

# *SQANTI3 report*

*Unique Genes: 656*  
*Unique Isoforms: 3925*

### *Transcript Classification*

Category	Isoforms, count
FSM	539
ISM	1138
NIC	744
NNC	1139
Genic Genomic	113
Antisense	26
Fusion	38
Intergenic	41
Genic Intron	147

### *Gene Classification*

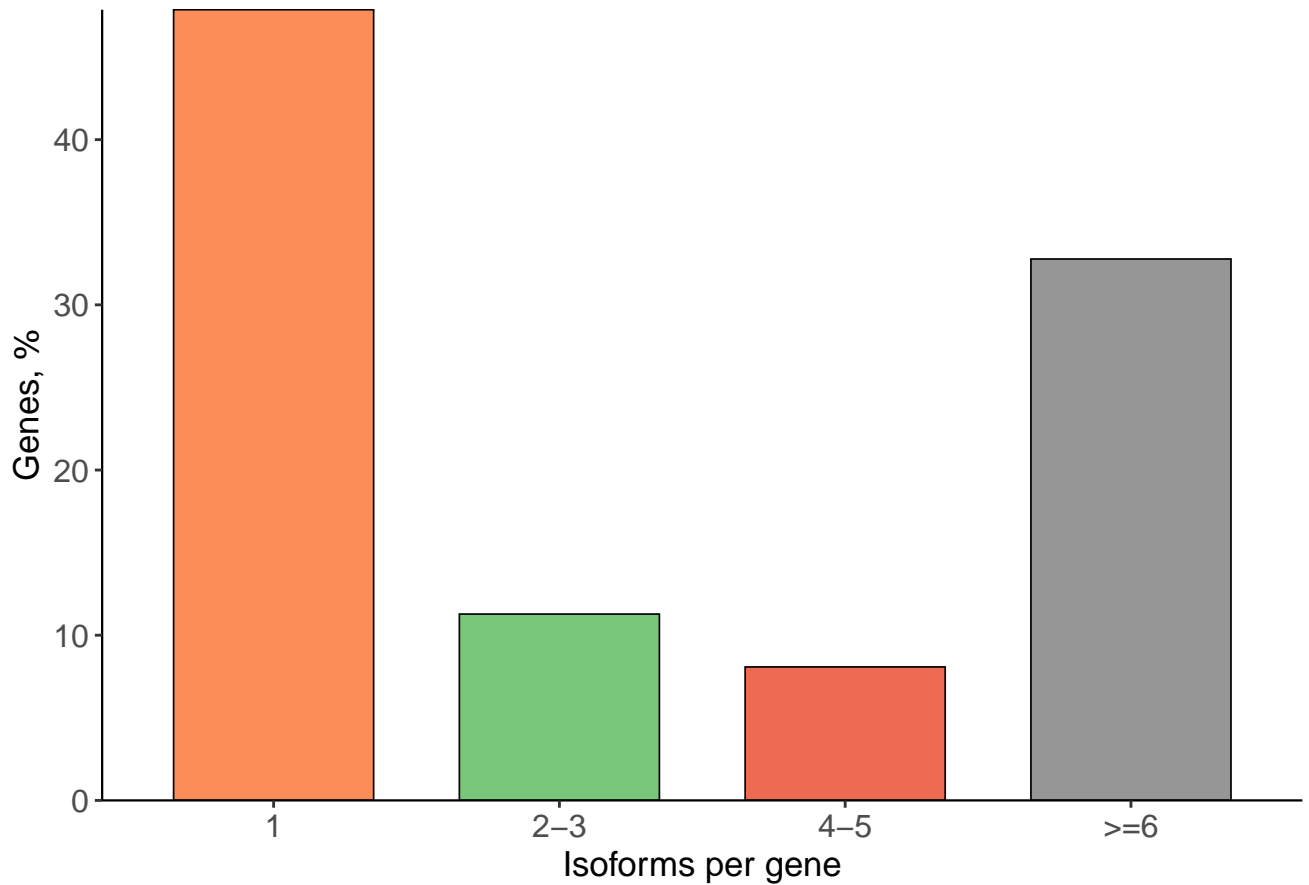
Category	Genes, count
Annotated Genes	443
Novel Genes	213

### *Splice Junction Classification*

Category	SJs, count	Percent
Known canonical	3374	69.74
Known Non–canonical	4	0.08
Novel canonical	1195	24.70
Novel Non–canonical	265	5.48

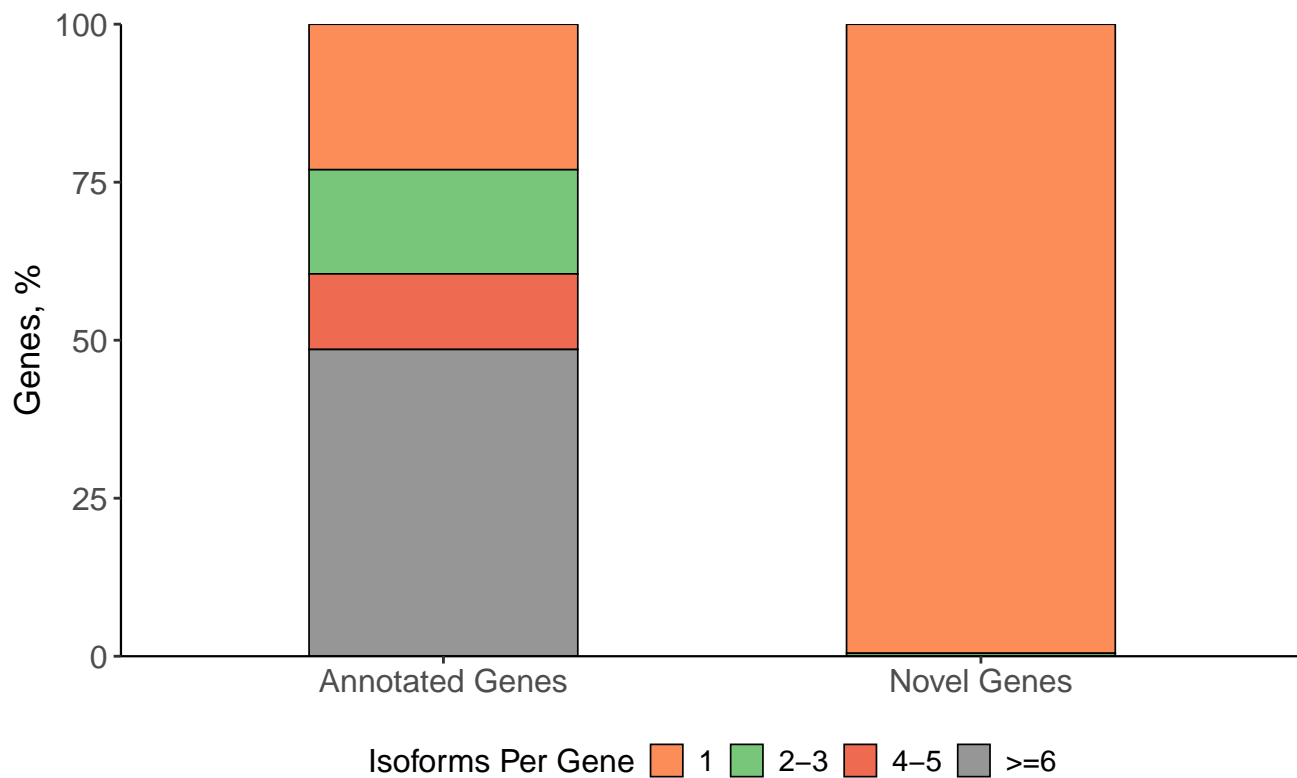
## *Gene Characterization*

## Number of Isoforms per Gene

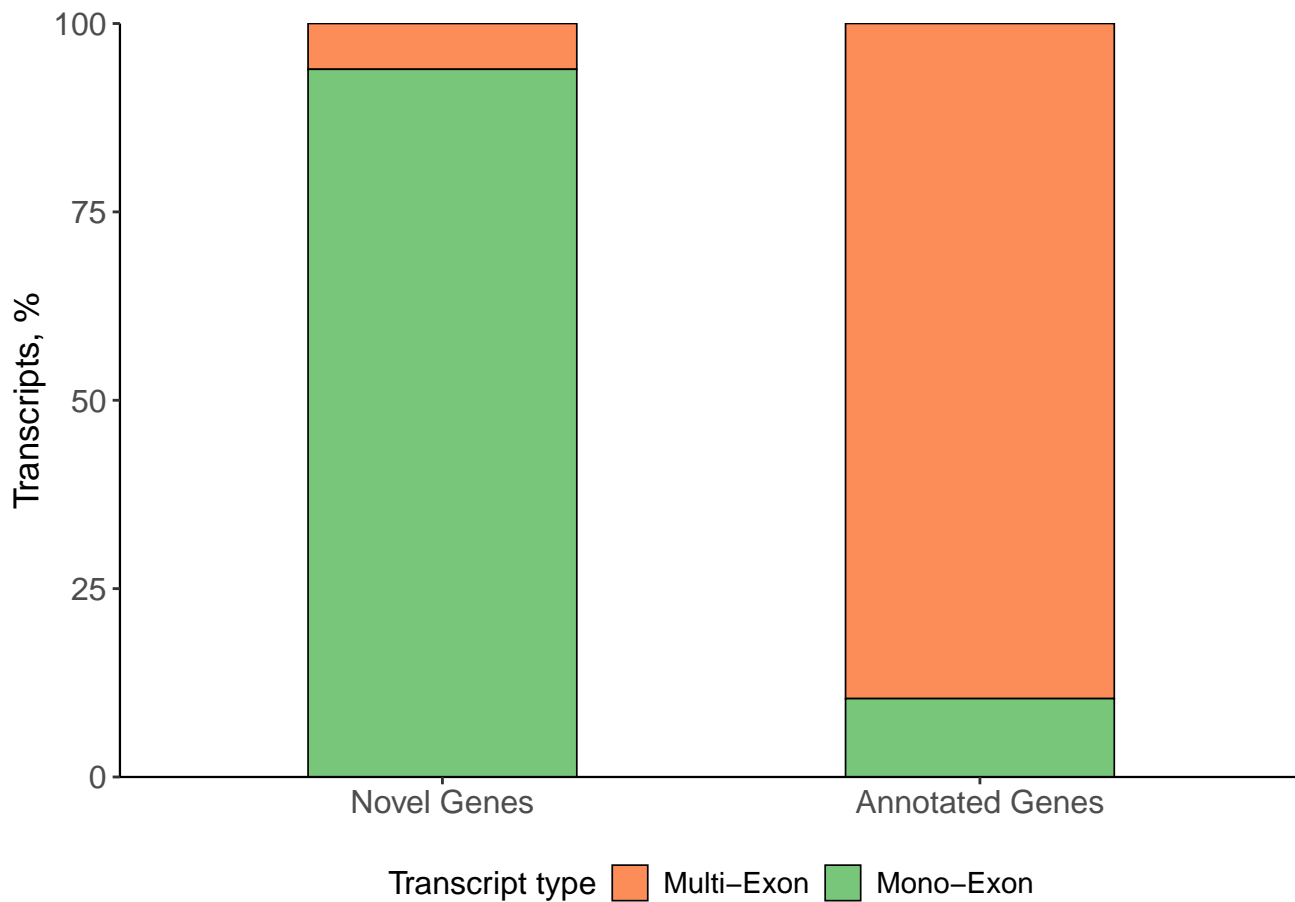


# Number of Isoforms per Gene

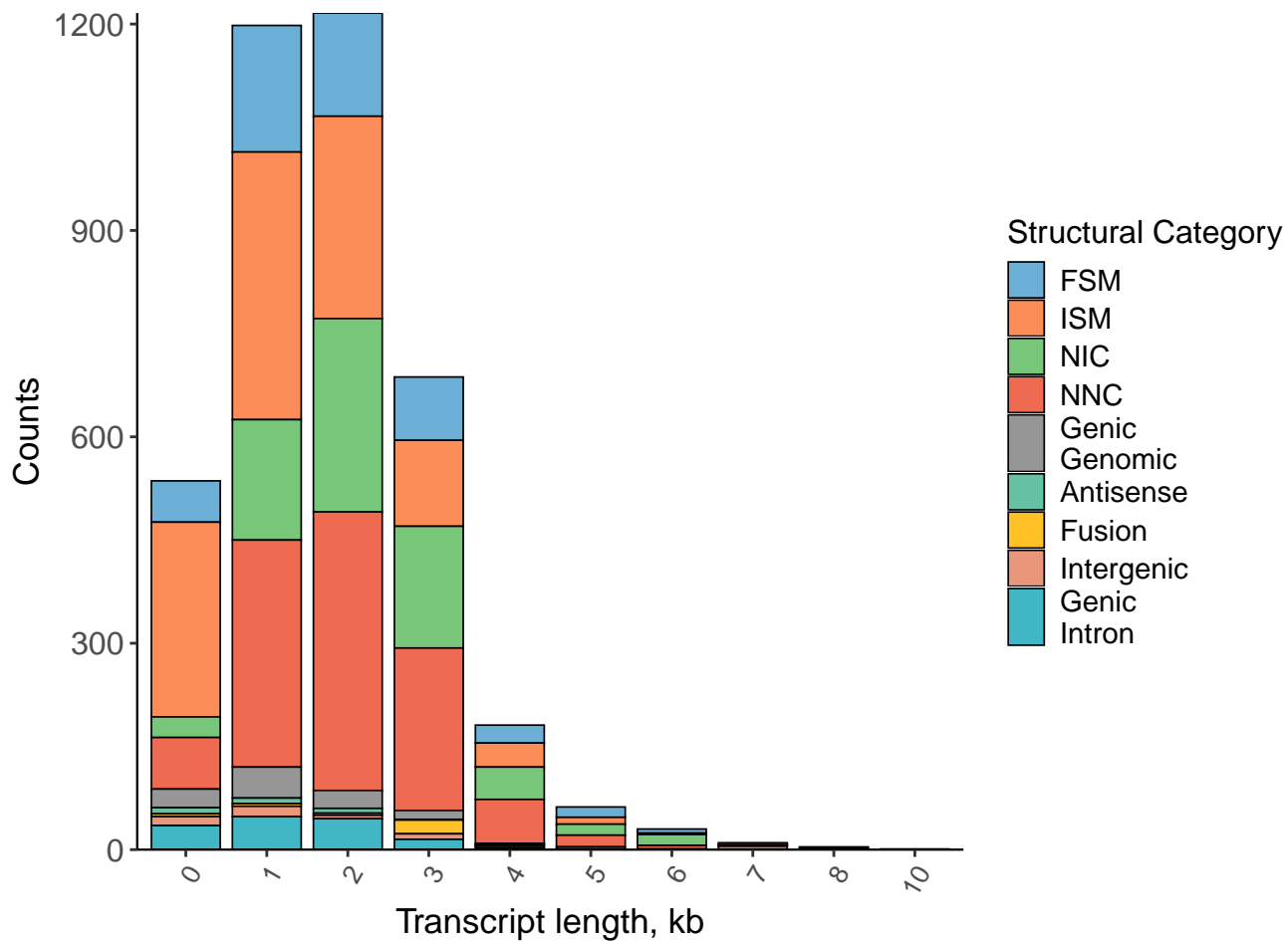
Known vs Novel Genes



Distribution of Mono- vs Multi-Exon Transcripts

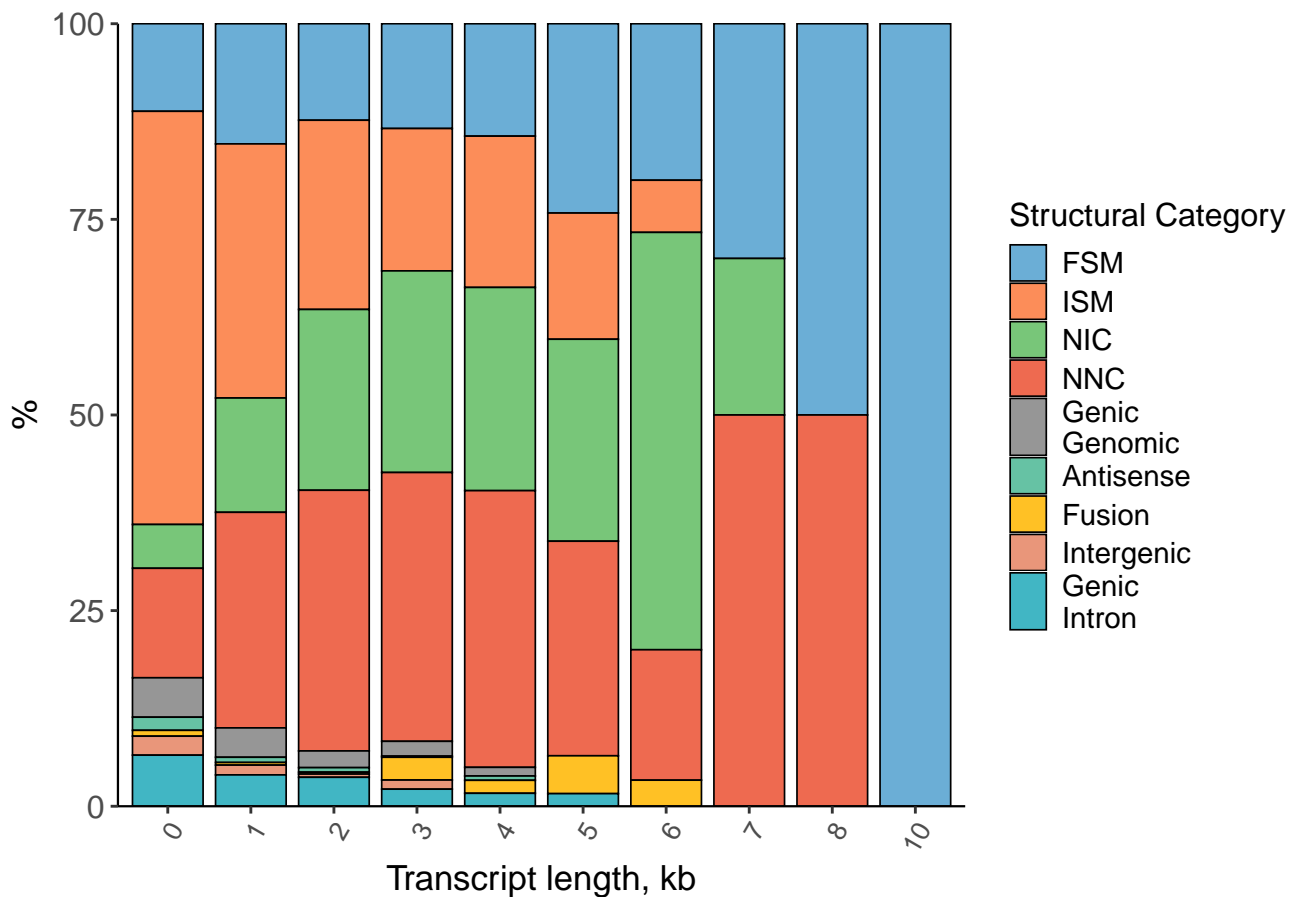


## Structural Categories by Transcript Length

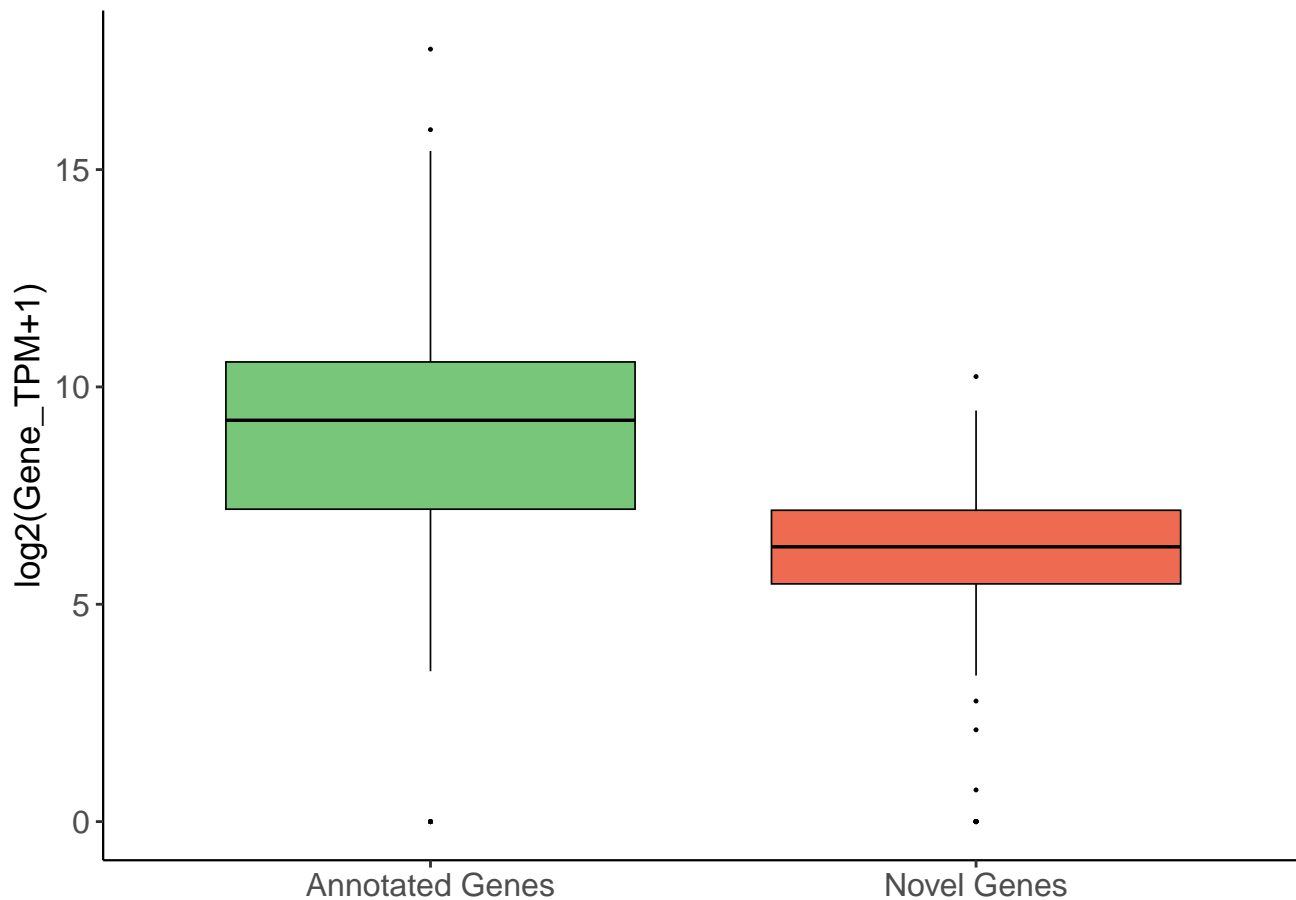




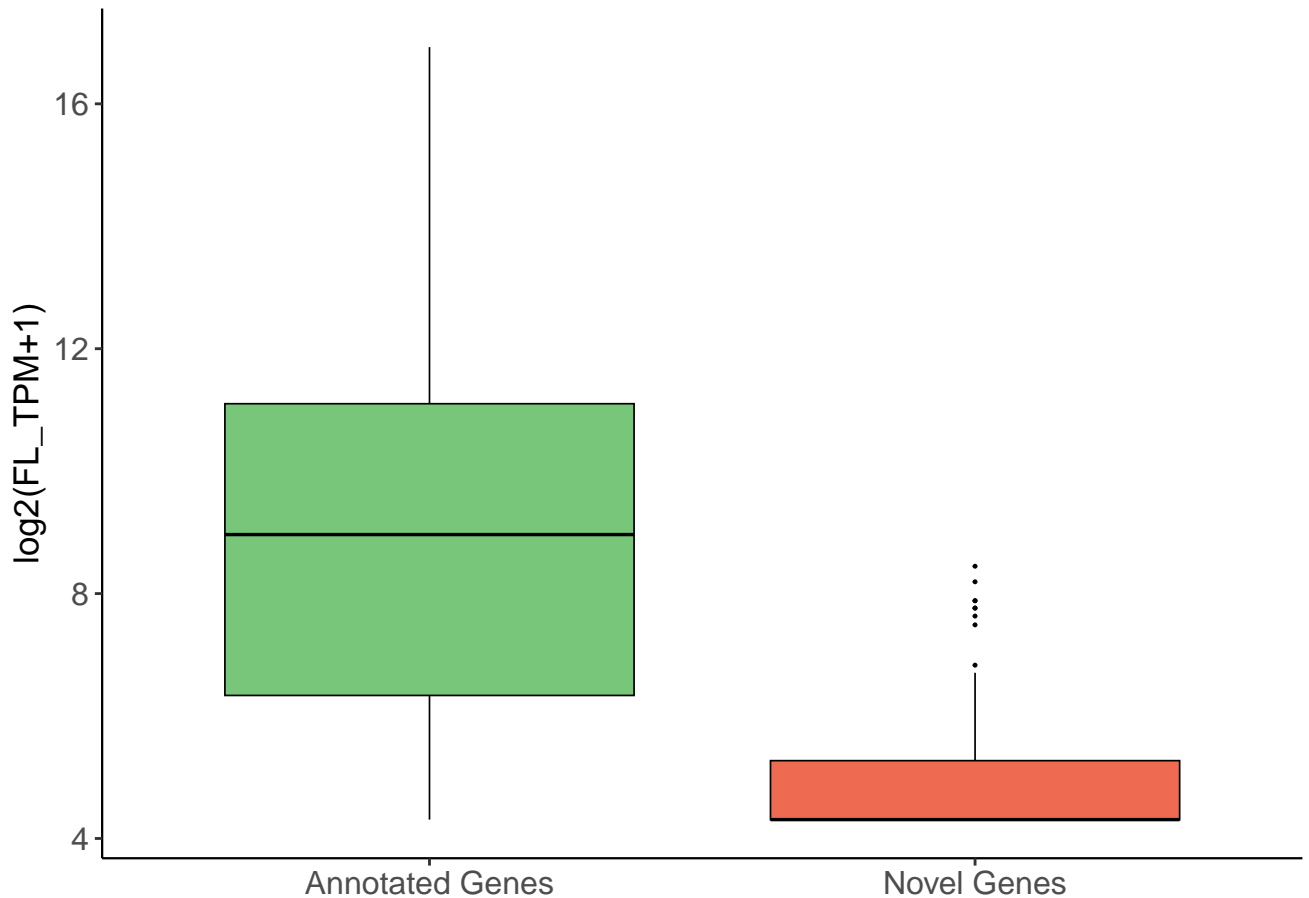
## Structural Categories by Transcript Length



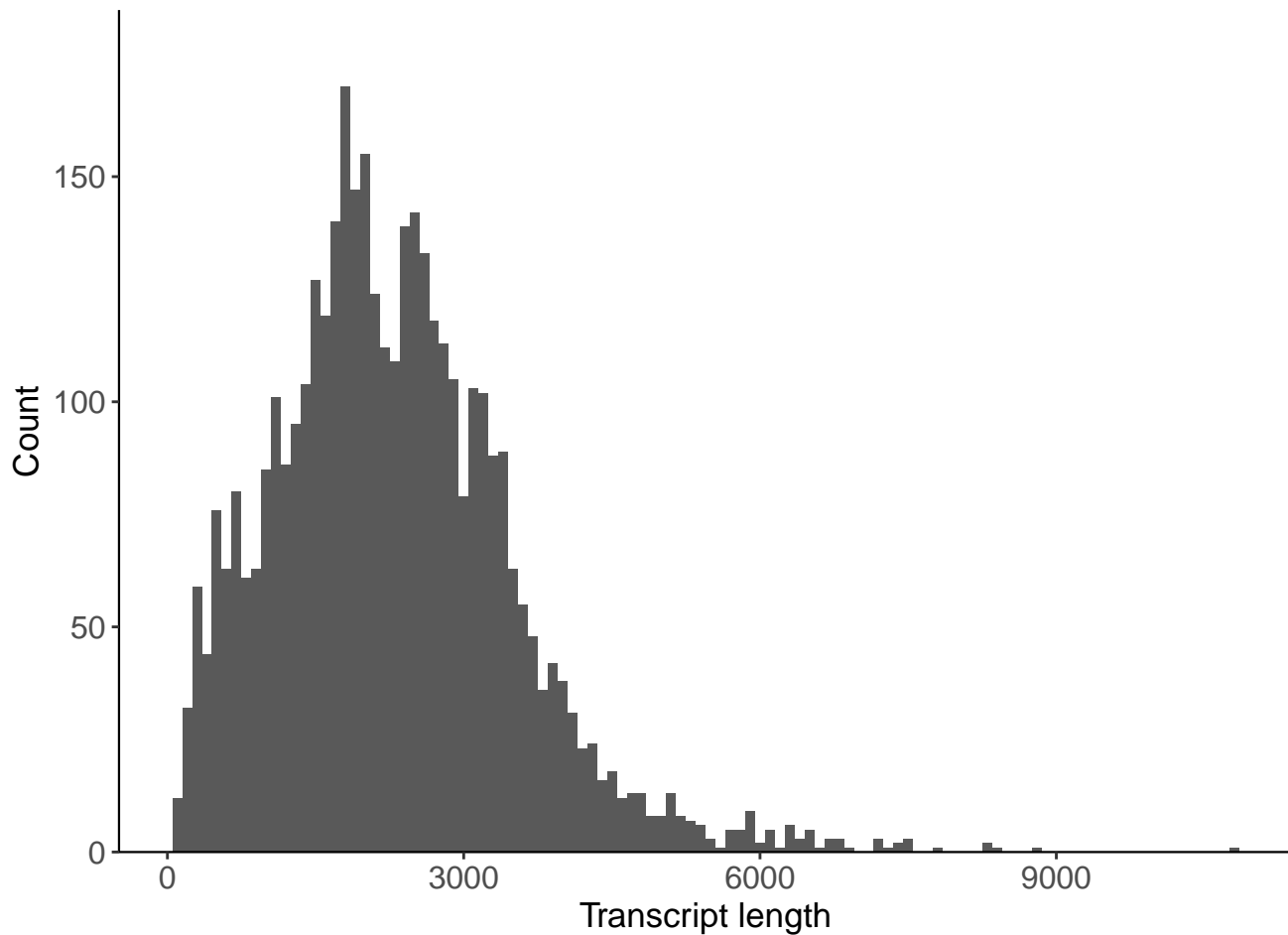
## Annotated vs Novel Gene Expression

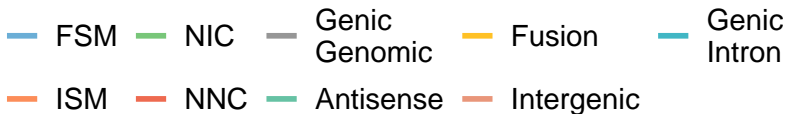


Number of FL reads per Gene by Type of Gene Annotation

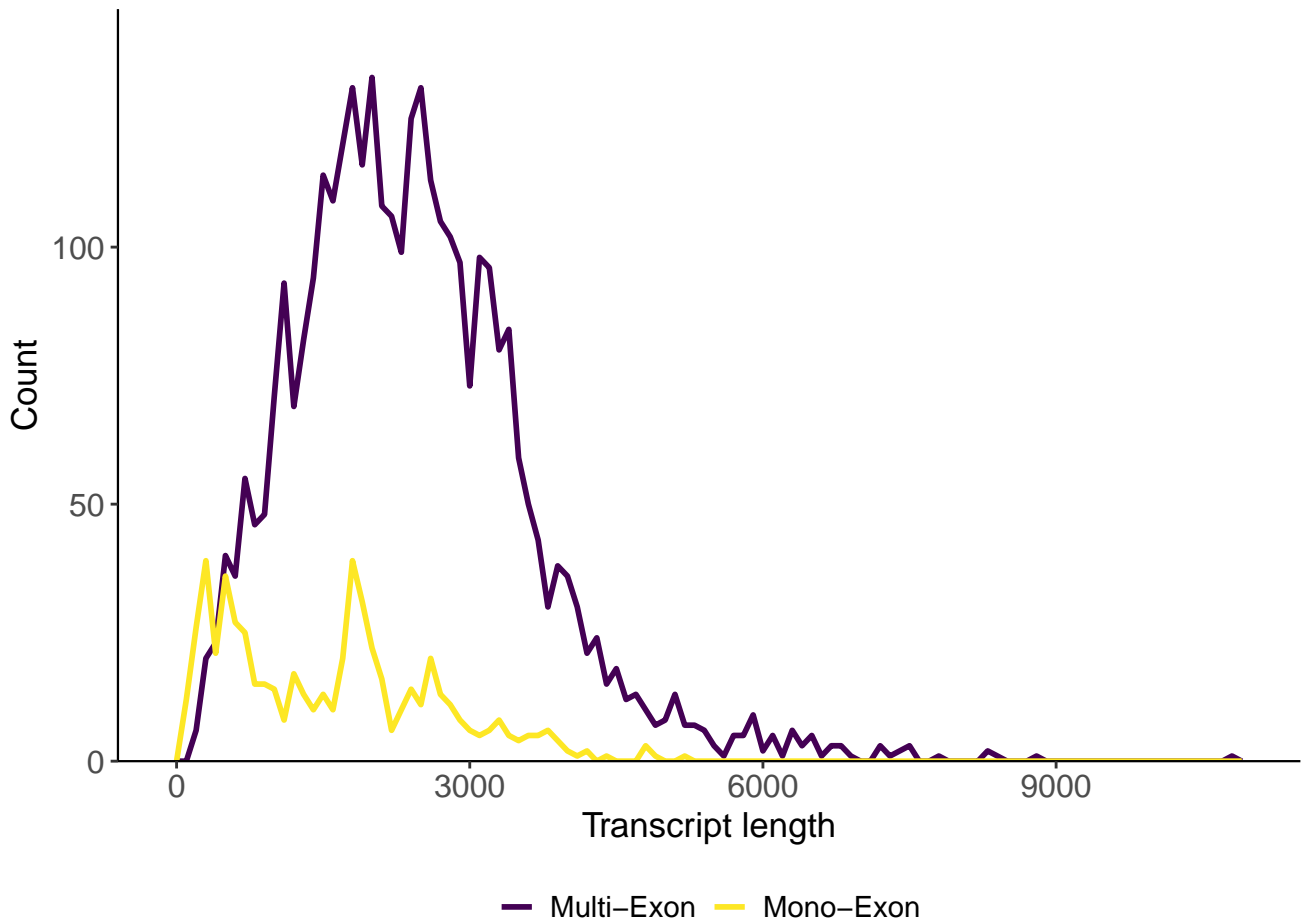


All Transcript Lengths Distribution



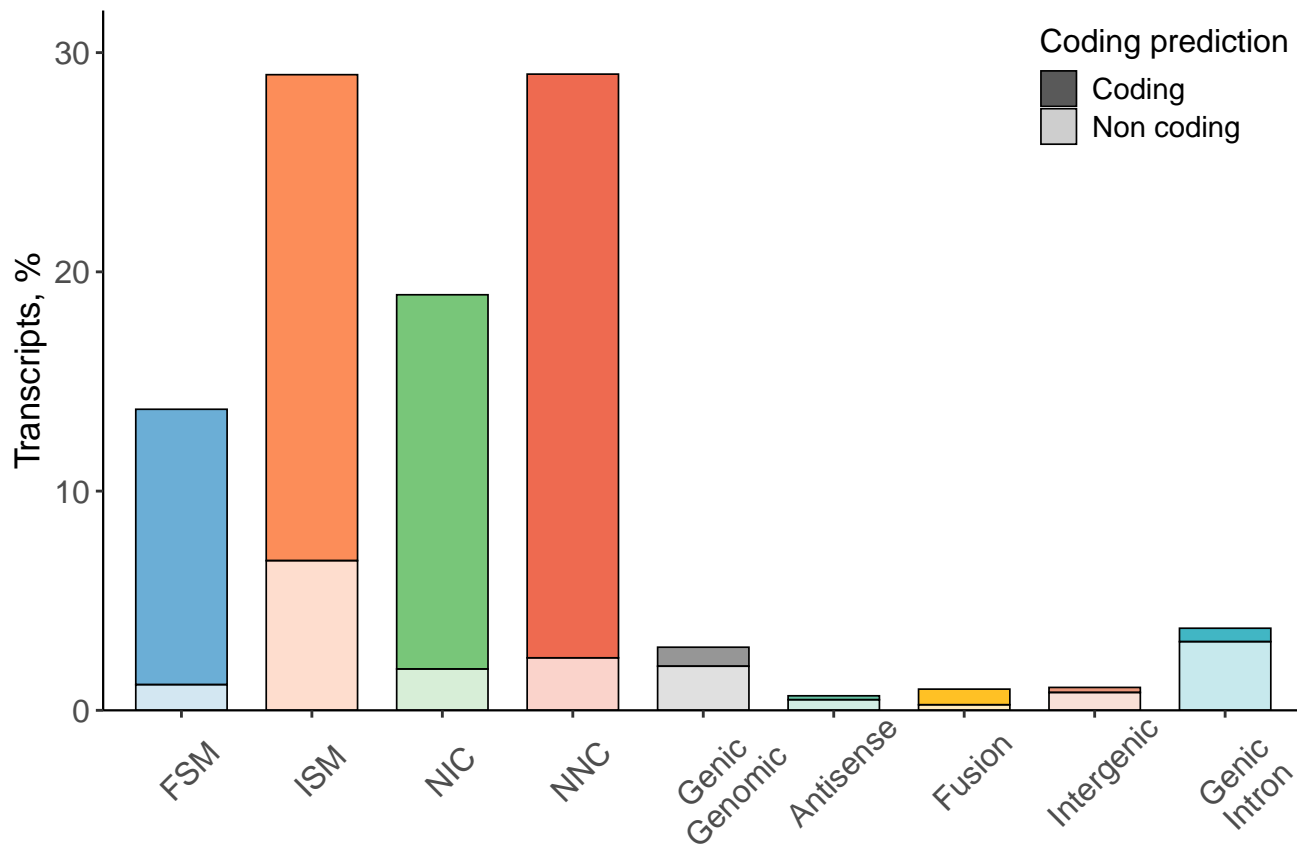


Mono- vs Multi- Exon Transcript Lengths Distribution



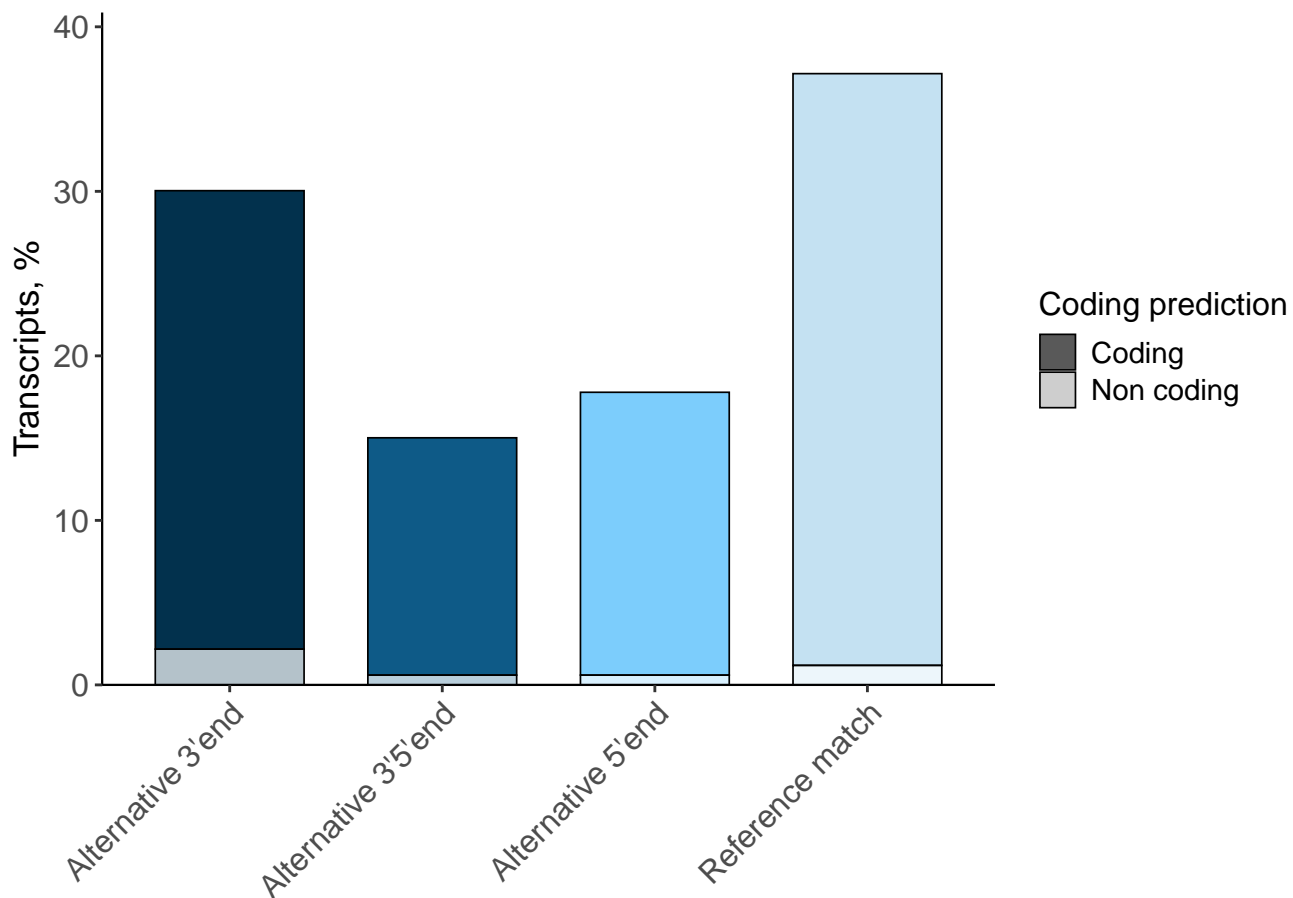
## *Structural Isoform Characterization*

## Isoform Distribution Across Structural Categories

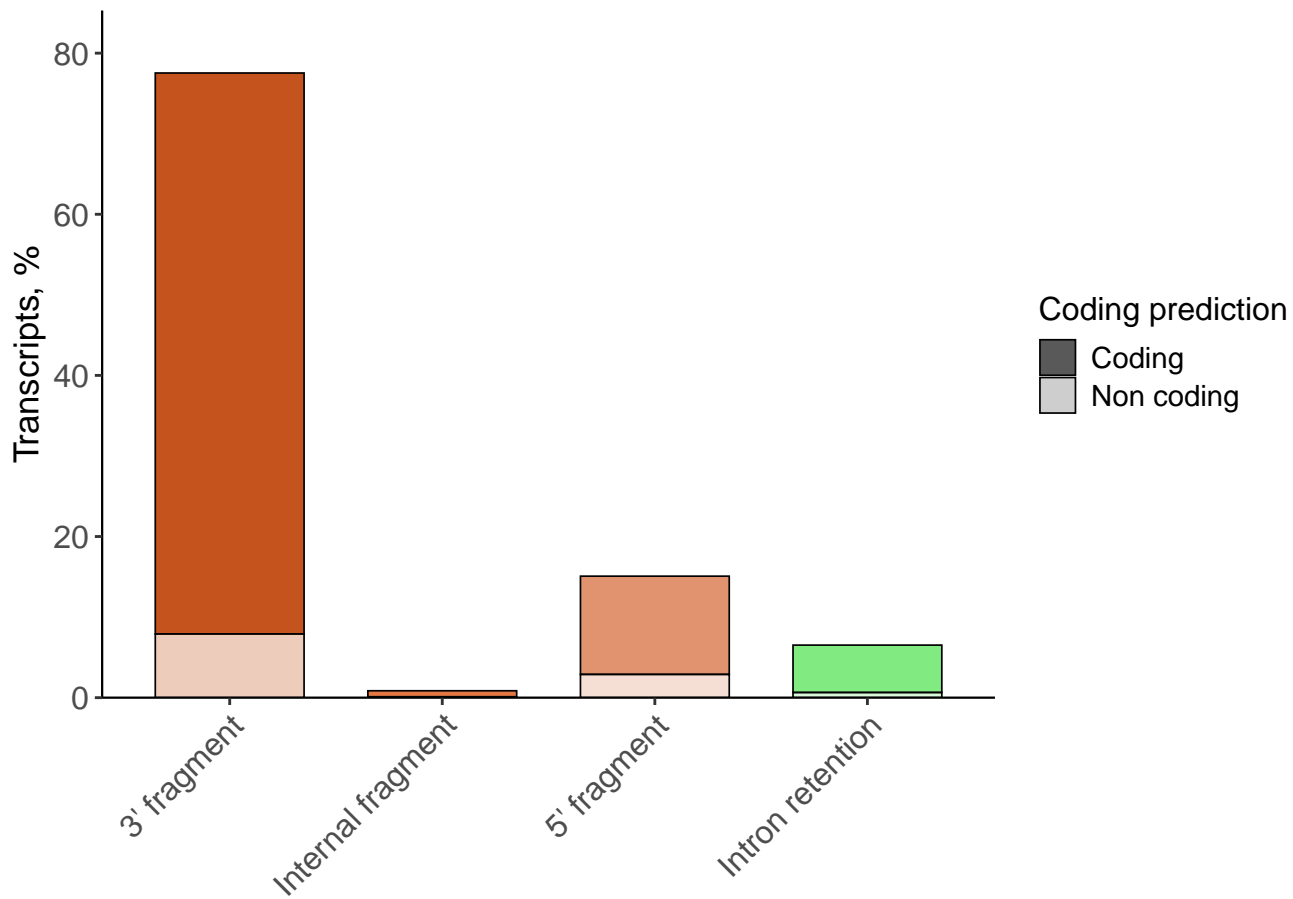




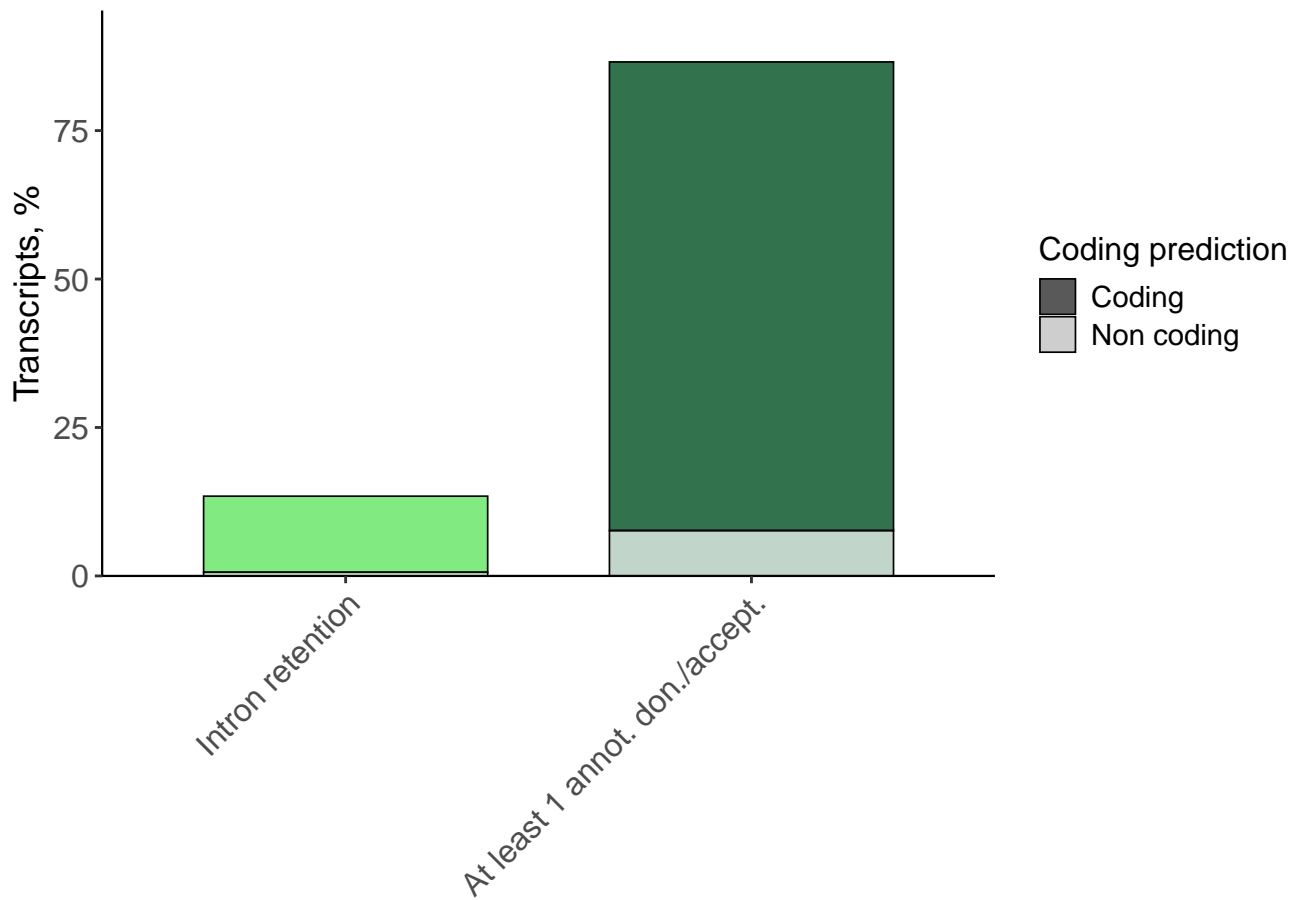
## Isoform Distribution Across FSM



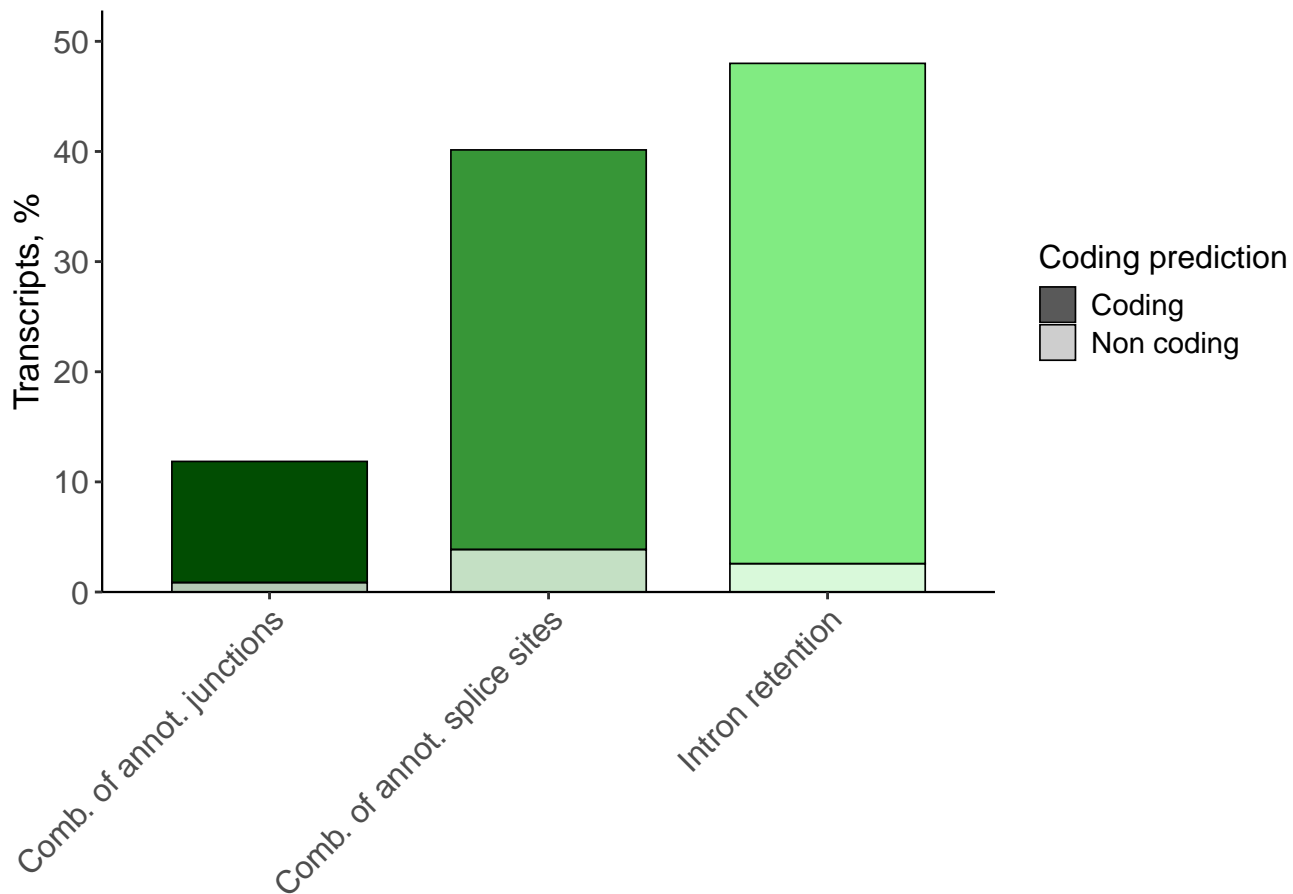
## Isoform Distribution Across ISM



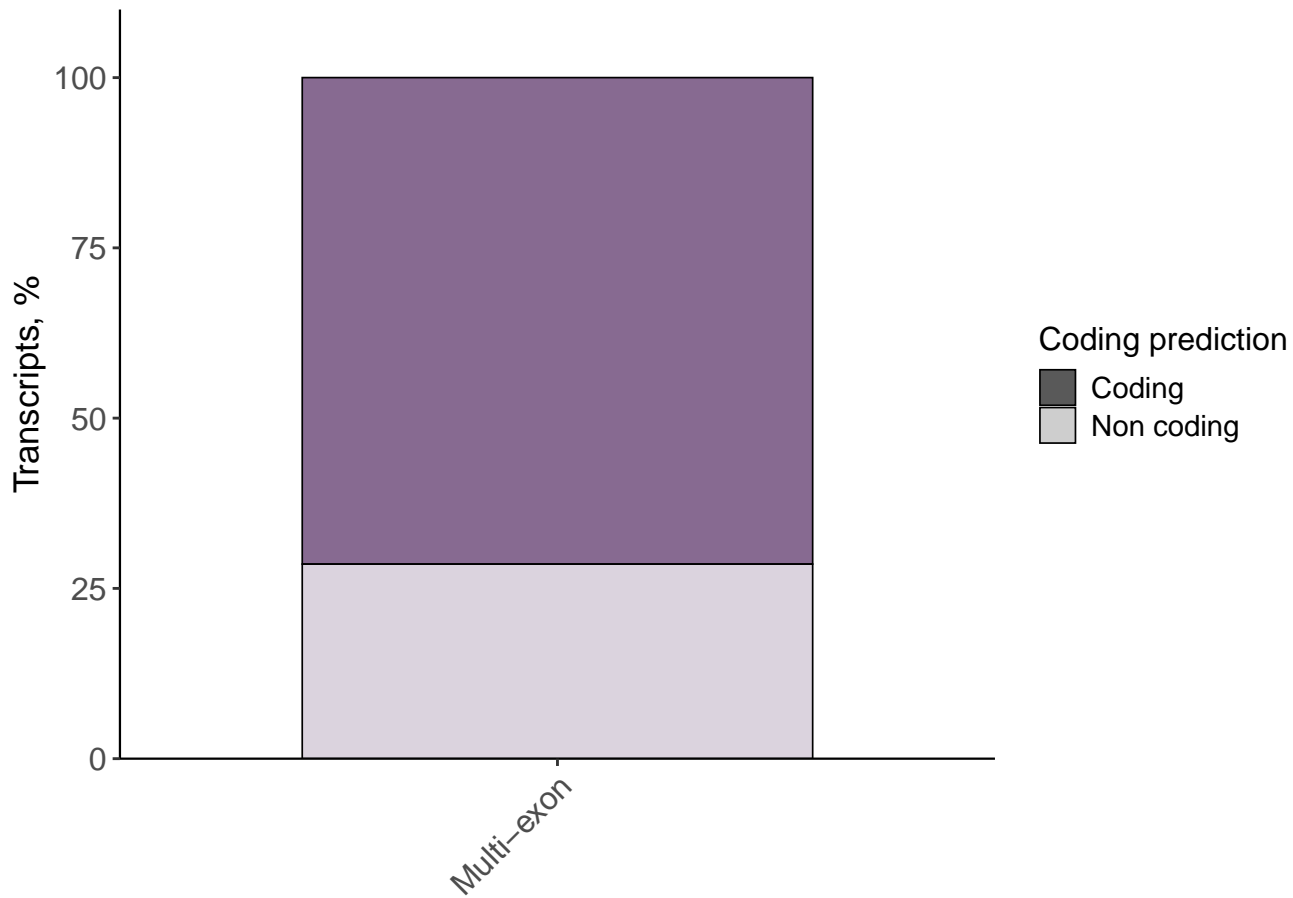
## Isoform Distribution Across NNC



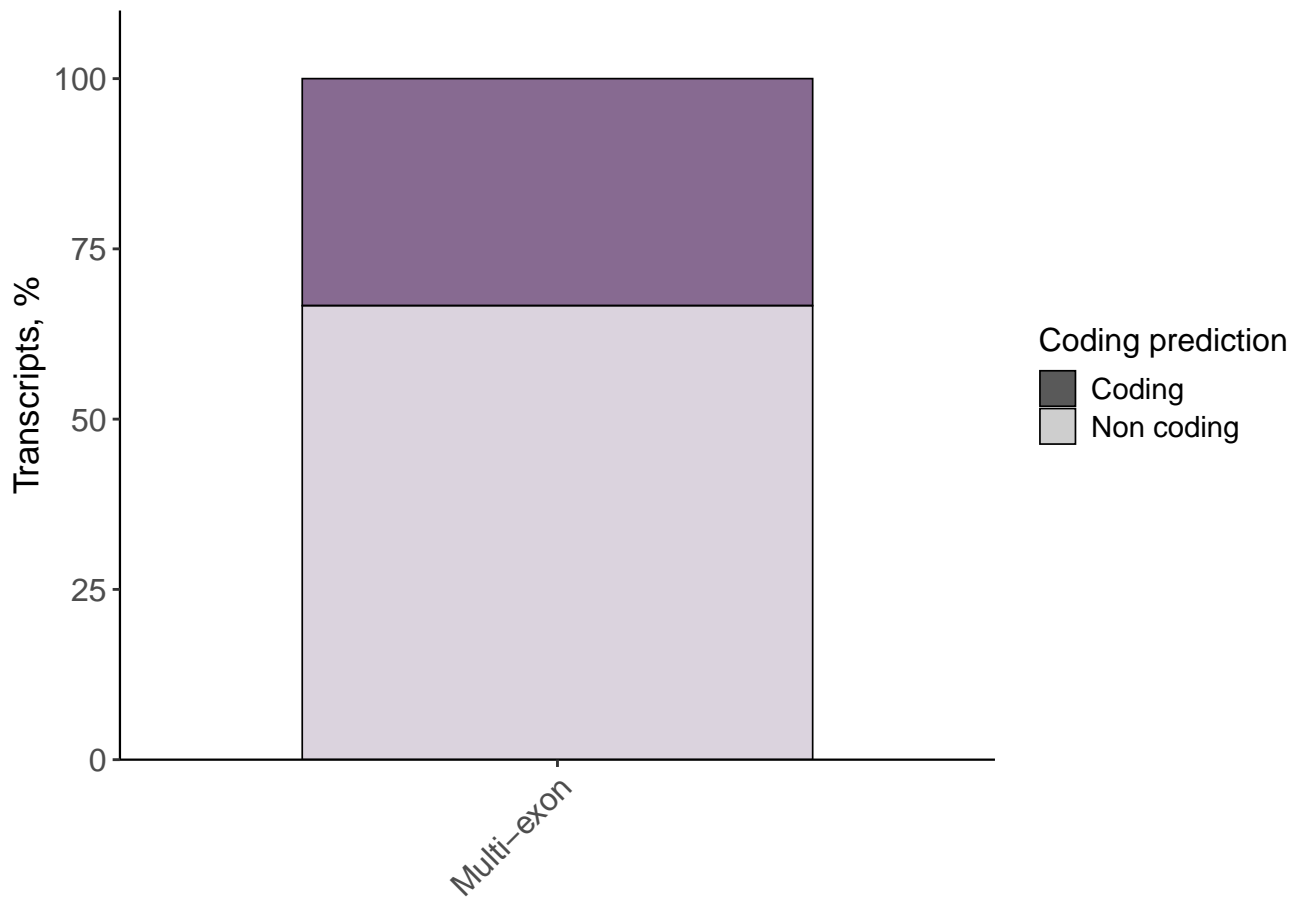
## Isoform Distribution Across NIC



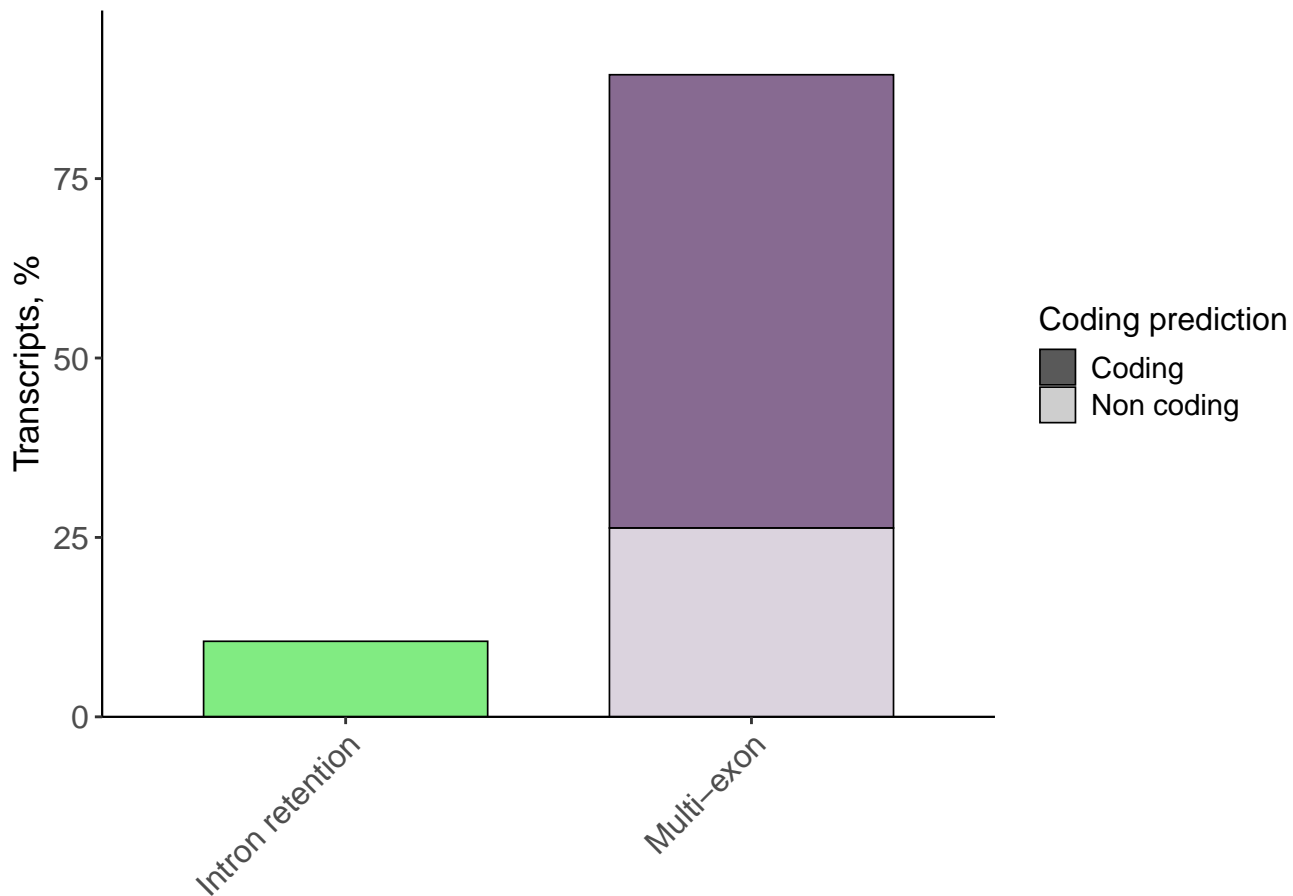
## Isoform Distribution Across Genic Genomic



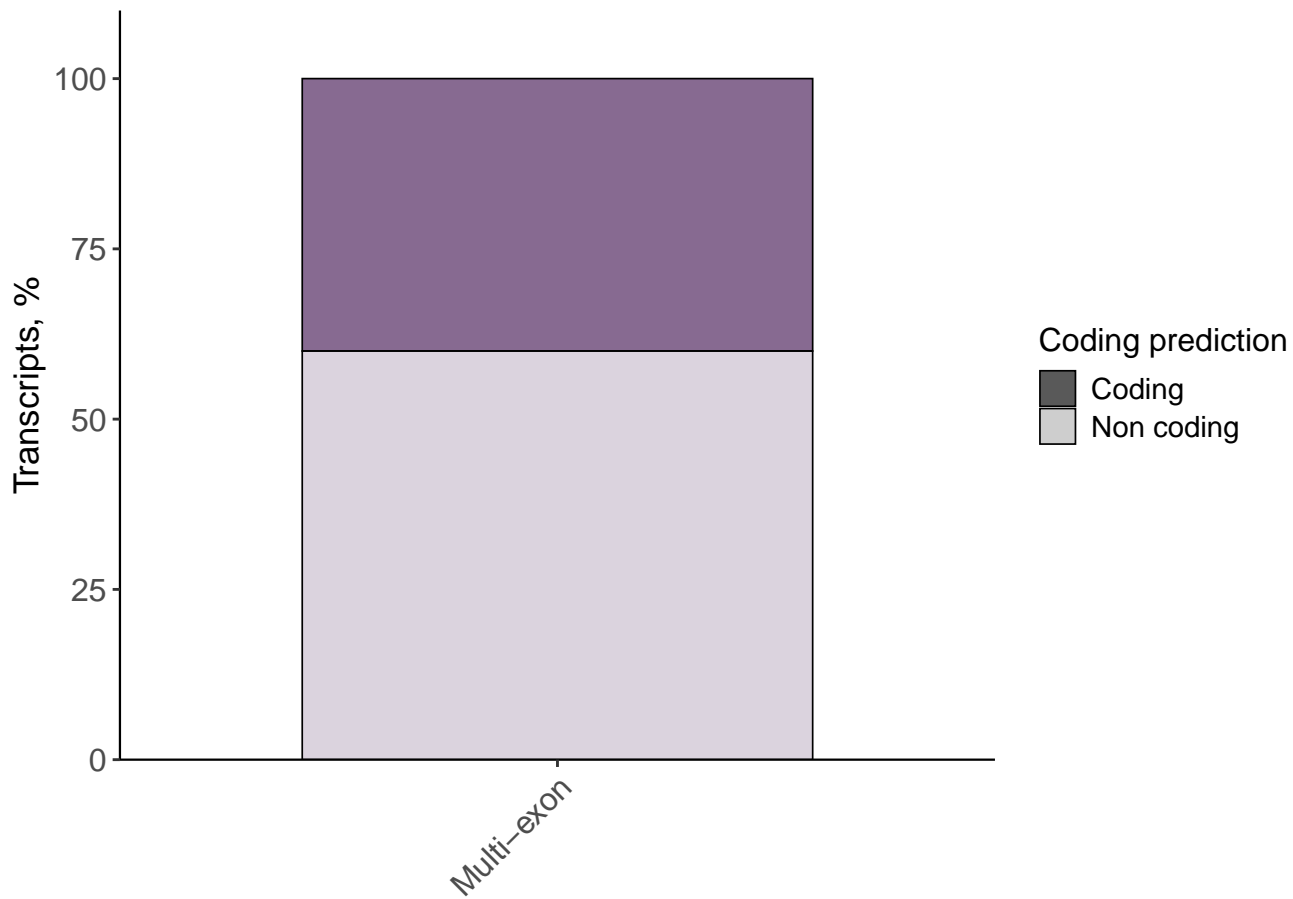
## Isoform Distribution Across Antisense



## Isoform Distribution Across Fusion

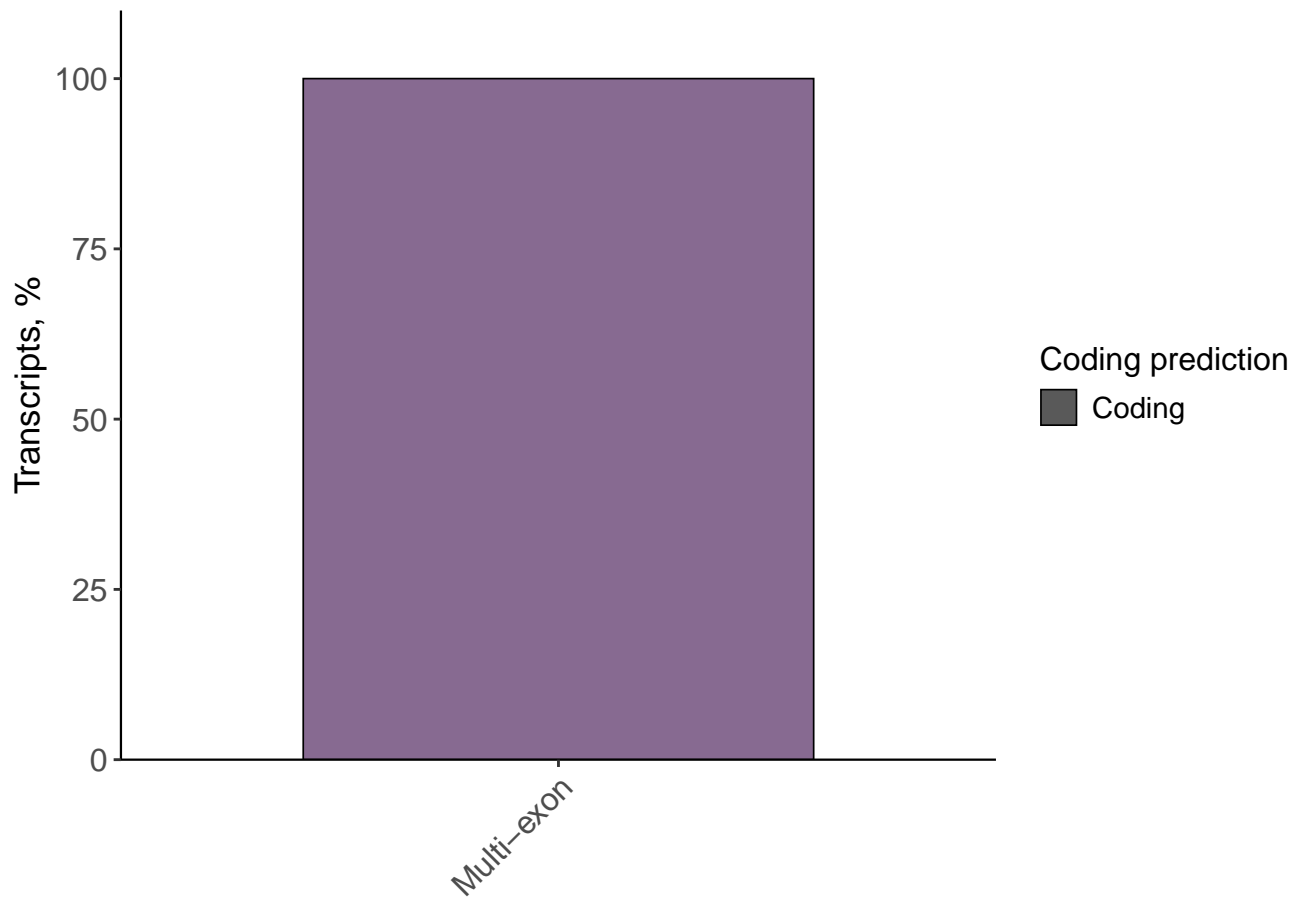


## Isoform Distribution Across Intergenic

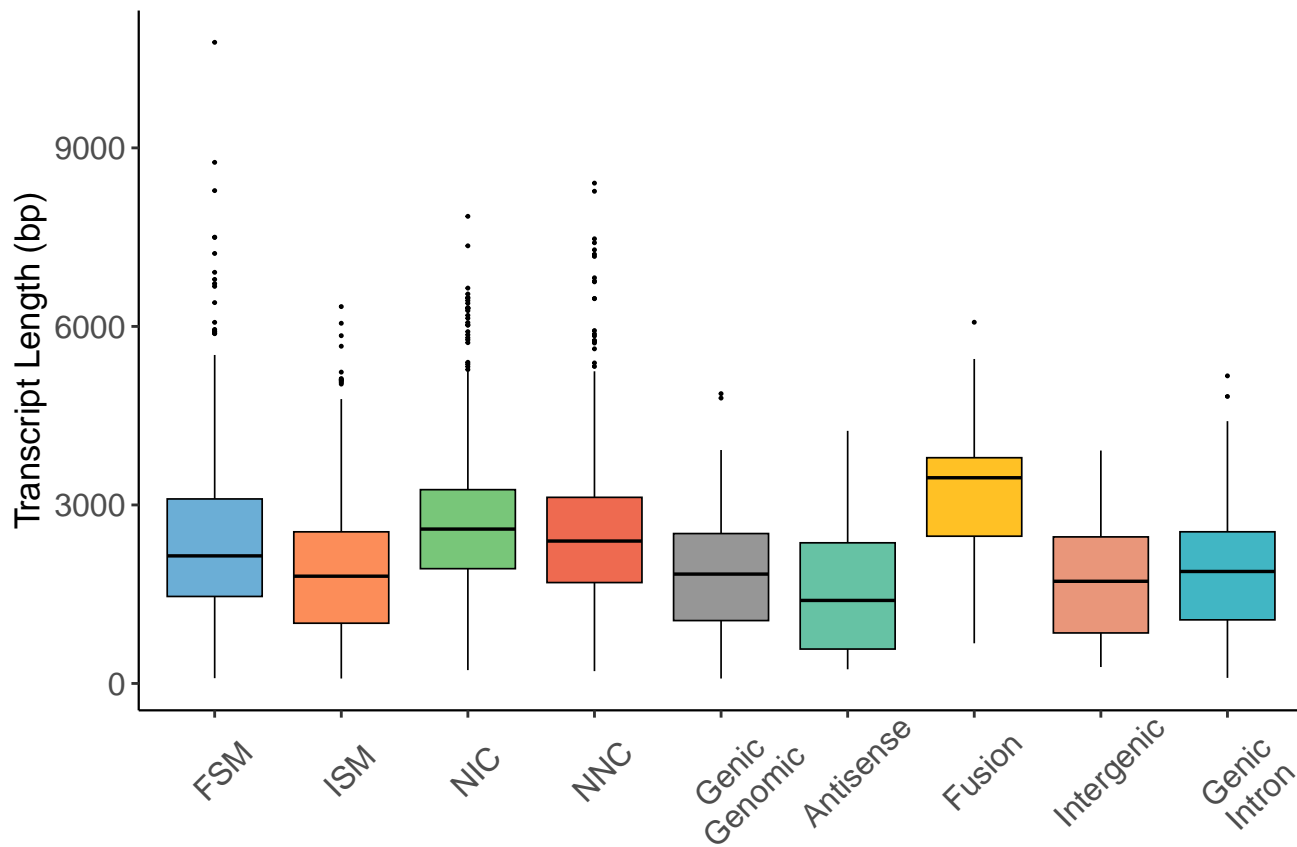




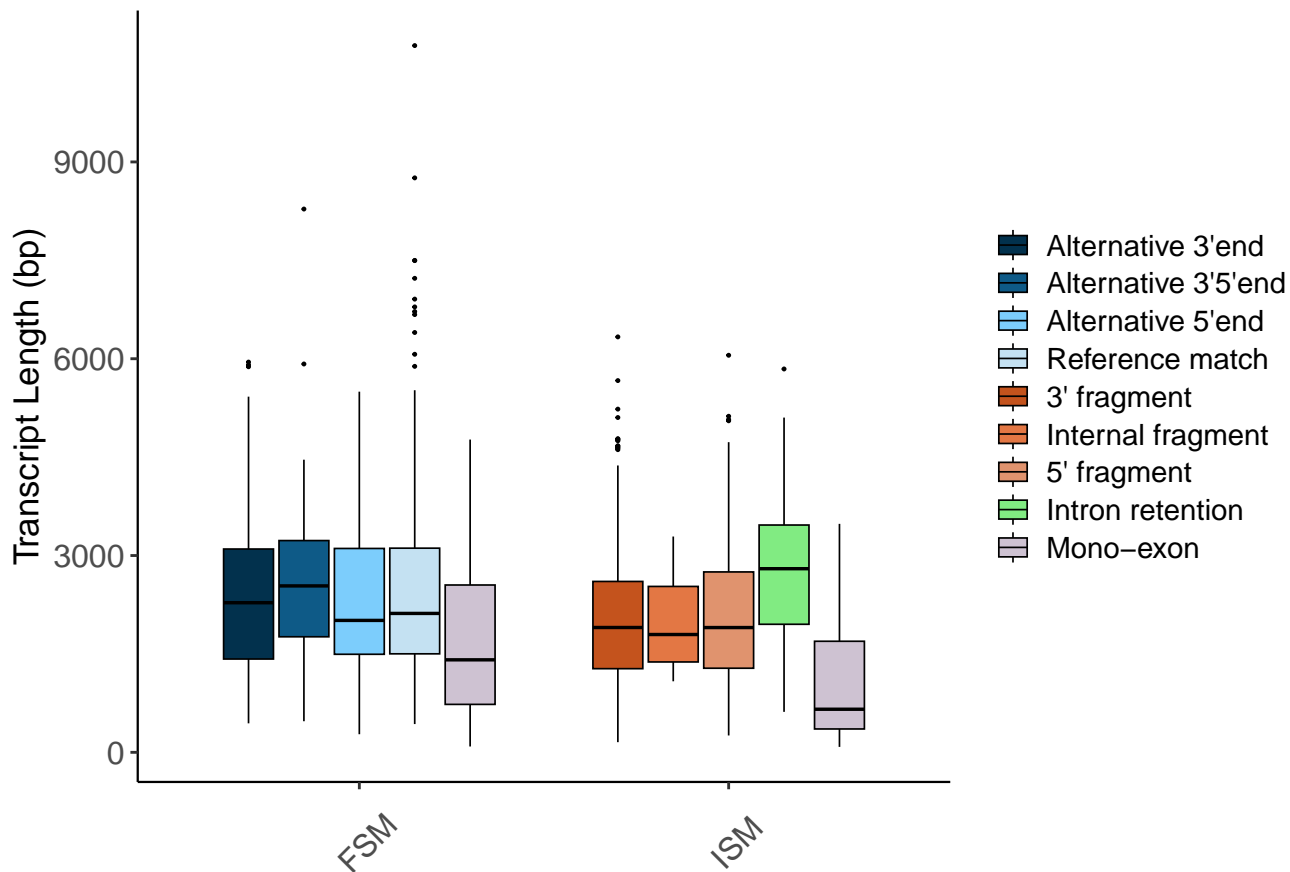
## Isoform Distribution Across Genic Intron



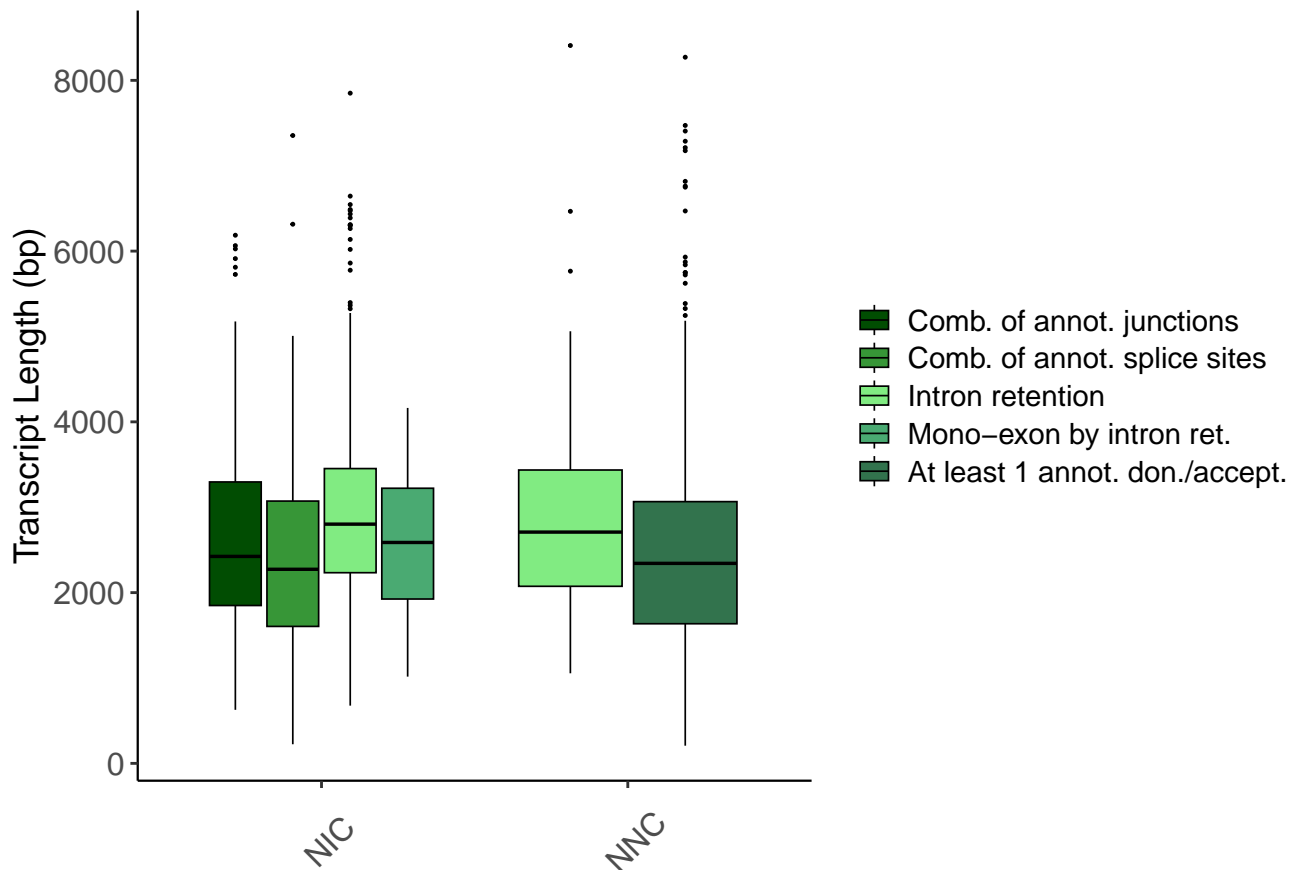
# Transcript Lengths by Structural Classification



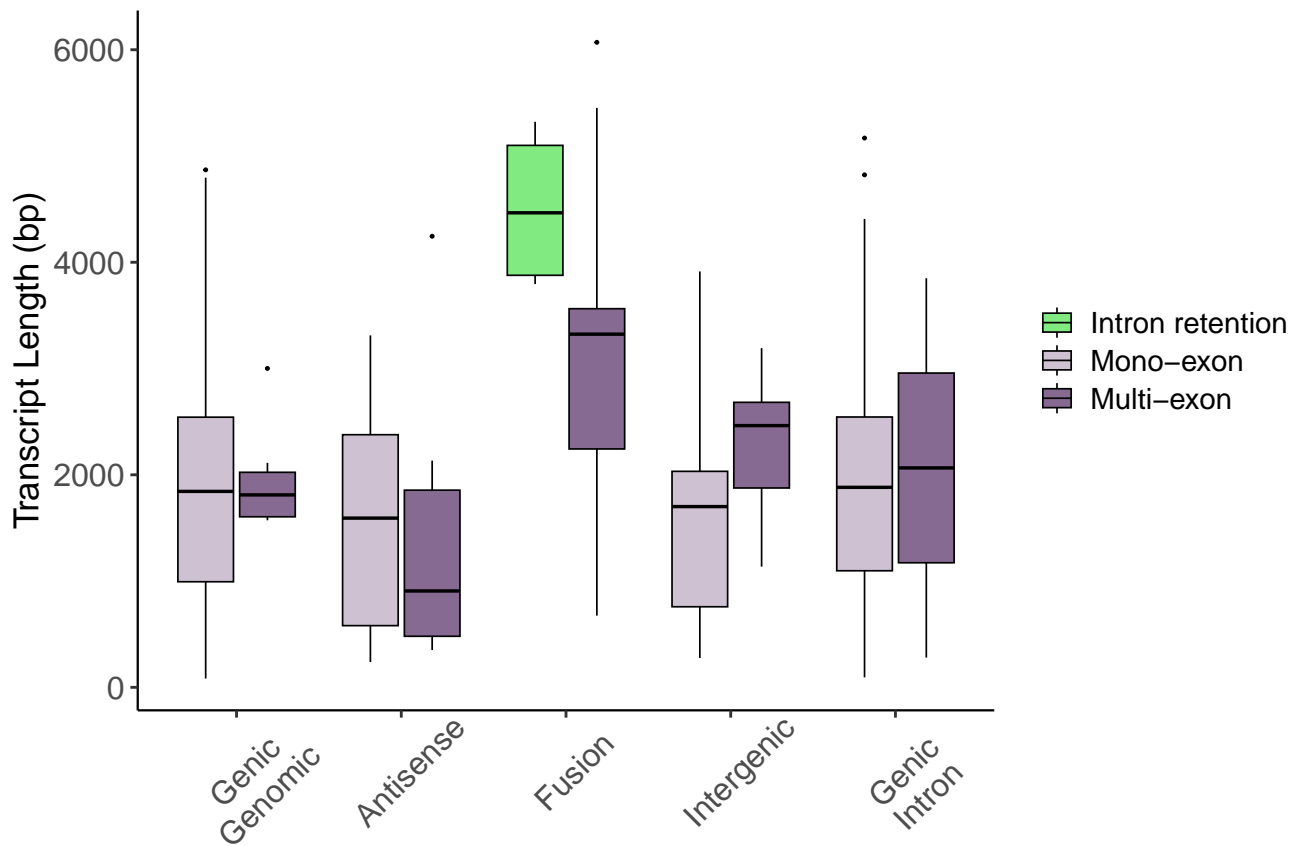
Transcript Lengths by Subcategory



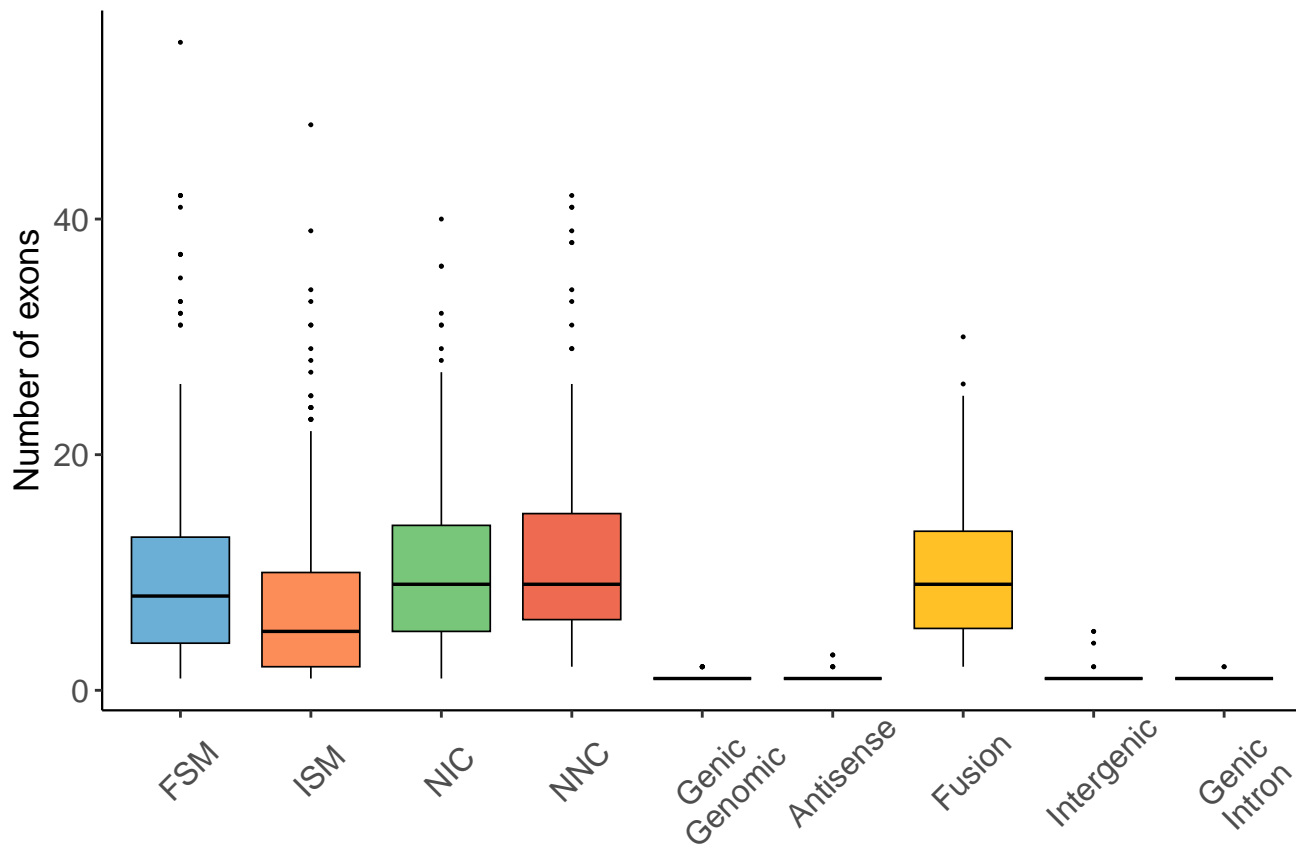
Transcript Lengths by Subcategory



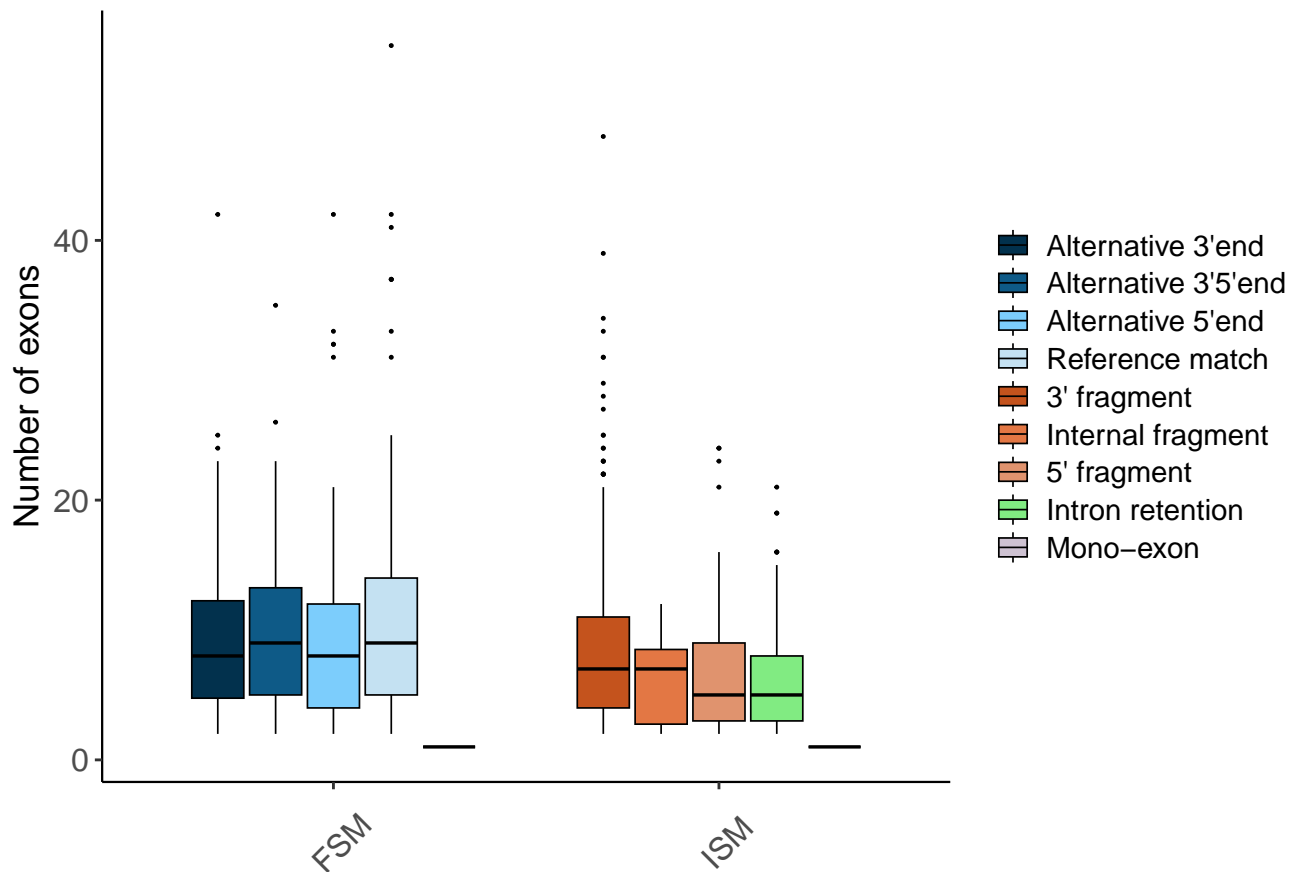
Transcript Lengths by Subcategory



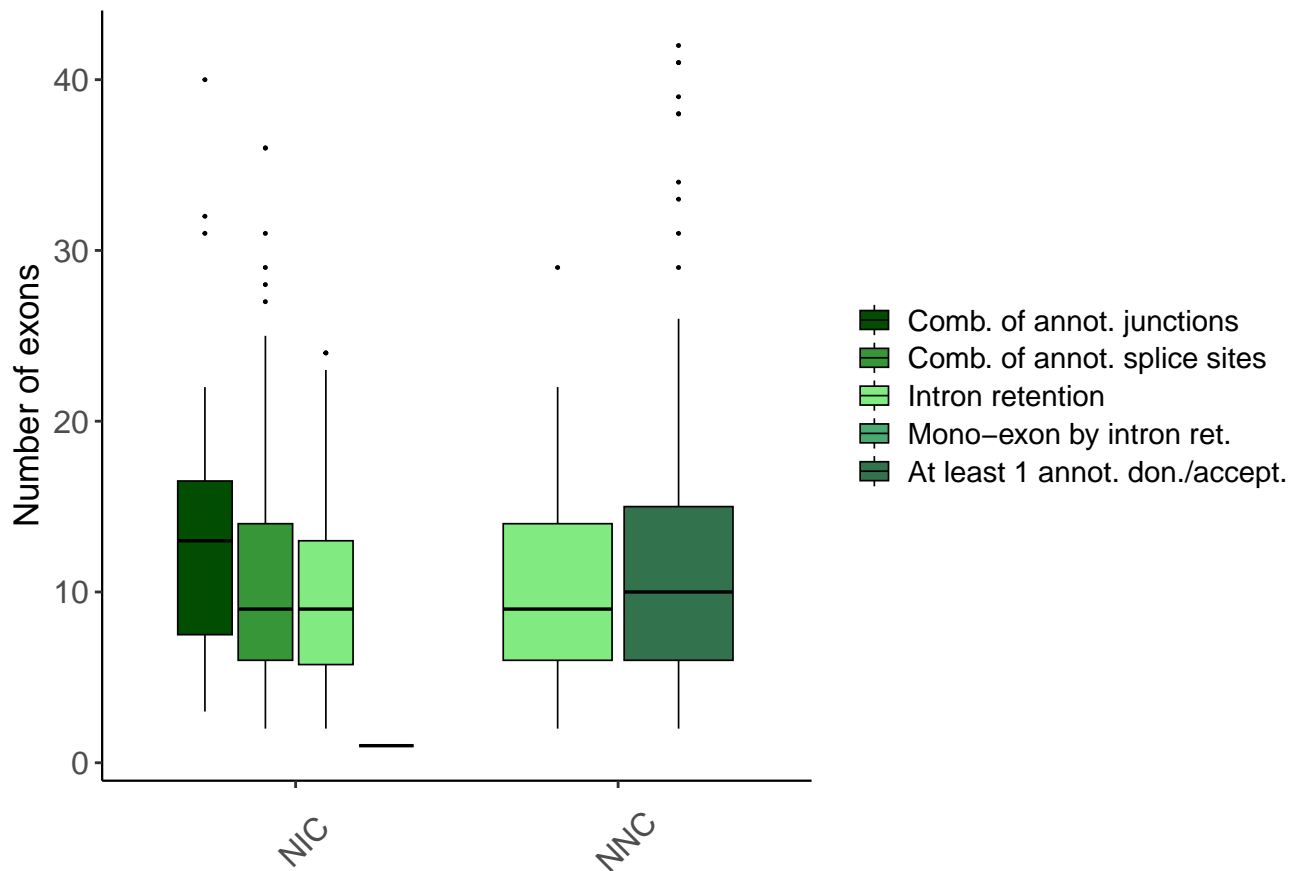
Exon Counts by Structural Classification



Exon Counts by Subcategory

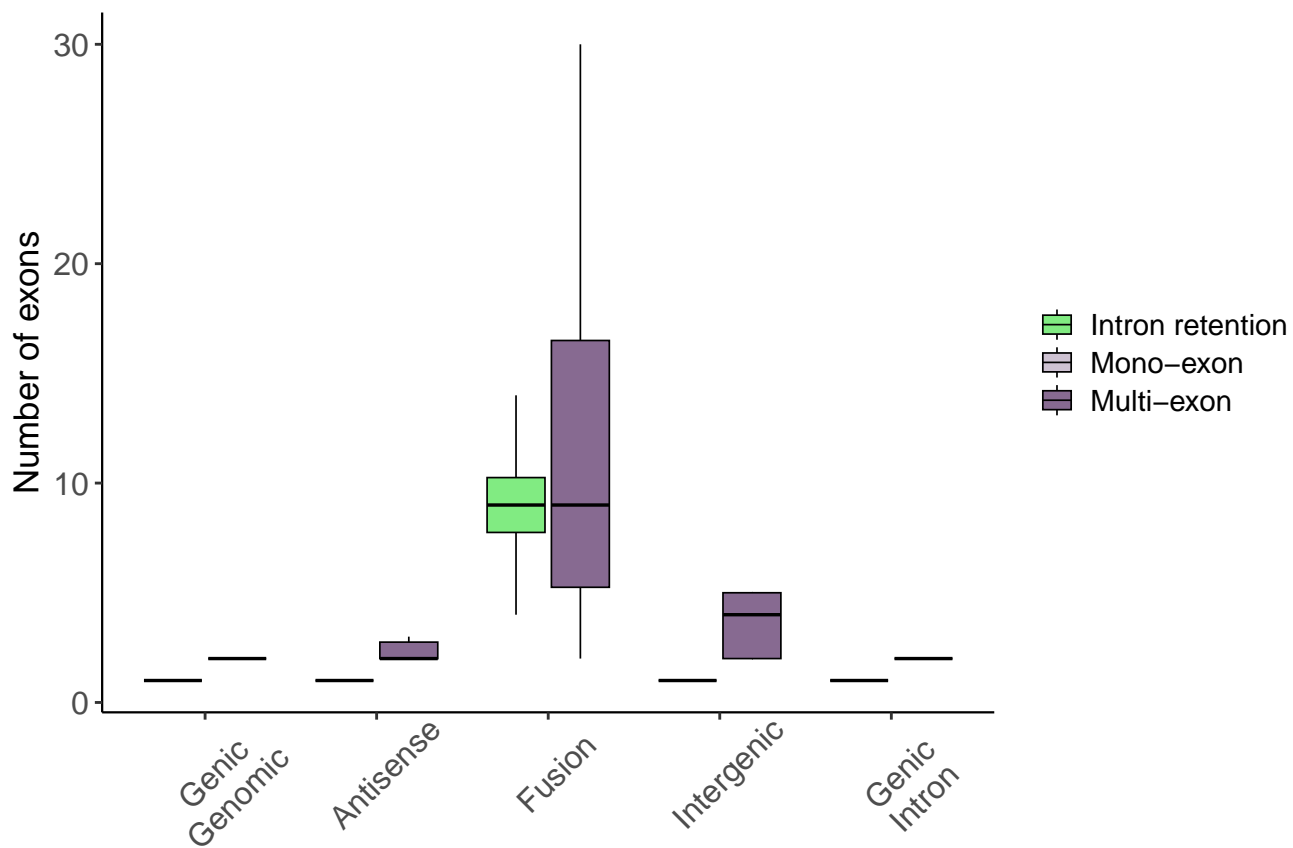


## Exon Counts by Subcategory

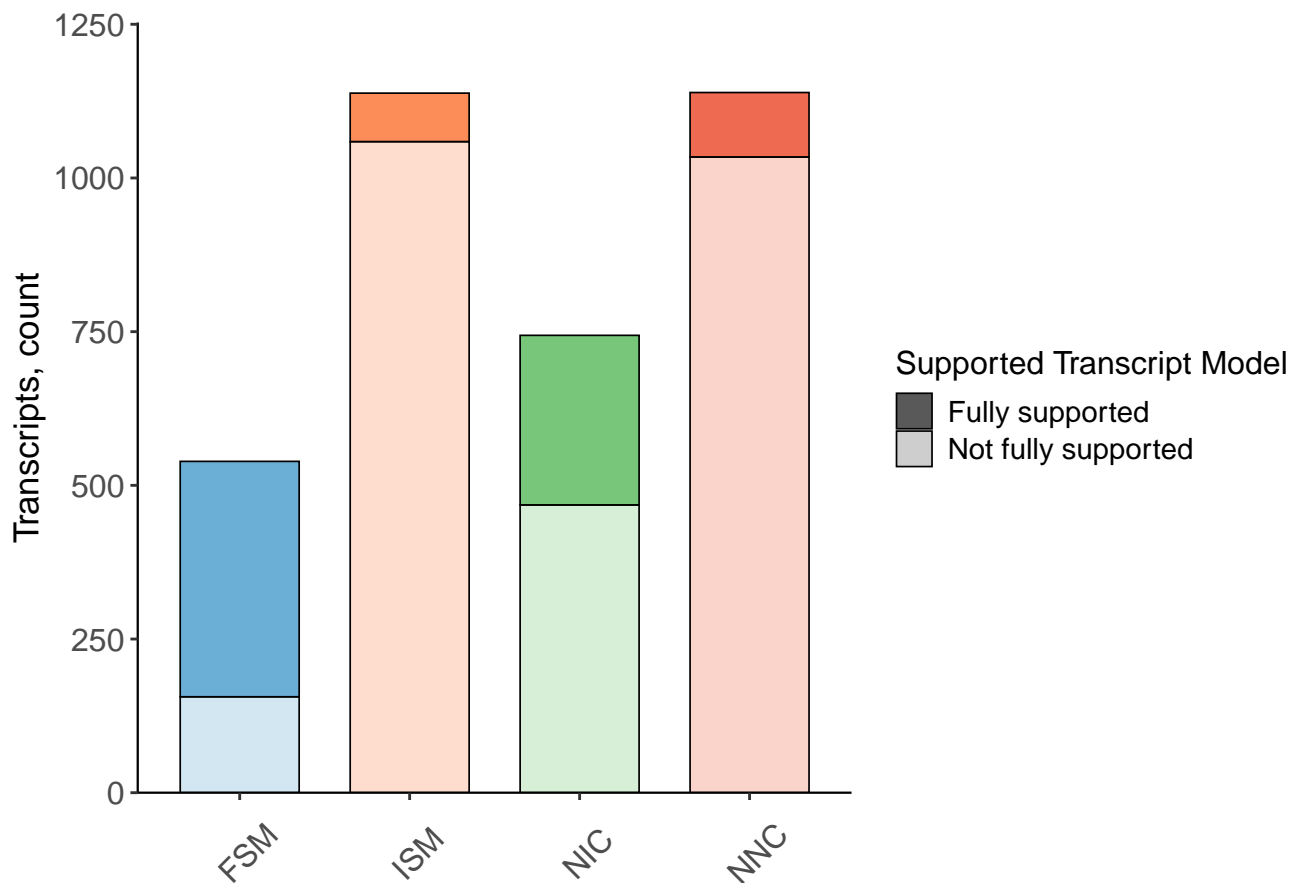




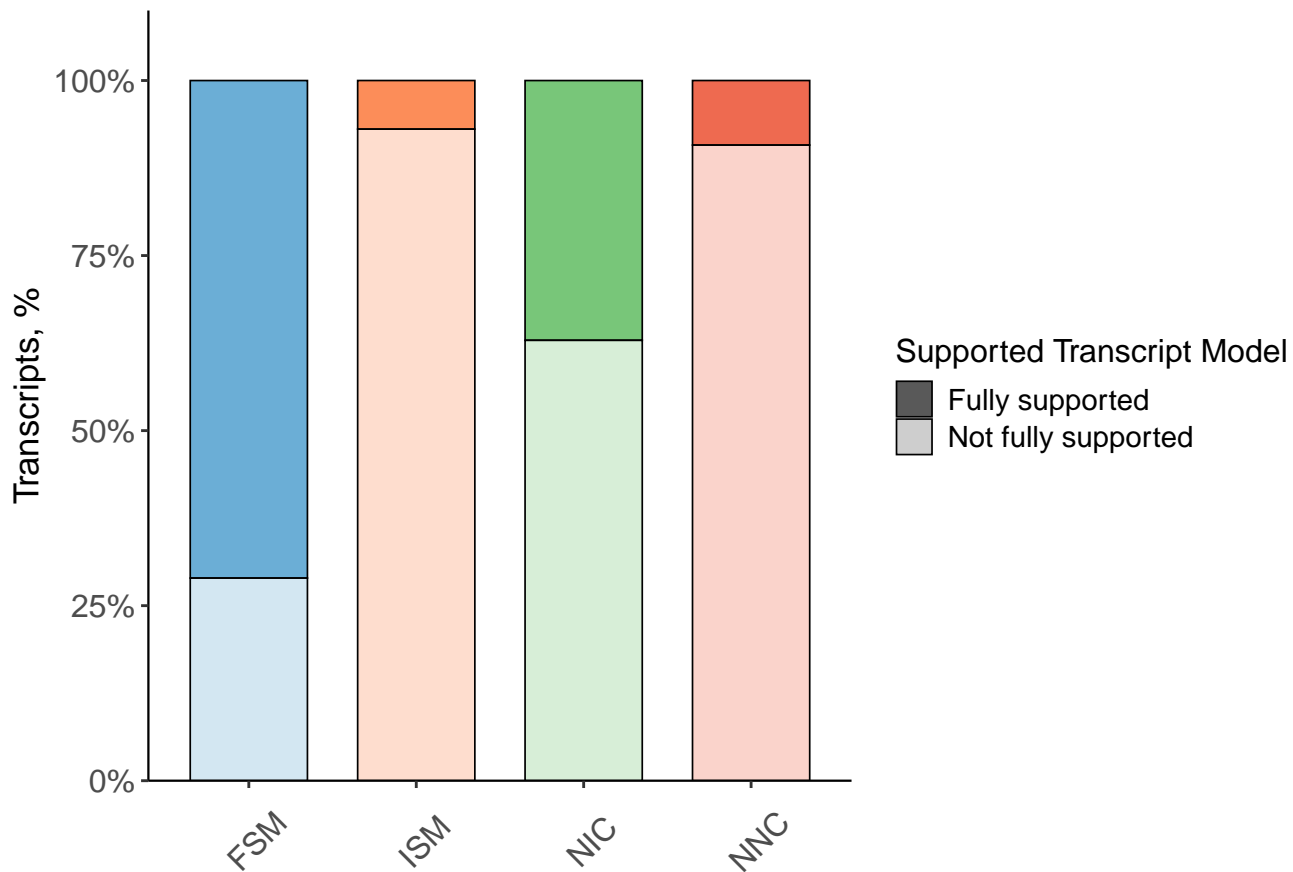
## Exon Counts by Subcategory



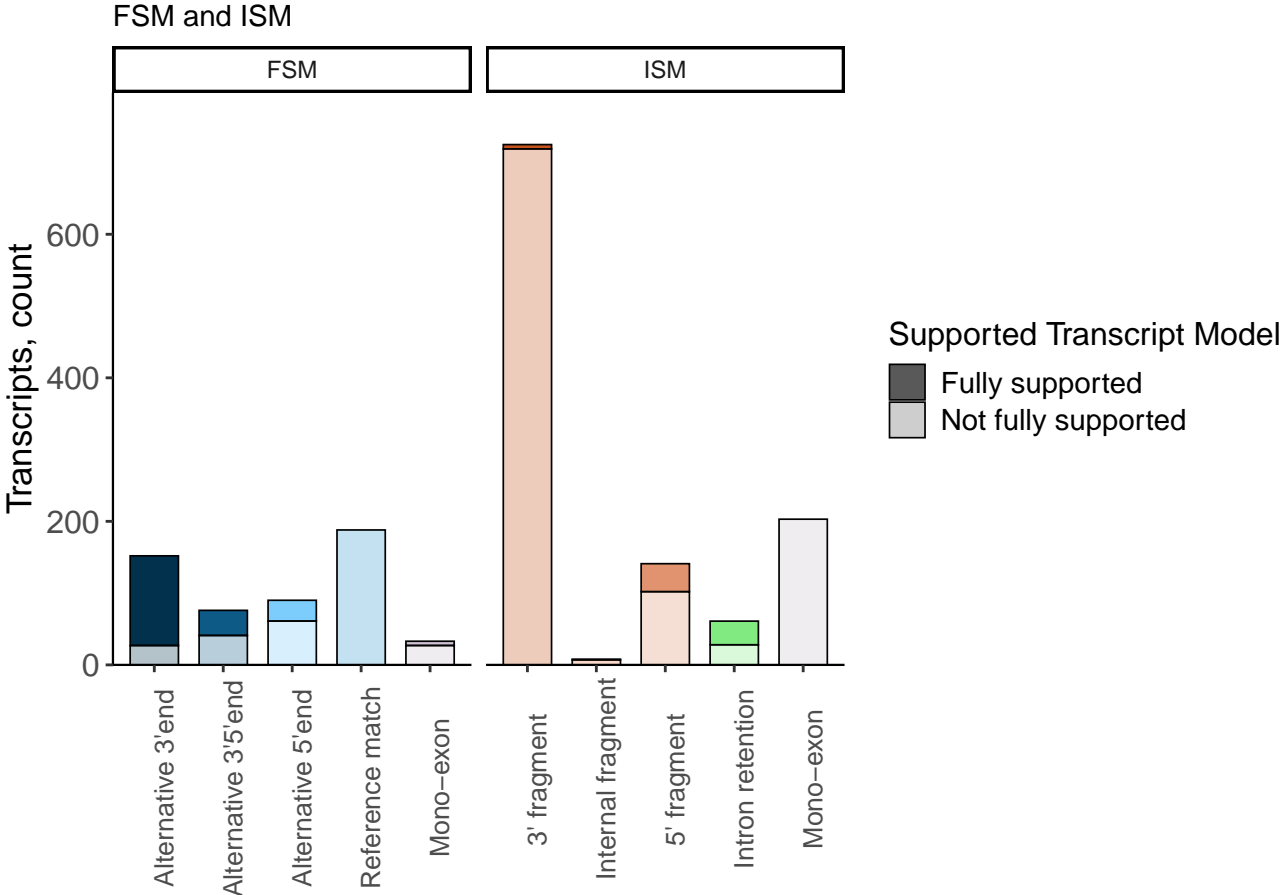
## Isoform Distribution Across Structural Categories



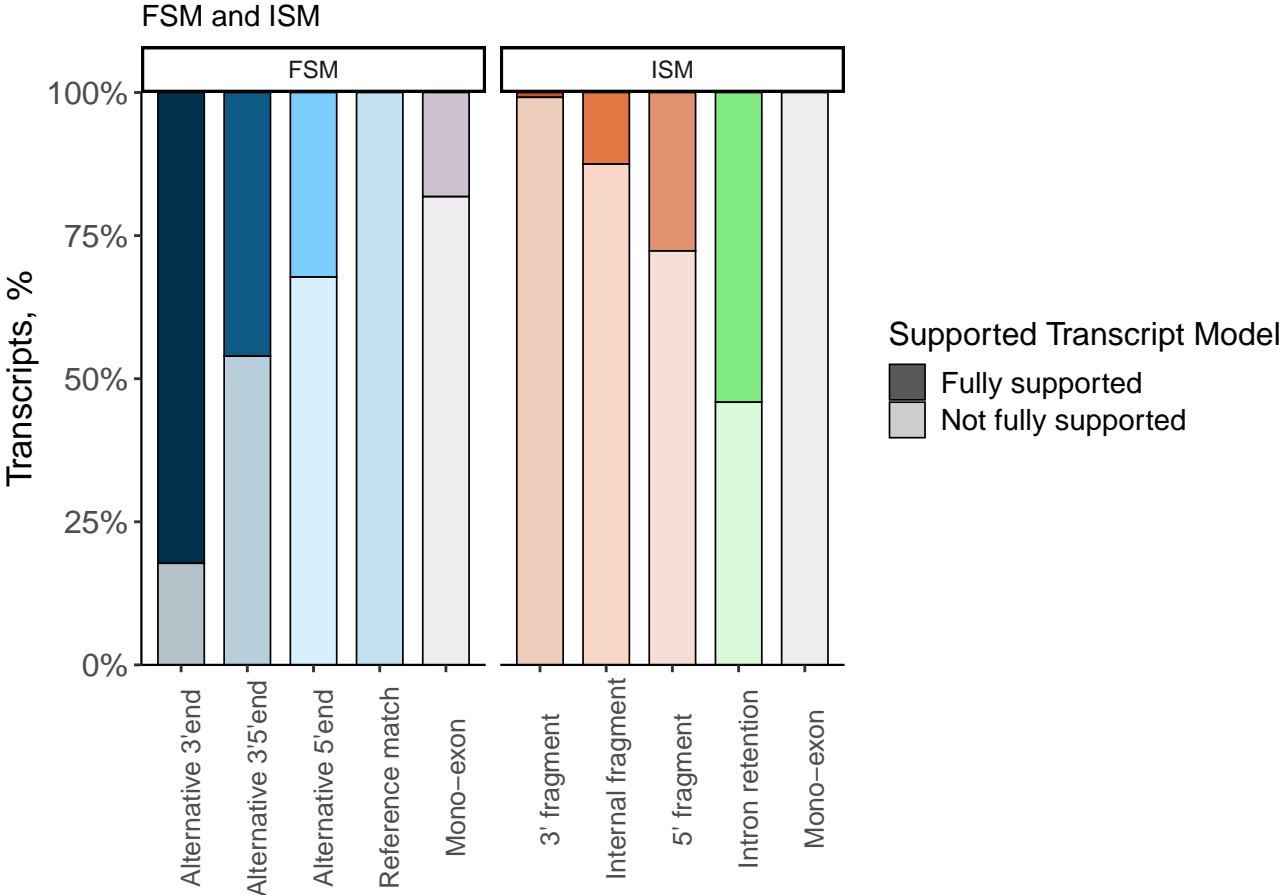
## Isoform Distribution Across Structural Categories



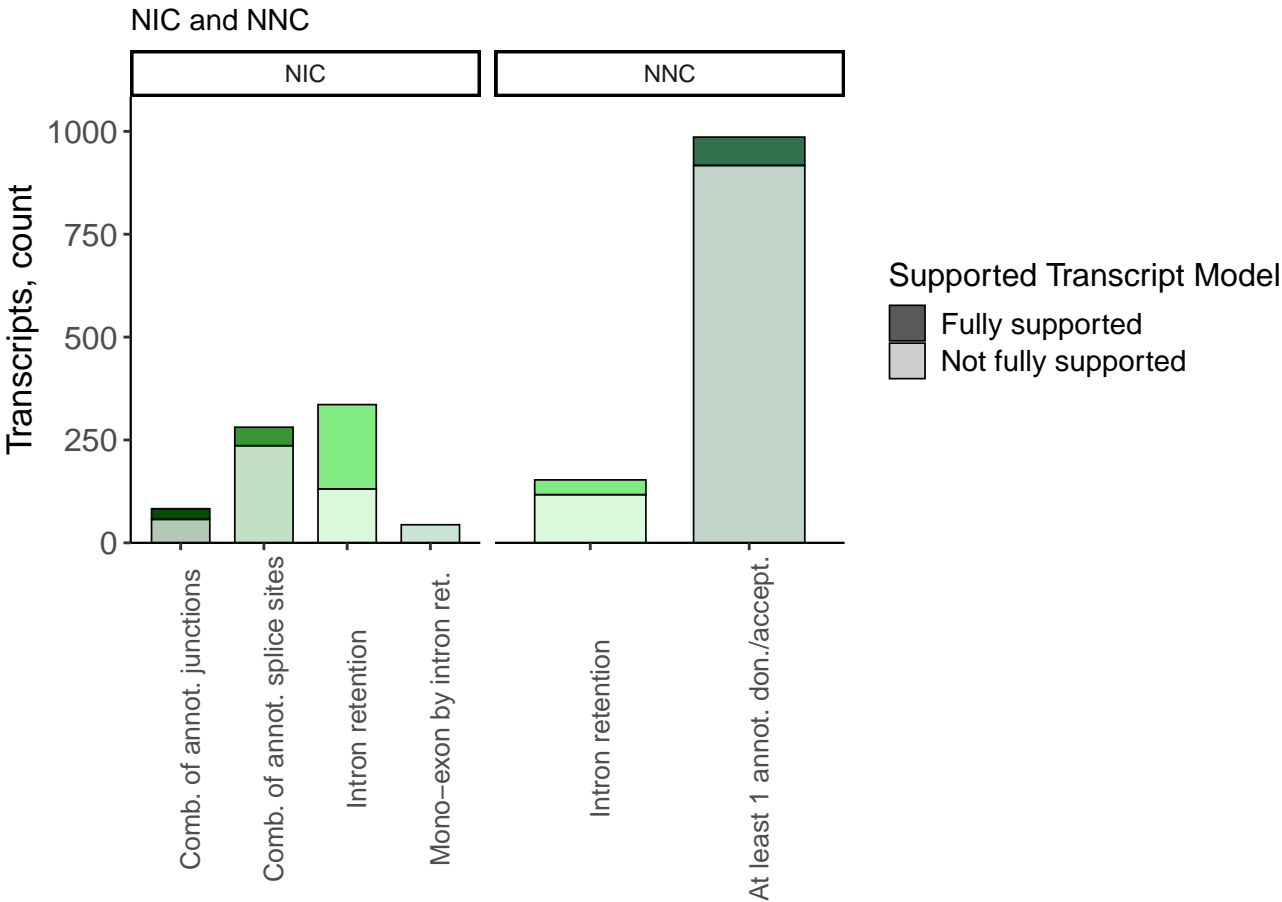
# Isoform Distribution Across Structural Subcategories



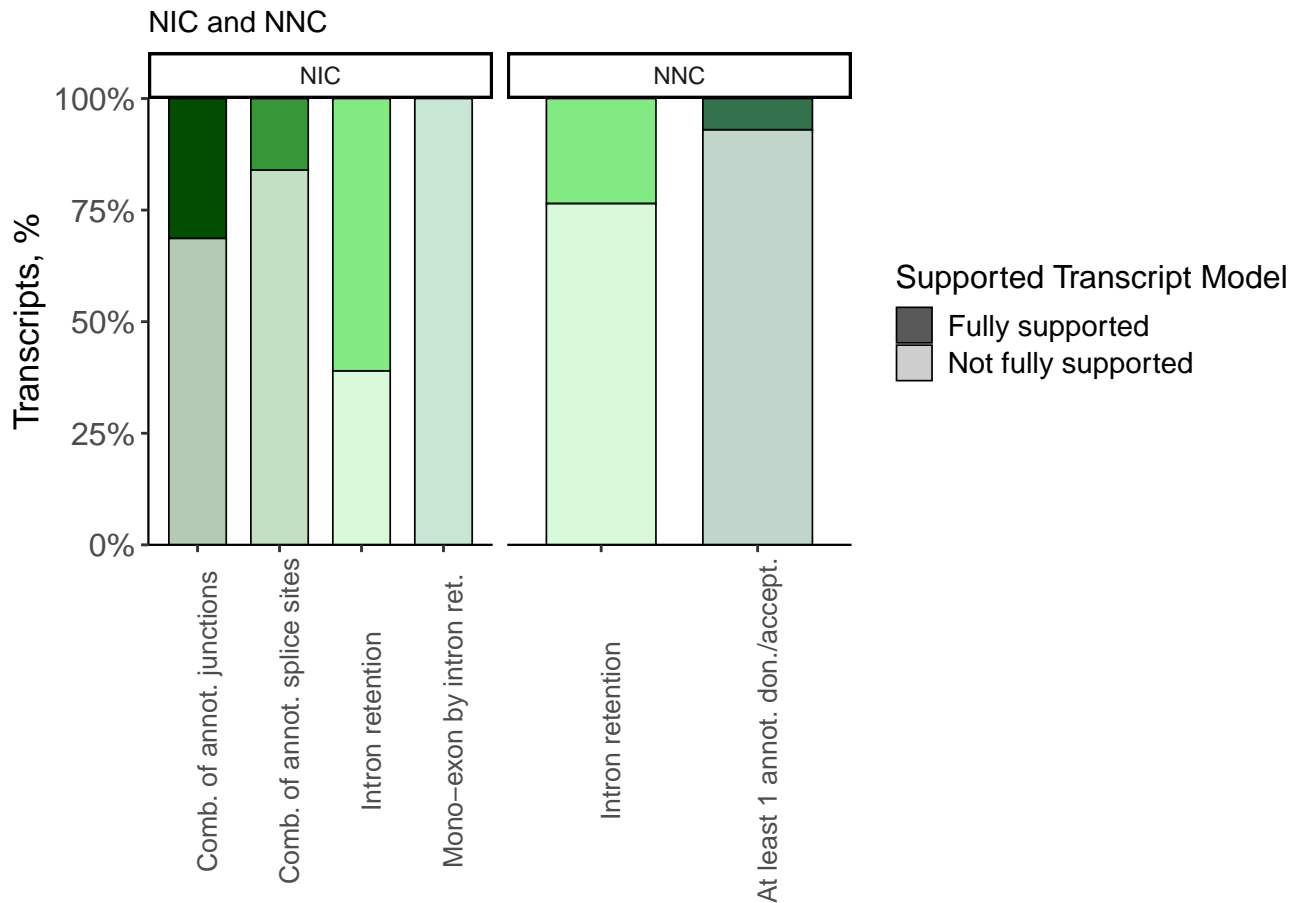
# Isoform Distribution Across Structural Subcategories



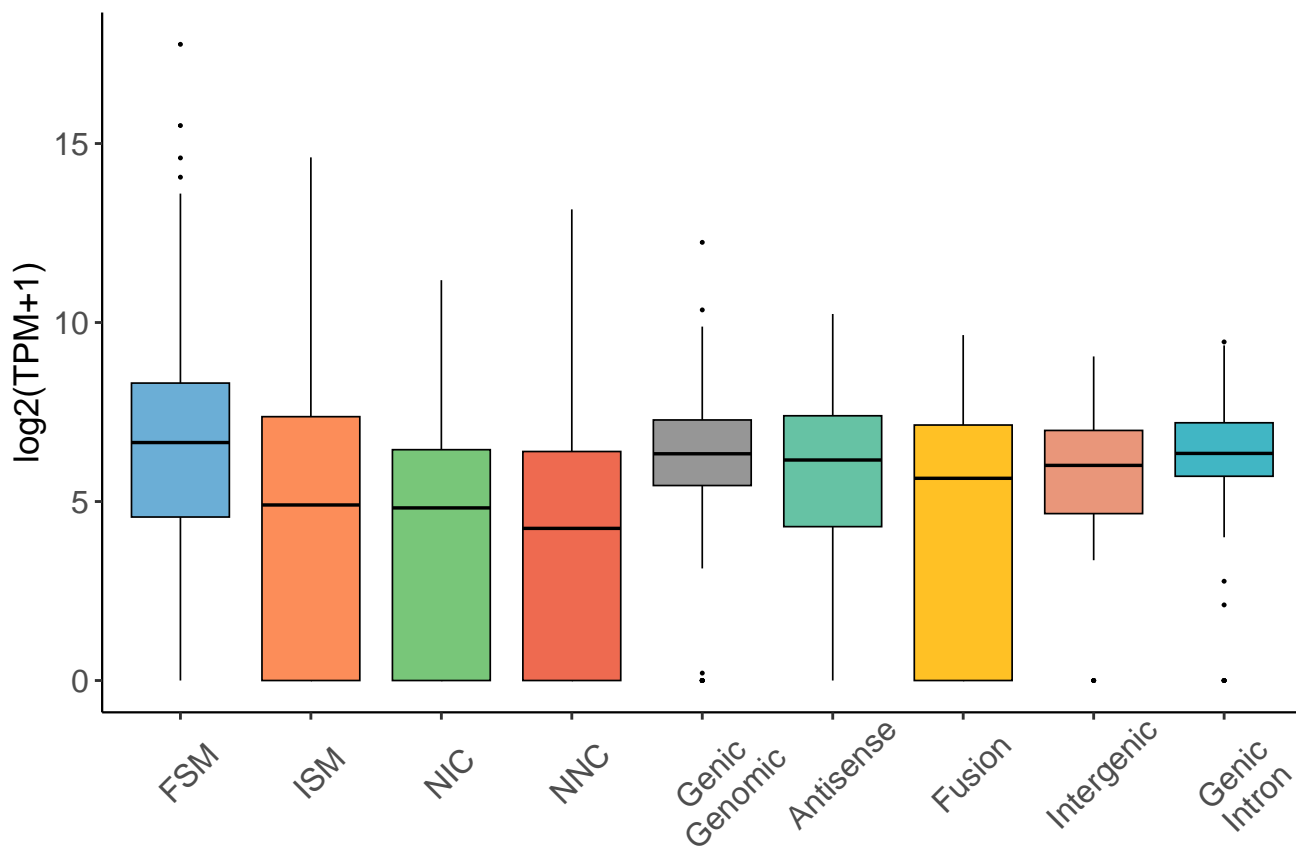
# Isoform Distribution Across Structural Subcategories



## Isoform Distribution Across Structural Subcategories

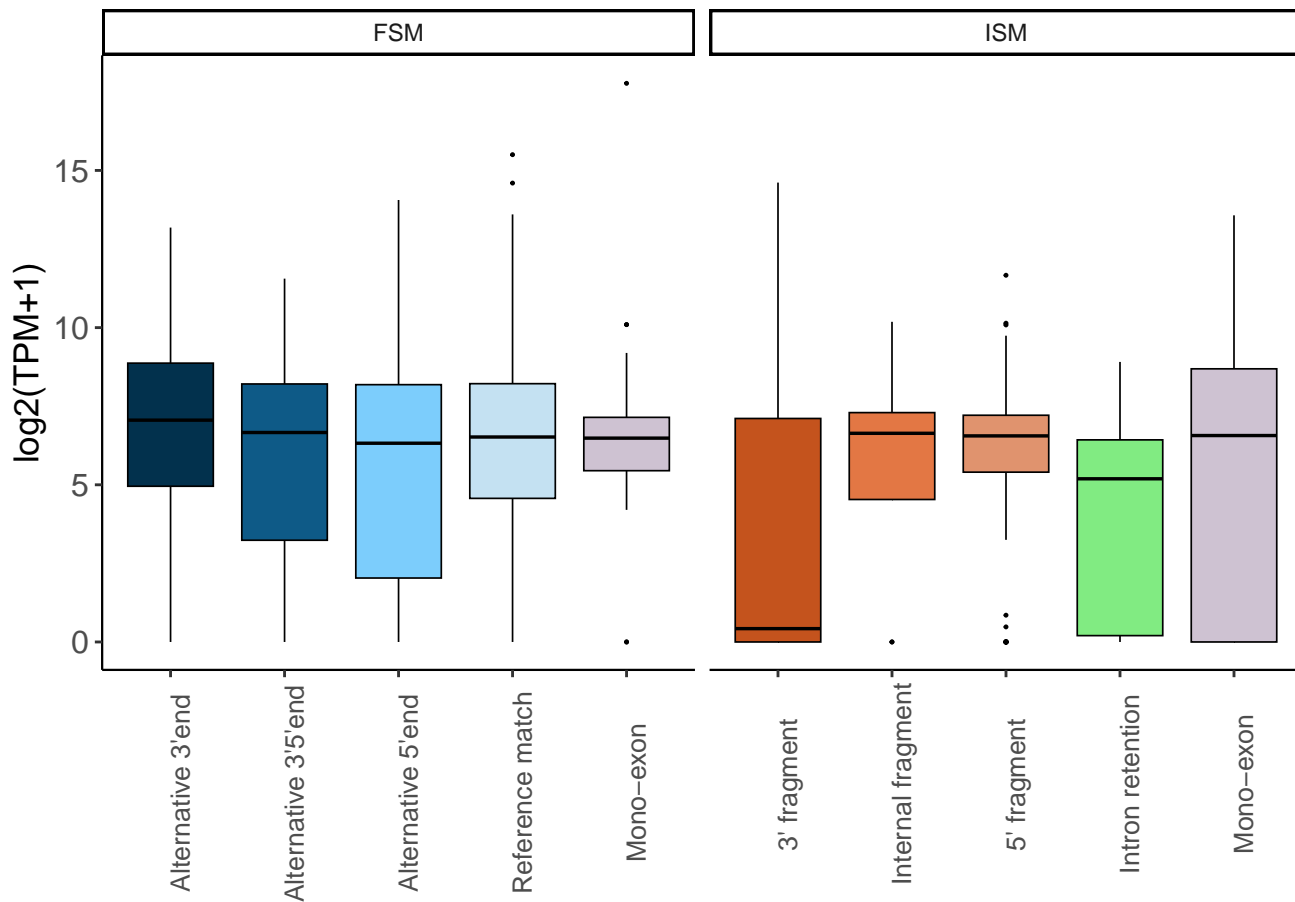


## Transcript Expression by Structural Category

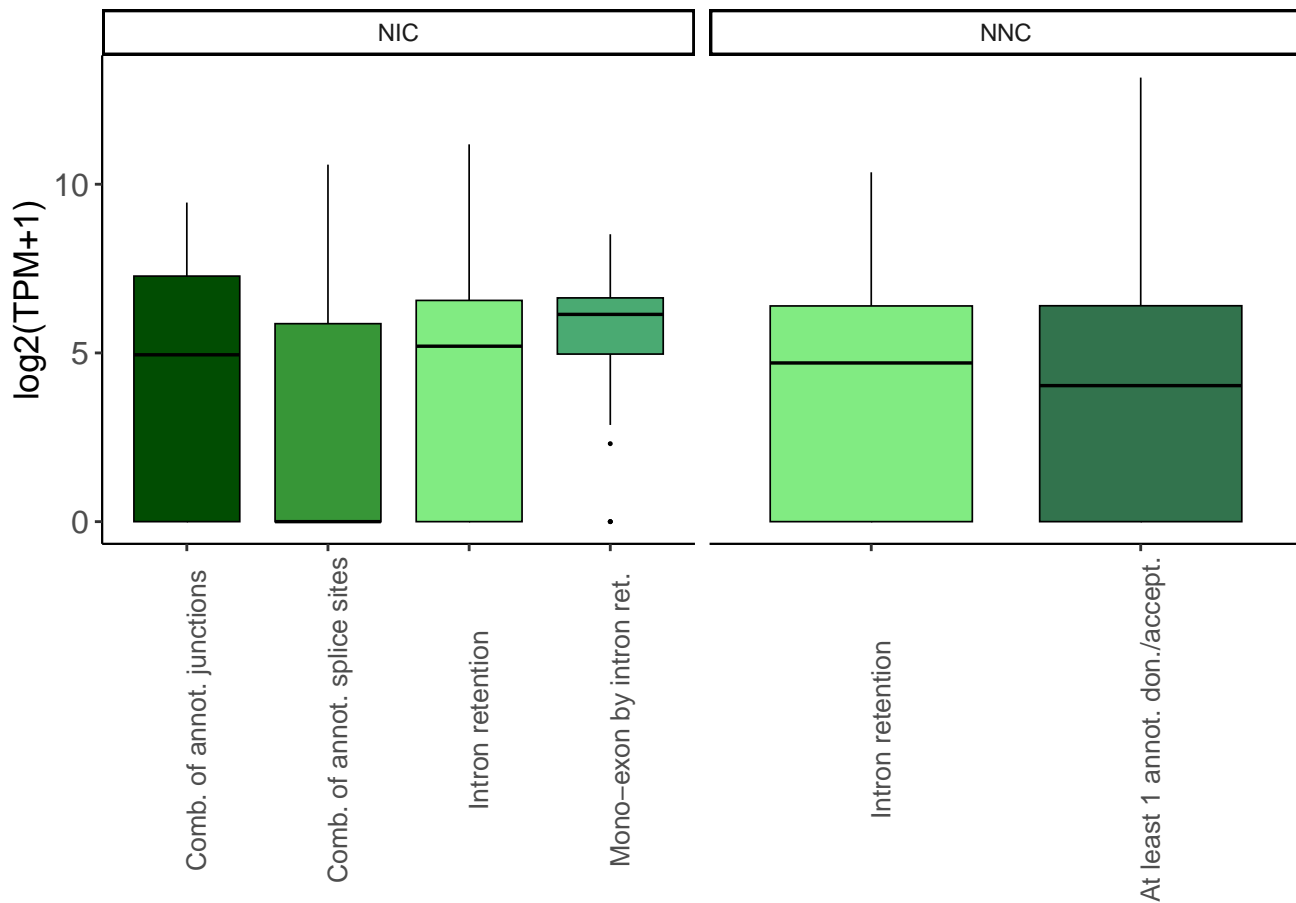




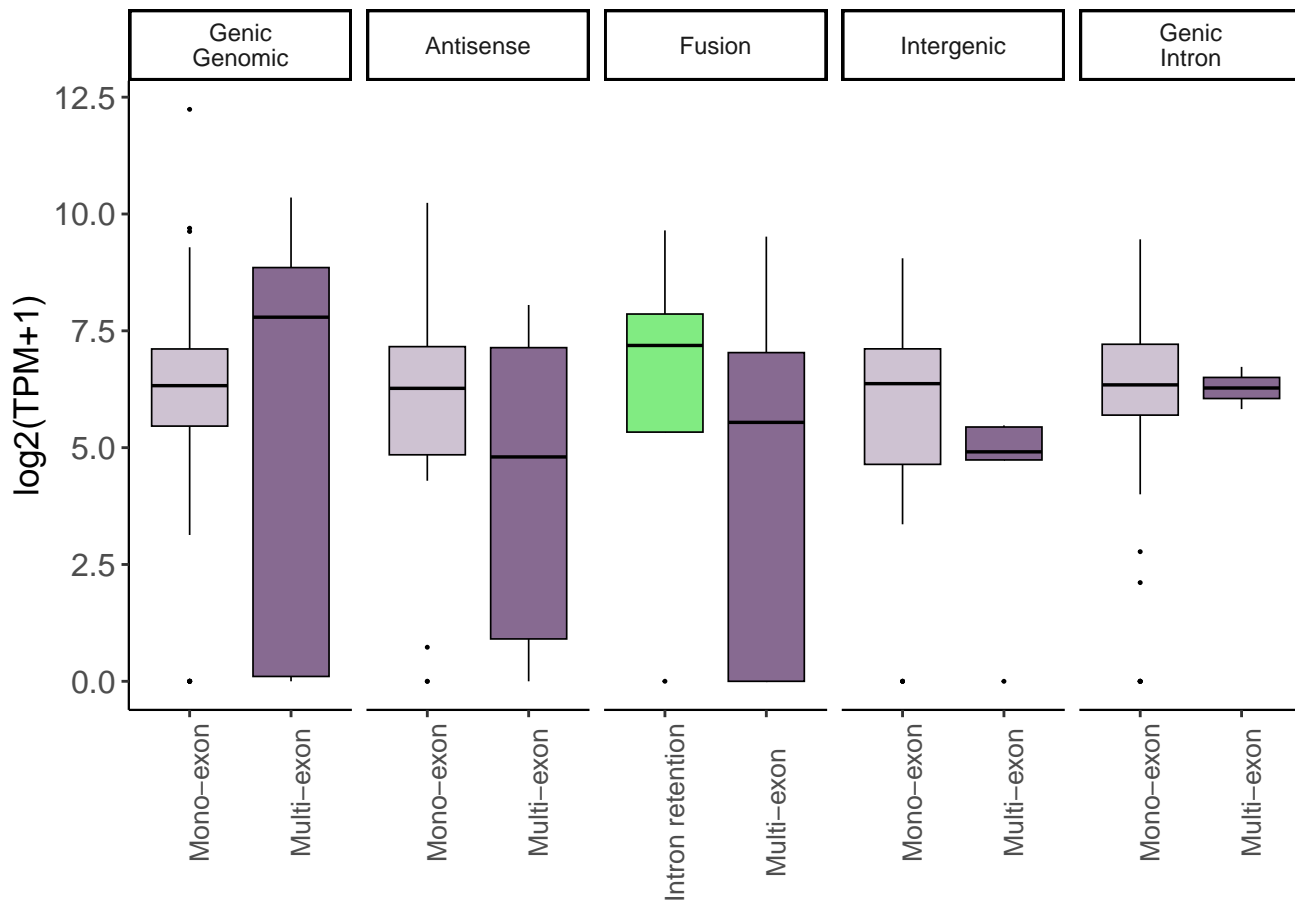
## Transcript Expression by Subcategory



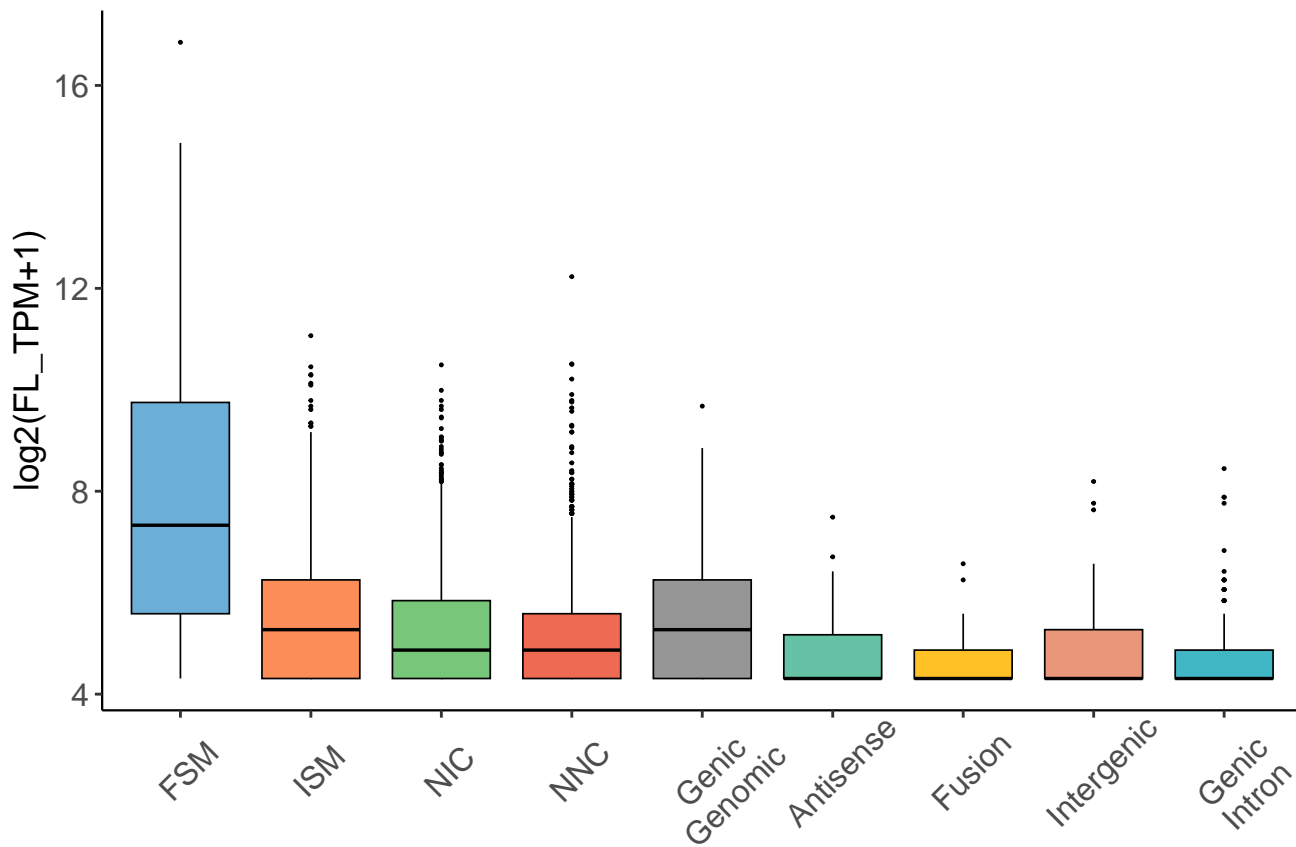
## Transcript Expression by Subcategory



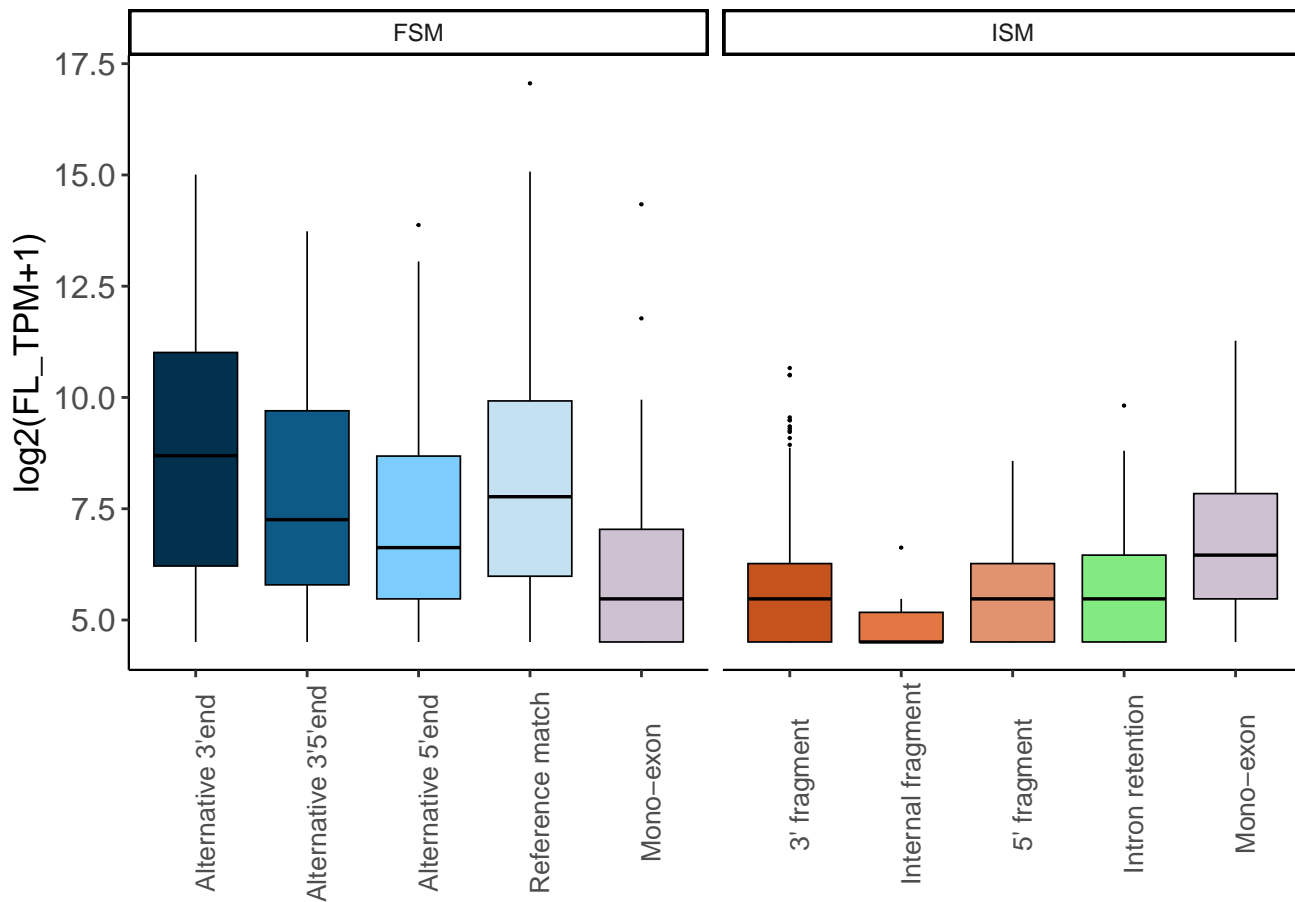
## Transcript Expression by Subcategory



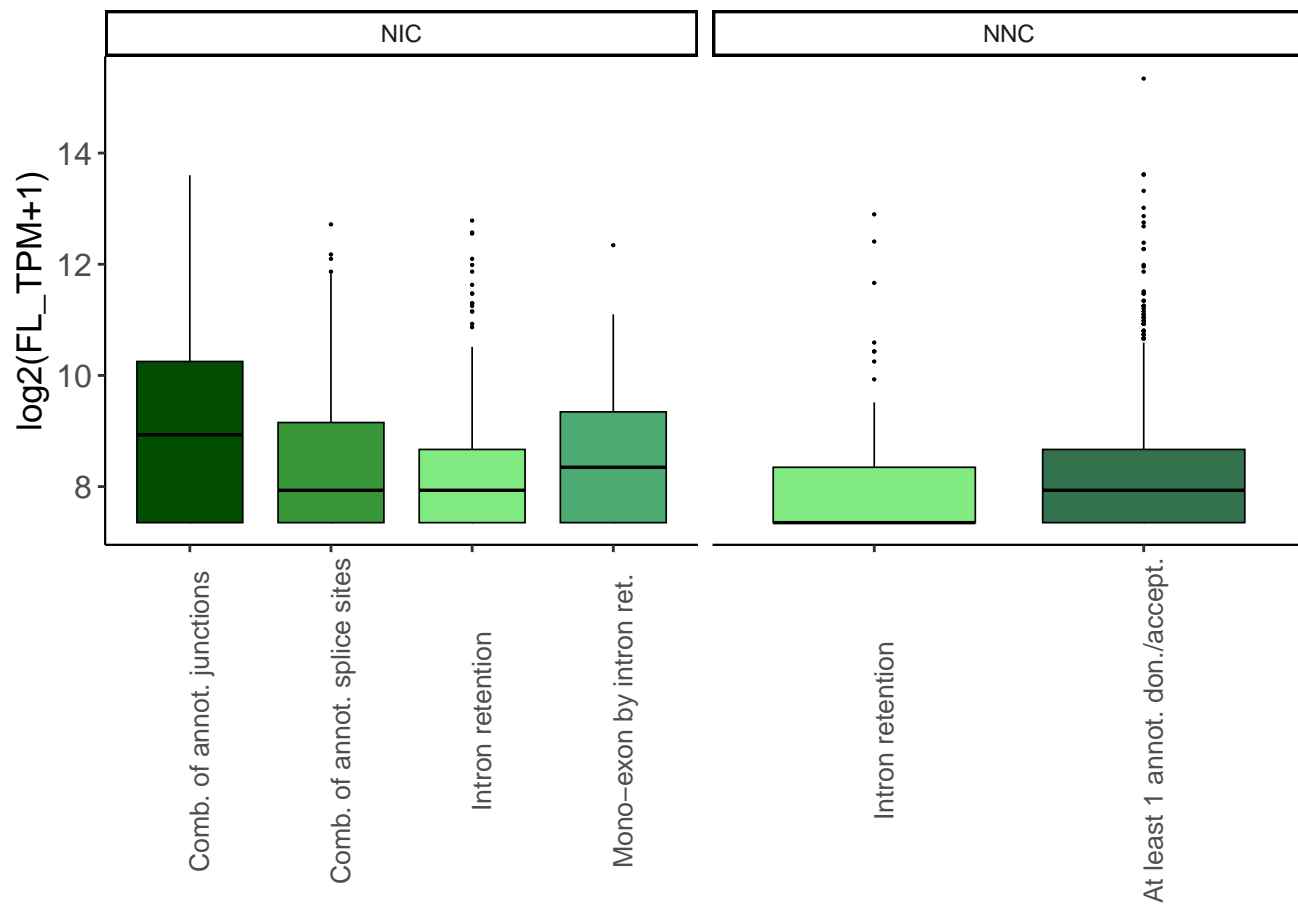
Long Reads Count by Structural Category



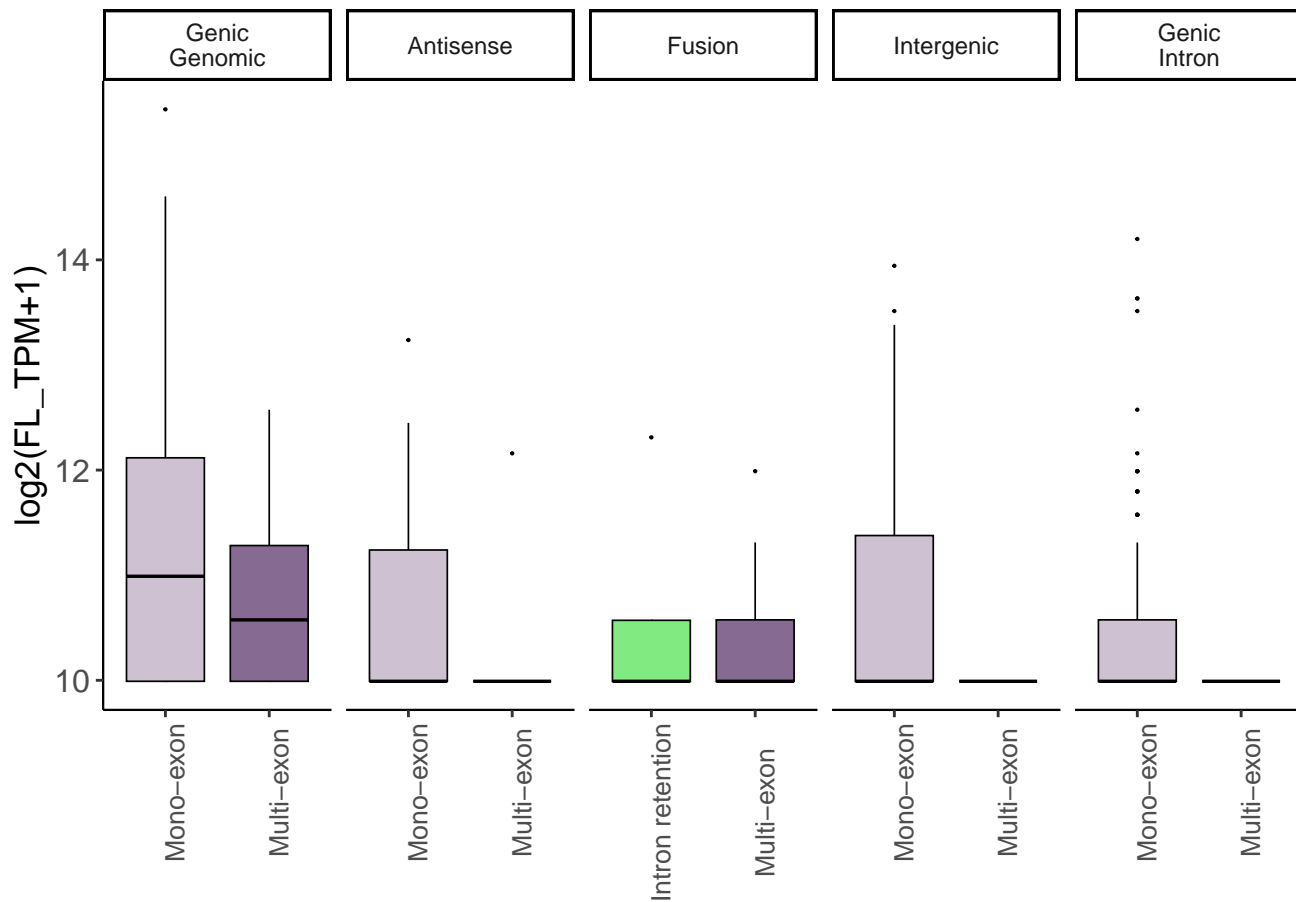
# Long Reads Count by Subcategory



## Long Reads Count by Subcategory

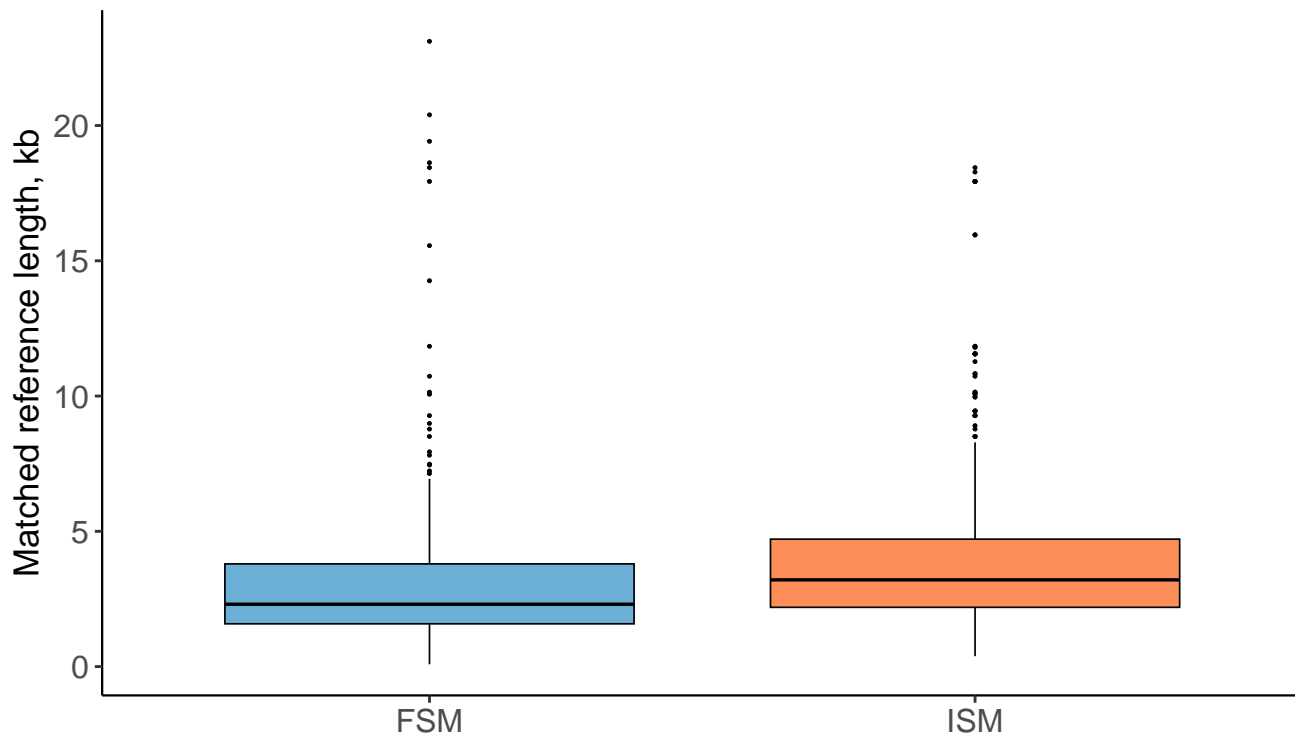


## Long Reads Count by Subcategory



# Length Distribution of Matched Reference Transcripts

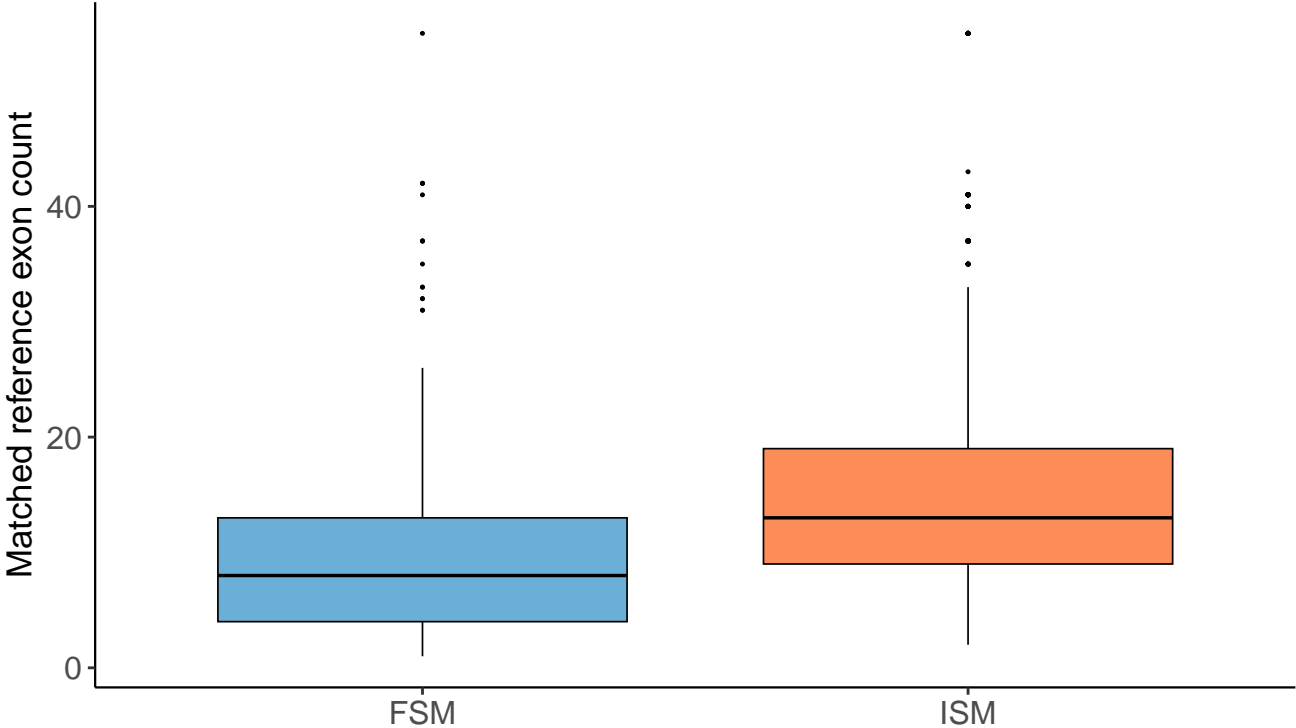
Applicable Only to FSM and ISM Categories



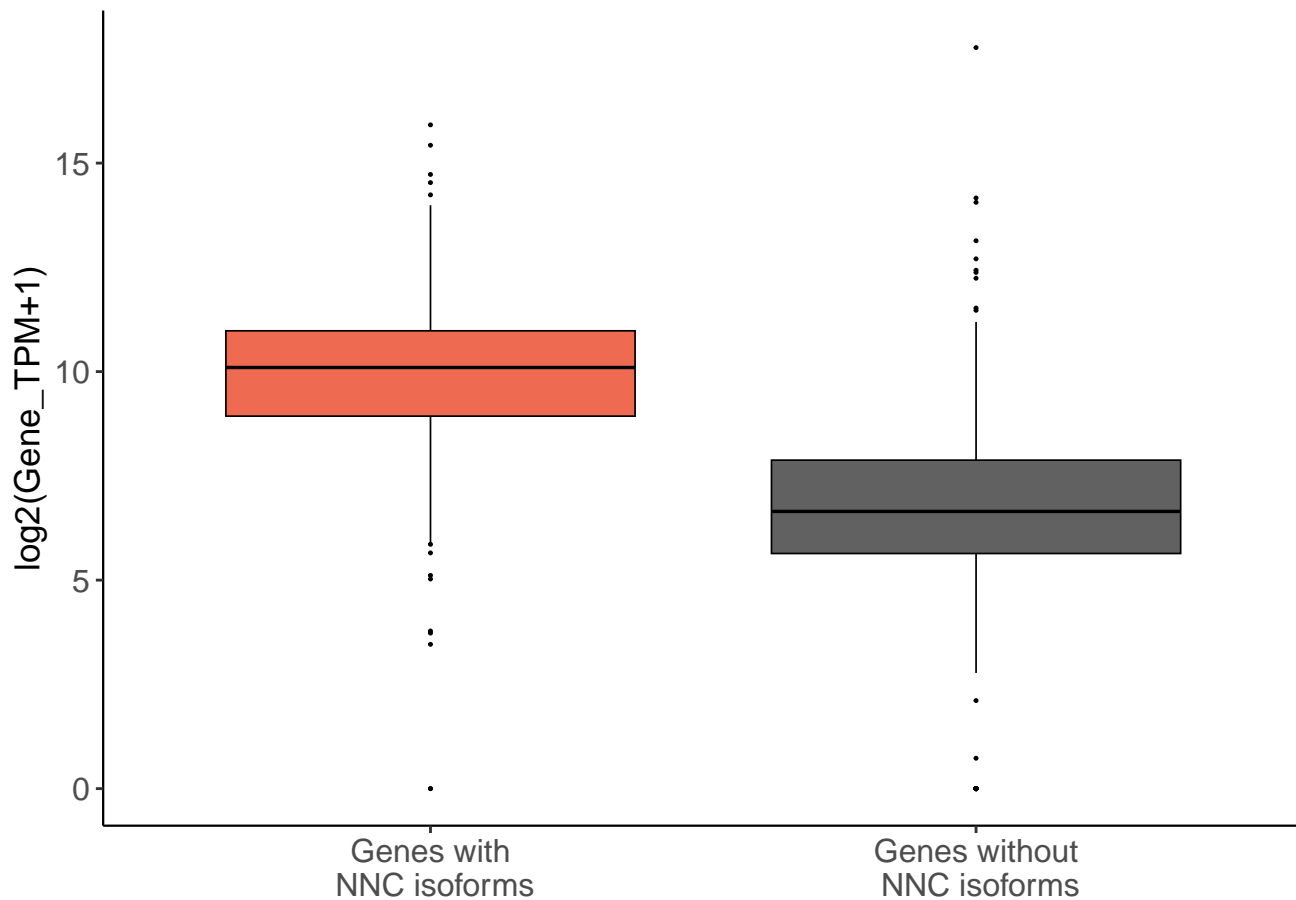


# Exon Count Distribution of Matched Reference Transcripts

Applicable Only to FSM and ISM Categories

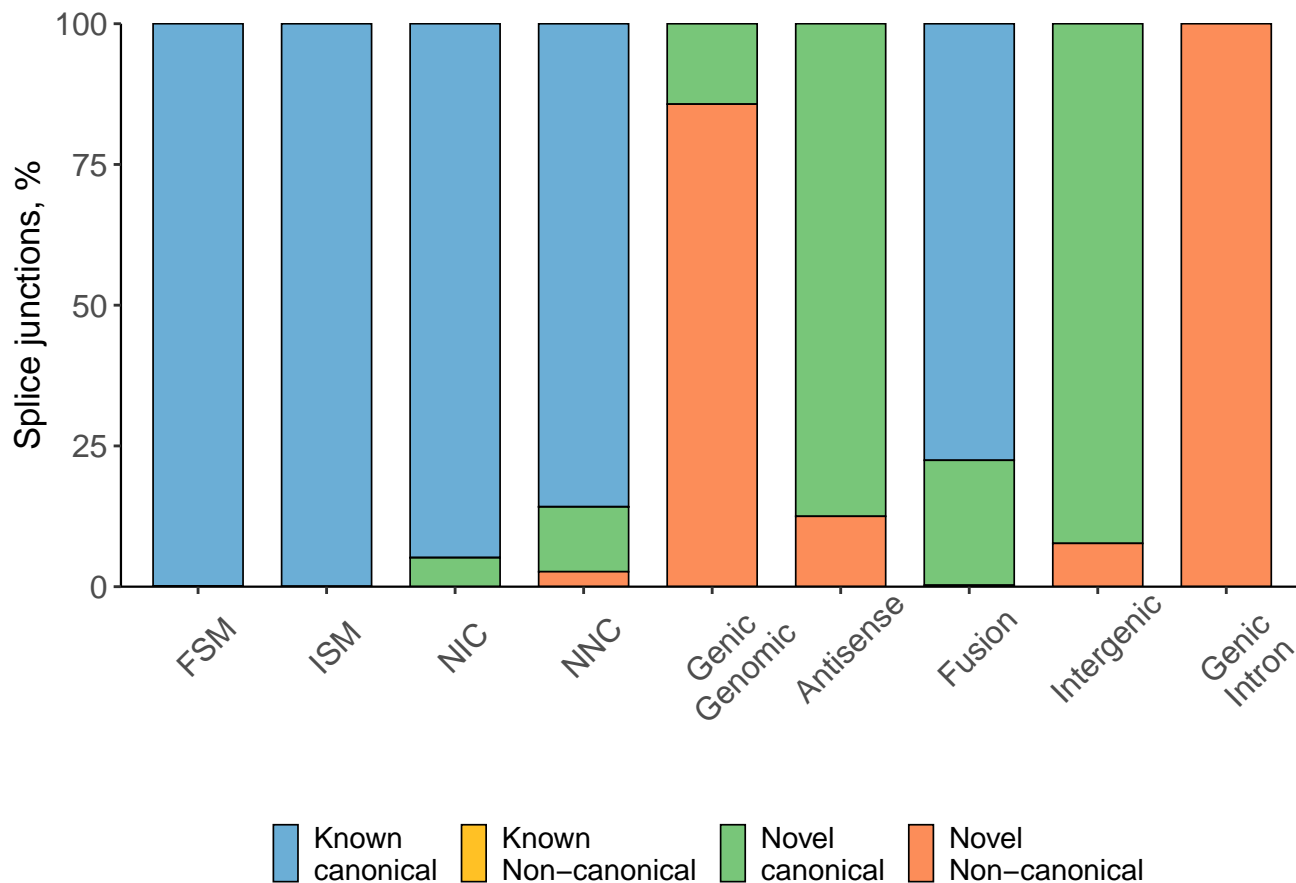


## Gene Expression of NNC And Not NNC Containing Genes

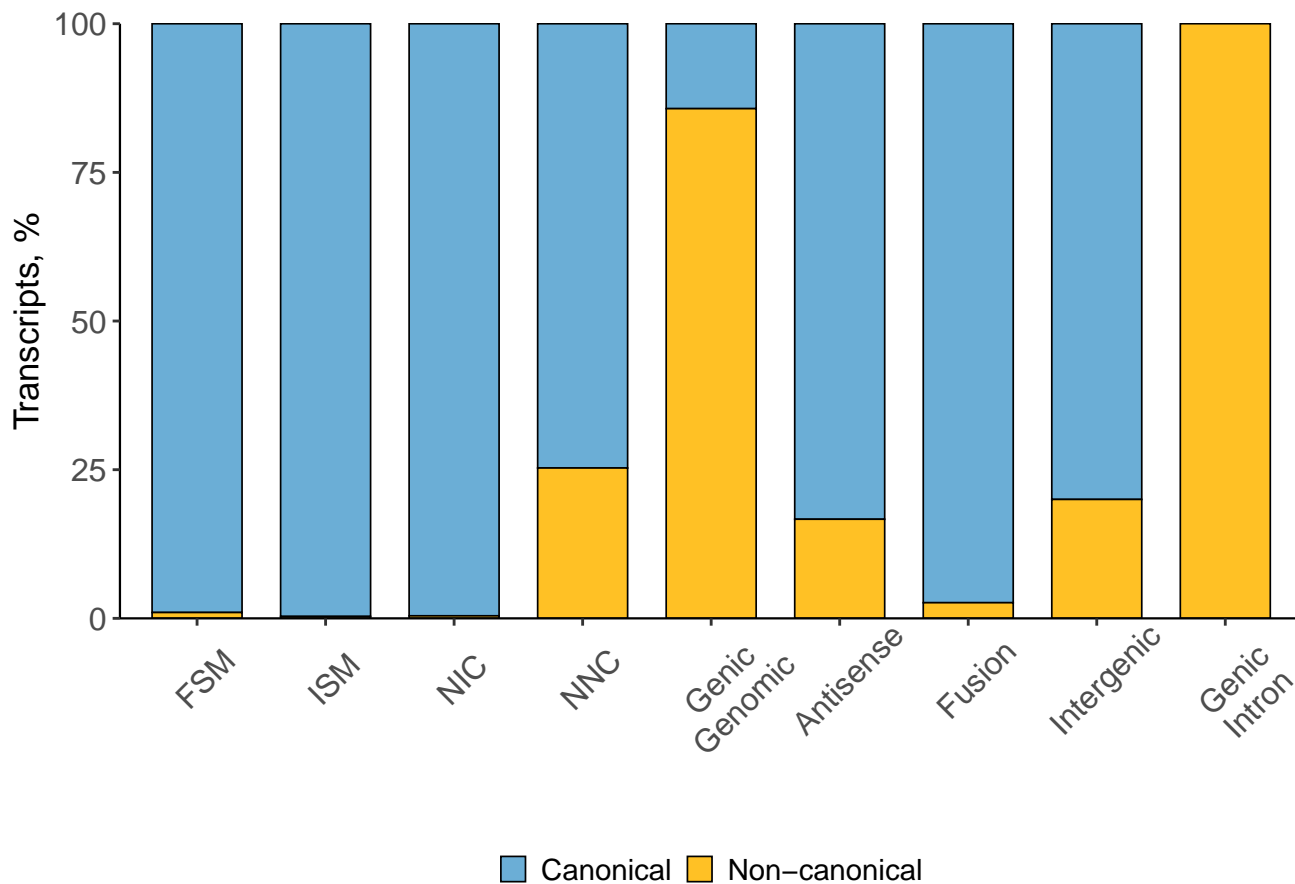


## *Splice Junction Characterization*

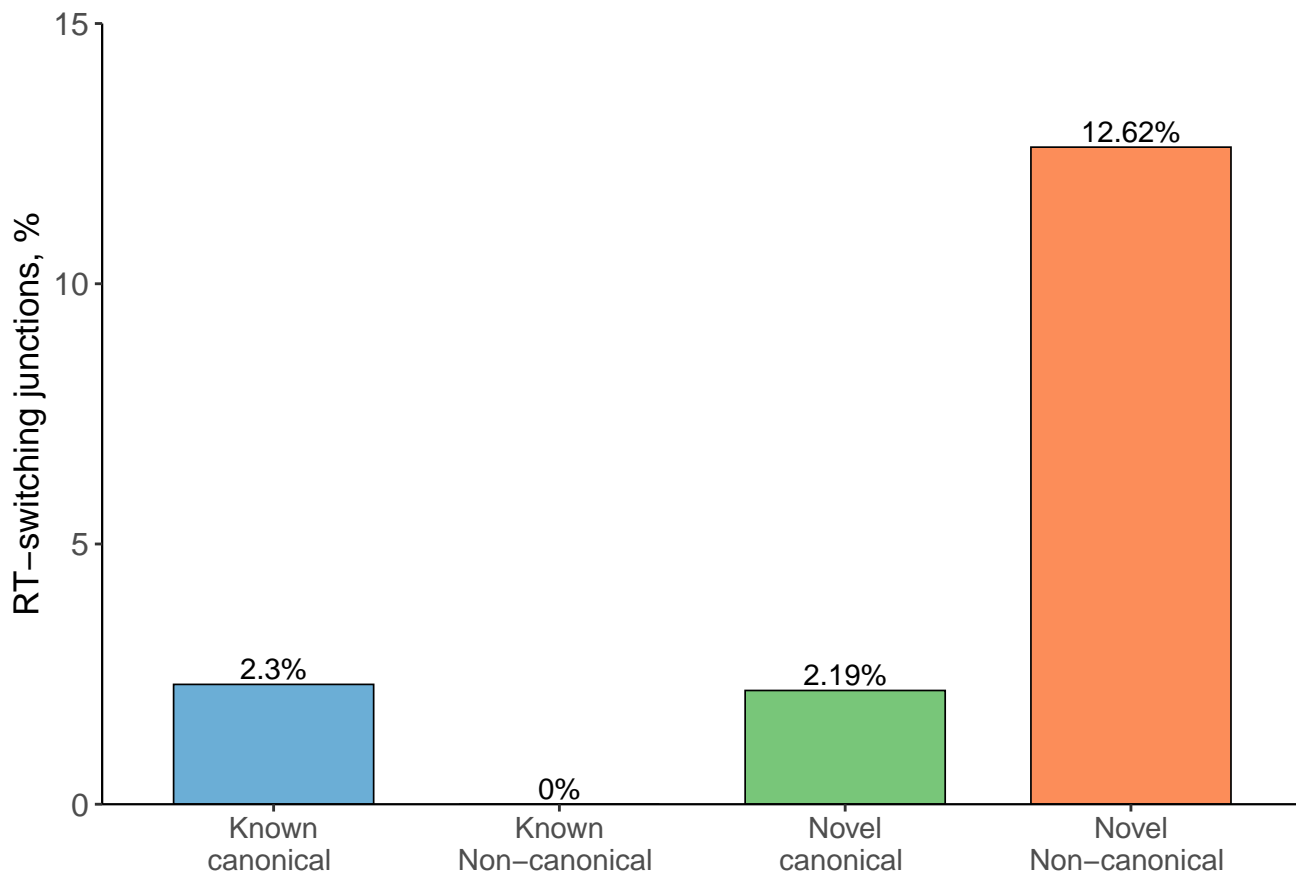
# Distribution of Splice Junctions by Structural Classification



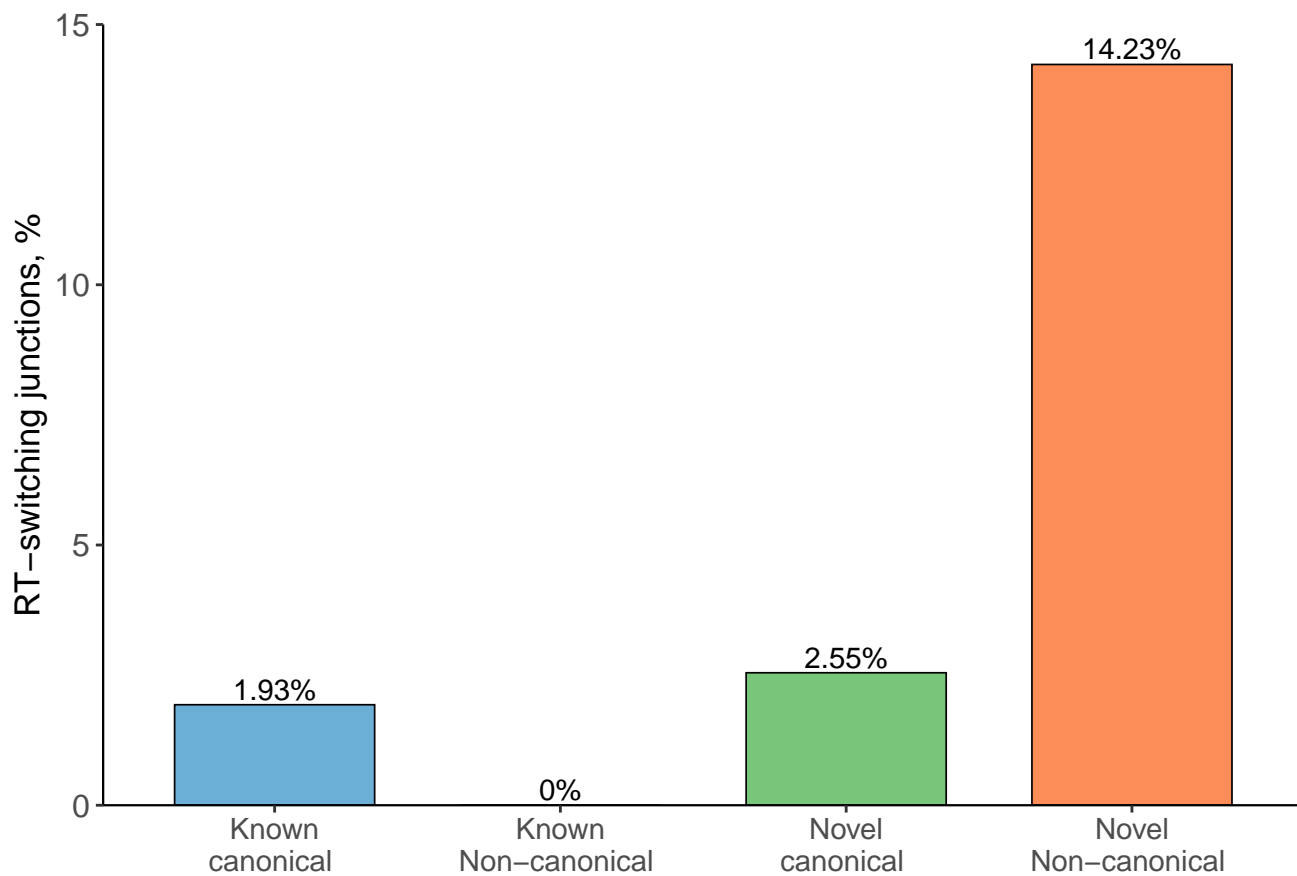
Distribution of Transcripts by Splice Junctions



## RT-Switching All Junctions



## Unique Junctions RT-switching

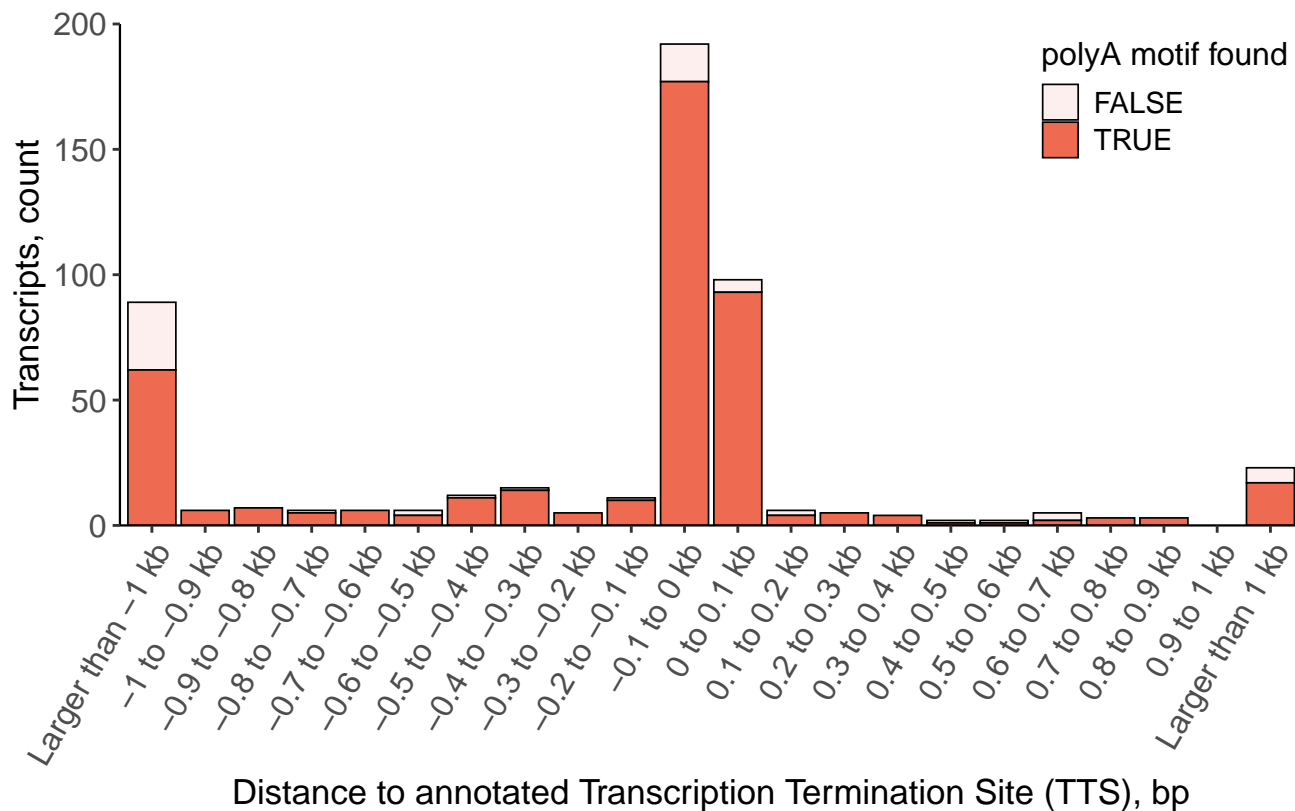


*Comparison With Annotated TSS and TTS*



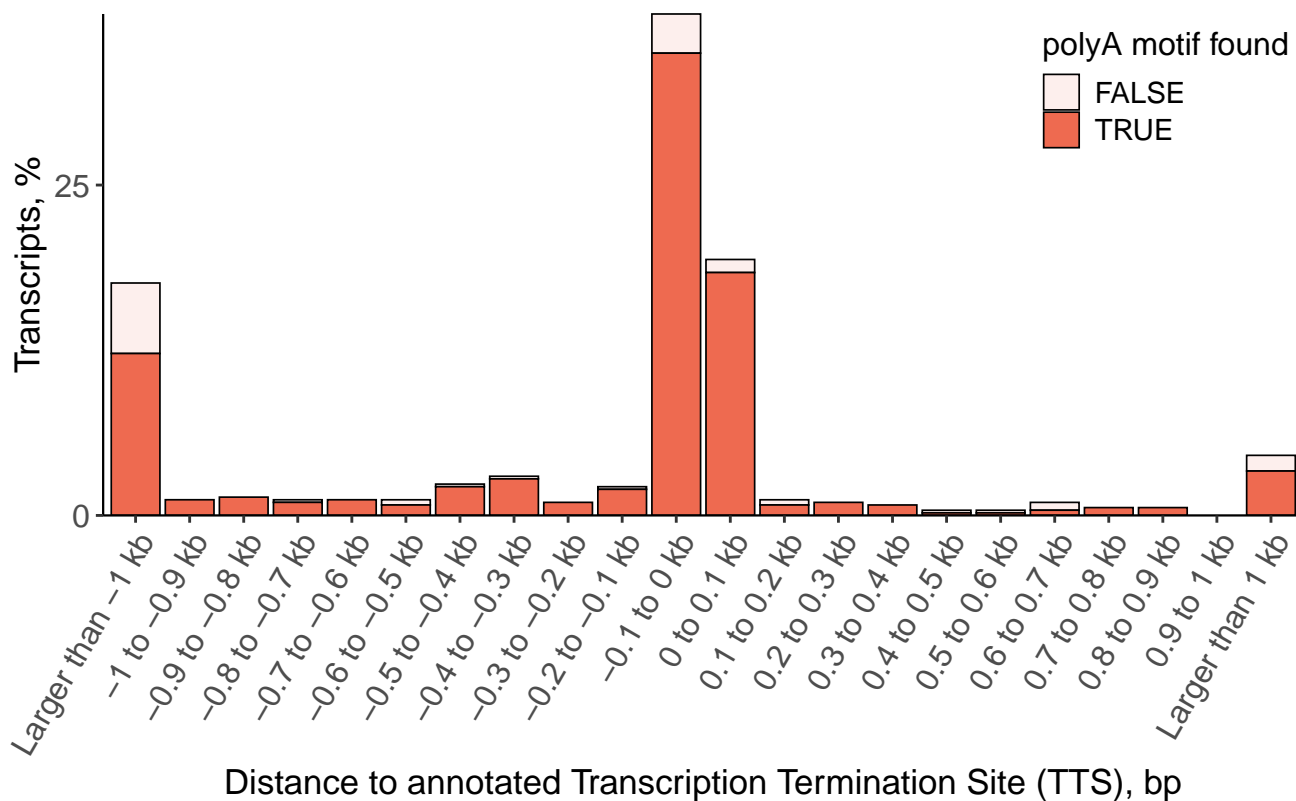
# Distance to annotated Transcription Termination Site (TTS) FSM

Negative values indicate upstream of annotated termination site



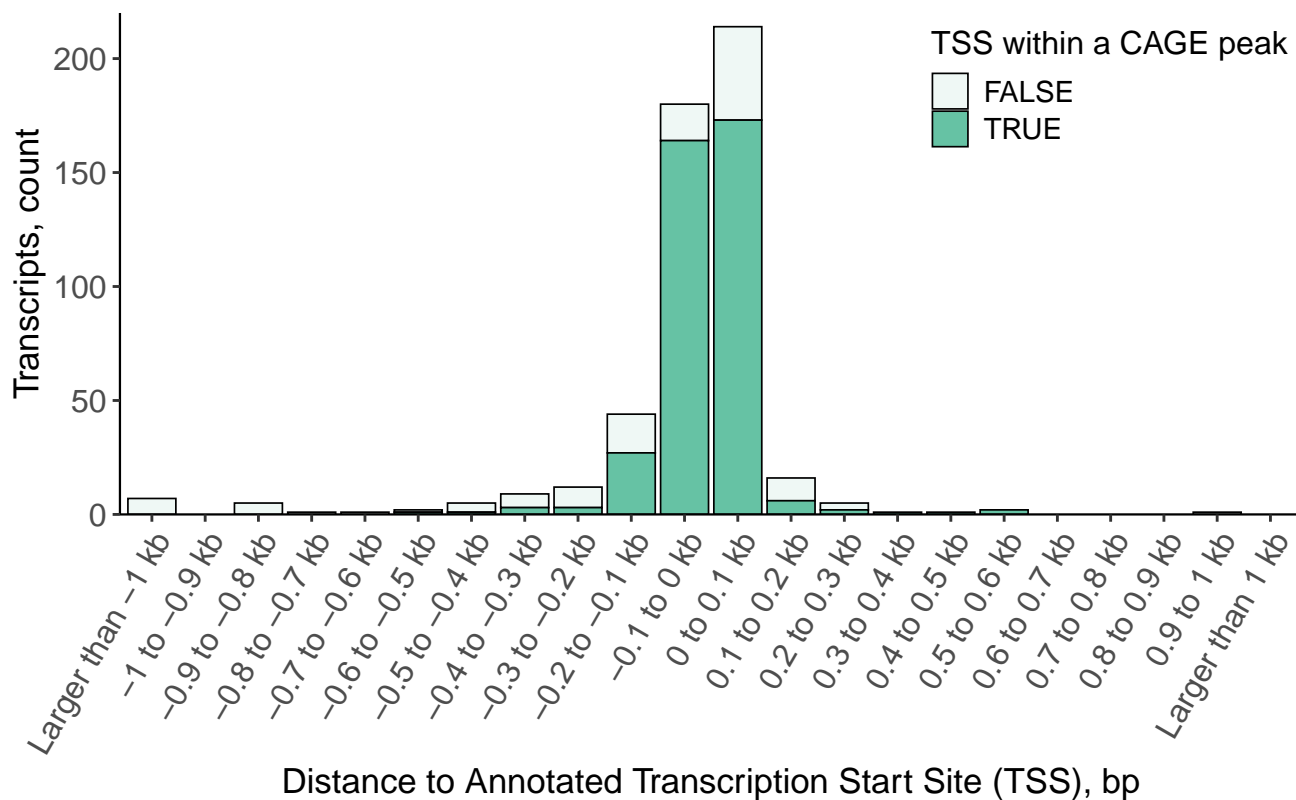
# Distance to annotated Transcription Termination Site (TTS) FSM

Negative values indicate upstream of annotated termination site



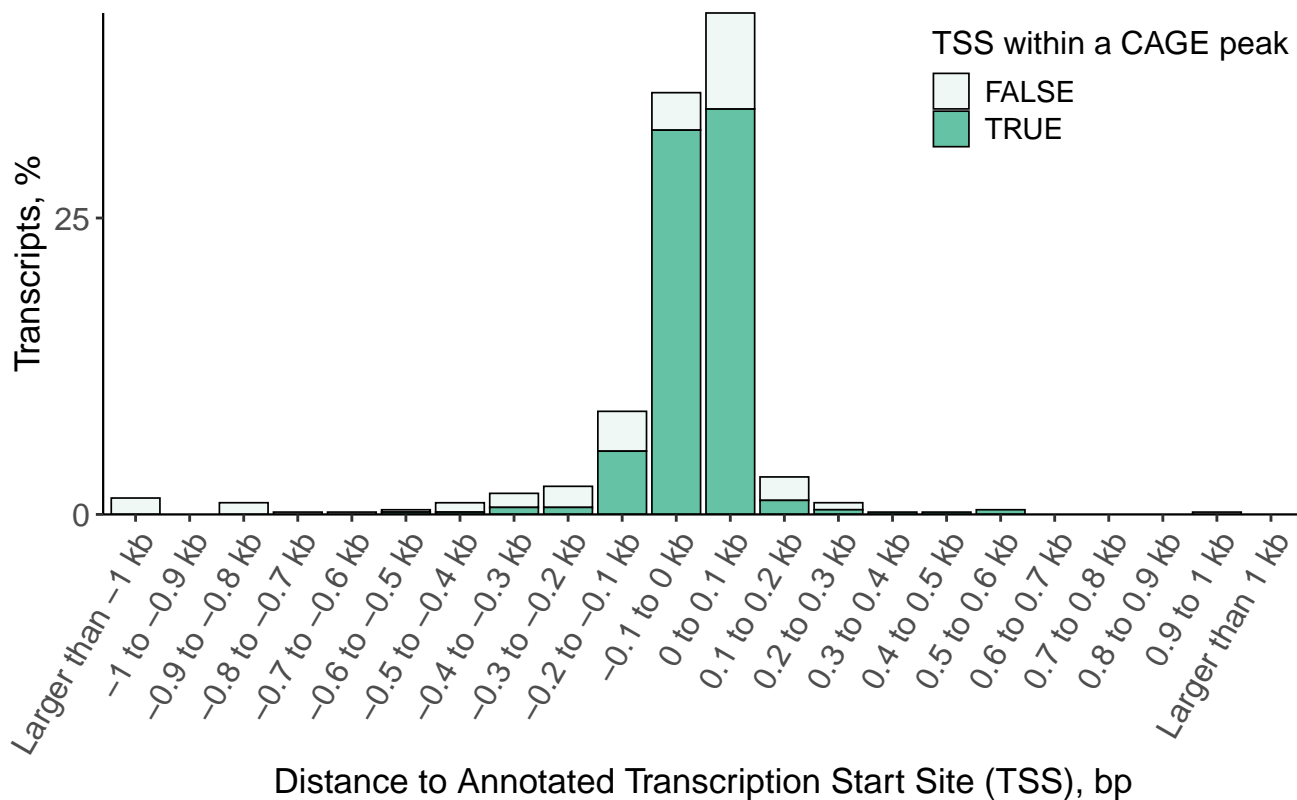
# Distance to Annotated Transcription Start Site for FSM

Negative values indicate downstream of annotated TSS



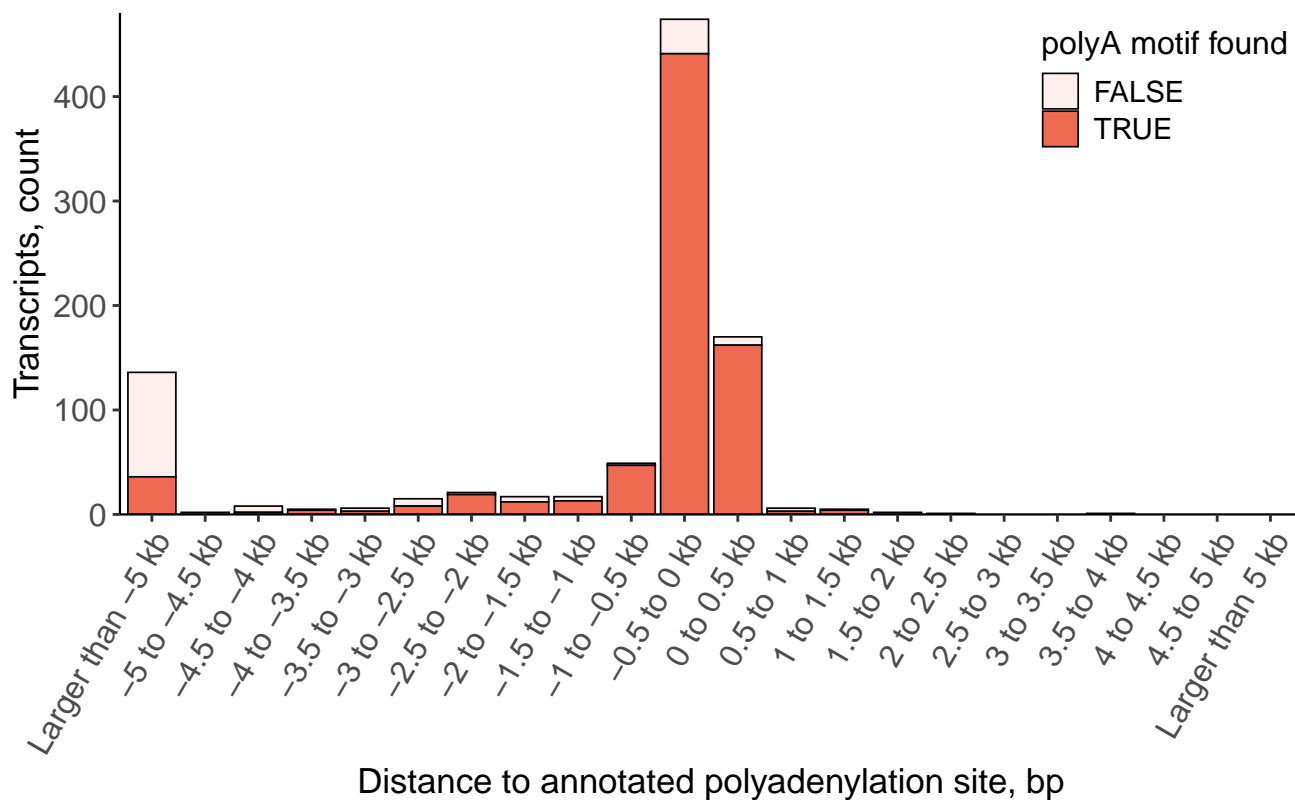
# Distance to Annotated Transcription Start Site for FSM

Negative values indicate downstream of annotated TSS



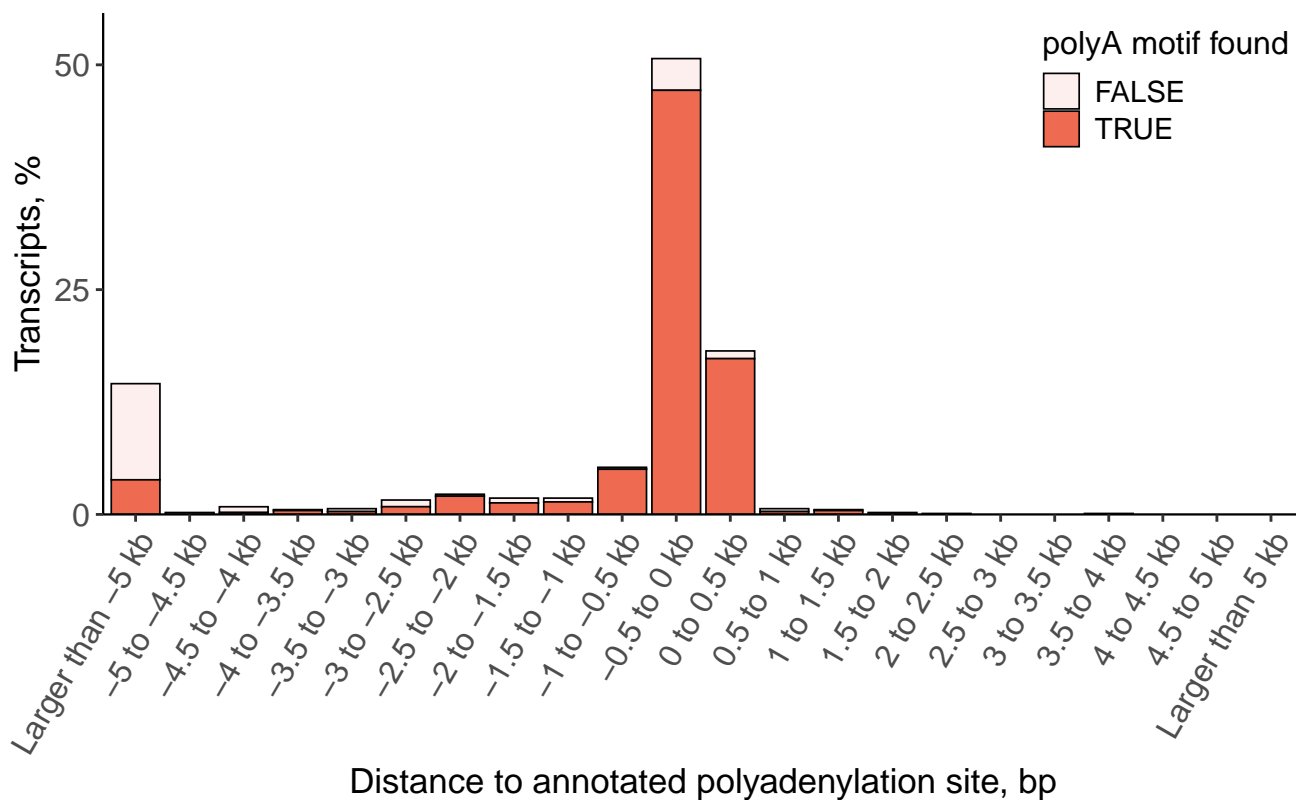
## Distance to Annotated Transcription Termination Site (TTS)

Negative values indicate upstream of annotated TTS



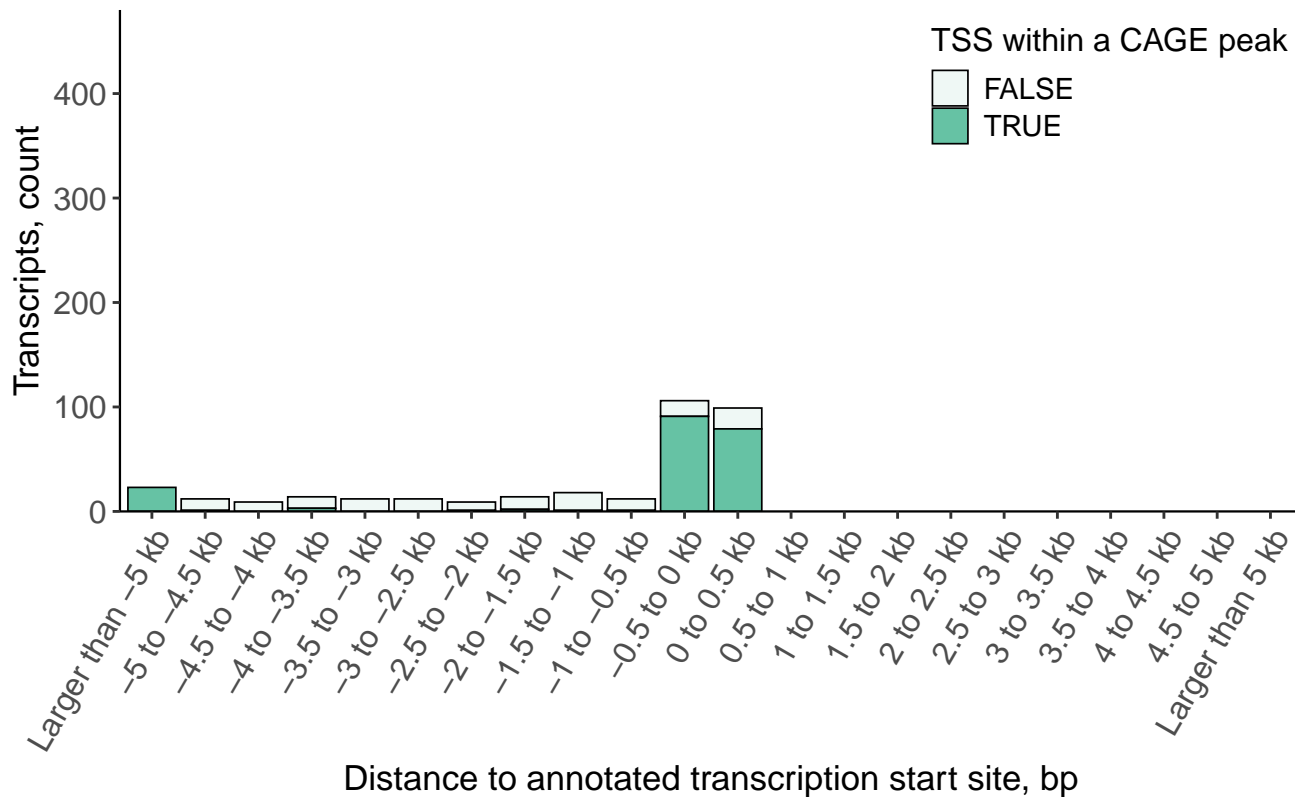
# Distance to Annotated Transcription Termination Site (TTS)

Negative values indicate upstream of annotated TTS



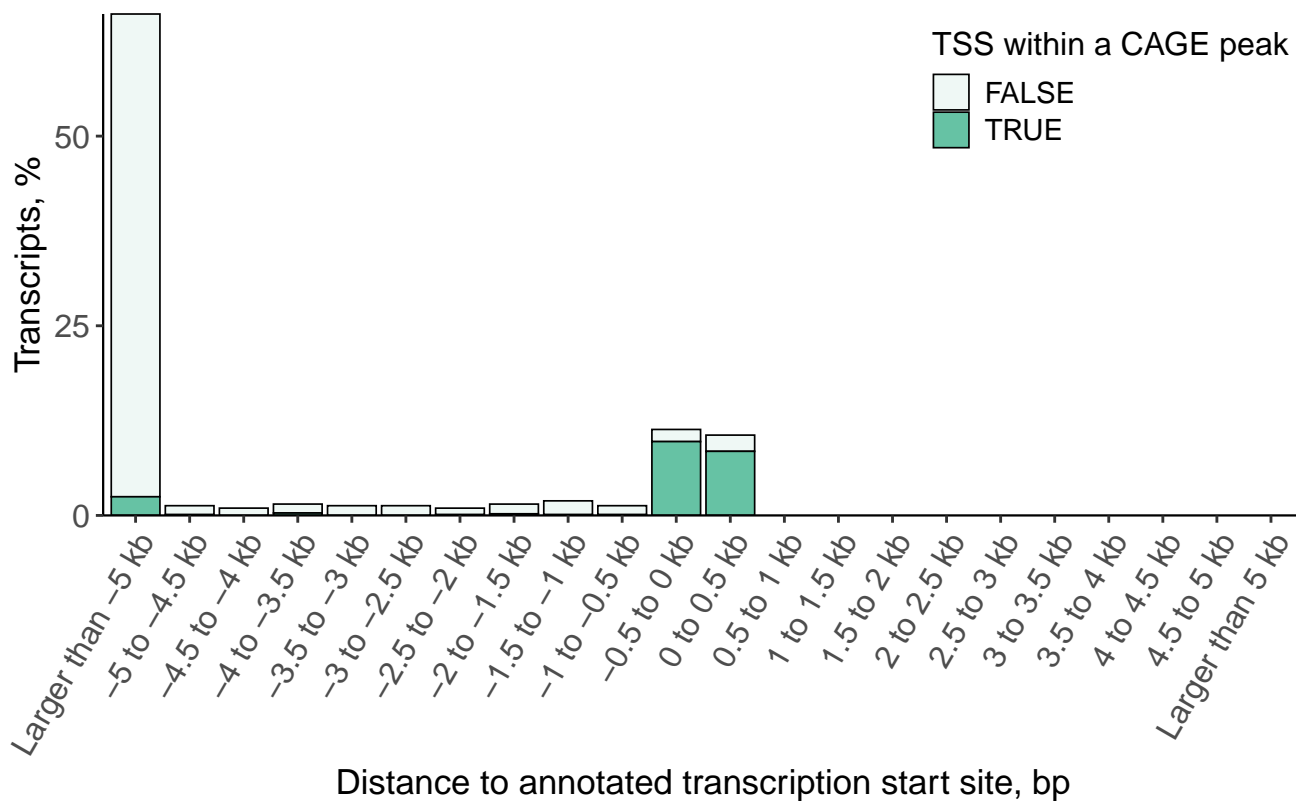
# Distance to Annotated Transcription Start Site for ISM

Negative values indicate downstream of annotated TSS



## Distance to Annotated Transcription Start Site for ISM

Negative values indicate downstream of annotated TSS

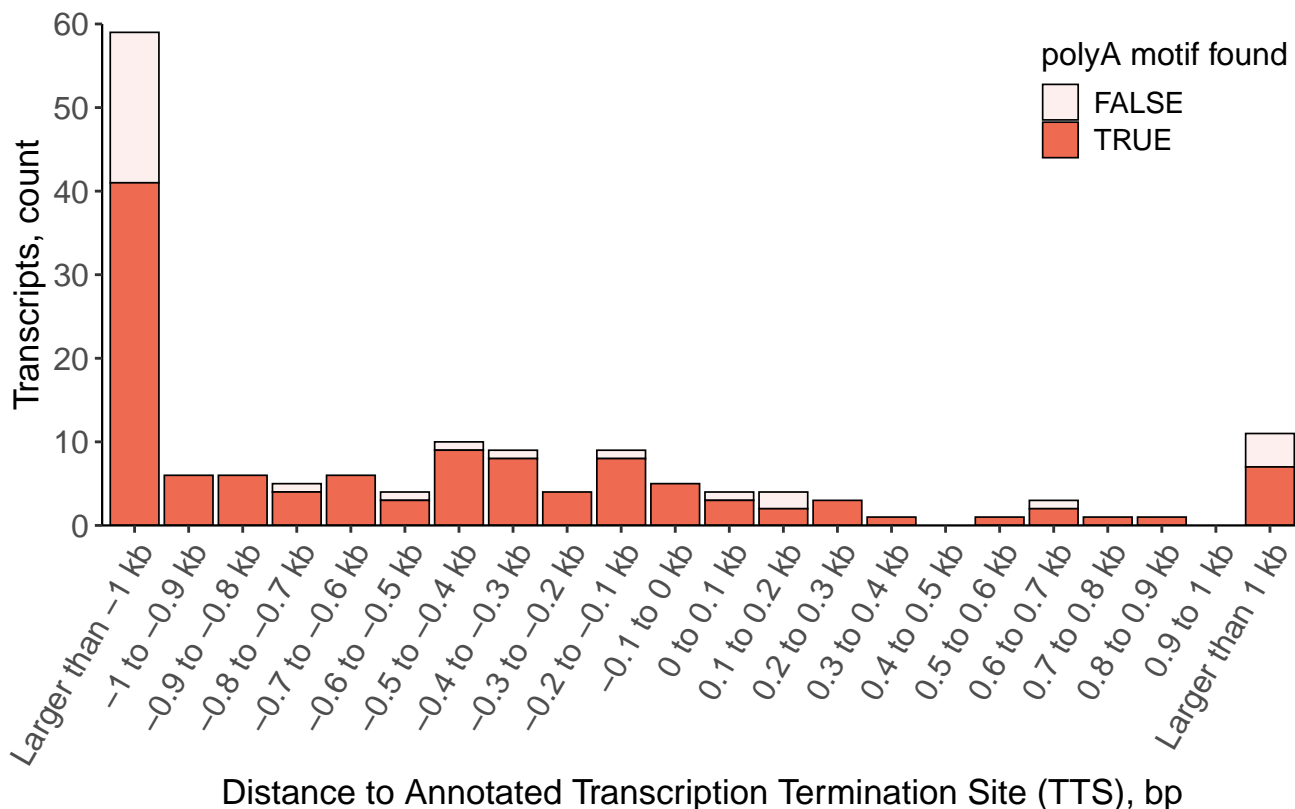




*Comparison With Annotated TSS and TTS  
by Subcategories*

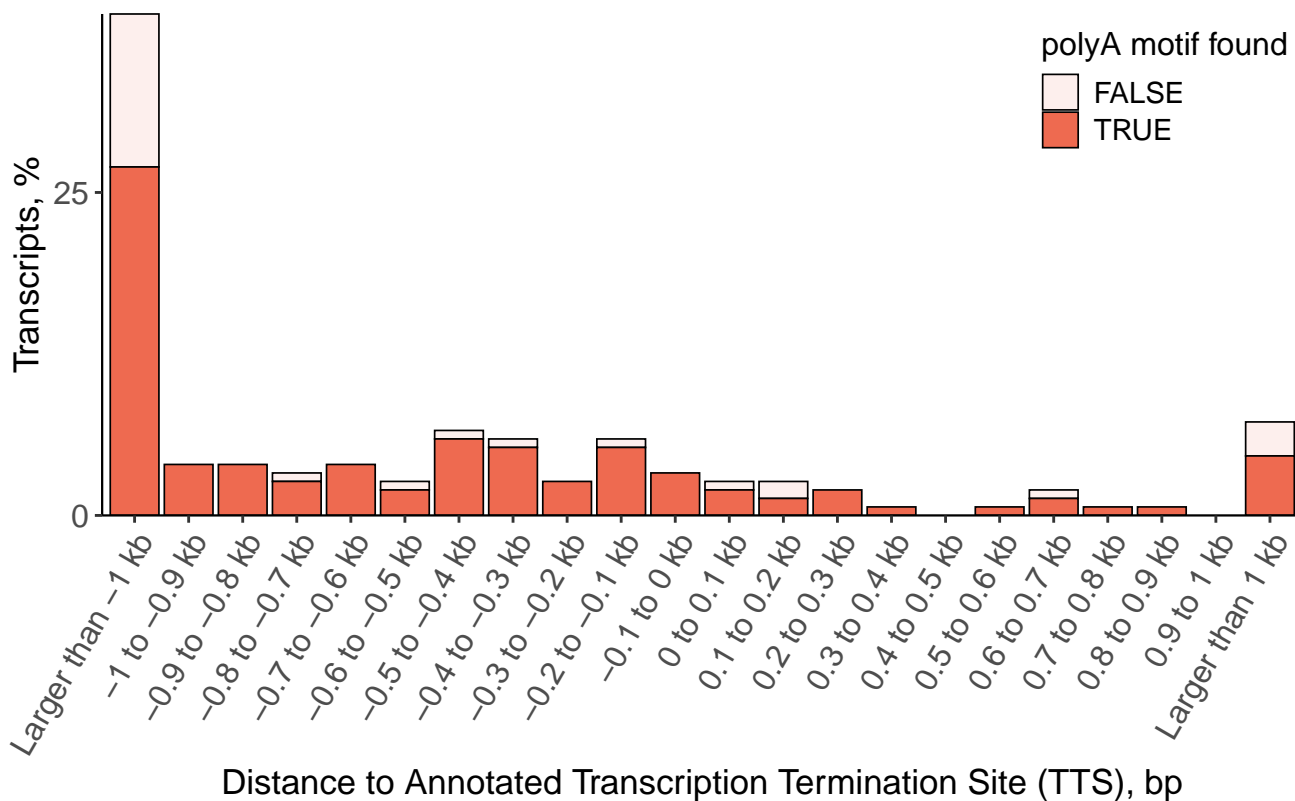
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 3'End

Negative values indicate upstream of annotated termination site



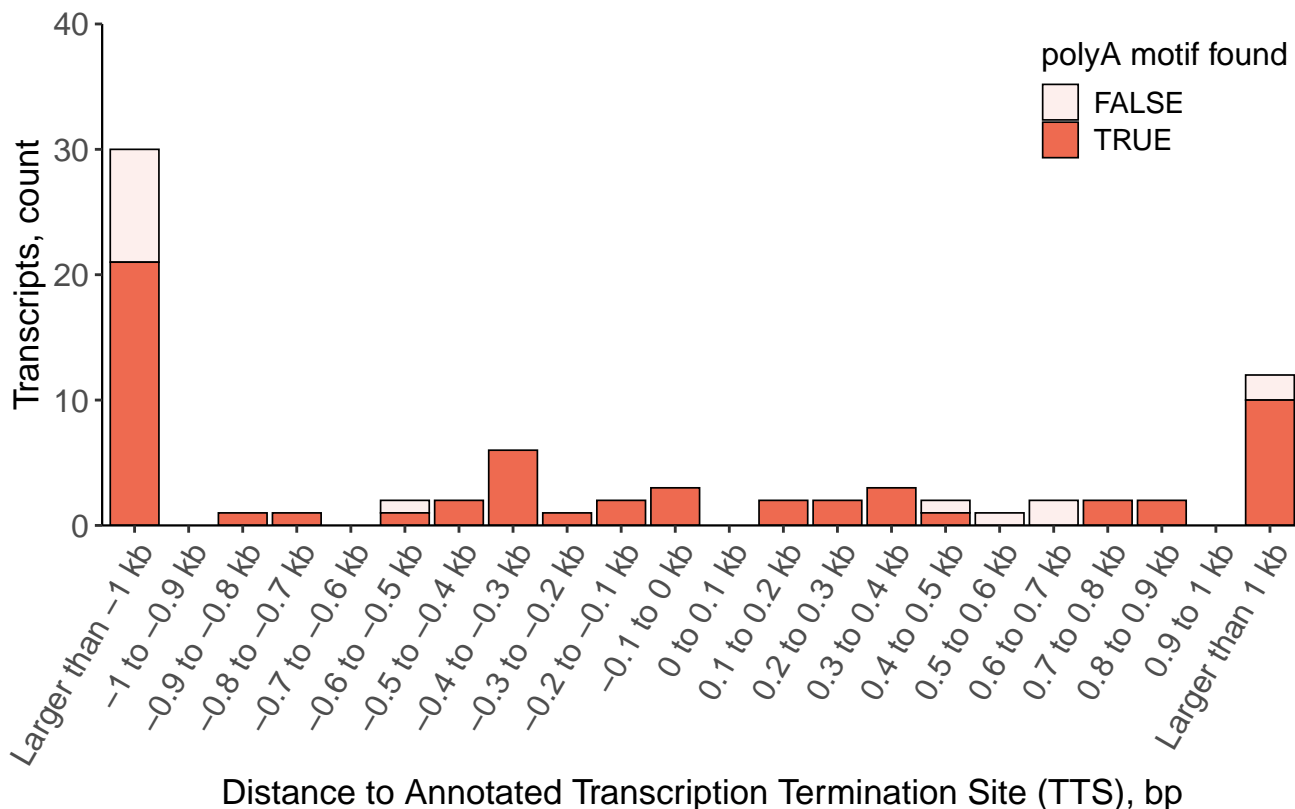
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 3'End

Negative values indicate upstream of annotated termination site



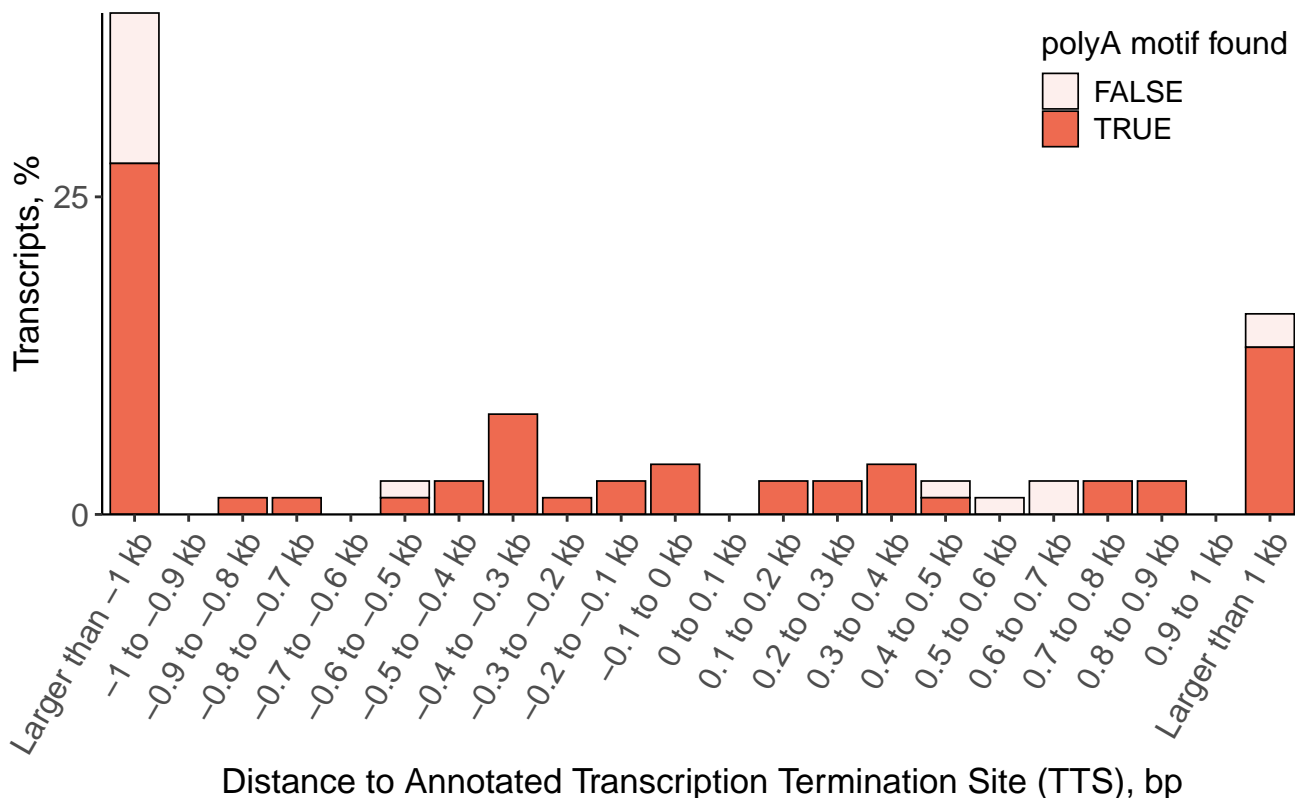
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 3'5'End

Negative values indicate upstream of annotated termination site



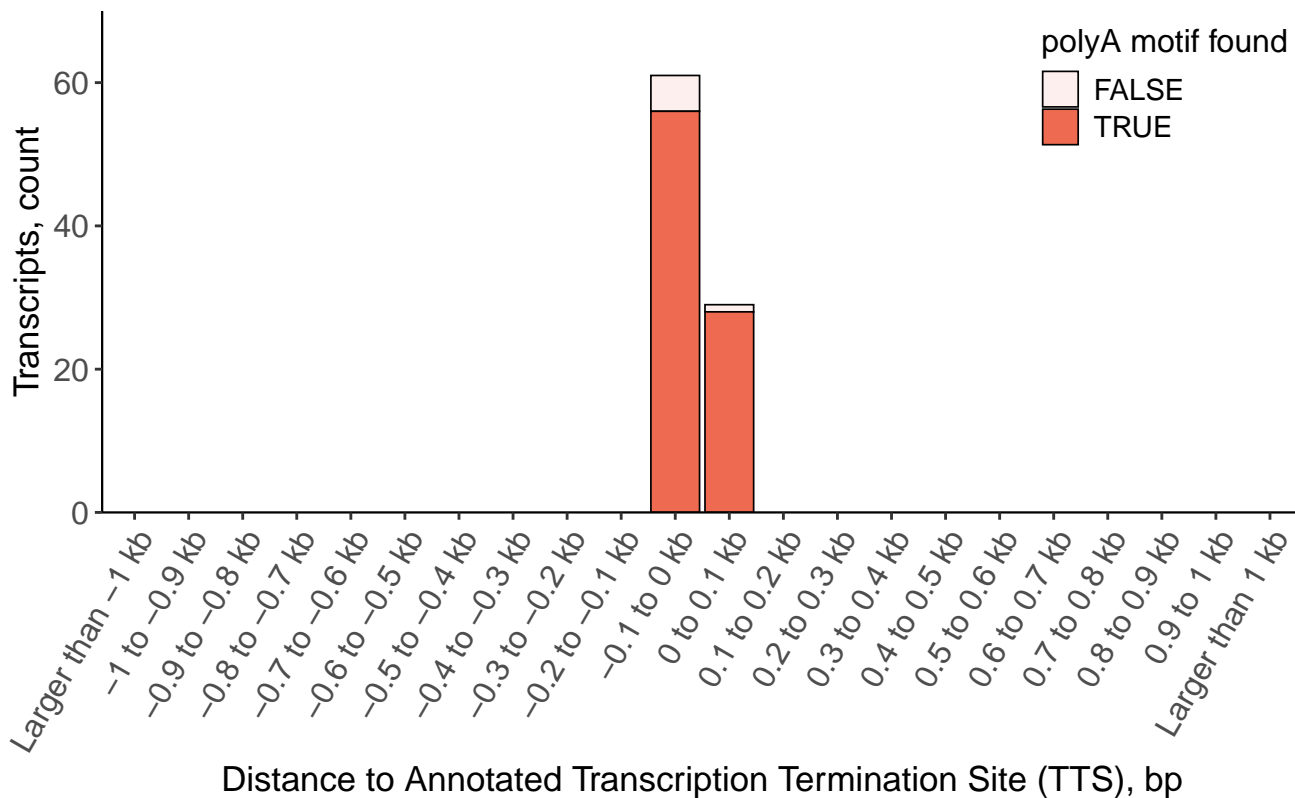
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 3'5'End

Negative values indicate upstream of annotated termination site



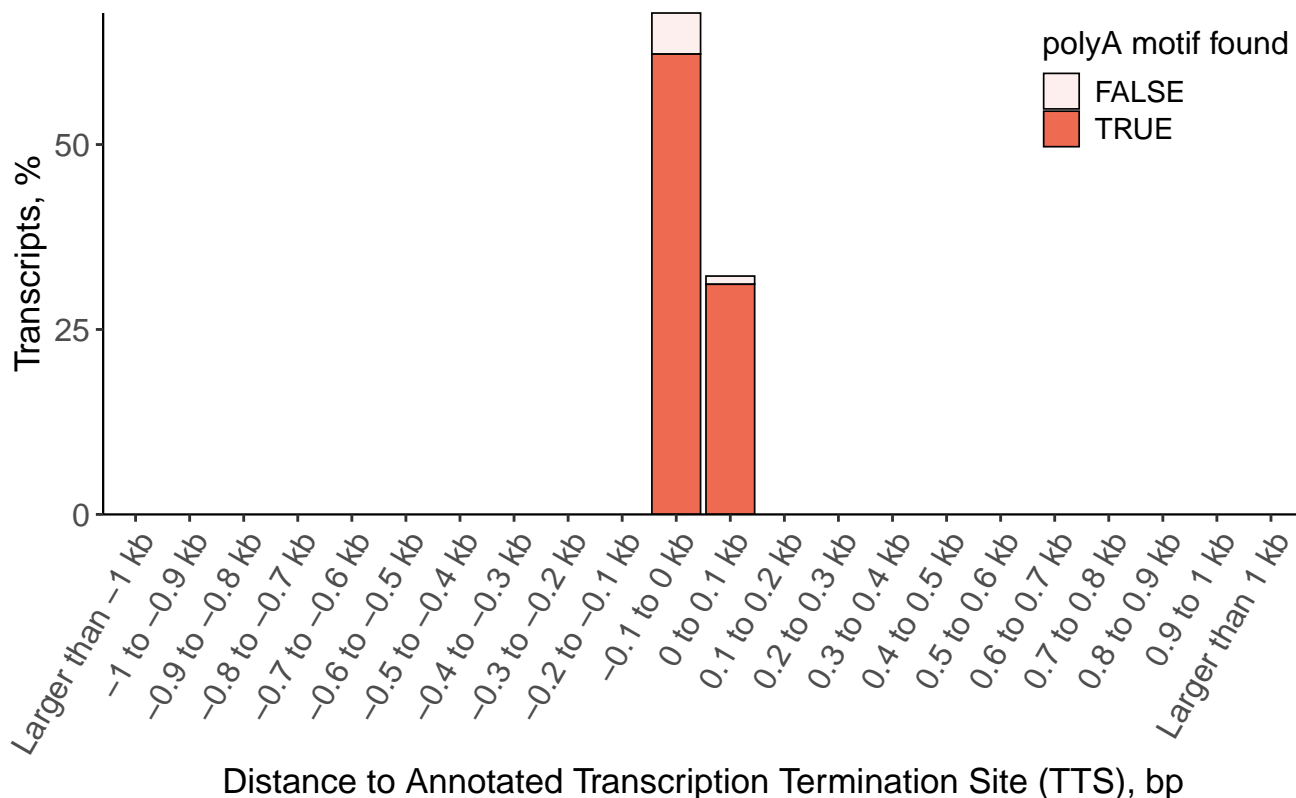
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 5'End

Negative values indicate upstream of annotated termination site



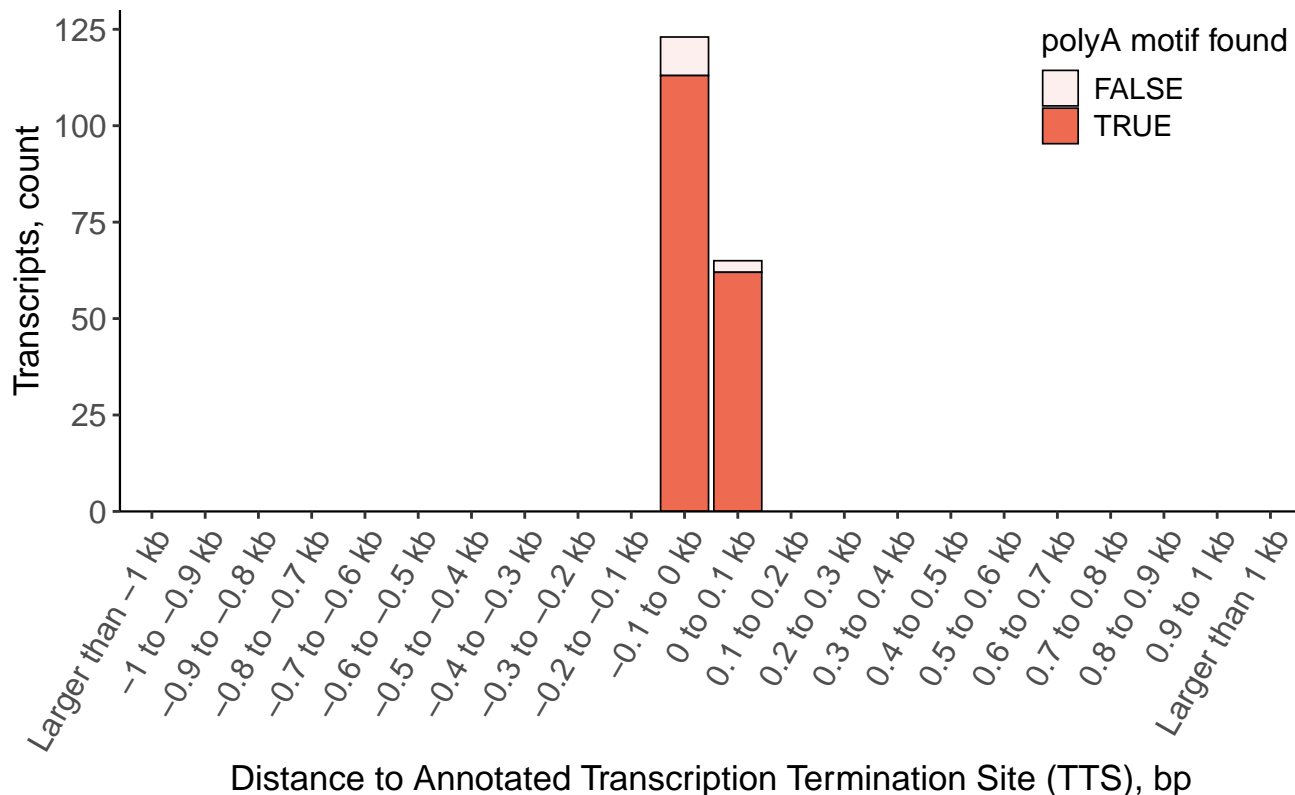
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 5'End

Negative values indicate upstream of annotated termination site



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

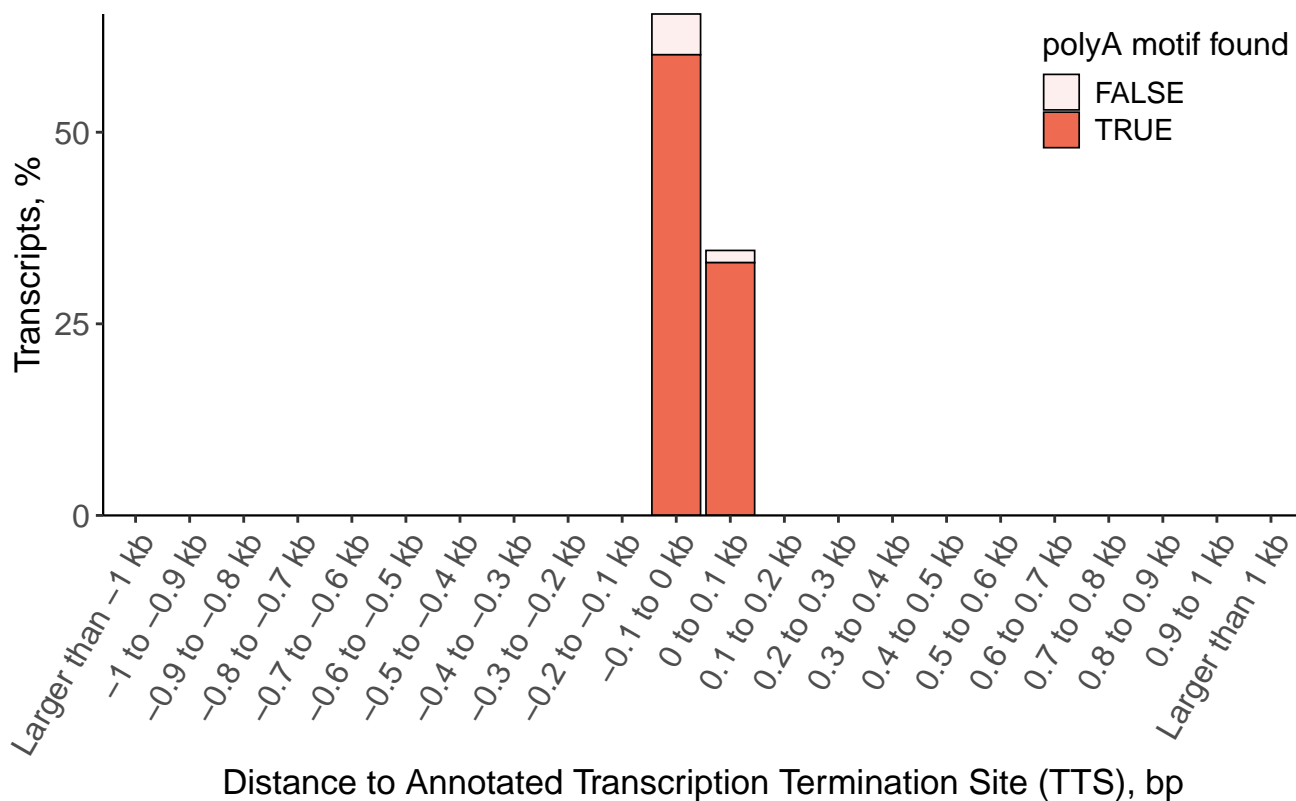
Negative values indicate upstream of annotated termination site





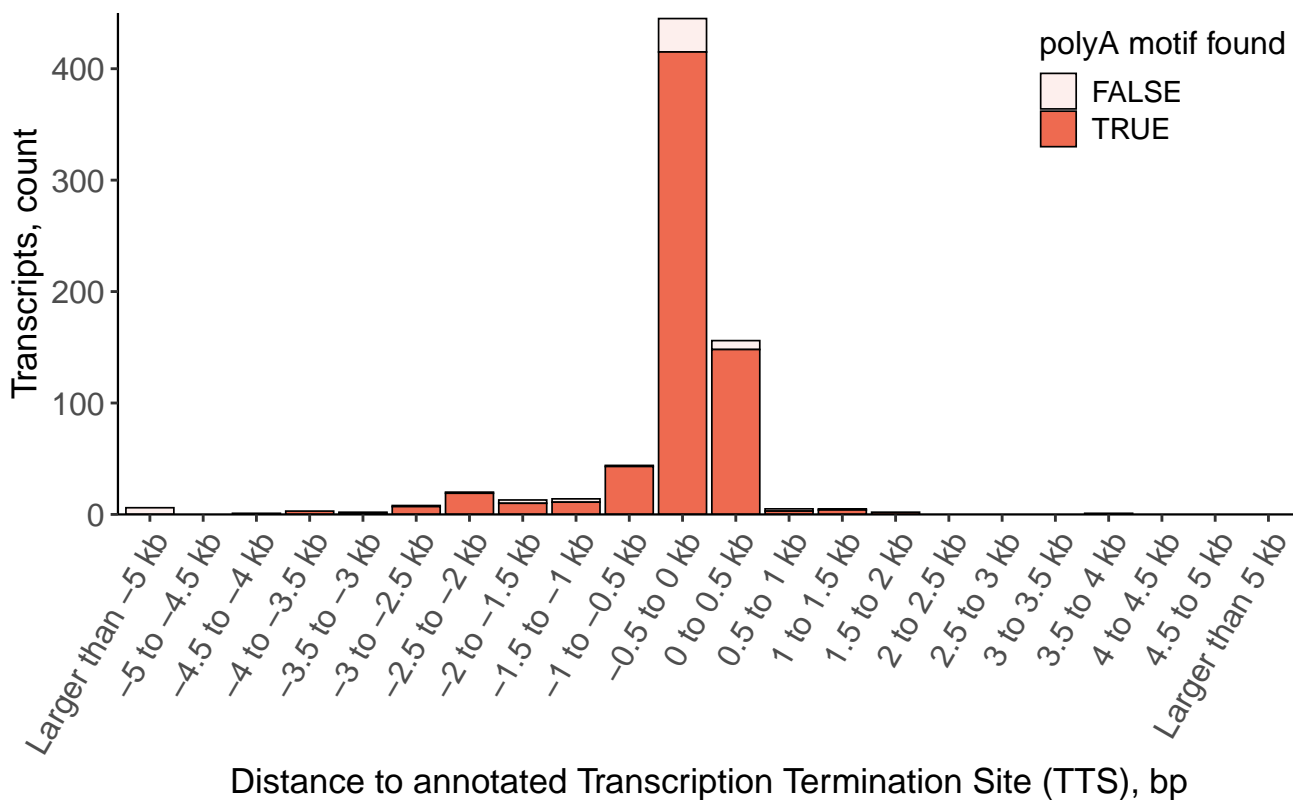
# Distance to annotated Transcription Termination Site (TTS) FSM Reference Match

Negative values indicate upstream of annotated termination site



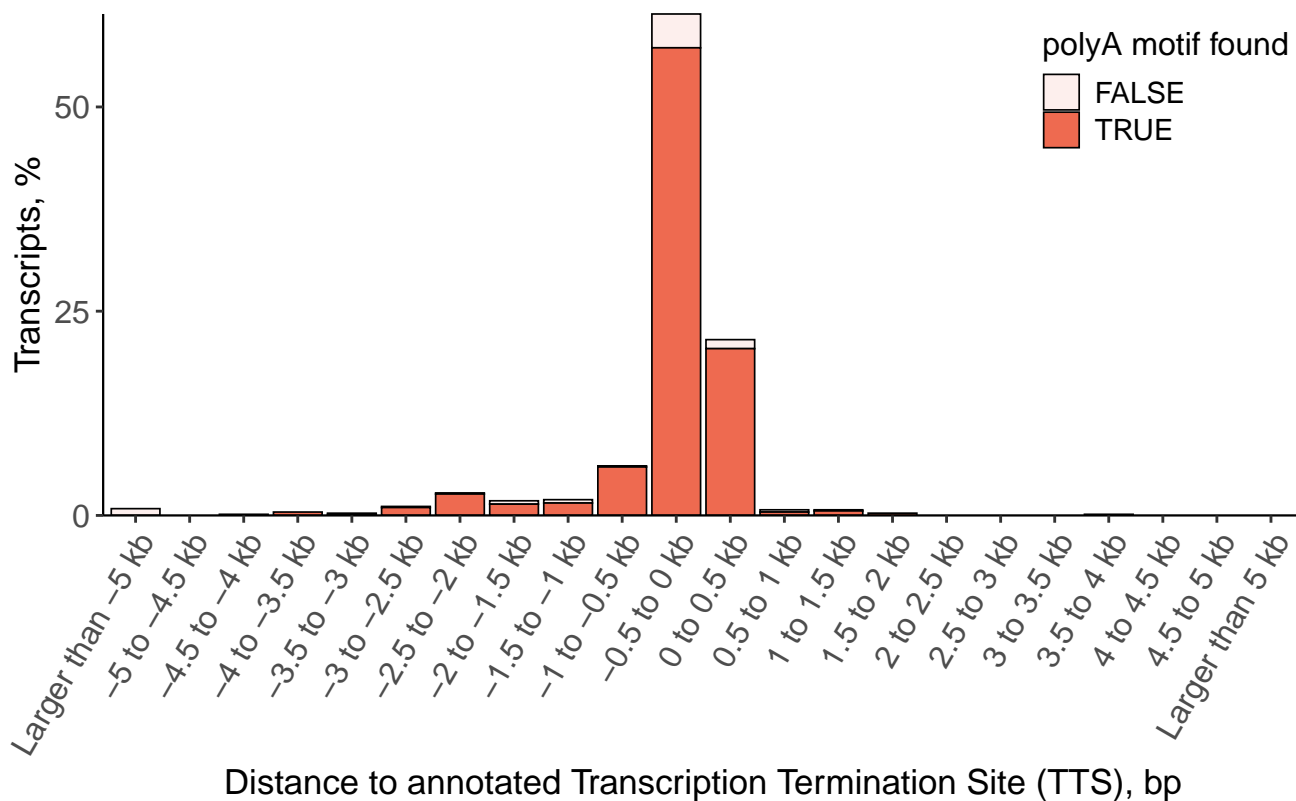
# Distance to Annotated Transcription Termination Site for ISM 3' Fragment

Negative values indicate upstream of annotated TTS



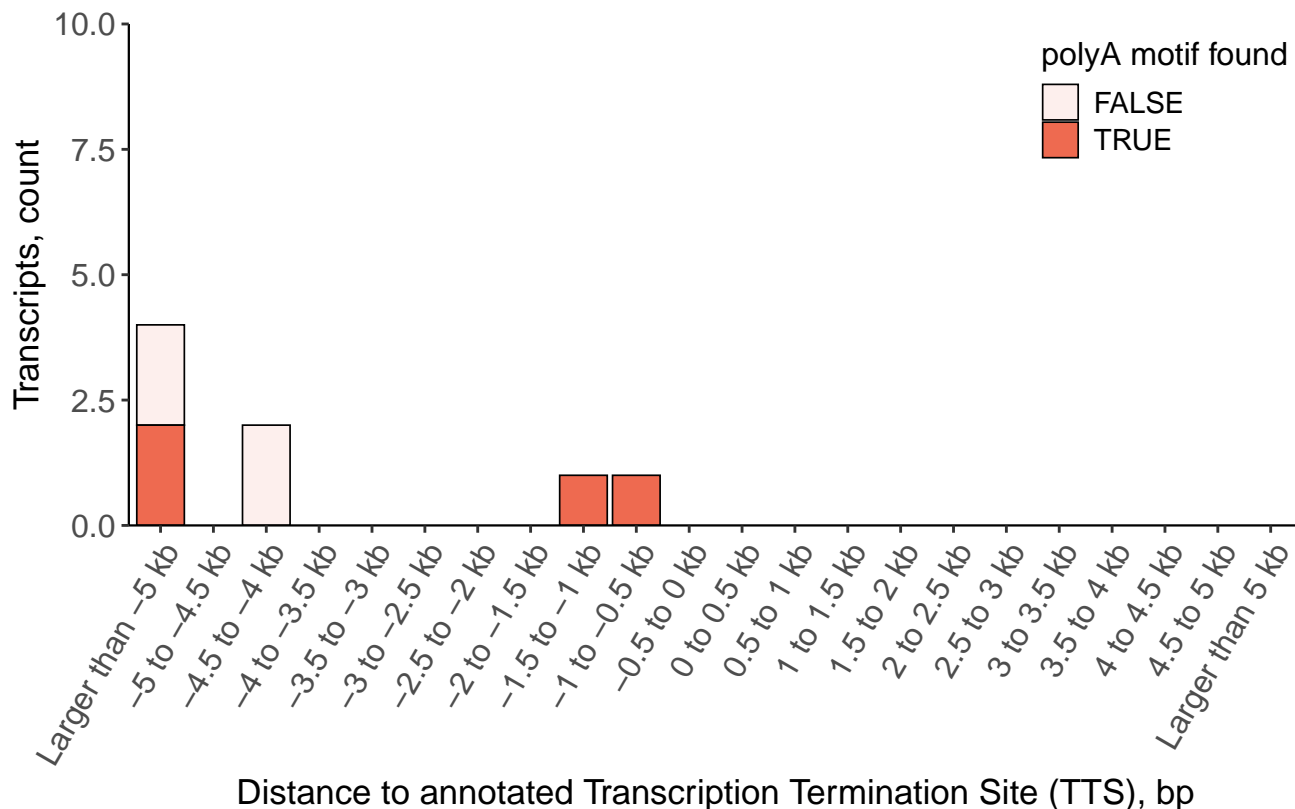
# Distance to Annotated Transcription Termination Site for ISM 3' Fragment

Negative values indicate upstream of annotated TTS



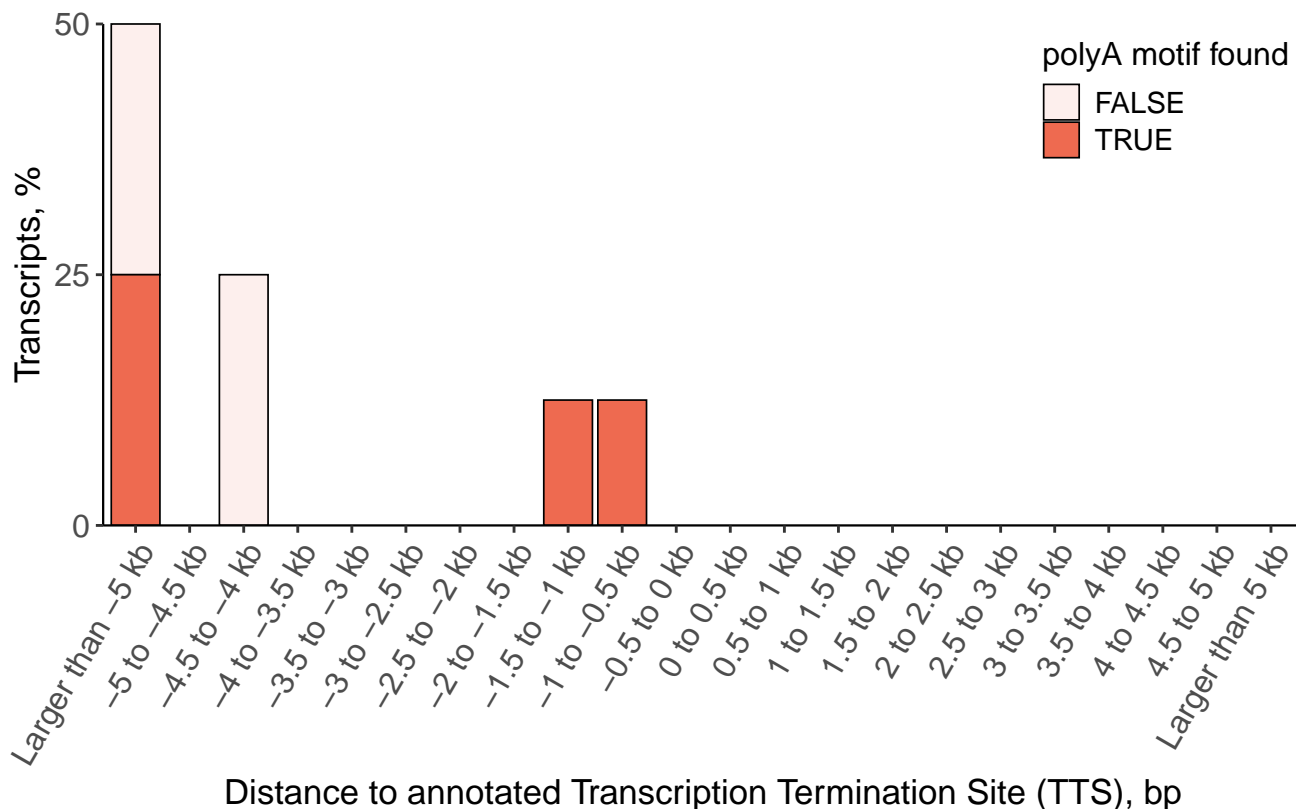
# Distance to Annotated Transcription Termination Site for ISM Internal Fragment

Negative values indicate upstream of annotated TTS



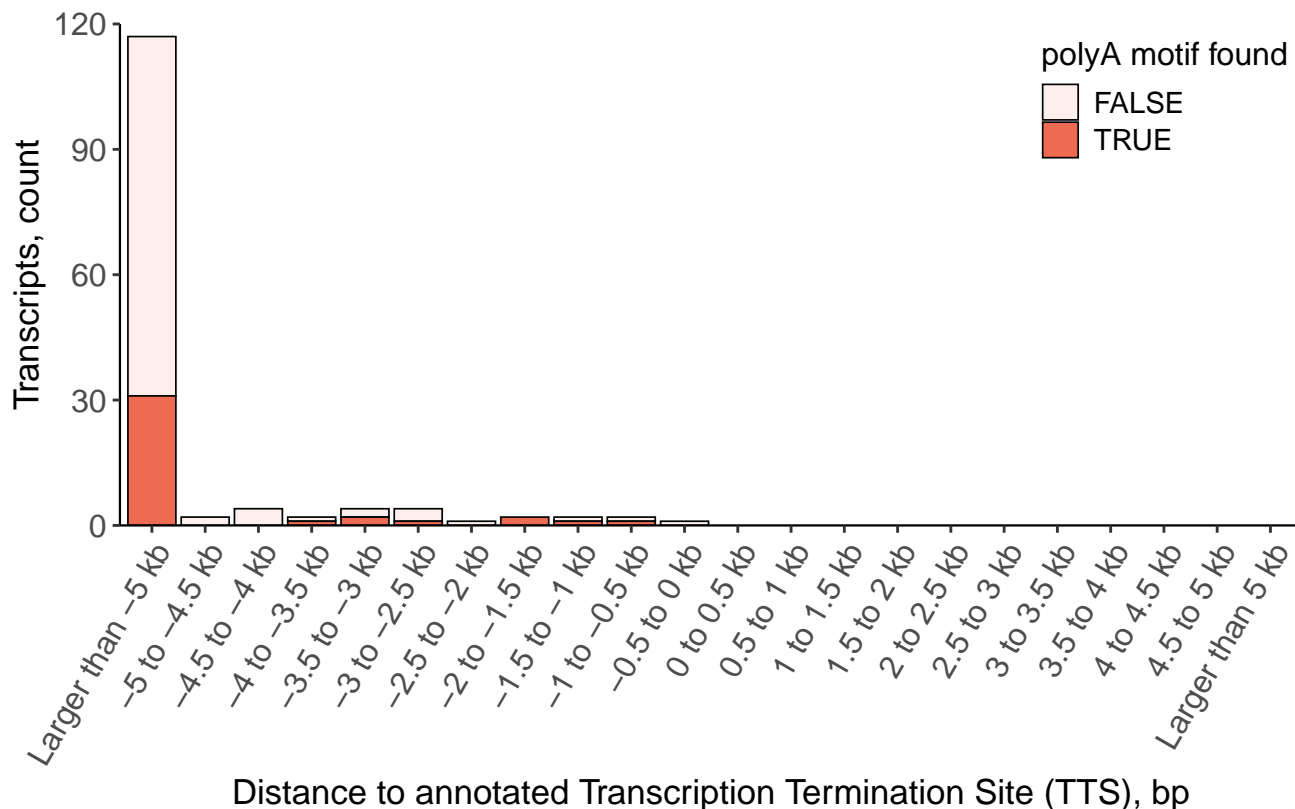
# Distance to Annotated Transcription Termination Site for ISM Internal Fragment

Negative values indicate upstream of annotated TTS



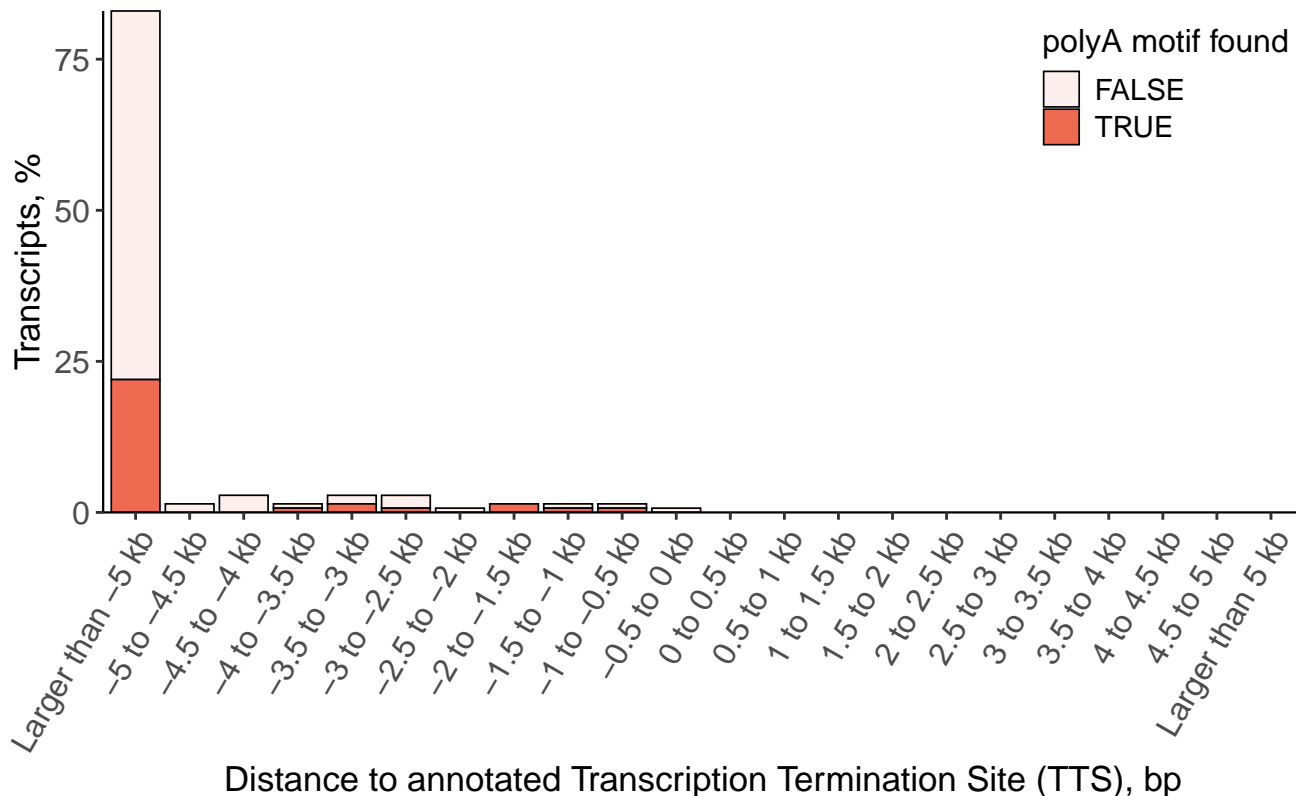
# Distance to Annotated Transcription Termination Site for ISM A5' Fragment

Negative values indicate upstream of annotated TTS



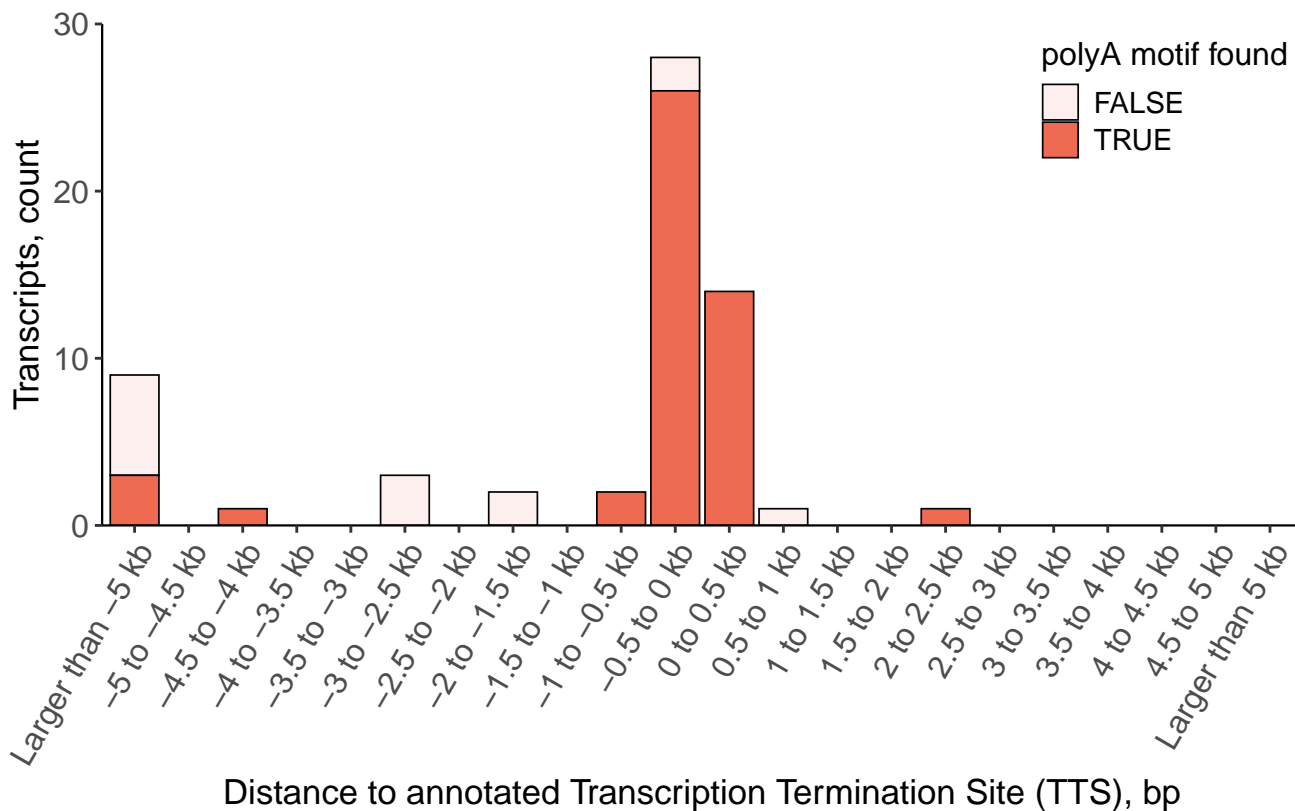
# Distance to Annotated Transcription Termination Site for ISM A5' Fragment

Negative values indicate upstream of annotated TTS



# Distance to Annotated Transcription Termination Site for ISM Intron Retention

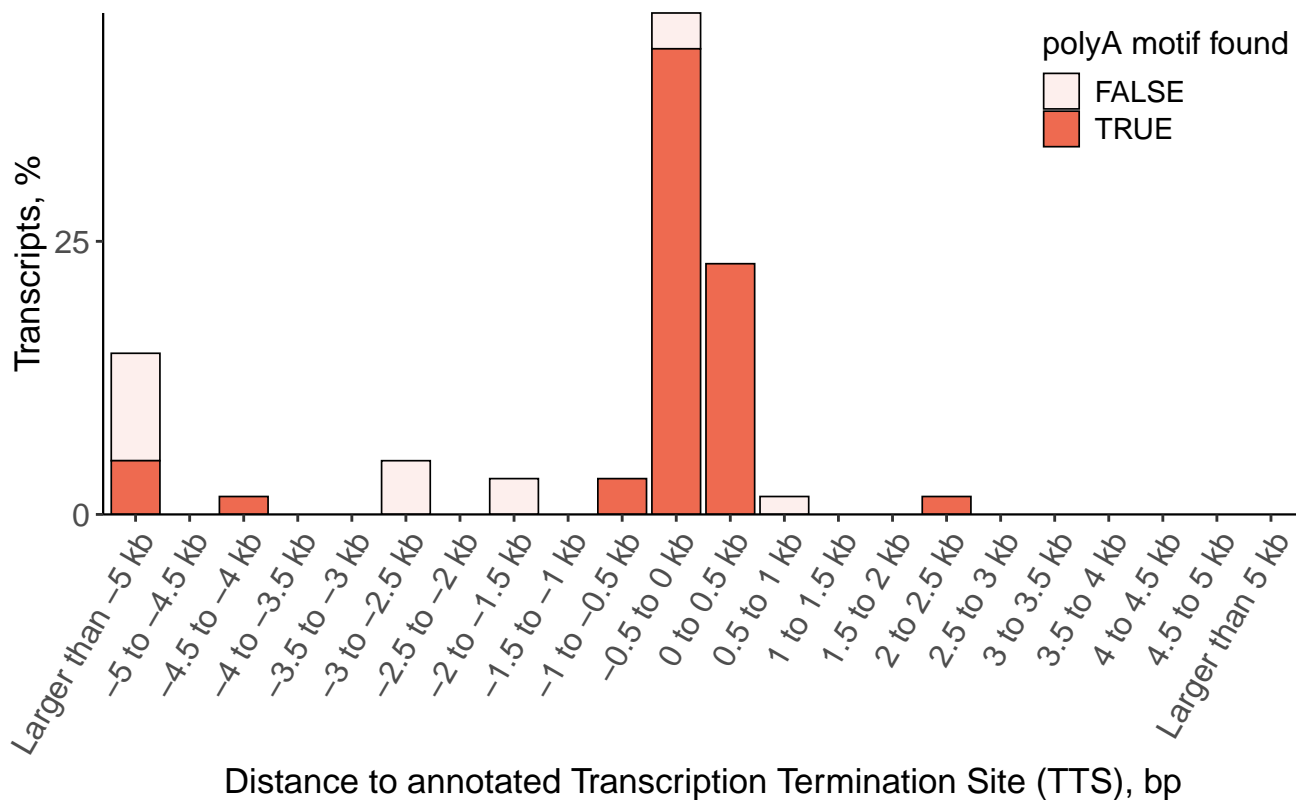
Negative values indicate upstream of annotated TTS





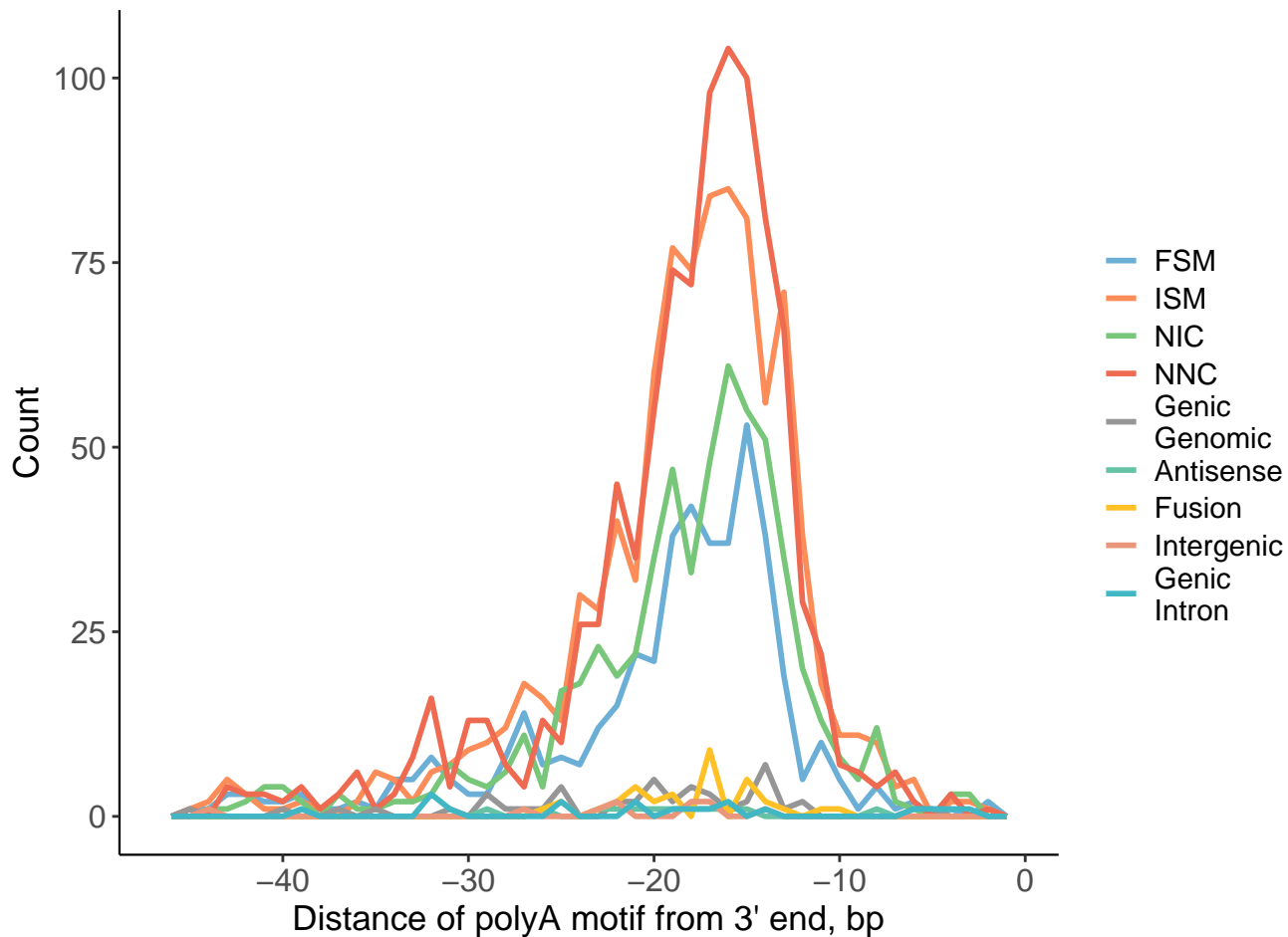
# Distance to Annotated Transcription Termination Site for ISM Intron Retention

Negative values indicate upstream of annotated TTS



## *PolyA Distance Analysis*

### Distance of Detected PolyA Motif From 3' end



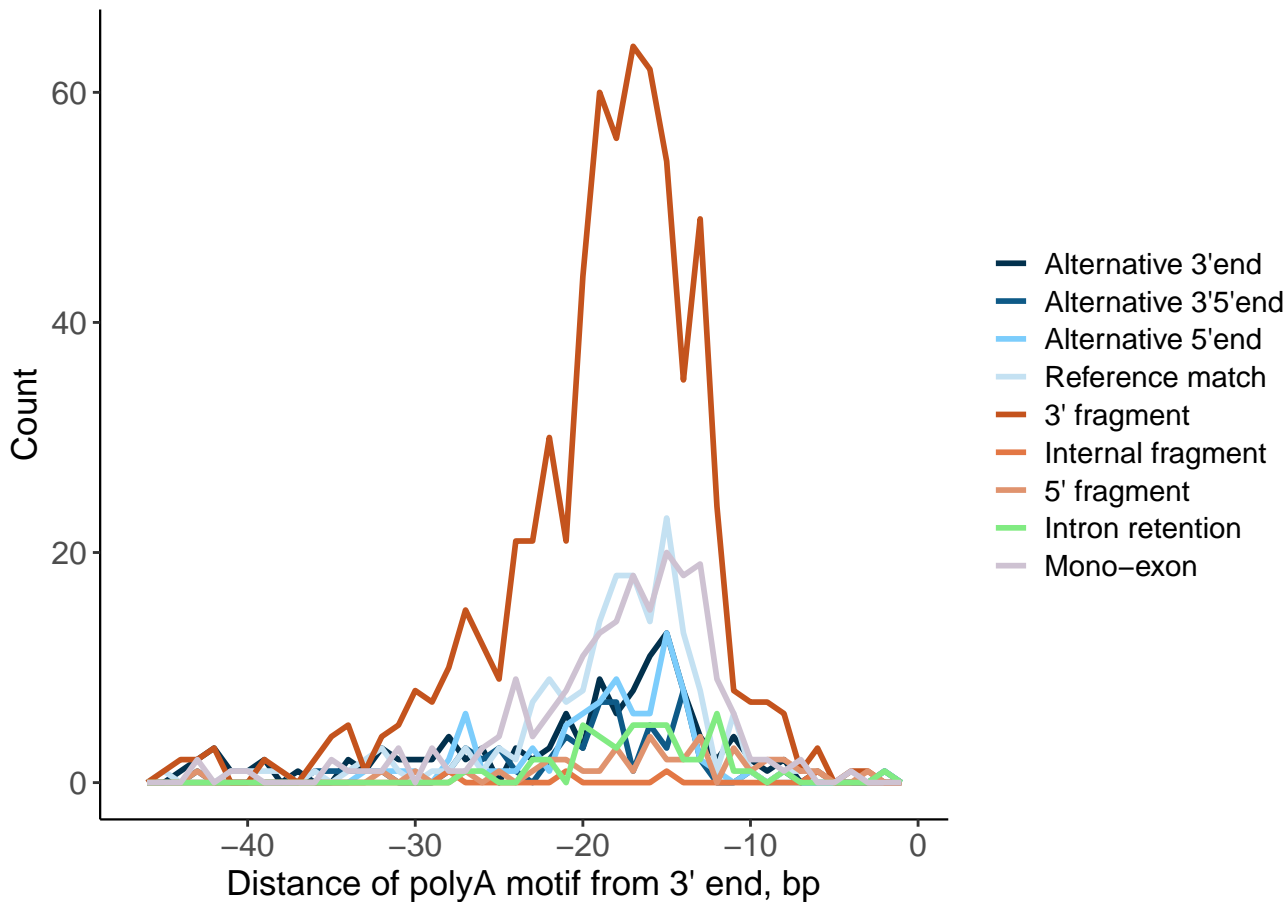
## Frequency of PolyA Motifs

### Number of polyA Motifs Detected

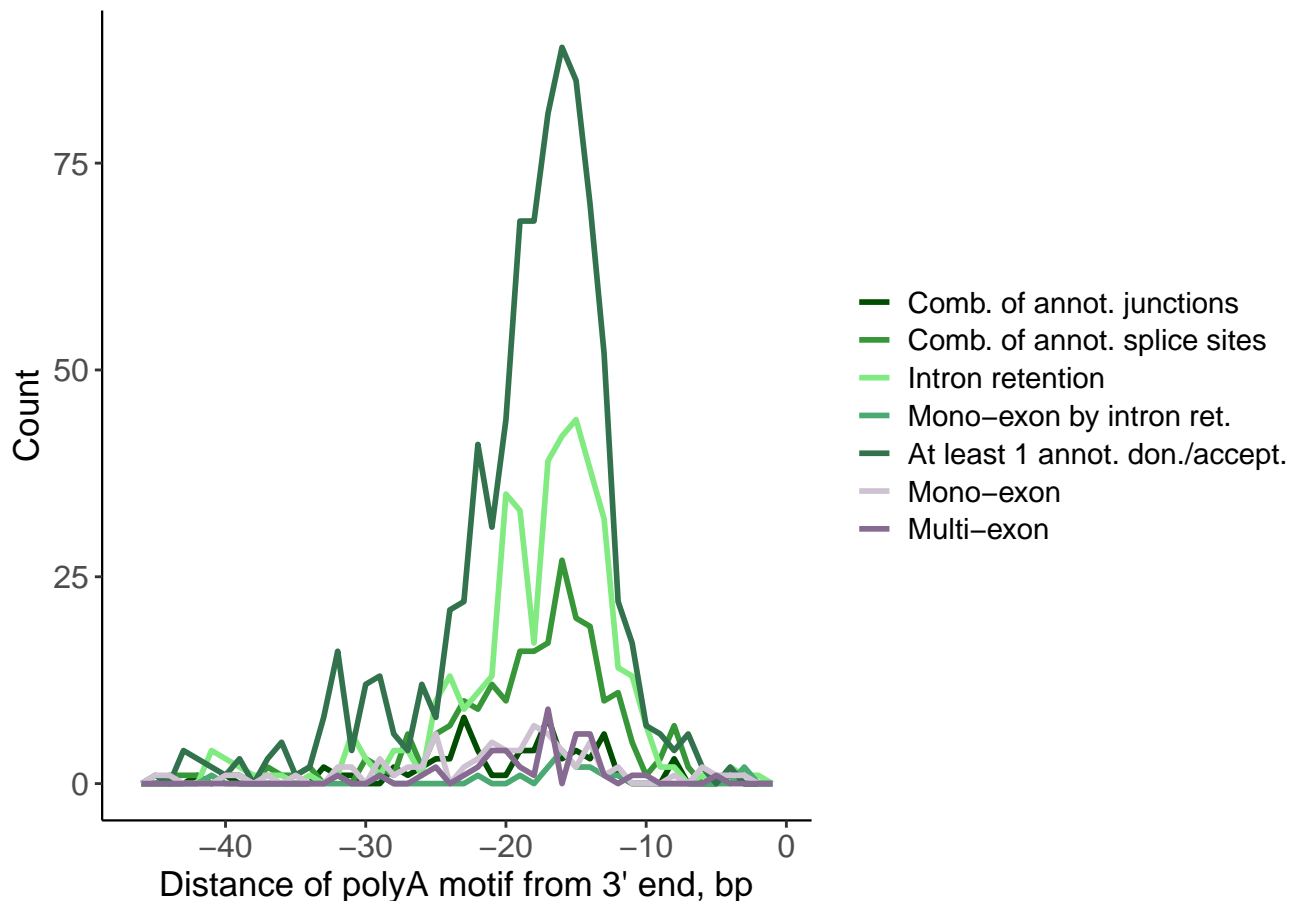
Category	Count	polyA Detected	%
FSM	539	458	85
ISM	1138	942	83
NIC	744	594	80
NNC	1139	981	86
Genic Genomic	113	48	42
Antisense	26	12	46
Fusion	38	34	89
Intergenic	41	10	24
Genic Intron	147	19	13

Motif	Count	%
AATAAA	2014	65.0
ATTAAA	529	17.1
AGTAAA	161	5.2
TATAAA	90	2.9
GATAAA	54	1.7
AAAACA	41	1.3
AATACA	38	1.2
AAAAAG	32	1.0
AATAGA	28	0.9
CATAAA	26	0.8
AAGAAA	24	0.8
AATATA	20	0.6
TTTAAA	16	0.5
AATGAA	13	0.4
ACTAAA	10	0.3
GGGGCT	2	0.1

### Distance of Detected PolyA Motif From 3'End by FSM and ISM Subcategories



Distance of Detected PolyA Motif From 3'End  
by Non-FSM/ISM Subcategories



## *Number of polyA Motifs Detected*

<b>Subcategory</b>	<b>Count</b>	<b>polyA Detected</b>	<b>%</b>
Alternative 3'end	152	121	80
Alternative 3'5'end	76	60	79
Alternative 5'end	90	84	93
Reference match	188	175	93
3' fragment	725	668	92
Internal fragment	8	4	50
5' fragment	141	39	28
Comb. of annot. junctions	83	68	82
Comb. of annot. splice sites	281	235	84
Intron retention	554	459	83
Mono-exon by intron ret.	44	18	41
At least 1 annot. don./accept.	986	846	86
Mono-exon	543	277	51
Multi-exon	54	44	81

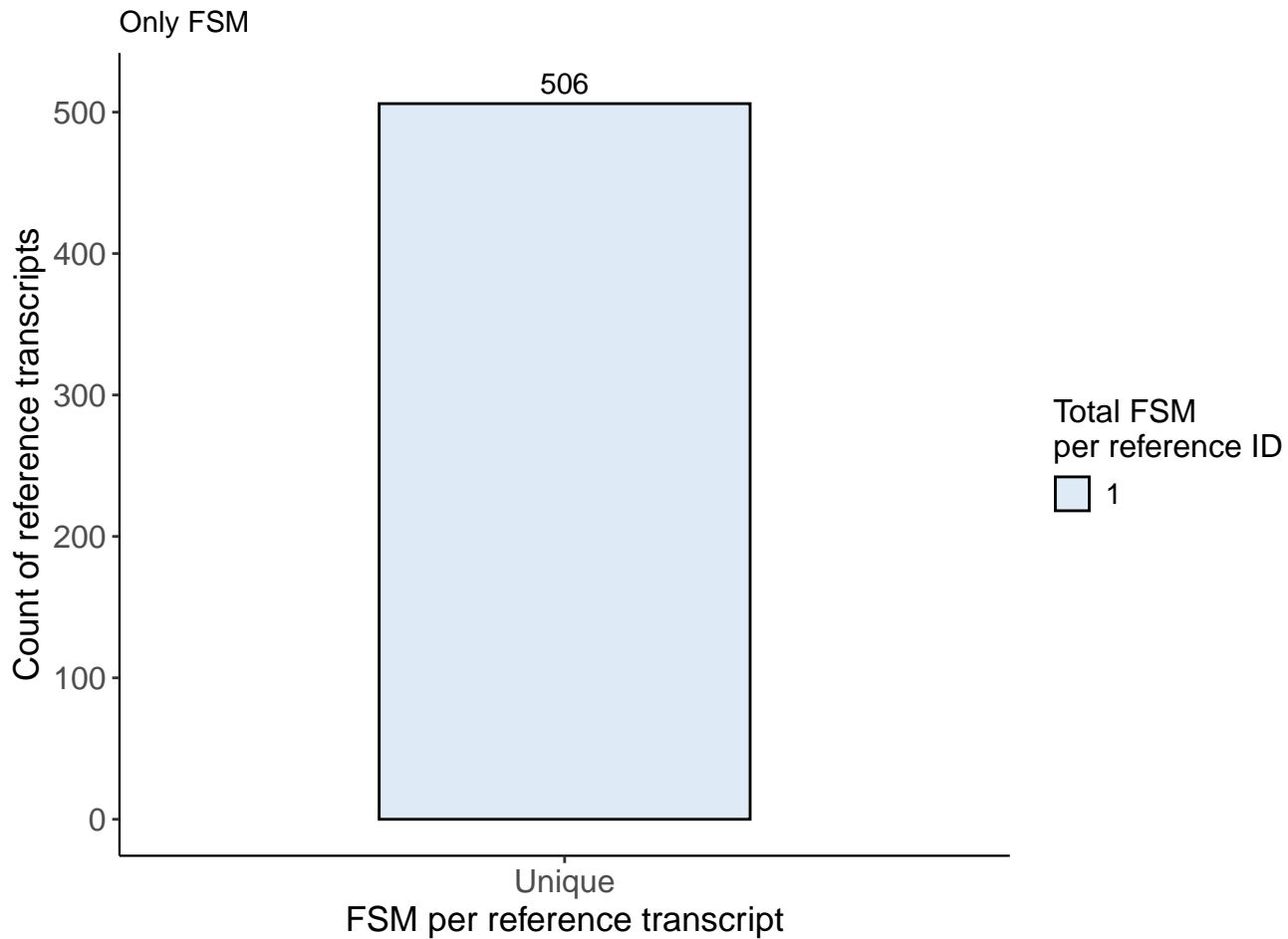
## *Frequency of PolyA Motifs*

<b>Motif</b>	<b>Count</b>	<b>%</b>
AATAAA	2014	65.0
ATTAAA	529	17.1
AGTAAA	161	5.2
TATAAA	90	2.9
GATAAA	54	1.7
AAAACA	41	1.3
AATACA	38	1.2
AAAAAG	32	1.0
AATAGA	28	0.9
CATAAA	26	0.8
AAGAAA	24	0.8
AATATA	20	0.6
TTTAAA	16	0.5
AATGAA	13	0.4
ACTAAA	10	0.3
GGGGCT	2	0.1

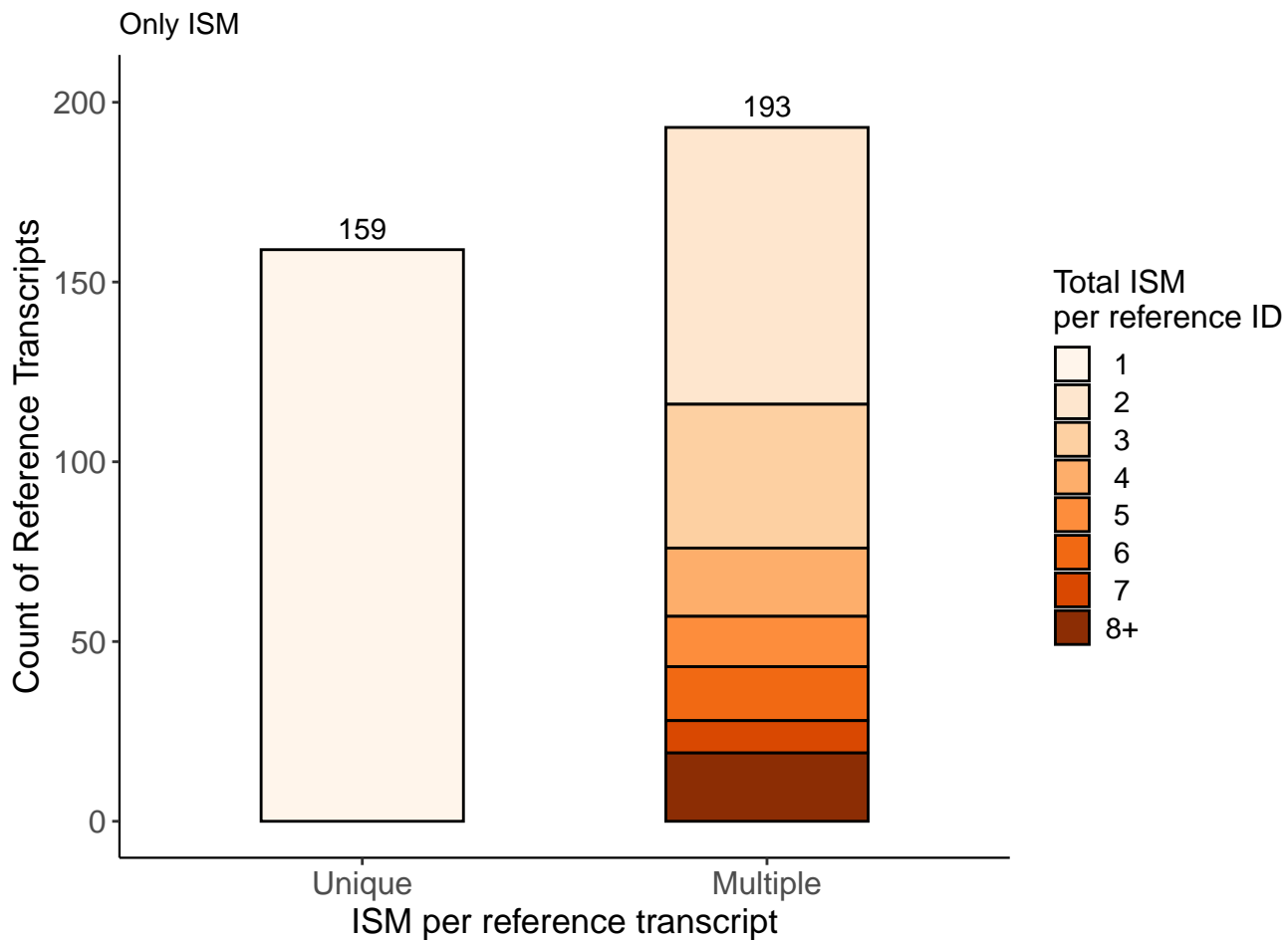


## *Redundancy Analysis*

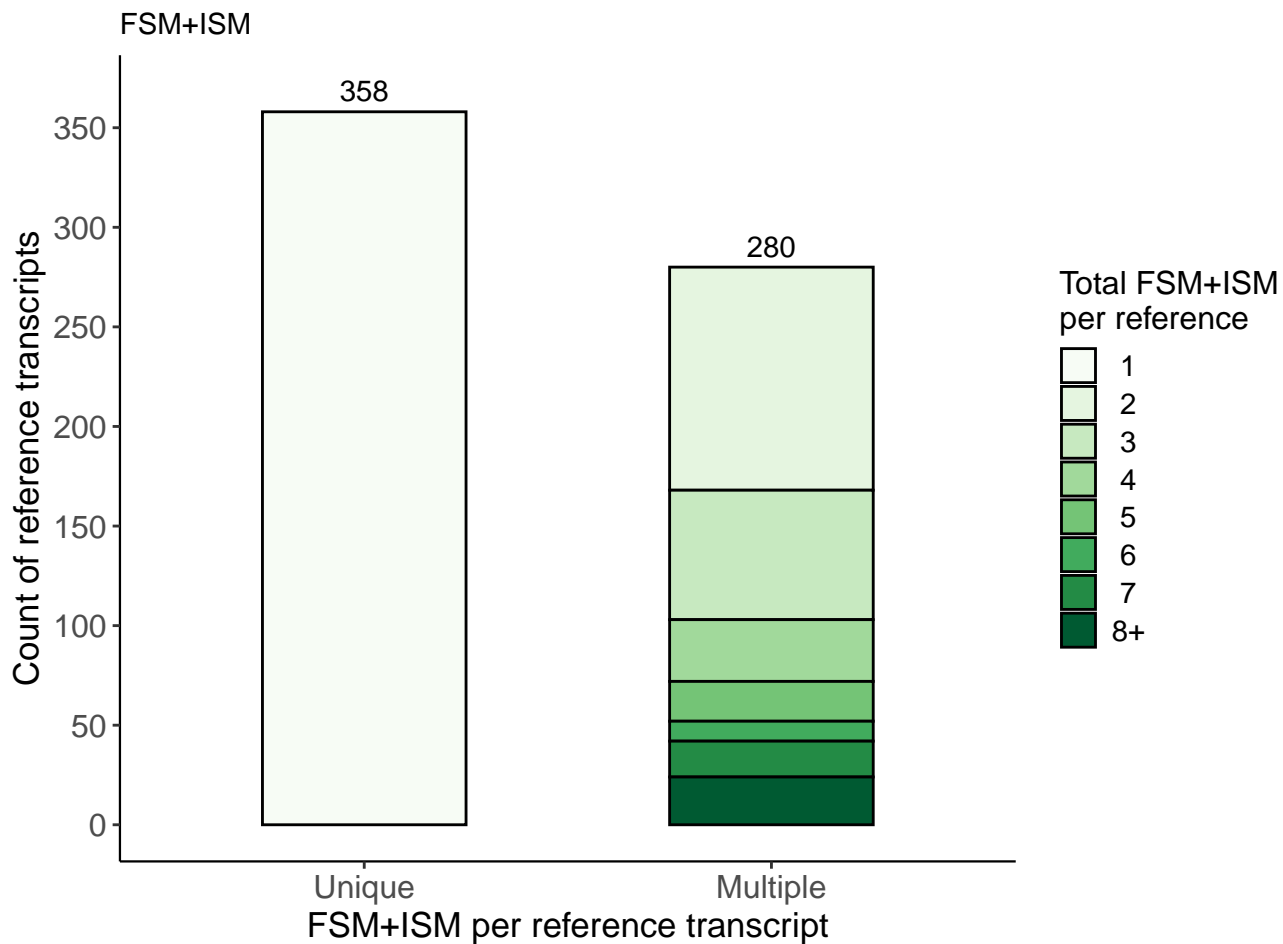
## Reference Transcript Redundancy



## Reference Transcript Redundancy

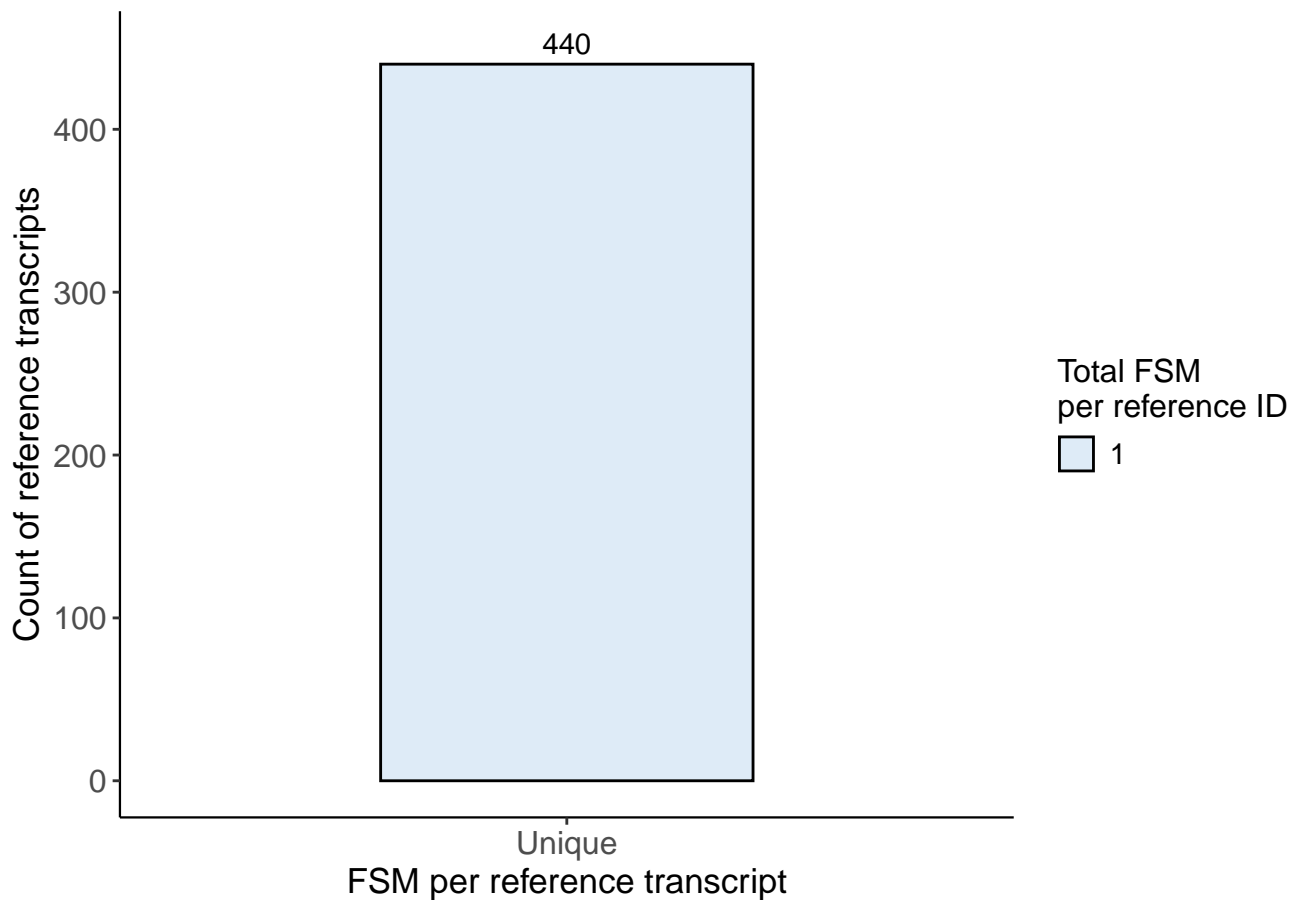


## Reference Transcript Redundancy



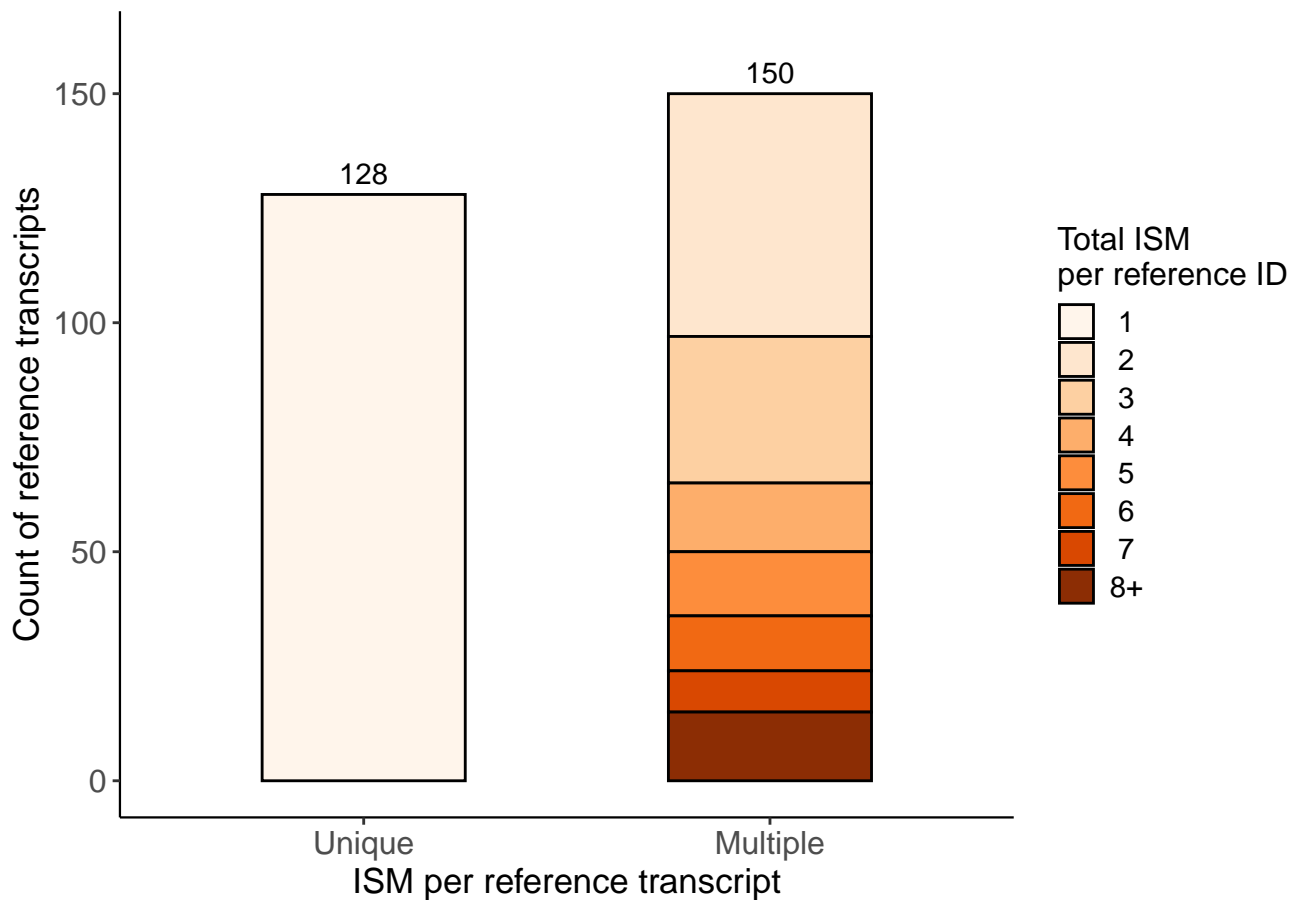
## Reference Transcript Redundancy

Only FSM with a polyA motif found



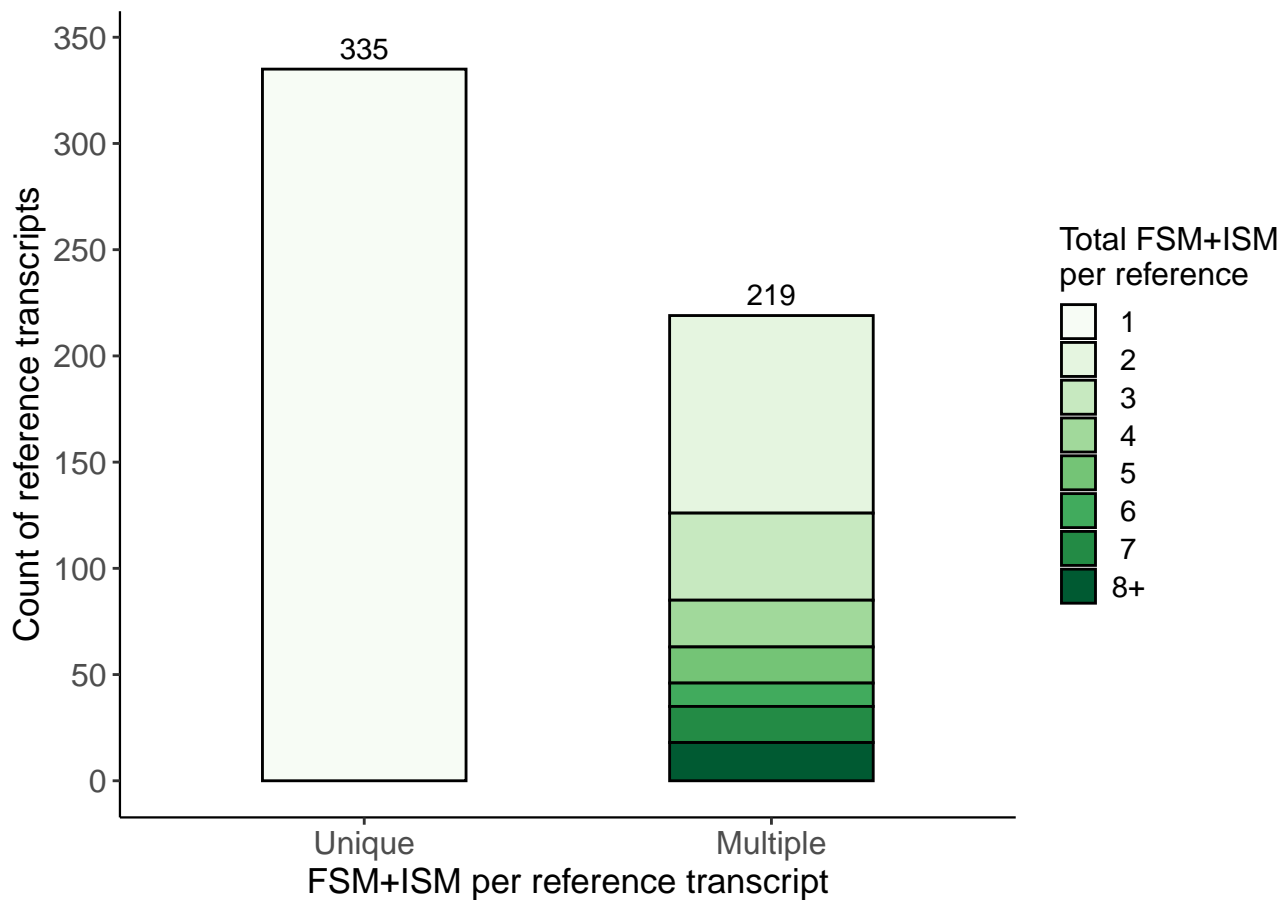
## Reference Transcript Redundancy

Only ISM with a polyA motif found



## Reference Transcript Redundancy

FSM+ISM with a polyA motif found

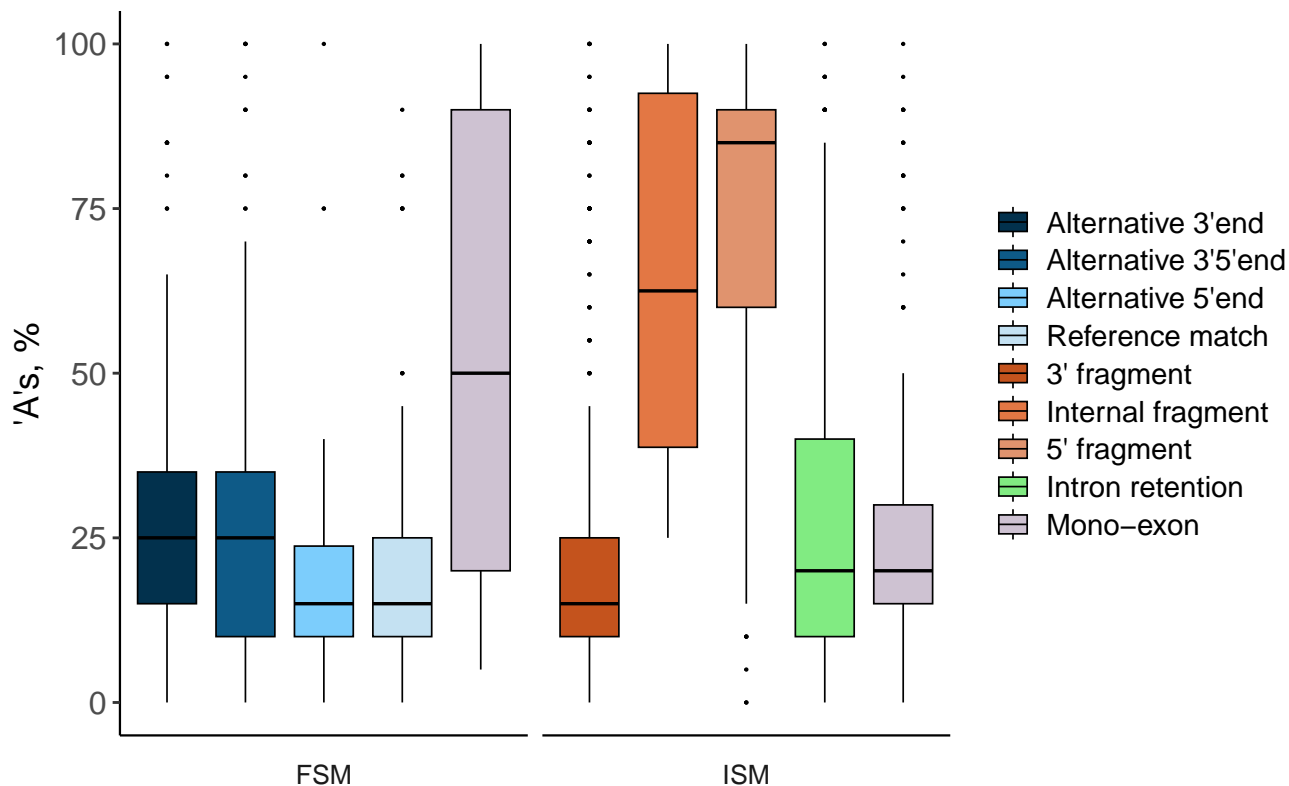


*Intra-Priming Quality Check*



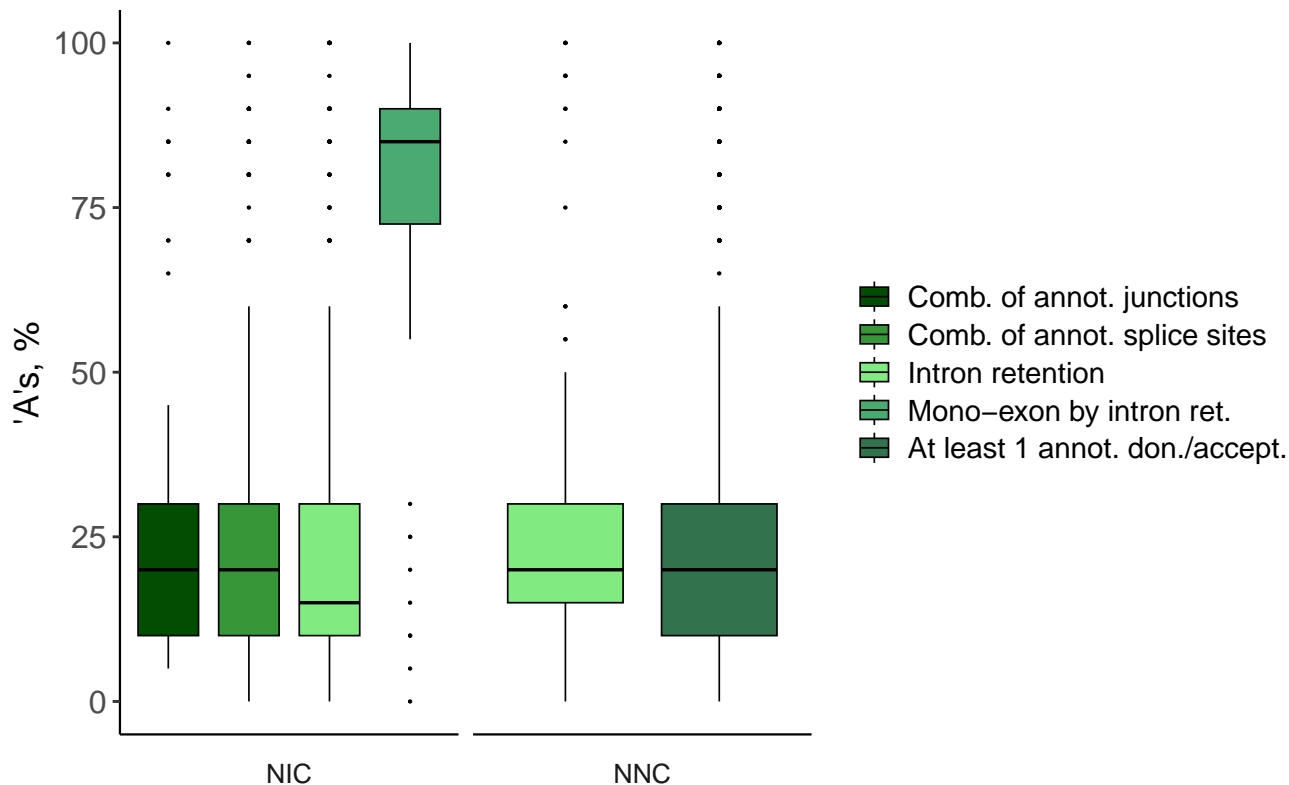
# Possible Intra-Priming by Structural Category

Percent of genomic 'A's in downstream 20 bp



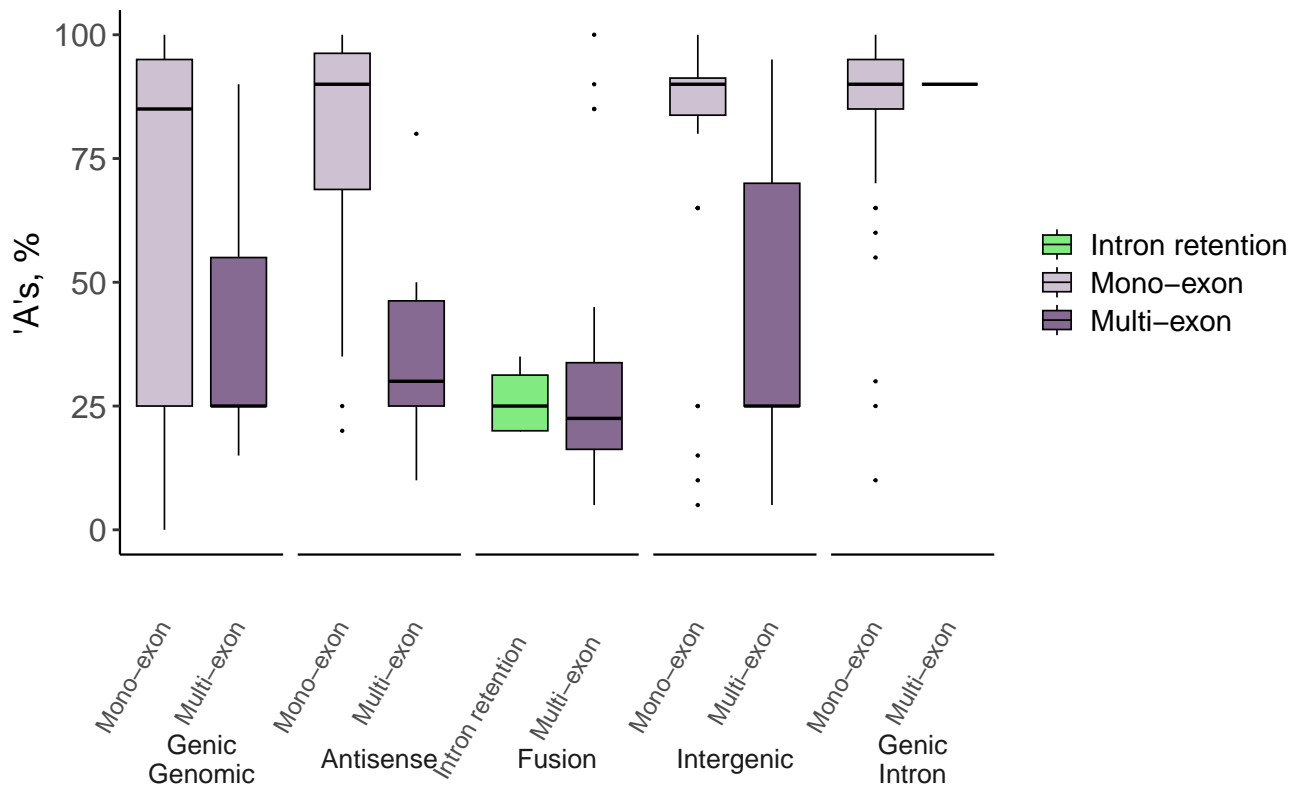
## Possible Intra-Priming by Structural Category

Percent of genomic 'A's in downstream 20 bp



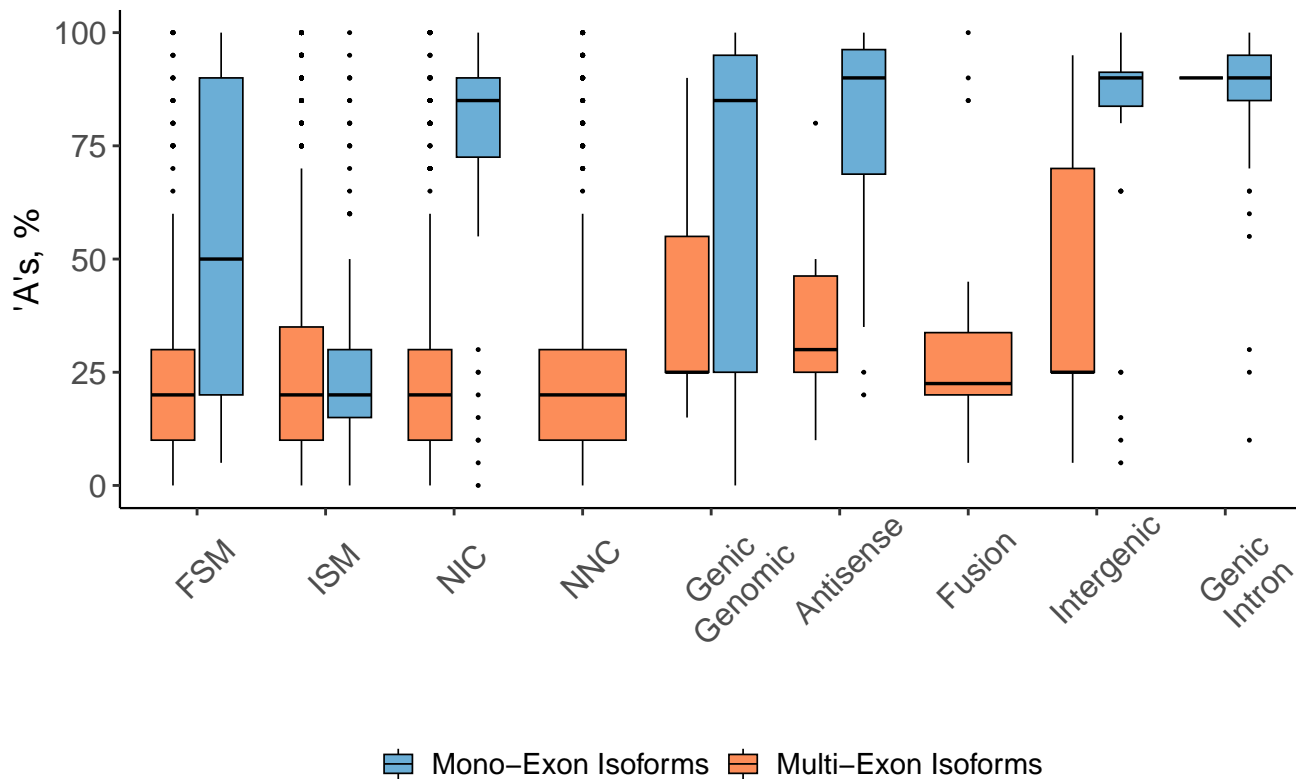
# Possible Intra-Priming by Structural Category

Percent of genomic 'A's in downstream 20 bp



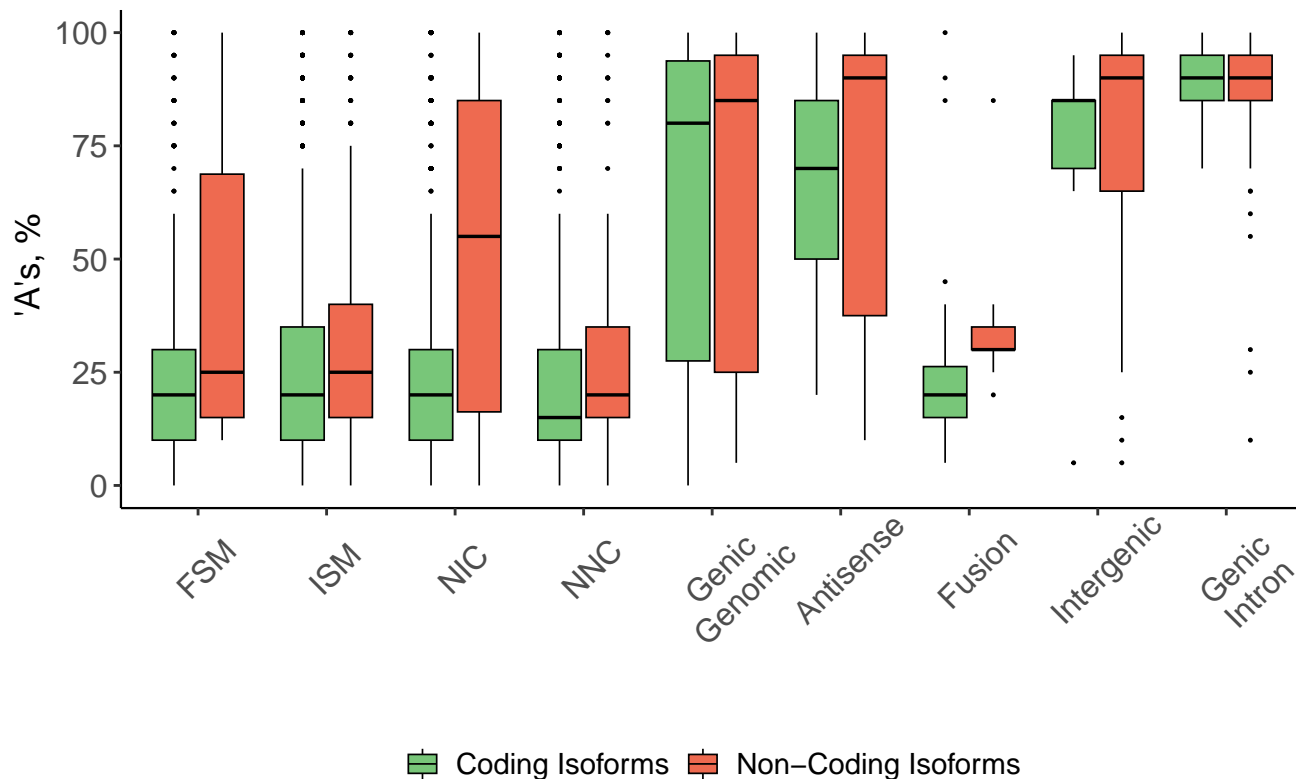
## Mono- vs Multi-Exon Possible Intra-Priming

Percent of genomic 'A's in downstream 20 bp



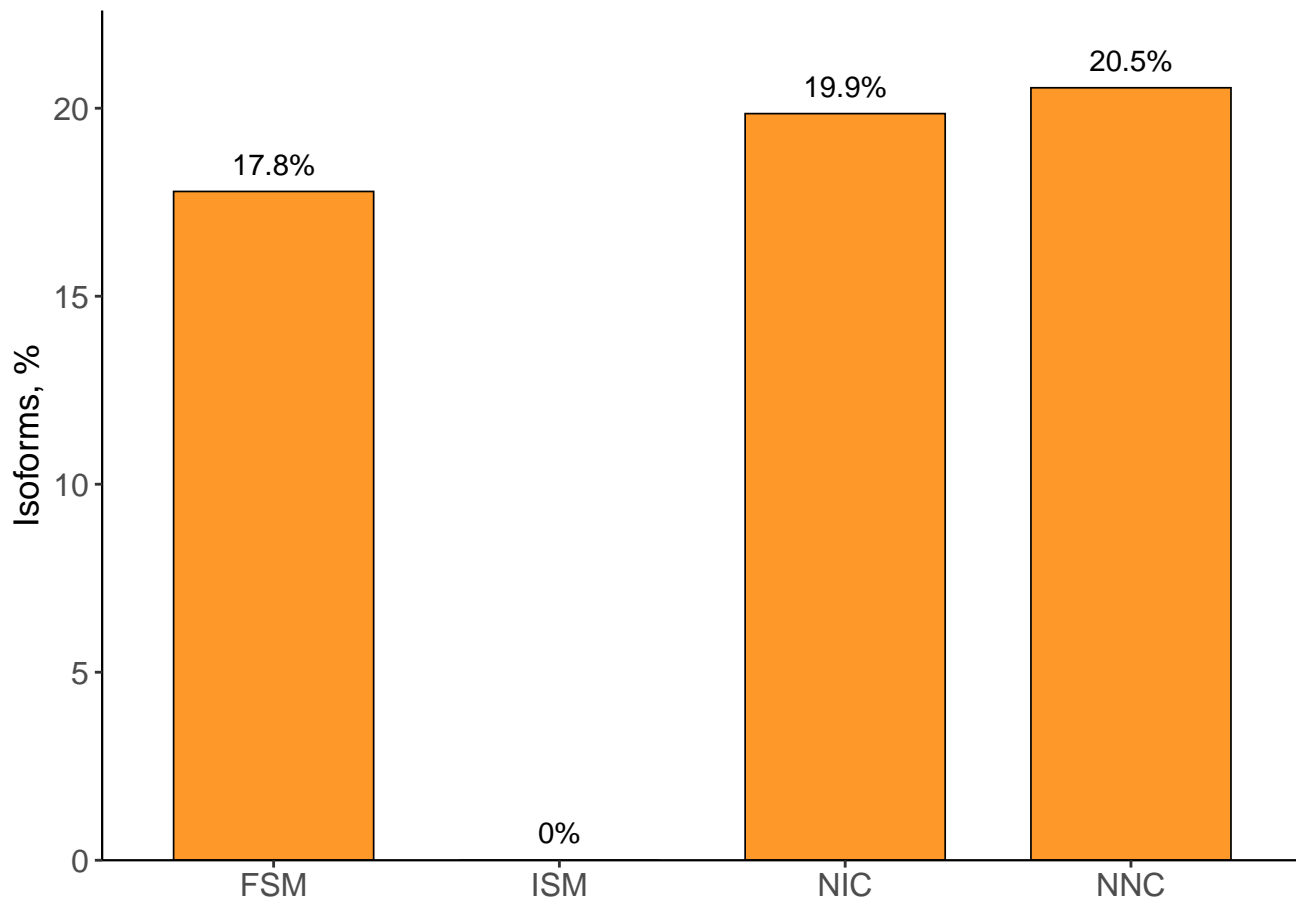
# Coding vs Non-Coding Possible Intra-Priming

Percent of genomic 'A's in downstream 20 bp

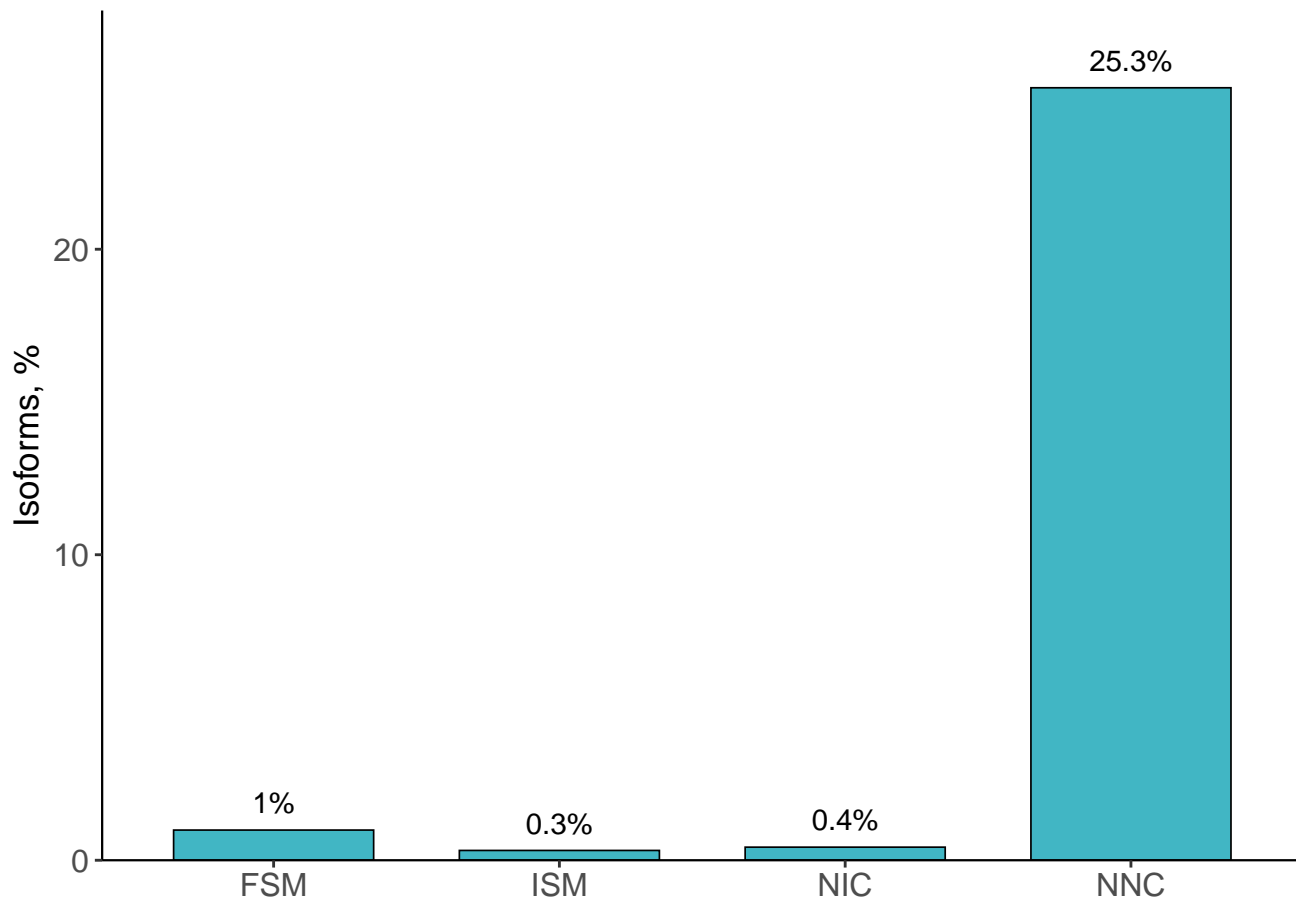


## *Features of Bad Quality*

## RT-switching

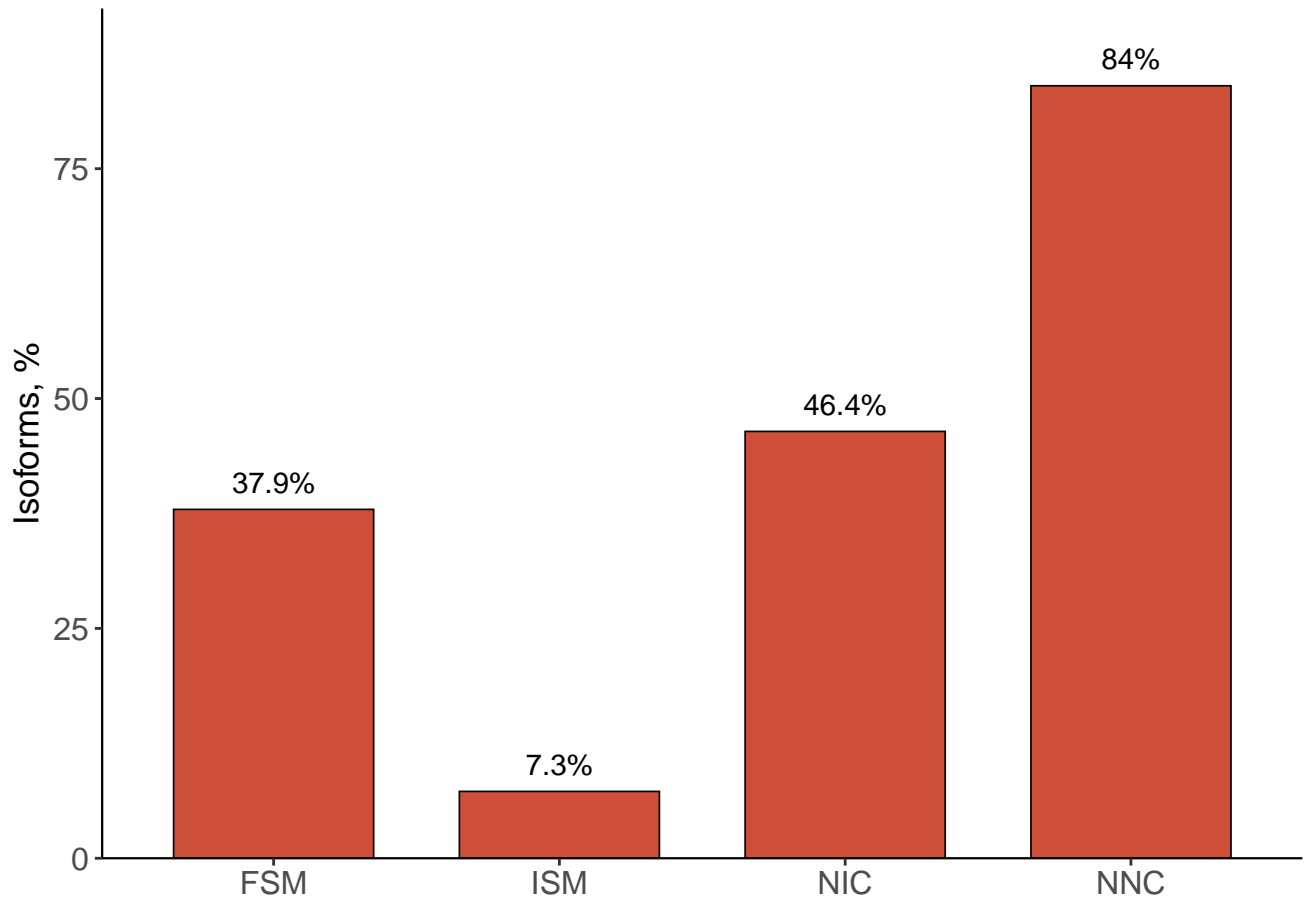


## Non-Canonical Junctions

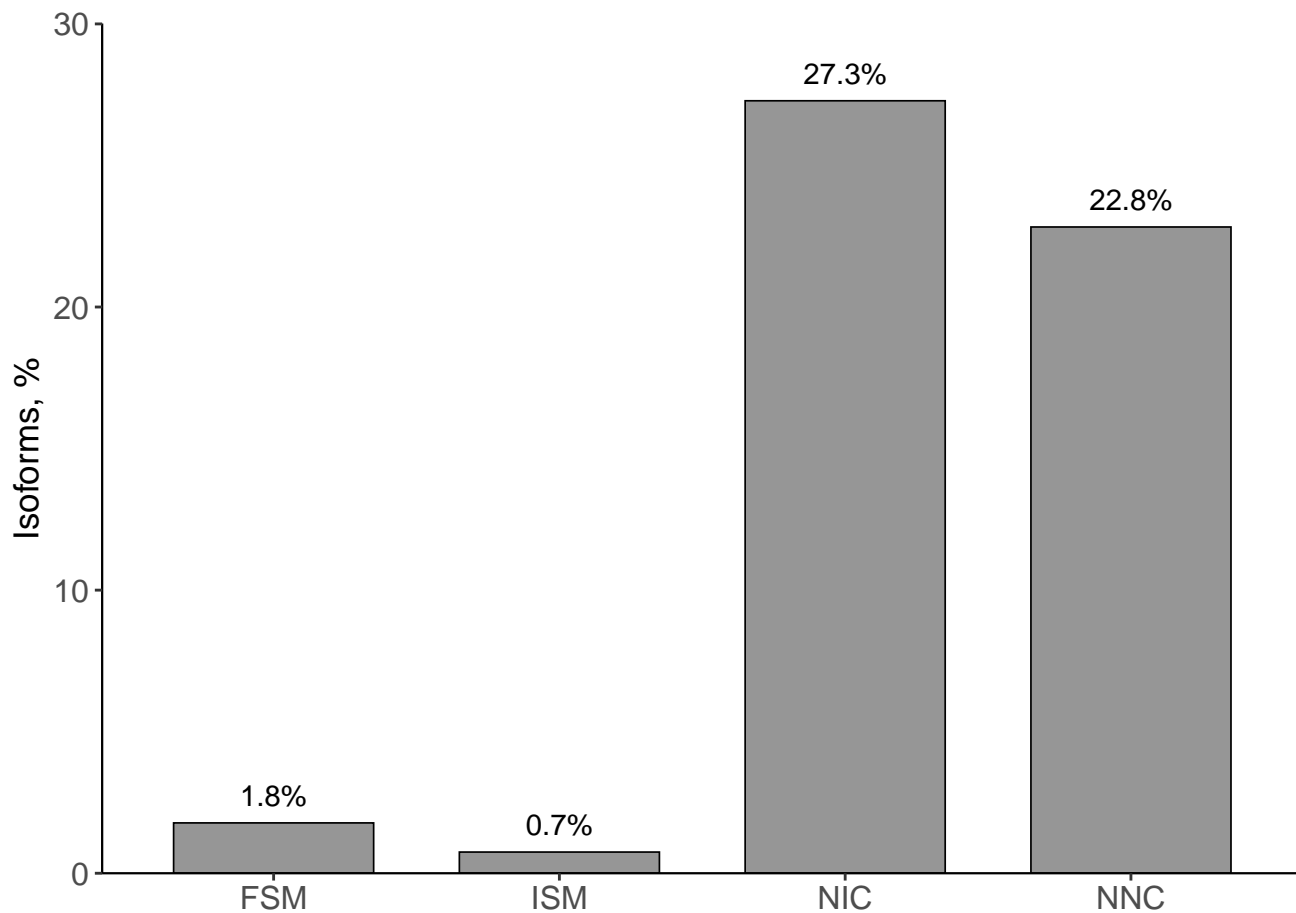




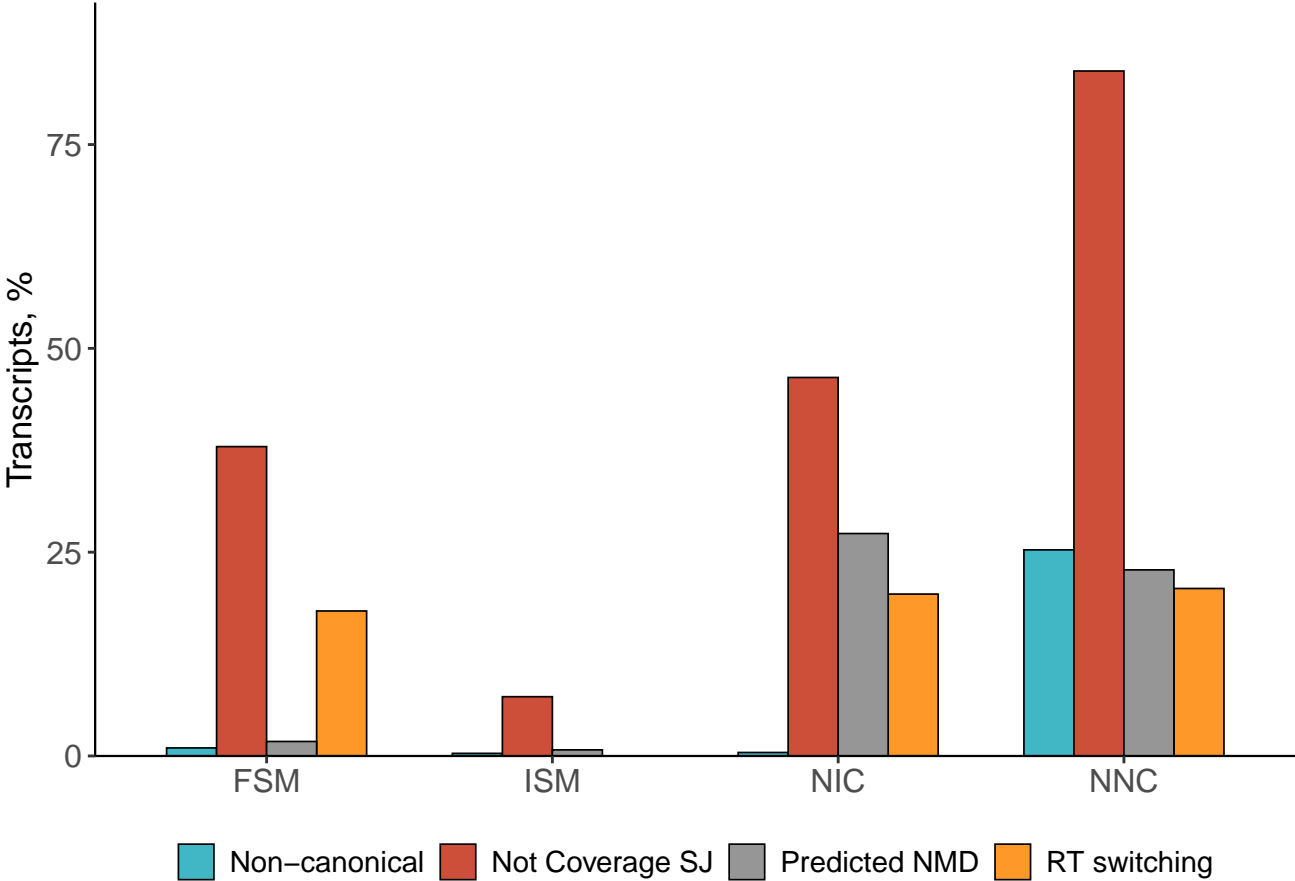
## Splice Junctions Without Short Read Coverage



## Nonsense-Mediated Decay by Structural Category

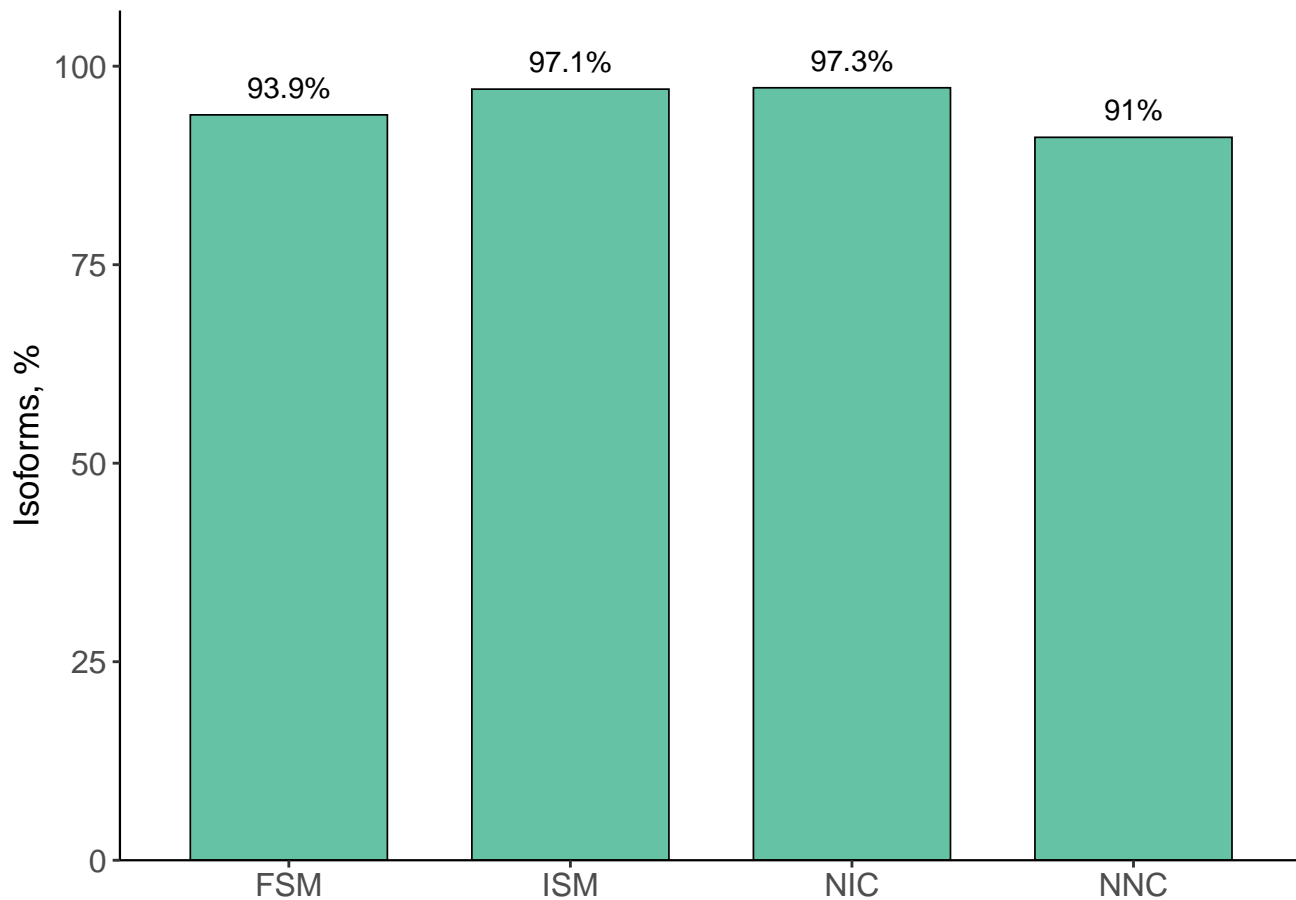


# Quality Control Attributes Across Structural Categories

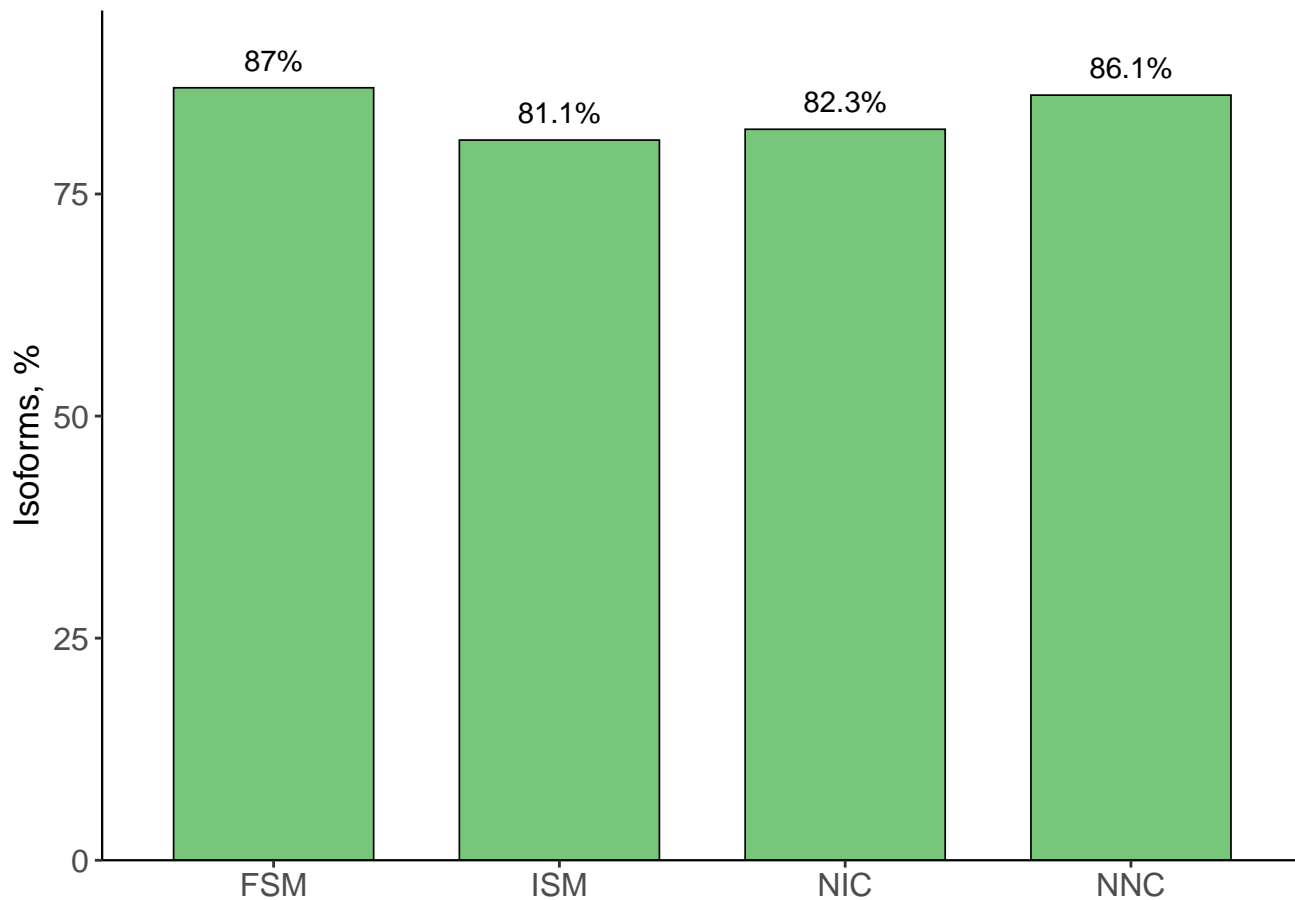


## *Features of Good Quality*

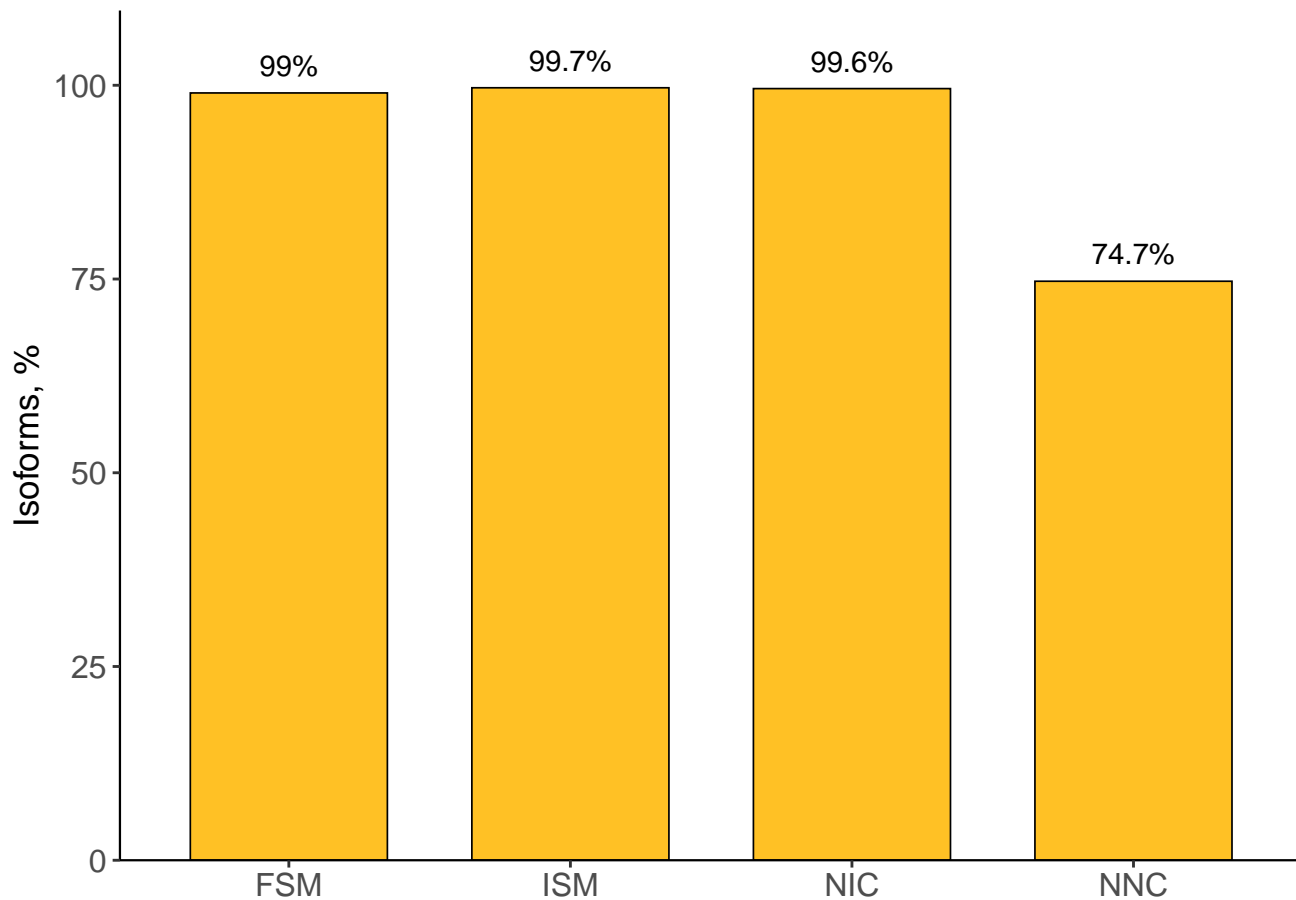
## Annotation Support



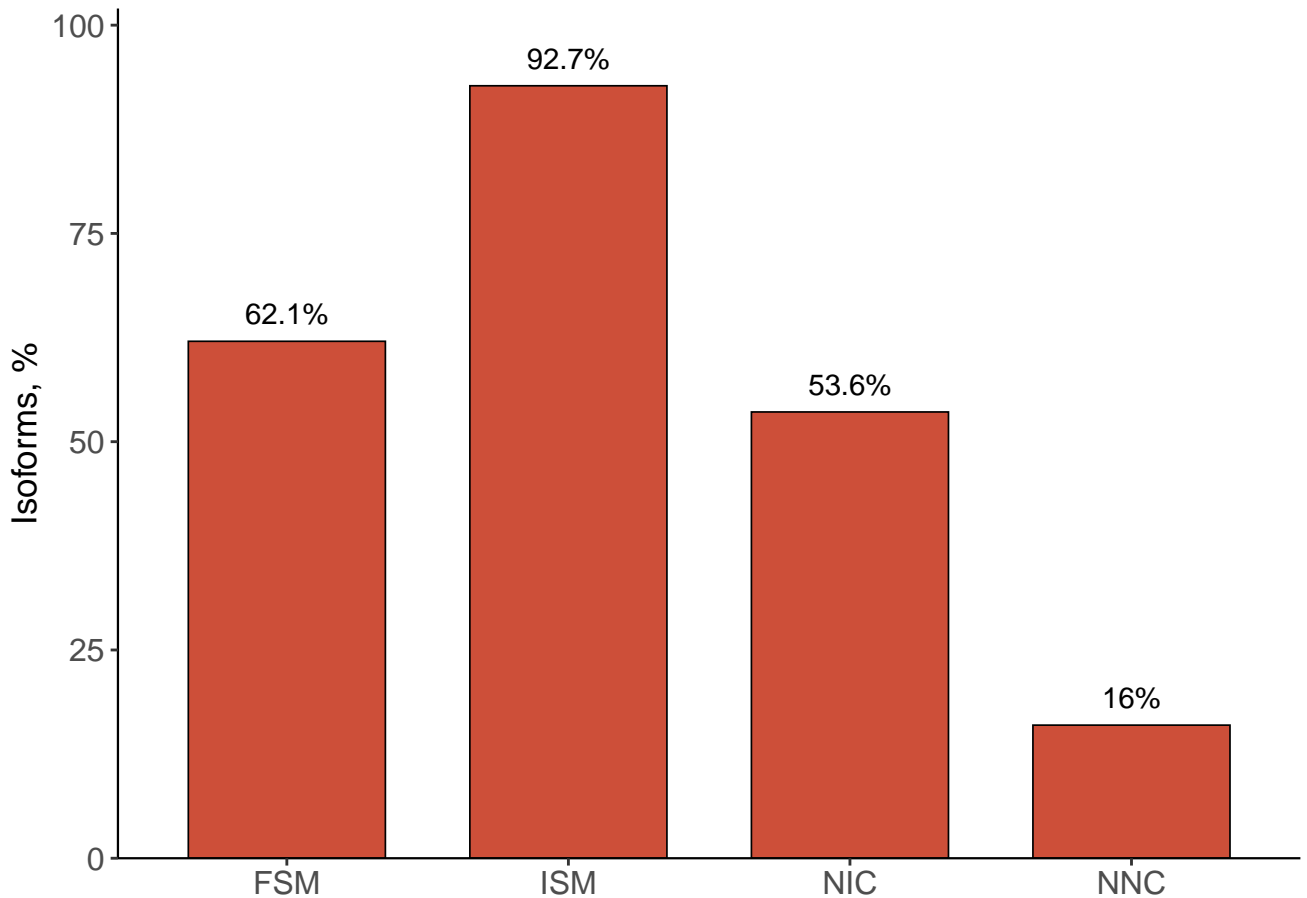
## PolyA Support



## All Canonical Junctions

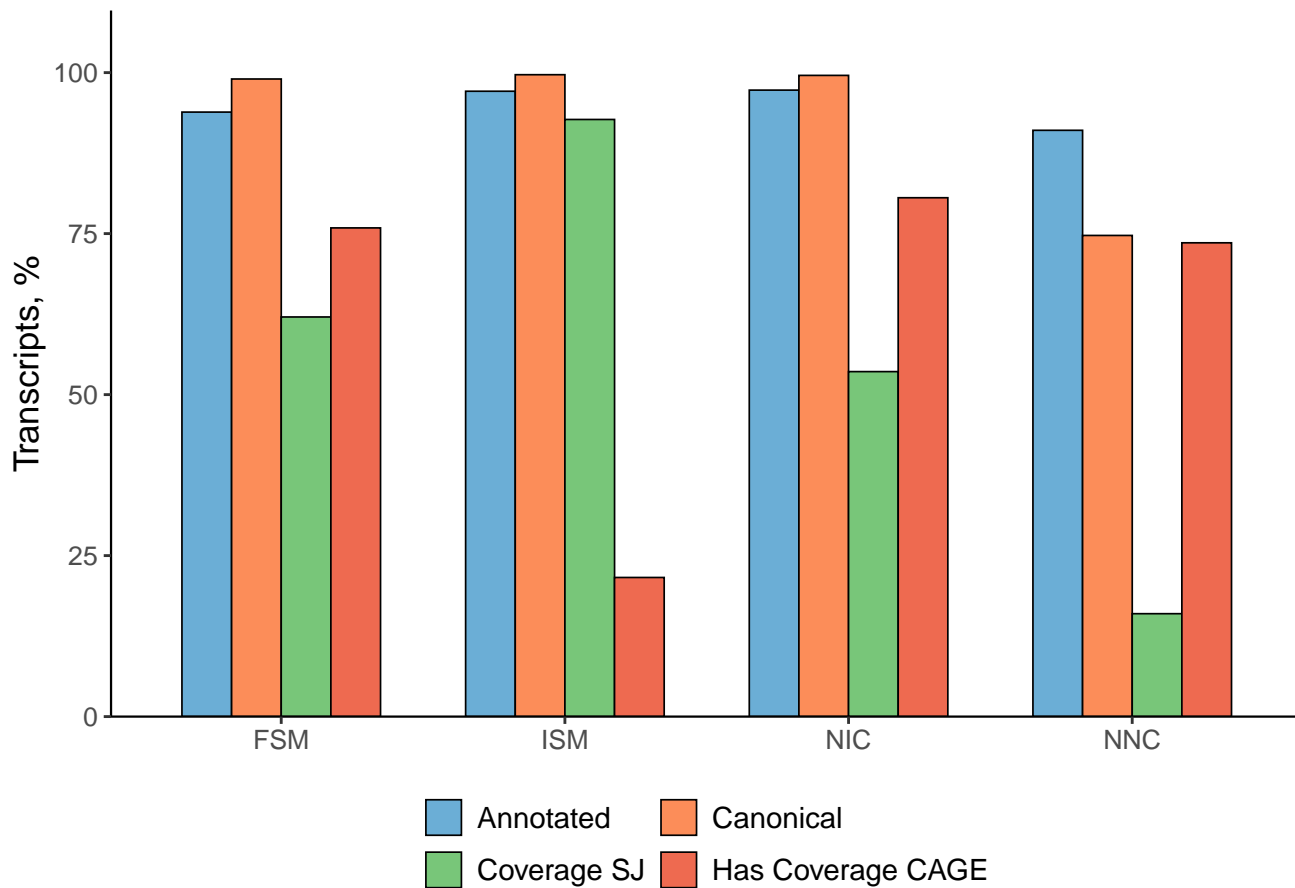


## Splice Junctions With Short Read Coverage





## Good Quality Control Attributes Across Structural Categories



# TSS Ratio FSM Reference Match vs ISM

