SQANTI *report*

Unique Genes: 12806

Unique Isoforms: 68553

Gene classification

Category	# Genes
Annotated Genes	12076
Novel Genes	730

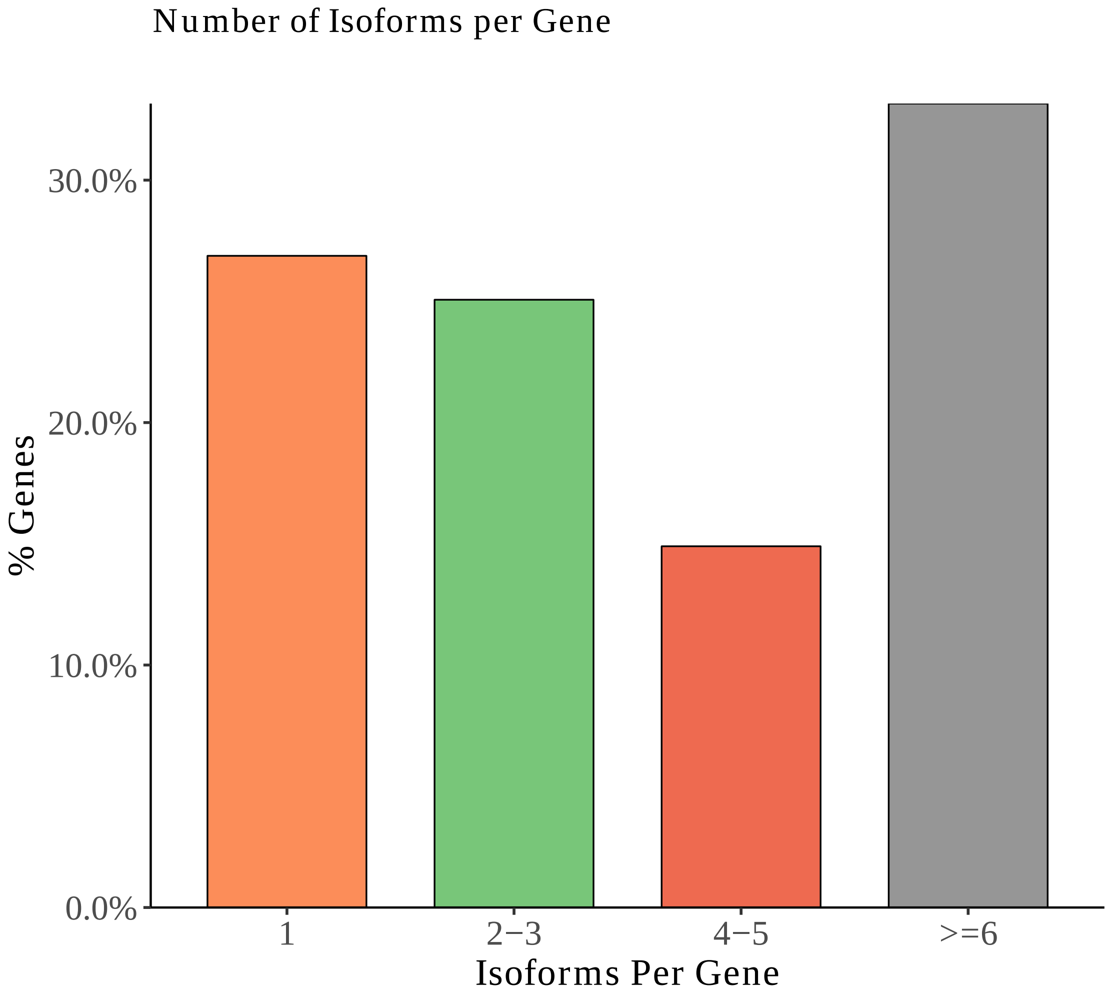
Characterization of transcripts based on splice junctions

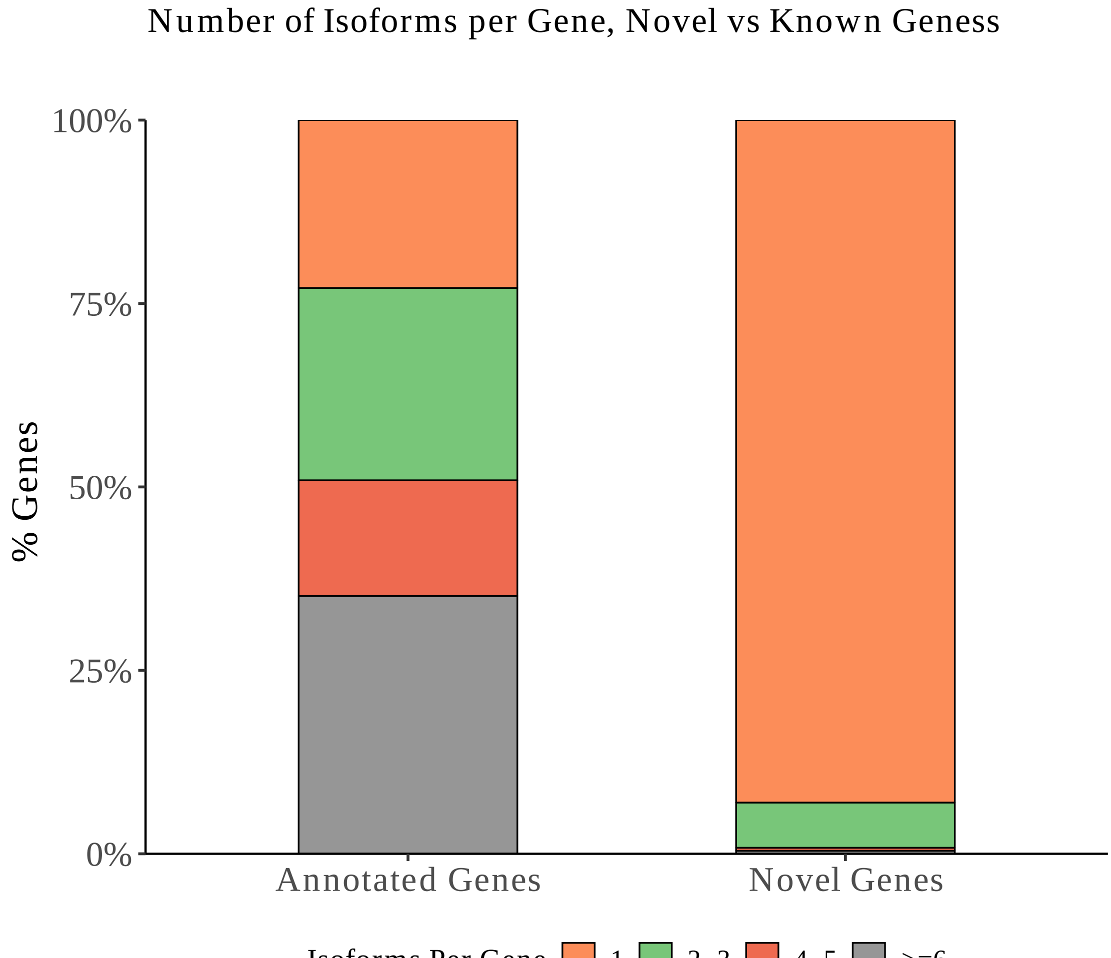
Category	# Isoforms
FSM	28988
ISM	5732
NIC	18868
NNC	13819
Genic Genomic	333
Antisense	463
Fusion	0
Intergenic	350
Genic Intron	0

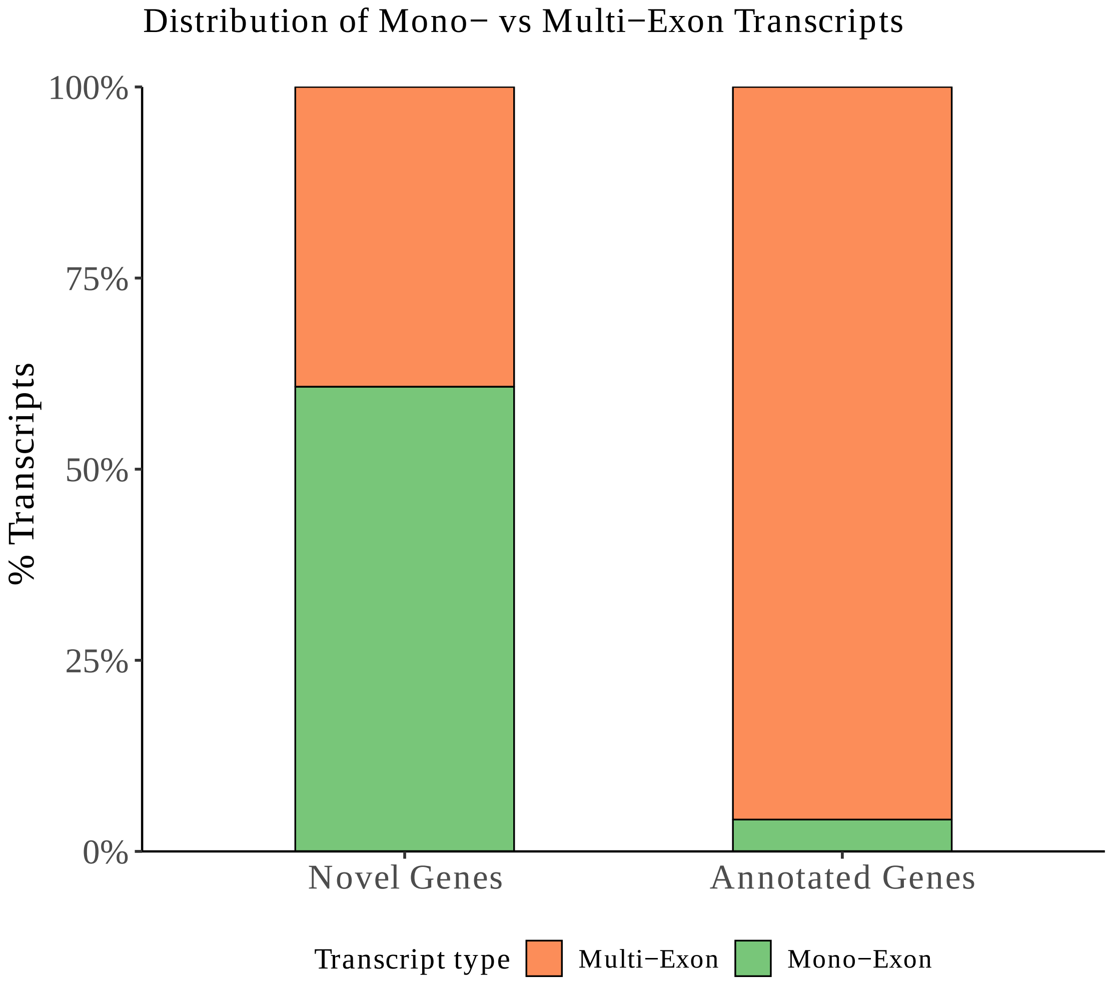
Splice Junction Classification

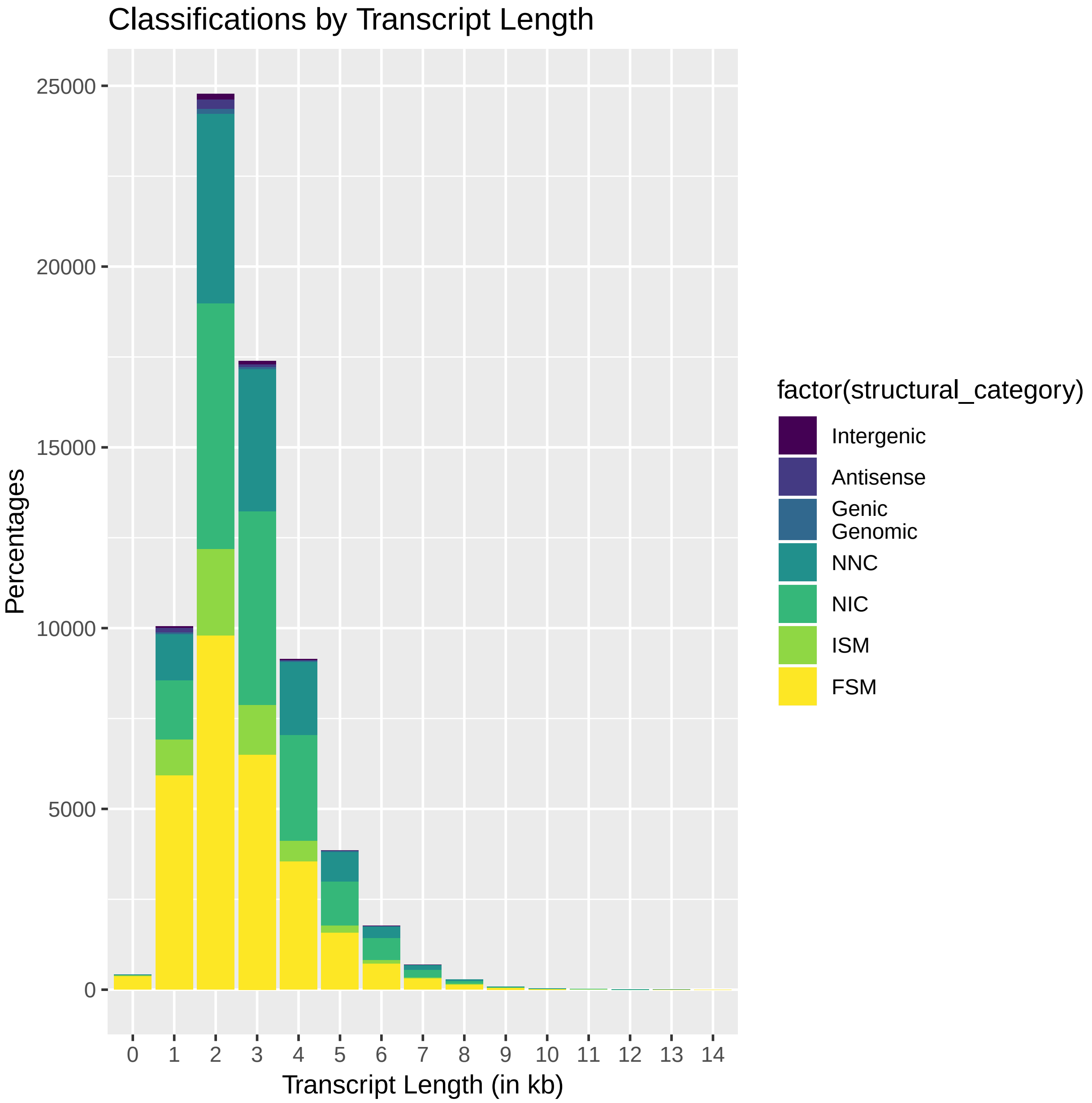
Category	# SJs	Percent
Known canonical	127162	87.97
Known Non−canonical	95	0.07
Novel canonical	13603	9.41
Novel Non−canonical	3697	2.56

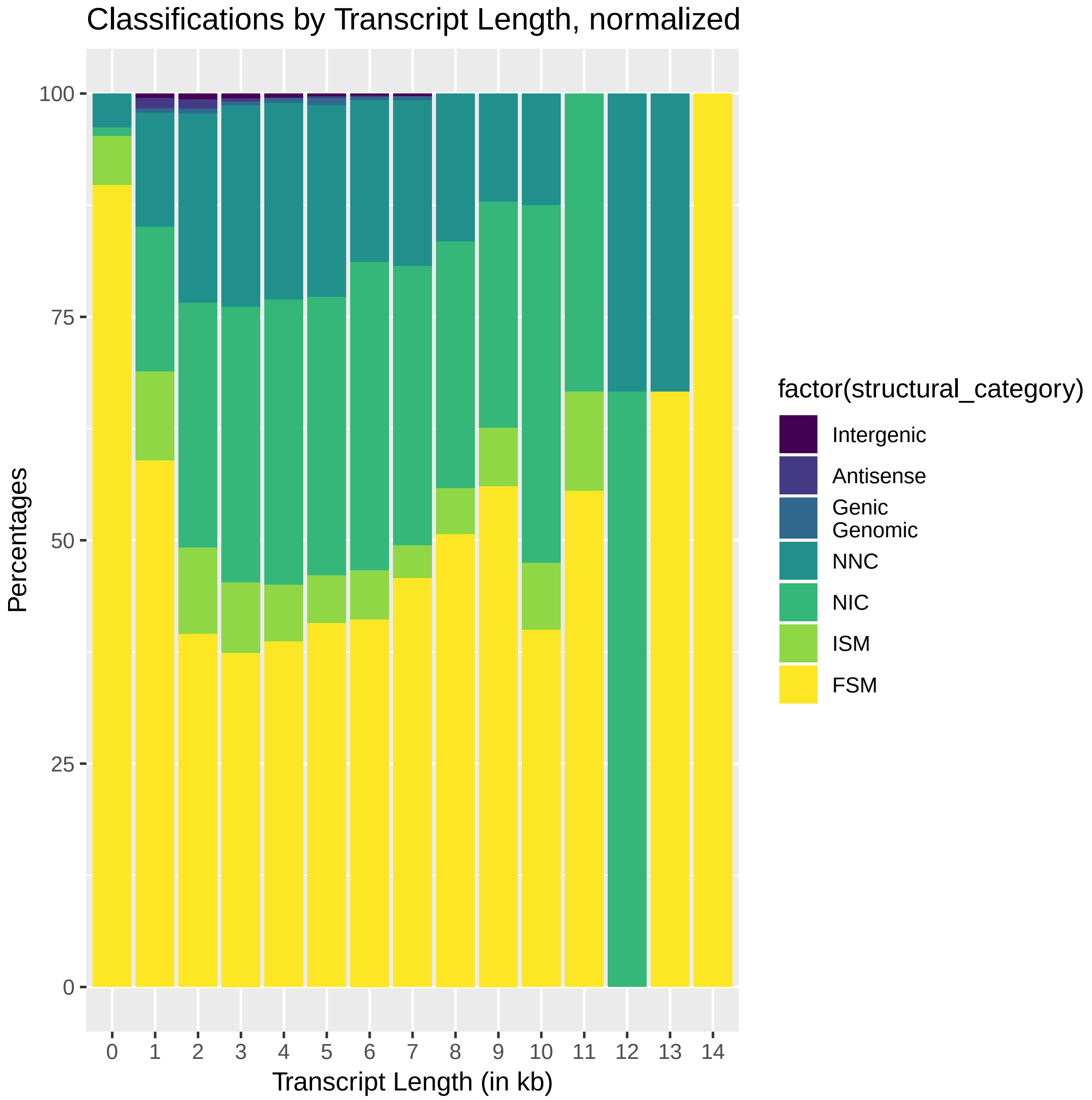
Gene Characterization

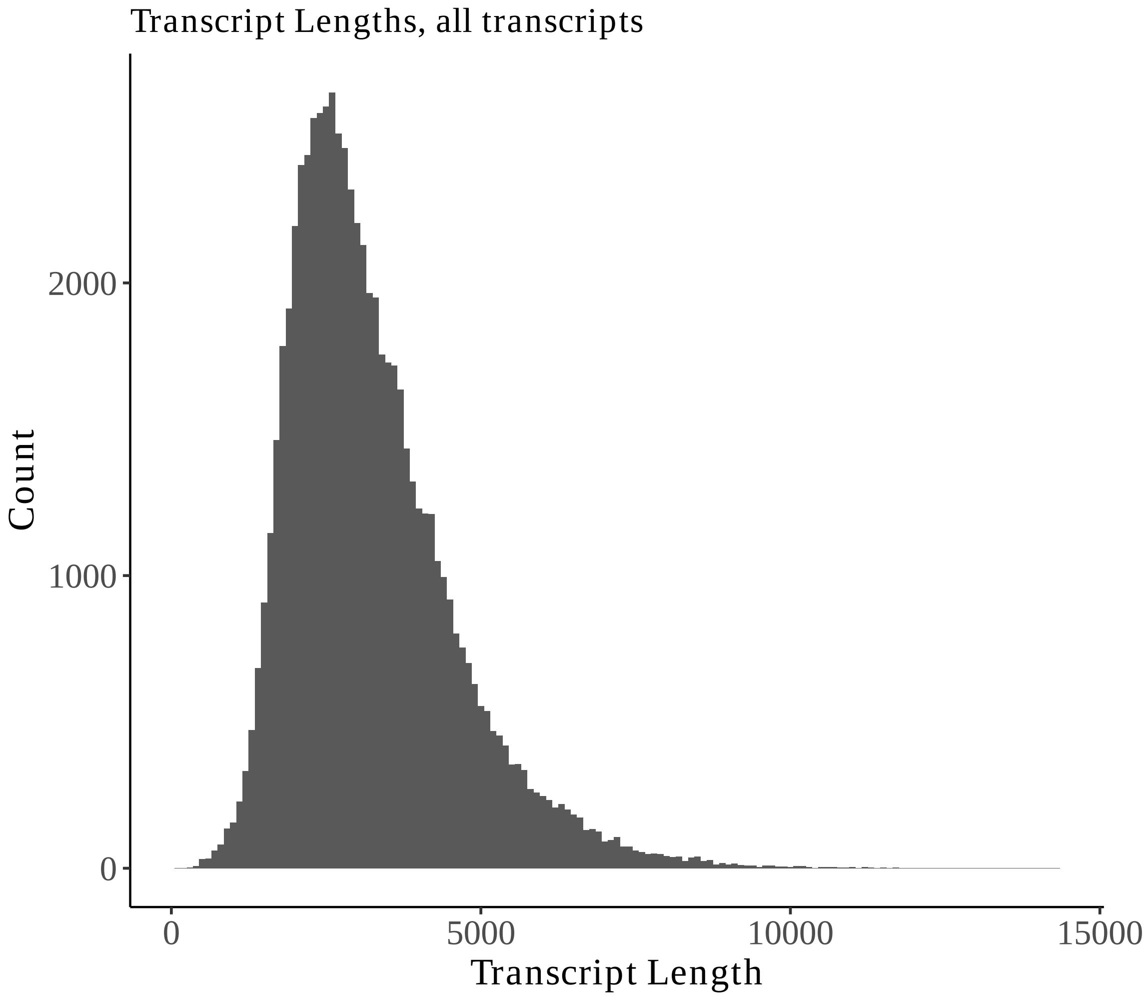

Number of Isoforms per Gene
% Genes
30.0%
20.0%
10.0%
0.0%
1
2−3
4−5
>=6
Isoforms Per Gene

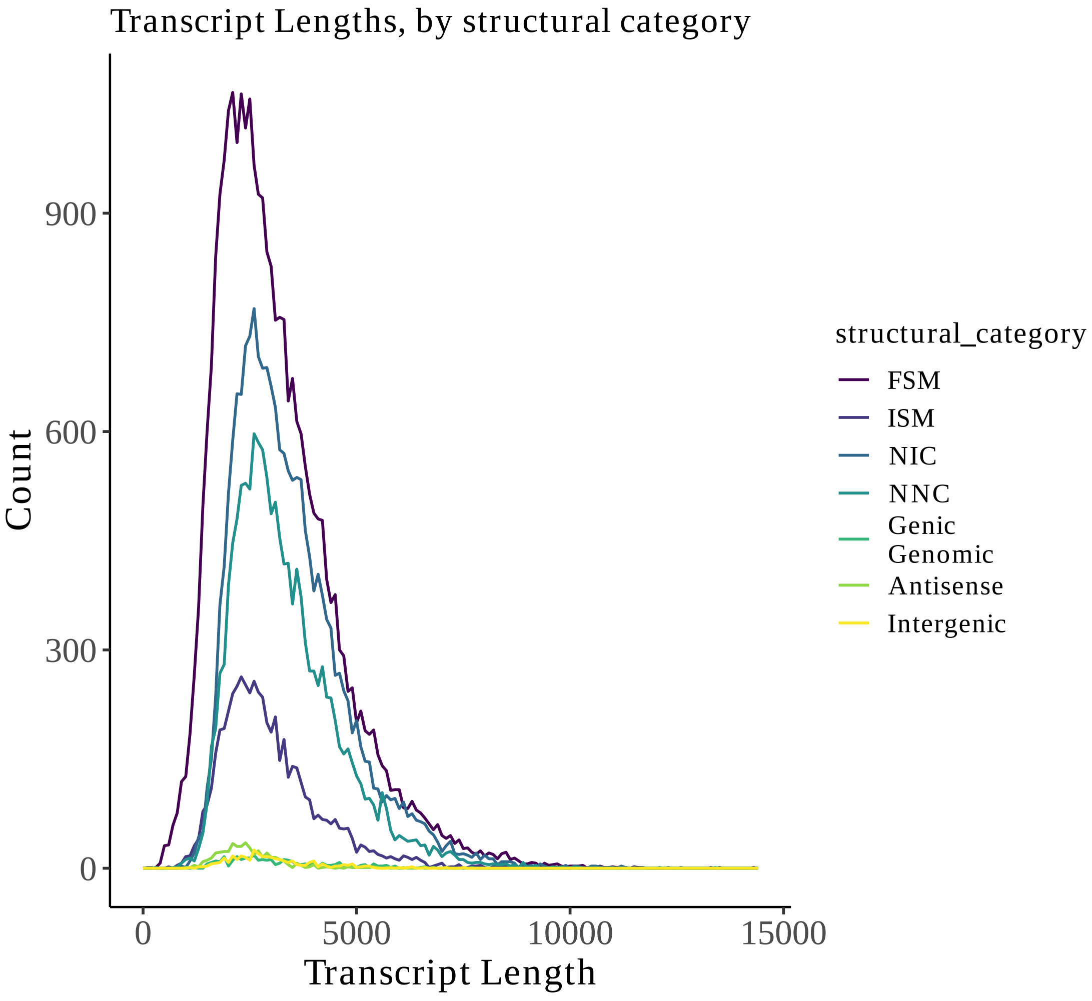

Number of Isoforms per Gene, Novel vs Known Geness
100%
75%
50%
25%
0%
% Genes
Annotated Genes
Novel Genes
Isoforms Per Gene
1
2−3
4−5
>=6

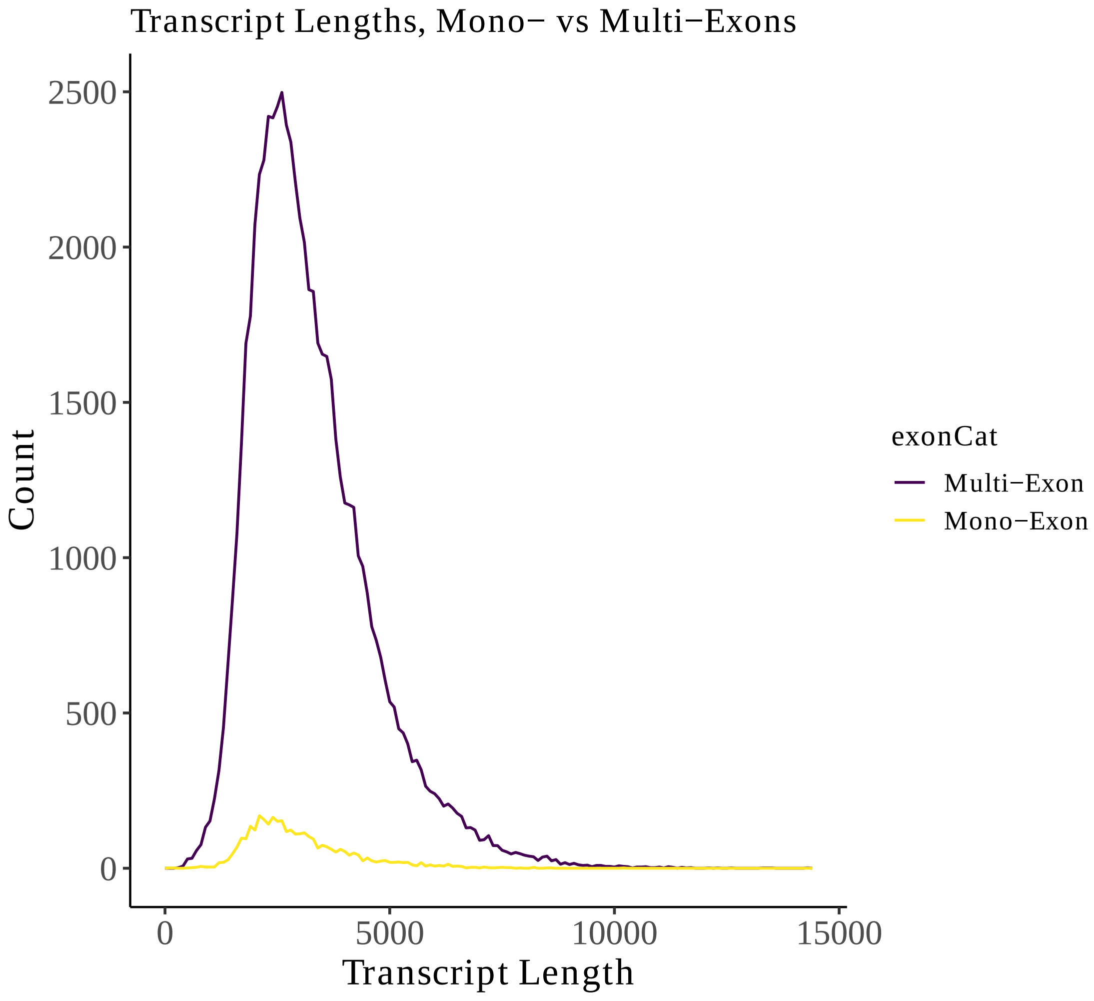

Distribution of Mono− vs Multi−Exon Transcripts
100%
75%
50%
25%
0%
% Transcripts
Novel Genes
Annotated Genes
Transcript type
Multi−Exon
Mono−Exon


Classifications by Transcript Length
Percentages
25000
20000
15000
10000
5000
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Transcript Length (in kb)
factor(structural_category)
Intergenic
Antisense
Genic
Genomic
NNC
NIC
ISM
FSM

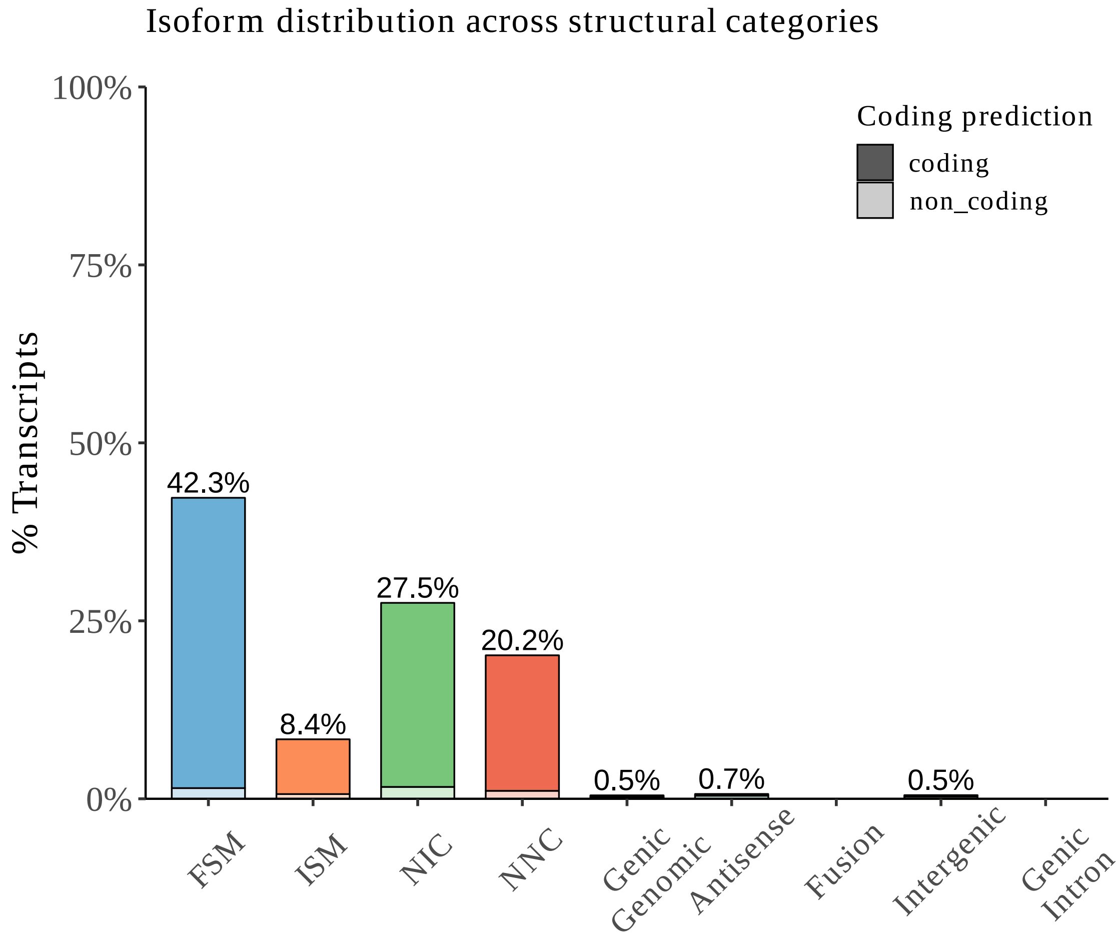

Classifications by Transcript Length, normalized
Percentages
100
75
50
25
0
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Transcript Length (in kb)
factor(structural_category)
Intergenic
Antisense
Genic
Genomic
NNC
NIC
ISM
FSM

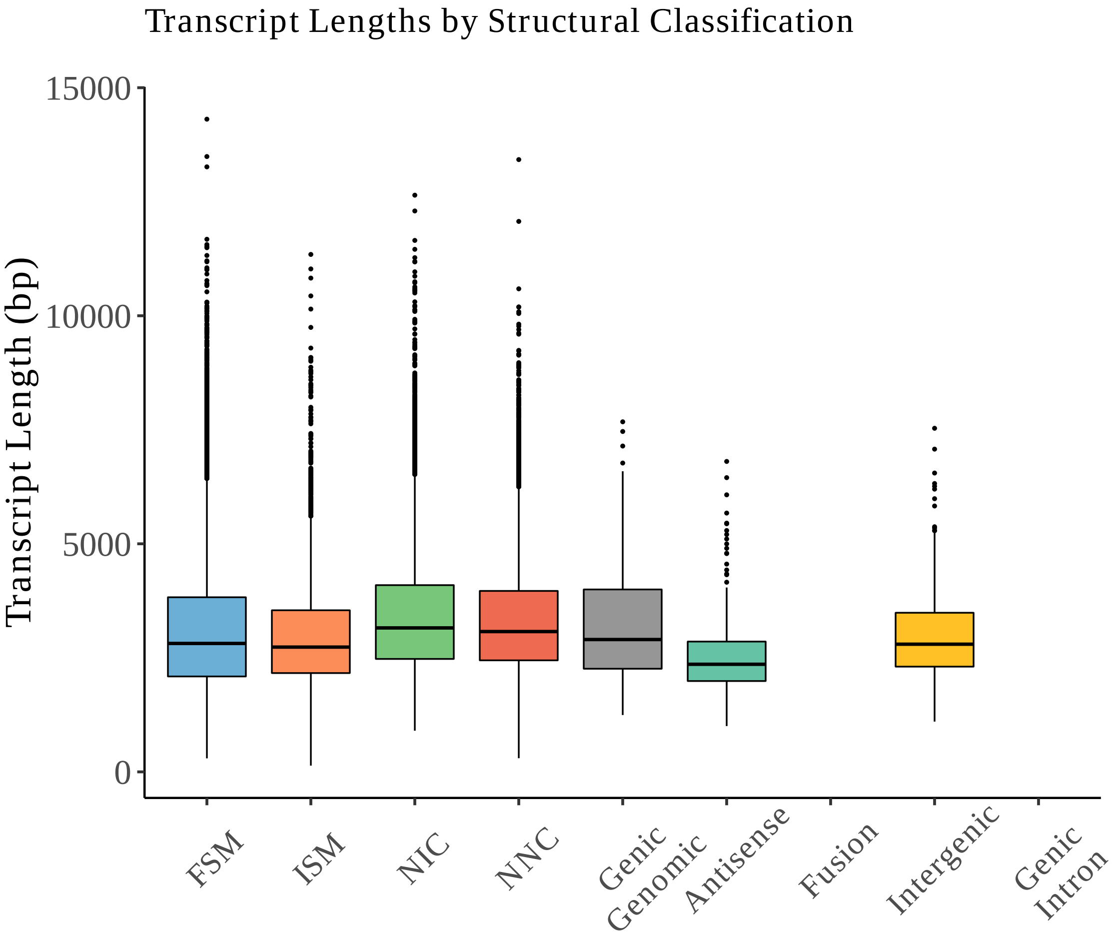

Transcript Lengths, all transcripts
Count
2000
1000
0
0
5000
10000
15000
Transcript Length


Transcript Lengths, by structural category
Count
900
600
300
0
0
5000
10000
15000
Transcript Length
structural_category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Intergenic

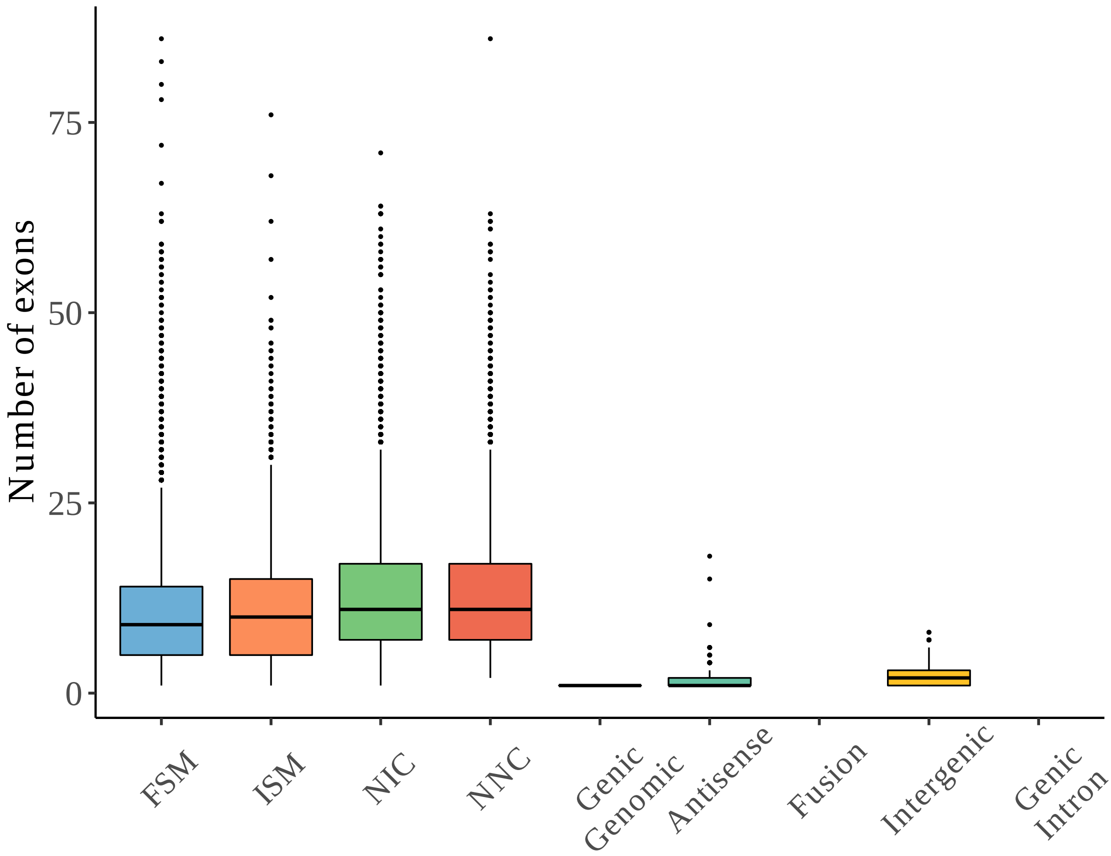

Transcript Lengths, Mono− vs Multi−Exons
Count
2500
2000
1500
1000
500
0
0
5000
10000
15000
Transcript Length
exonCat
Multi−Exon
Mono−Exon

Structrual Isoform Characterization by Splice Junctions


Isoform distribution across structural categories
Coding prediction
coding
non_coding
% Transcripts
100%
75%
50%
25%
0%
42.3%
8.4%
27.5%
20.2%
0.5%
0.7%
0.5%
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

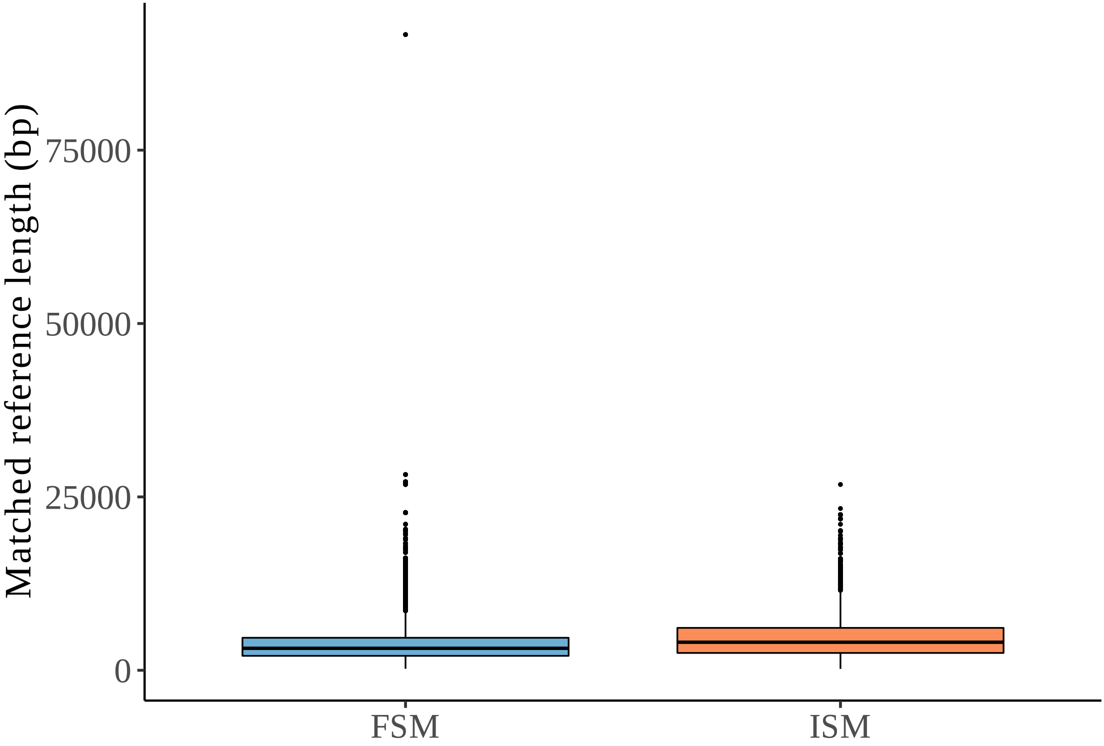

Transcript Lengths by Structural Classification
Transcript Length (bp)
15000
10000
5000
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

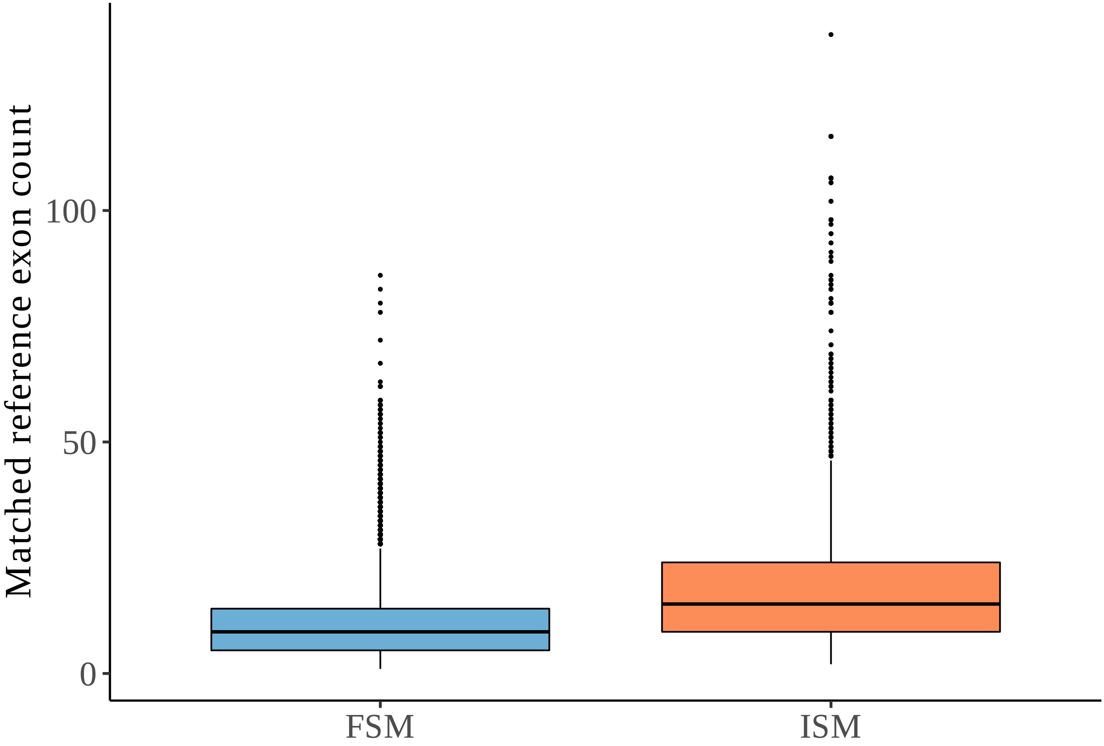
Exon Counts by Structural Classification



Length distribution of matched reference transcripts

Applicable only to FSM and ISM categories

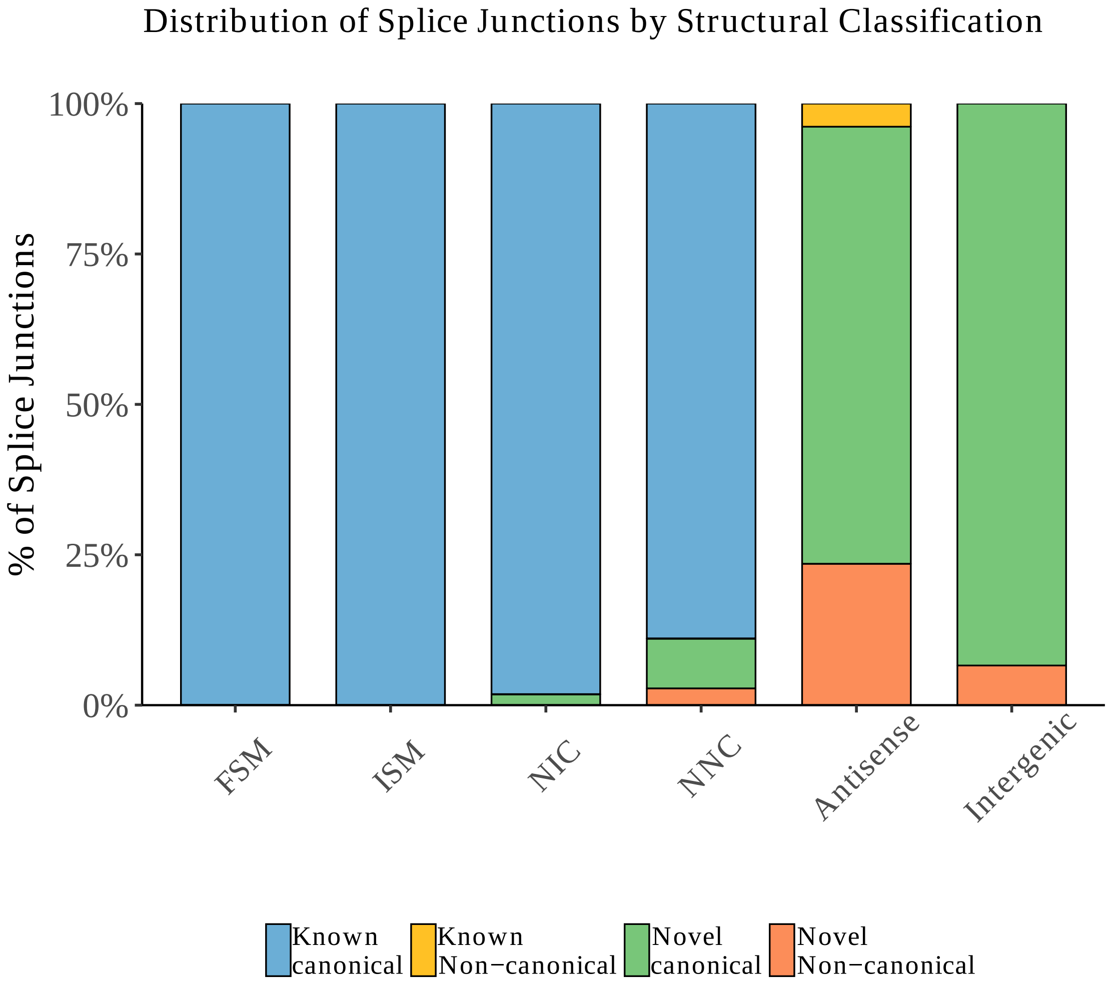


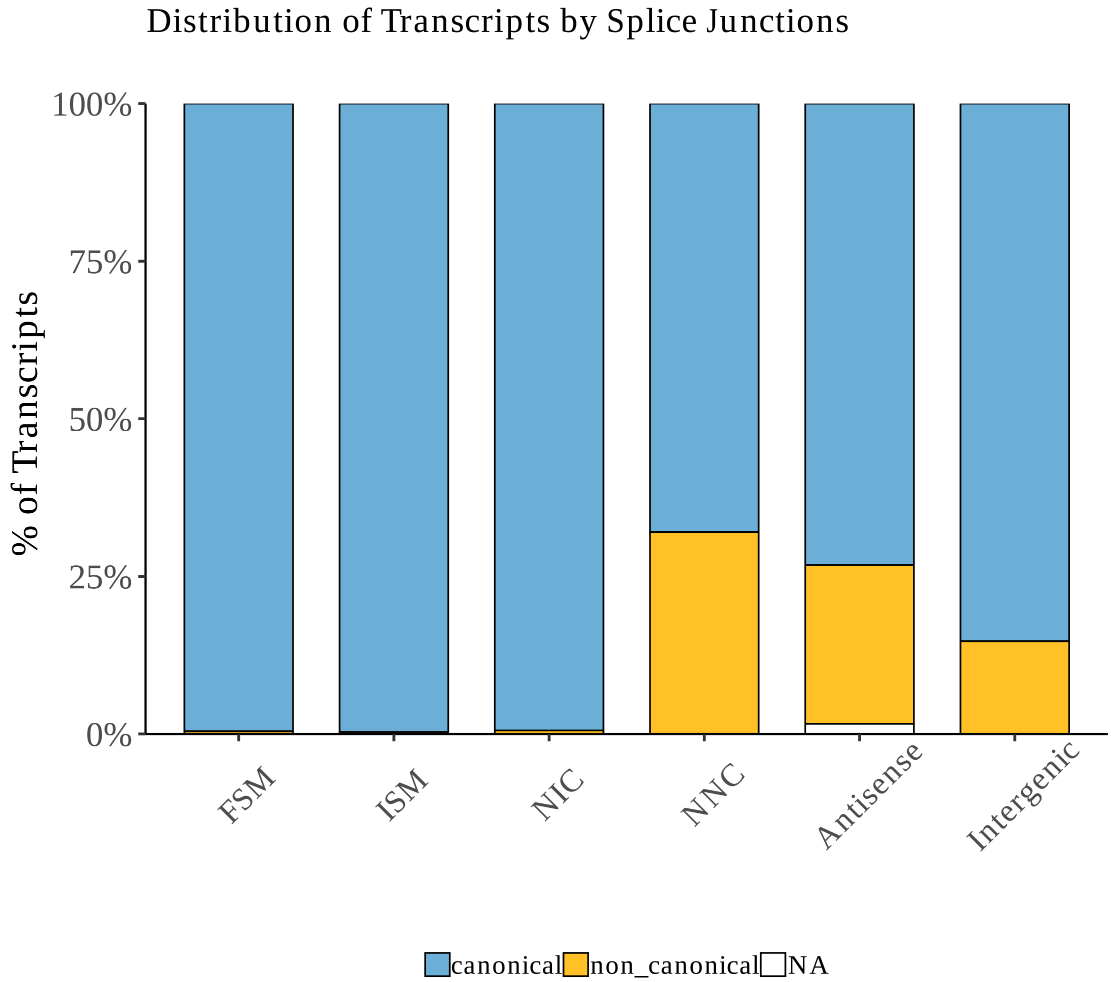


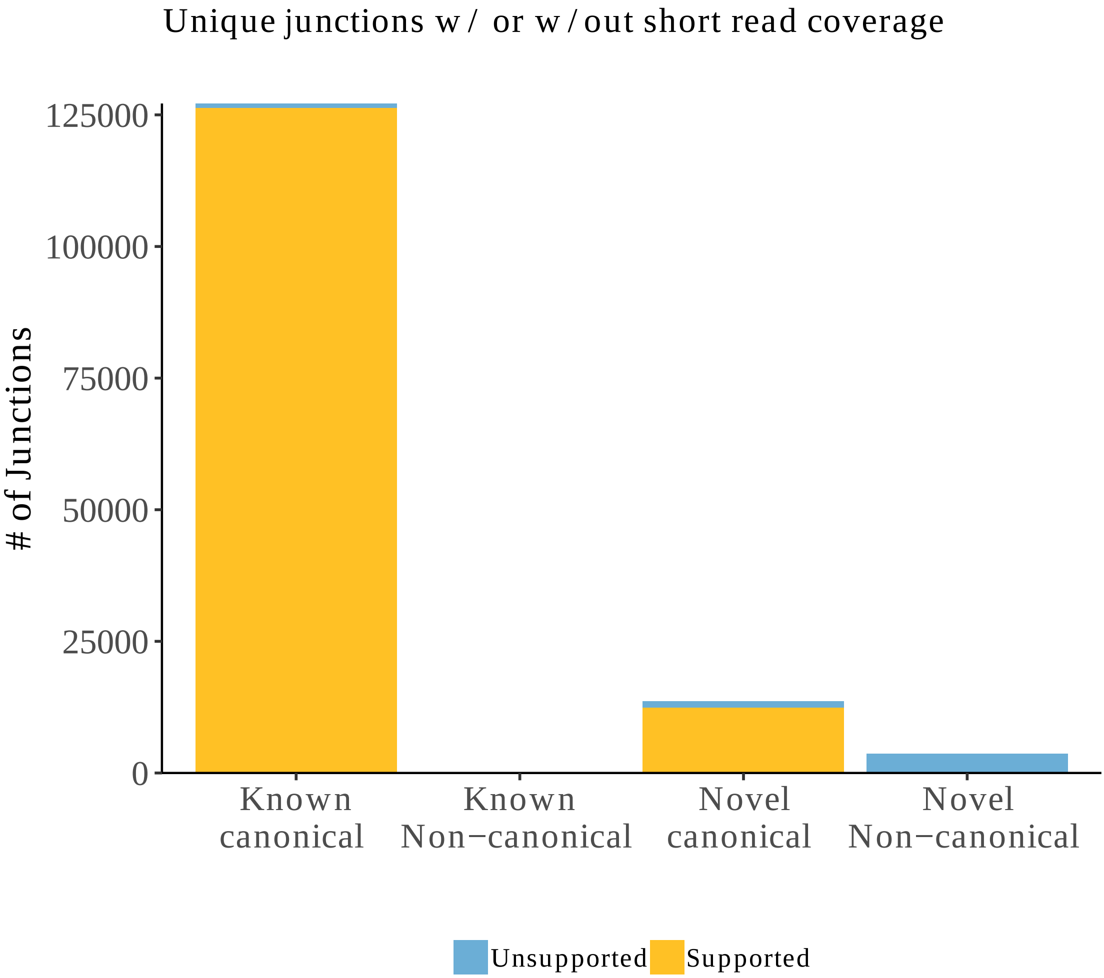
Exon number distribution of matched reference transcripts

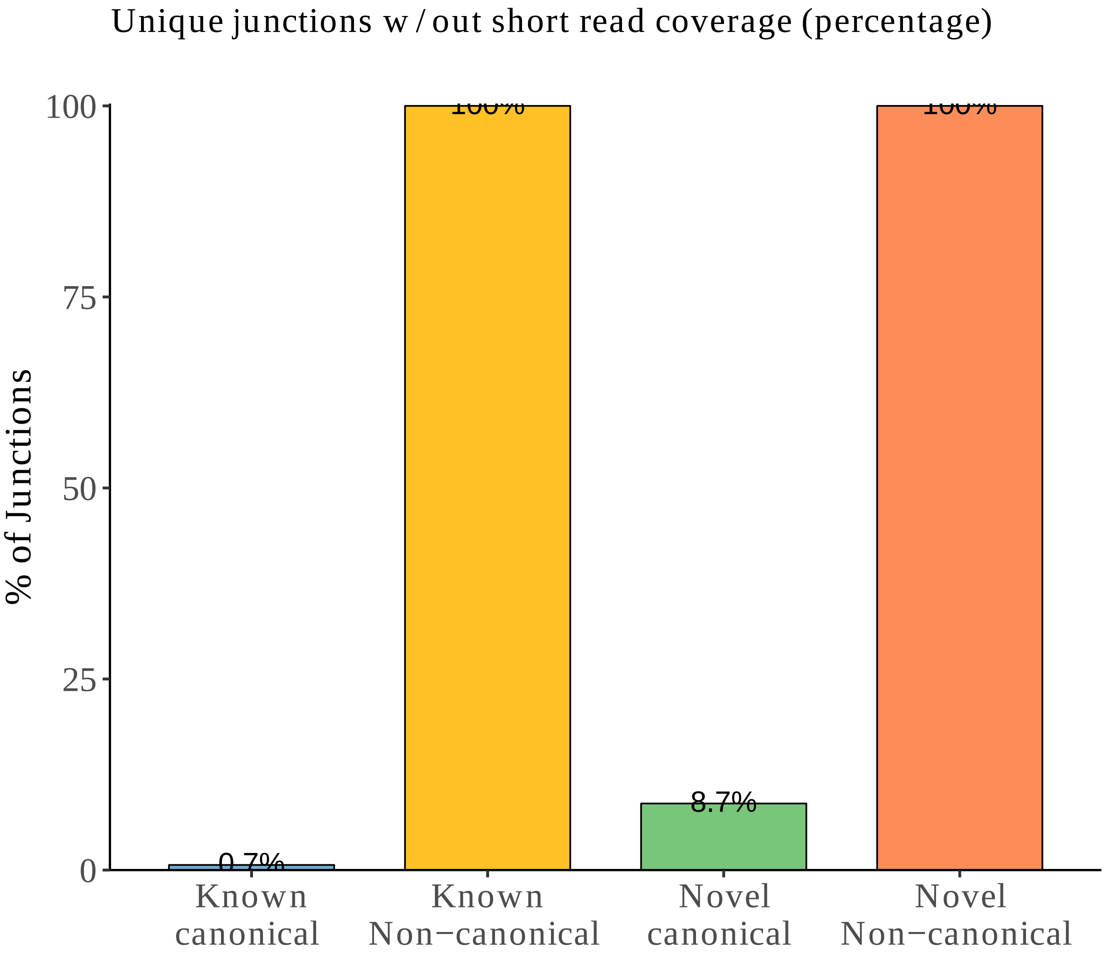
Applicable only to FSM and ISM categories

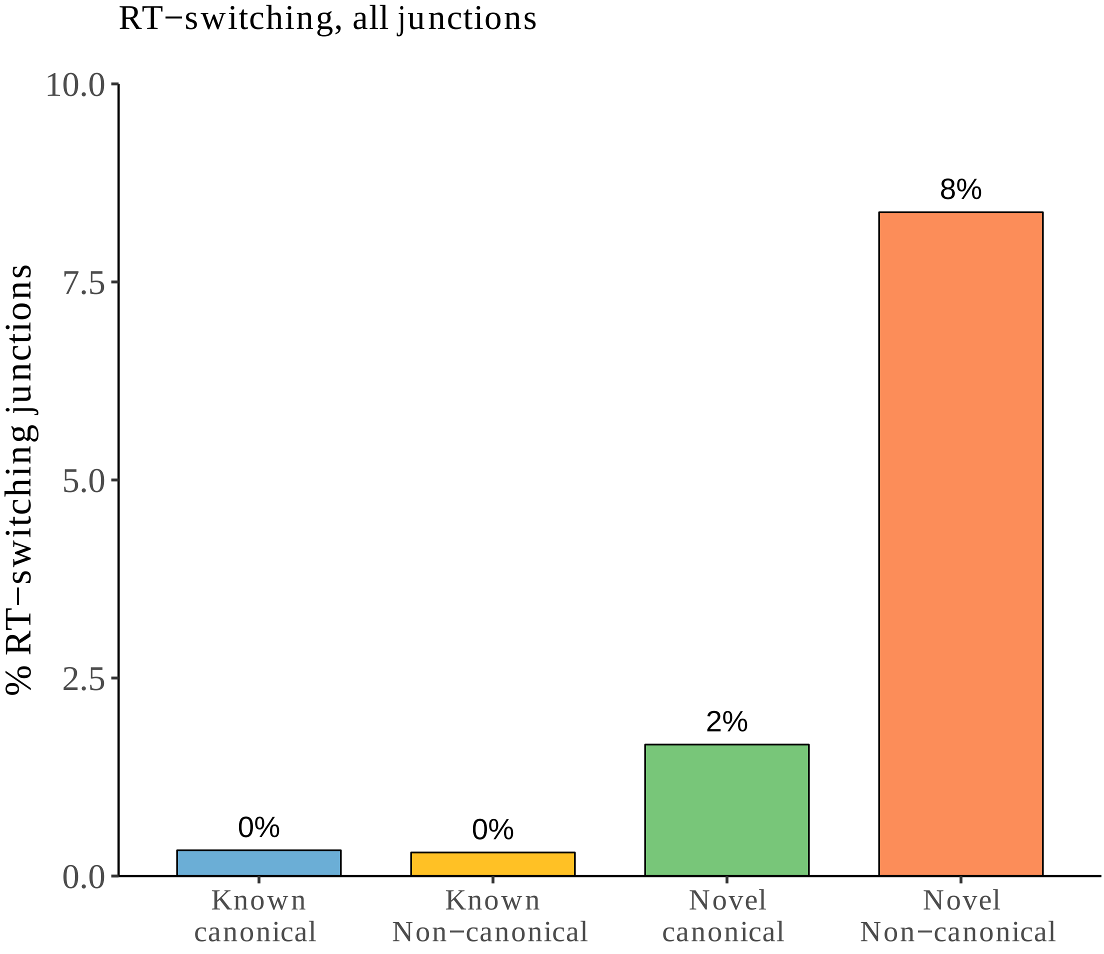
Splice Junction Characterization

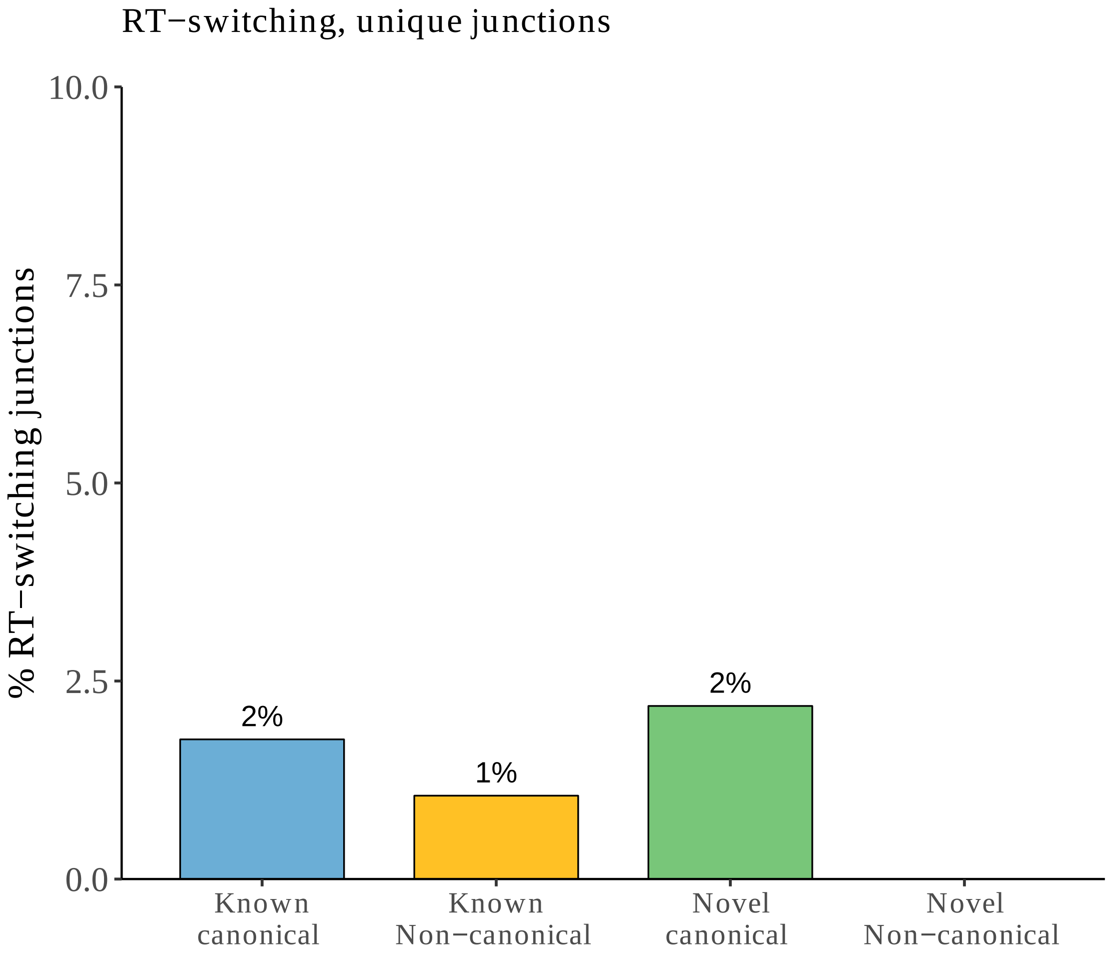

Distribution of Splice Junctions by Structural Classification
% of Splice Junctions
100%
75%
50%
25%
0%
FSM
ISM
NIC
NNC
Antisense
Intergenic
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical


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


Unique junctions w/ or w/out short read coverage
# of Junctions
125000
100000
75000
50000
25000
0
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical
Unsupported
Supported


Unique junctions w/out short read coverage (percentage)
% of Junctions
100
75
50
25
0
0.7%
100%
8.7%
100%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

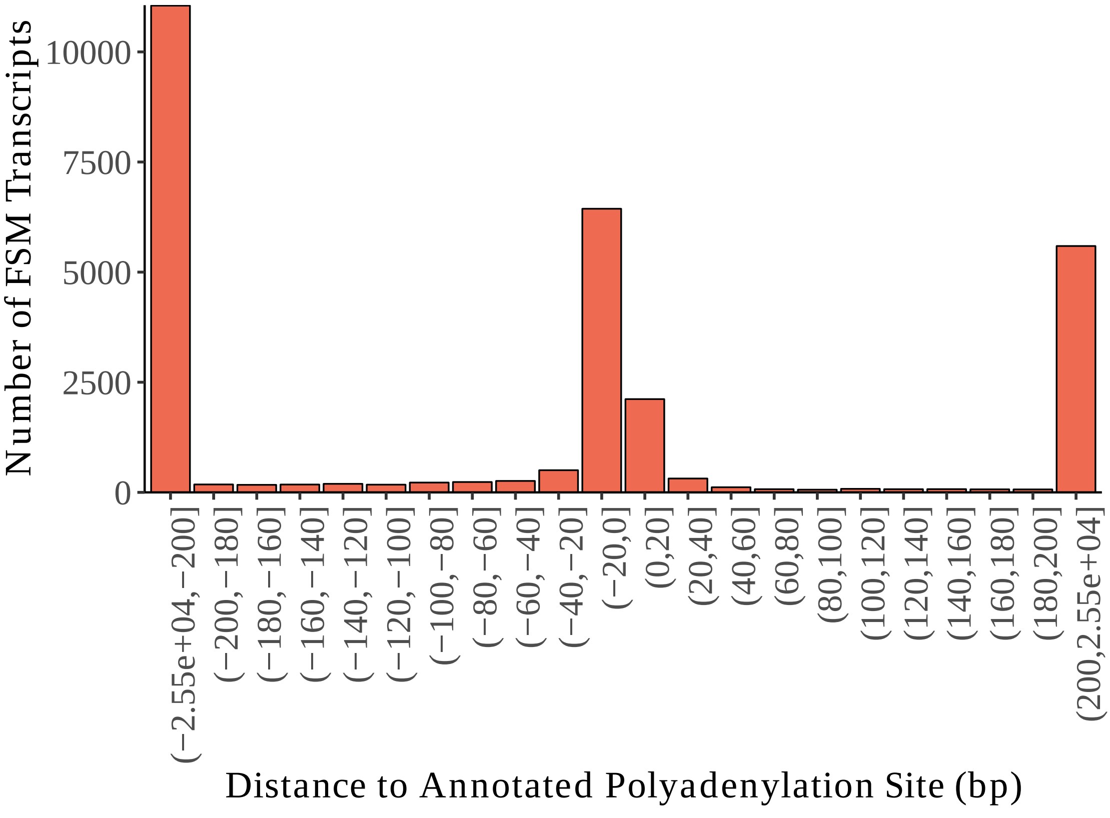

RT−switching, all junctions
% RT−switching junctions
10.0
7.5
5.0
2.5
0.0
0%
0%
2%
8%
Known
canonical
Known
Non−canonical
Novel
canonical
Novel
Non−canonical


RT−switching, unique junctions
% RT−switching junctions
10.0
7.5
5.0
2.5
0.0
2%
1%
2%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

Comparison with Annotated TSS and PolyA Sites

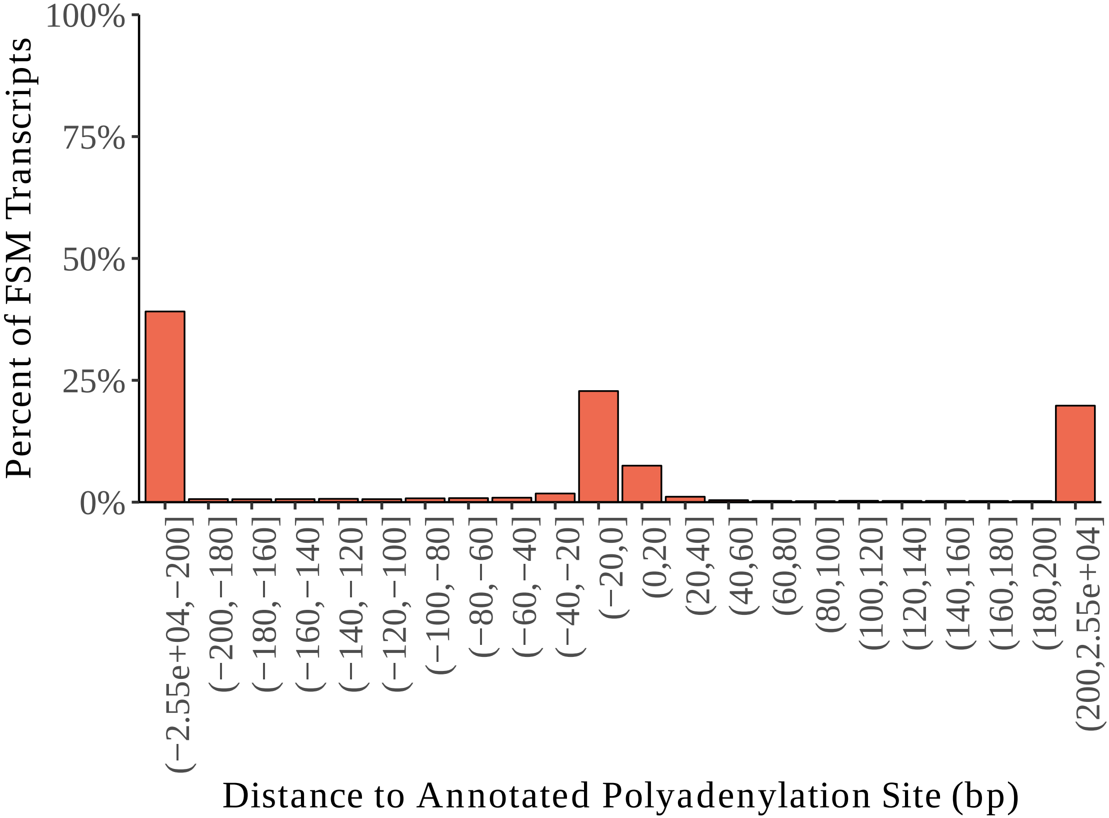
Distance to Annotated Polyadenylation Site, FSM only

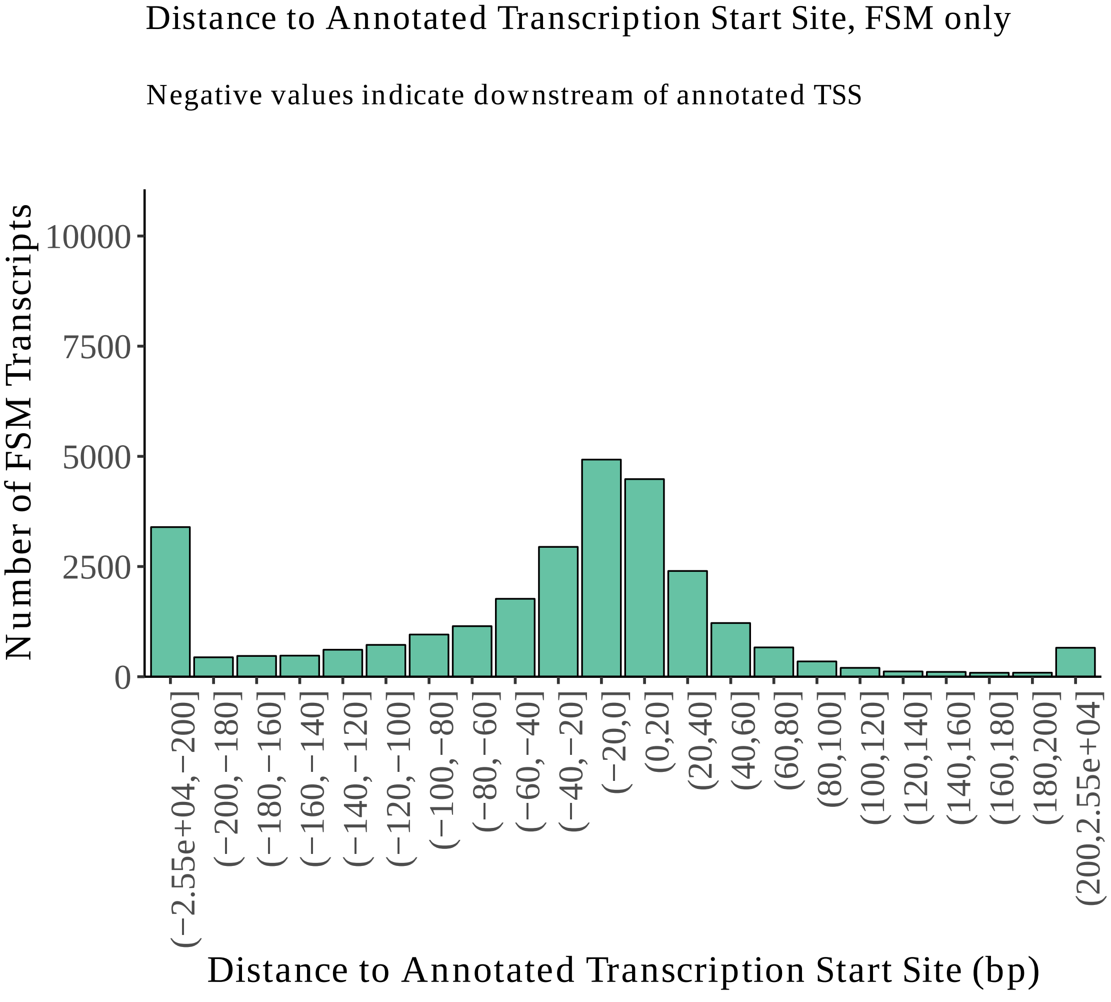
Negative values indicate upstream of annotated polyA site



Distance to Annotated Polyadenylation Site, FSM only

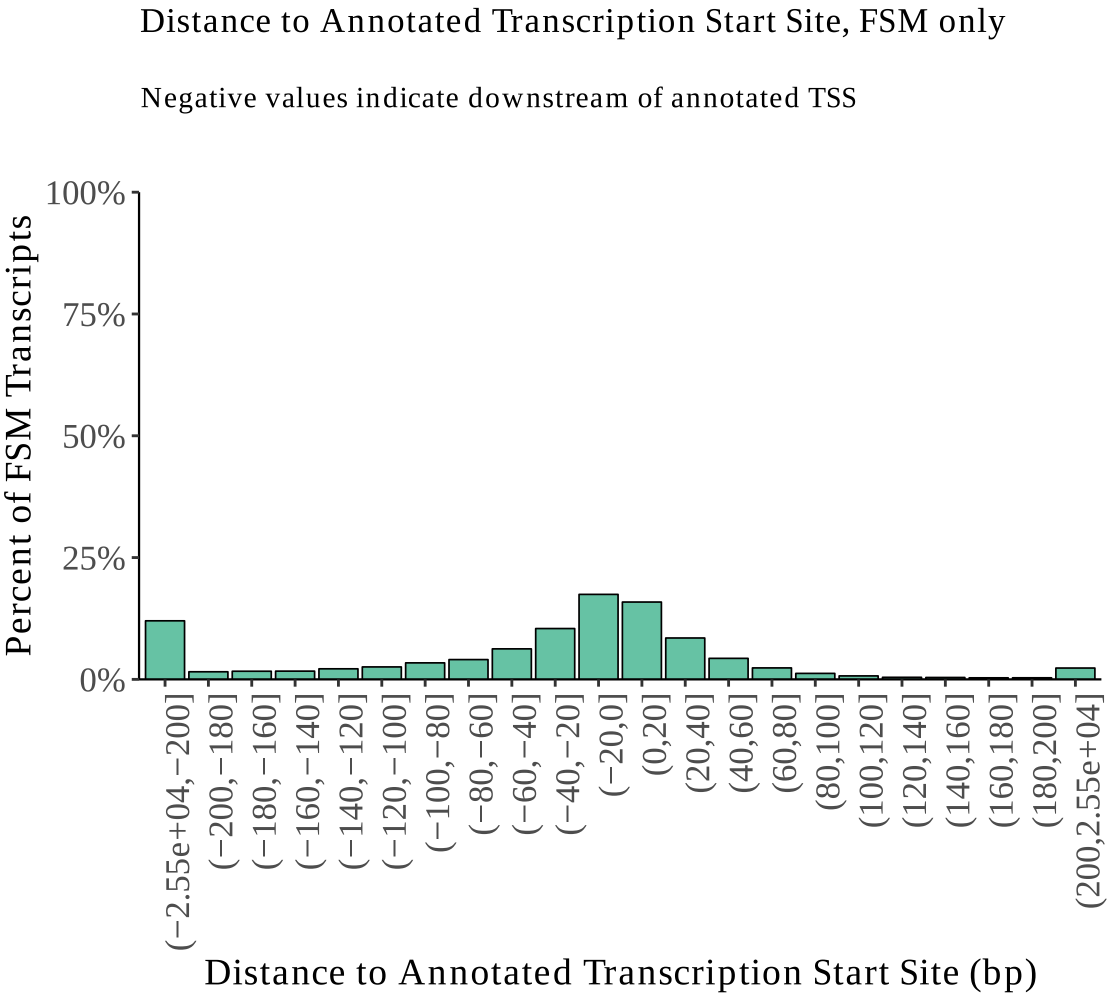
Negative values indicate upstream of annotated polyA site





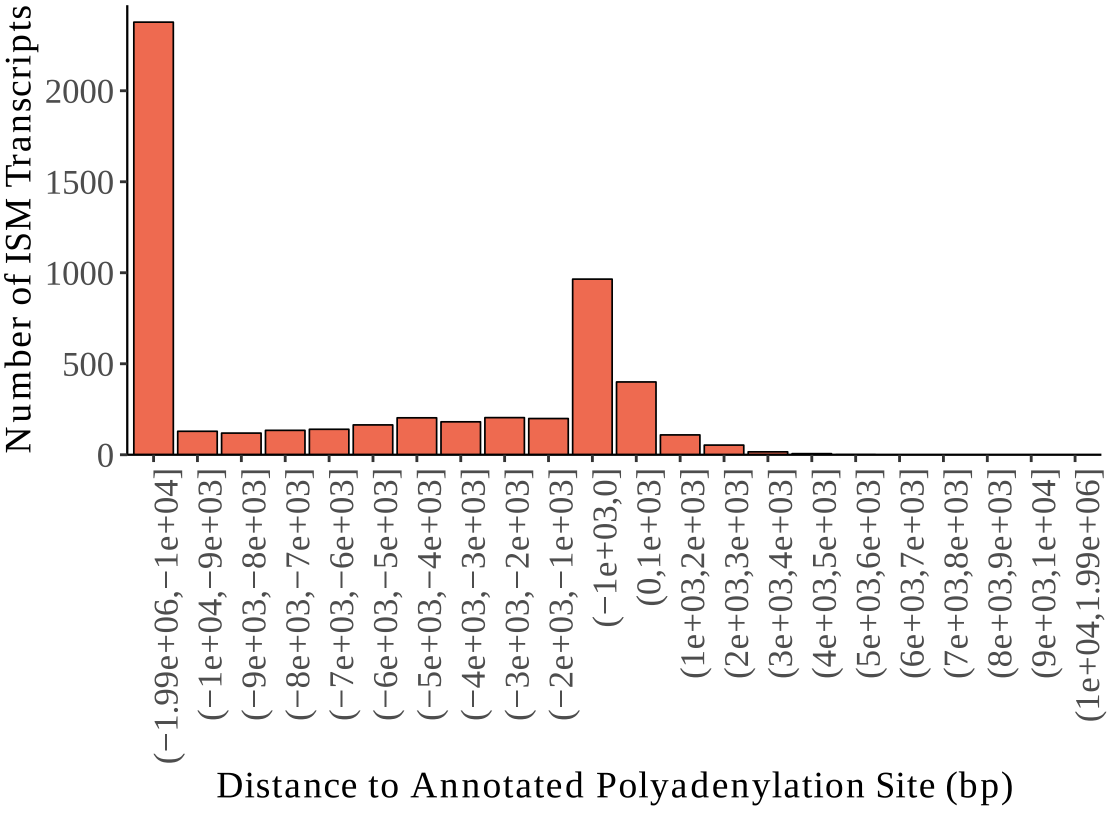
Distance to Annotated Transcription Start Site, FSM only

Negative values indicate downstream of annotated TSS



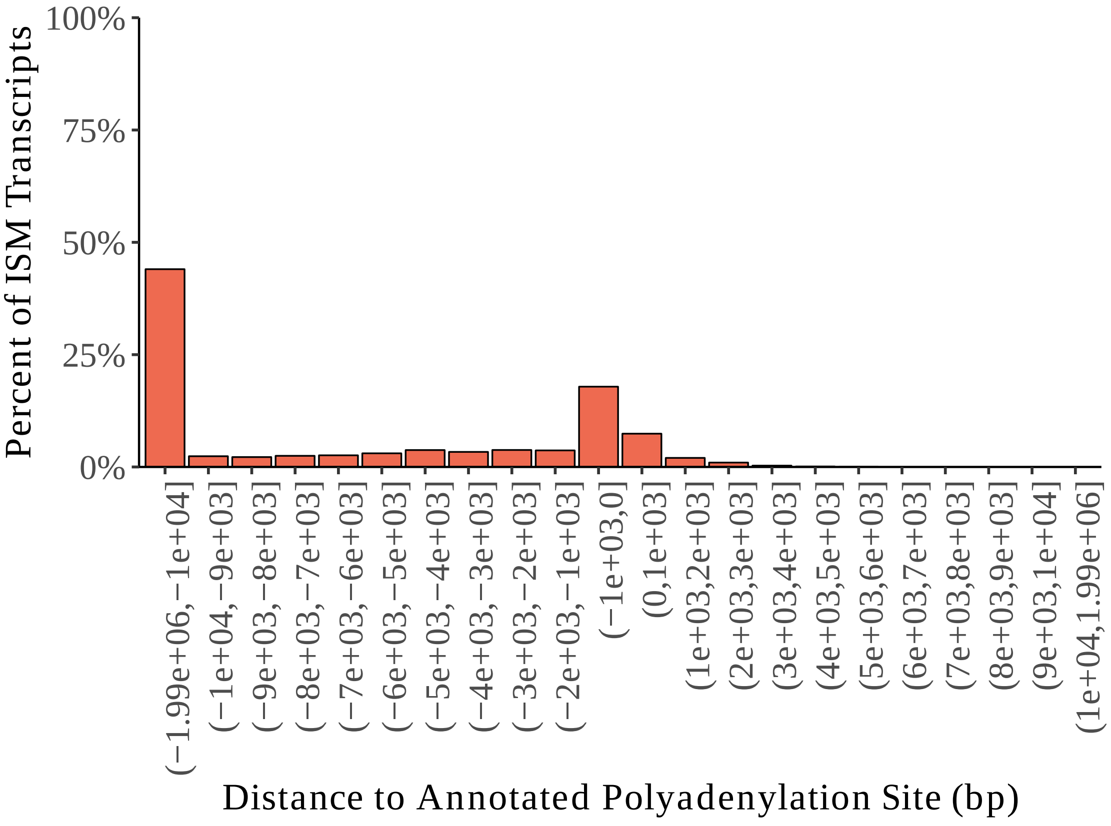
Distance to Annotated Transcription Start Site, FSM only

Negative values indicate downstream of annotated TSS



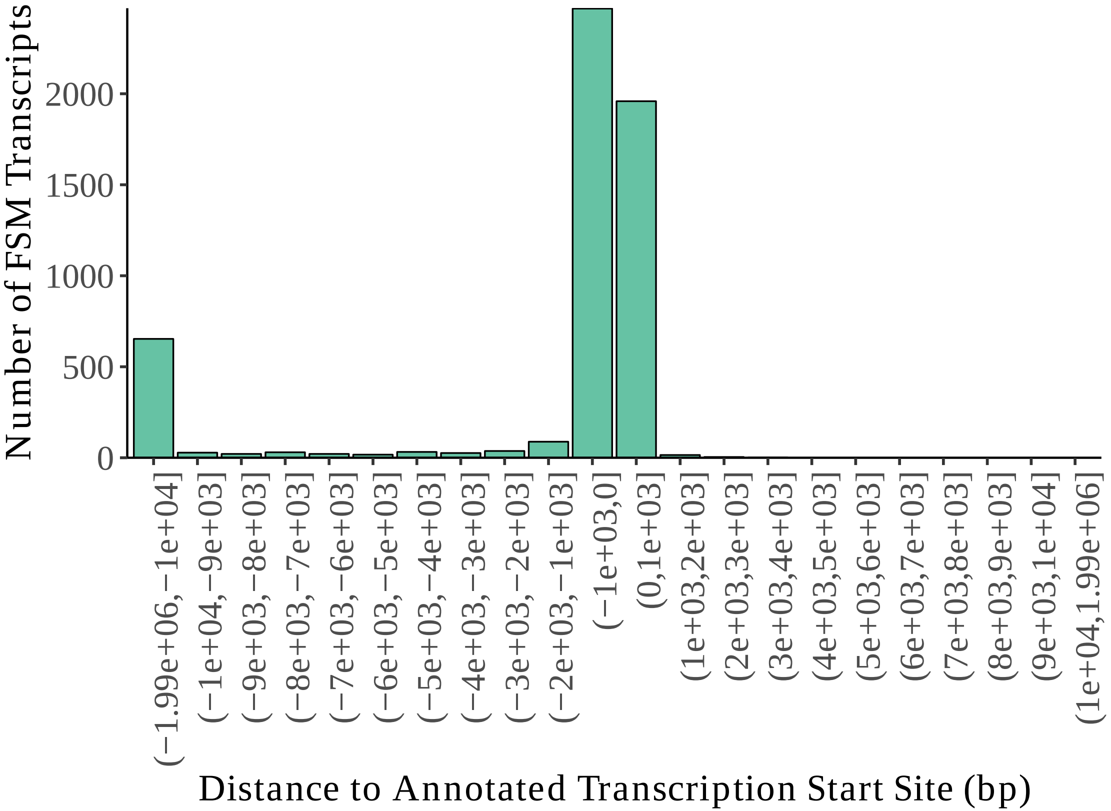
Distance to Annotated Polyadenylation Site, ISM only

Negative values indicate upstream of annotated polyA site



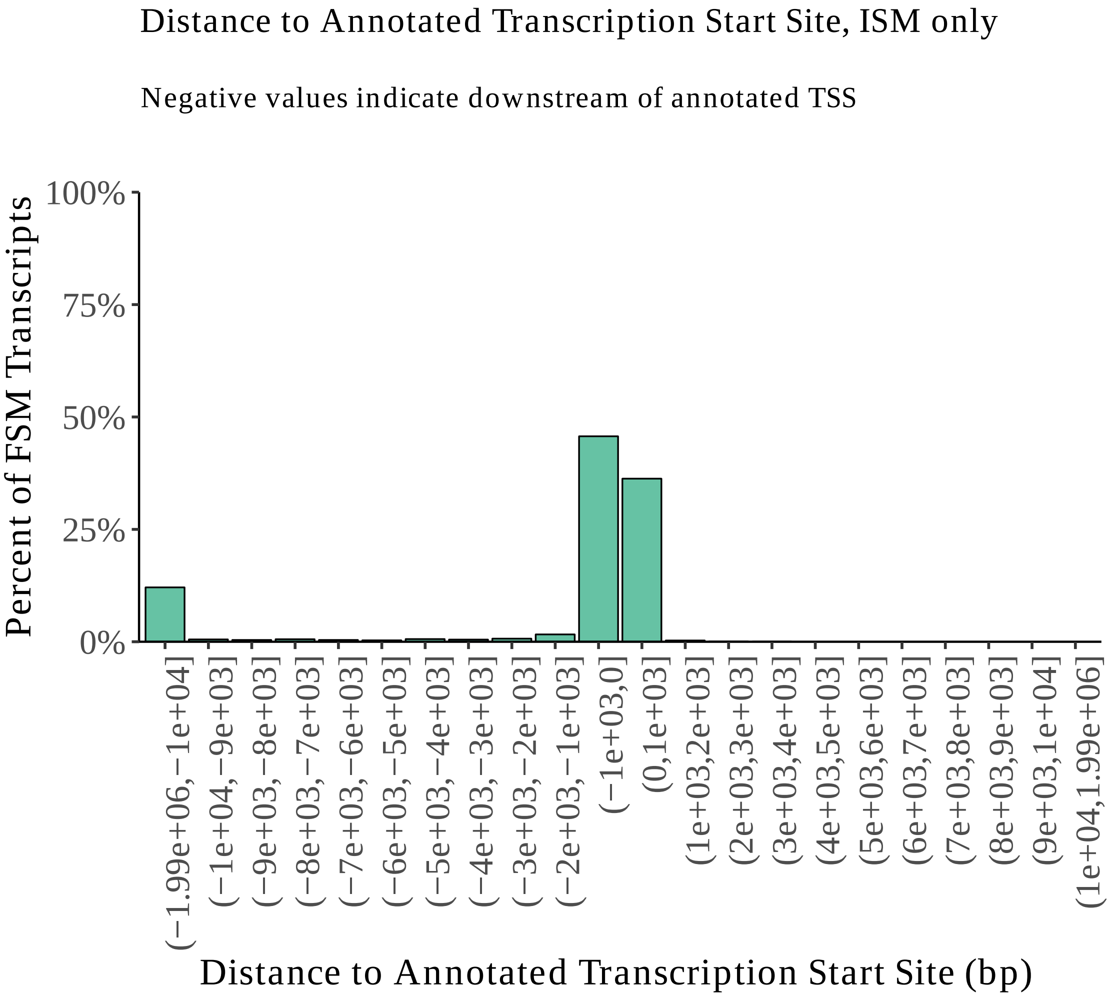
Distance to Annotated Polyadenylation Site, ISM only

Negative values indicate upstream of annotated polyA site

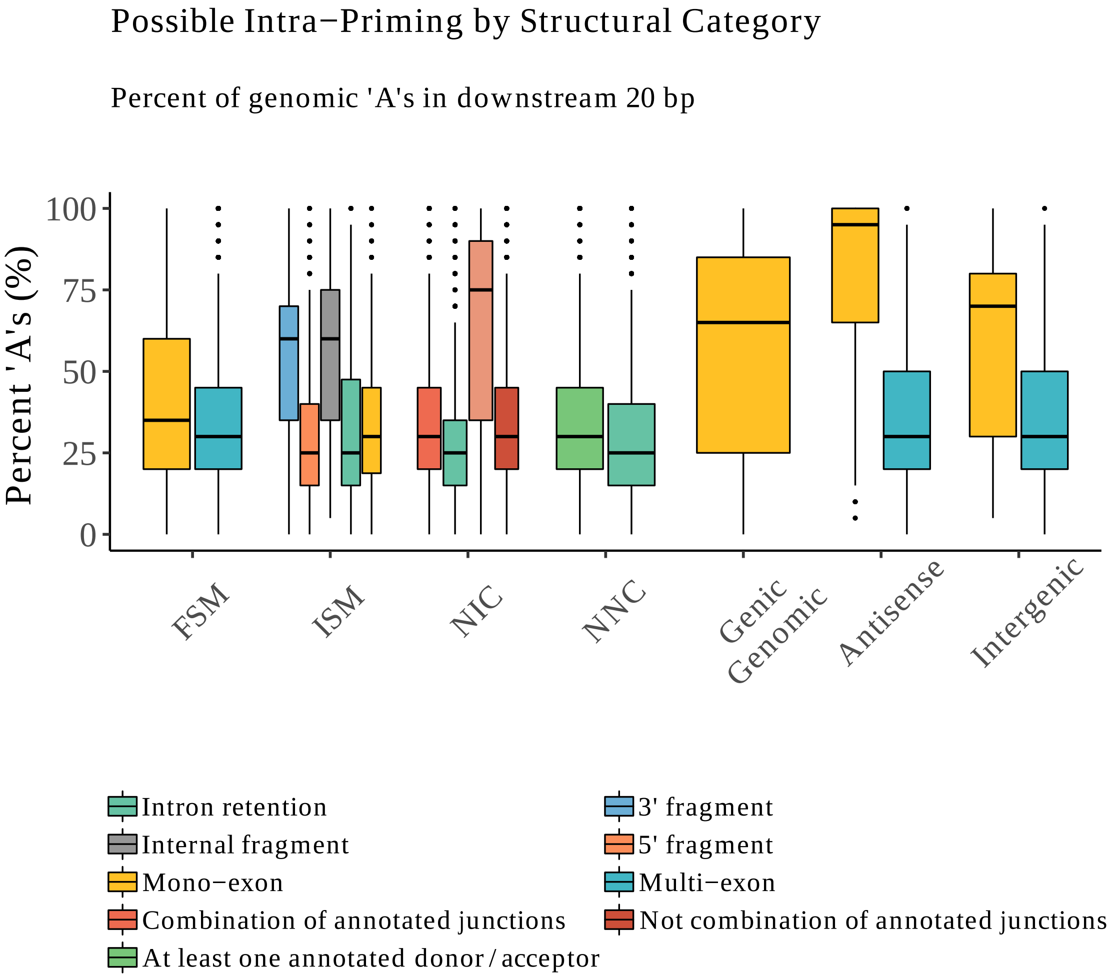


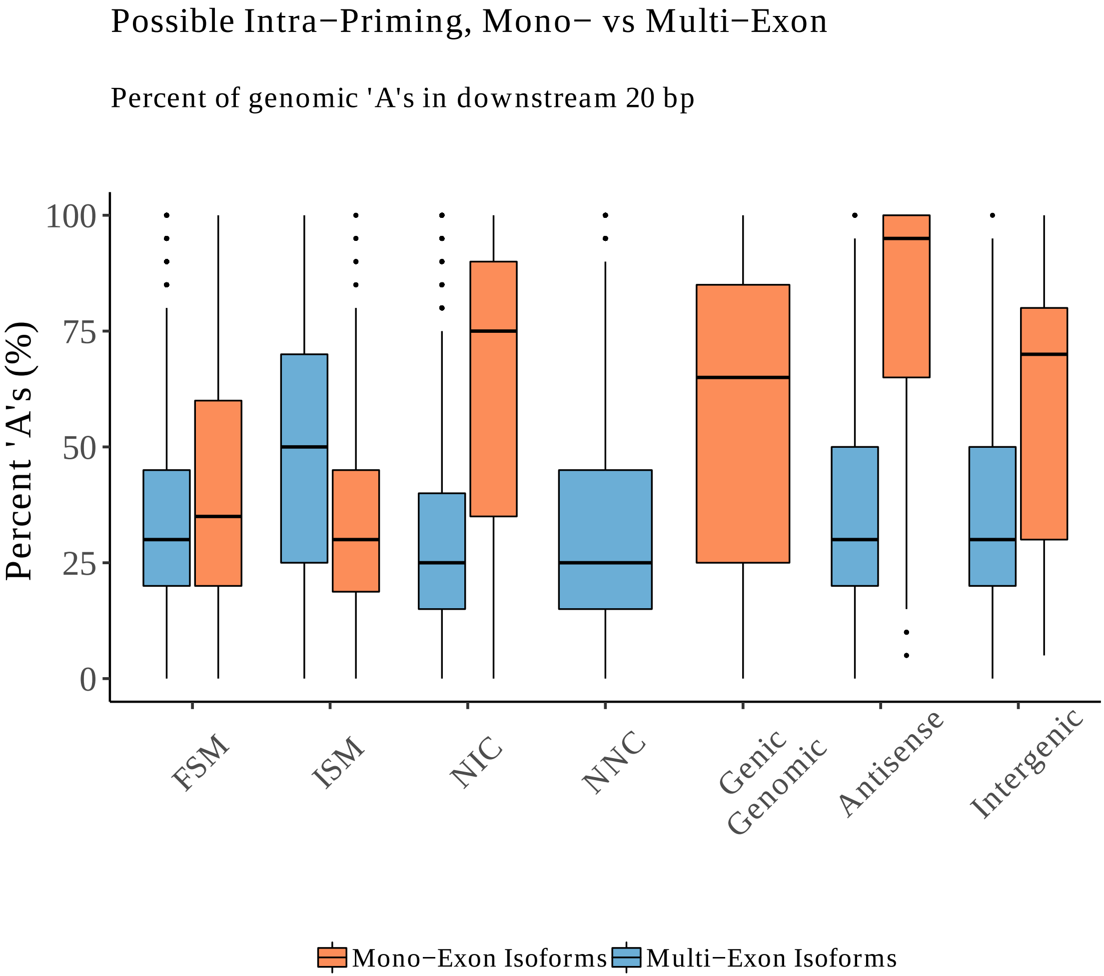
Distance to Annotated Transcription Start Site, ISM only

Negative values indicate downstream of annotated TSS


Distance to Annotated Transcription Start Site, ISM only
Negative values indicate downstream of annotated TSS
Percent of FSM Transcripts
100%
75%
50%
25%
0%
(−1.99e+06,−1e+04]
(−1e+04,−9e+03]
(−9e+03,−8e+03]
(−8e+03,−7e+03]
(−7e+03,−6e+03]
(−6e+03,−5e+03]
(−5e+03,−4e+03]
(−4e+03,−3e+03]
(−3e+03,−2e+03]
(−2e+03,−1e+03]
(−1e+03,0]
(0,1e+03]
(1e+03,2e+03]
(2e+03,3e+03]
(3e+03,4e+03]
(4e+03,5e+03]
(5e+03,6e+03]
(6e+03,7e+03]
(7e+03,8e+03]
(8e+03,9e+03]
(9e+03,1e+04]
(1e+04,1.99e+06]
Distance to Annotated Transcription Start Site (bp)

Intra−Priming Quality Check


Possible Intra−Priming by Structural Category
Percent of genomic 'A's in downstream 20 bp
Percent 'A's (%)
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Intergenic
Intron retention
3' fragment
Internal fragment
5' fragment
Mono−exon
Multi−exon
Combination of annotated junctions
Not combination of annotated junctions
At least one annotated donor/acceptor

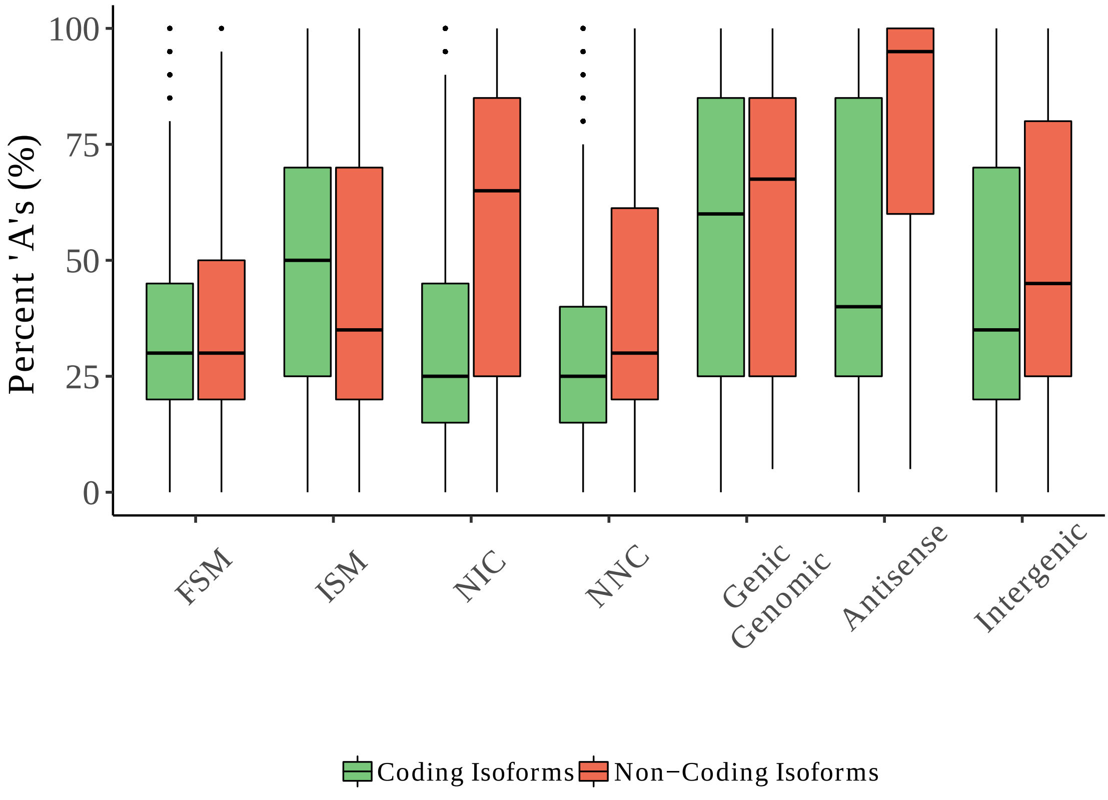


Possible Intra−Priming, Mono− vs Multi−Exon

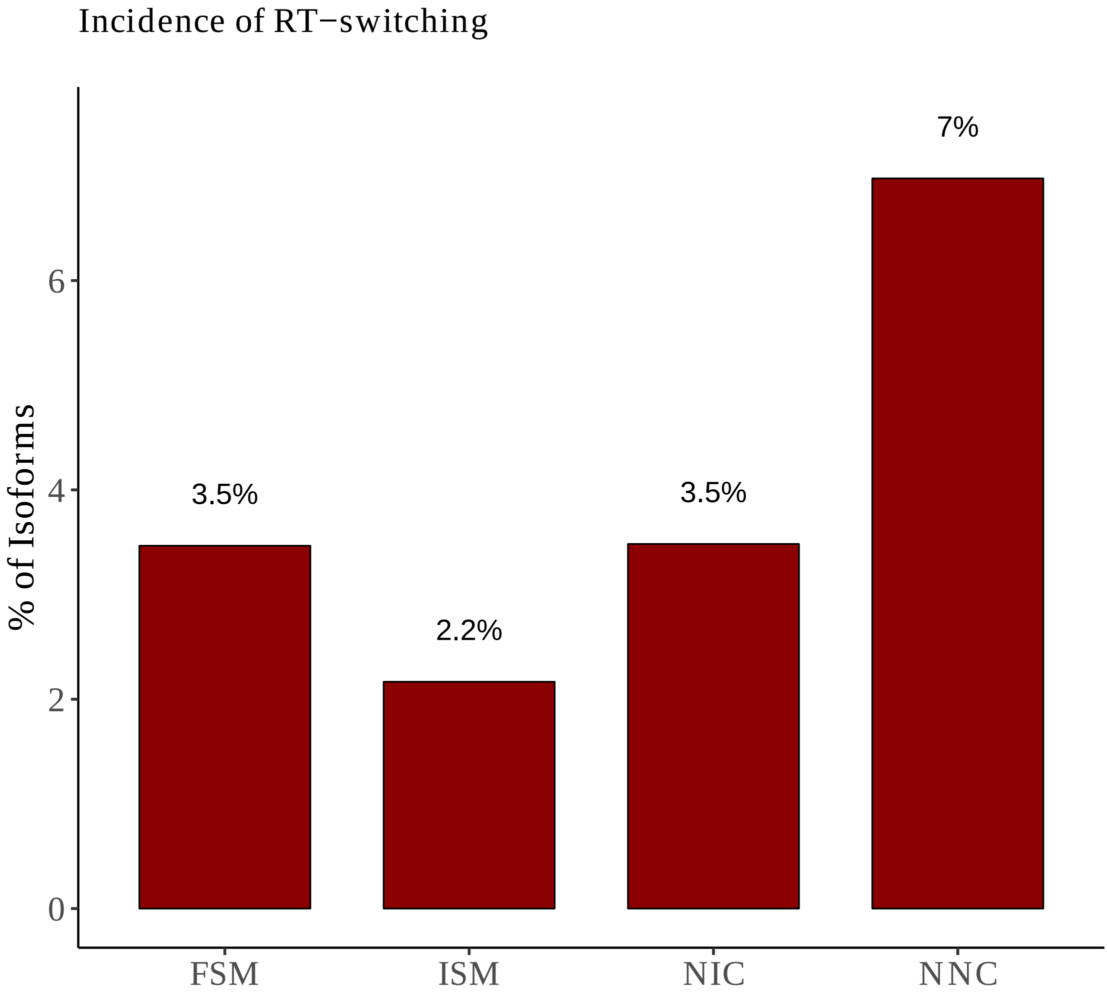
Percent of genomic 'A's in downstream 20 bp

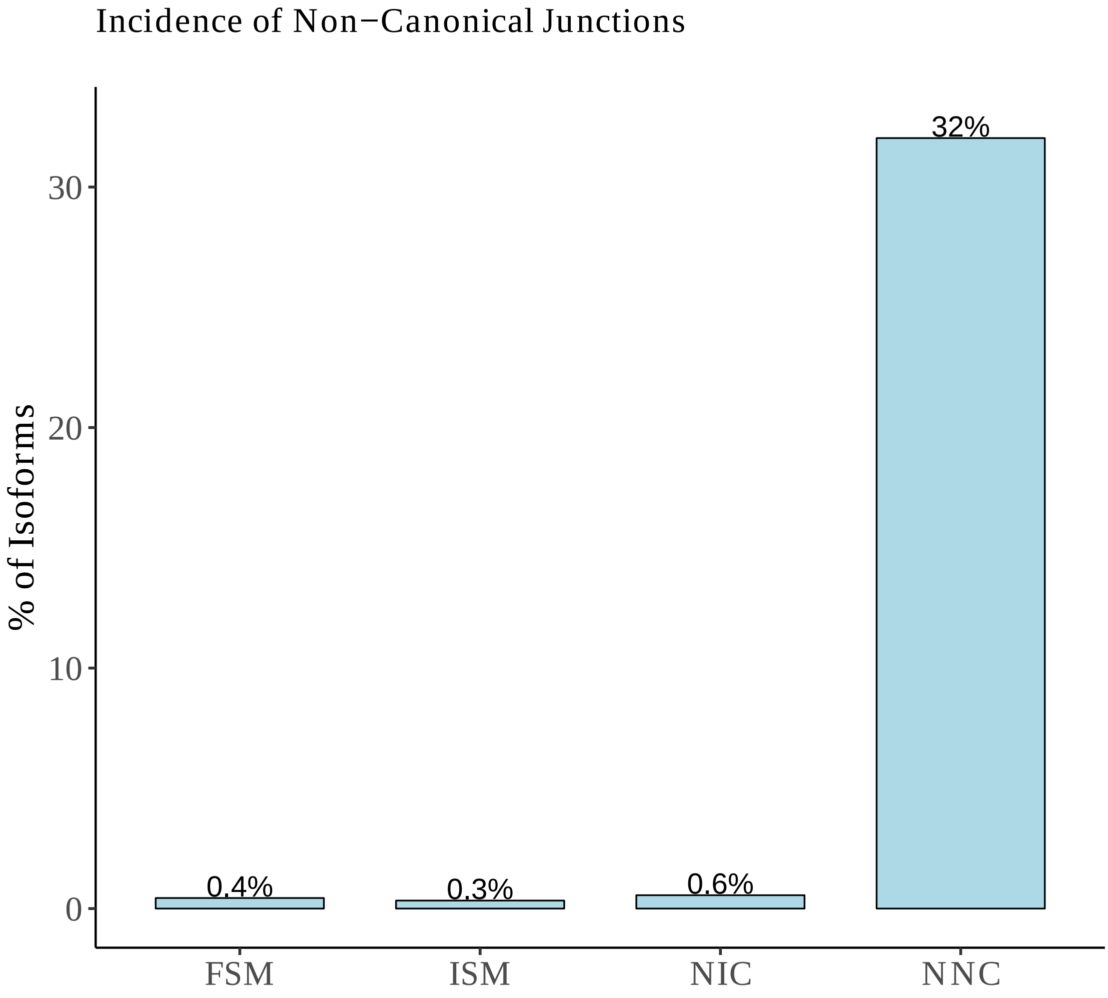
Possible Intra−Priming, Coding vs Non−Coding

Percent of genomic 'A's in downstream 20 bp



Quality Controls


Incidence of RT−switching
% of Isoforms
0
2
4
6
3.5%
2.2%
3.5%
7%
FSM
ISM
NIC
NNC


Incidence of Non−Canonical Junctions
% of Isoforms
0
10
20
30
0.4%
0.3%
0.6%
32%
FSM
ISM
NIC
NNC