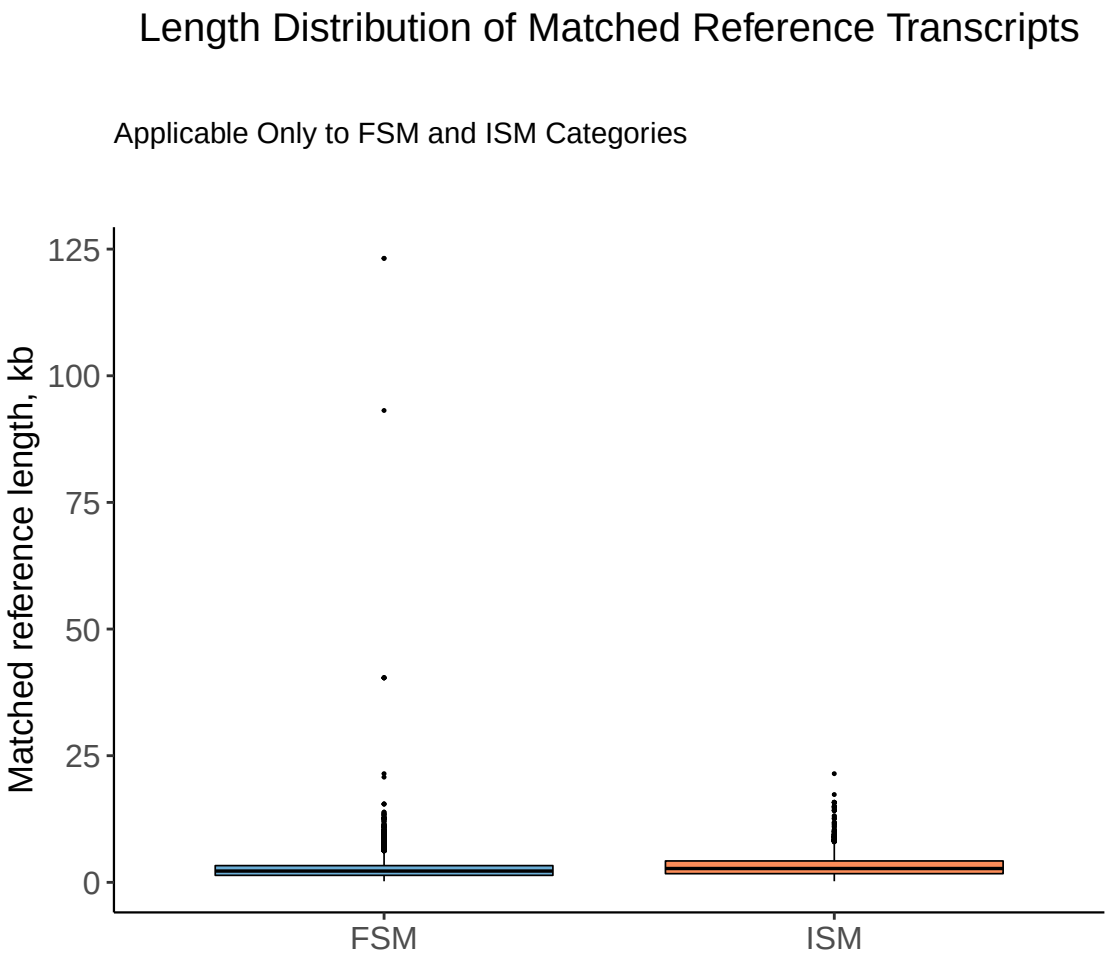

Transcript Lengths by Subcategory
Transcript Length (bp)
6000
4000
2000
Genic Genomic
Antisense
Fusion
Intergenic
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not 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


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

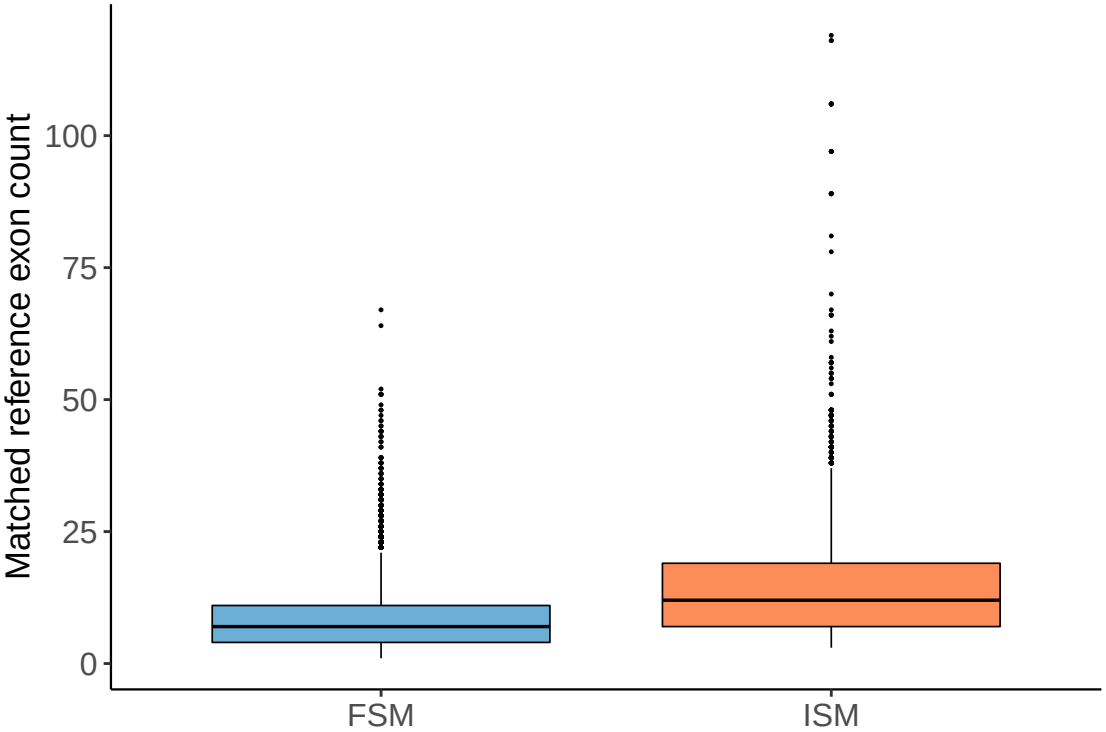

Exon Counts by Subcategory
Number of exons
0
20
40
60
NIC
NNC
Alternative 3'end
Alternative 3'5'end
Alterantive 5'end
Reference match
3' fragment
Internal fragment
5' fragment
Comb. of annot. junctions
Comb. of annot. splice sites
Intron retention
Not comb. of annot. junctions
Mono−exon by intron ret.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


Exon Counts by Subcategory
Number of exons
10
5
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

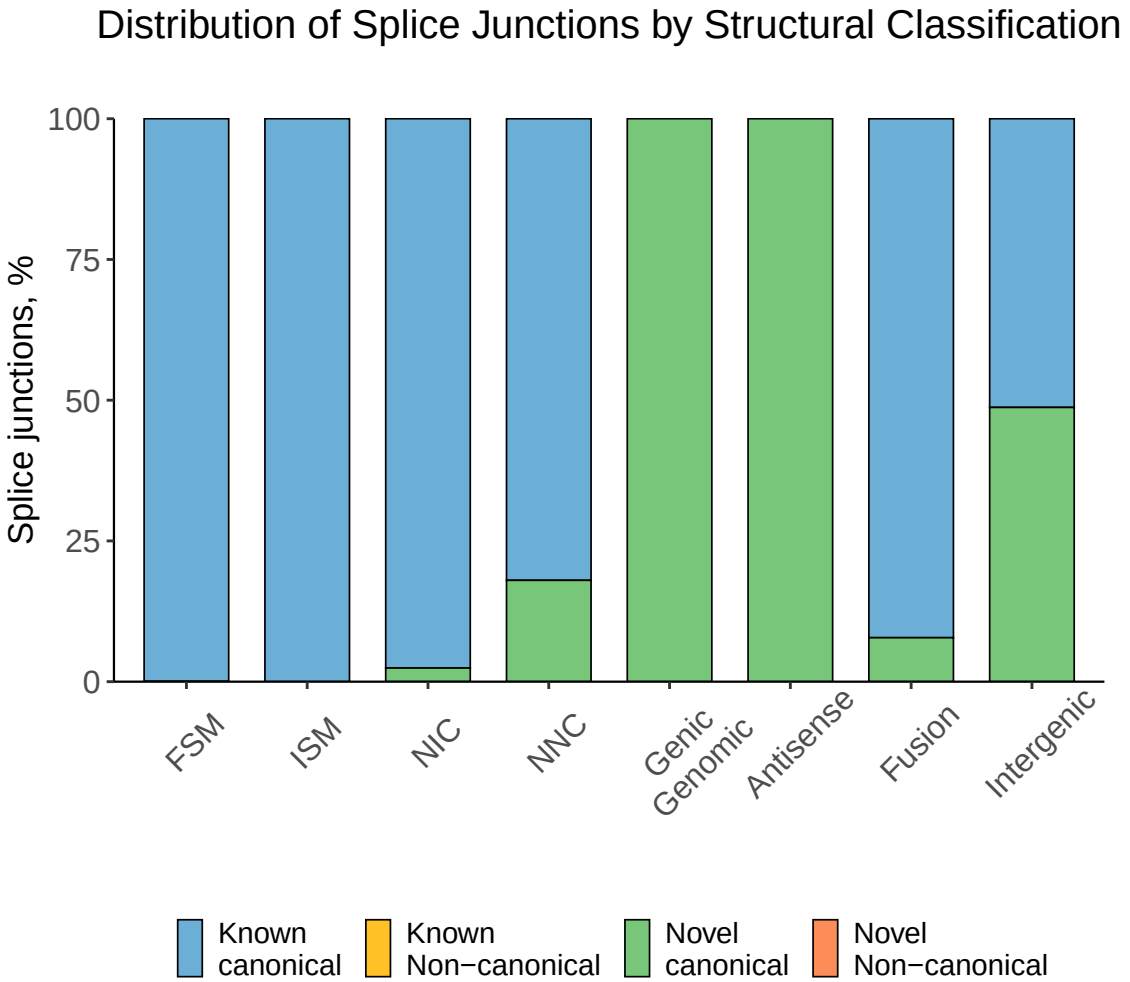

Length Distribution of Matched Reference Transcripts
Applicable Only to FSM and ISM Categories
Matched reference length, kb
125
100
75
50
25
0
FSM
ISM

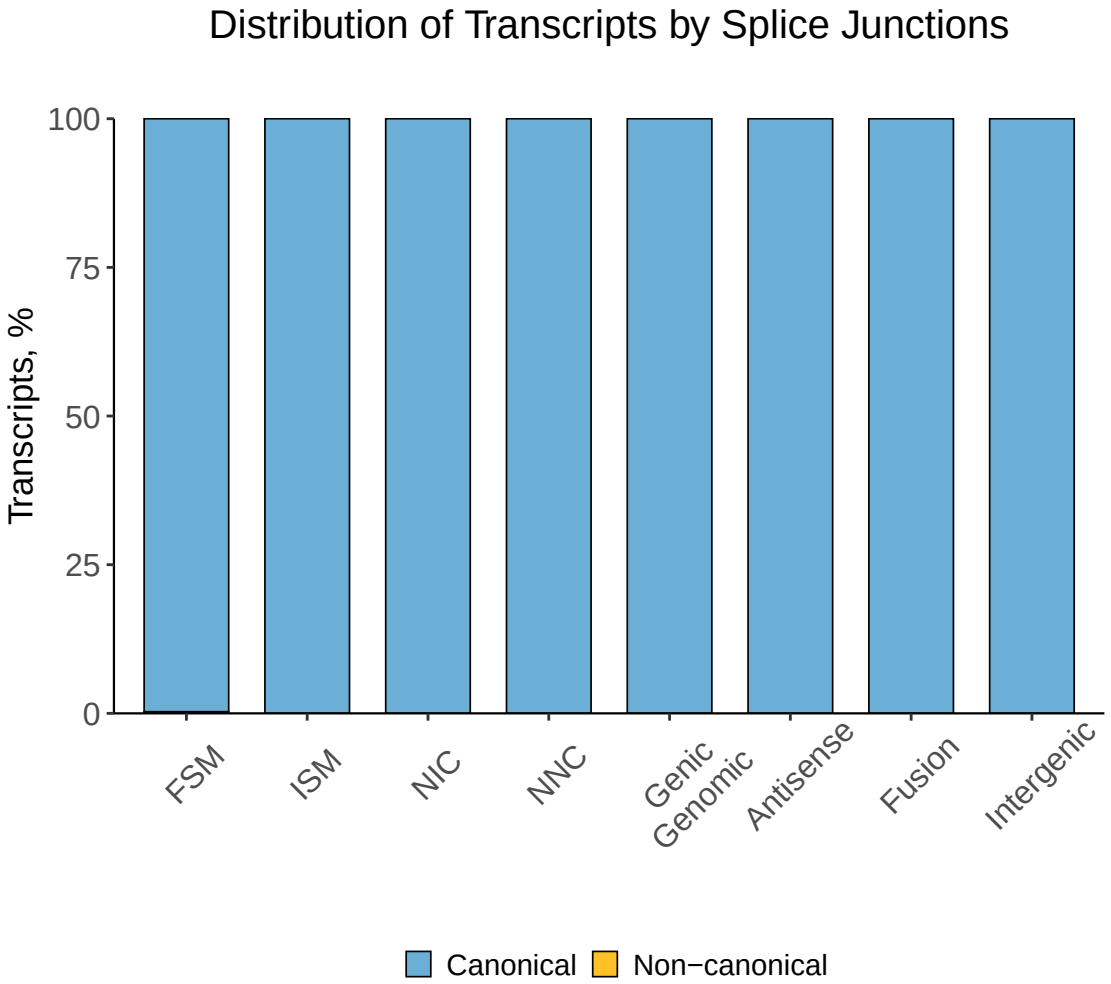
Exon Count Distribution of Matched Reference Transcripts

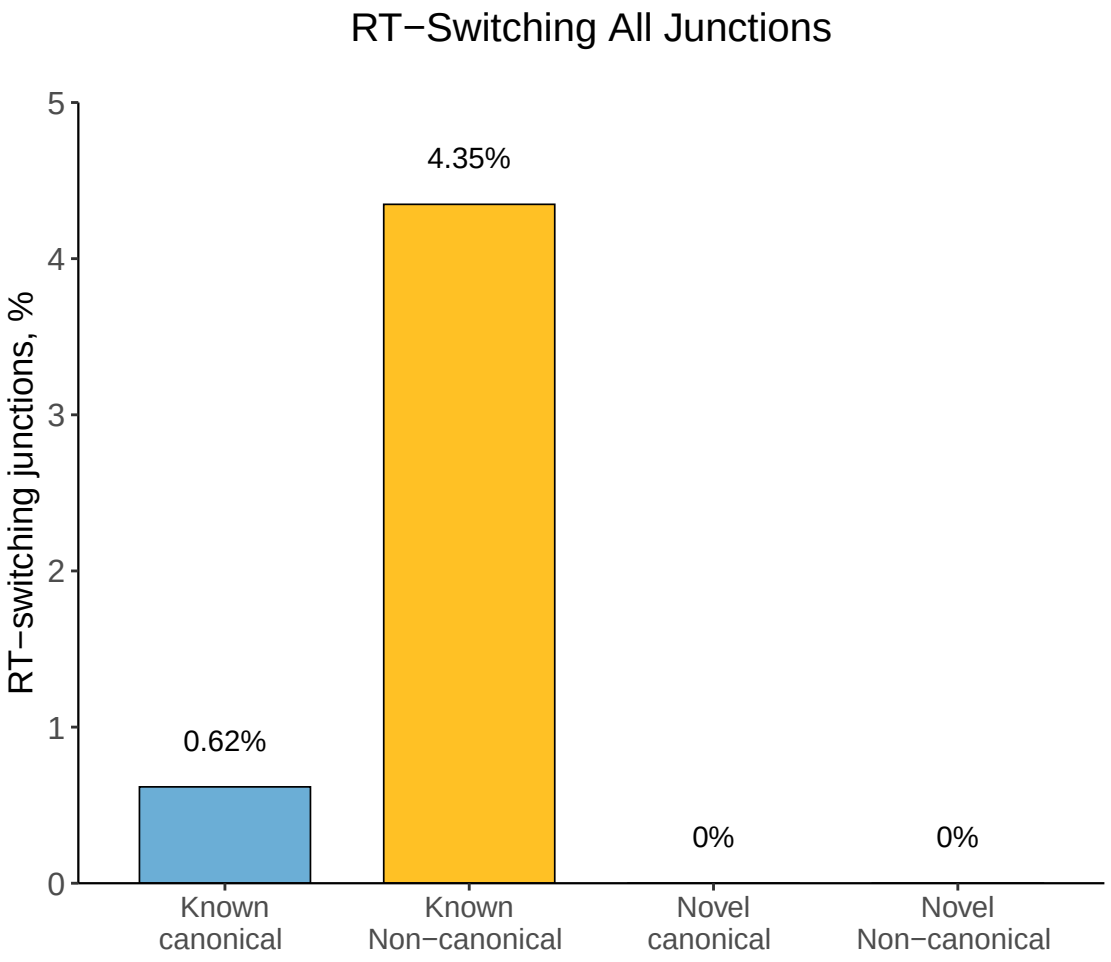
Applicable Only to FSM and ISM Categories

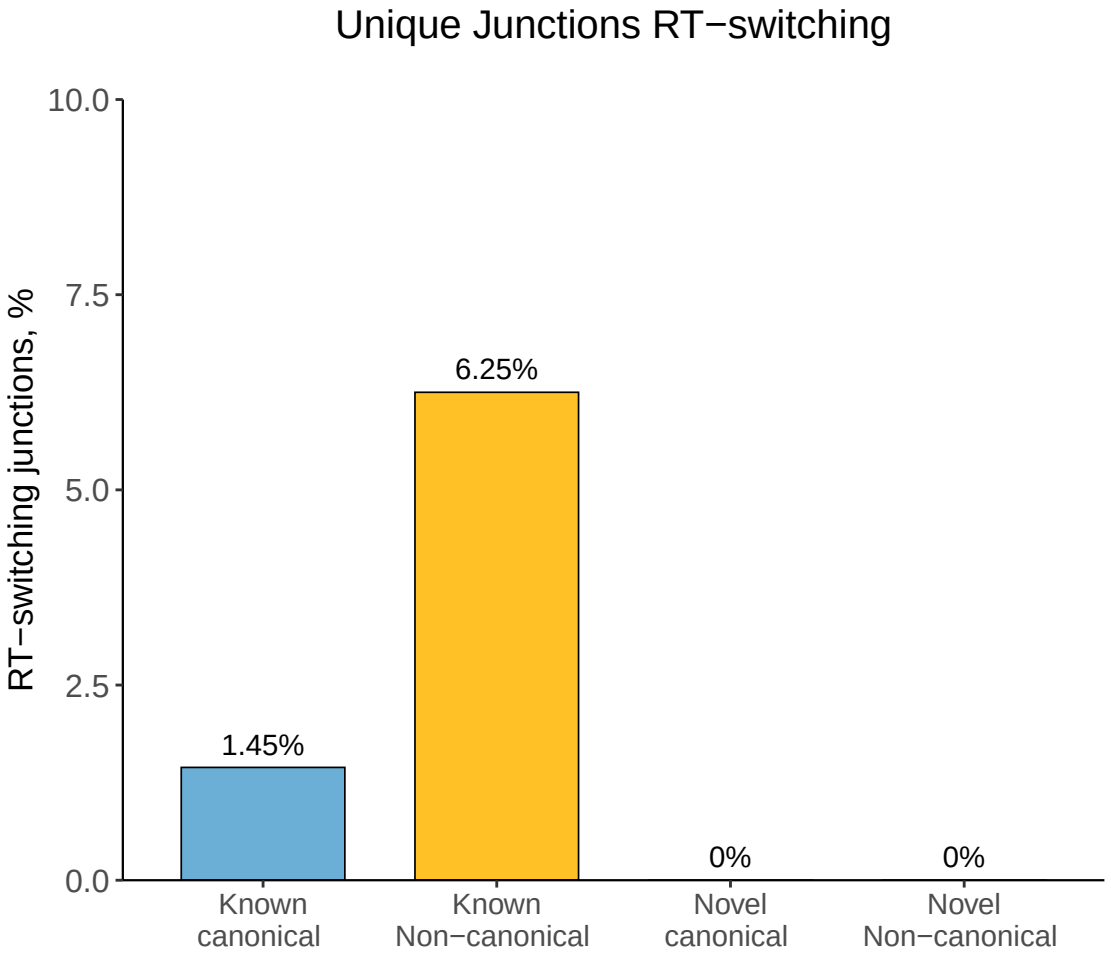


Splice Junction Characterization

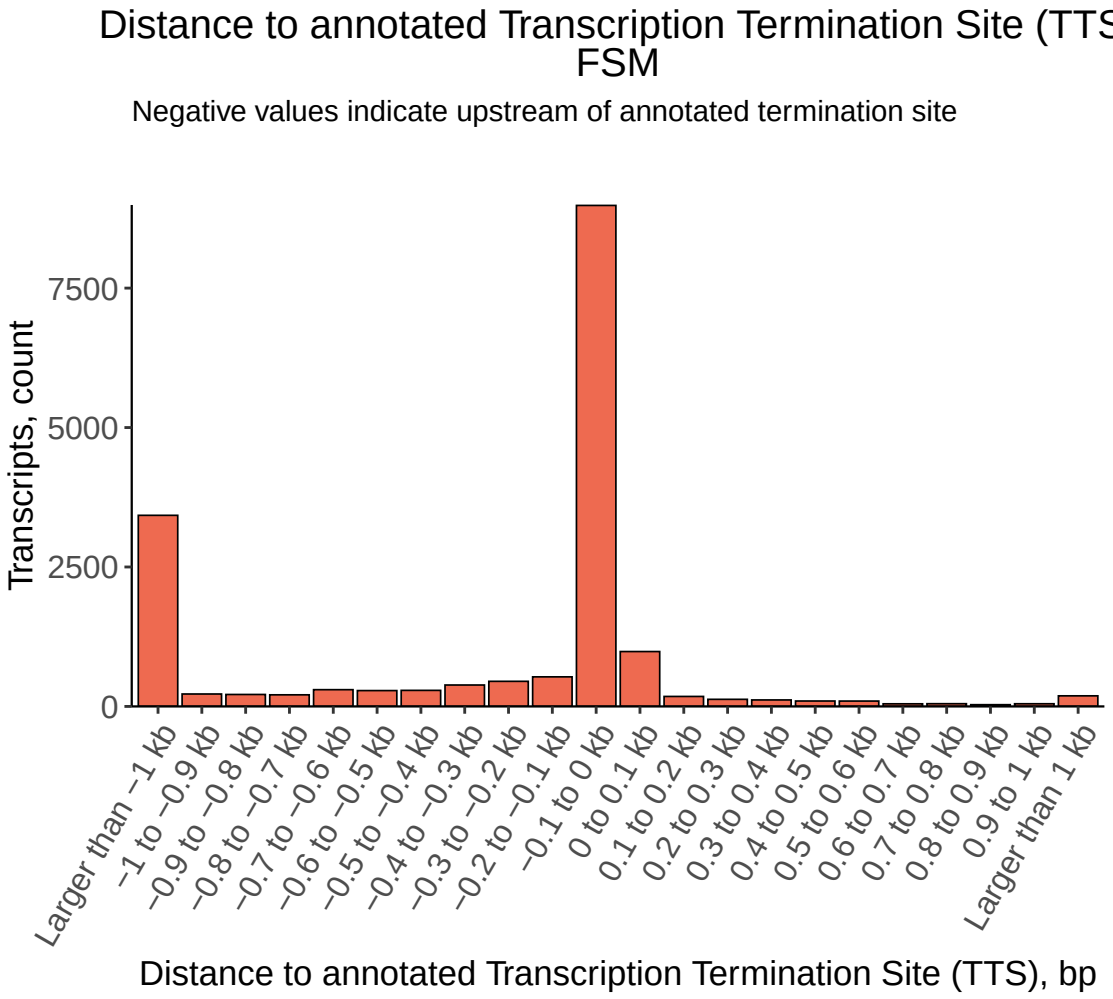

Distribution of Splice Junctions by Structural Classification
Splice junctions, %
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

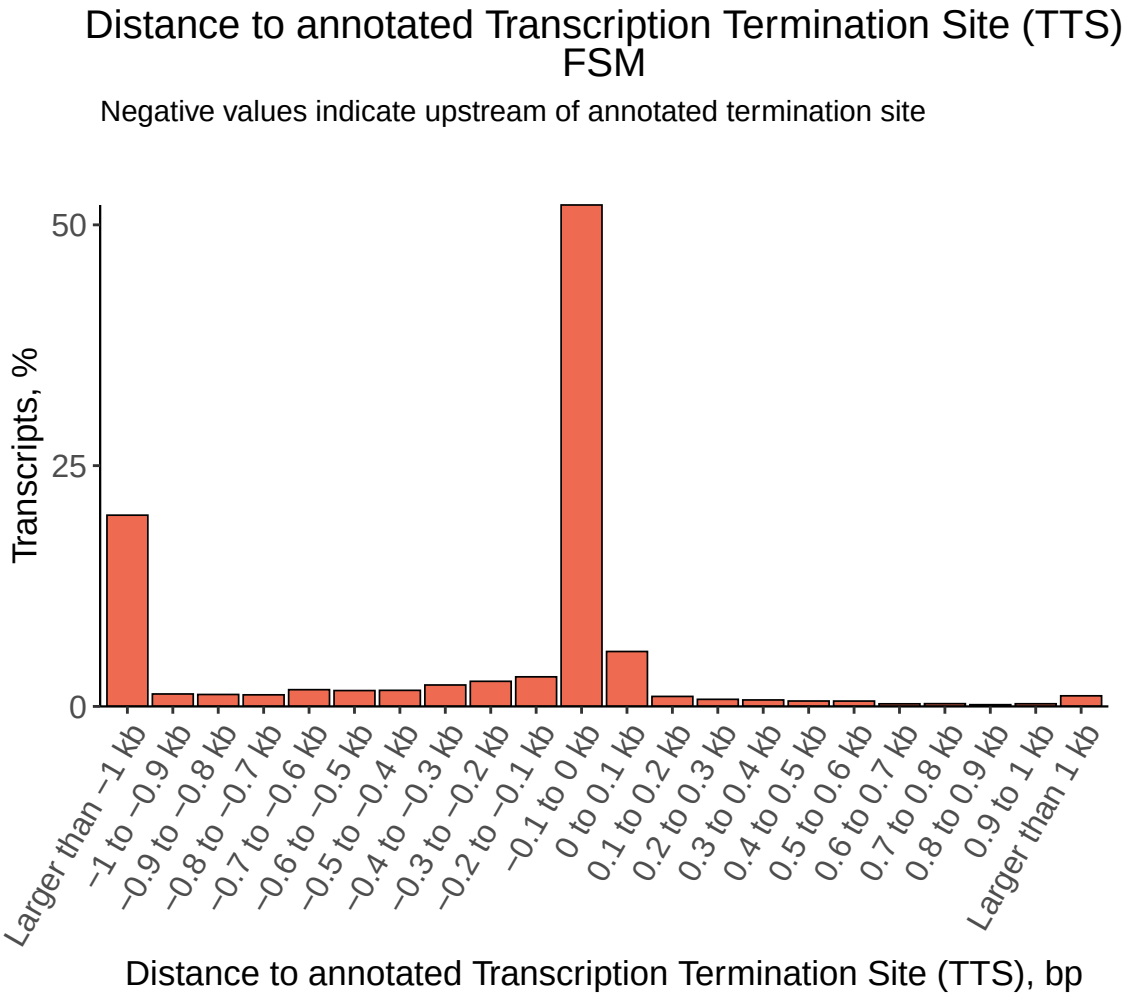

Distribution of Transcripts by Splice Junctions
Transcripts, %
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Canonical
Non−canonical

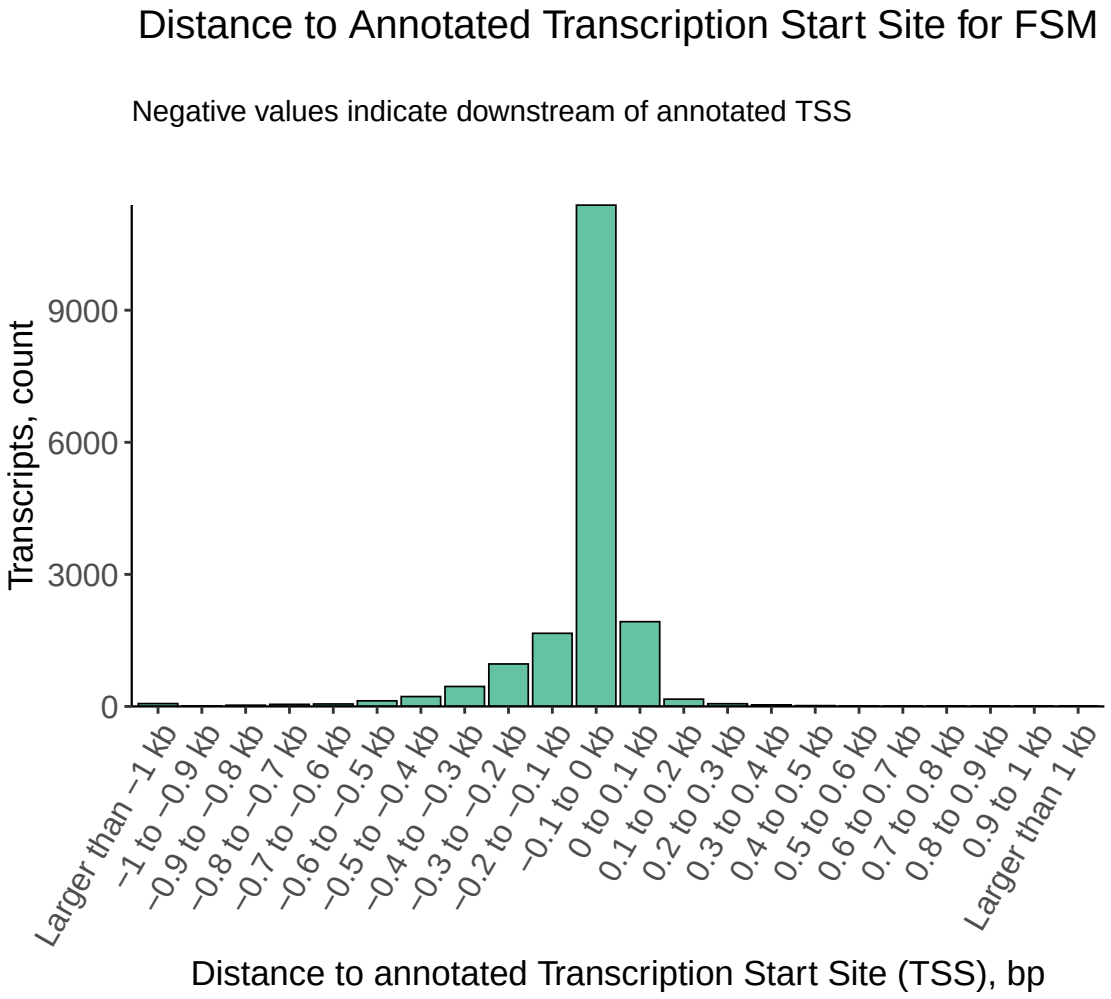

RT−Switching All Junctions
RT−switching junctions, %
5
4
3
2
1
0
0.62%
4.35%
0%
0%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

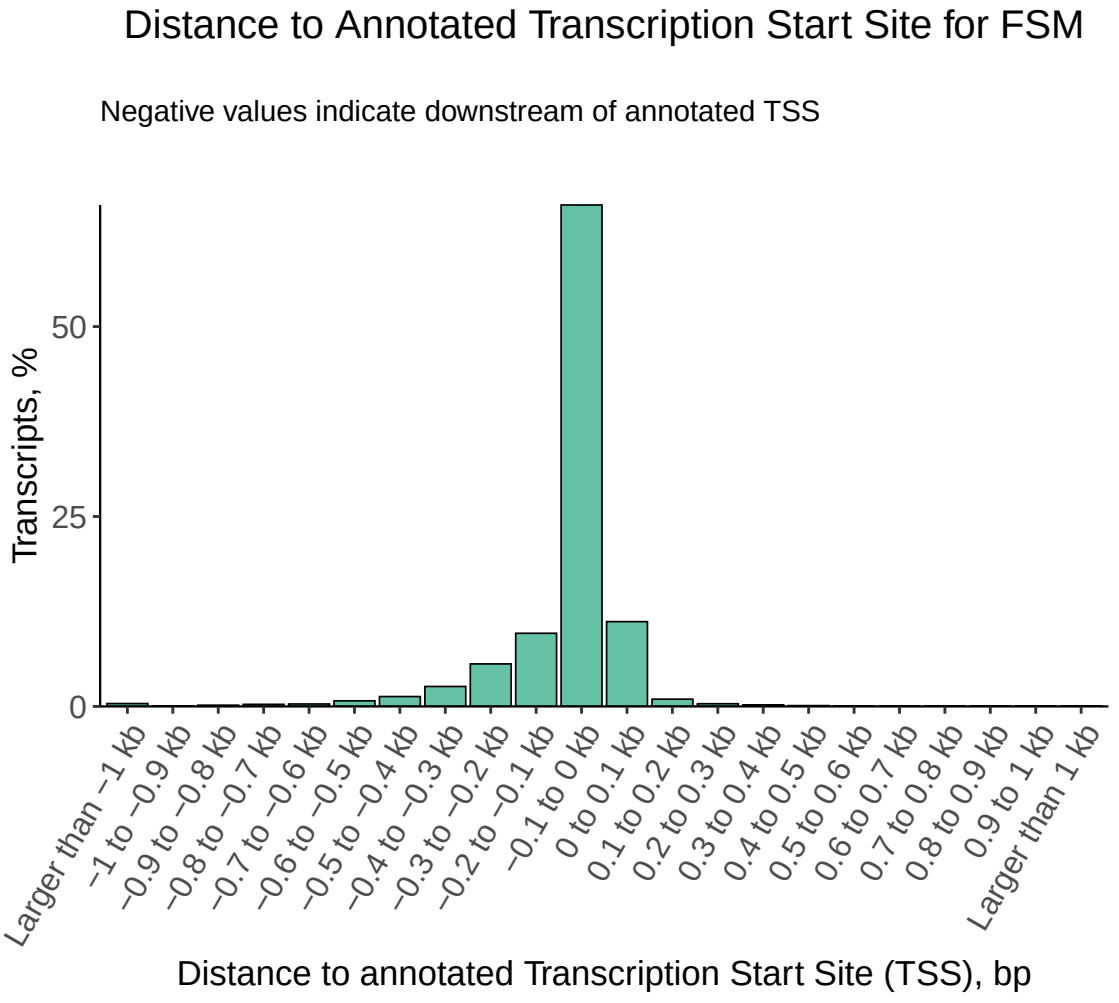

Unique Junctions RT−switching
RT−switching junctions, %
10.0
7.5
5.0
2.5
0.0
1.45%
6.25%
0%
0%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

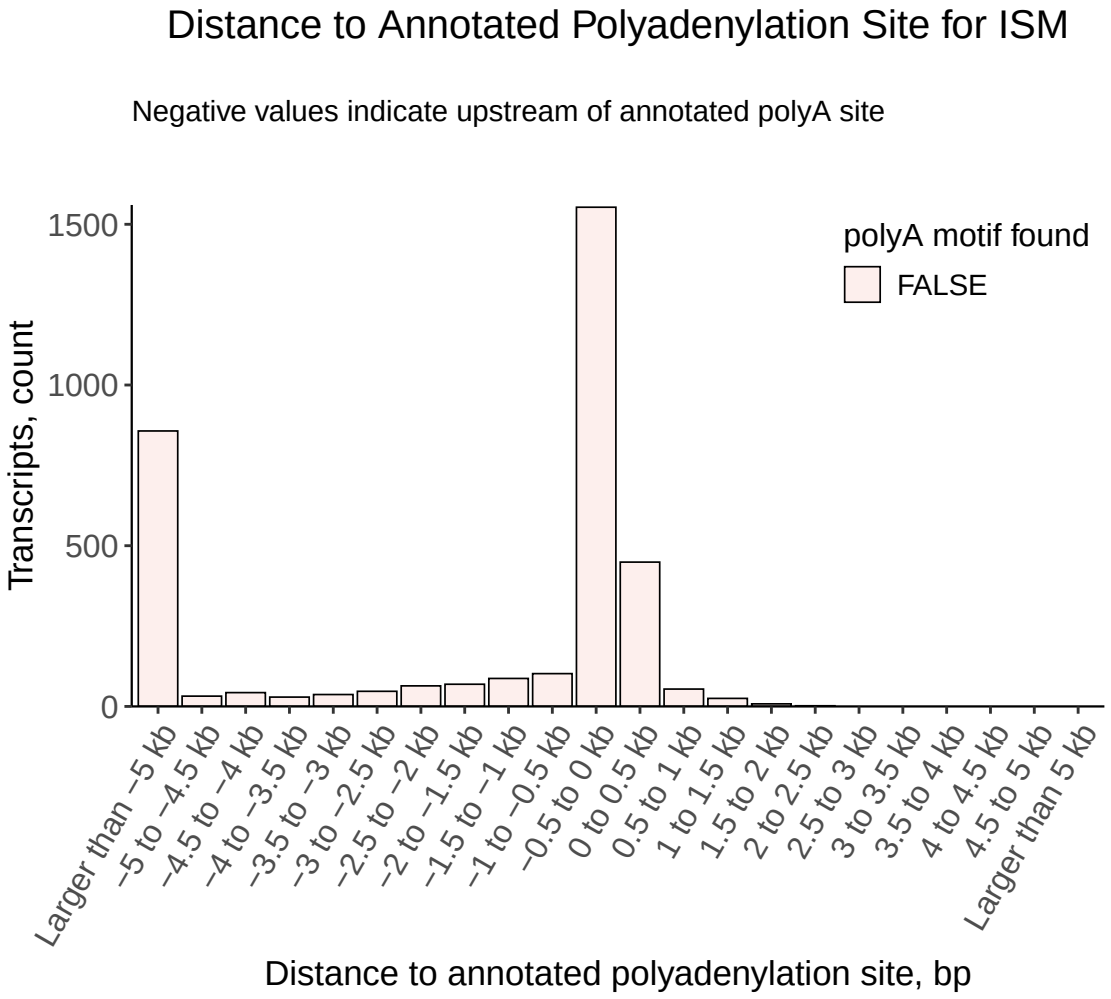
Comparison With Annotated TSS and TTS

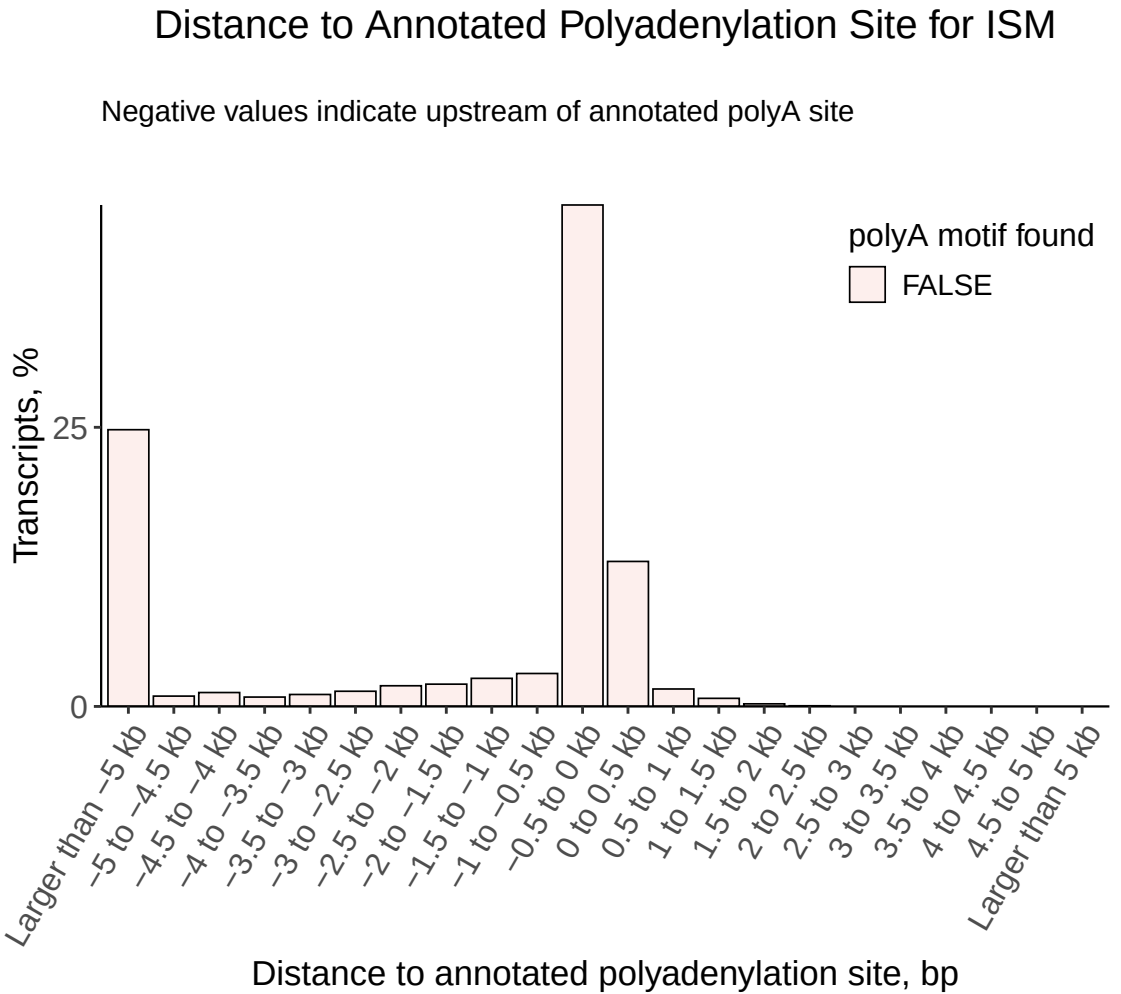

Distance to annotated Transcription Termination Site (TTS)
FSM
Negative values indicate upstream of annotated termination site
Transcripts, count
7500
5000
2500
0
Larger than −1 kb
−1 to −0.9 kb
−0.9 to −0.8 kb
−0.8 to −0.7 kb
−0.7 to −0.6 kb
−0.6 to −0.5 kb
−0.5 to −0.4 kb
−0.4 to −0.3 kb
−0.3 to −0.2 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

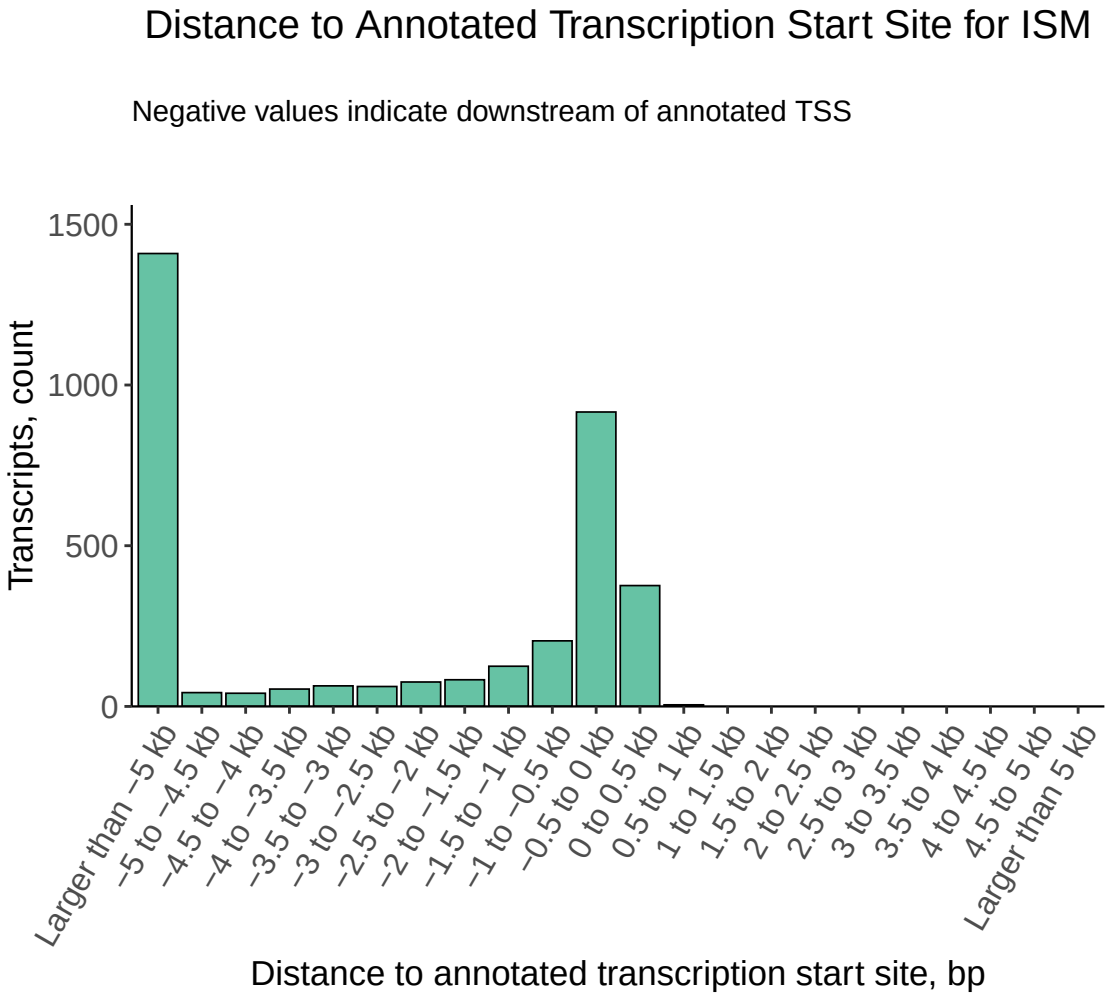

Distance to annotated Transcription Termination Site (TTS)
FSM
Negative values indicate upstream of annotated termination site
Transcripts, %
50
25
0
Larger than −1 kb
−1 to −0.9 kb
−0.9 to −0.8 kb
−0.8 to −0.7 kb
−0.7 to −0.6 kb
−0.6 to −0.5 kb
−0.5 to −0.4 kb
−0.4 to −0.3 kb
−0.3 to −0.2 kb
−0.2 to −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

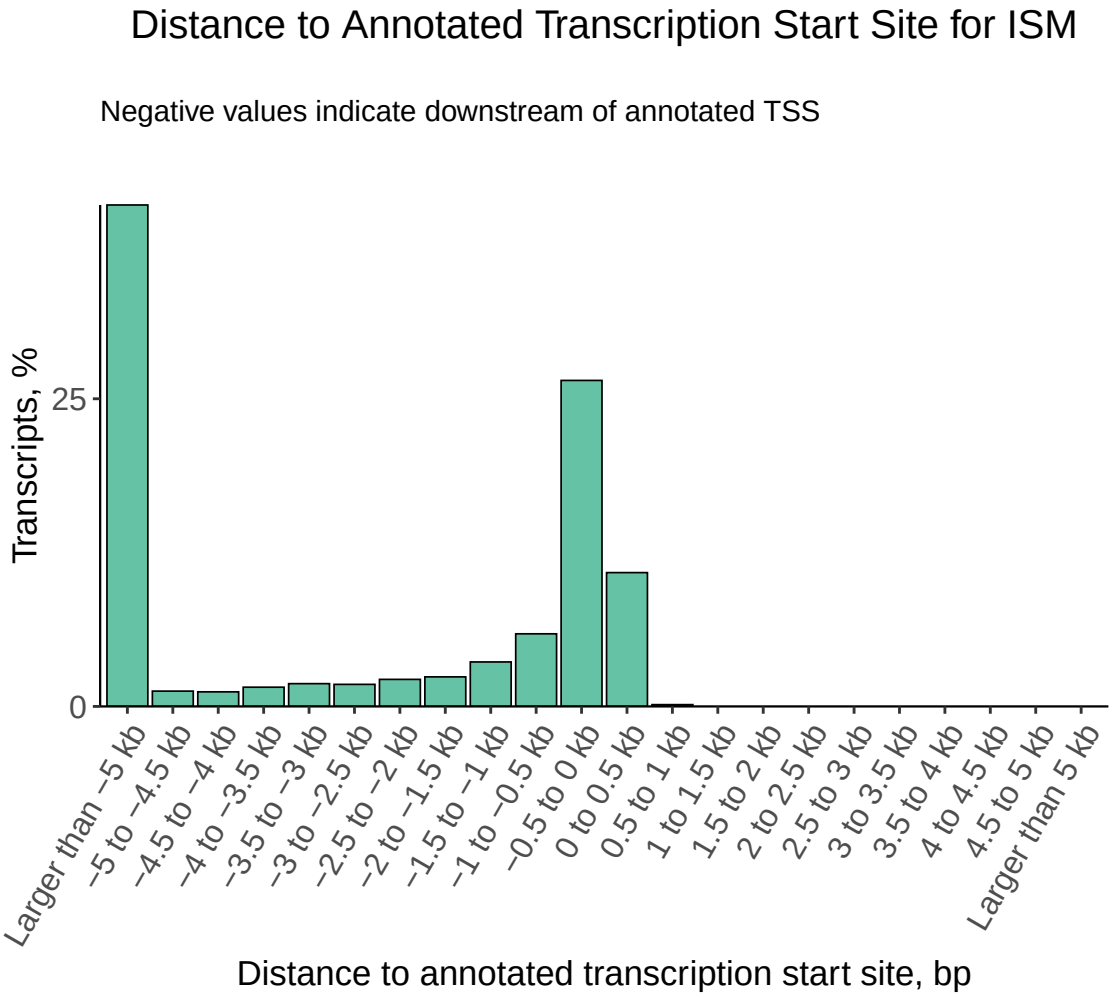

Distance to Annotated Transcription Start Site for FSM
Negative values indicate downstream of annotated TSS
Transcripts, count
9000
6000
3000
0
Larger than −1 kb
−1 to −0.9 kb
−0.9 to −0.8 kb
−0.8 to −0.7 kb
−0.7 to −0.6 kb
−0.6 to −0.5 kb
−0.5 to −0.4 kb
−0.4 to −0.3 kb
−0.3 to −0.2 kb
−0.2 to −0.1 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 Transcription Start Site for FSM
Negative values indicate downstream of annotated TSS
Transcripts, %
50
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
1500
1000
500
0
polyA motif found
FALSE
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
polyA motif found
FALSE
Transcripts, %
25
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

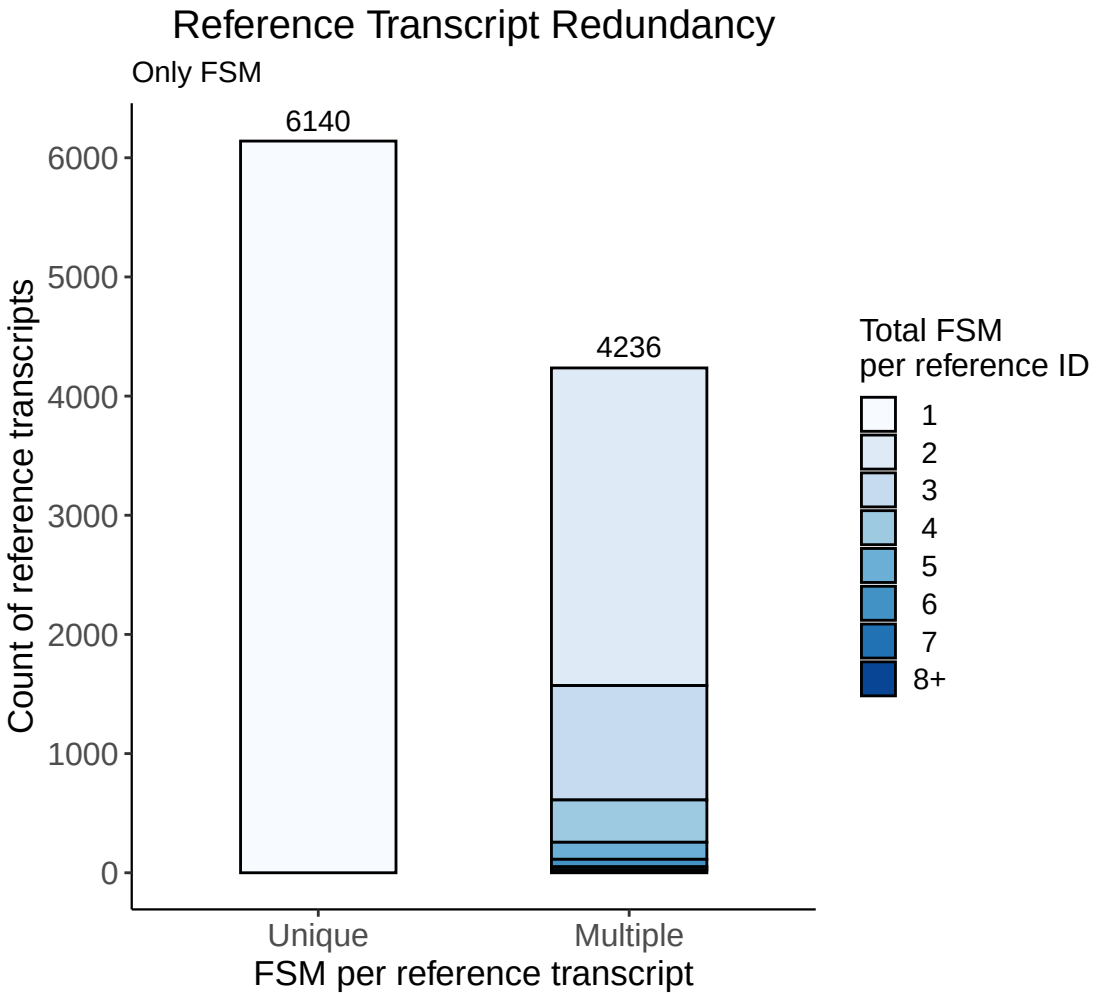

Distance to Annotated Transcription Start Site for ISM
Negative values indicate downstream of annotated TSS
Transcripts, count
1500
1000
500
0
Larger than −5 kb
−5 to −4.5 kb
−4.5 to −4 kb
−4 to −3.5 kb
−3.5 to −3 kb
−3 to −2.5 kb
−2.5 to −2 kb
−2 to −1.5 kb
−1.5 to −1 kb
−1 to −0.5 kb
−0.5 to 0 kb
0 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 transcription start site, bp

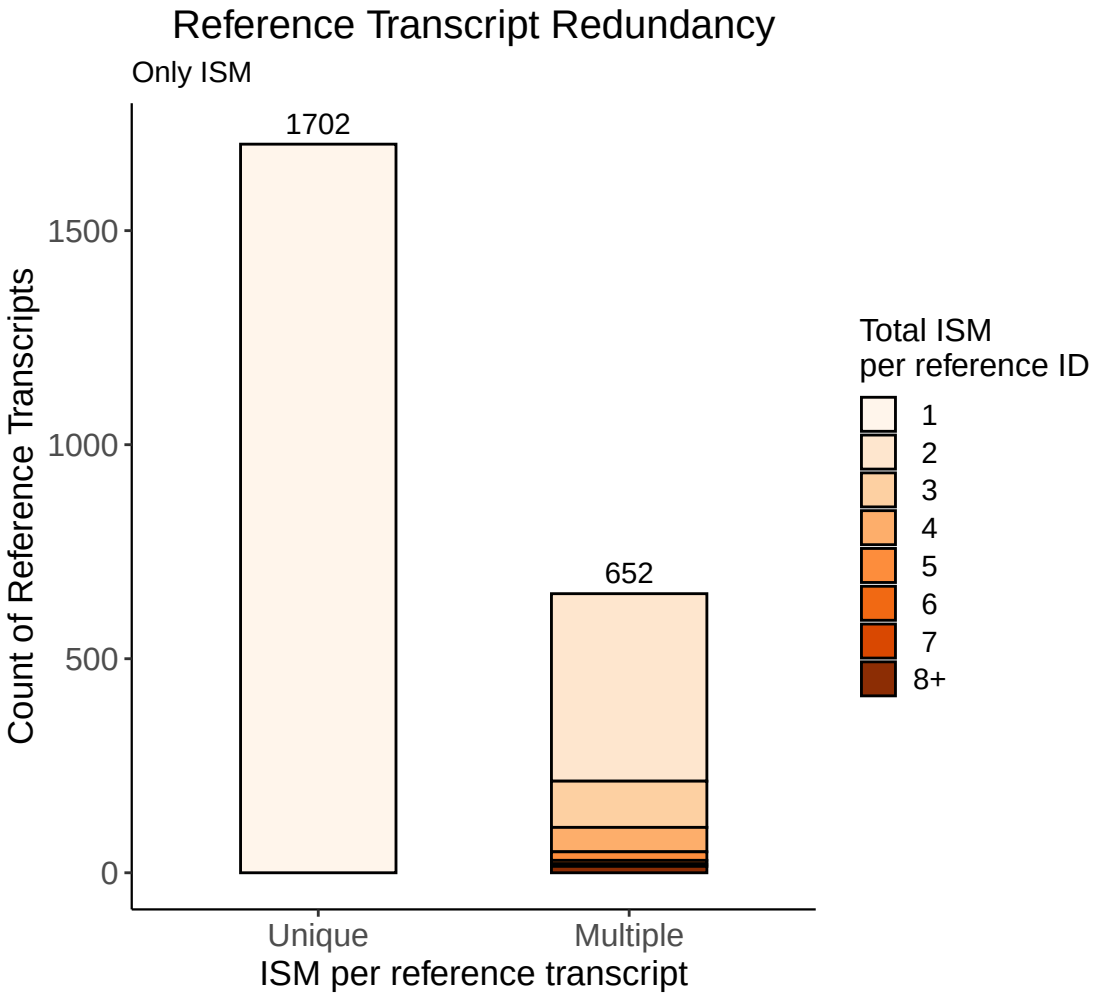

Distance to Annotated Transcription Start Site for ISM
Negative values indicate downstream of annotated TSS
Transcripts, %
25
0
Larger than −5 kb
−5 to −4.5 kb
−4.5 to −4 kb
−4 to −3.5 kb
−3.5 to −3 kb
−3 to −2.5 kb
−2.5 to −2 kb
−2 to −1.5 kb
−1.5 to −1 kb
−1 to −0.5 kb
−0.5 to 0 kb
0 to 0.5 kb
0.5 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 transcription start site, bp

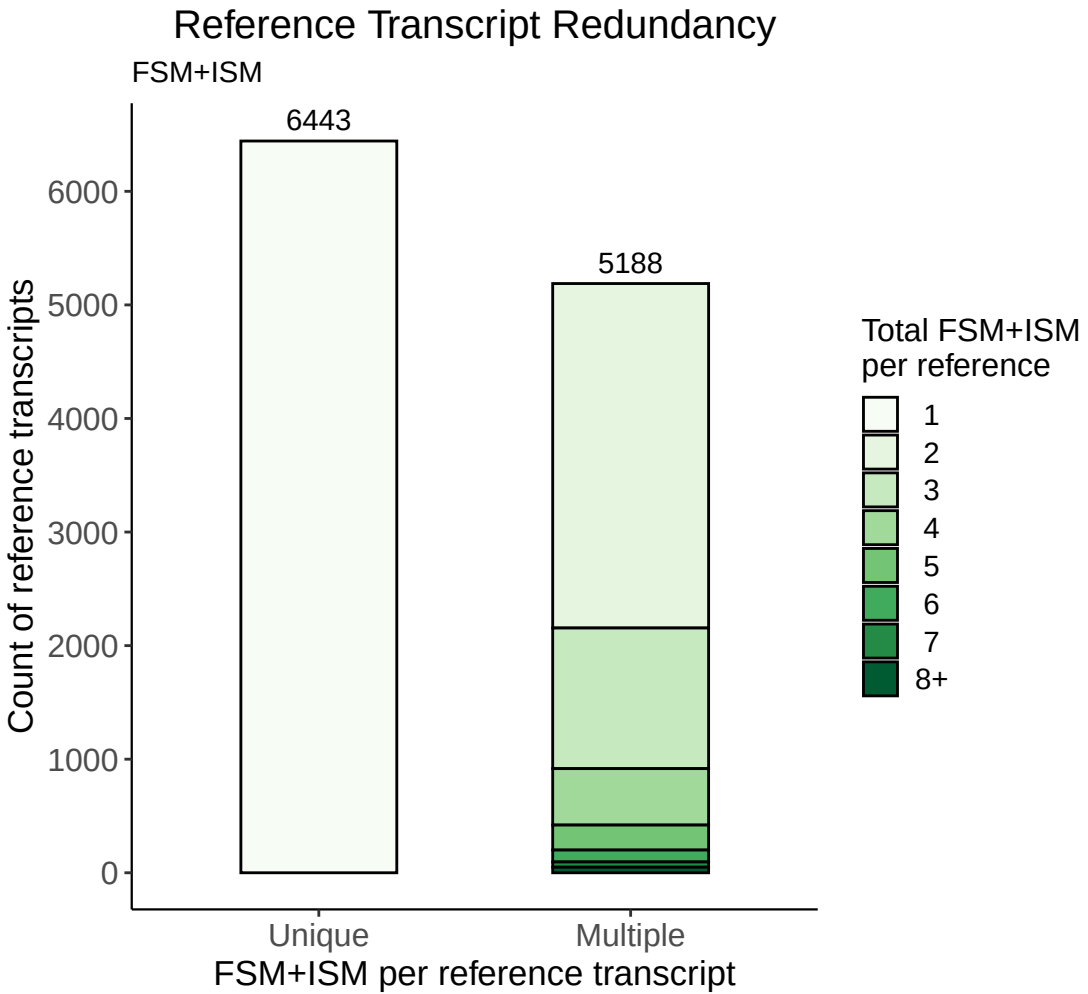
Comparison With Annotated TSS and TTS by Subcategories

PolyA Distance Analysis

Redundancy Analysis

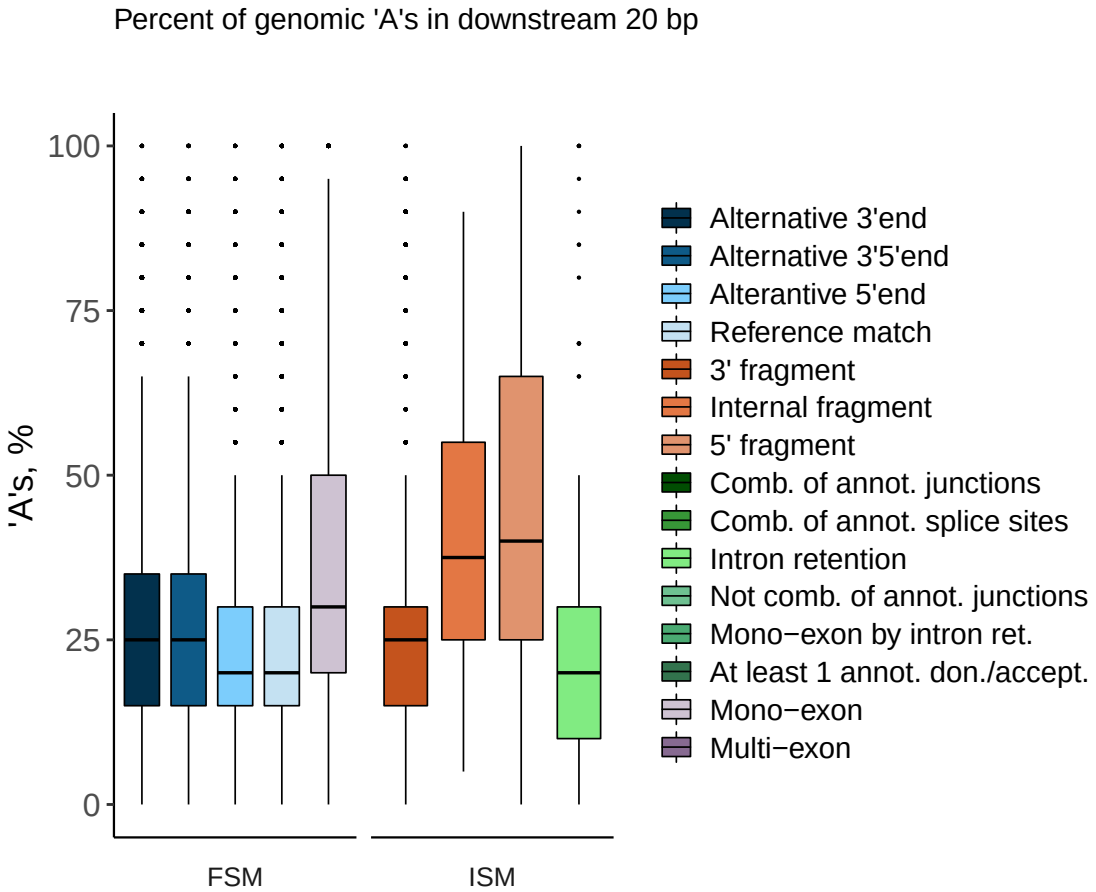

Reference Transcript Redundancy
Only FSM
6140
4236
6000
5000
4000
3000
2000
1000
0
Count of reference transcripts
Unique
Multiple
FSM per reference transcript
Total FSM
per reference ID
1
2
3
4
5
6
7
8+


Reference Transcript Redundancy
Only ISM
1702
652
1500
1000
500
0
Count of Reference Transcripts
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
5
6
7
8+

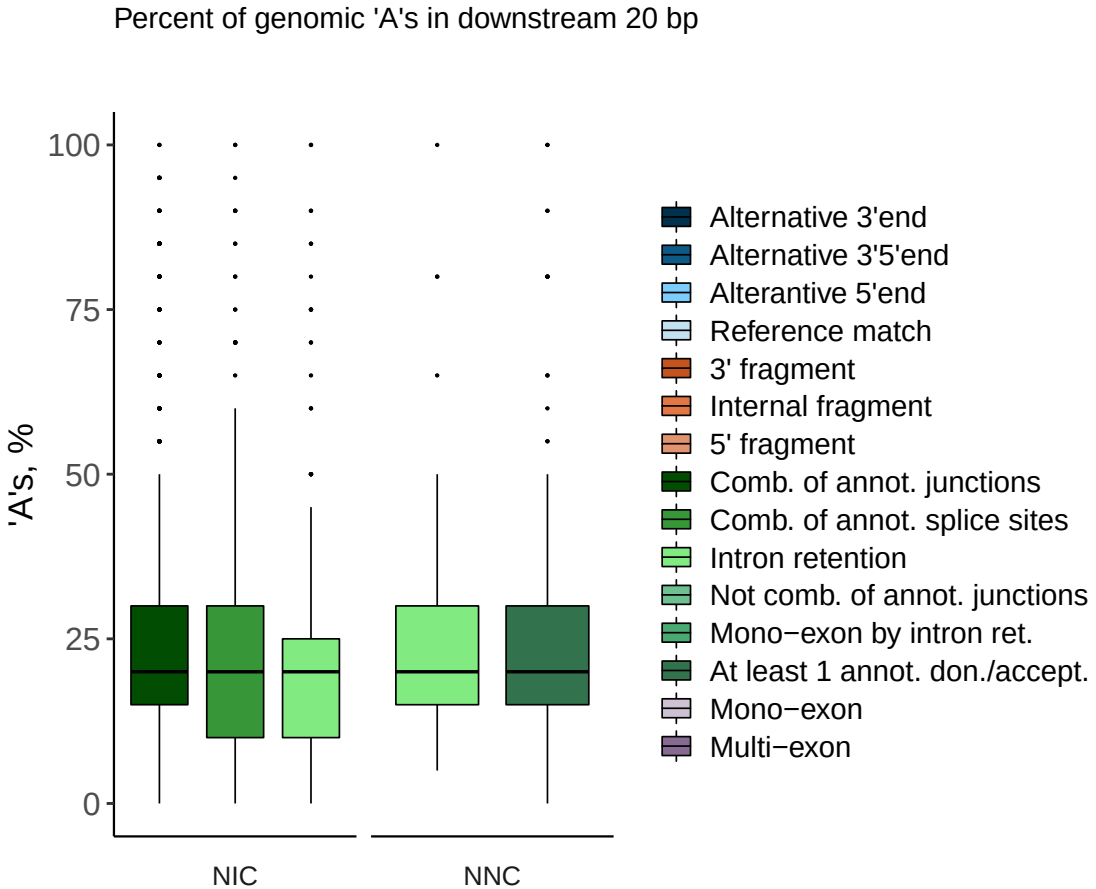

Reference Transcript Redundancy
FSM+ISM
6443
5188
6000
5000
4000
3000
2000
1000
0
Count of reference transcripts
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference
1
2
3
4
5
6
7
8+

Intra−Priming Quality Check

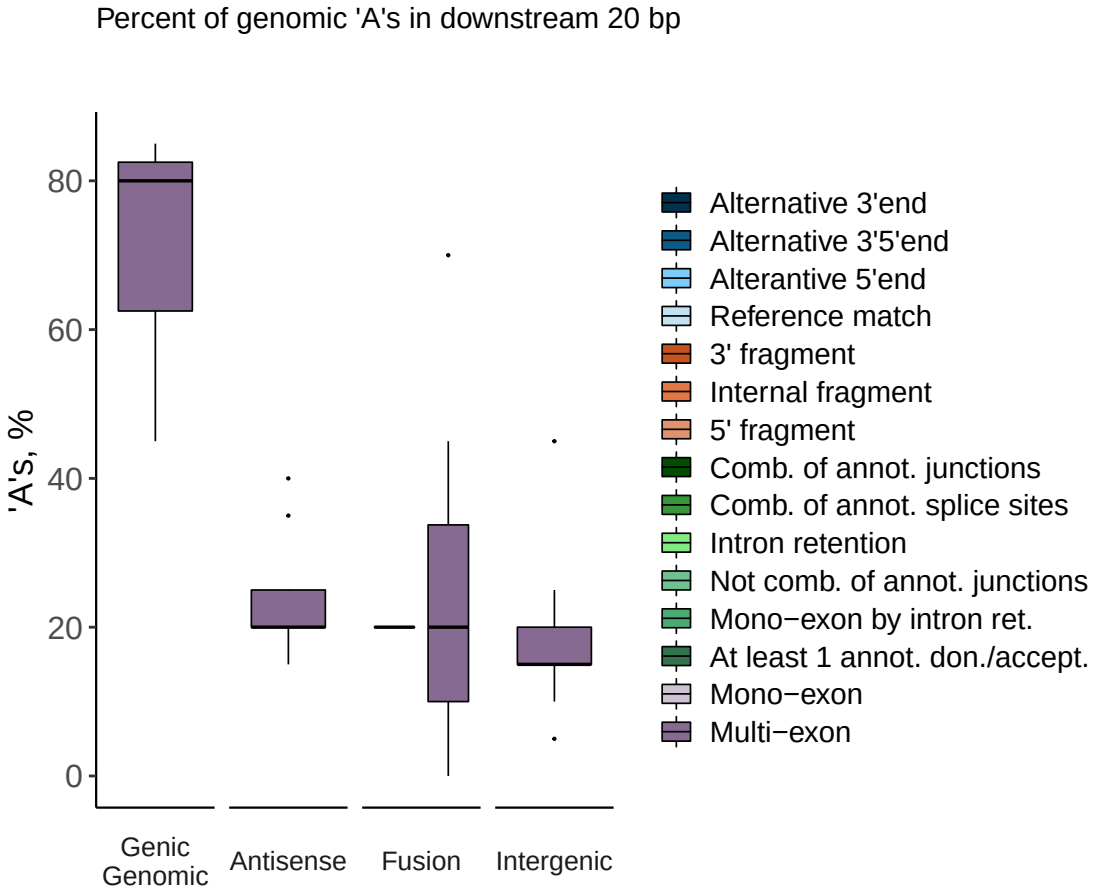
Possible Intra−Priming by Structural Category

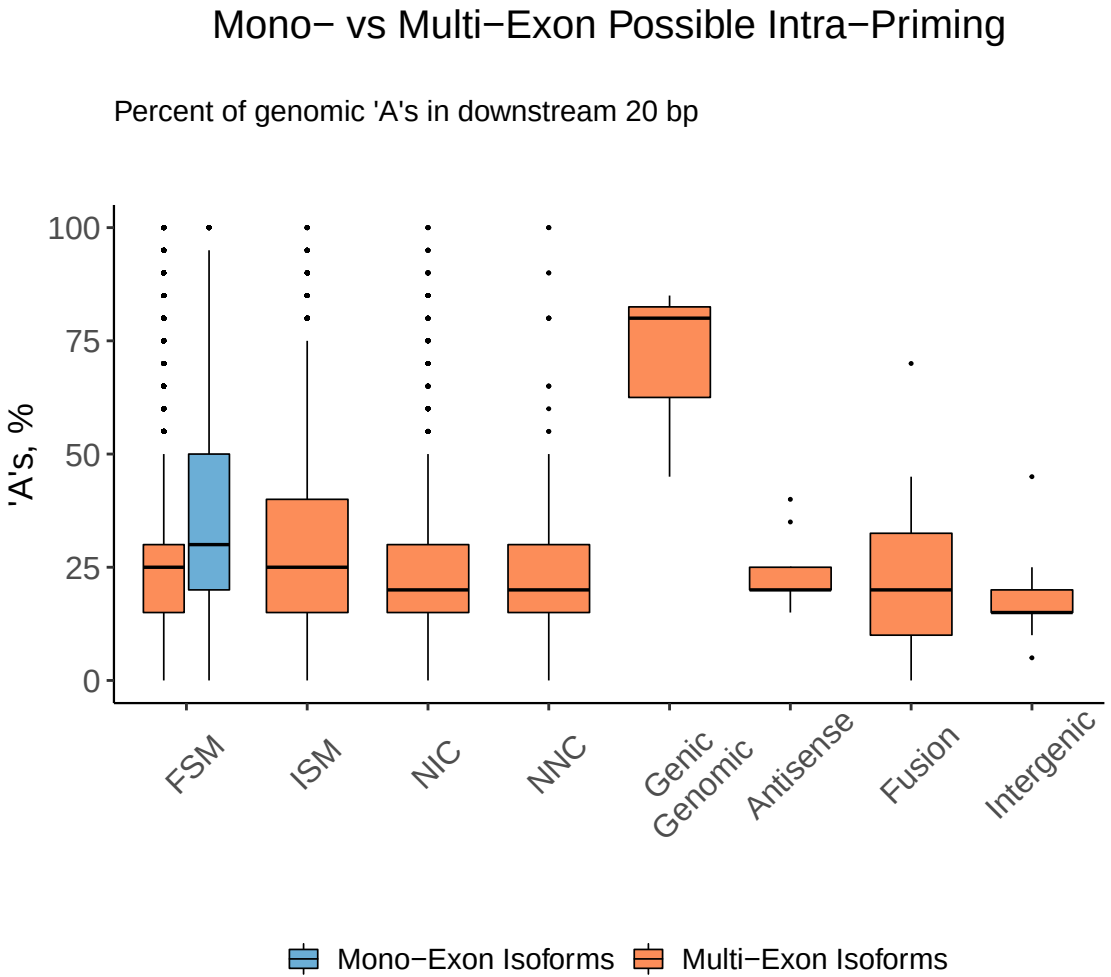


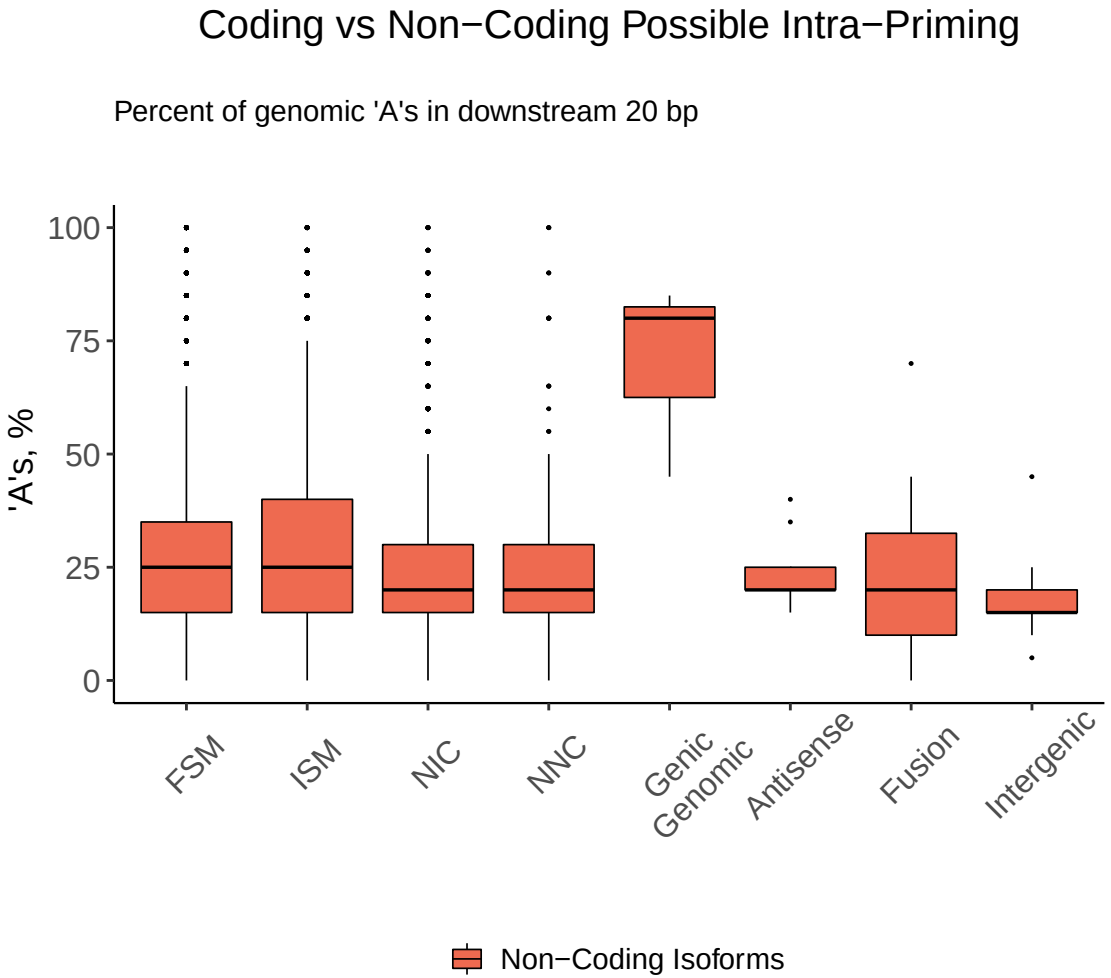
Possible Intra−Priming by Structural Category



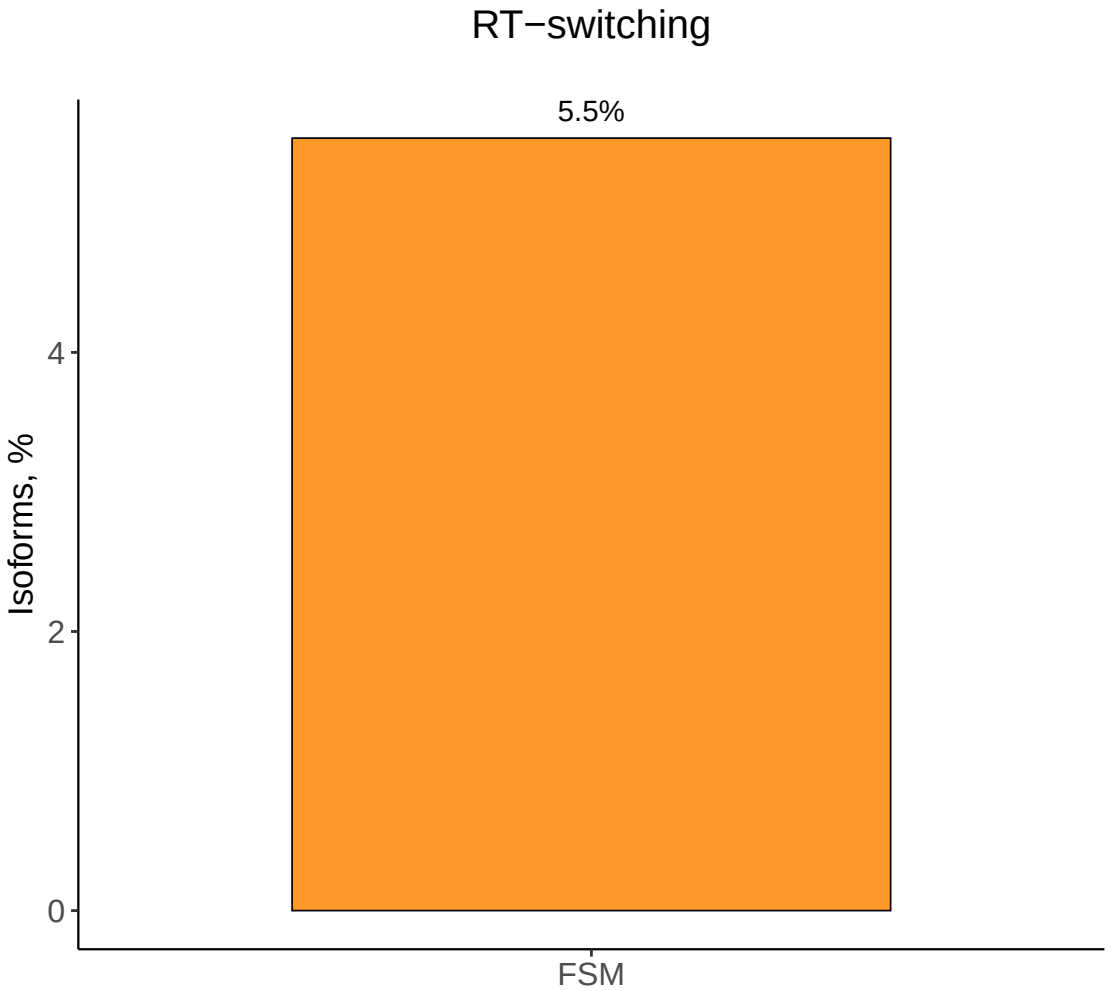
Possible Intra−Priming by Structural Category

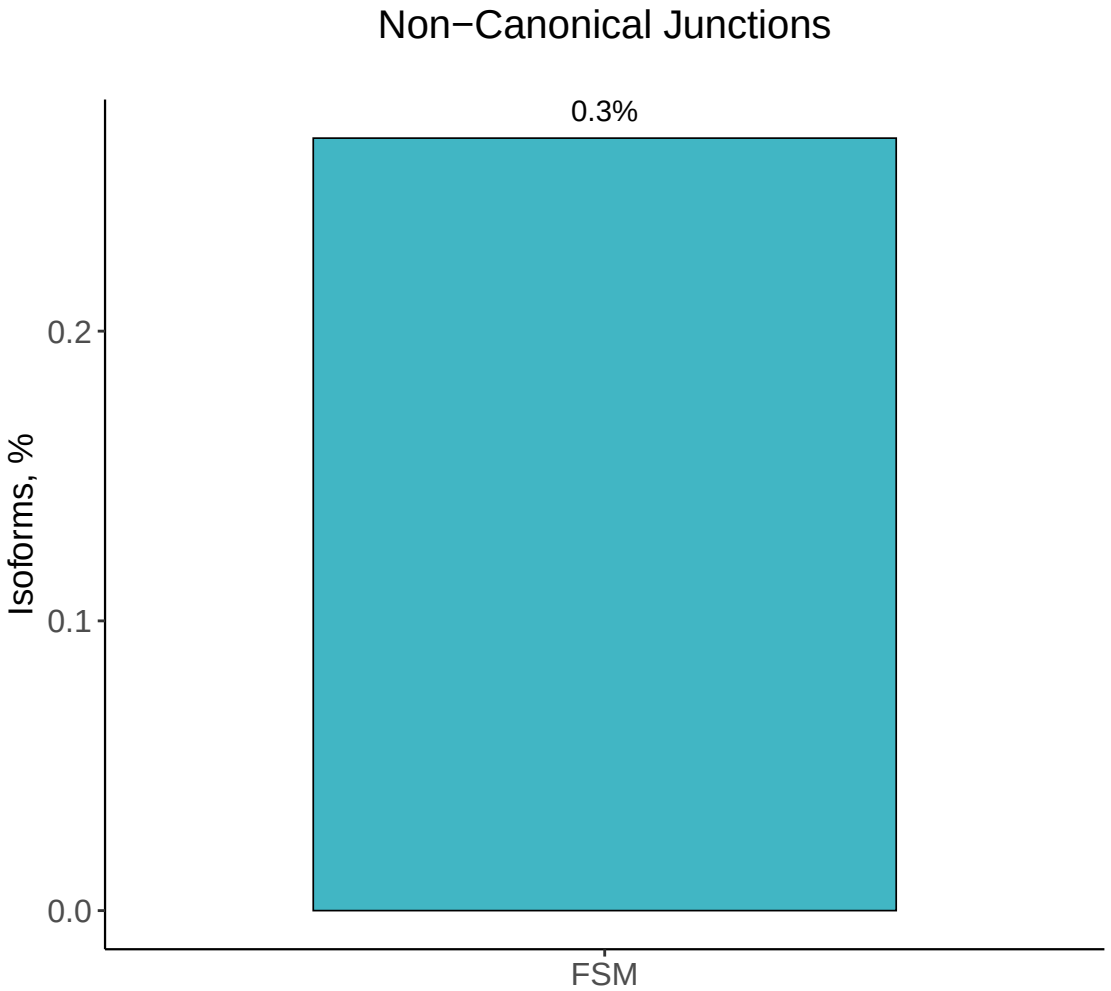


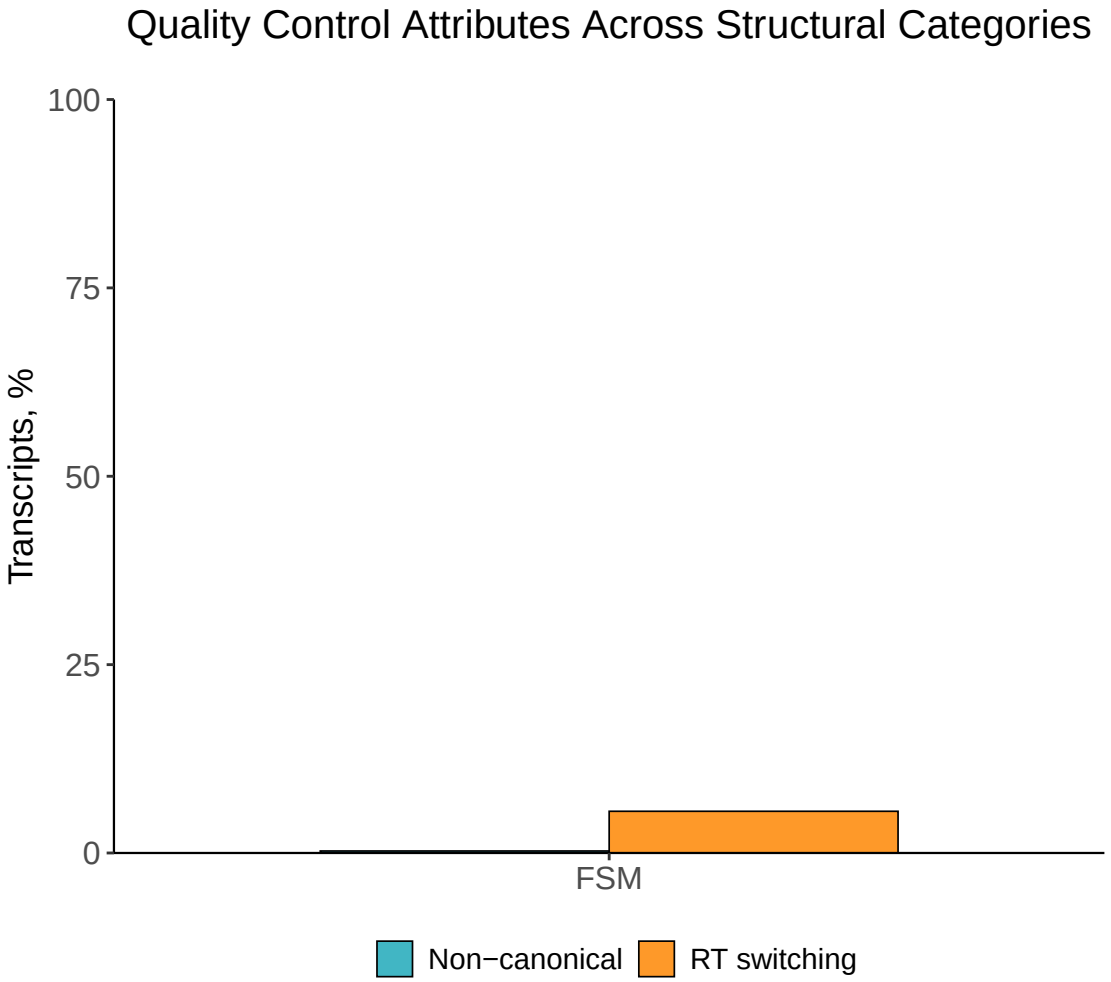

Mono− vs Multi−Exon Possible Intra−Priming
Percent of genomic 'A's in downstream 20 bp
'A's, %
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Mono−Exon Isoforms
Multi−Exon Isoforms


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
Non−Coding Isoforms

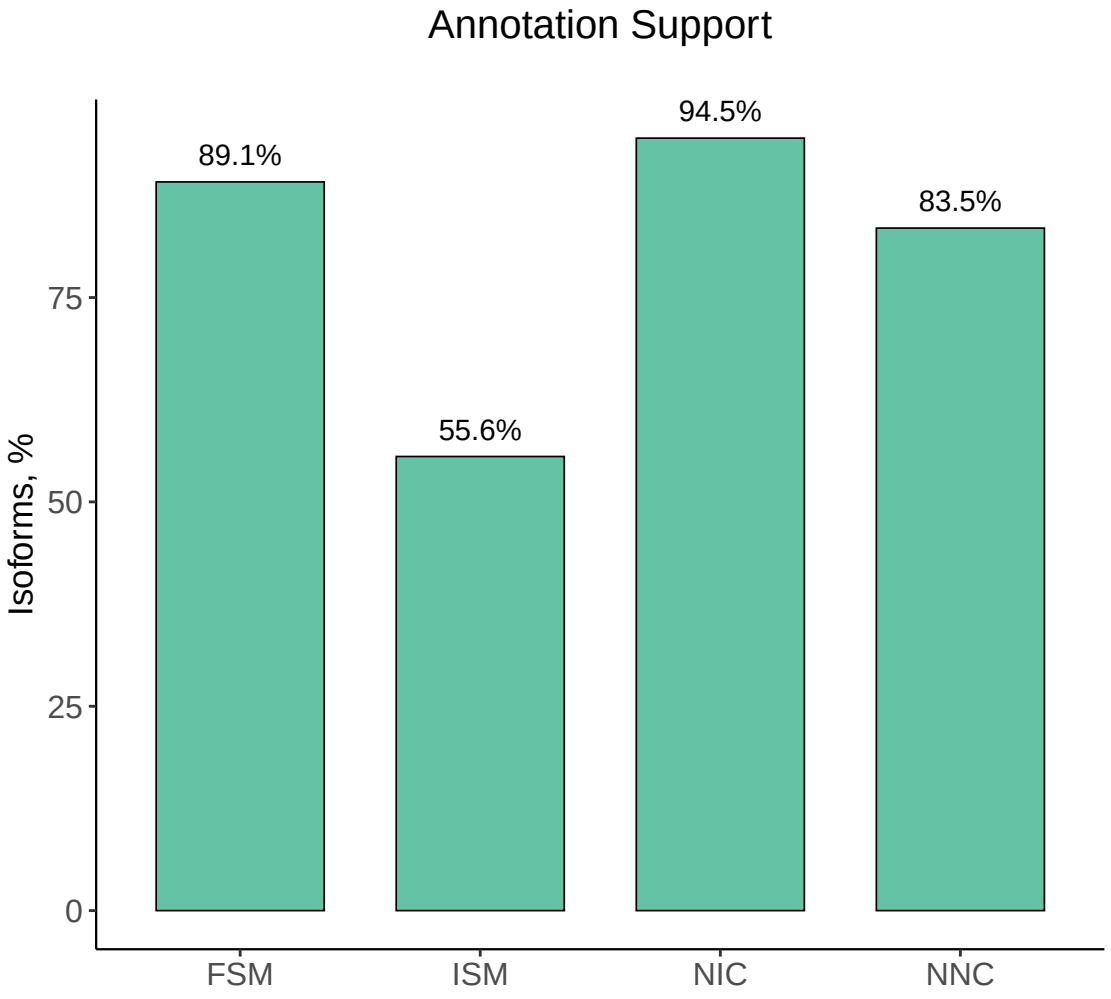
Features of Bad Quality

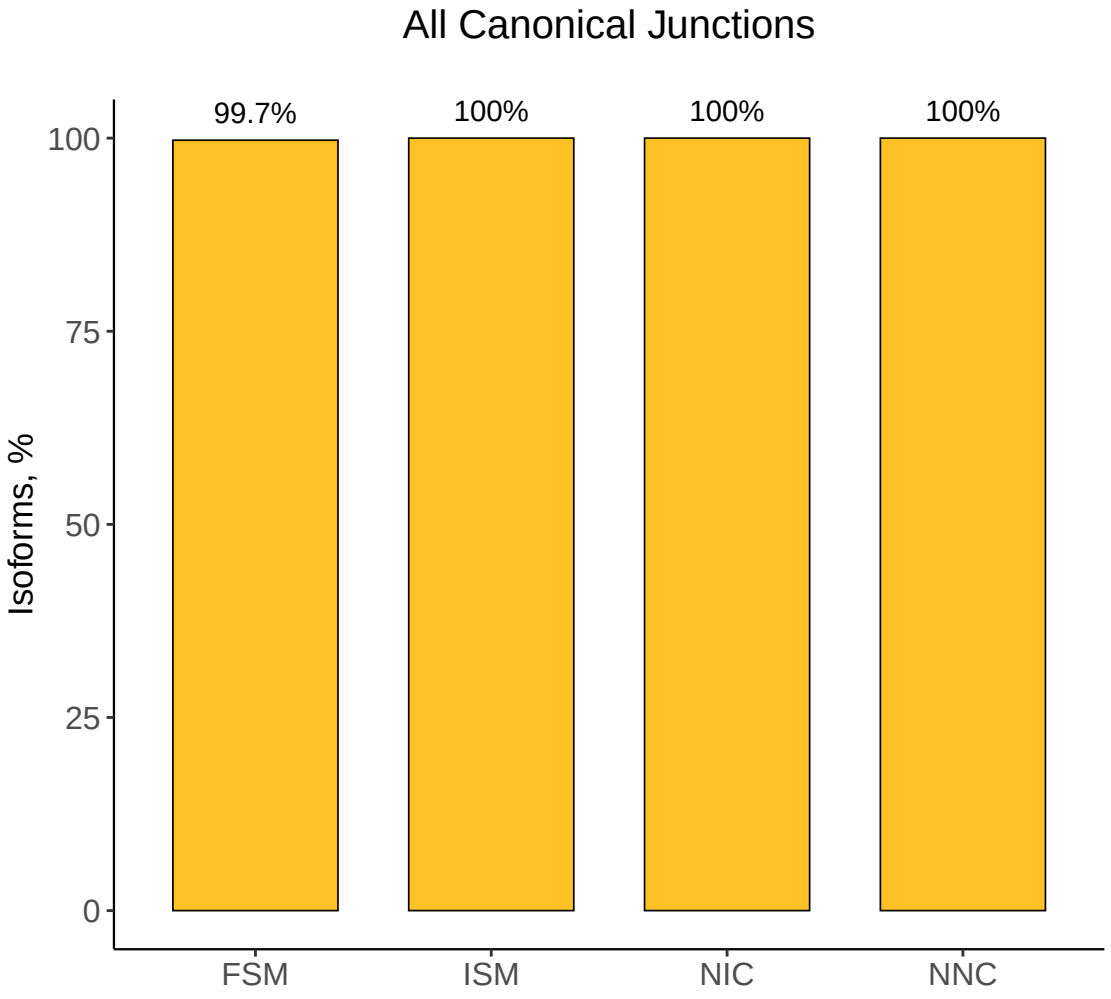

RT−switching
5.5%
Isoforms, %
4
2
0
FSM

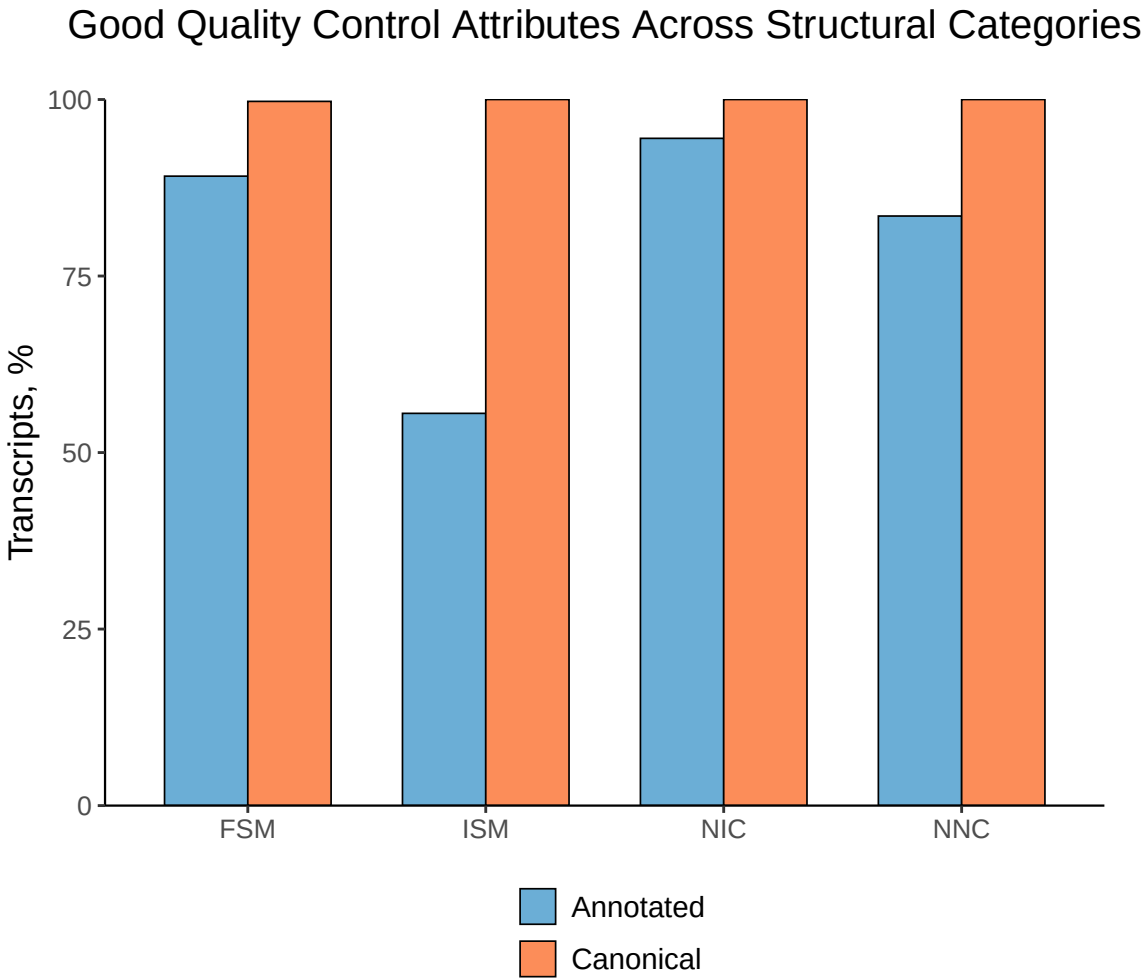

Non−Canonical Junctions
0.3%
Isoforms, %
0.2
0.1
0.0
FSM


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

Features of Good Quality


Annotation Support
89.1%
55.6%
94.5%
83.5%
Isoforms, %
0
25
50
75
FSM
ISM
NIC
NNC


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


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