SQANTI3 report

Unique Genes: 248

Unique Isoforms: 1609

Gene Classification

Category	Genes, count
Annotated Genes	224
Novel Genes	24

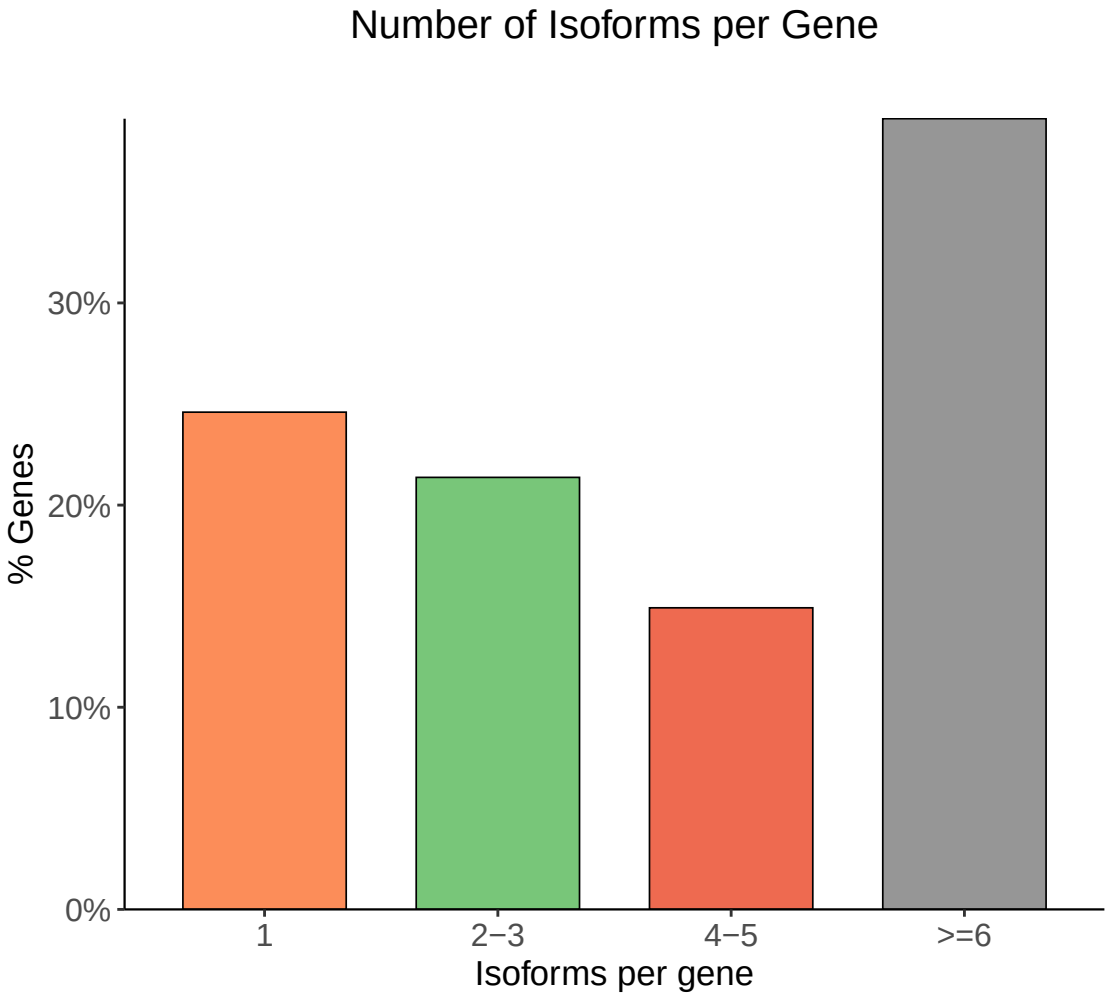
Splice Junction Classification

Category	SJs, count	Percent
Known canonical	2465	82.86
Known Non−canonical	3	0.10
Novel canonical	443	14.89
Novel Non−canonical	64	2.15

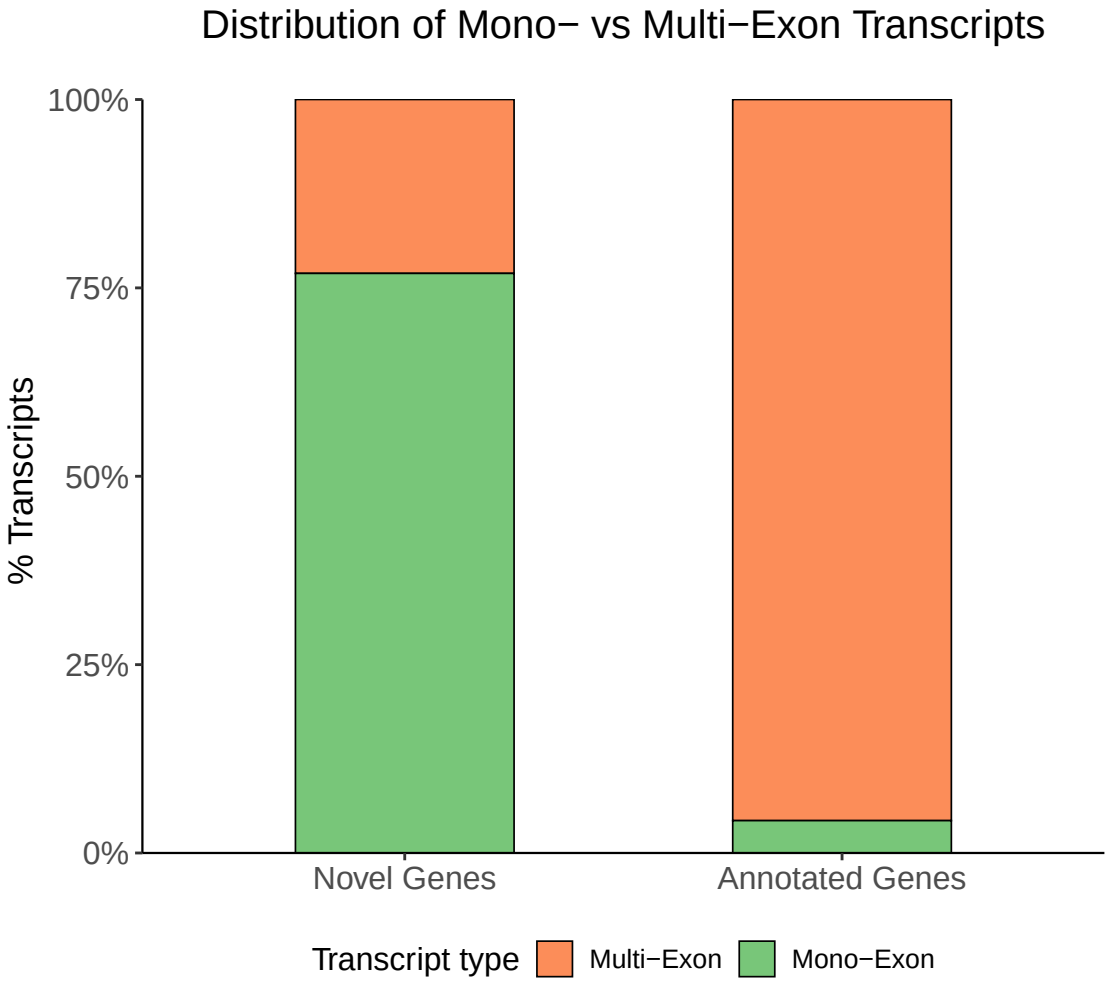
Transcript Classification

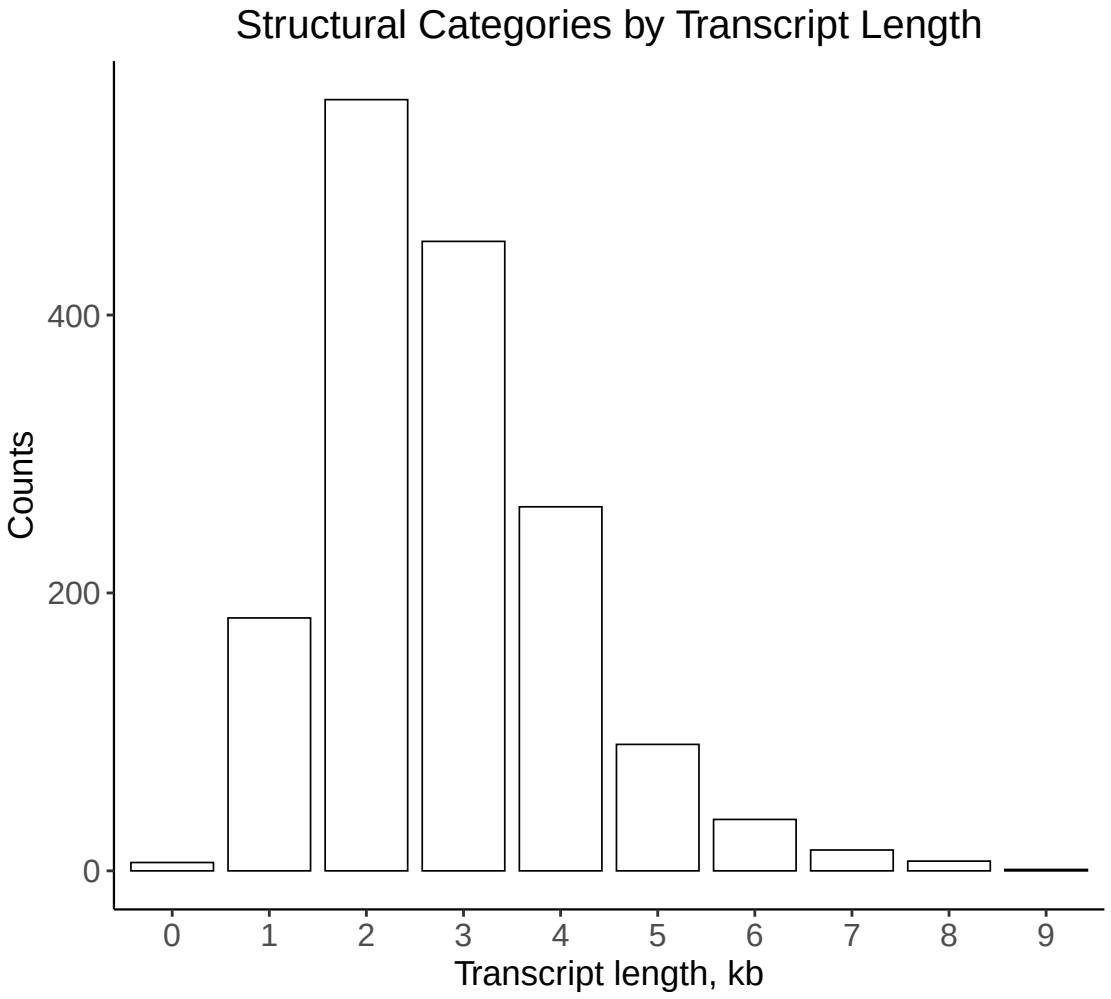
Category	Isoforms, count
FSM	0
ISM	0
NIC	0
NNC	0
Genic Genomic	0
Antisense	0
Fusion	0
Intergenic	0
Genic Intron	0

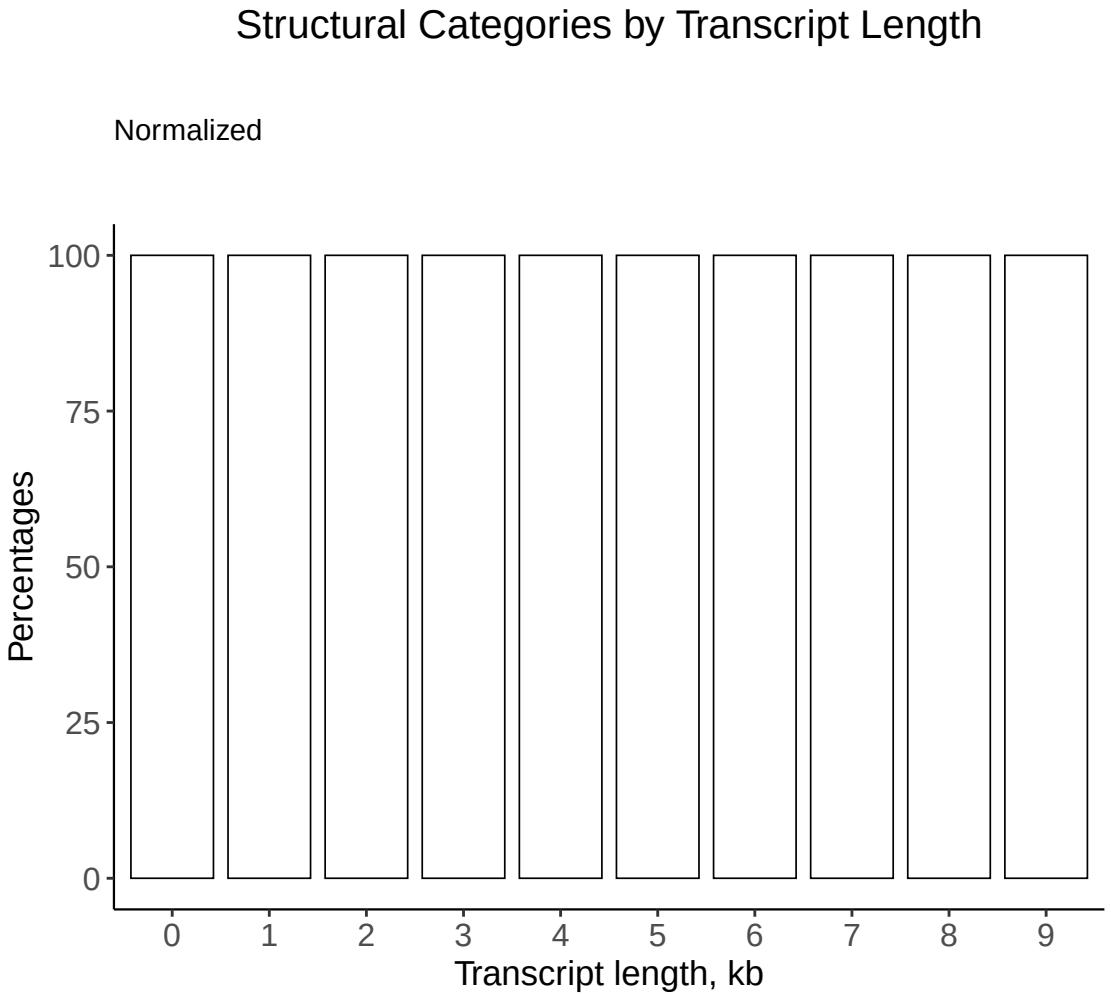
Gene Characterization

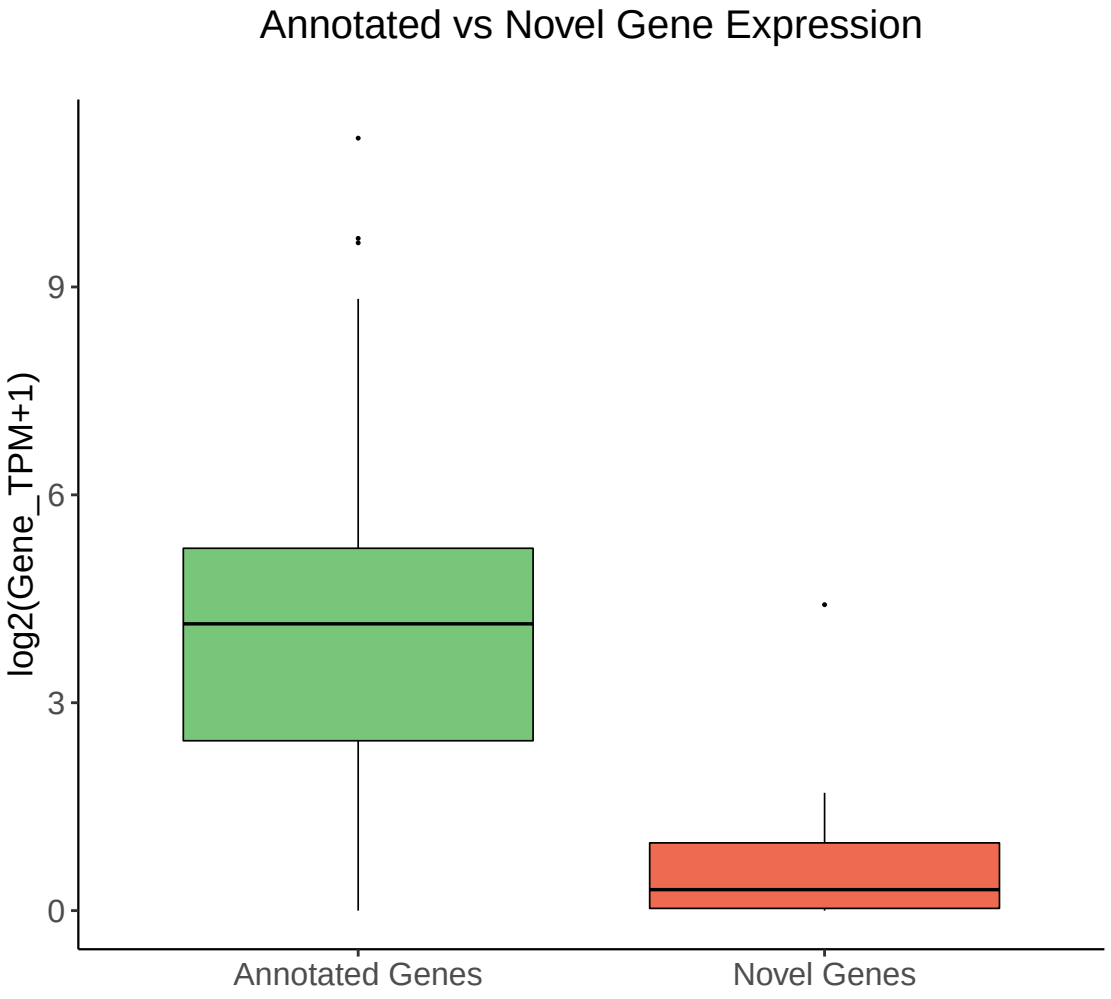

Number of Isoforms per Gene
% Genes
30%
20%
10%
0%
1
2−3
4−5
>=6
Isoforms per gene

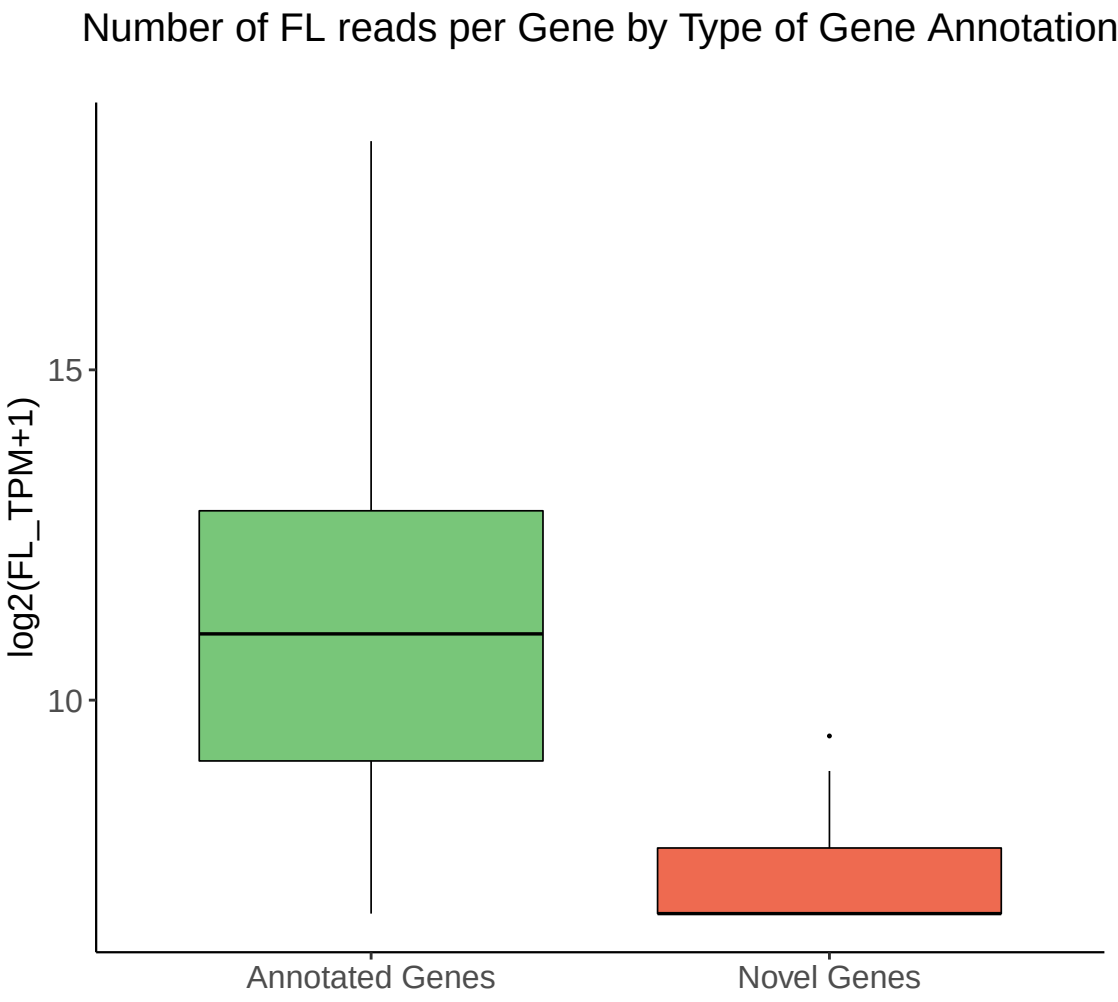
Number of Isoforms per Gene
Known vs Novel Genes
100%
75%
50%
25%
0%
Genes, %
Annotated Genes
Novel Genes
Isoforms Per Gene
1
2−3
4−5
>=6

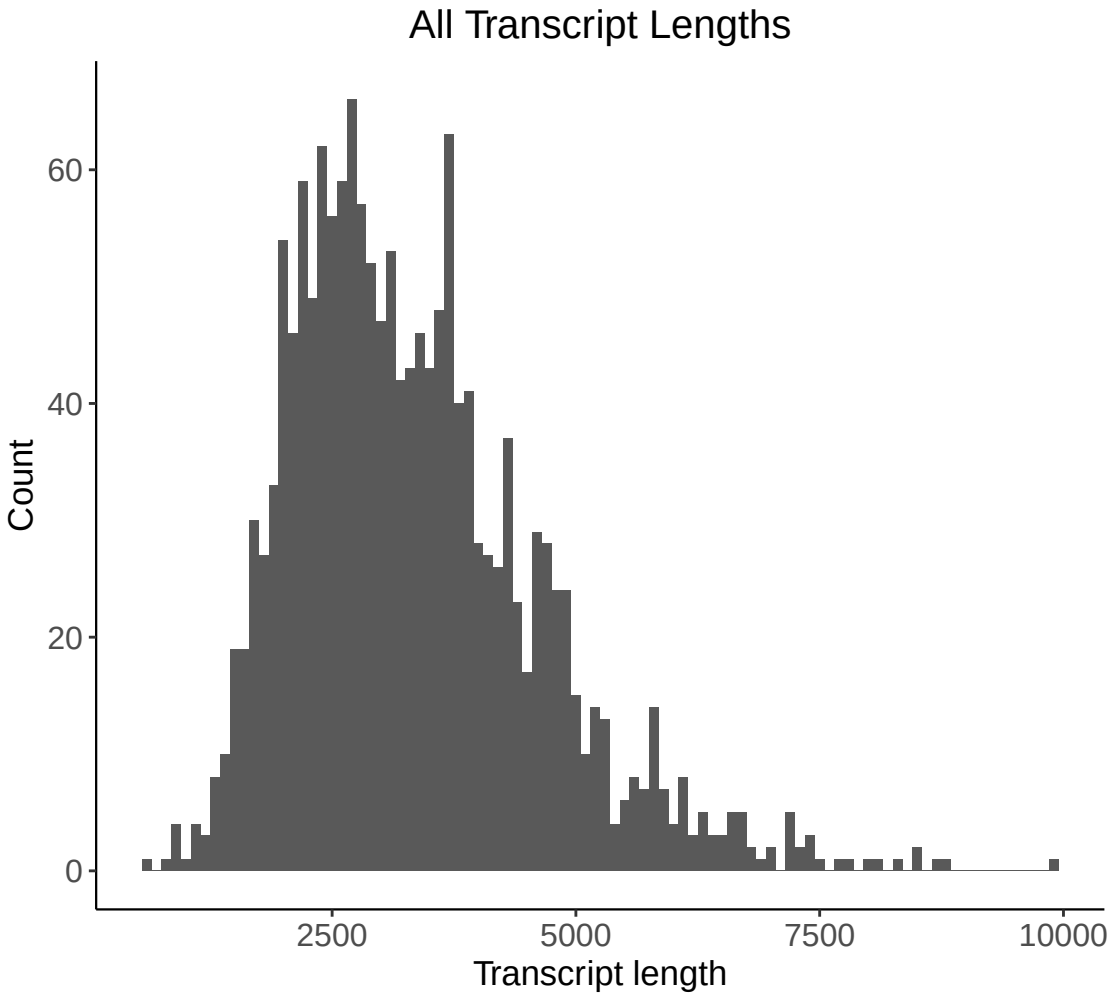

Distribution of Mono− vs Multi−Exon Transcripts
100%
75%
50%
25%
0%
% Transcripts
Novel Genes
Annotated Genes
Transcript type
Multi−Exon
Mono−Exon

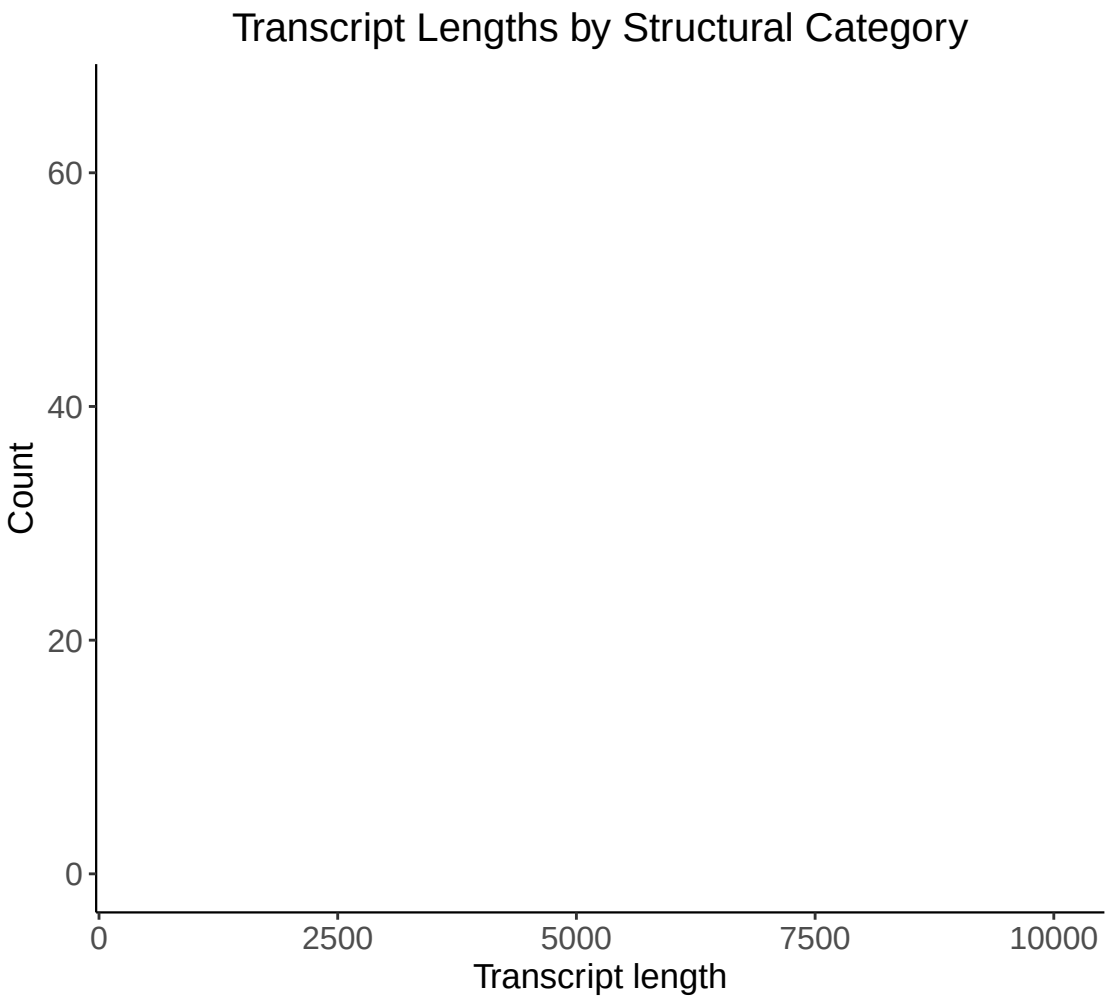

Structural Categories by Transcript Length
Counts
400
200
0
0
1
2
3
4
5
6
7
8
9
Transcript length, kb

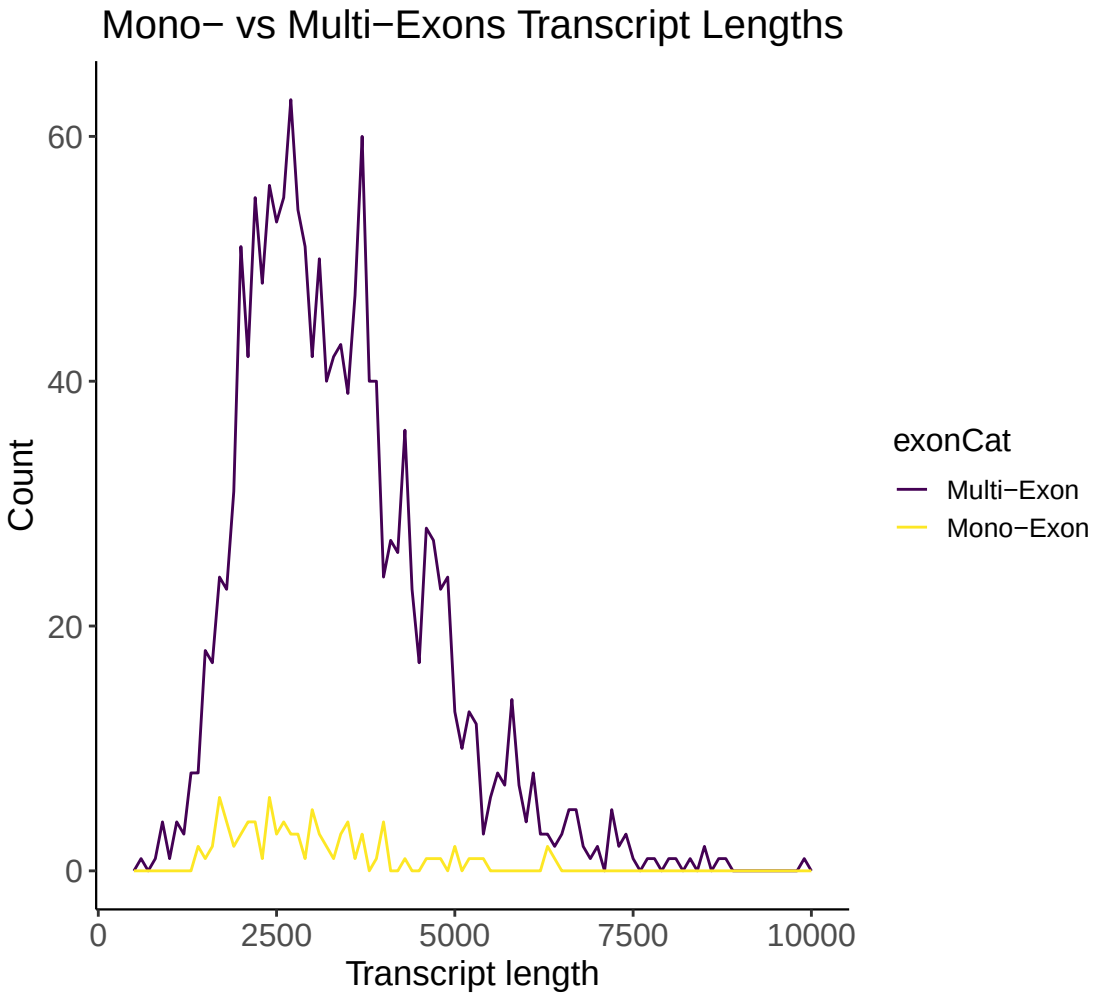

Structural Categories by Transcript Length
Normalized
Percentages
100
75
50
25
0
0
1
2
3
4
5
6
7
8
9
Transcript length, kb

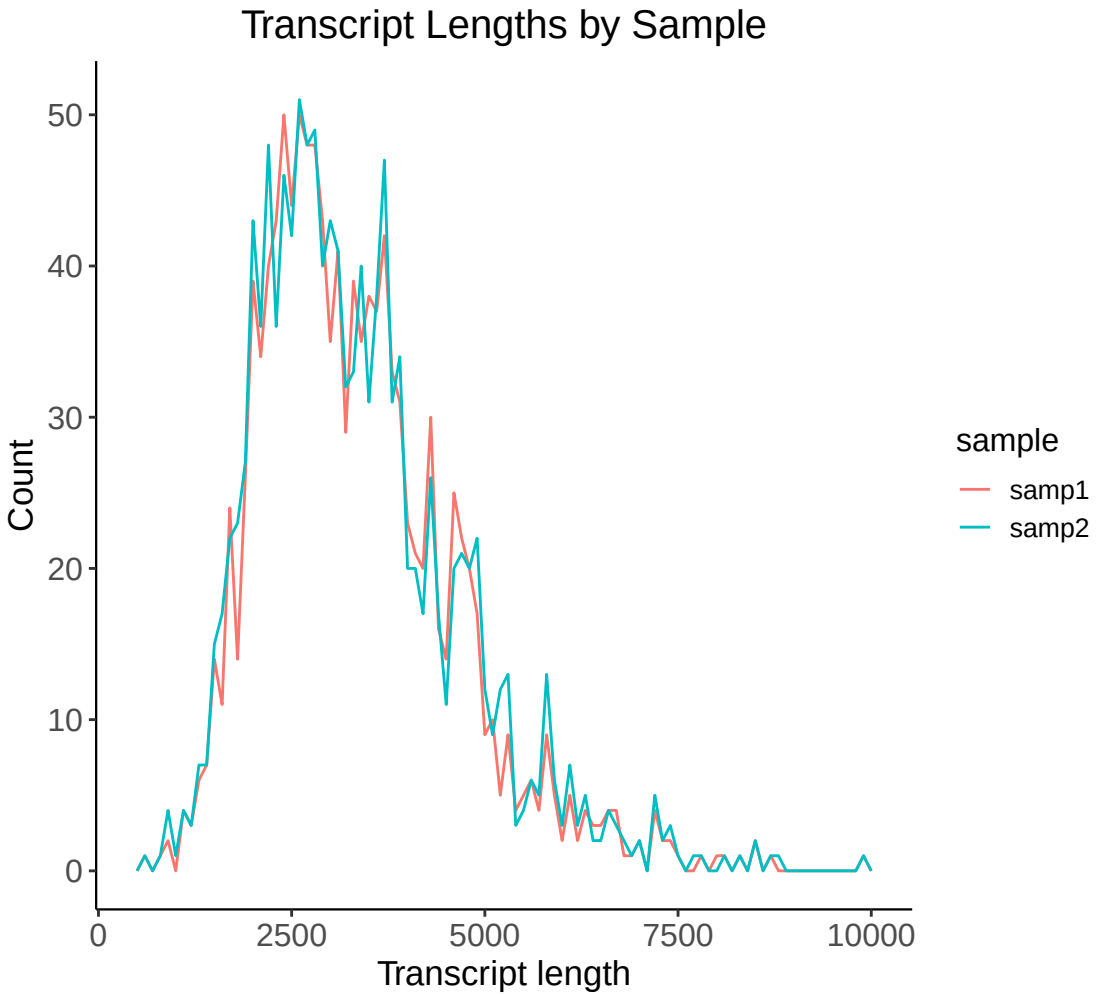

Annotated vs Novel Gene Expression
log2(Gene_TPM+1)
9
6
3
0
Annotated Genes
Novel Genes

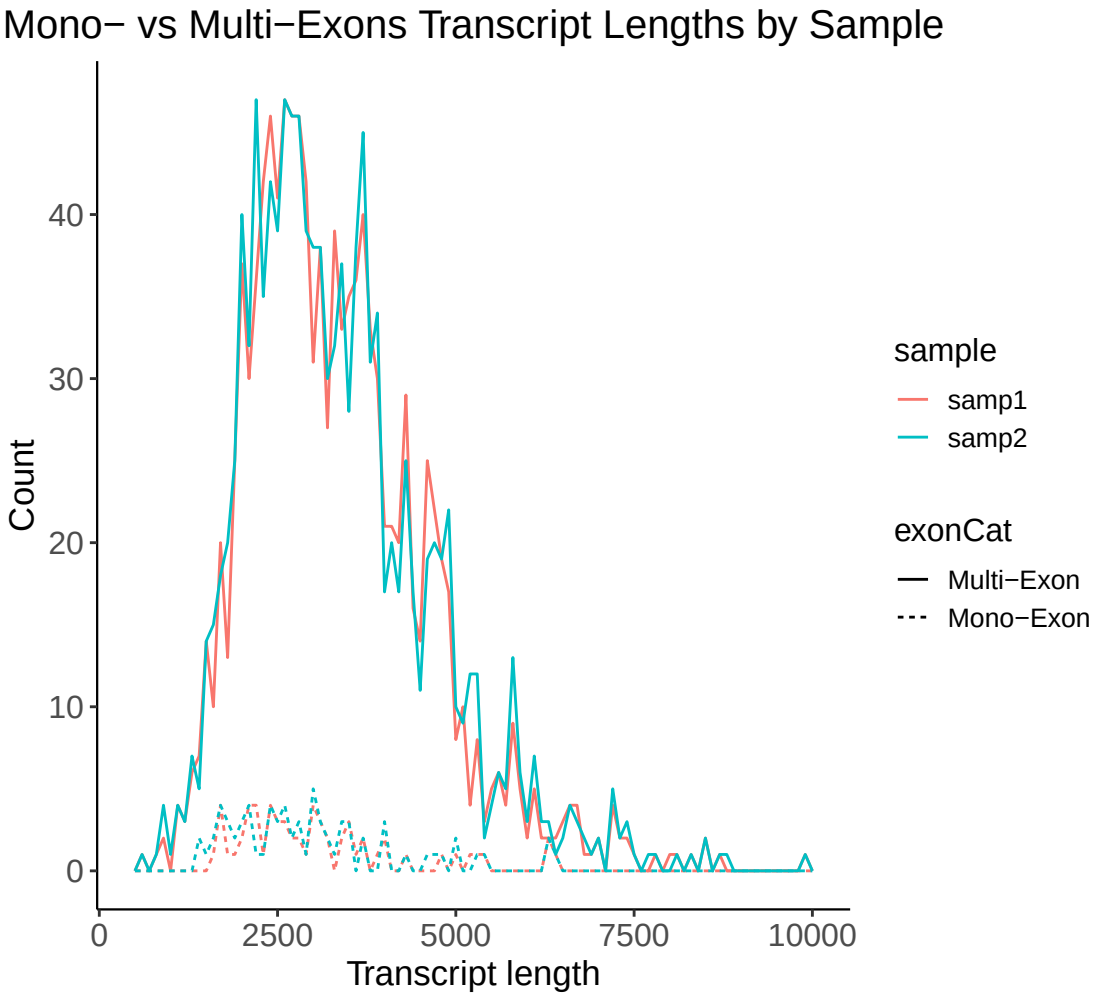

Number of FL reads per Gene by Type of Gene Annotation
log2(FL_TPM+1)
15
10
Annotated Genes
Novel Genes


All Transcript Lengths
Count
60
40
20
0
2500
5000
7500
10000
Transcript length

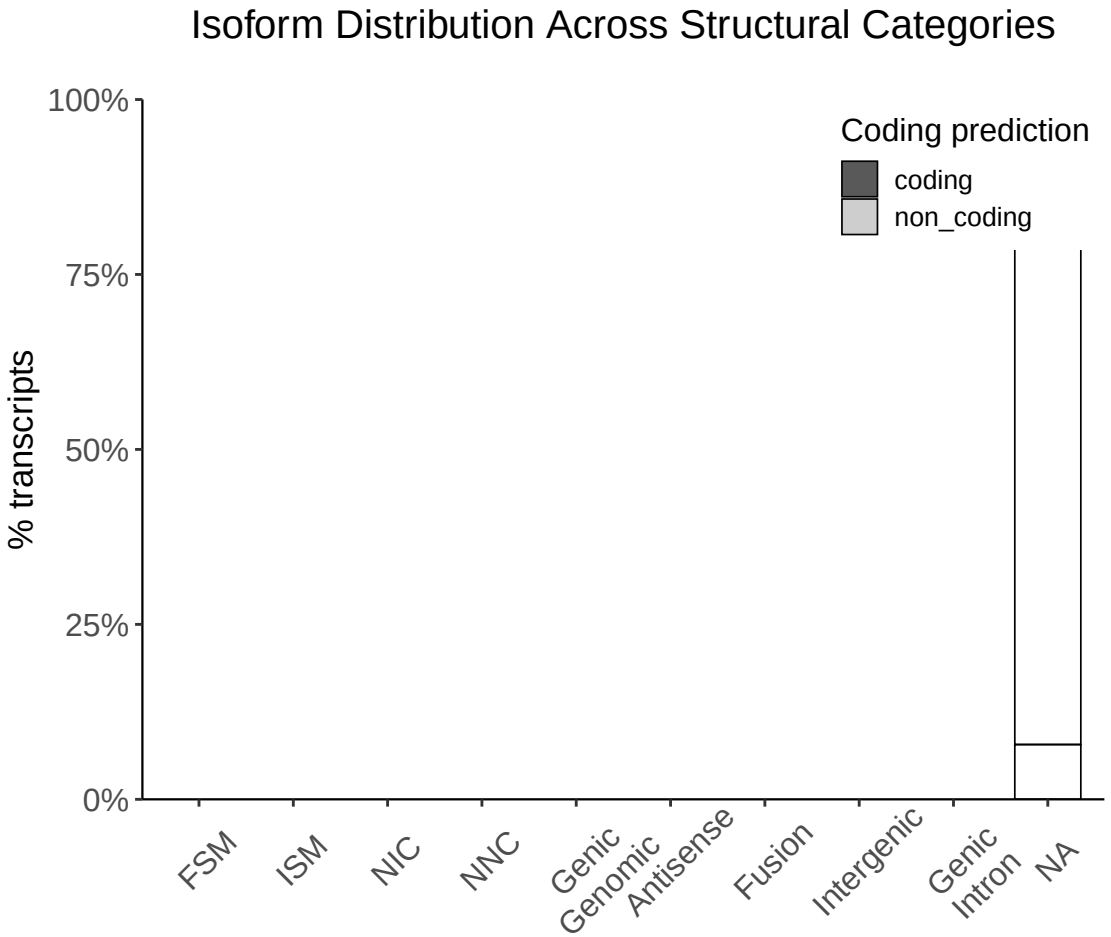

Transcript Lengths by Structural Category
Count
60
40
20
0
0
2500
5000
7500
10000
Transcript length

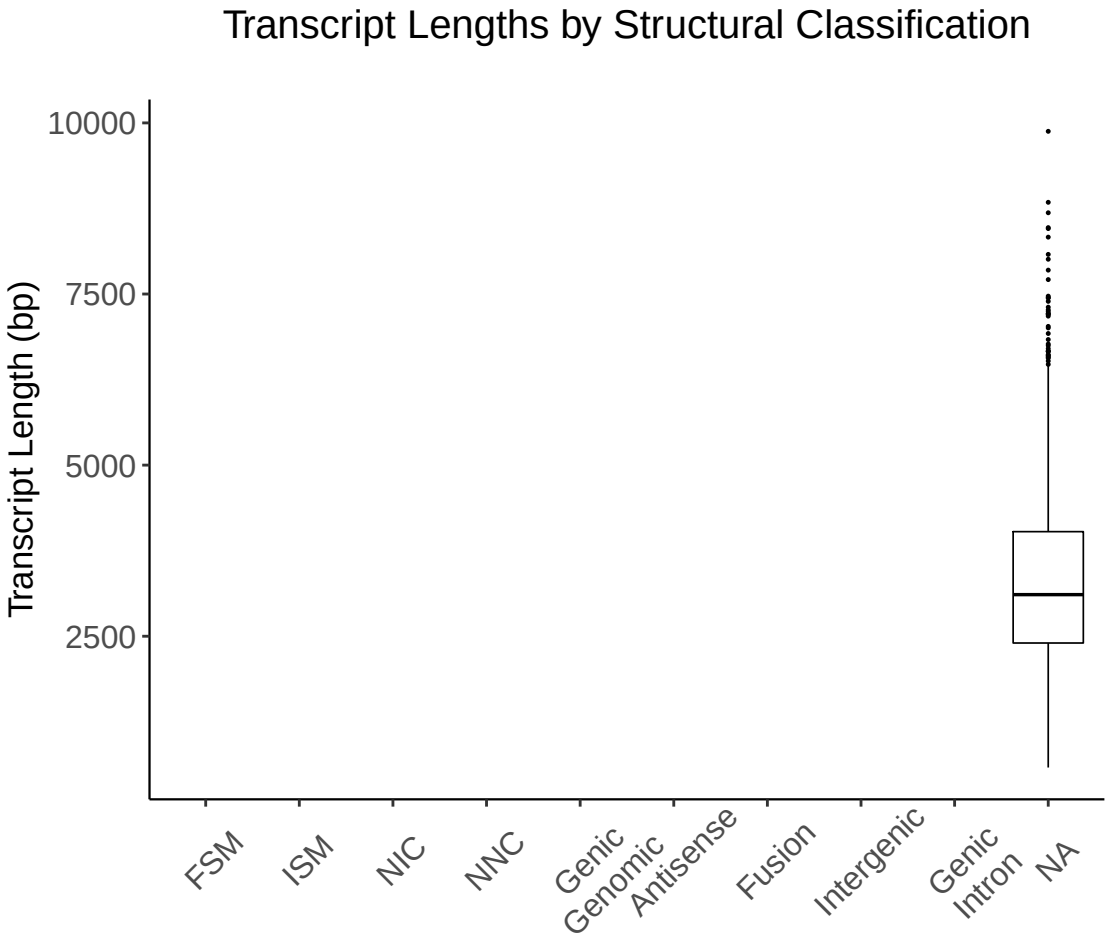

Mono− vs Multi−Exons Transcript Lengths
Count
60
40
20
0
0
2500
5000
7500
10000
Transcript length
exonCat
Multi−Exon
Mono−Exon

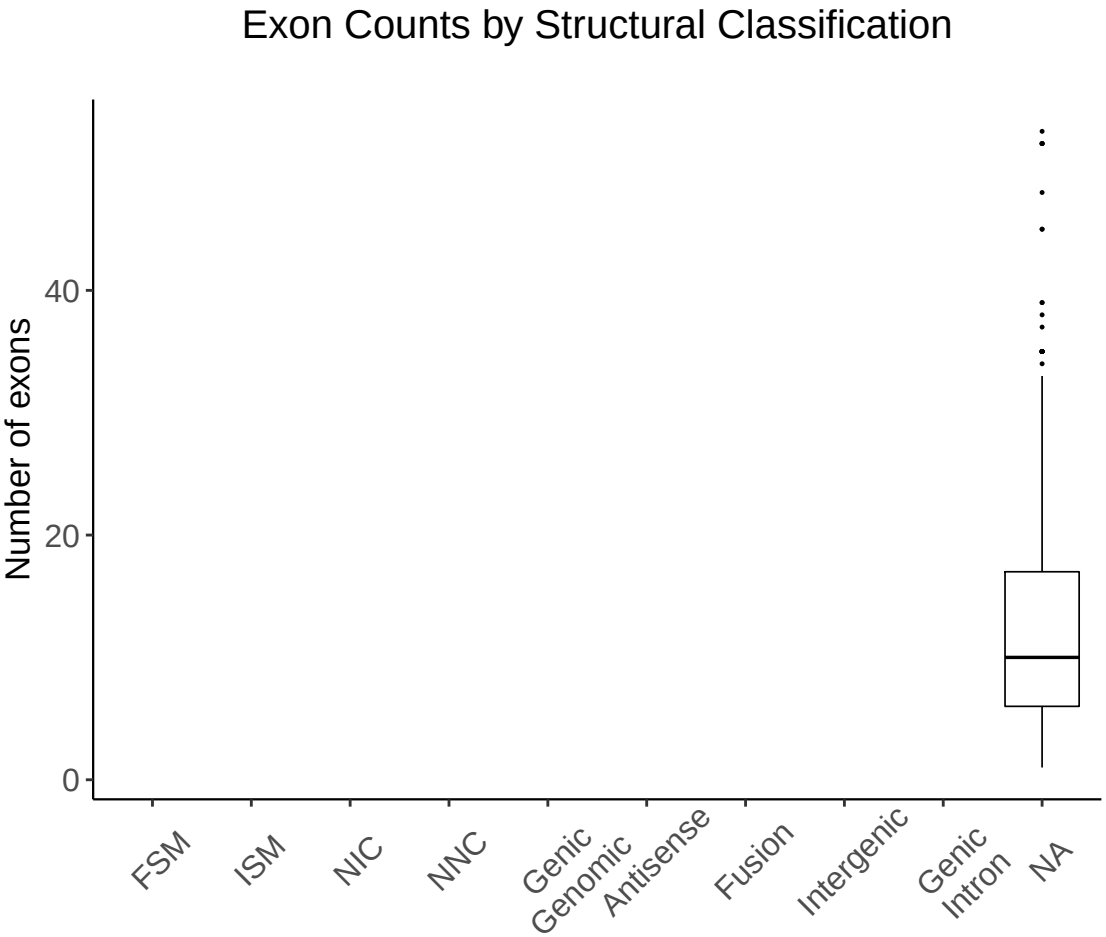

Transcript Lengths by Sample
Count
50
40
30
20
10
0
0
2500
5000
7500
10000
Transcript length
sample
samp1
samp2

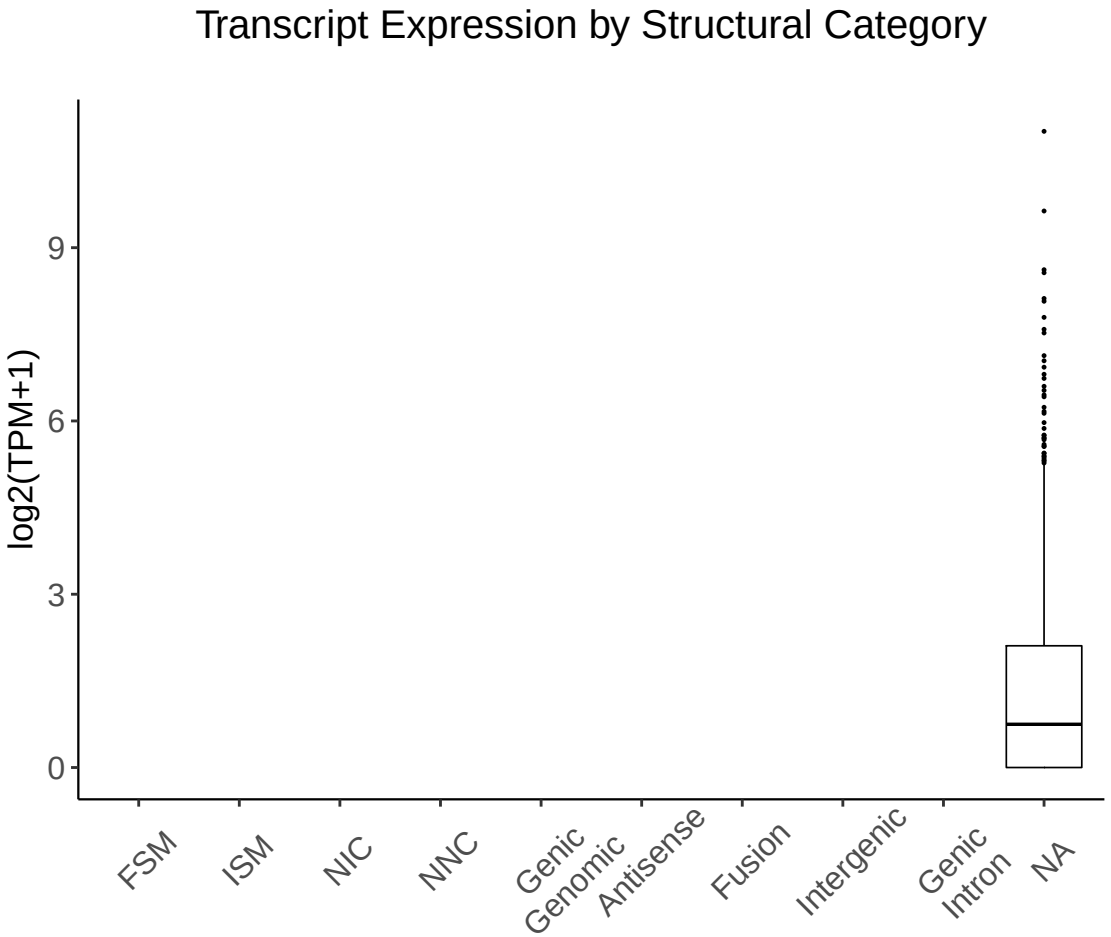

Mono− vs Multi−Exons Transcript Lengths by Sample
Count
0
10
20
30
40
0
2500
5000
7500
10000
Transcript length
sample
samp1
samp2
exonCat
Multi−Exon
Mono−Exon

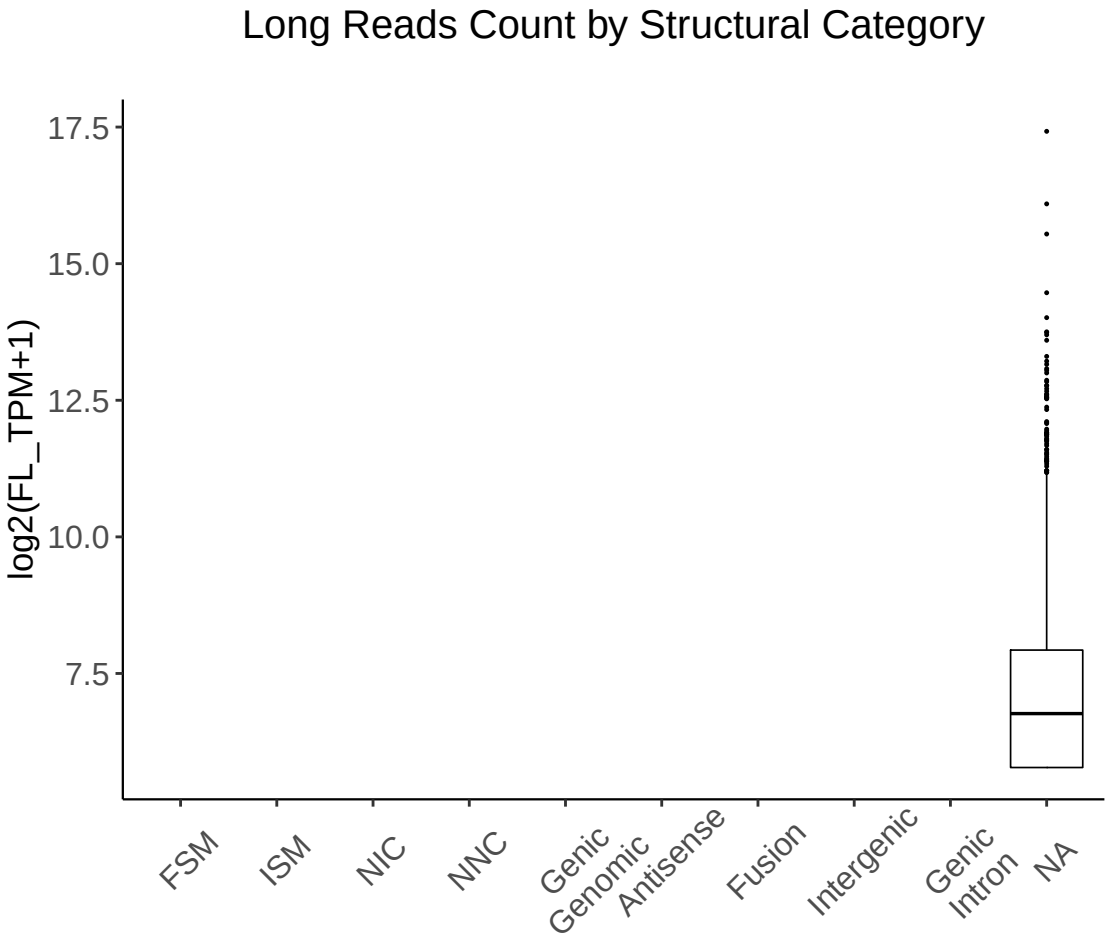
Structural Isoform Characterization

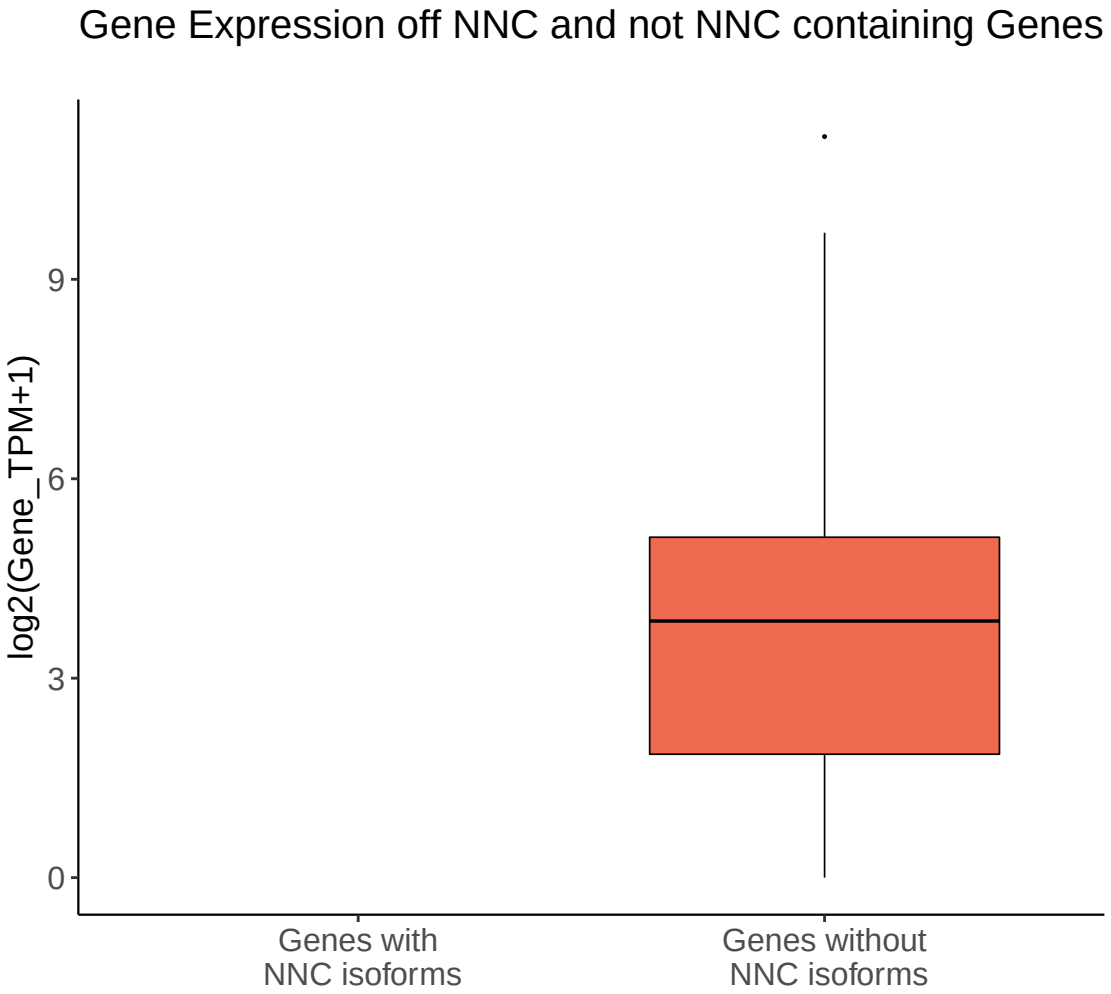

Isoform Distribution Across Structural Categories
100%
75%
50%
25%
0%
% transcripts
Coding prediction
coding
non_coding
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron
NA


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

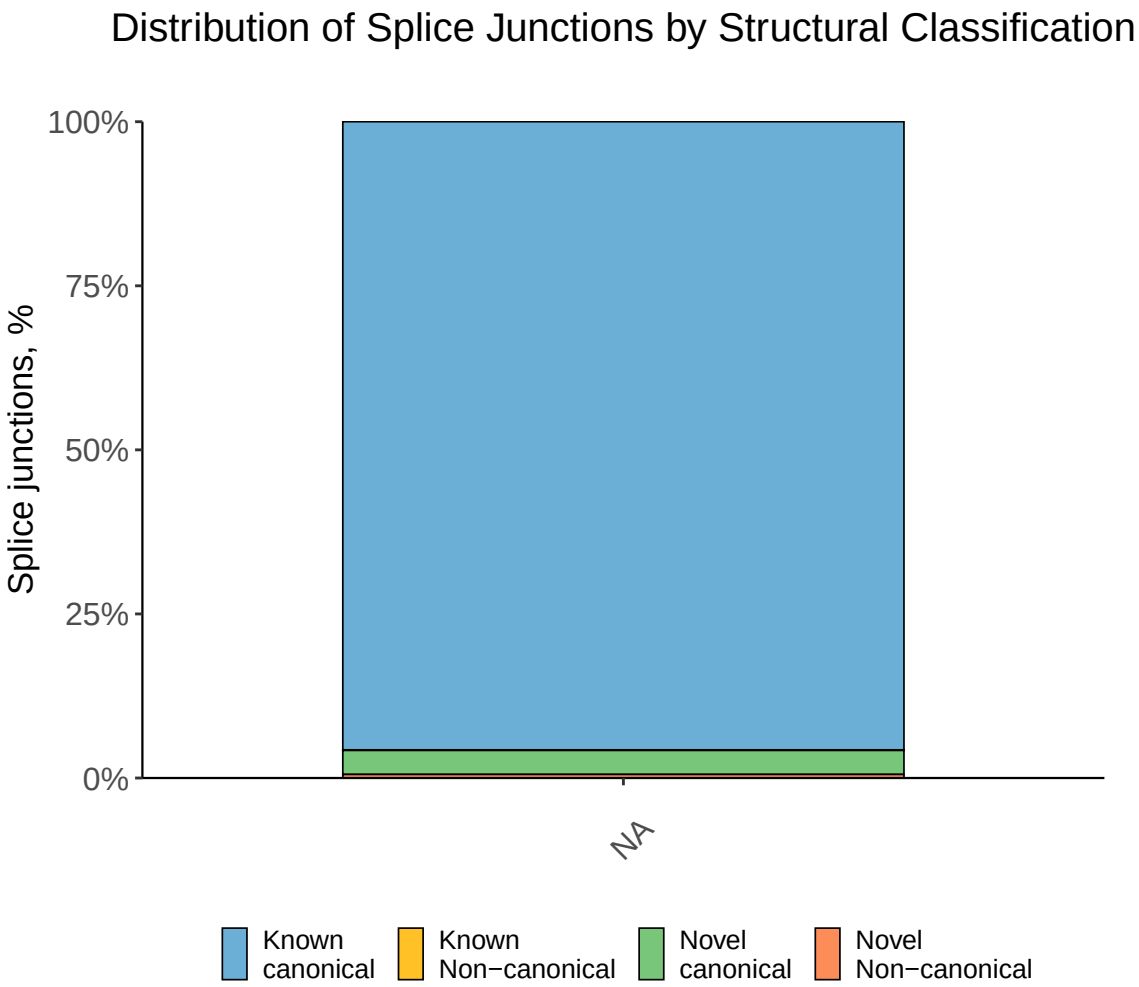

Exon Counts by Structural Classification
Number of exons
0
20
40
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron
NA

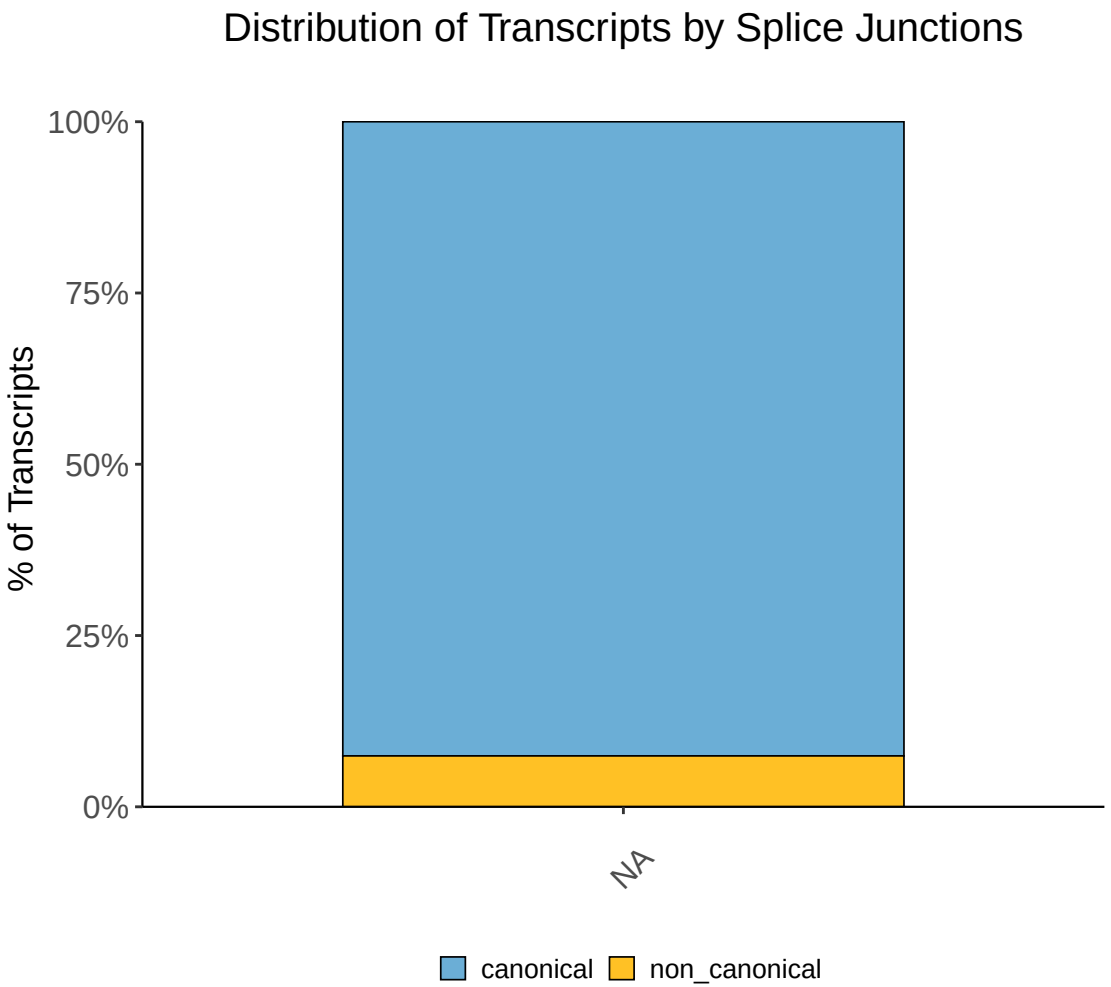

Transcript Expression by Structural Category
log2(TPM+1)
0
3
6
9
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron
NA

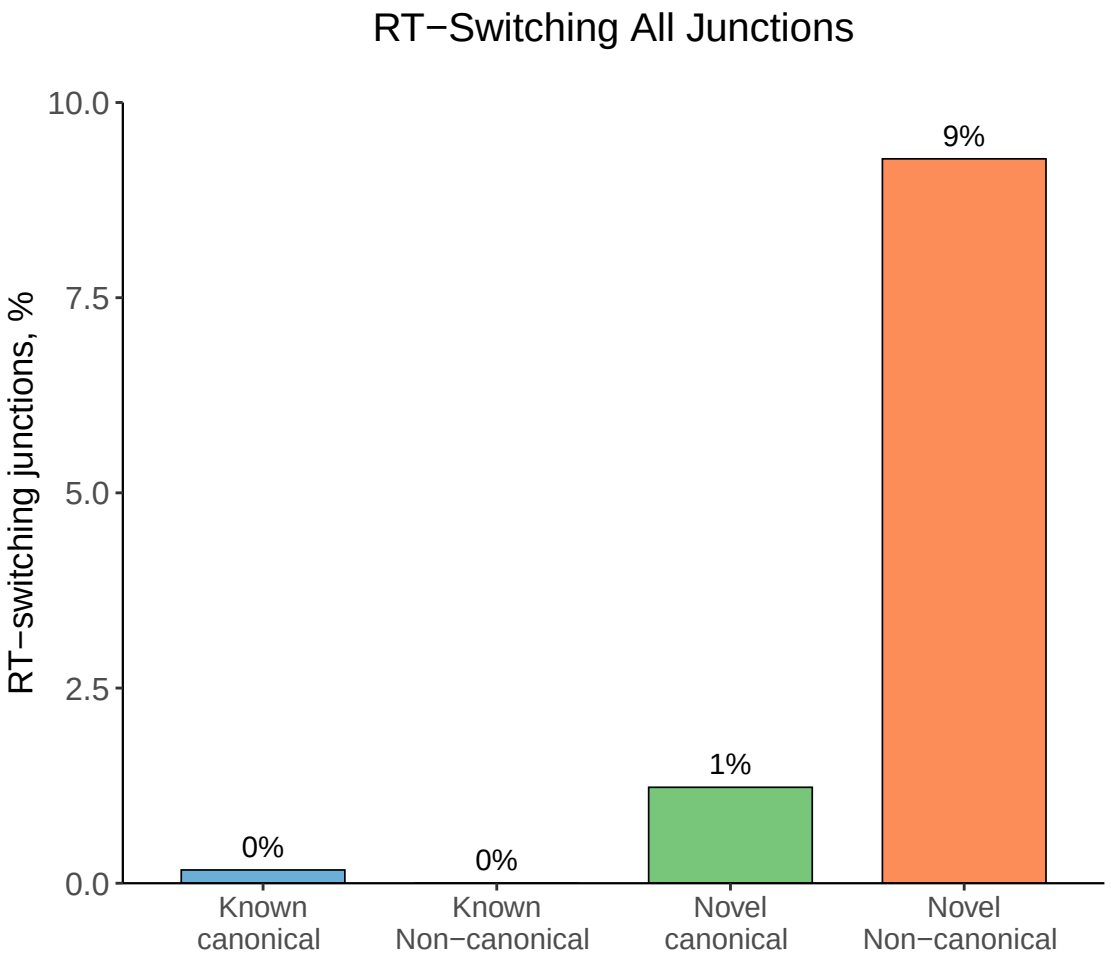

Long Reads Count by Structural Category
log2(FL_TPM+1)
17.5
15.0
12.5
10.0
7.5
FSM
ISM
NIC
NNC
Genic Genomic
Antisense
Fusion
Intergenic
Genic Intron
NA

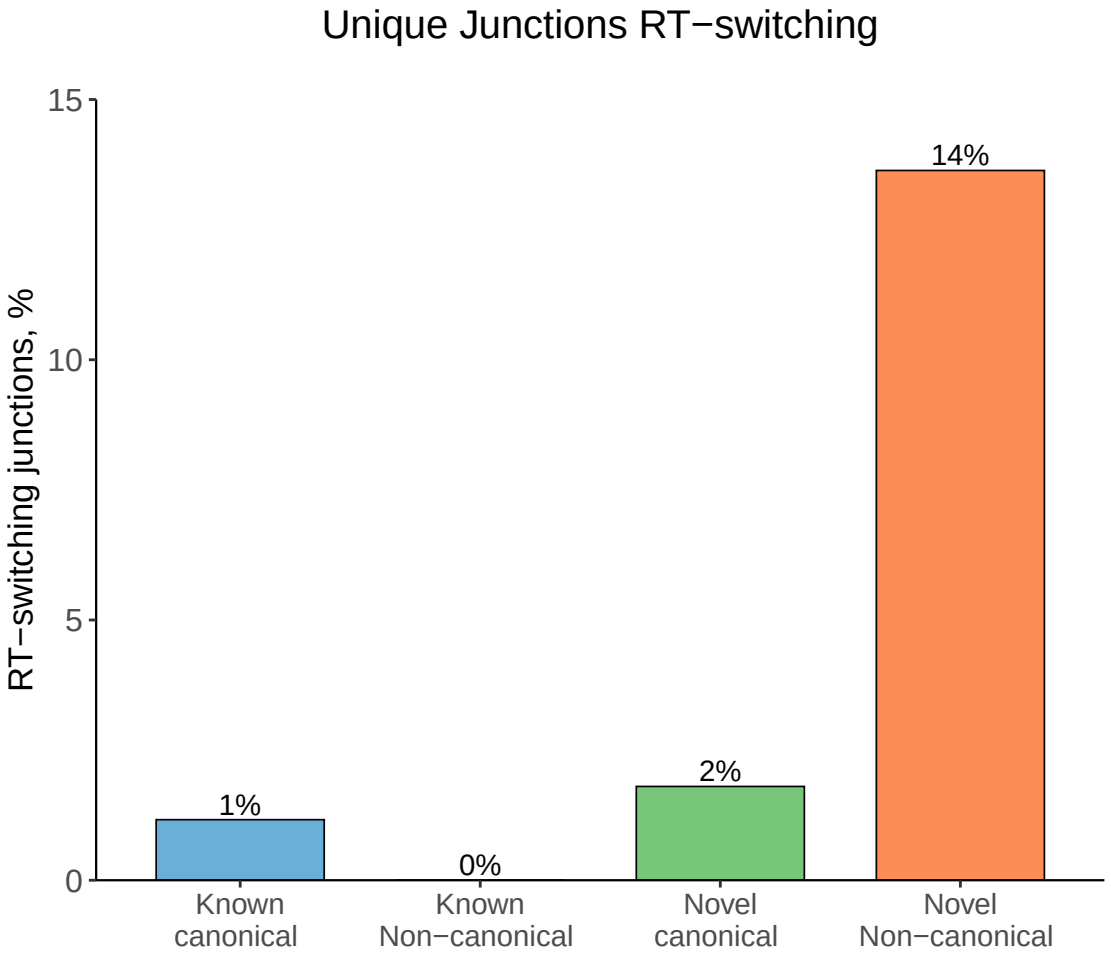

Gene Expression off NNC and not NNC containing Genes
log2(Gene_TPM+1)
9
6
3
0
Genes with
NNC isoforms
Genes without
NNC isoforms

Splice Junction Characterization


Distribution of Splice Junctions by Structural Classification
Splice junctions, %
100%
75%
50%
25%
0%
NA
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical


Distribution of Transcripts by Splice Junctions
% of Transcripts
100%
75%
50%
25%
0%
NA
canonical
non_canonical

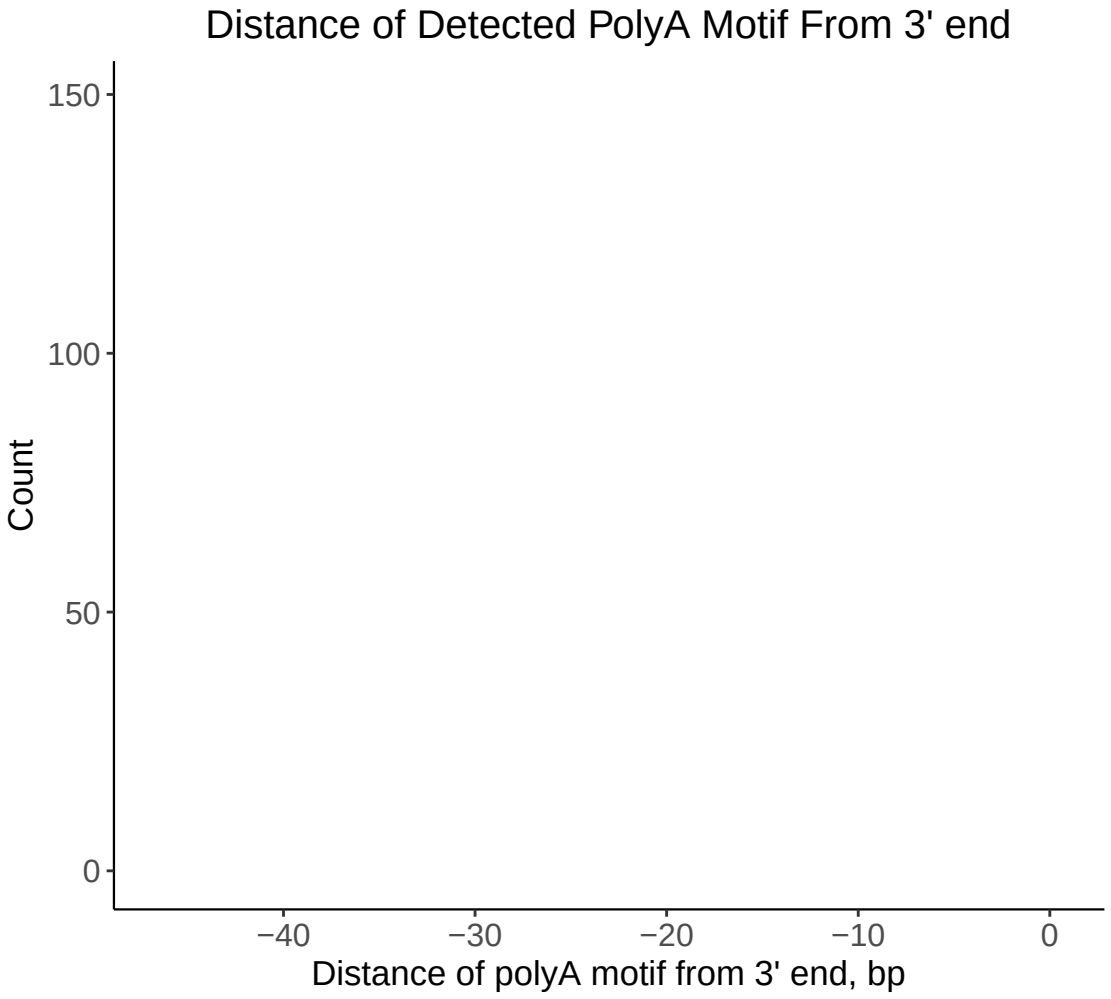

RT−Switching All Junctions
RT−switching junctions, %
10.0
7.5
5.0
2.5
0.0
0%
0%
1%
9%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical


Unique Junctions RT−switching
RT−switching junctions, %
15
10
5
0
1%
0%
2%
14%
Known canonical
Known Non−canonical
Novel canonical
Novel Non−canonical

Comparison With Annotated TSS and TTS

Comparison With Annotated TSS and TTS by Subcategories

PolyA Distance Analysis

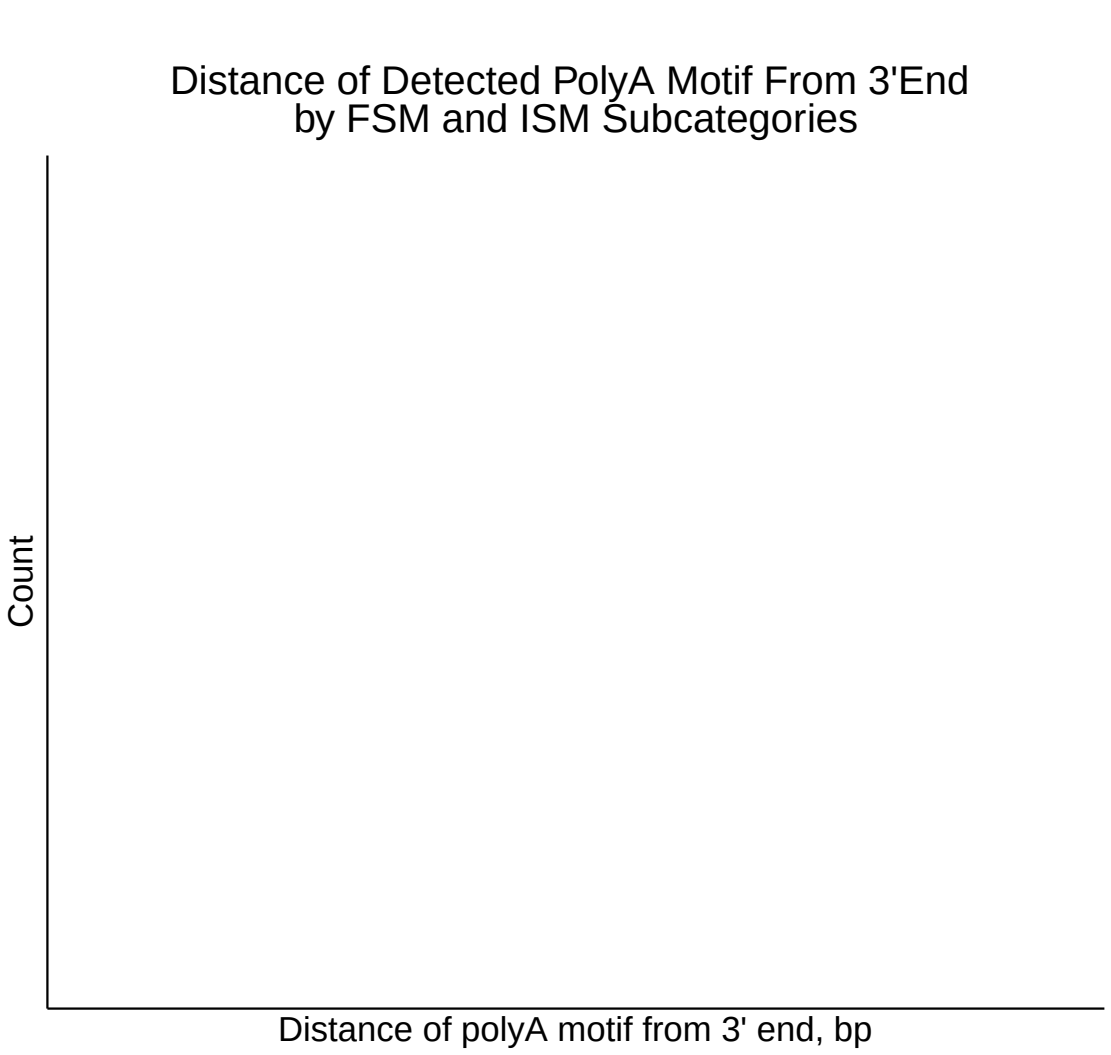

Distance of Detected PolyA Motif From 3' end
150
100
50
0
Count
−40
−30
−20
−10
0
Distance of polyA motif from 3' end, bp

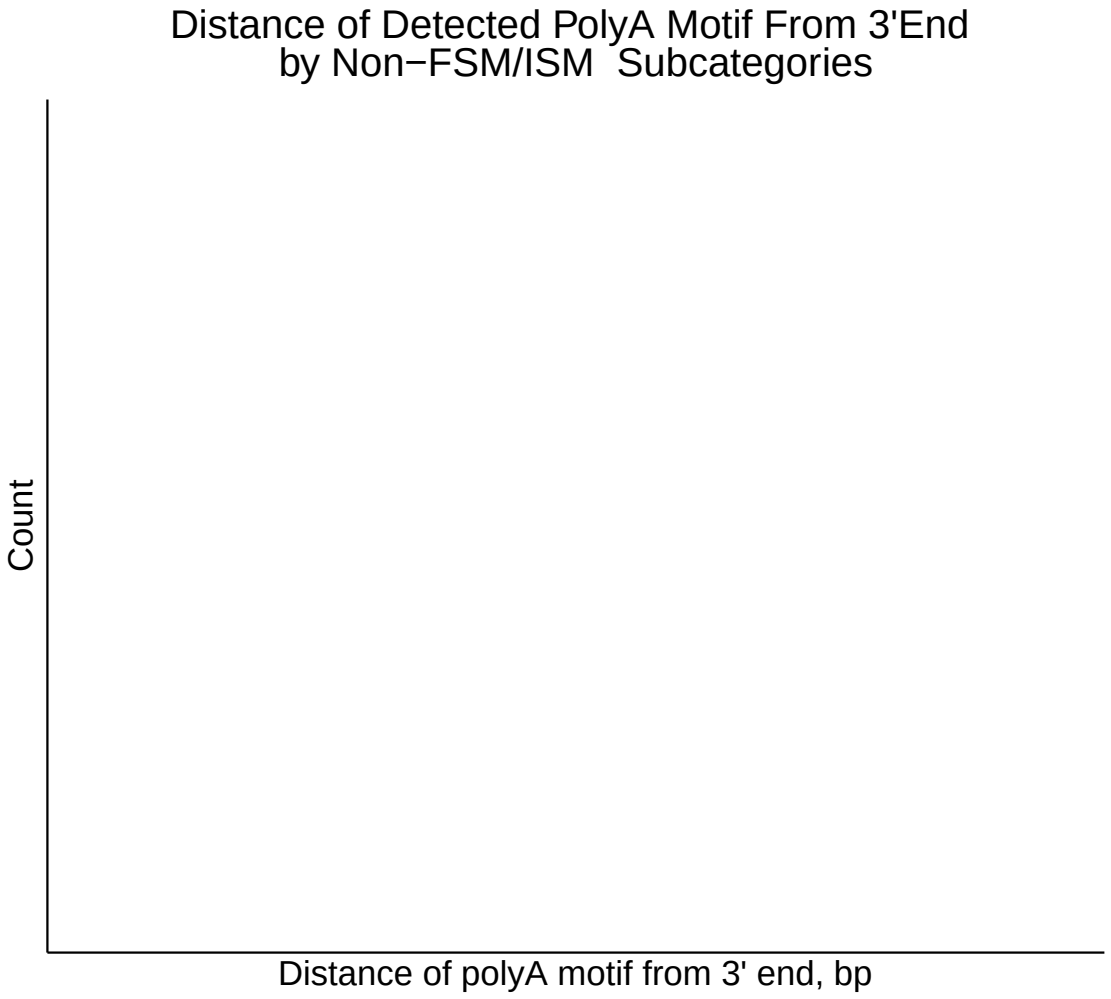
Frequency of PolyA Motifs

Motif	Count	%
AATAAA	844	62.5
ATTAAA	255	18.9
AATACA	38	2.8
TATAAA	34	2.5
AGTAAA	29	2.1
TTTAAA	28	2.1
AAGAAA	25	1.9
AAAAAG	20	1.5
AATATA	18	1.3
CATAAA	18	1.3
GATAAA	12	0.9
AATGAA	9	0.7
ACTAAA	8	0.6
AATAGA	7	0.5
AAAACA	6	0.4

Number of polyA Motifs Detected

Category	Count	polyA Detected	%
NA	1609	1351	84


Distance of Detected PolyA Motif From 3'End
by FSM and ISM Subcategories
Count
Distance of polyA motif from 3' end, bp


Distance of Detected PolyA Motif From 3'End
by Non−FSM/ISM Subcategories
Count
Distance of polyA motif from 3' end, bp

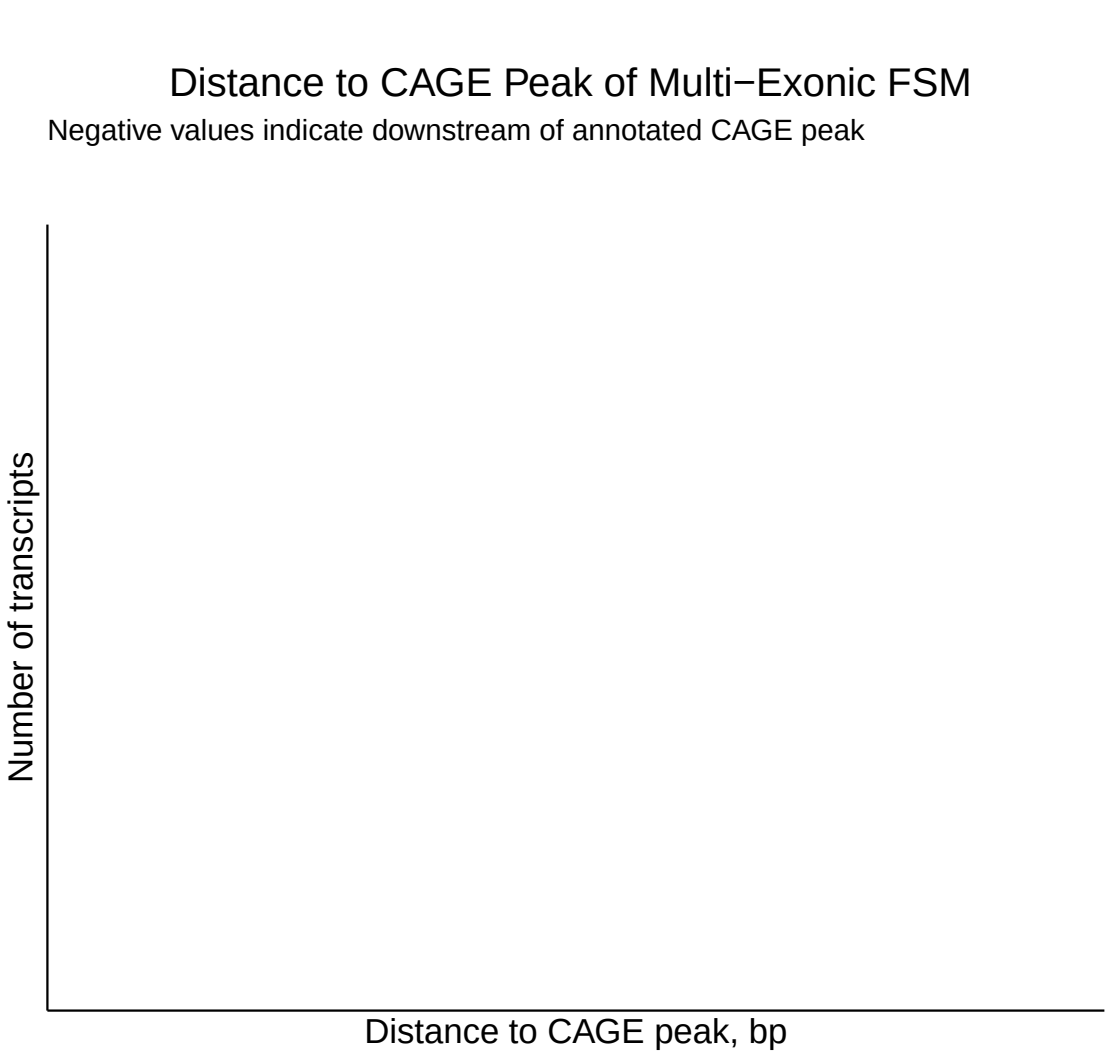
Number of polyA Motifs Detected

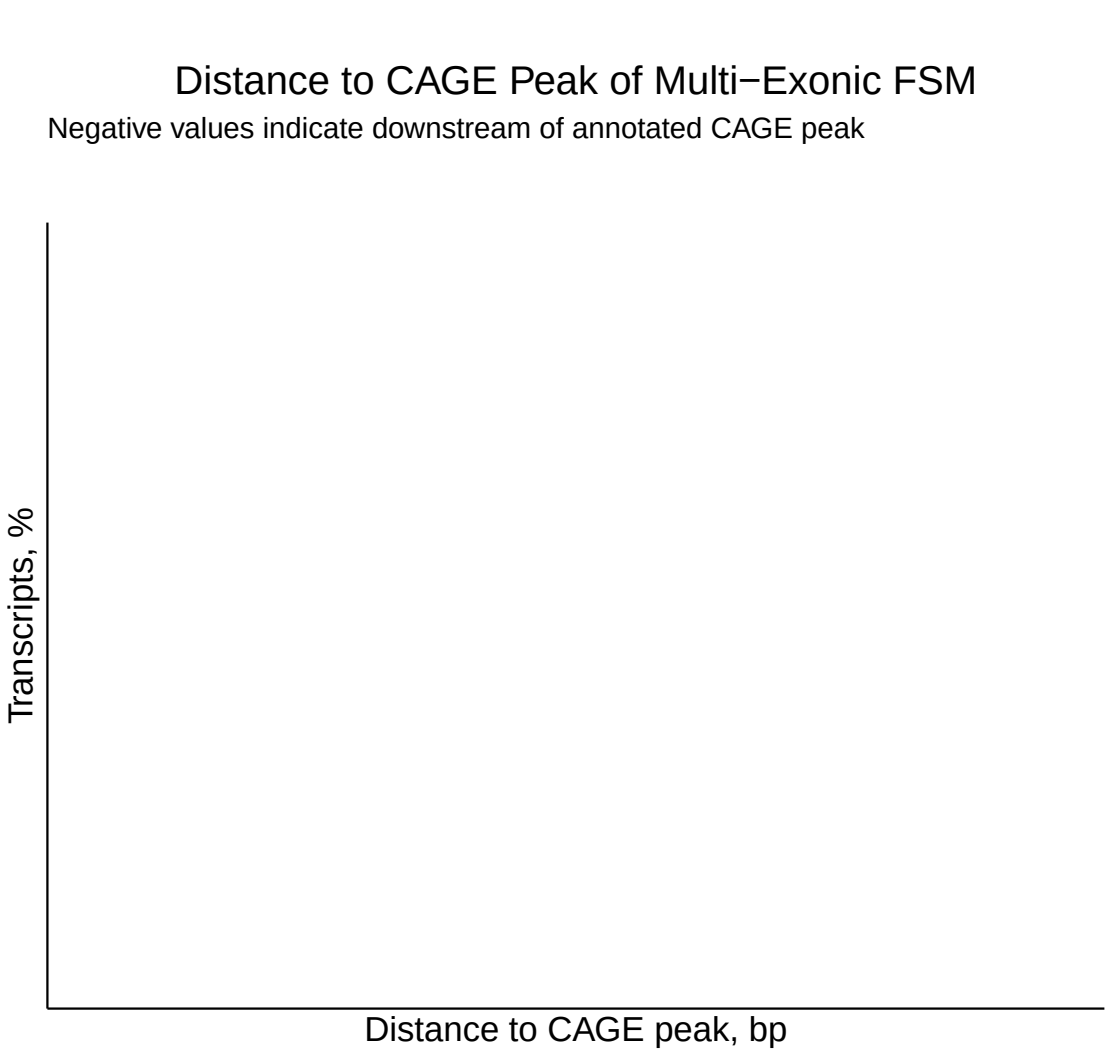
Subcategory	Count	polyA Detected	%
3' fragment	22	20	91
Internal fragment	2	1	50
5' fragment	109	77	71
Combin. of annot. junctions	119	99	83
Combin. of annot. splice sites	166	145	87
Intron retention	186	162	87
Mono−exon by intron retention	8	0	0
At least 1 annot. donor/accept.	339	311	92
Mono−exon	80	40	50
Multi−exon	578	496	86

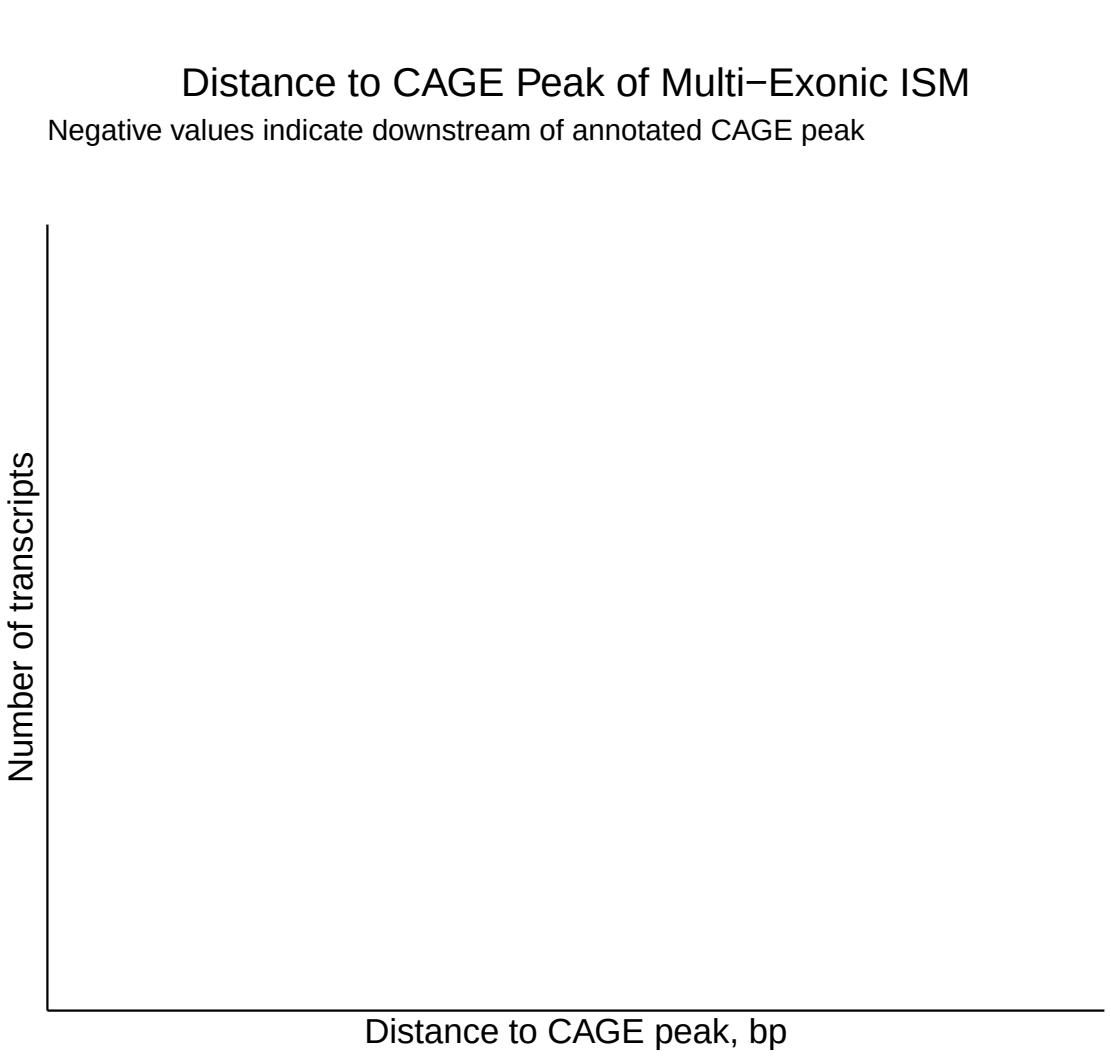
Frequency of PolyA Motifs

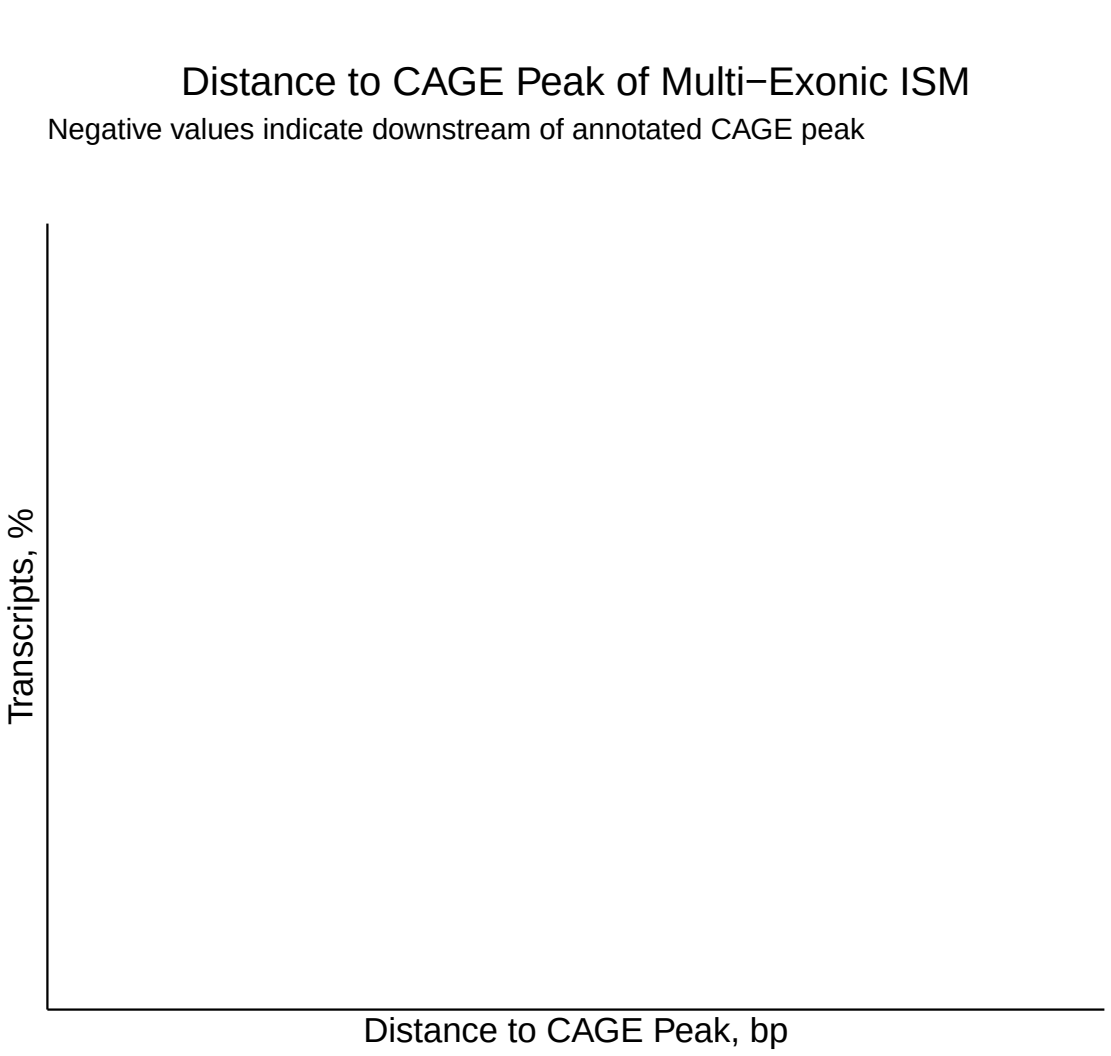
Motif	Count	%
AATAAA	844	62.5
ATTAAA	255	18.9
AATACA	38	2.8
TATAAA	34	2.5
AGTAAA	29	2.1
TTTAAA	28	2.1
AAGAAA	25	1.9
AAAAAG	20	1.5
AATATA	18	1.3
CATAAA	18	1.3
GATAAA	12	0.9
AATGAA	9	0.7
ACTAAA	8	0.6
AATAGA	7	0.5
AAAACA	6	0.4

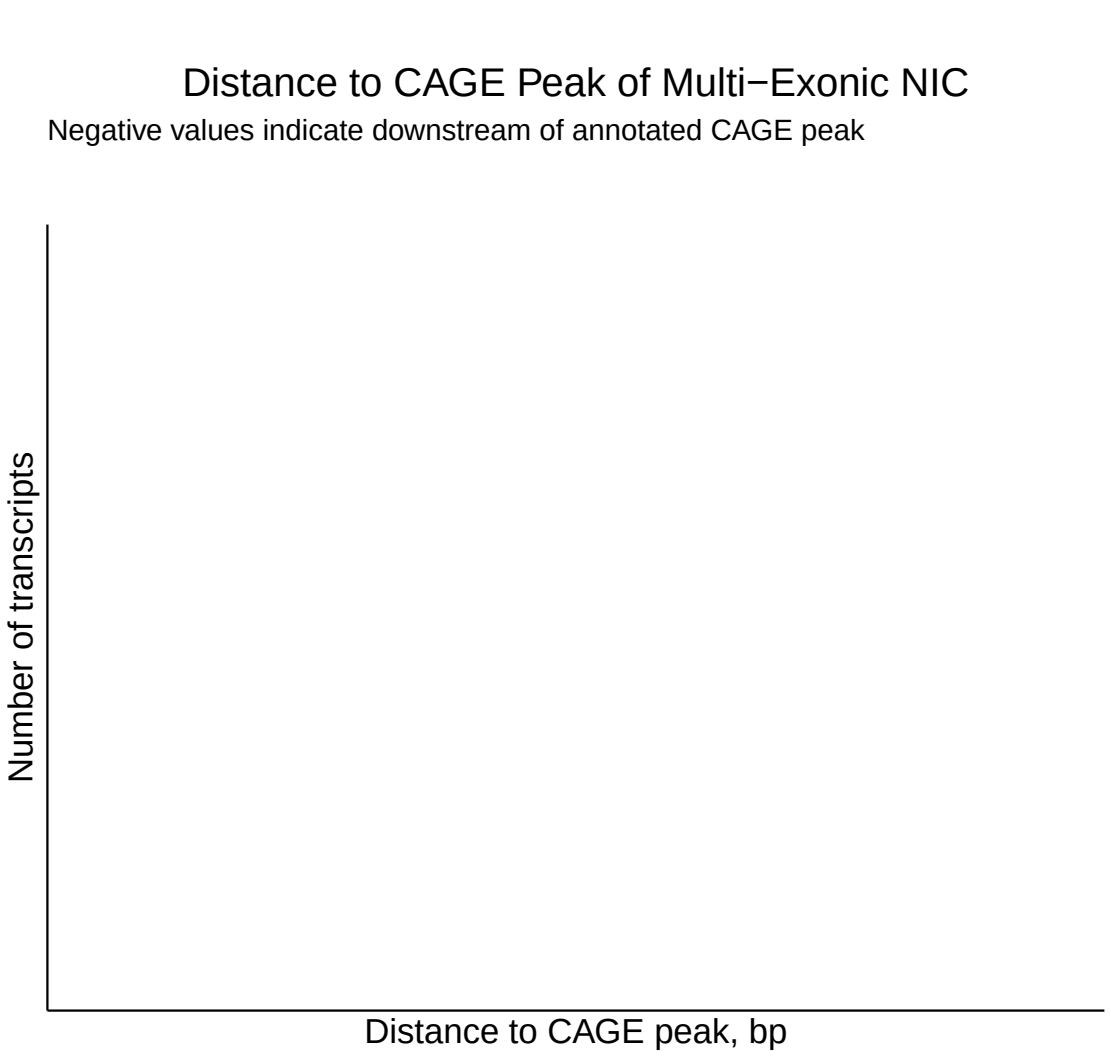
CAGE Distances Analysis

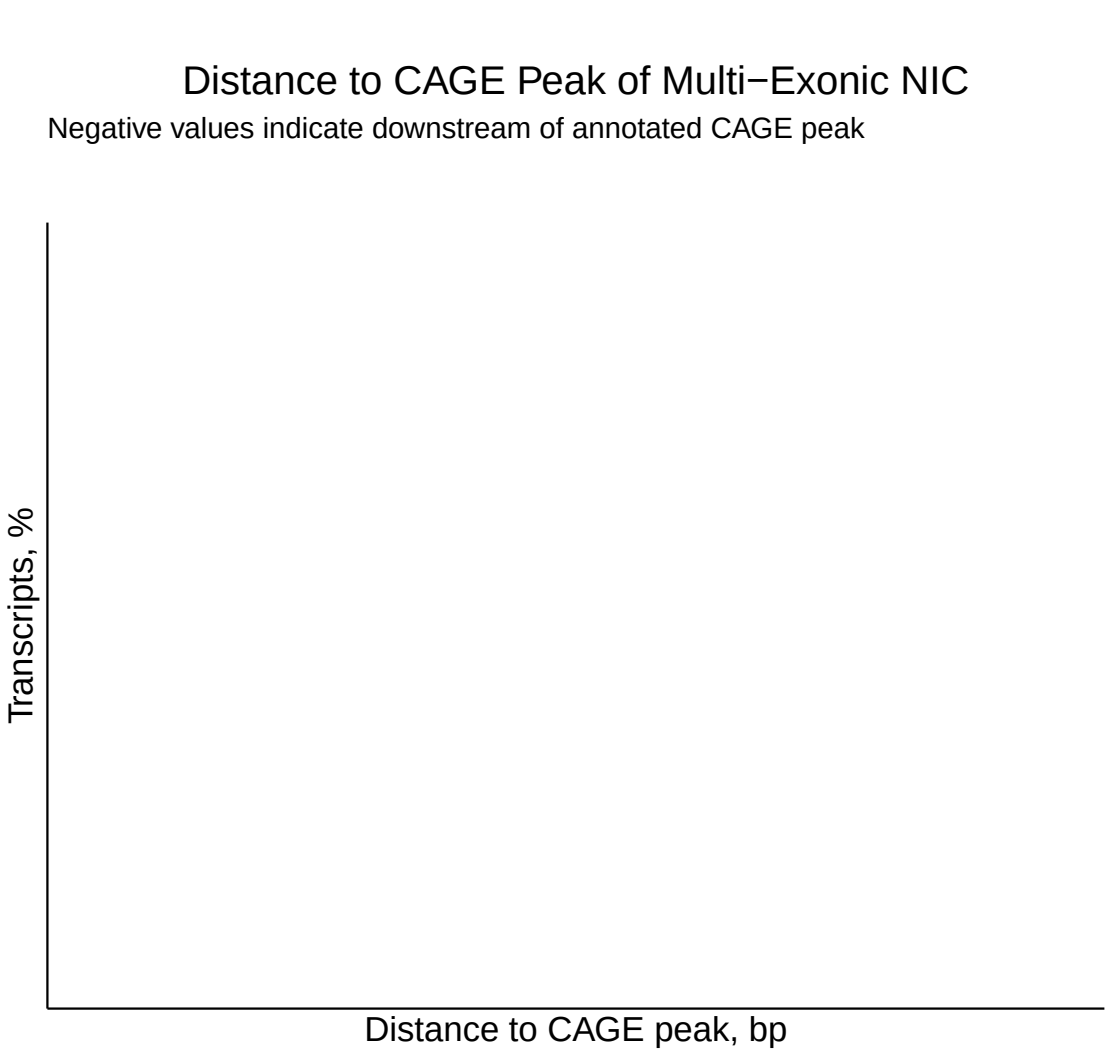

Distance to CAGE Peak of Multi−Exonic FSM
Negative values indicate downstream of annotated CAGE peak
Number of transcripts
Distance to CAGE peak, bp

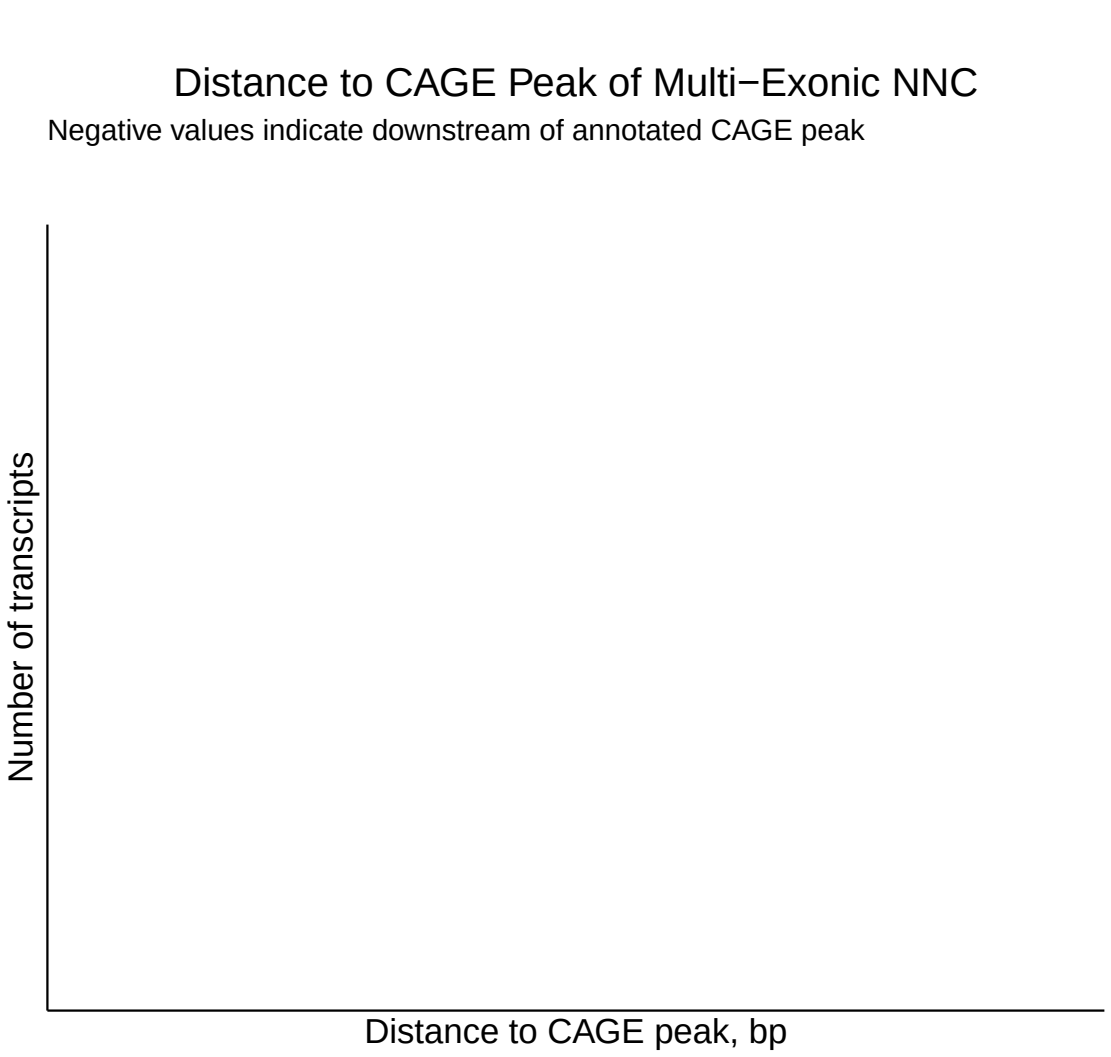

Distance to CAGE Peak of Multi−Exonic FSM
Negative values indicate downstream of annotated CAGE peak
Transcripts, %
Distance to CAGE peak, bp

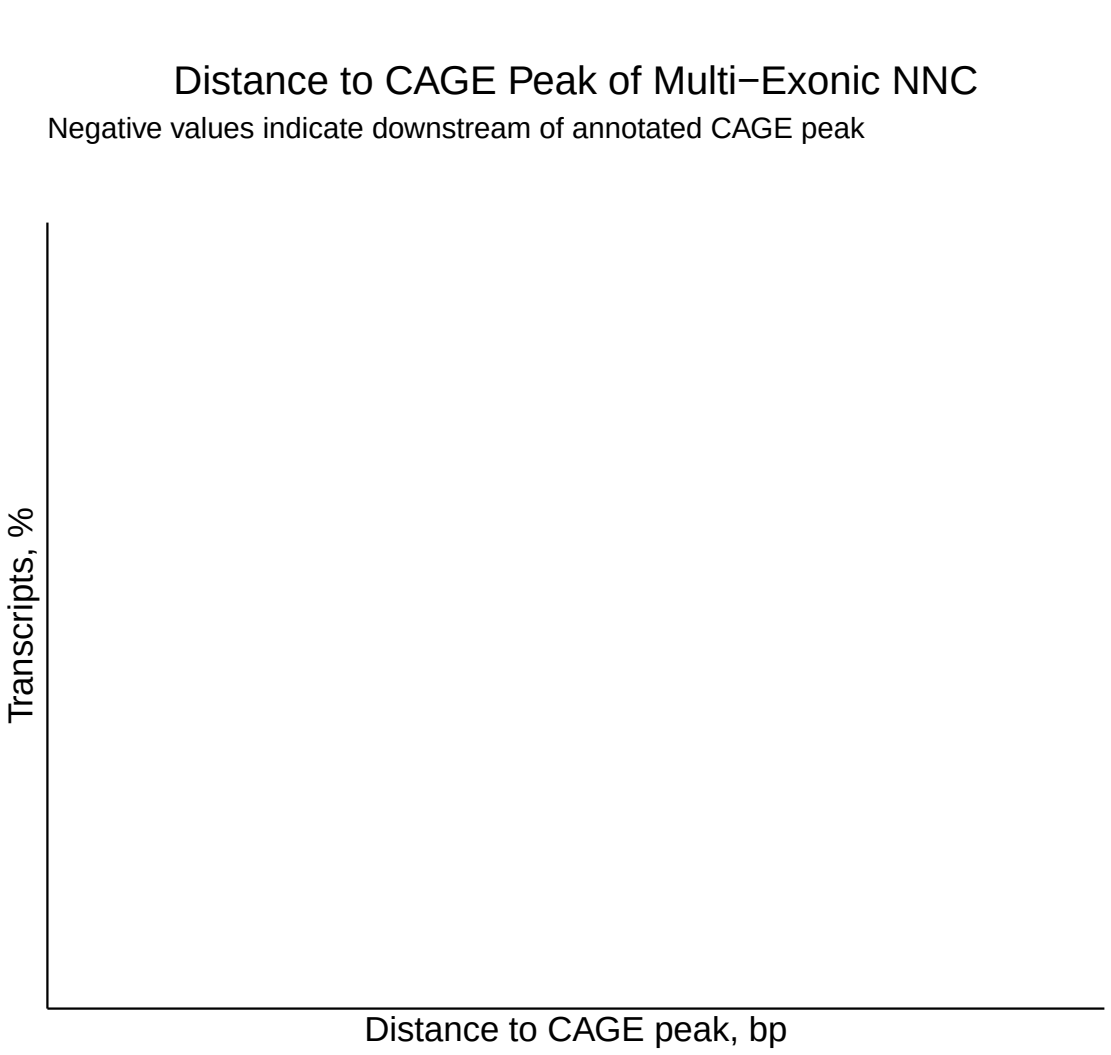

Distance to CAGE Peak of Multi−Exonic ISM
Negative values indicate downstream of annotated CAGE peak
Number of transcripts
Distance to CAGE peak, bp


Distance to CAGE Peak of Multi−Exonic ISM
Negative values indicate downstream of annotated CAGE peak
Transcripts, %
Distance to CAGE Peak, bp


Distance to CAGE Peak of Multi−Exonic NIC
Negative values indicate downstream of annotated CAGE peak
Number of transcripts
Distance to CAGE peak, bp


Distance to CAGE Peak of Multi−Exonic NIC
Negative values indicate downstream of annotated CAGE peak
Transcripts, %
Distance to CAGE peak, bp


Distance to CAGE Peak of Multi−Exonic NNC
Negative values indicate downstream of annotated CAGE peak
Number of transcripts
Distance to CAGE peak, bp


Distance to CAGE Peak of Multi−Exonic NNC
Negative values indicate downstream of annotated CAGE peak
Transcripts, %
Distance to CAGE peak, bp

Number of CAGE Detected

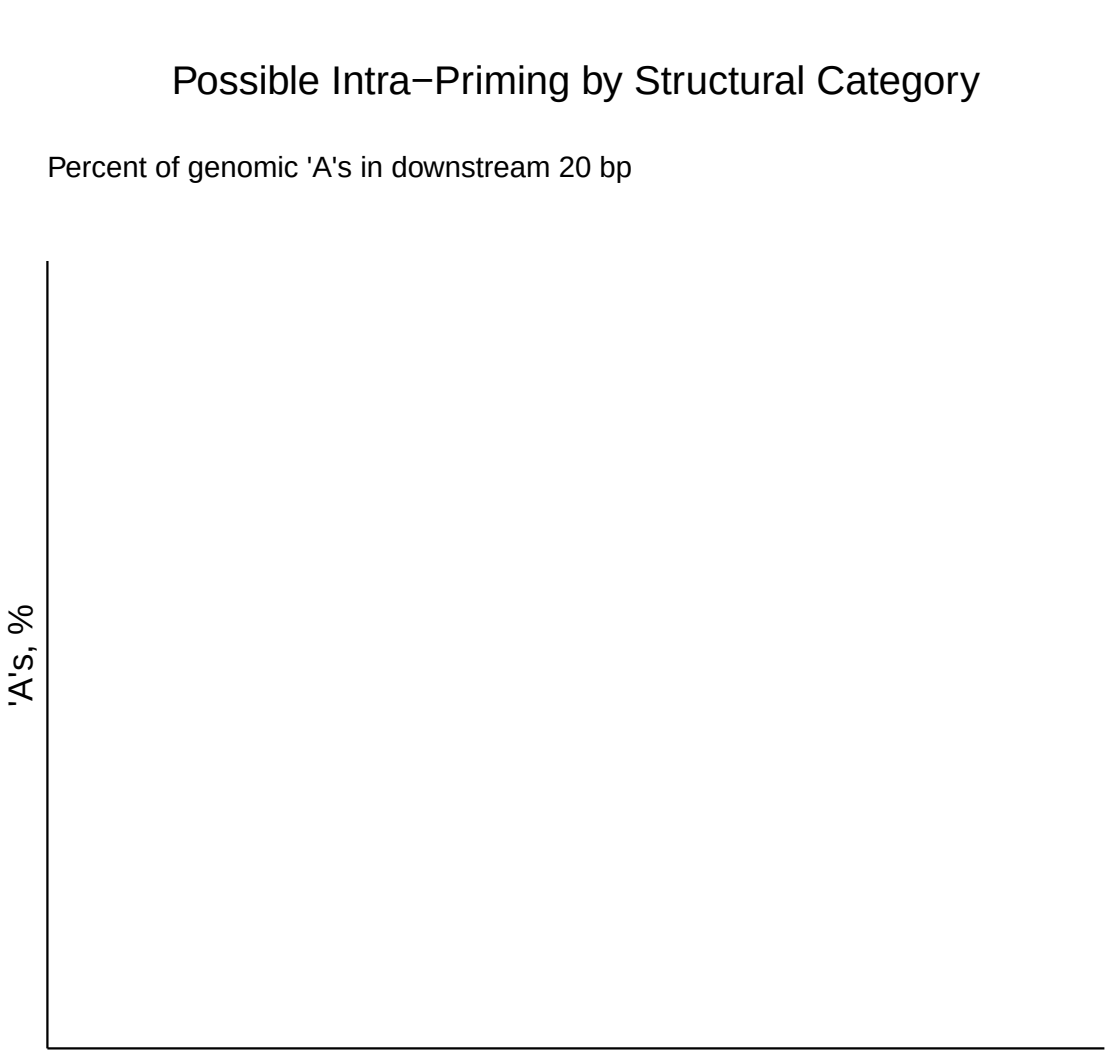
Category	Count	CAGE Detected	%
NA	1609	1242	77

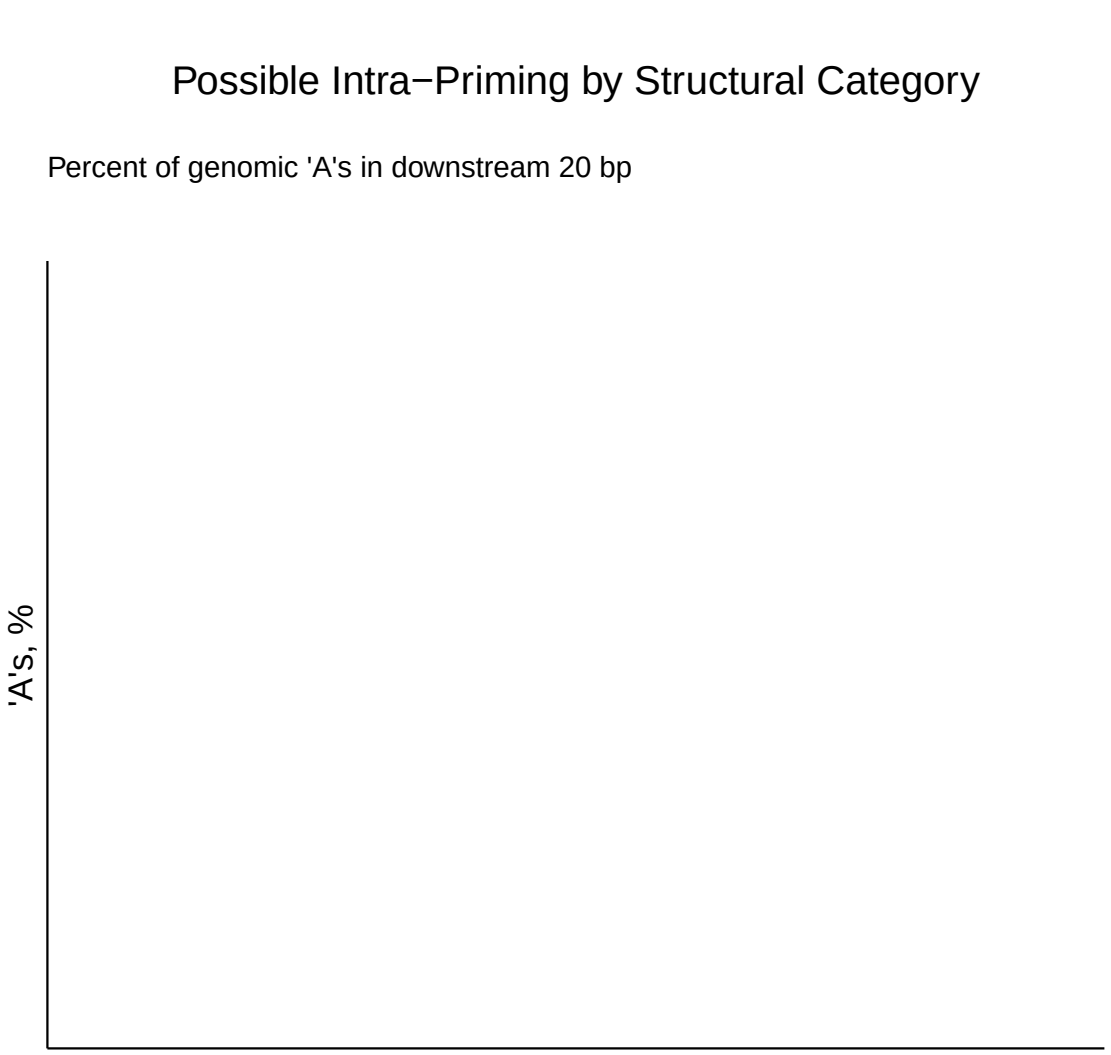
Number of CAGE Detected

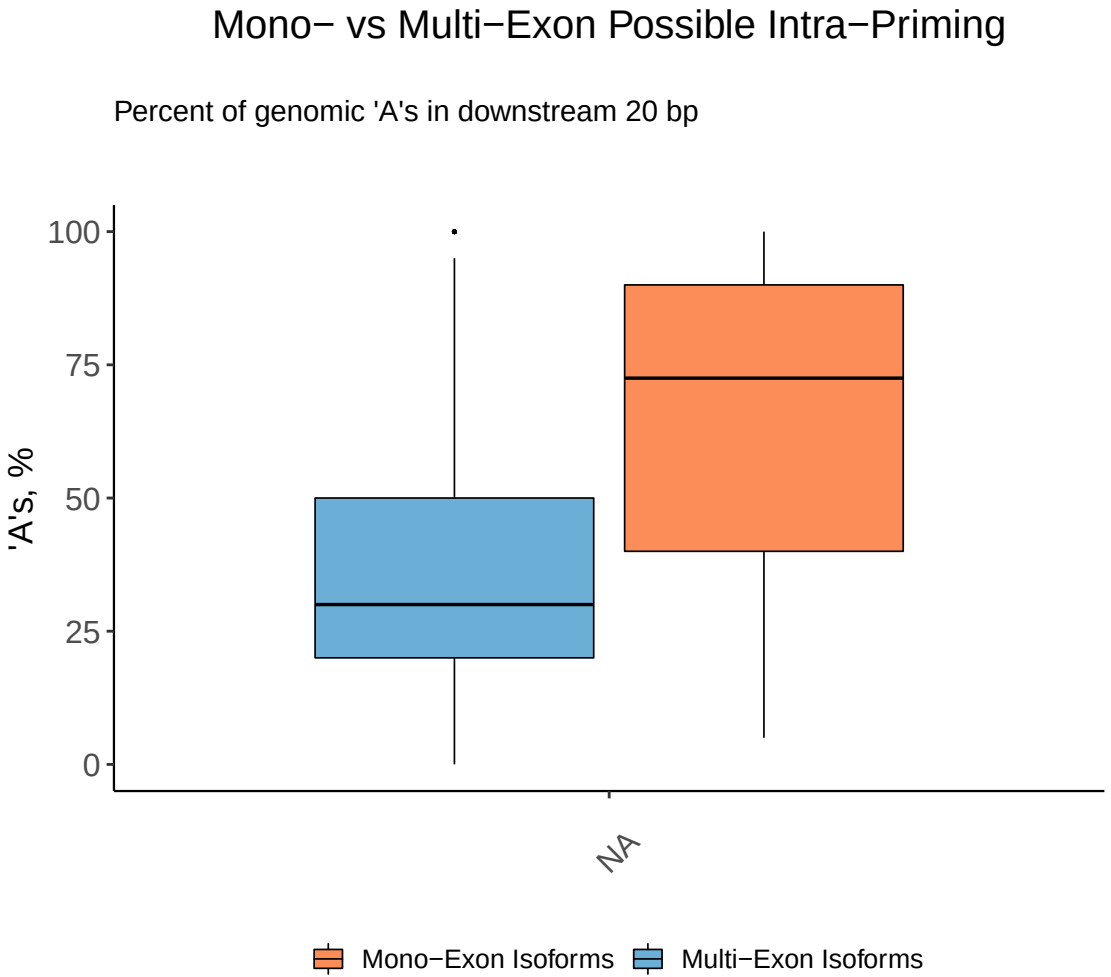
Subcategory	Count	CAGE Detected	%
3' fragment	22	3	14
Internal fragment	2	0	0
5' fragment	109	90	83
Combin. of annot. junctions	119	98	82
Combin. of annot. splice sites	166	135	81
Intron retention	186	155	83
Mono−exon by intron retention	8	6	75
At least 1 annot. donor/accept.	339	252	74
Mono−exon	80	40	50
Multi−exon	578	463	80

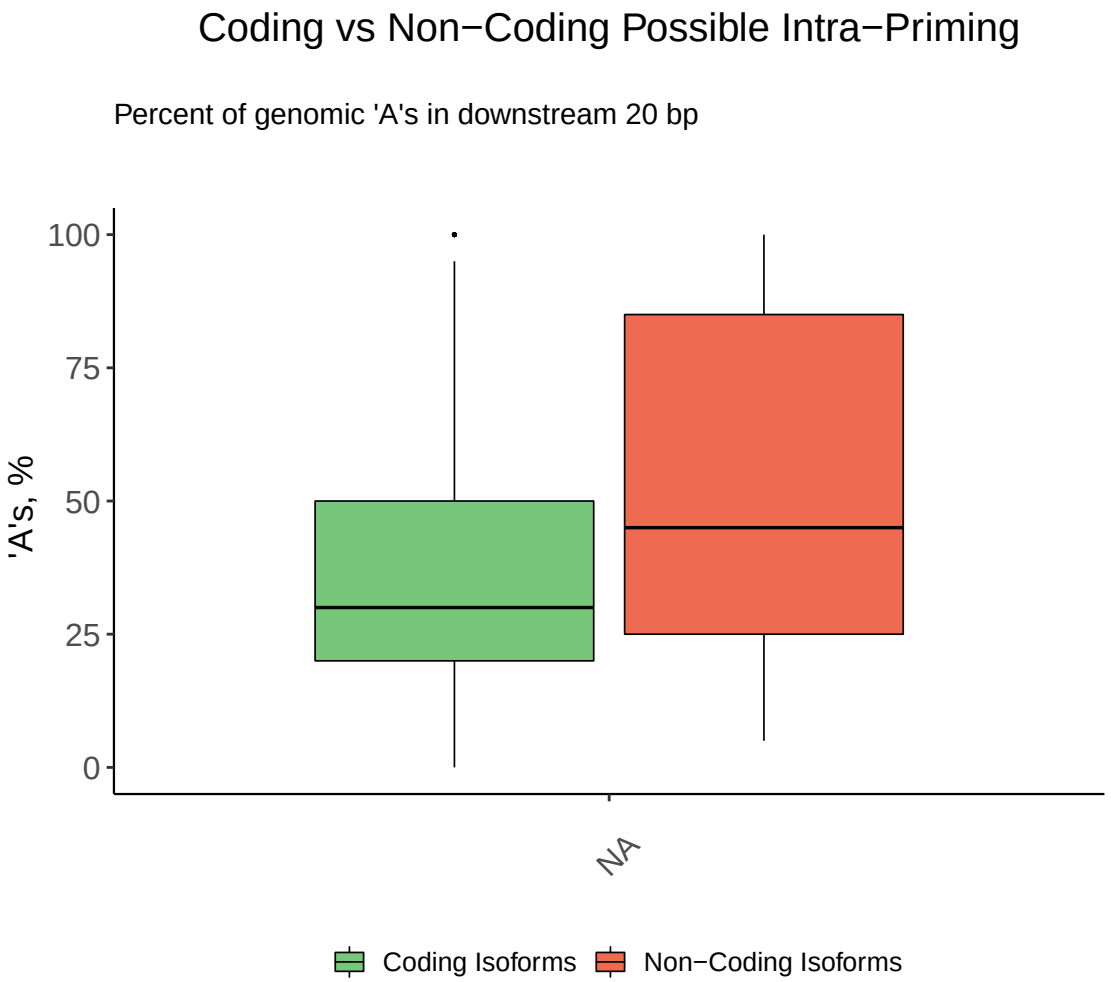
Redundancy Analysis

Intra−Priming Quality Check

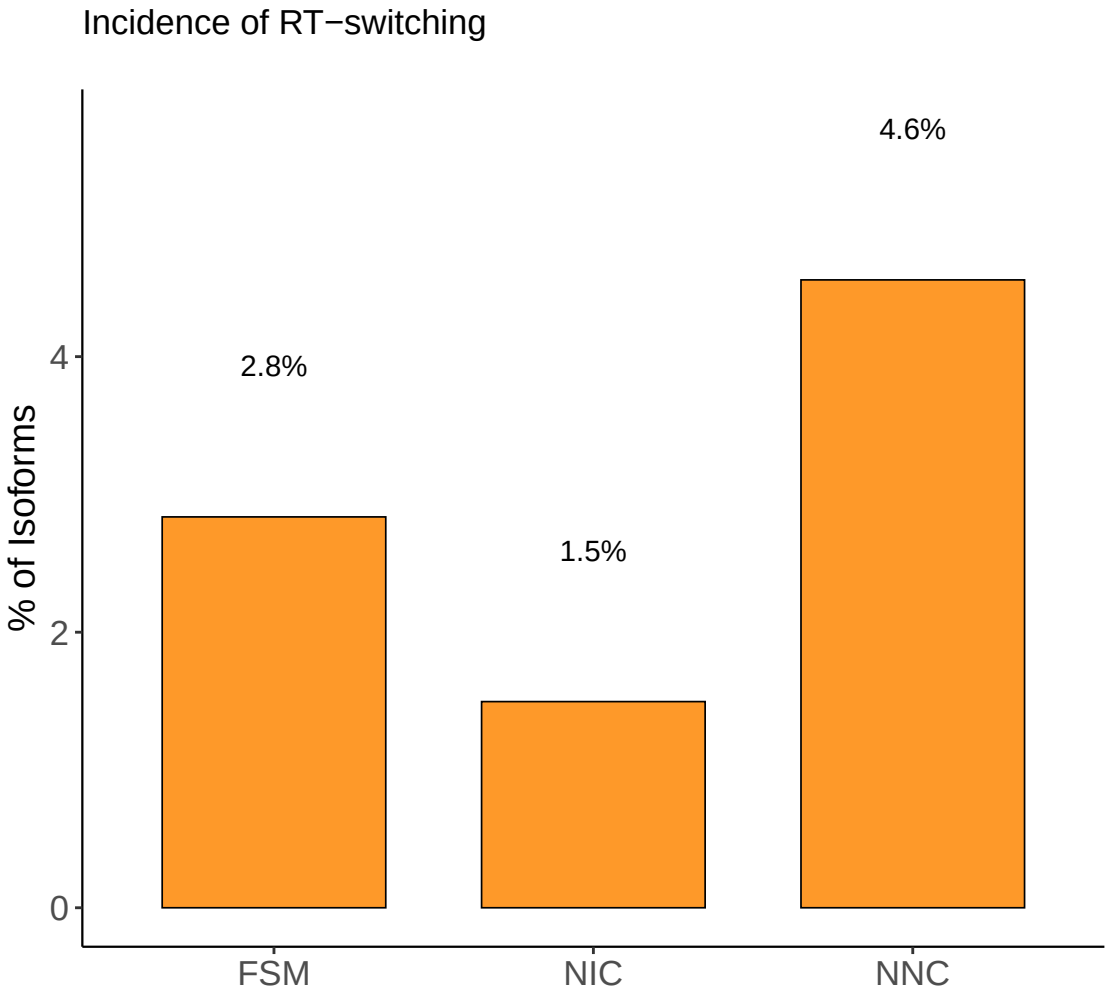

Possible Intra−Priming by Structural Category
Percent of genomic 'A's in downstream 20 bp
'A's, %

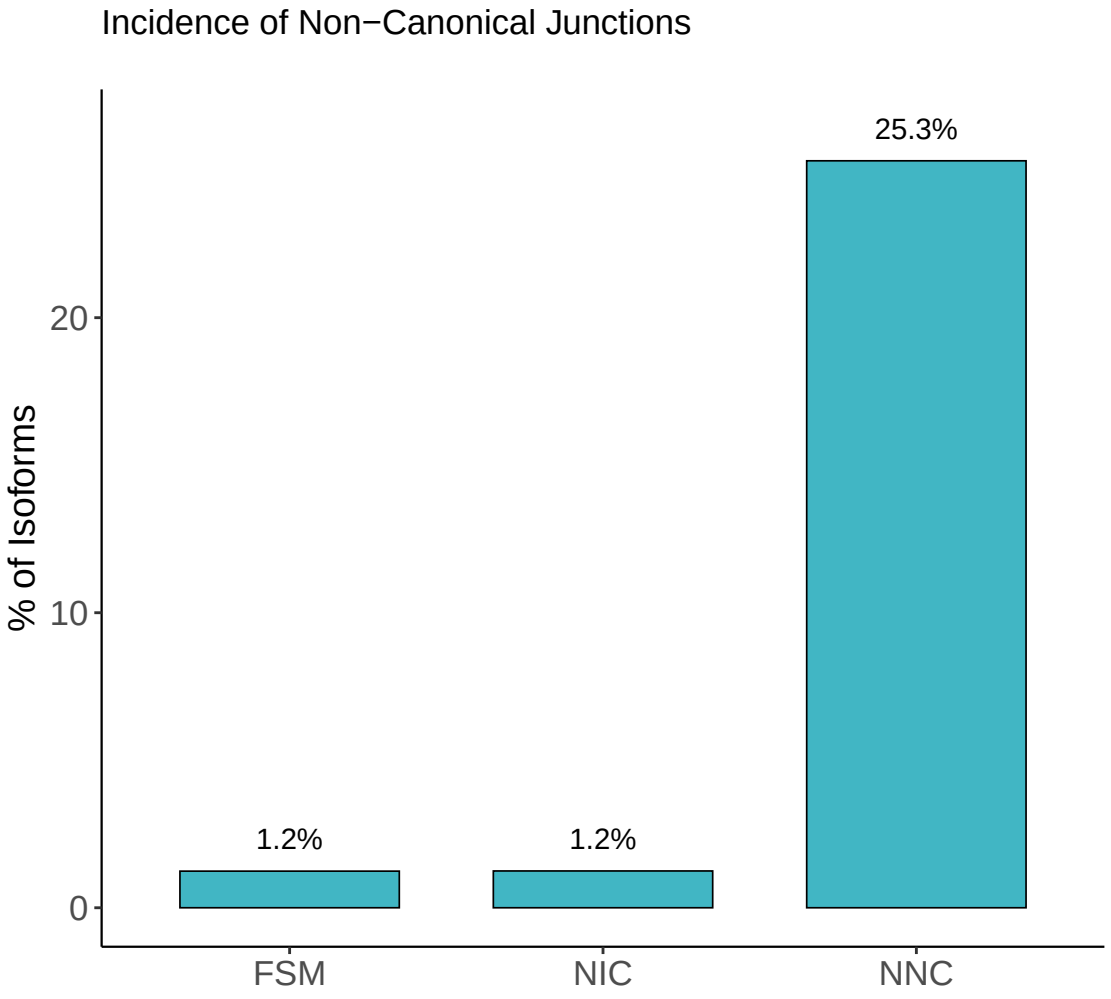

Possible Intra−Priming by Structural Category
Percent of genomic 'A's in downstream 20 bp
'A's, %


Mono− vs Multi−Exon Possible Intra−Priming
Percent of genomic 'A's in downstream 20 bp
'A's, %
100
75
50
25
0
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
NA
Coding Isoforms
Non−Coding Isoforms

Bad Quality Controls


Incidence of RT−switching
% of Isoforms
0
2
4
2.8%
1.5%
4.6%
FSM
NIC
NNC


Incidence of Non−Canonical Junctions
% of Isoforms
0
10
20
1.2%
1.2%
25.3%
FSM
NIC
NNC

Good Quality Controls