SQANTI2 report

Version	7.3.2
Input	flair.collapse.isoforms.gtf
Annotation	Homo_sapiens.GRCh38.99.gtf
Genome	Homo_sapiens.GRCh38.dna.primary_assembly.fa
Aligner	minimap2
FLCount	NA
Expression	NA
Junction	NA
CagePeak	NA
PolyA	NA
PolyAPeak	NA
IsFusion	False

Unique Genes: 20576

Unique Isoforms: 61069

Gene classification

Category	# Genes
Annotated Genes	15306
Novel Genes	5270

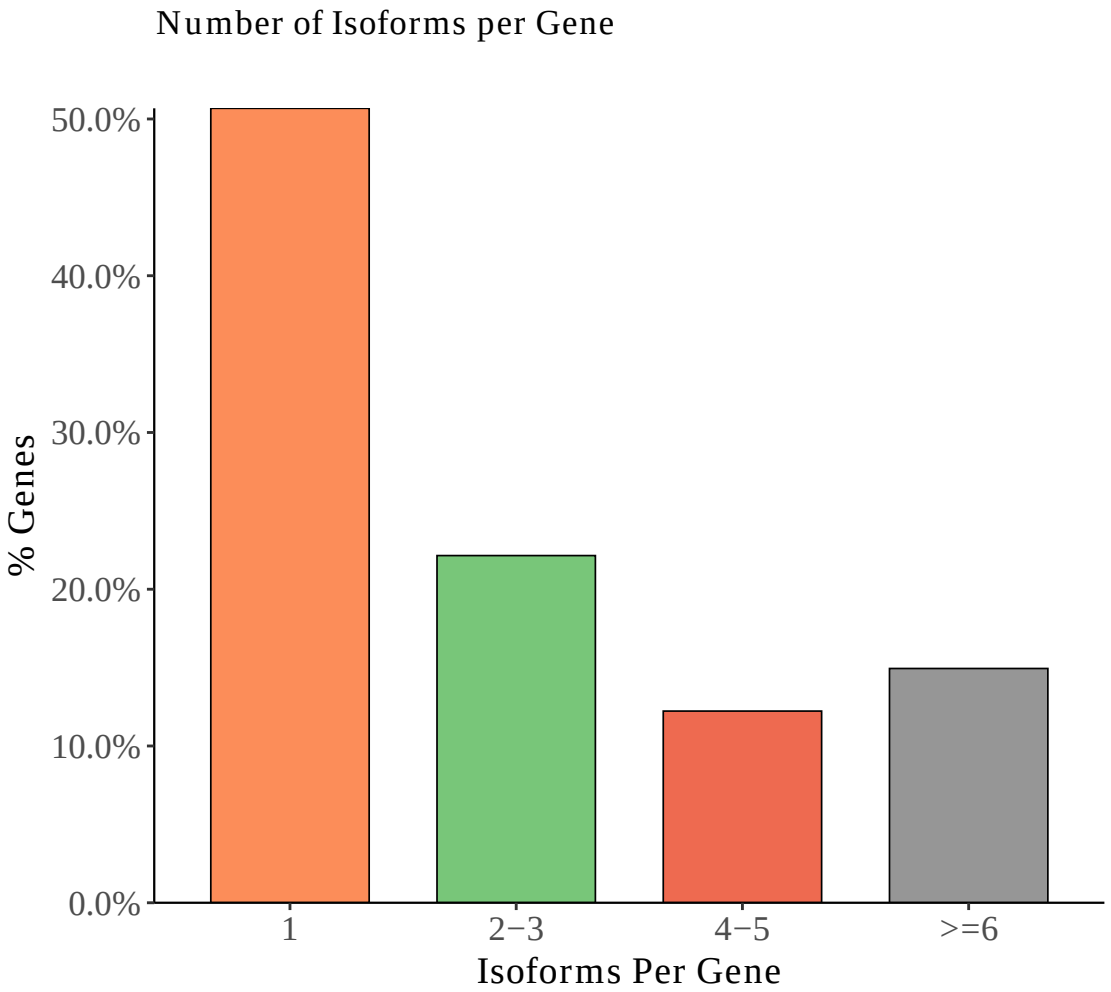
Splice Junction Classification

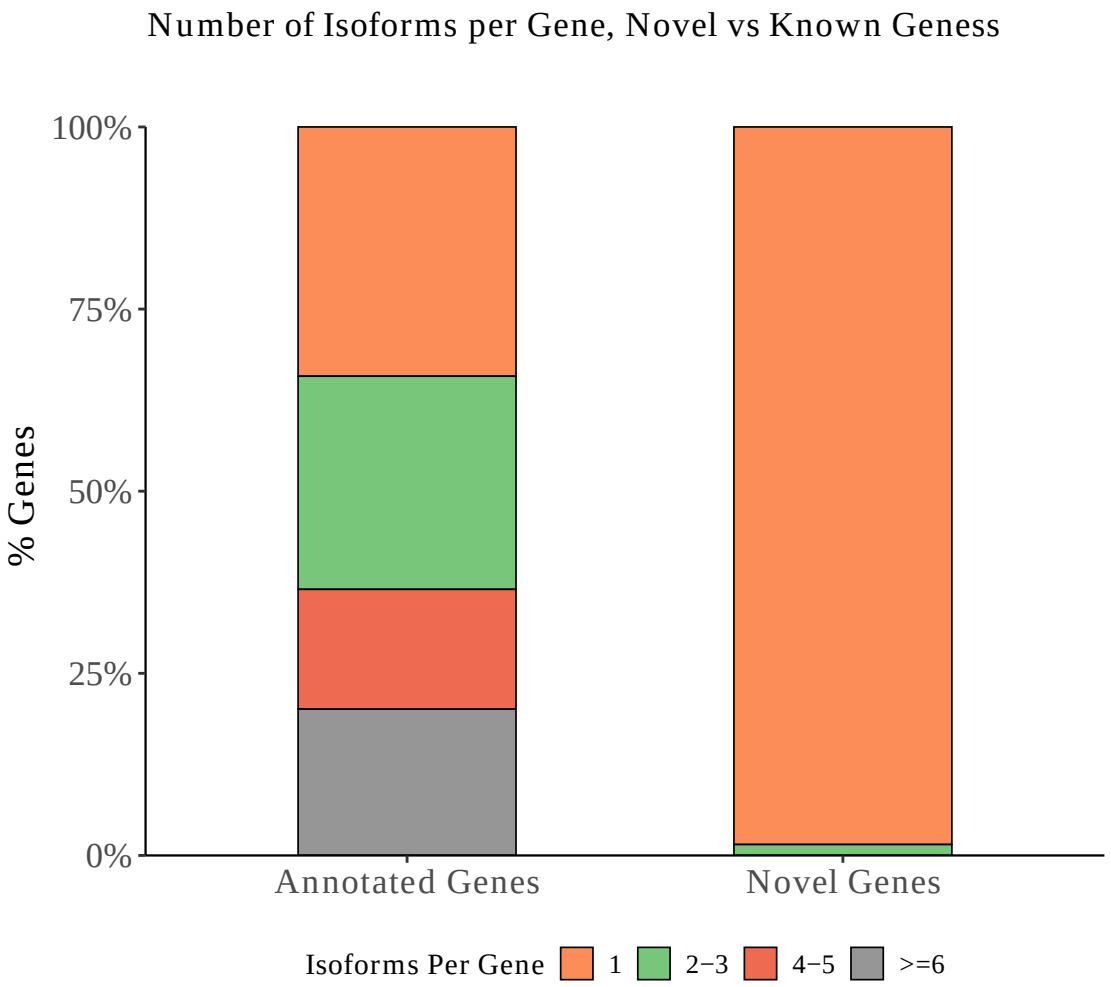
Category	# SJs	Percent
Known canonical	138490	95.32
Known Non−canonical	134	0.09
Novel canonical	6543	4.50
Novel Non−canonical	124	0.09

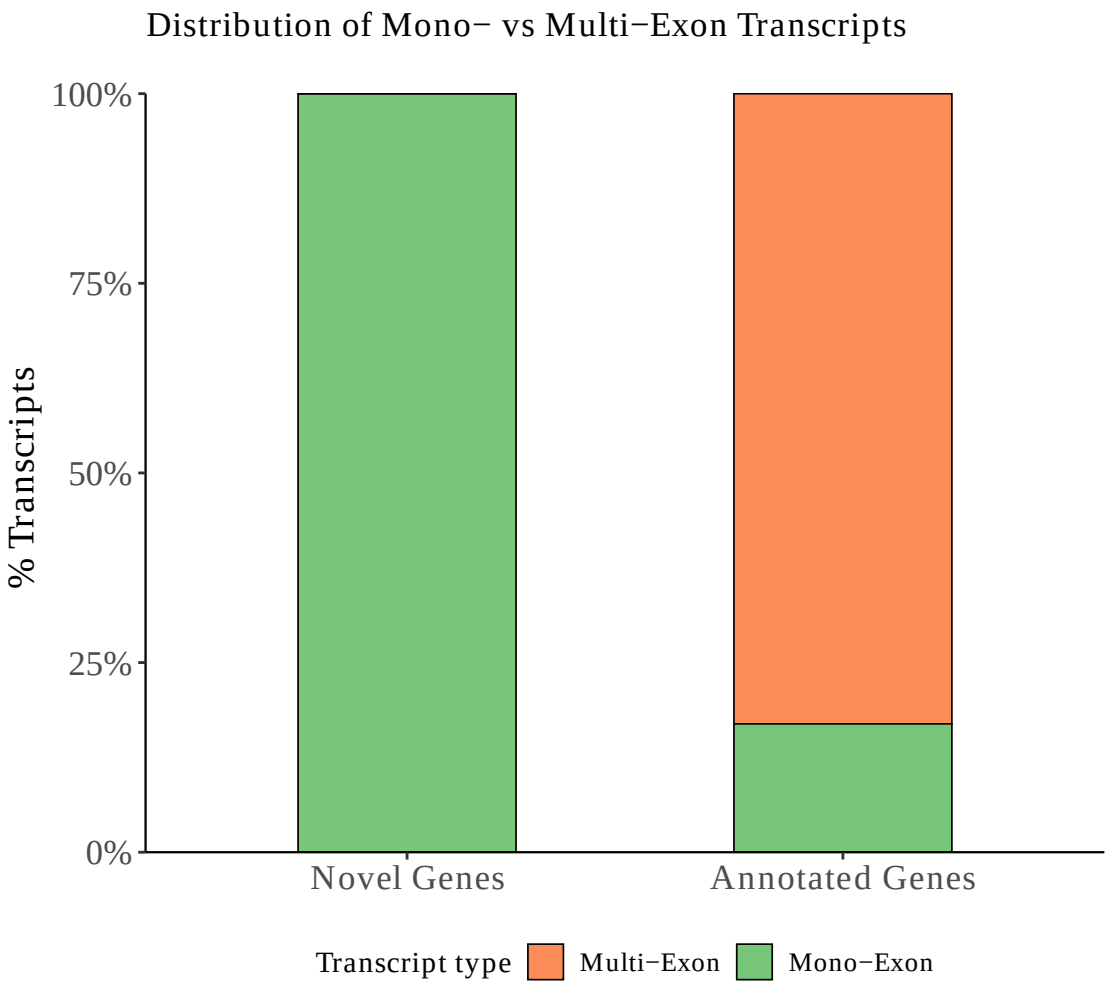
Characterization of transcripts based on splice junctions

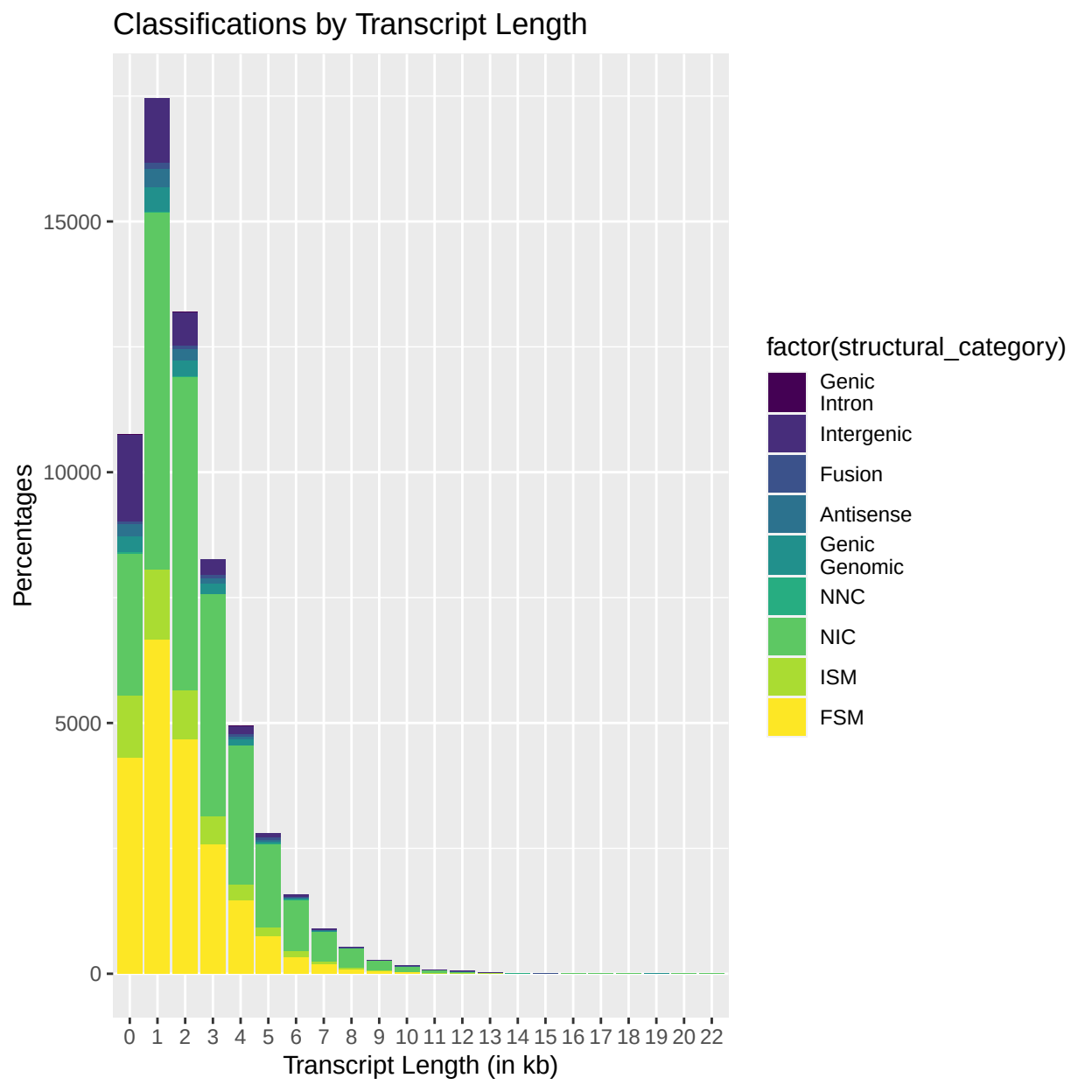
Category	# Isoforms	# Genes
FSM	21173	12513
ISM	4919	3780
NIC	27495	8576
NNC	72	58
Genic Genomic	1546	1437
Antisense	1071	954
Fusion	476	346
Intergenic	4313	4313
Genic Intron	4	4

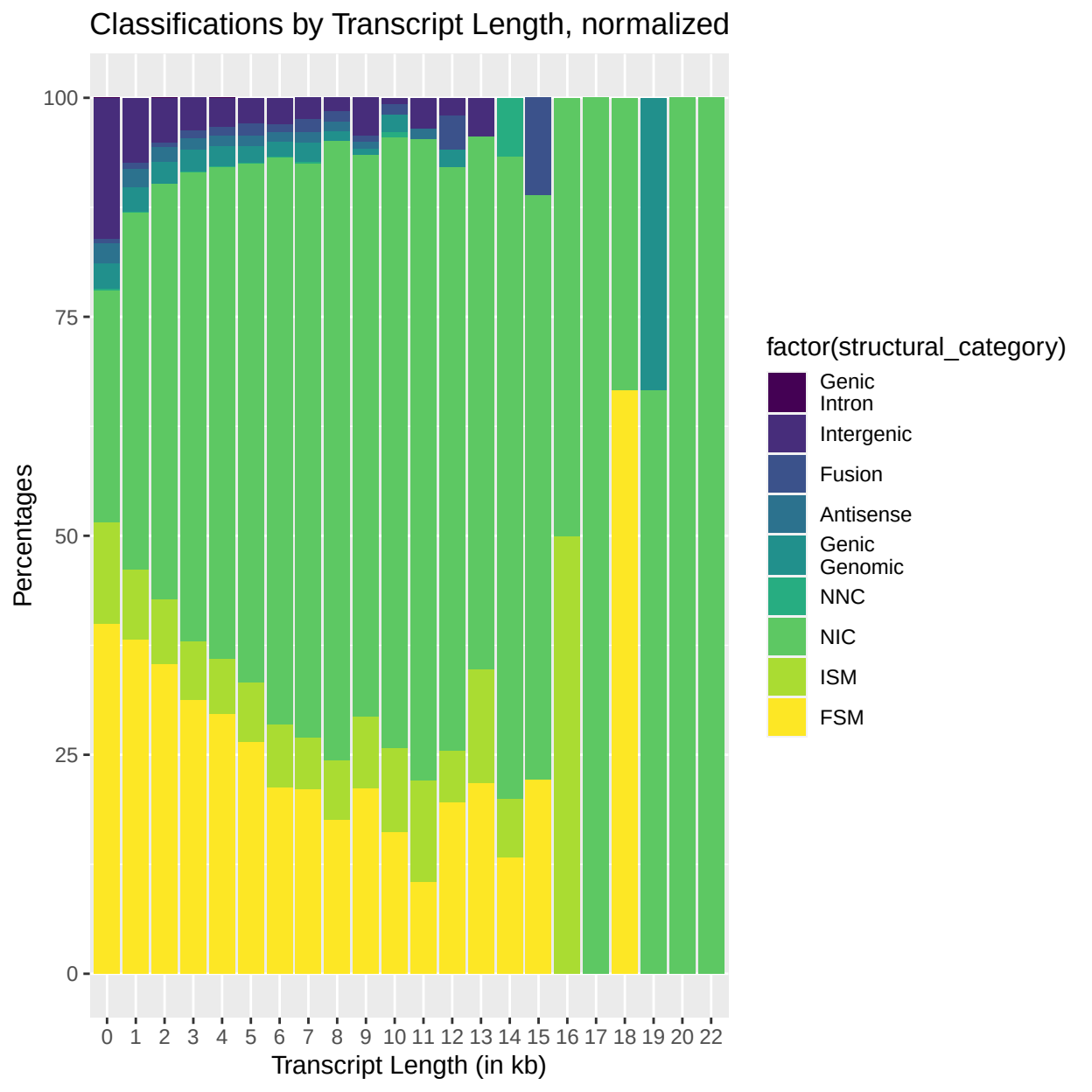
Gene Characterization

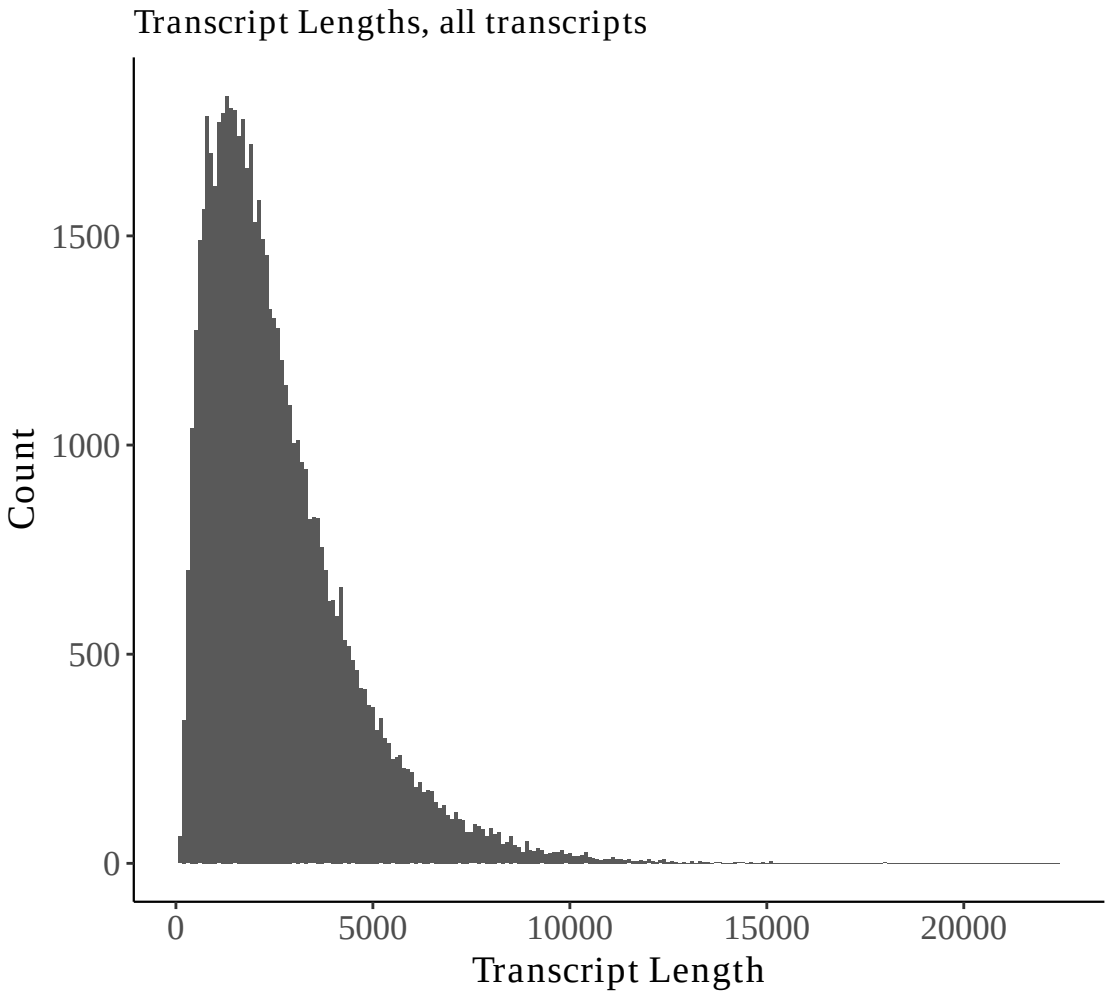

Number of Isoforms per Gene
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
% Genes
1
2−3
4−5
>=6
Isoforms Per Gene

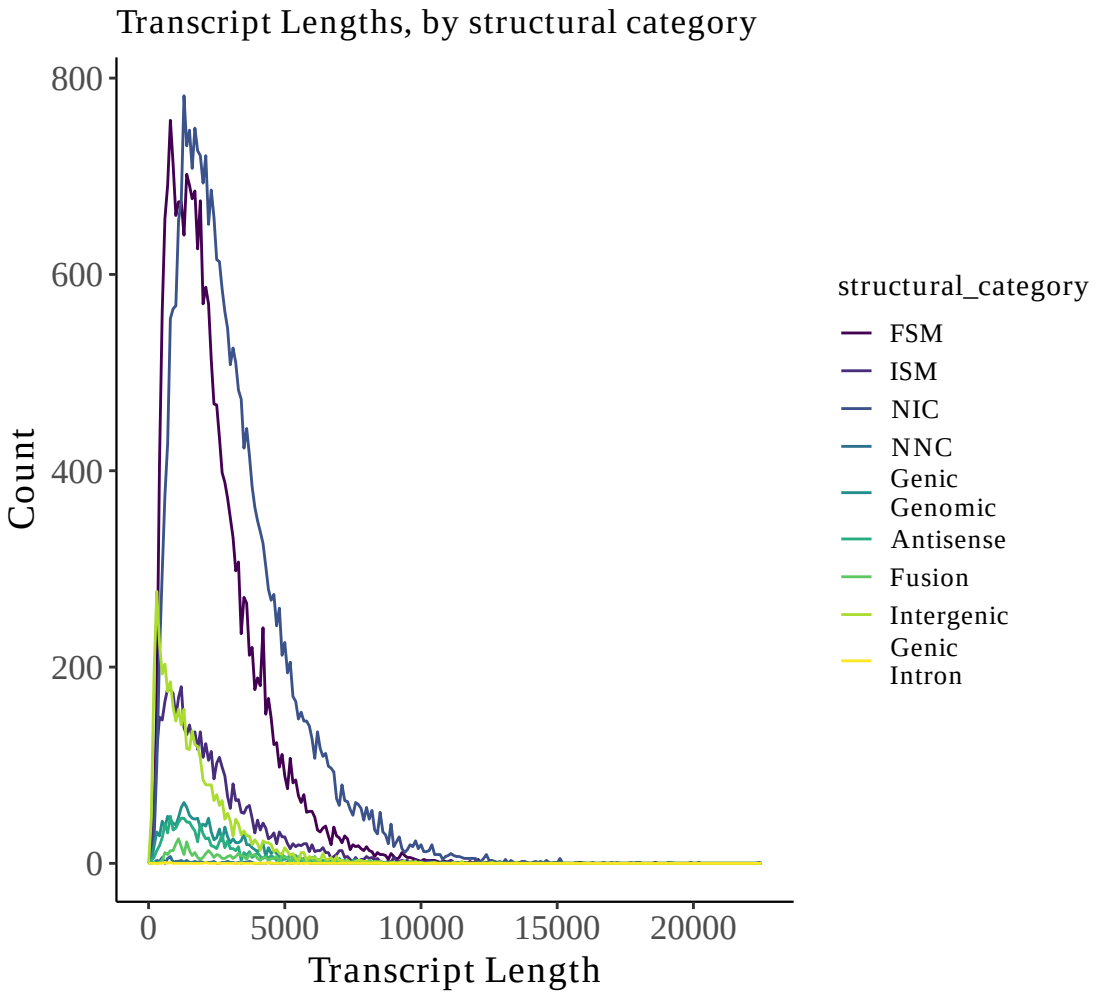

Number of Isoforms per Gene, Novel vs Known Geness
100%
75%
50%
25%
0%
% Genes
Annotated Genes
Novel Genes
Isoforms Per Gene
1
2−3
4−5
>=6

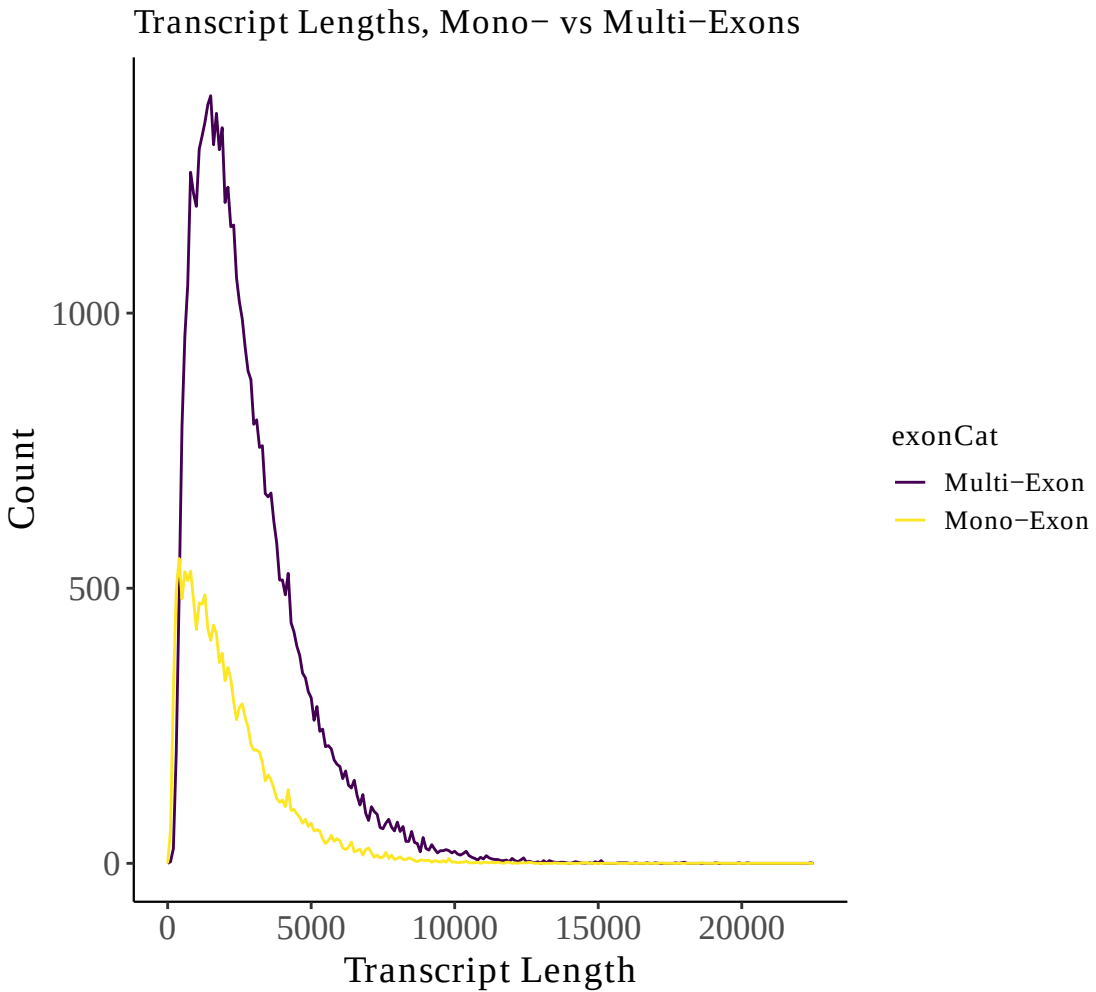

Distribution of Mono− vs Multi−Exon Transcripts
100%
75%
50%
25%
0%
% Transcripts
Novel Genes
Annotated Genes
Transcript type
Multi−Exon
Mono−Exon


Classifications by Transcript Length
Percentages
15000
10000
5000
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 22
Transcript Length (in kb)
factor(structural_category)
Genic Intron
Intergenic
Fusion
Antisense
Genic Genomic
NNC
NIC
ISM
FSM

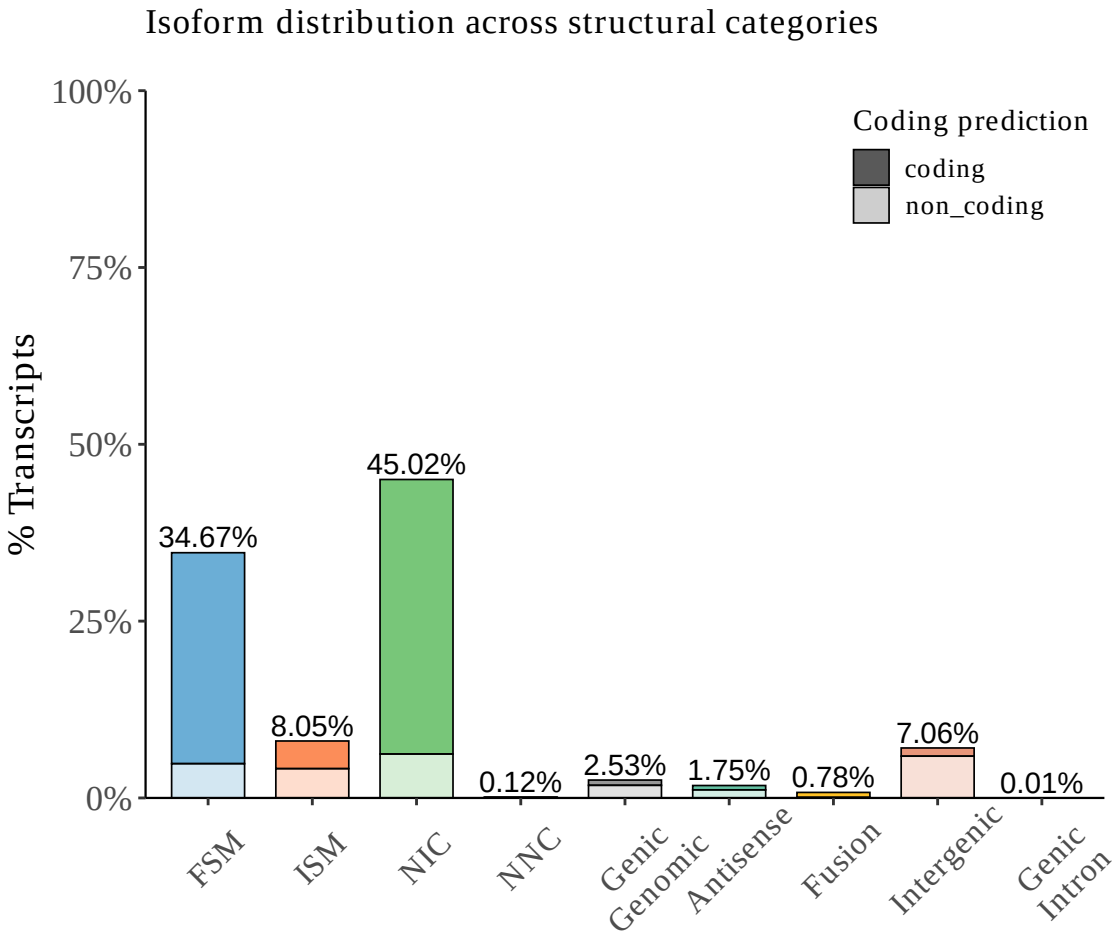

Classifications by Transcript Length, normalized
Percentages
100
75
50
25
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 22
Transcript Length (in kb)
factor(structural_category)
Genic Intron
Intergenic
Fusion
Antisense
Genic Genomic
NNC
NIC
ISM
FSM

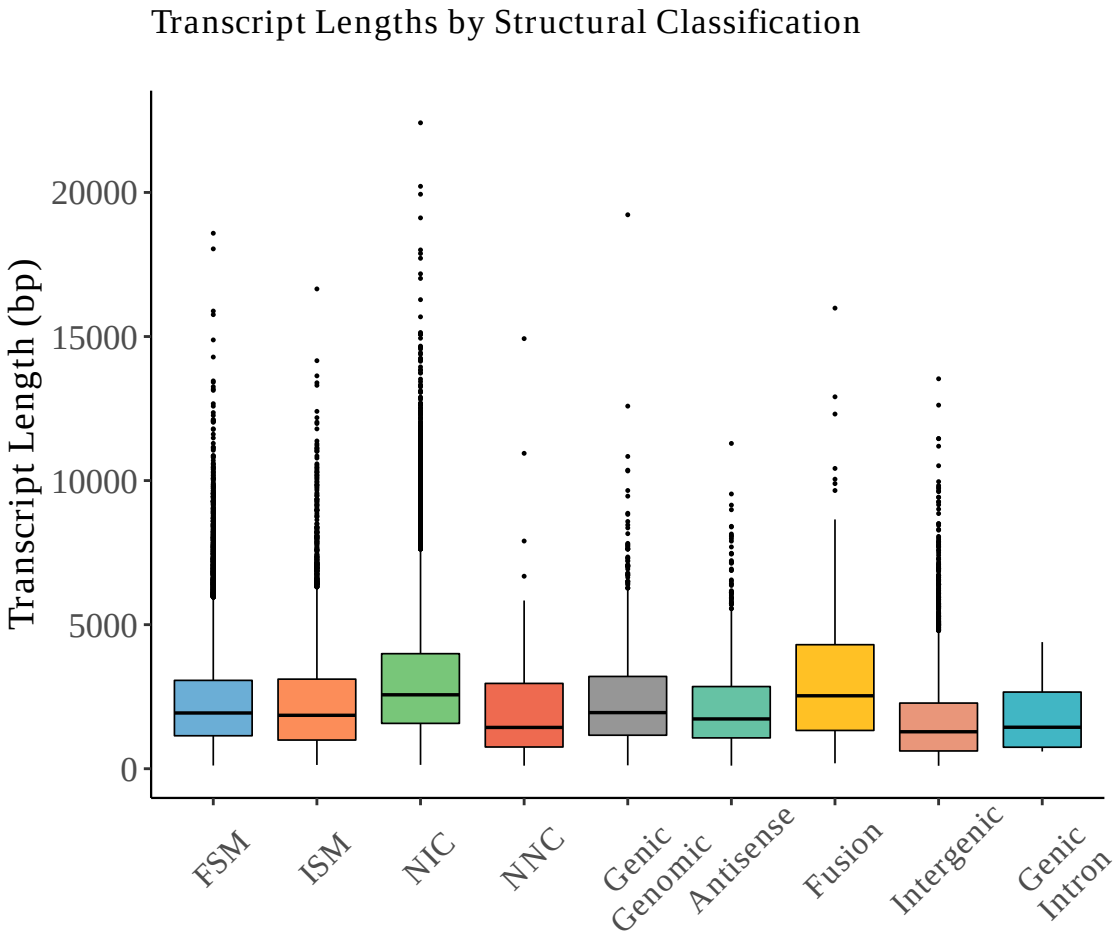

Transcript Lengths, all transcripts
Count
1500
1000
500
0
0
5000
10000
15000
20000
Transcript Length

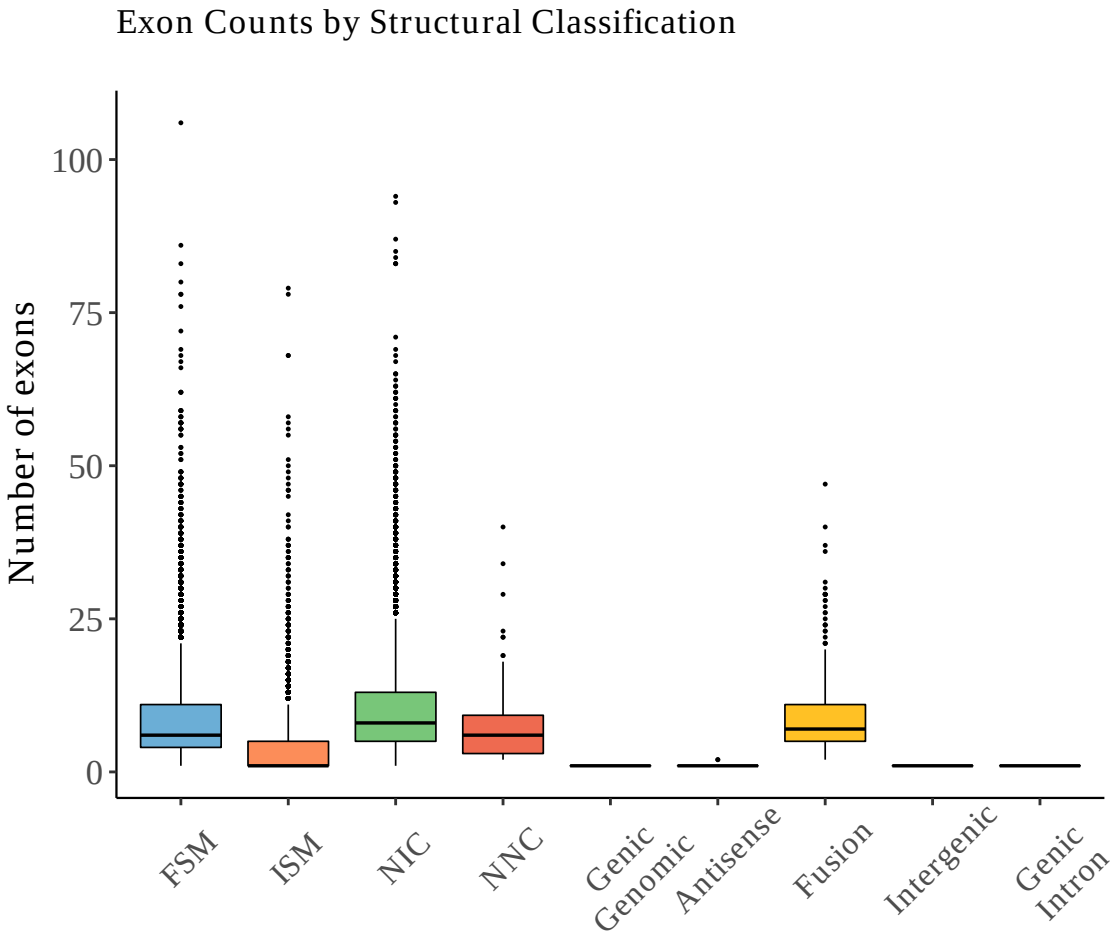

Transcript Lengths, by structural category
800
600
400
200
0
Count
0
5000
10000
15000
20000
Transcript Length
structural_category
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron


Transcript Lengths, Mono− vs Multi−Exons
Count
1000
500
0
0
5000
10000
15000
20000
Transcript Length
exonCat
Multi−Exon
Mono−Exon

Structural Isoform Characterization by Splice Junctions

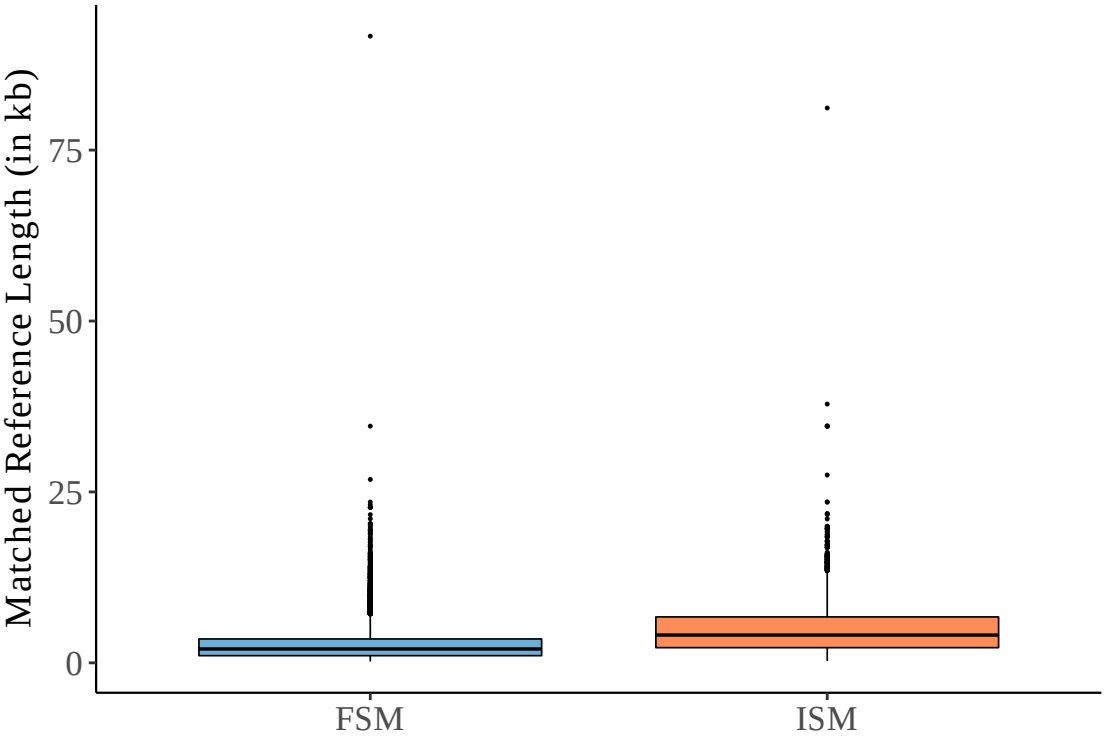

Isoform distribution across structural categories
100%
75%
50%
25%
0%
% Transcripts
Coding prediction
coding
non_coding
34.67%
8.05%
45.02%
0.12%
2.53%
1.75%
0.78%
7.06%
0.01%
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

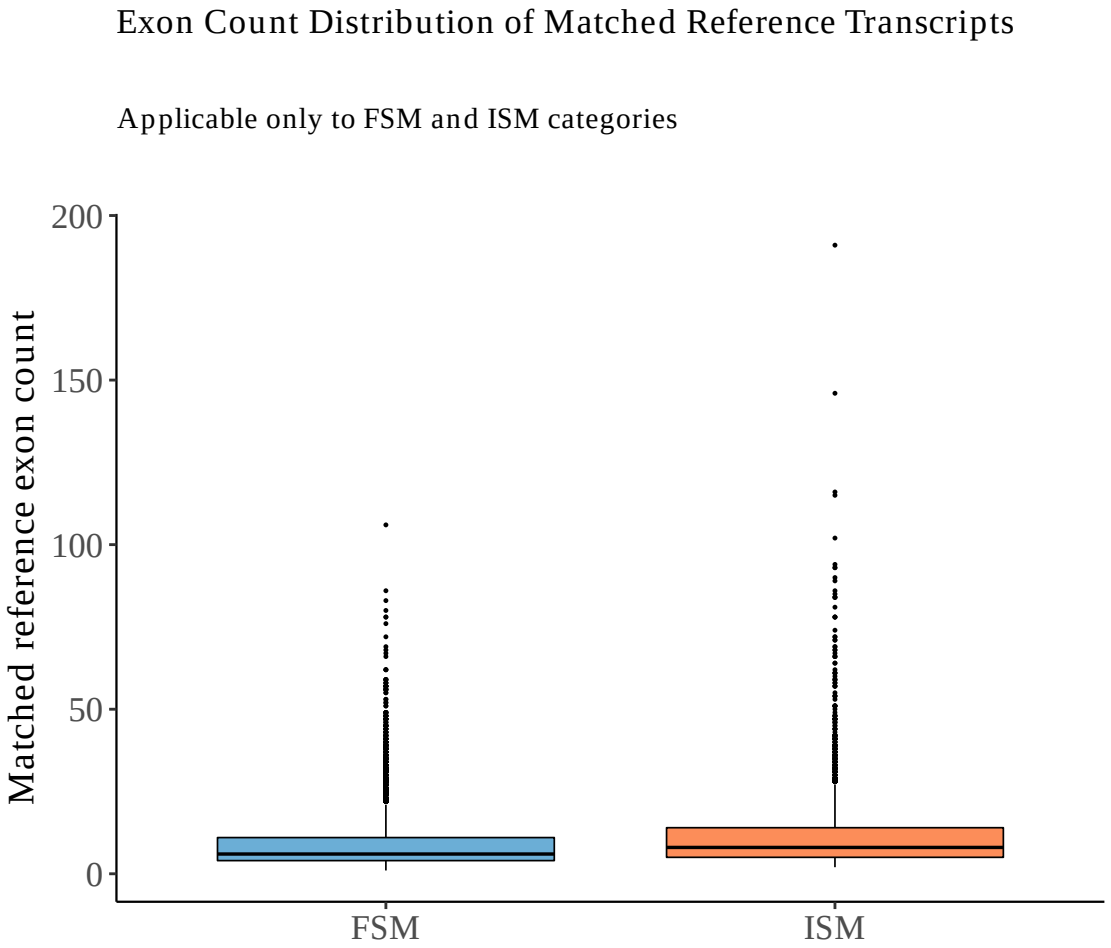

Transcript Lengths by Structural Classification
Transcript Length (bp)
20000
15000
10000
5000
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron


Exon Counts by Structural Classification
Number of exons
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron

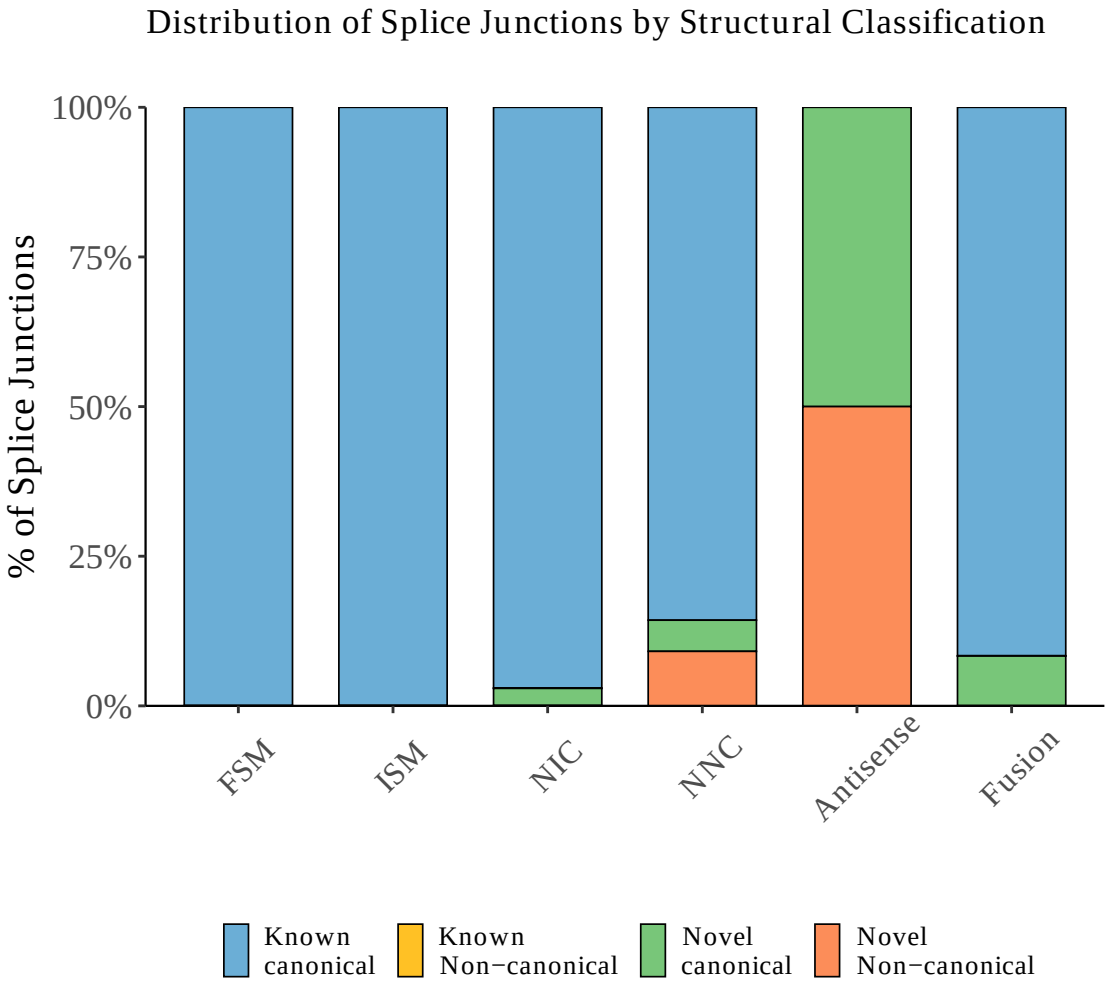
Length Distribution of Matched Reference Transcripts

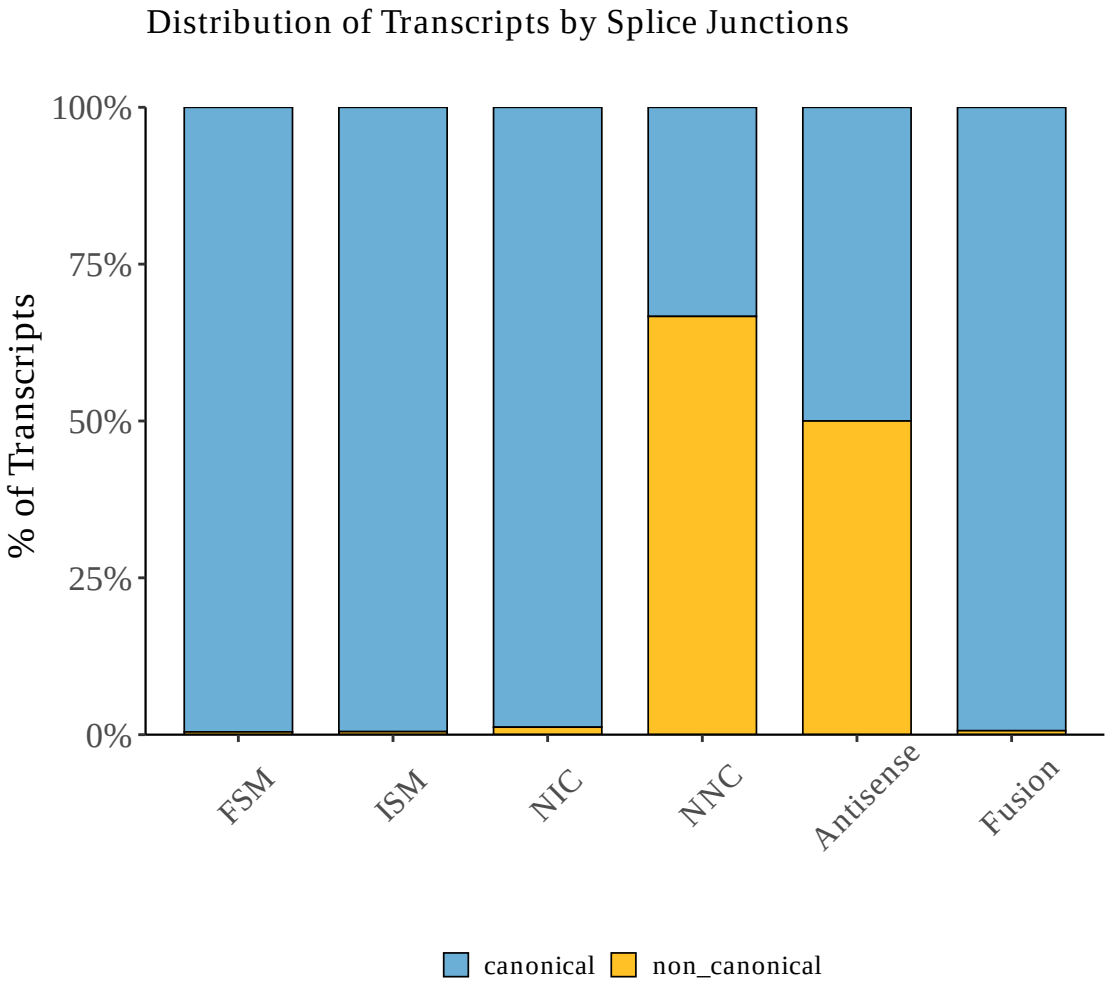
Applicable only to FSM and ISM categories

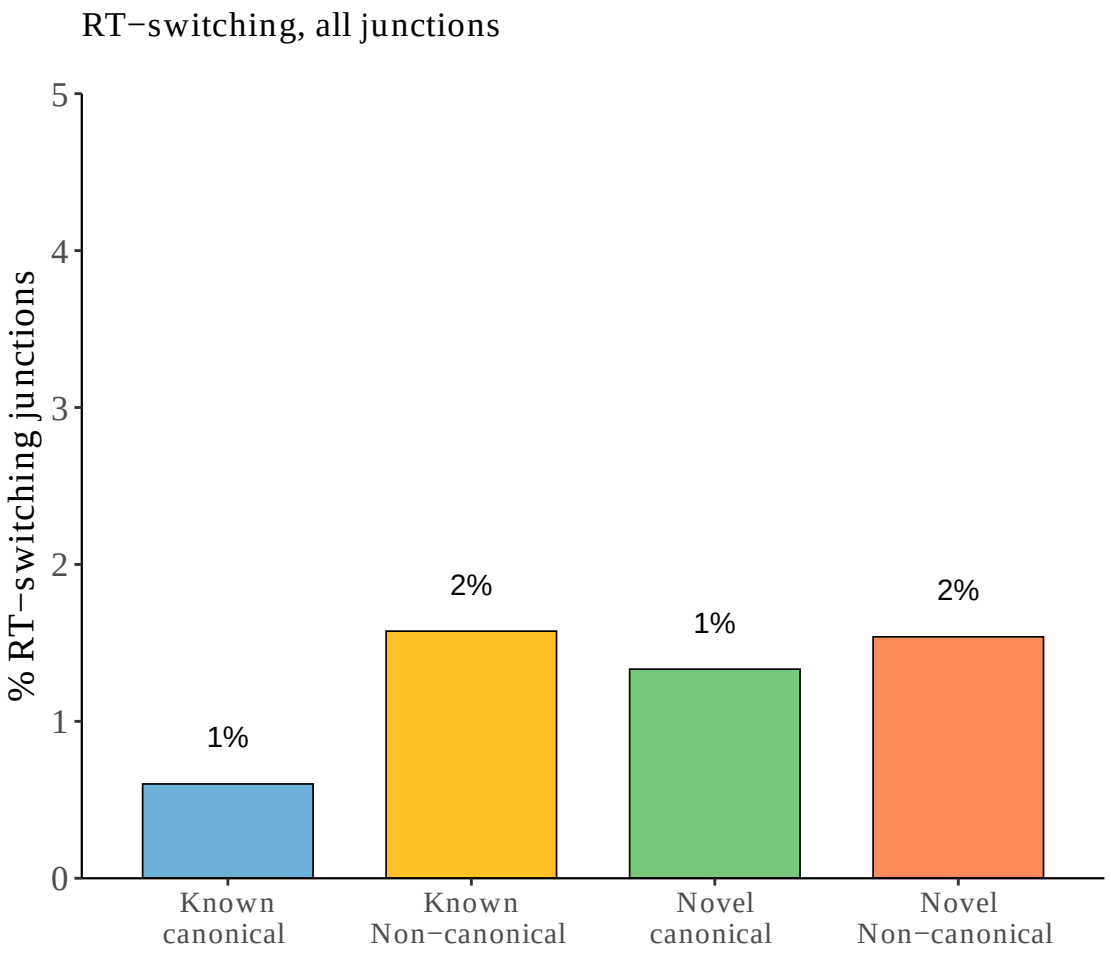


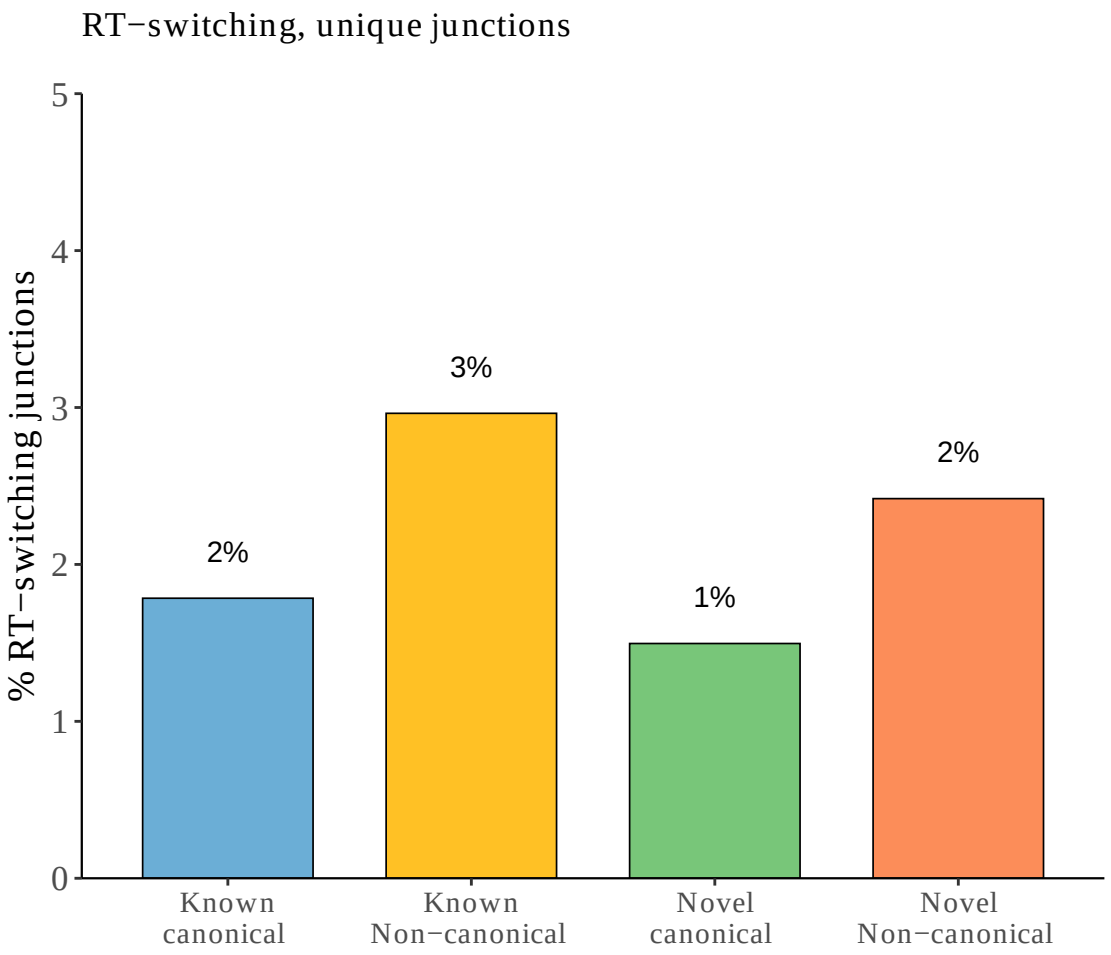

Exon Count Distribution of Matched Reference Transcripts
Applicable only to FSM and ISM categories
Matched reference exon count
200
150
100
50
0
FSM
ISM

Splice Junction Characterization

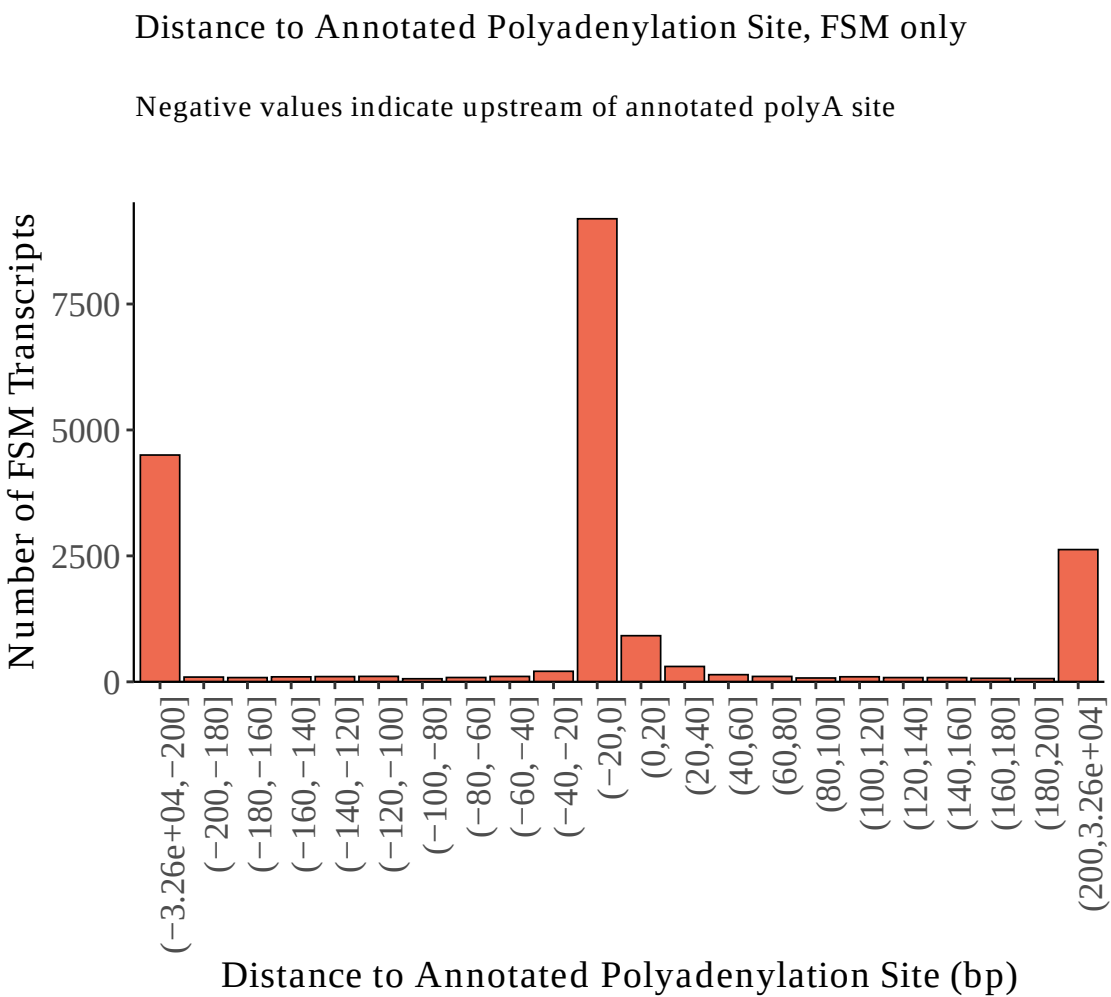

Distribution of Splice Junctions by Structural Classification
% of Splice Junctions
100%
75%
50%
25%
0%
FSM
ISM
NIC
NNC
Antisense
Fusion
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical


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


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

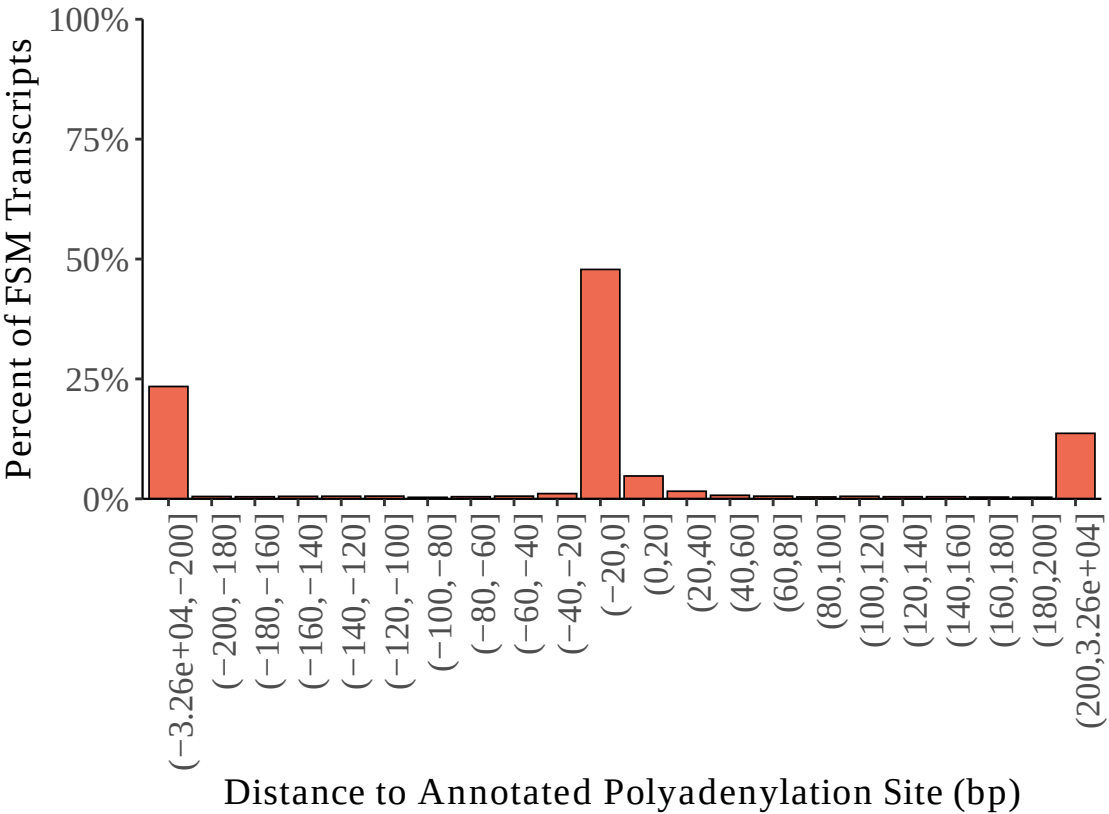

RT−switching, unique junctions
% RT−switching junctions
5
4
3
2
1
0
2%
3%
1%
2%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

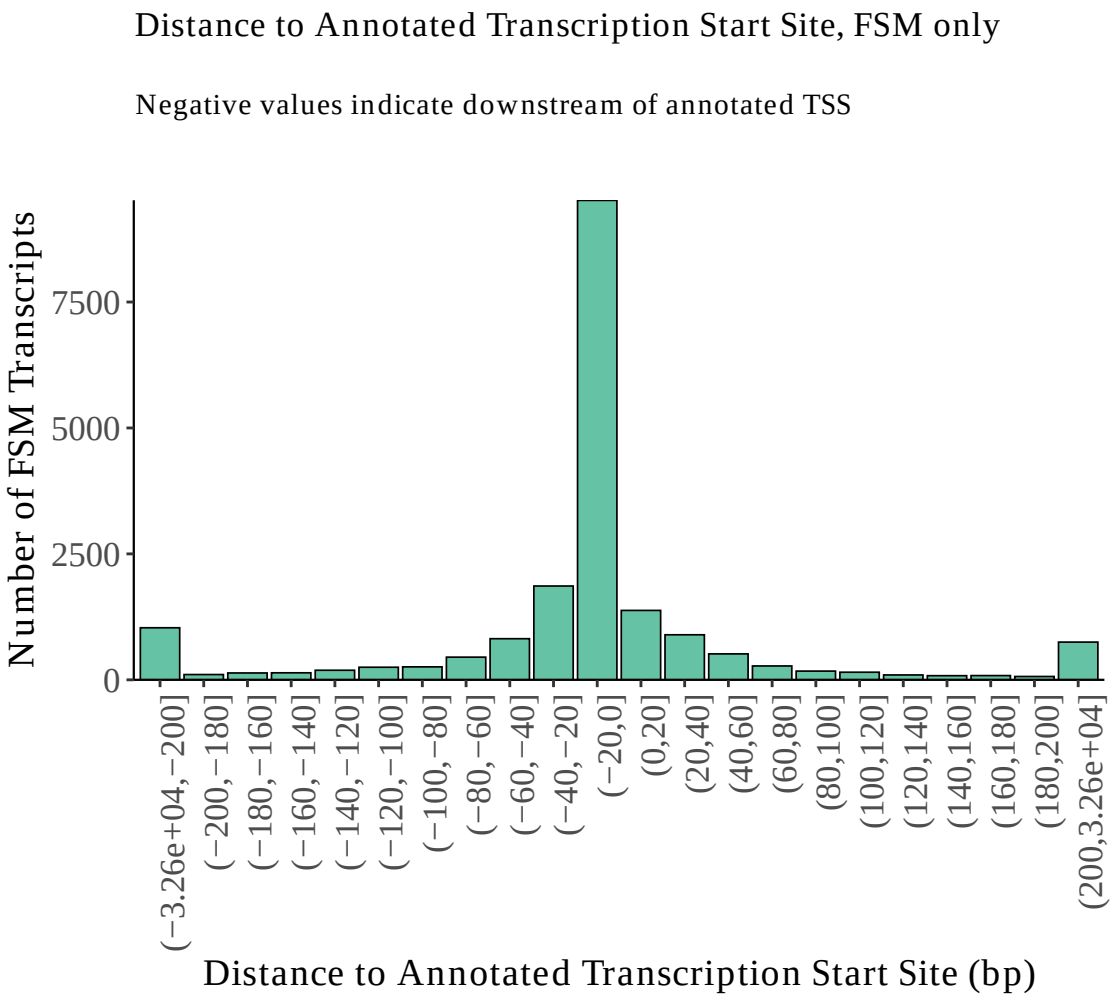
Comparison with Annotated TSS and PolyA Sites

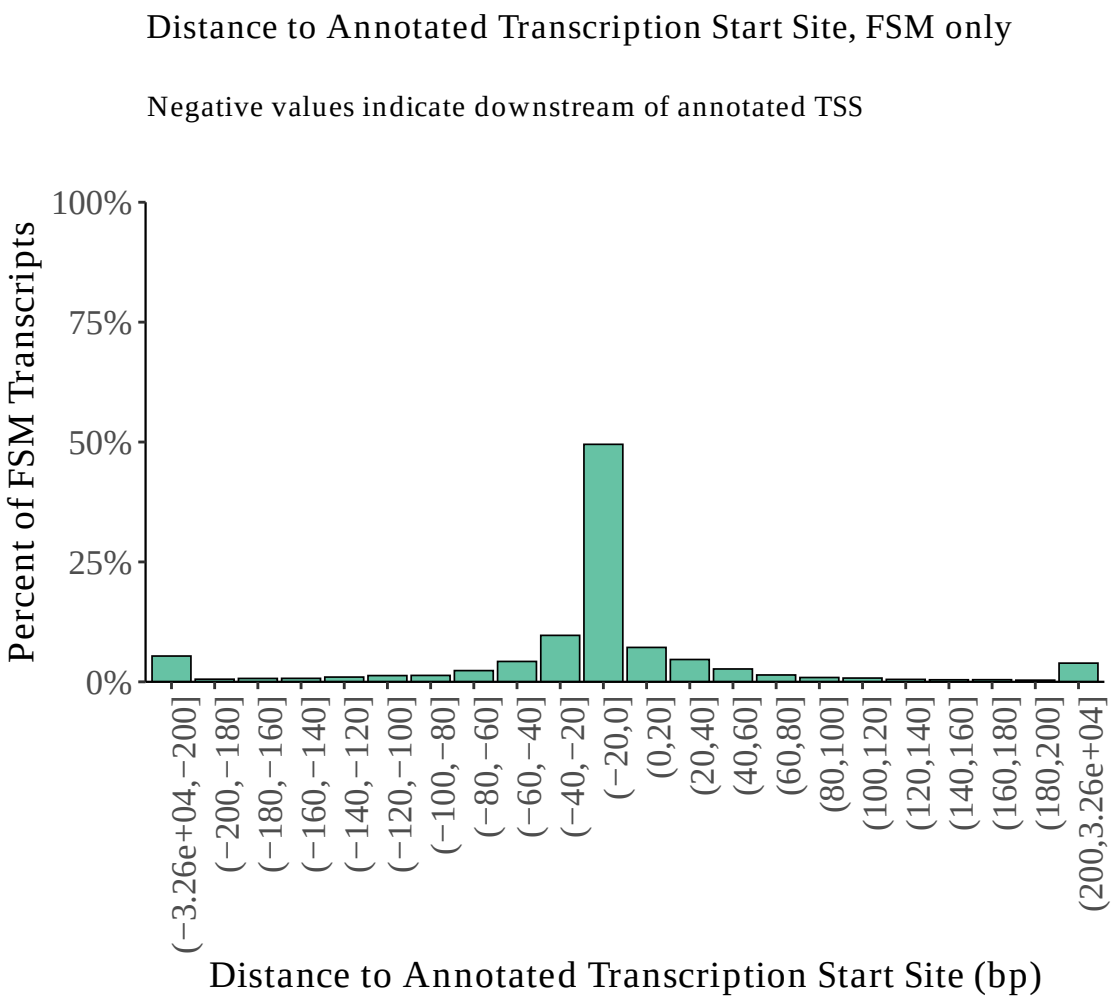

Distance to Annotated Polyadenylation Site, FSM only
Negative values indicate upstream of annotated polyA site
Number of FSM Transcripts
0
2500
5000
7500
(−3.26e+04,−200]
(−200,−180]
(−180,−160]
(−160,−140]
(−140,−120]
(−120,−100]
(−100,−80]
(−80,−60]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(40,60]
(60,80]
(80,100]
(100,120]
(120,140]
(140,160]
(160,180]
(180,200]
(200,3.26e+04]
Distance to Annotated Polyadenylation Site (bp)

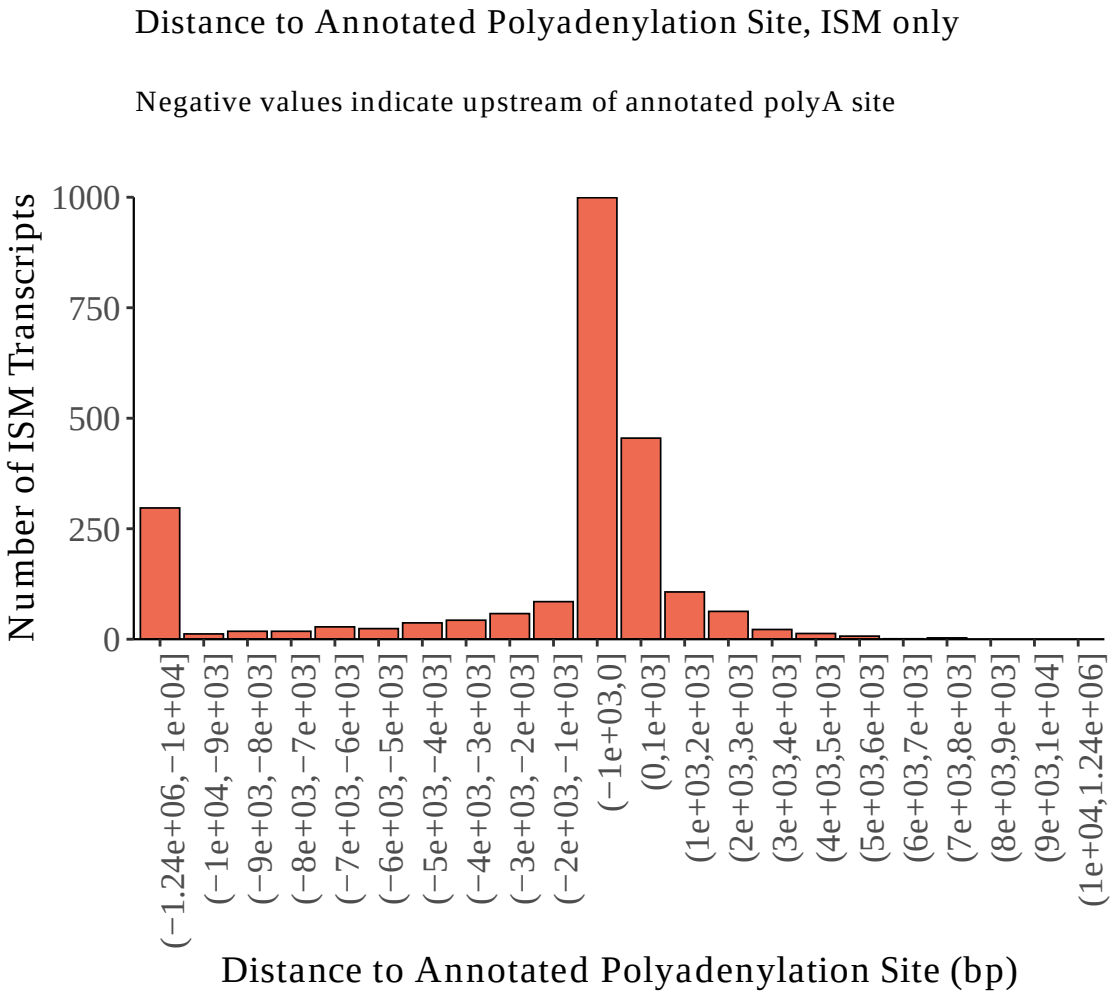
Distance to Annotated Polyadenylation Site, FSM only

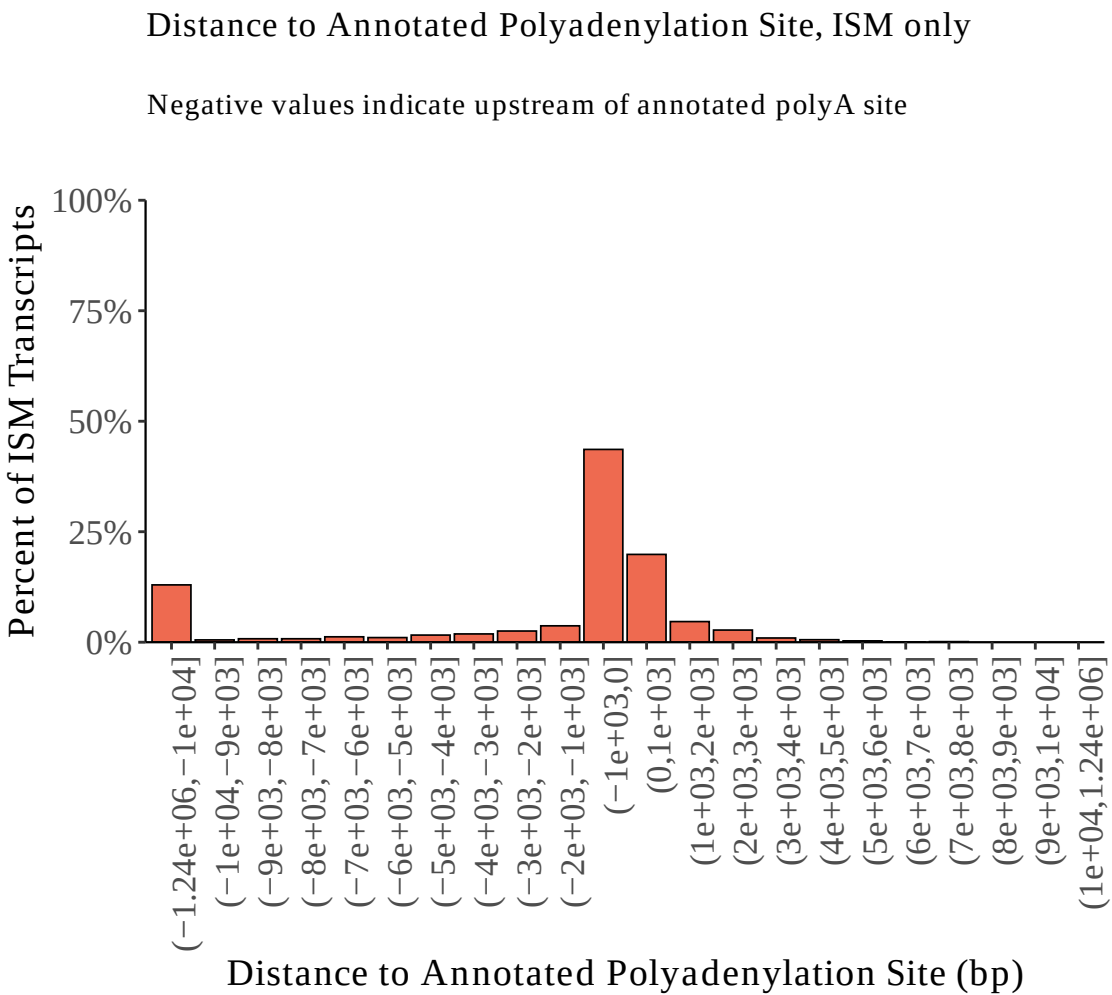
Negative values indicate upstream of annotated polyA site




Distance to Annotated Transcription Start Site, FSM only
Negative values indicate downstream of annotated TSS
Number of FSM Transcripts
0
2500
5000
7500
(−3.26e+04,−200]
(−200,−180]
(−180,−160]
(−160,−140]
(−140,−120]
(−120,−100]
(−100,−80]
(−80,−60]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(40,60]
(60,80]
(80,100]
(100,120]
(120,140]
(140,160]
(160,180]
(180,200]
(200,3.26e+04]
Distance to Annotated Transcription Start Site (bp)

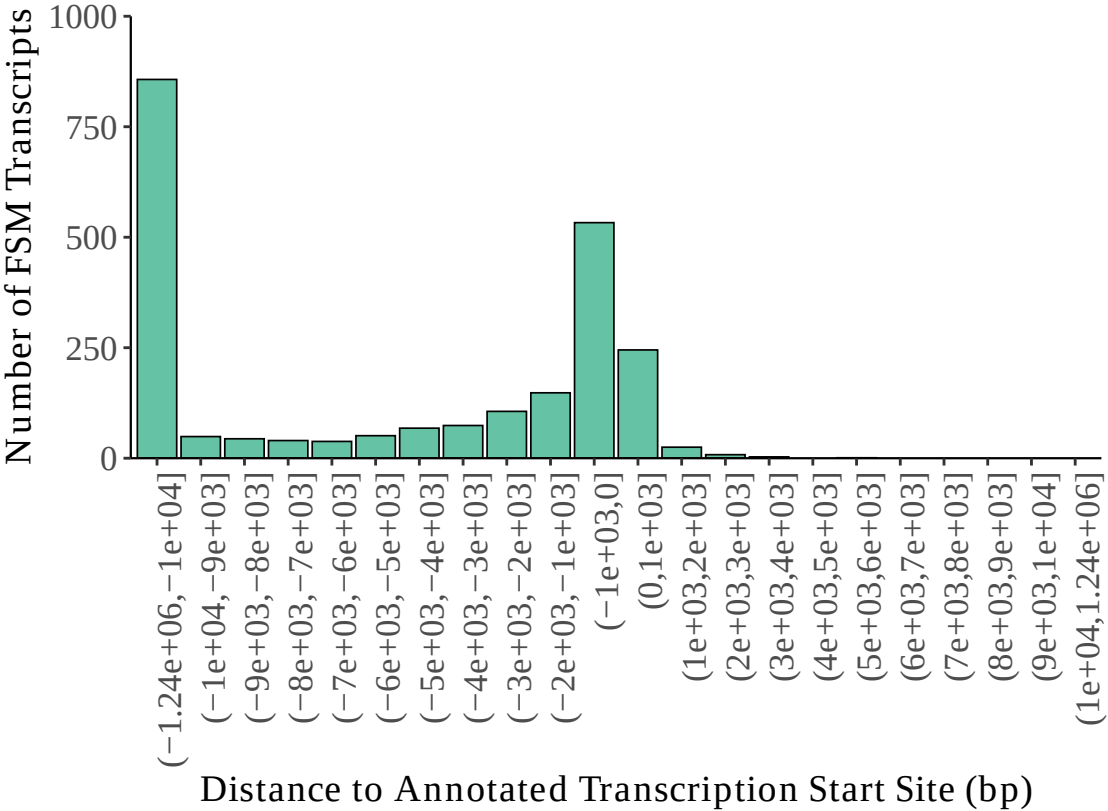

Distance to Annotated Transcription Start Site, FSM only
Negative values indicate downstream of annotated TSS
Percent of FSM Transcripts
100%
75%
50%
25%
0%
(−3.26e+04,−200]
(−200,−180]
(−180,−160]
(−160,−140]
(−140,−120]
(−120,−100]
(−100,−80]
(−80,−60]
(−60,−40]
(−40,−20]
(−20,0]
(0,20]
(20,40]
(40,60]
(60,80]
(80,100]
(100,120]
(120,140]
(140,160]
(160,180]
(180,200]
(200,3.26e+04]
Distance to Annotated Transcription Start Site (bp)

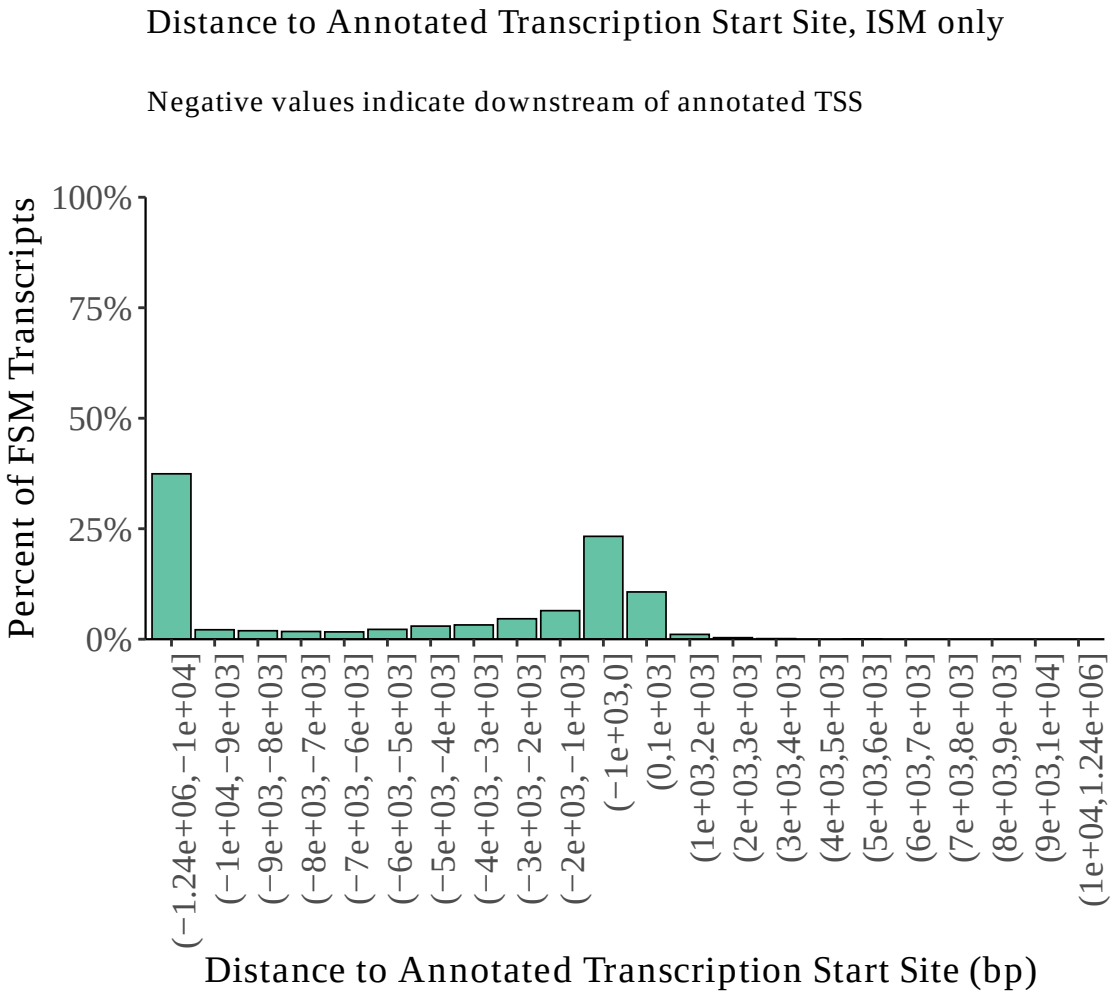

Distance to Annotated Polyadenylation Site, ISM only
Negative values indicate upstream of annotated polyA site
Number of ISM Transcripts
1000
750
500
250
0
(−1.24e+06,−1e+04]
(−1e+04,−9e+03]
(−9e+03,−8e+03]
(−8e+03,−7e+03]
(−7e+03,−6e+03]
(−6e+03,−5e+03]
(−5e+03,−4e+03]
(−4e+03,−3e+03]
(−3e+03,−2e+03]
(−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+04,1.24e+06]
Distance to Annotated Polyadenylation Site (bp)


Distance to Annotated Polyadenylation Site, ISM only
Negative values indicate upstream of annotated polyA site
Percent of ISM Transcripts
100%
75%
50%
25%
0%
(−1.24e+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+04,1.24e+06]
Distance to Annotated Polyadenylation Site (bp)

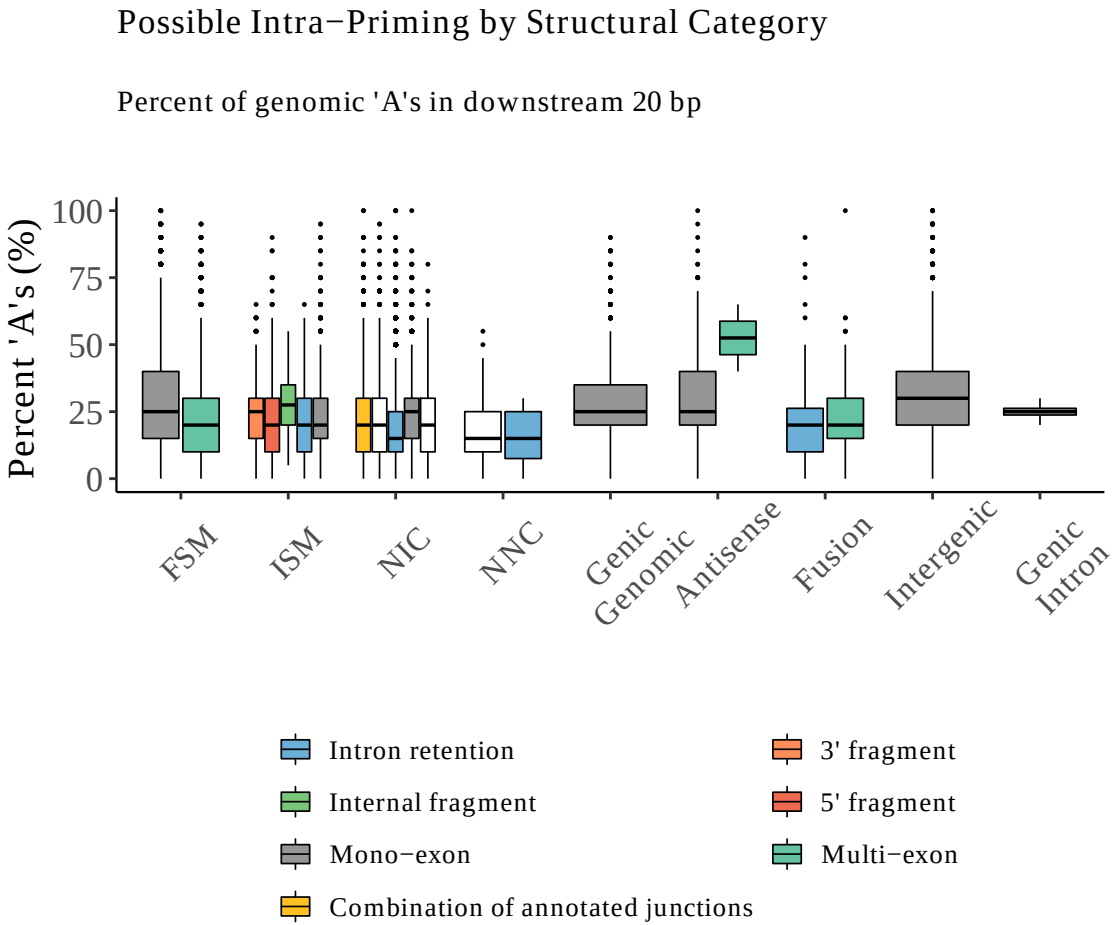
Distance to Annotated Transcription Start Site, ISM only

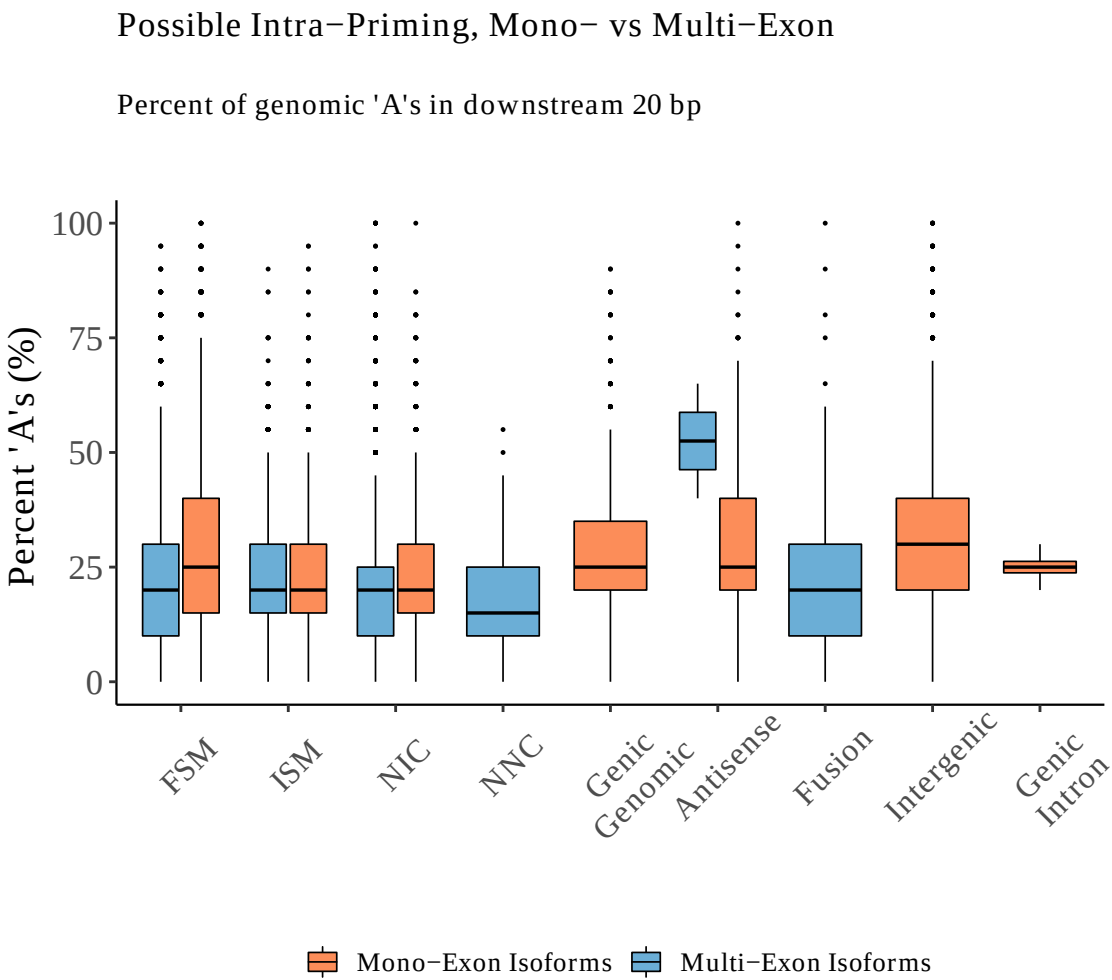
Negative values indicate downstream of annotated TSS

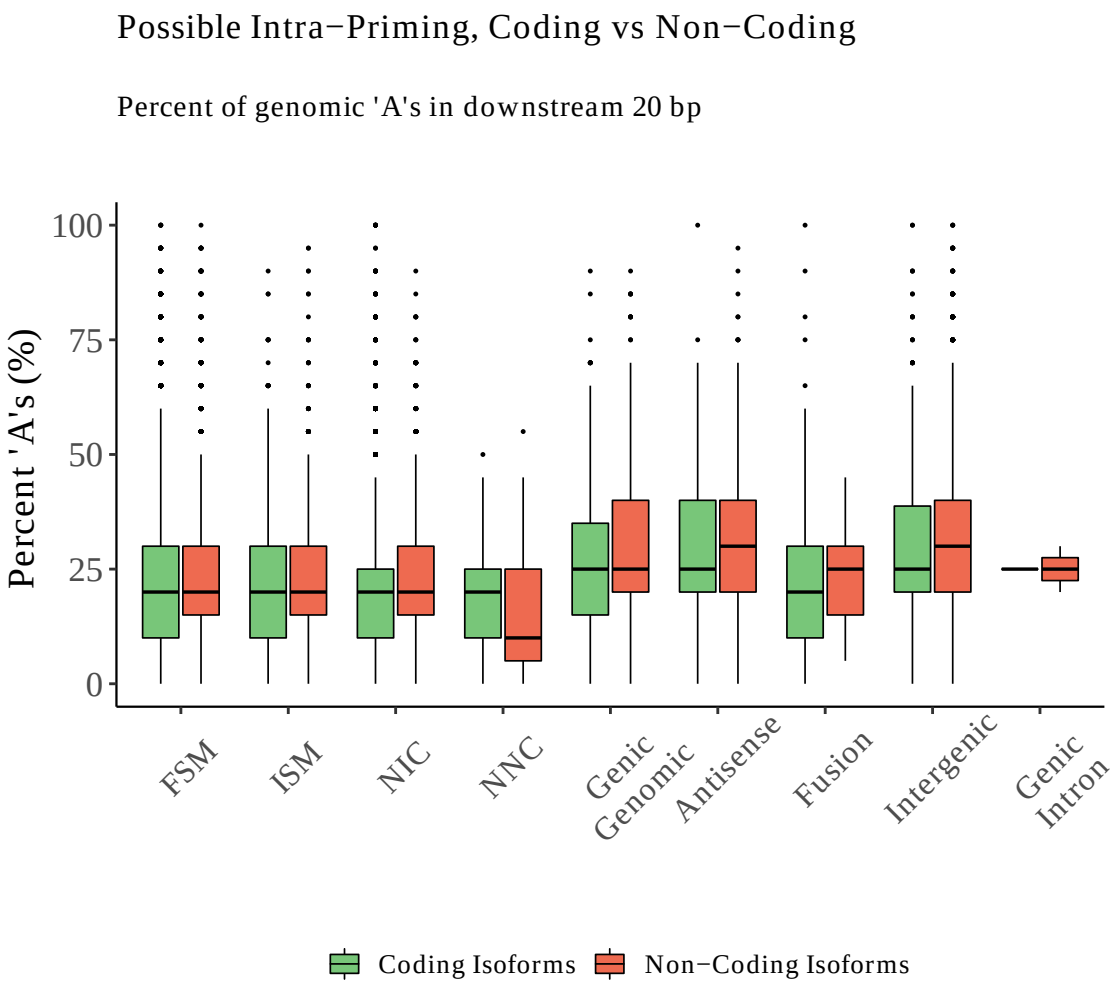



Distance to Annotated Transcription Start Site, ISM only
Negative values indicate downstream of annotated TSS
Percent of FSM Transcripts
100%
75%
50%
25%
0%
(−1.24e+06,−1e+04]
(−1e+04,−9e+03]
(−9e+03,−8e+03]
(−8e+03,−7e+03]
(−7e+03,−6e+03]
(−6e+03,−5e+03]
(−5e+03,−4e+03]
(−4e+03,−3e+03]
(−3e+03,−2e+03]
(−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+04,1.24e+06]
Distance to Annotated Transcription Start Site (bp)

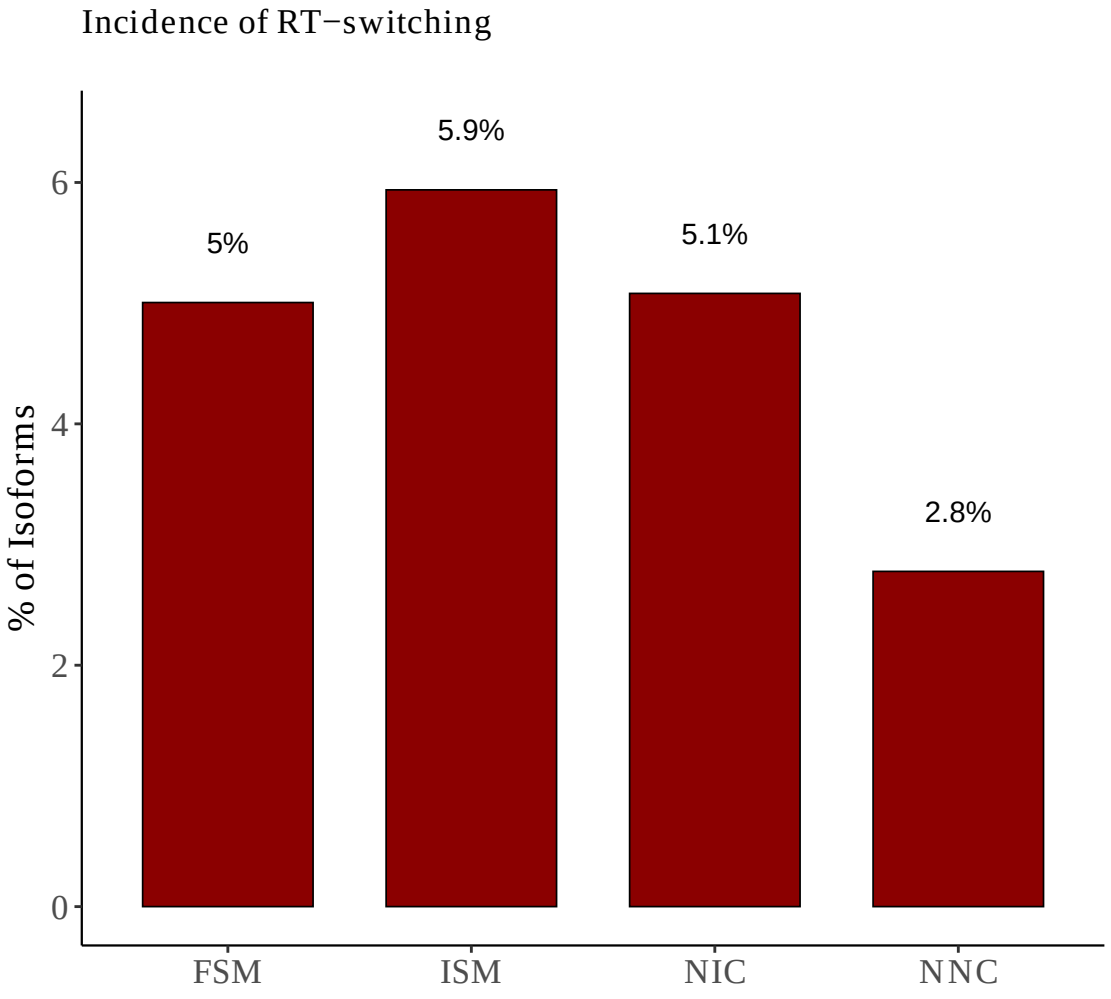
Intra−Priming Quality Check

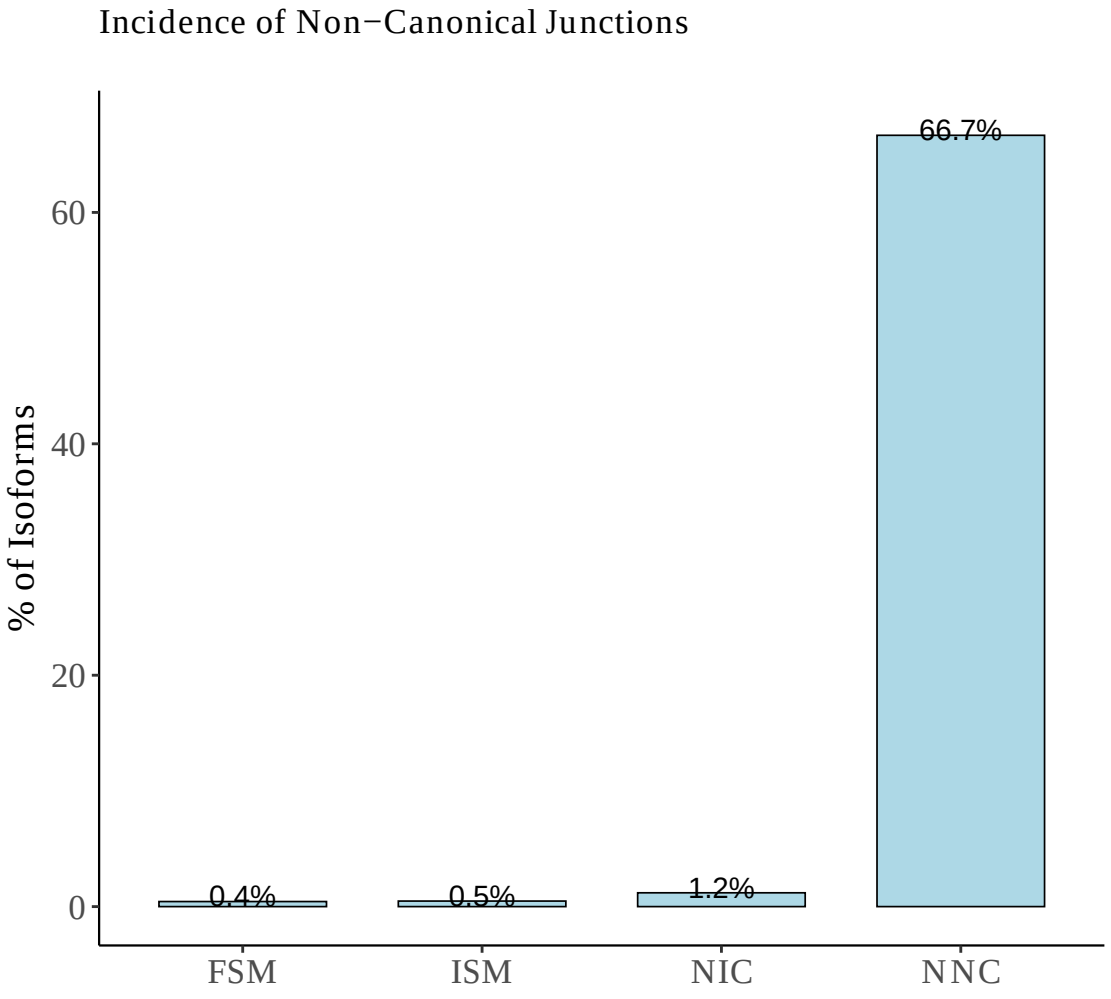

Possible Intra−Priming by Structural Category
Percent of genomic 'A's in downstream 20 bp
Percent 'A's (%)
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron
Intron retention
Internal fragment
Mono−exon
Combination of annotated junctions
3' fragment
5' fragment
Multi−exon


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


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

Quality Controls


Incidence of RT−switching
% of Isoforms
0
2
4
6
5%
5.9%
5.1%
2.8%
FSM
ISM
NIC
NNC


Incidence of Non−Canonical Junctions
% of Isoforms
0
20
40
60
0.4%
0.5%
1.2%
66.7%
FSM
ISM
NIC
NNC