

# *SQANTI3 report*

*Unique Genes: 23820*  
*Unique Isoforms: 555113*

### *Transcript Classification*

<b>Category</b>	<b>Isoforms, count</b>
FSM	138528
ISM	204661
NIC	114899
NNC	88389
Genic Genomic	704
Antisense	2035
Fusion	2792
Intergenic	3041
Genic Intron	0

### *Gene Classification*

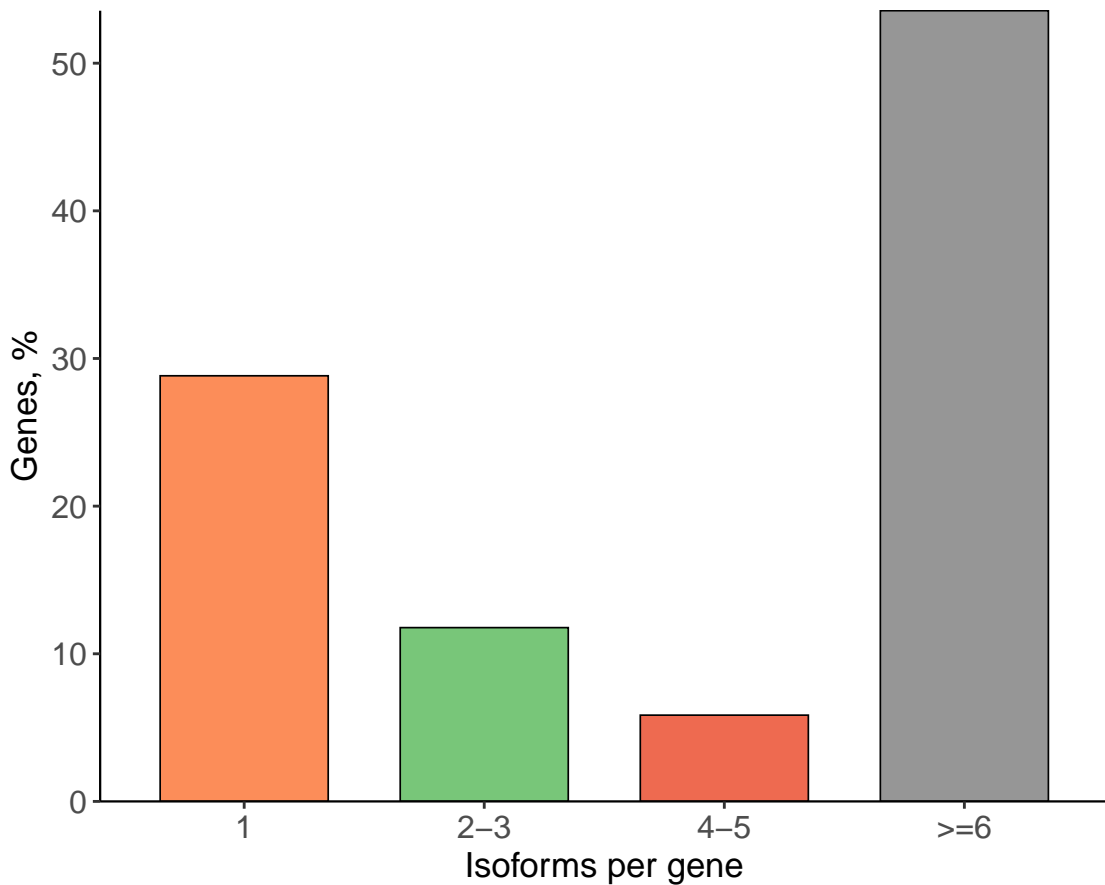
<b>Category</b>	<b>Genes, count</b>
Annotated Genes	20052
Novel Genes	3768

## *Splice Junction Classification*

<b>Category</b>	<b>SJs, count</b>	<b>Percent</b>
Known canonical	184080	66.84
Known Non-canonical	86	0.03
Novel canonical	91232	33.13
Novel Non-canonical	0	0.00

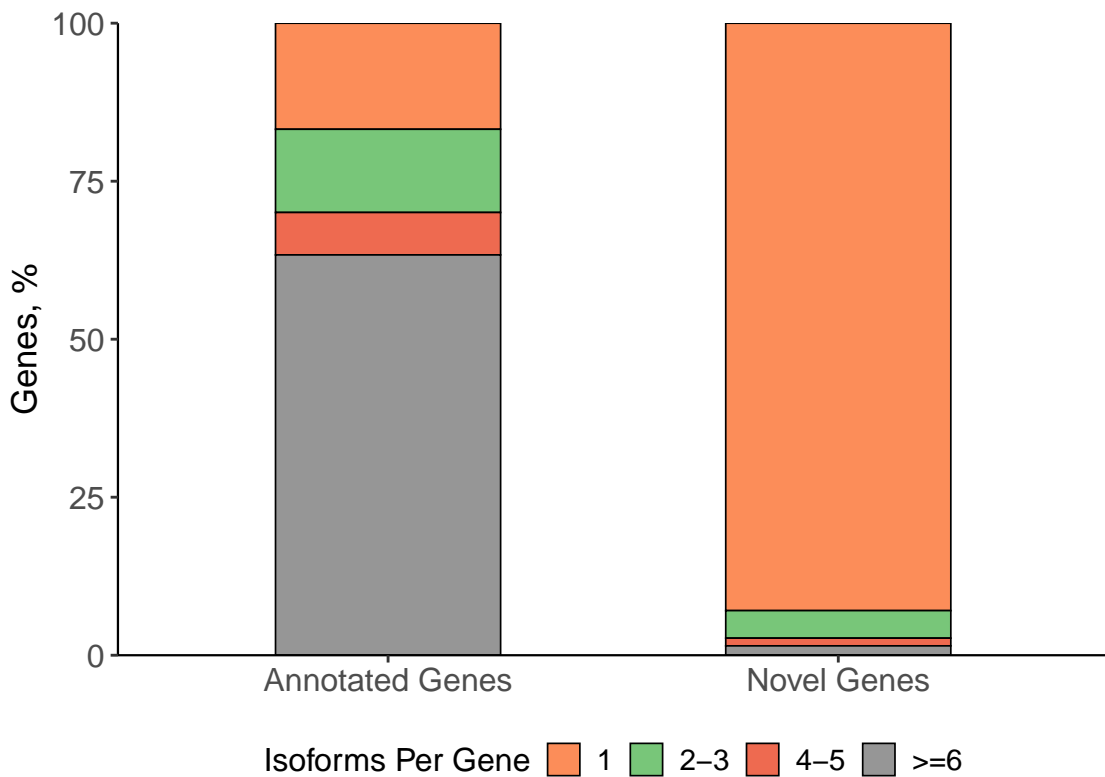
## *Gene Characterization*

## Number of Isoforms per Gene

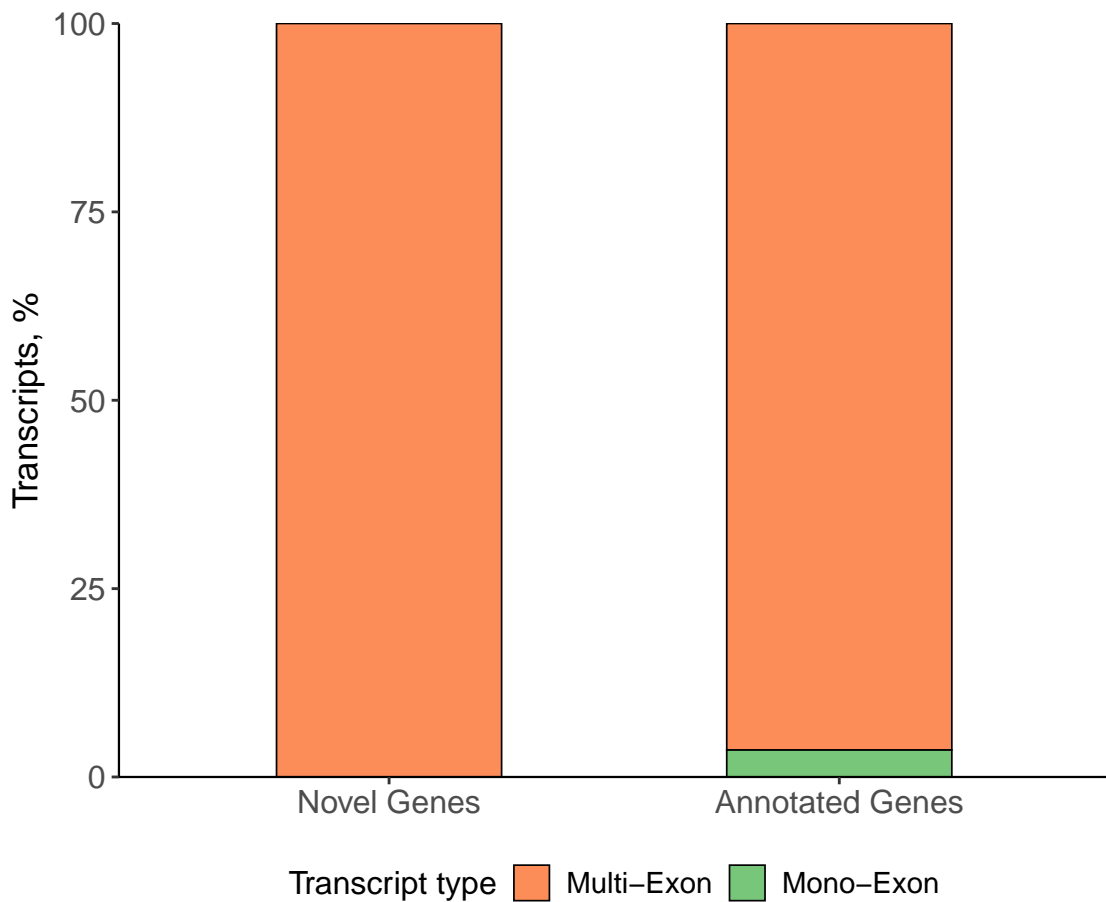


# Number of Isoforms per Gene

Known vs Novel Genes



## Distribution of Mono- vs Multi-Exon Transcripts

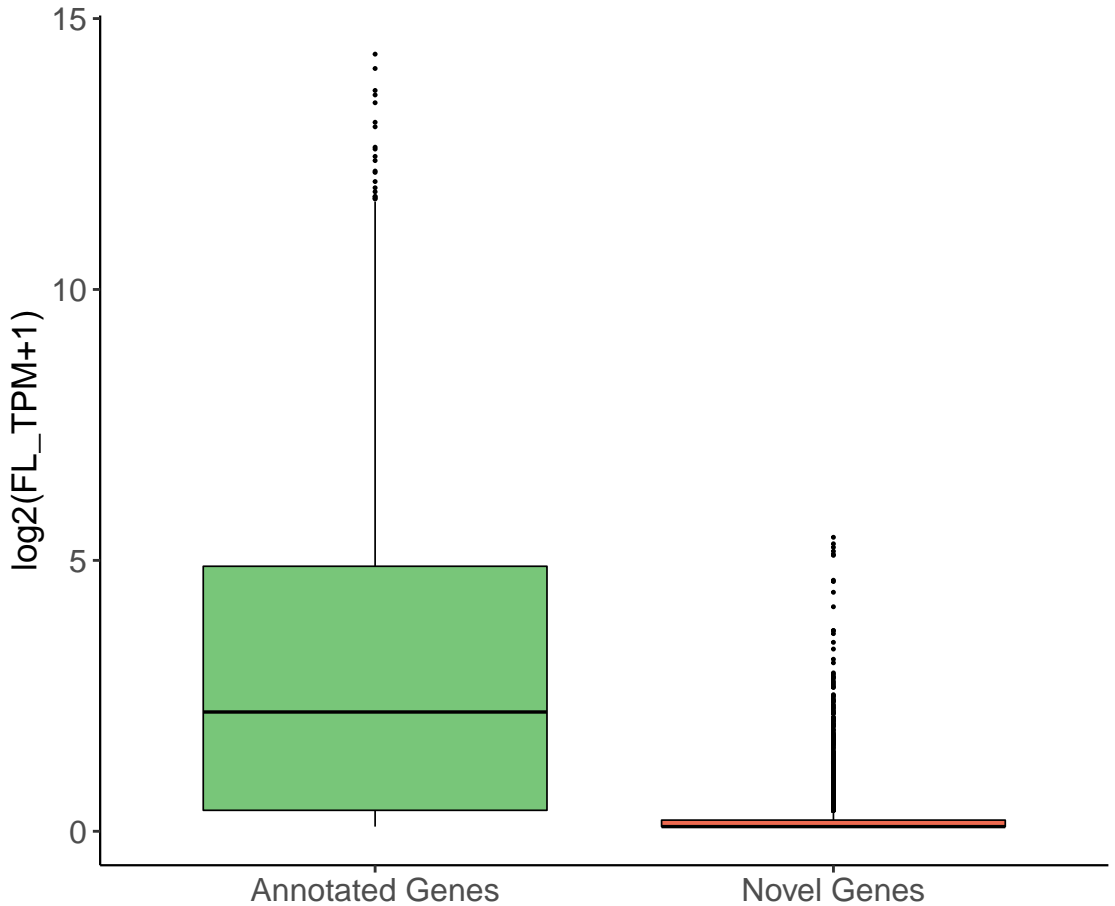




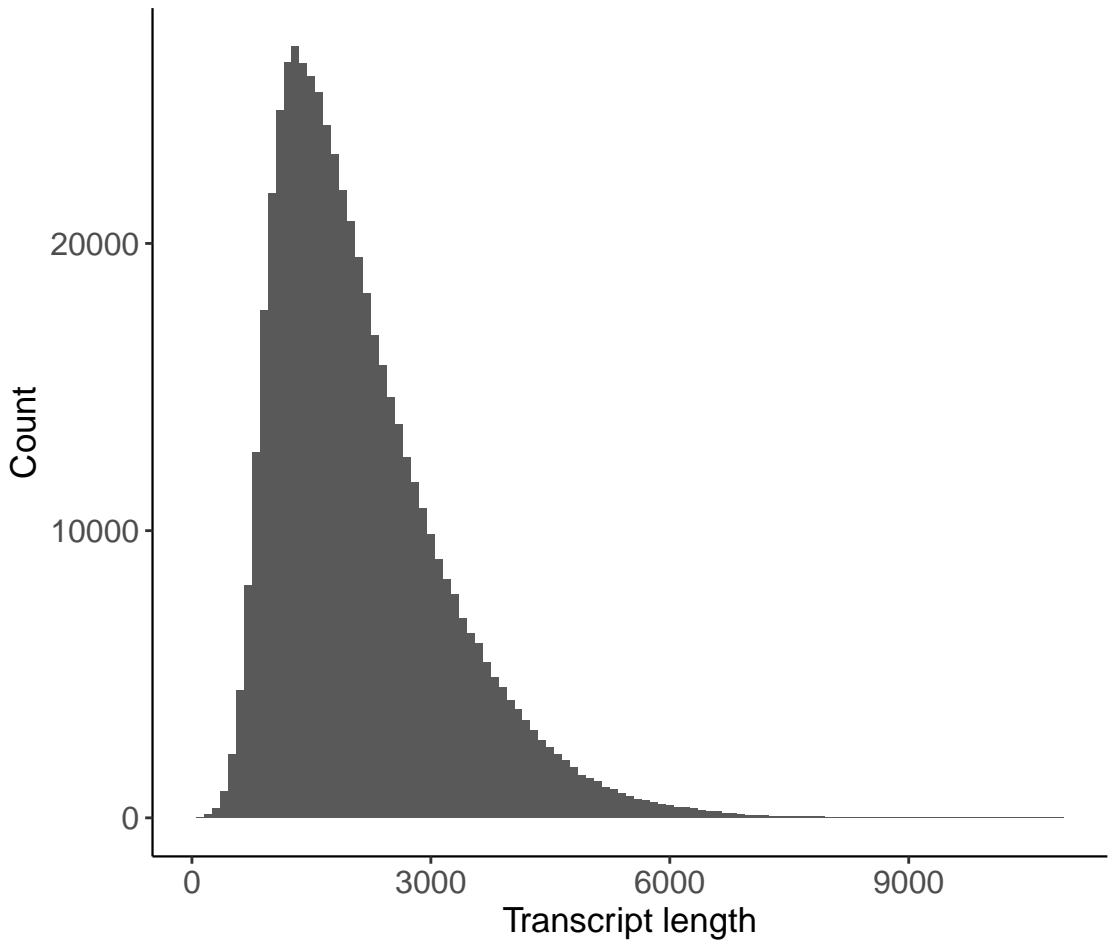




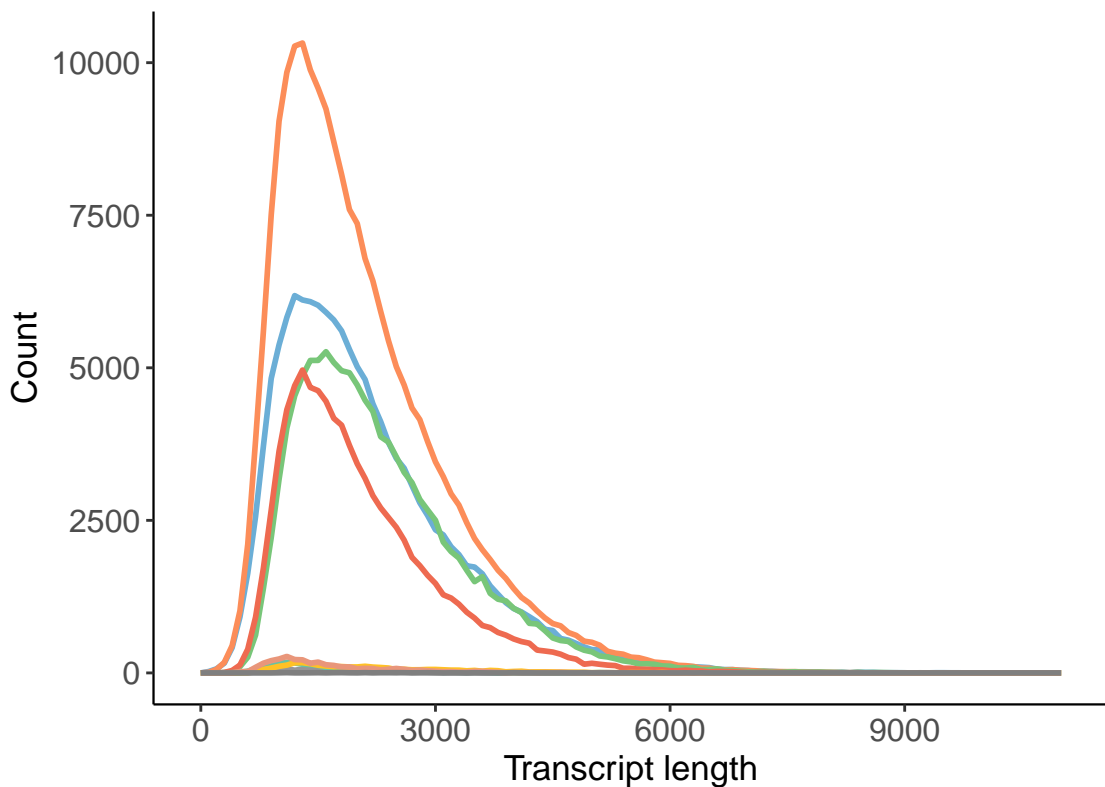
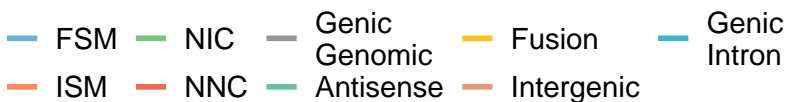
Number of FL reads per Gene by Type of Gene Annotation



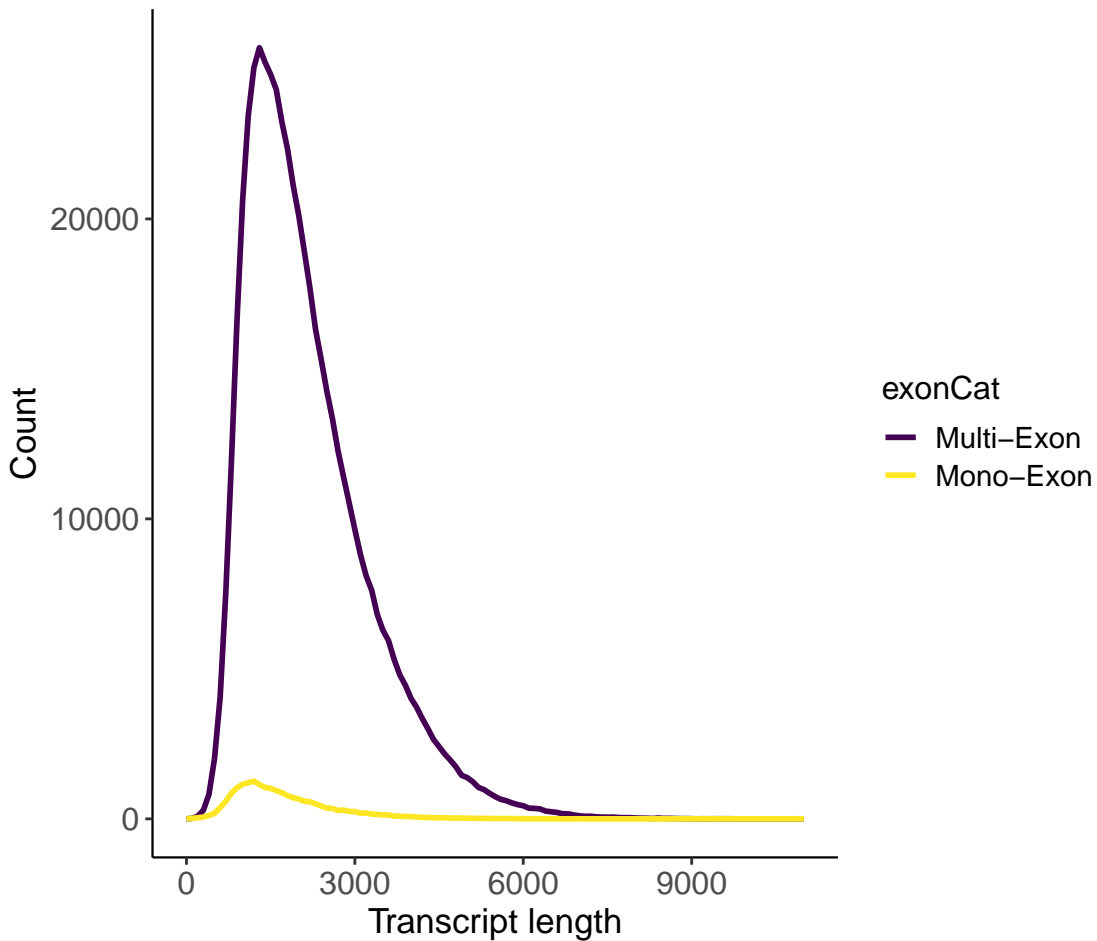
# All Transcript Lengths Distribution



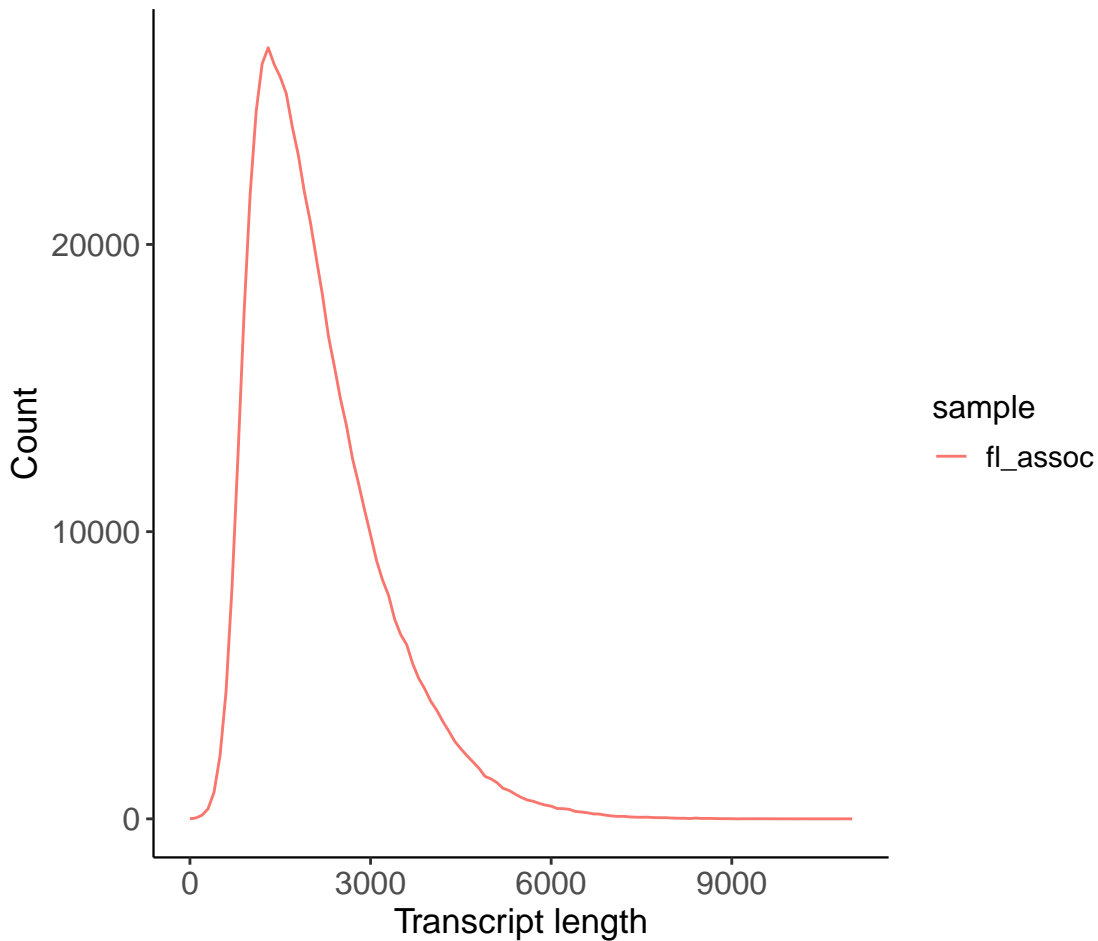
# Transcript Lengths Distribution by Structural Category



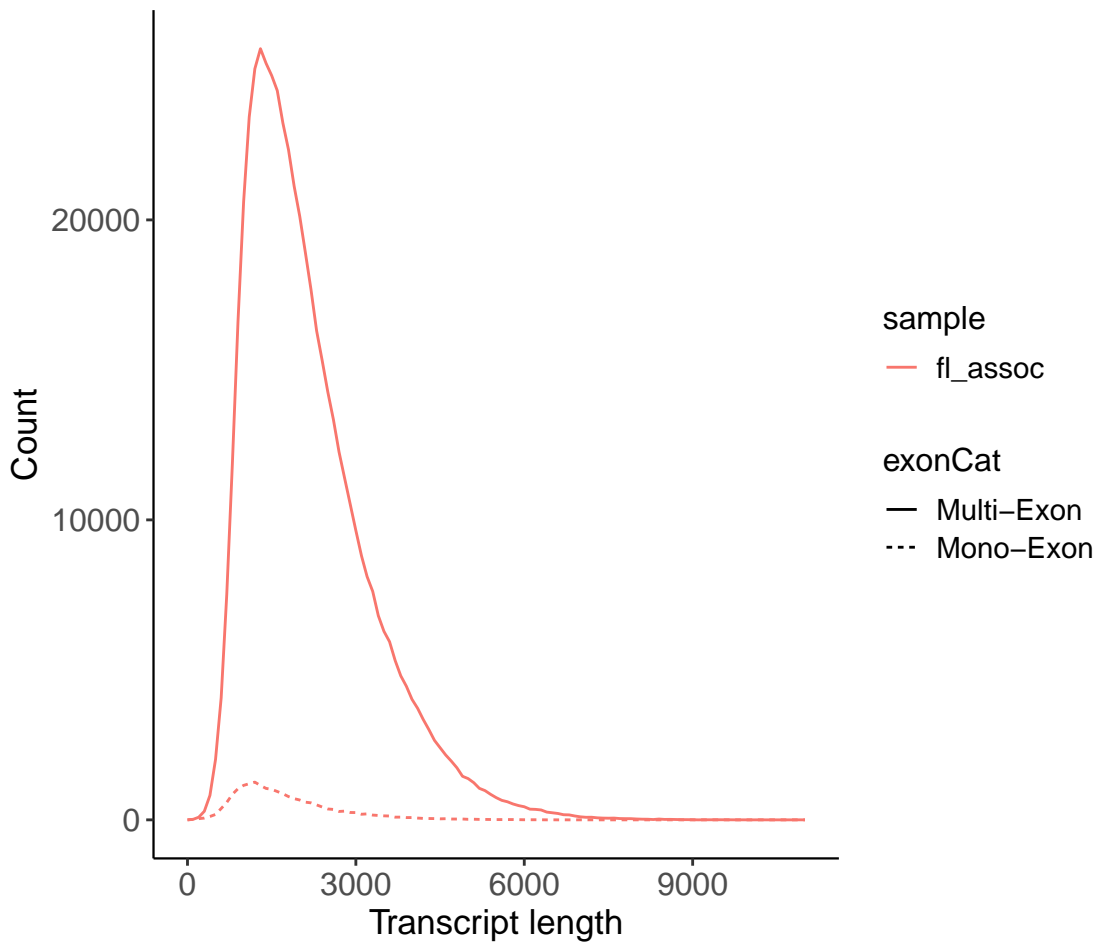
# Mono- vs Multi- Exon Transcript Lengths Distribution



# Transcript Lengths by Sample



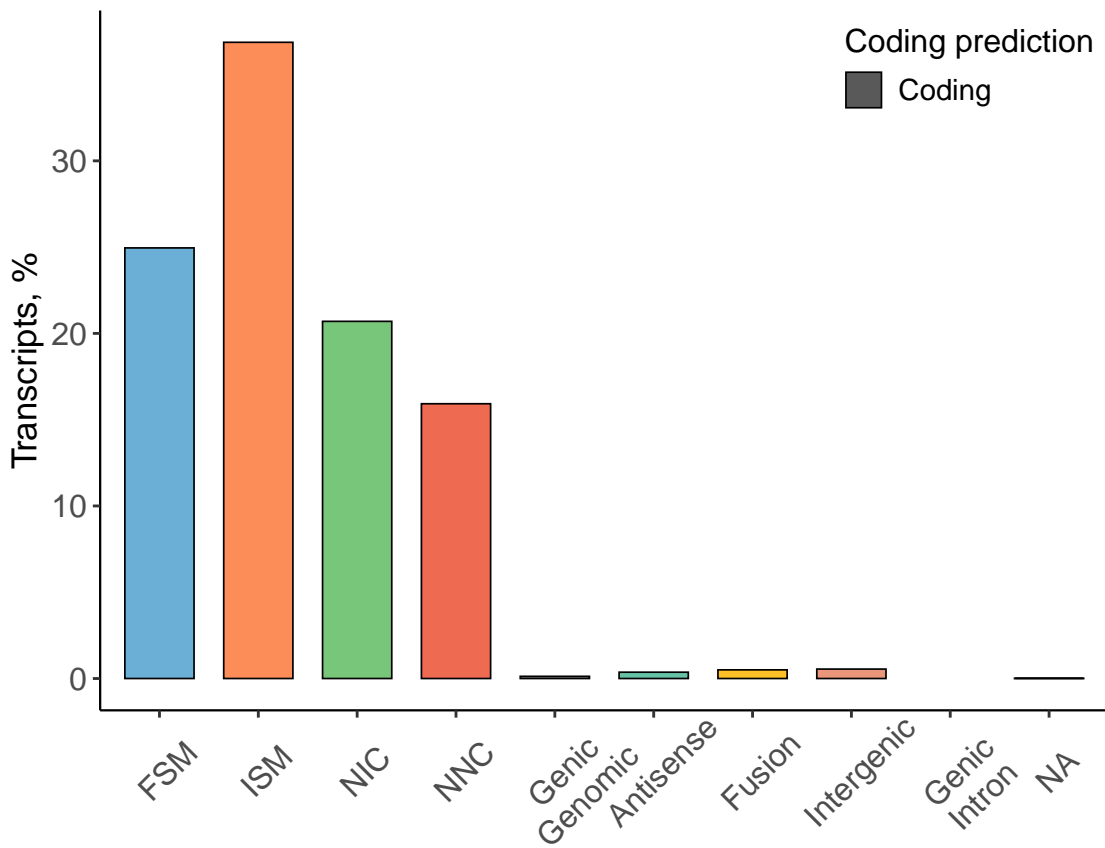
# Mono- vs Multi-Exons Transcript Lengths by Sample



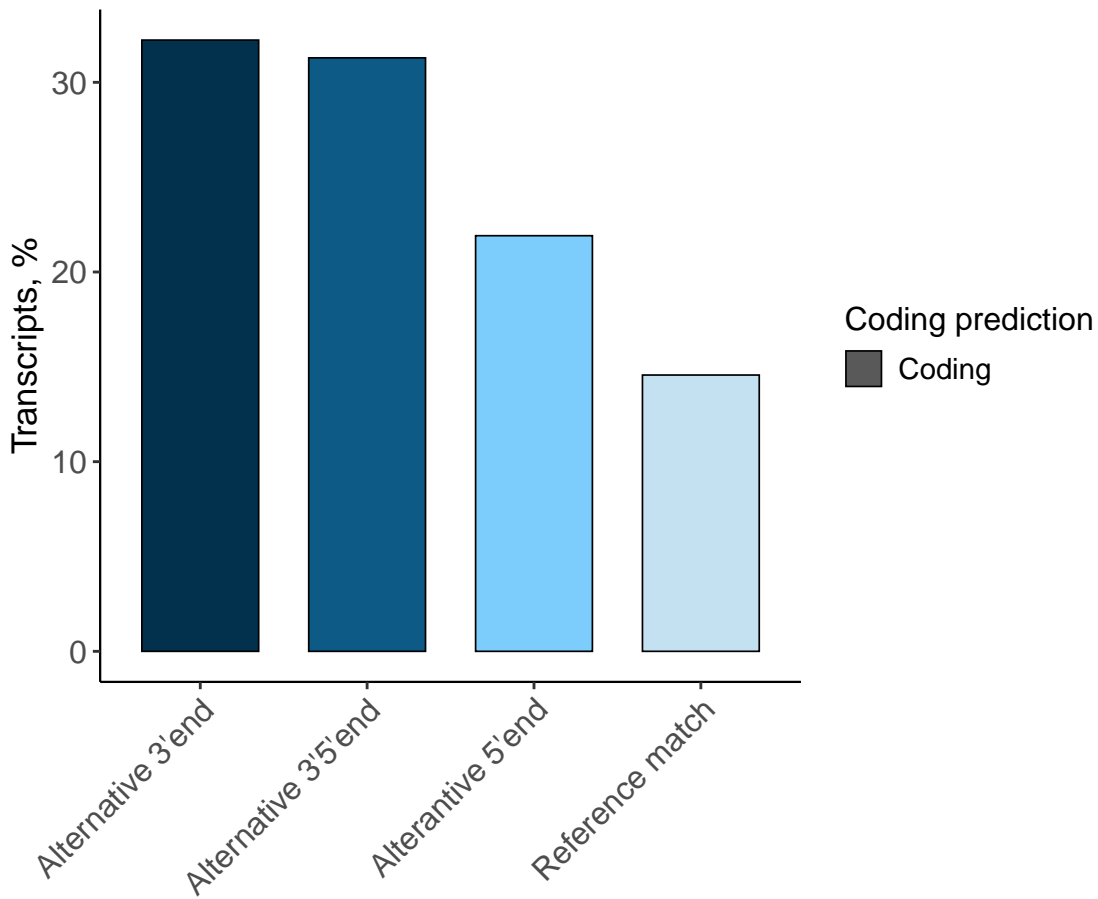
## *Structural Isoform Characterization*



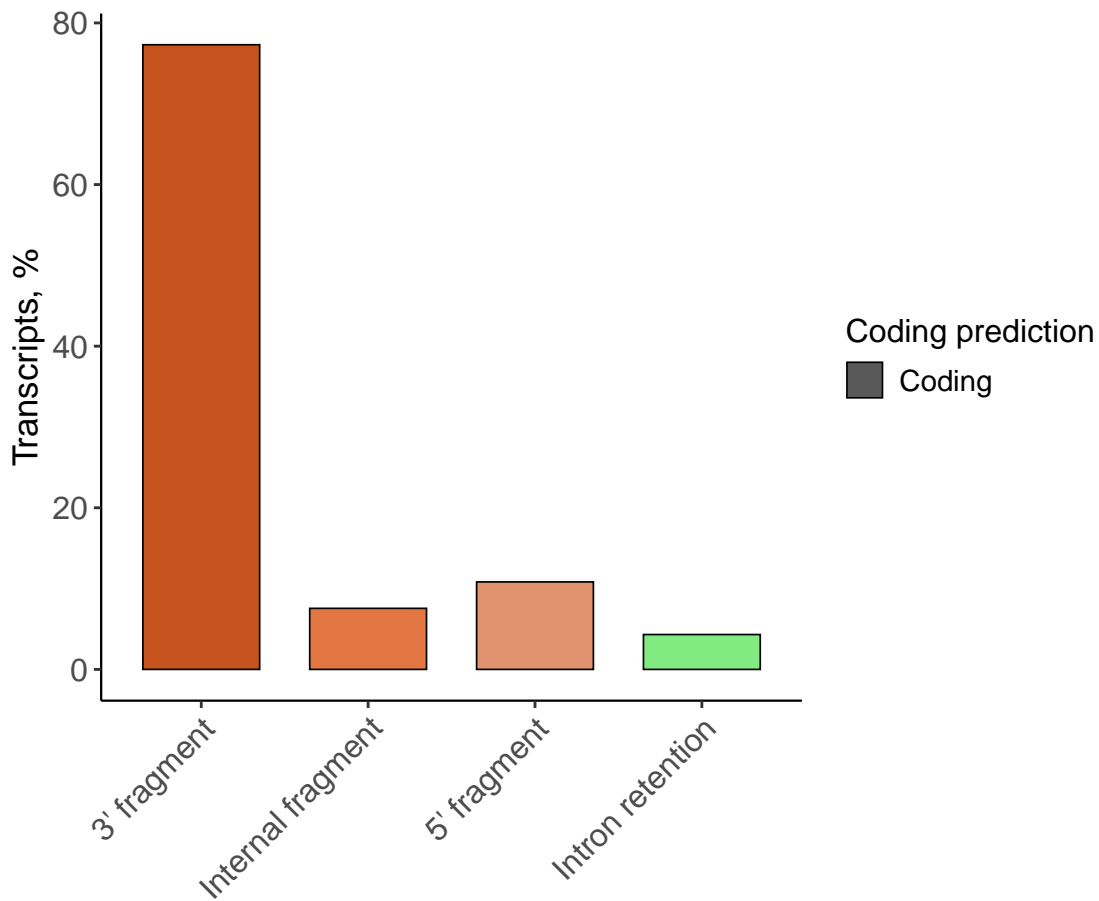
## Isoform Distribution Across Structural Categories



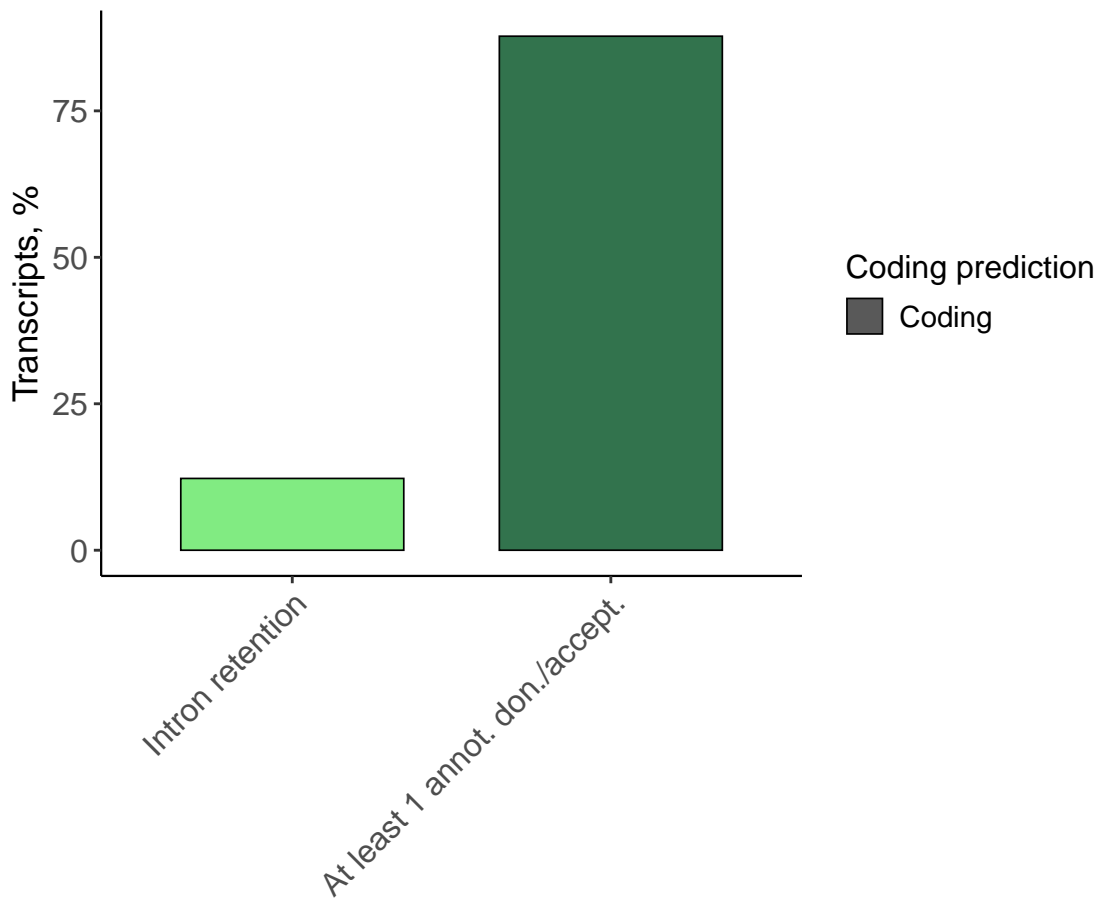
## Isoform Distribution Across FSM



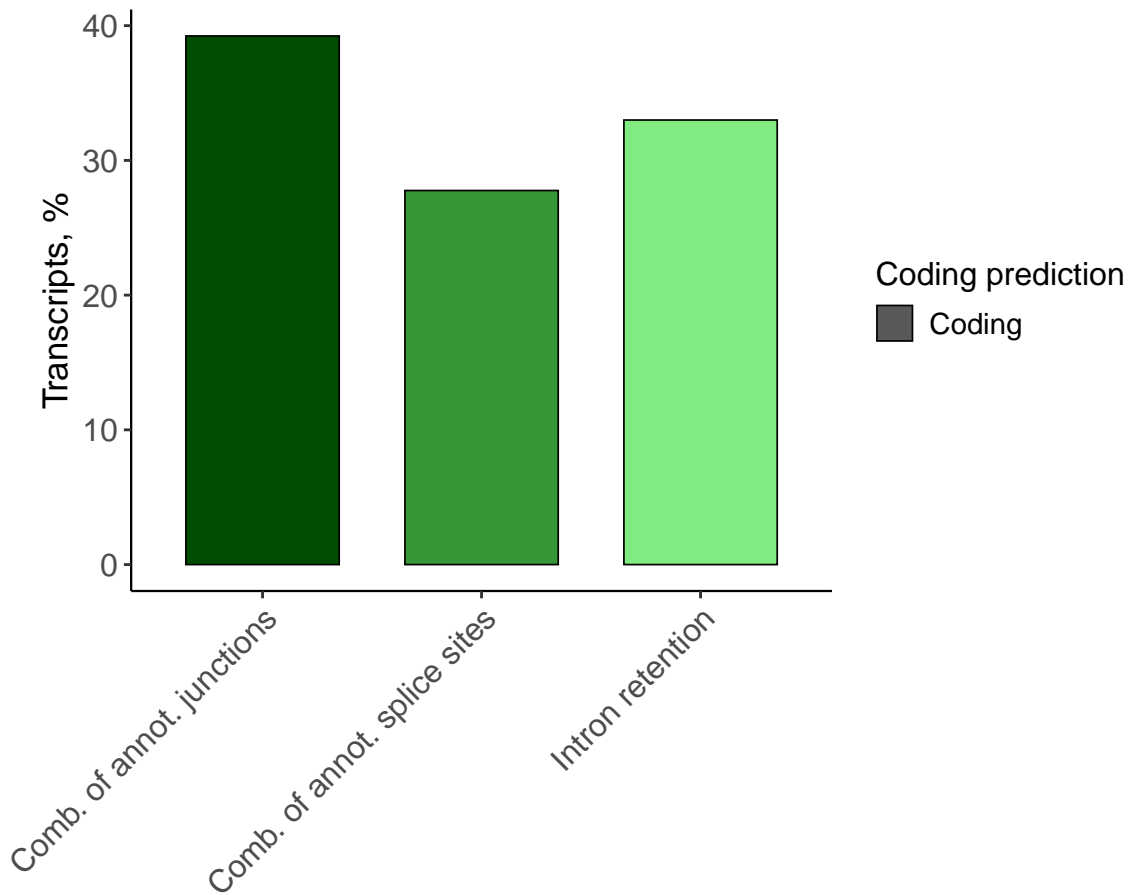
## Isoform Distribution Across ISM



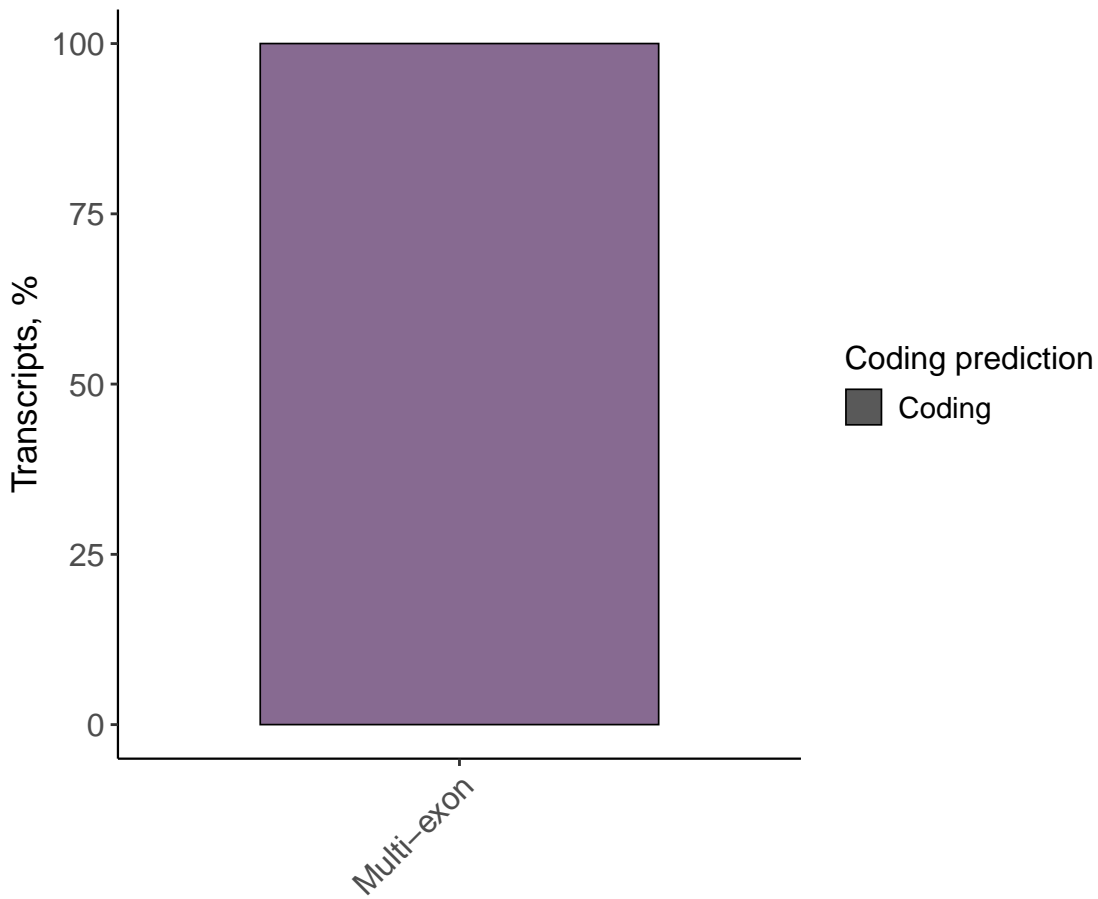
## Isoform Distribution Across NNC



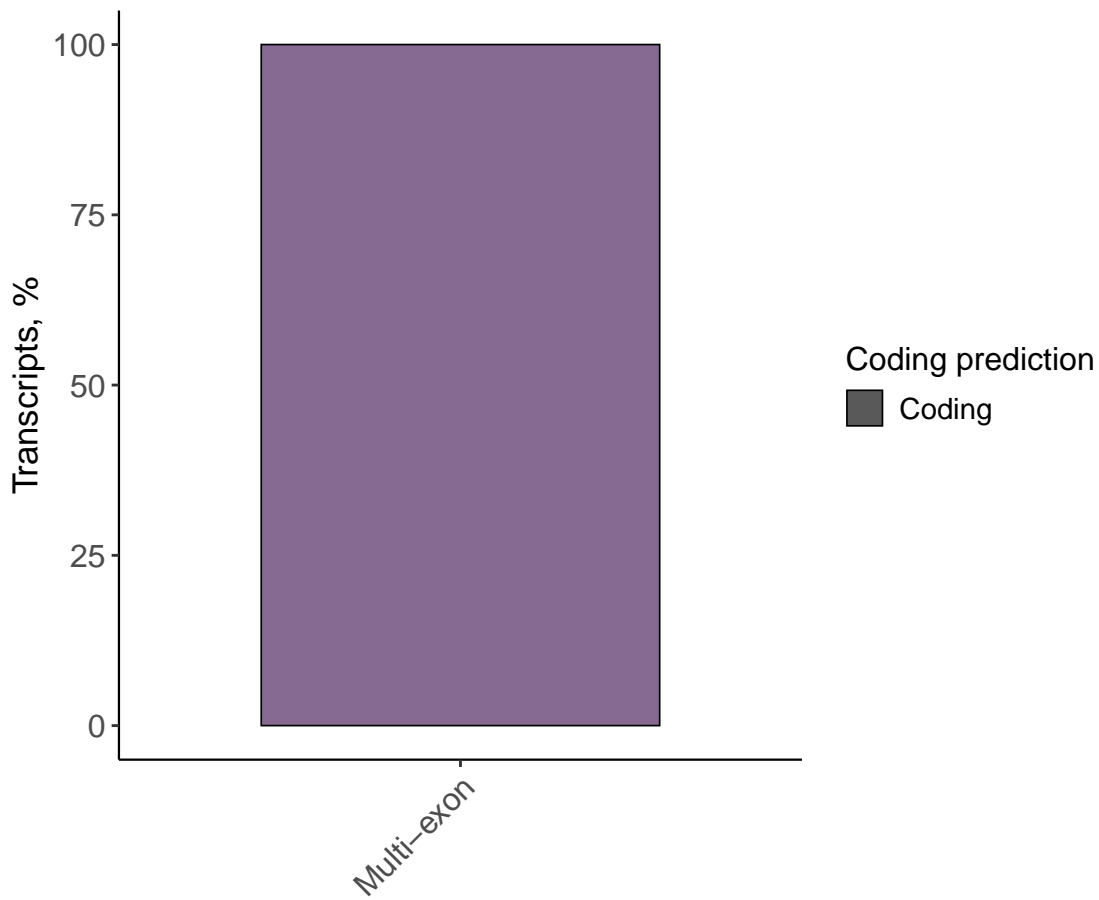
## Isoform Distribution Across NIC



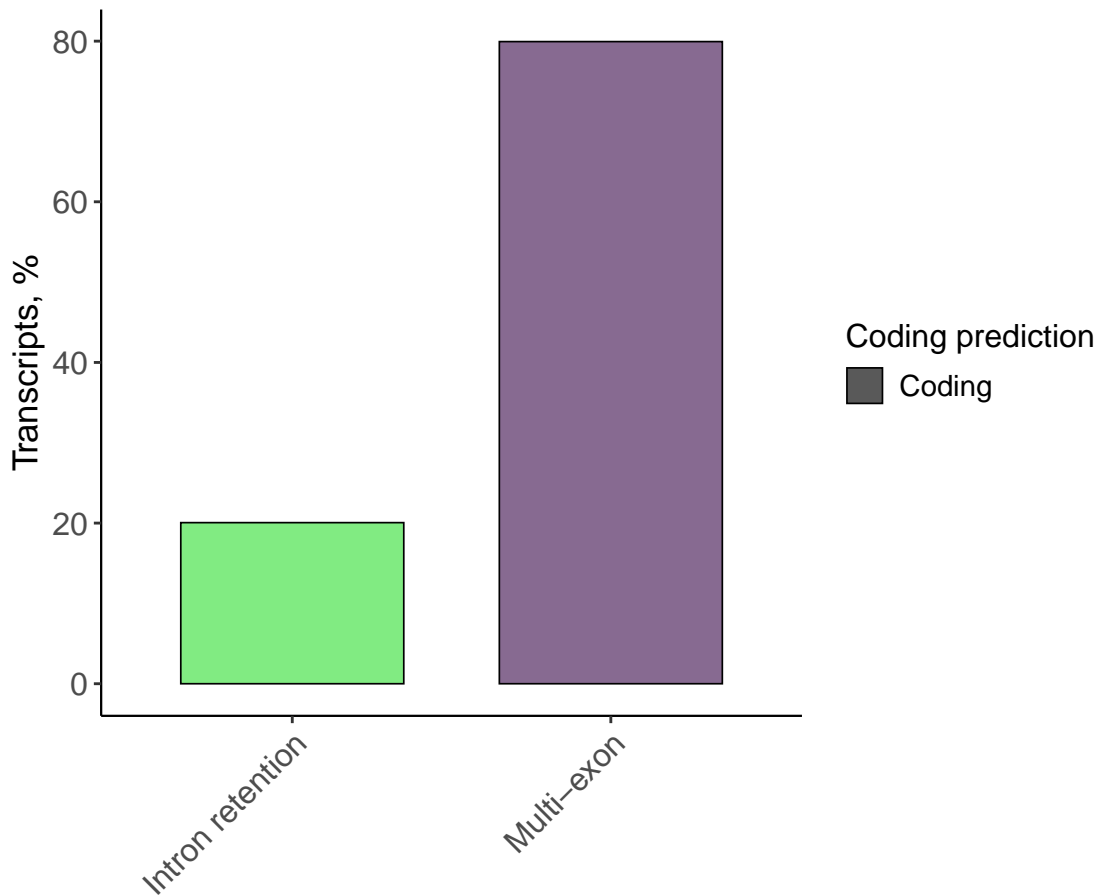
# Isoform Distribution Across Genic Genomic



## Isoform Distribution Across Antisense

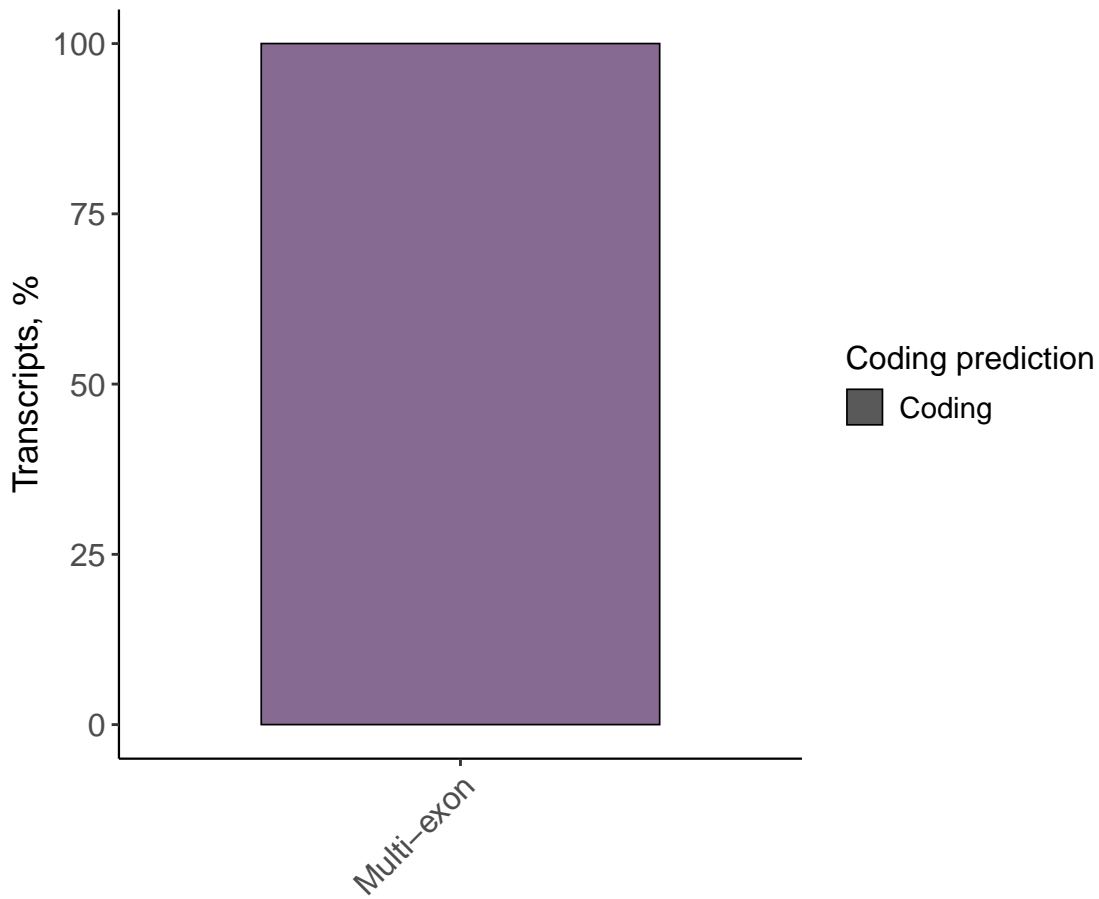


## Isoform Distribution Across Fusion

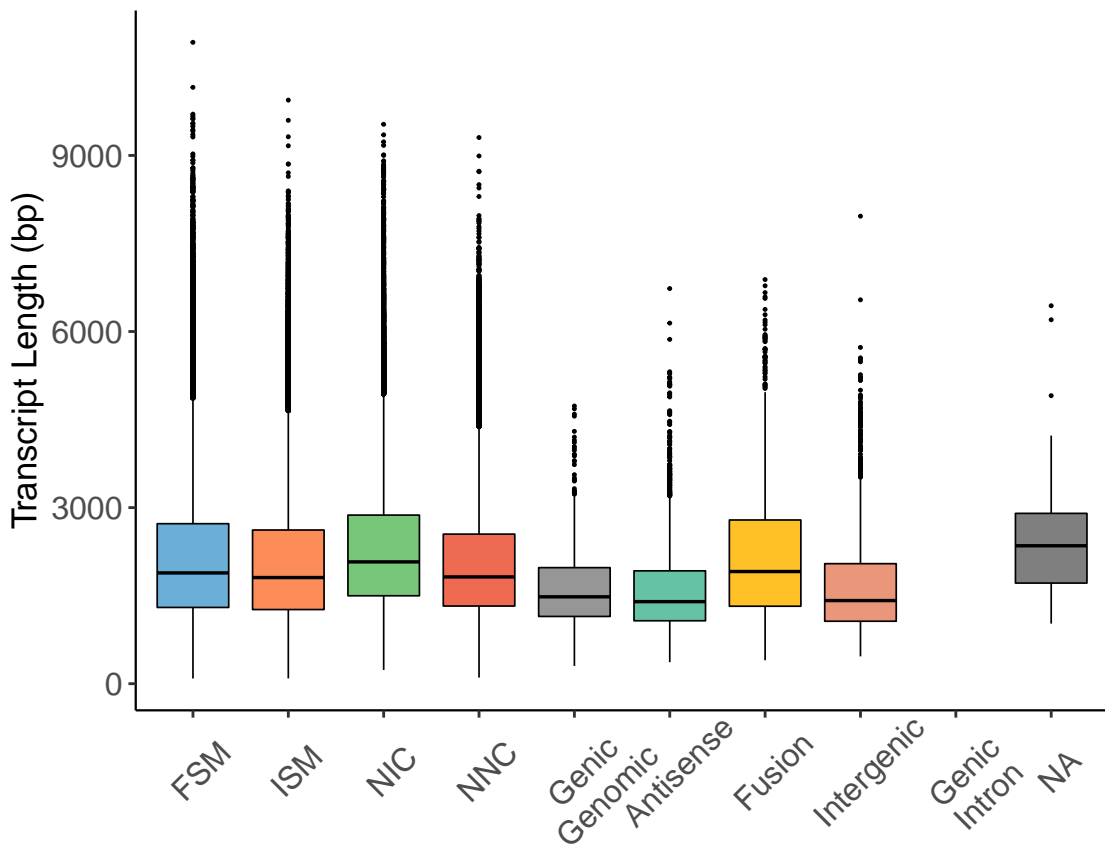




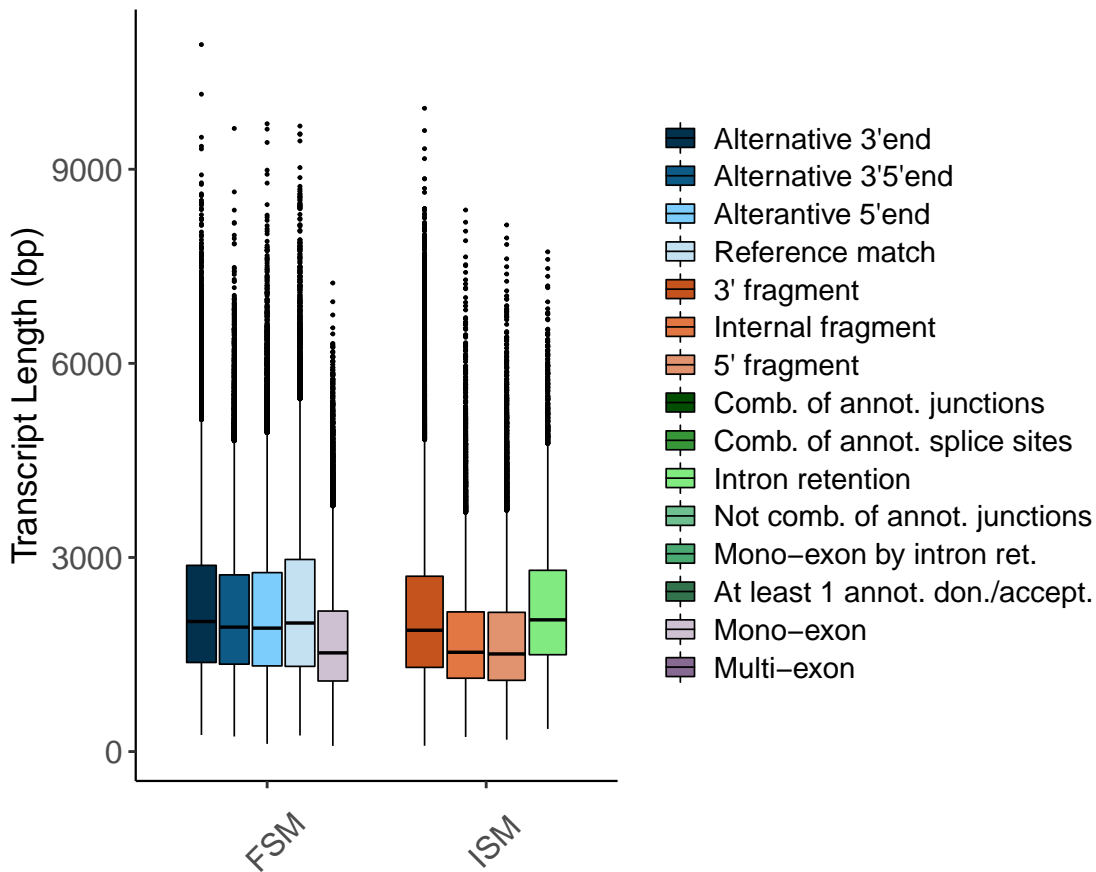
## Isoform Distribution Across Intergenic



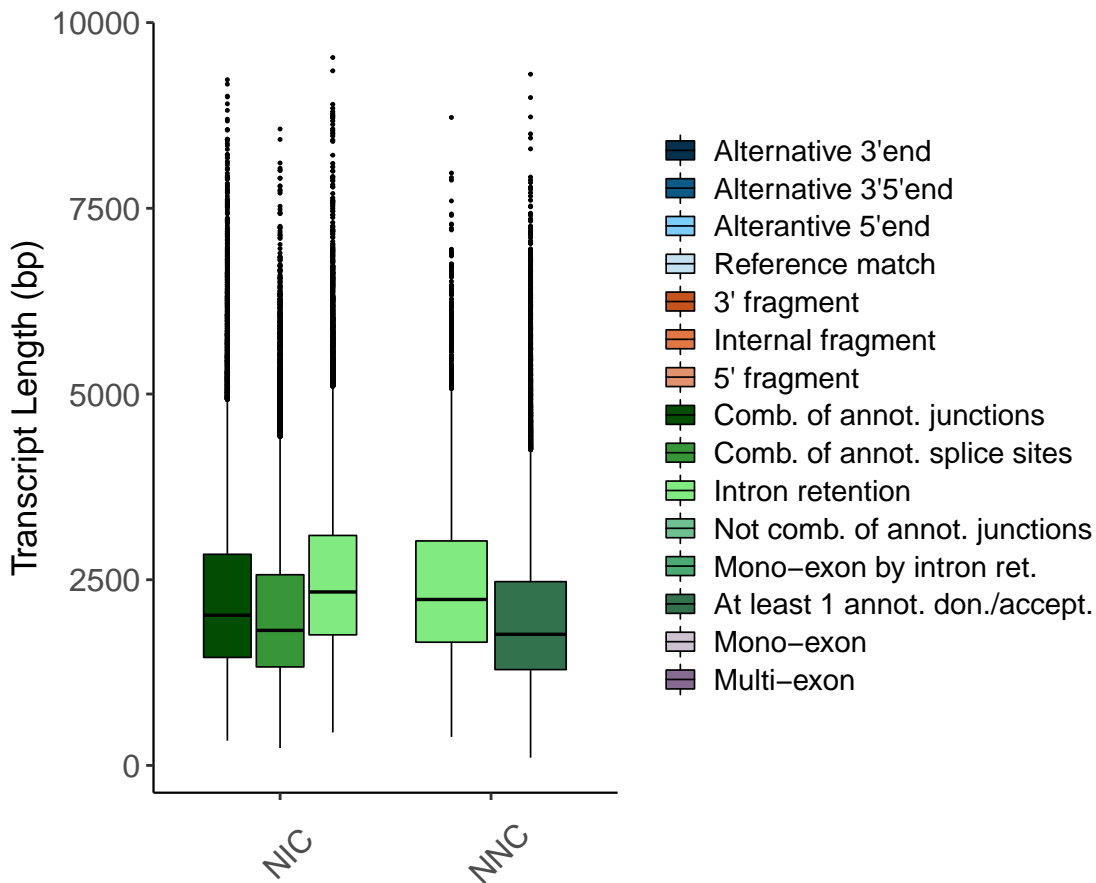
# Transcript Lengths by Structural Classification



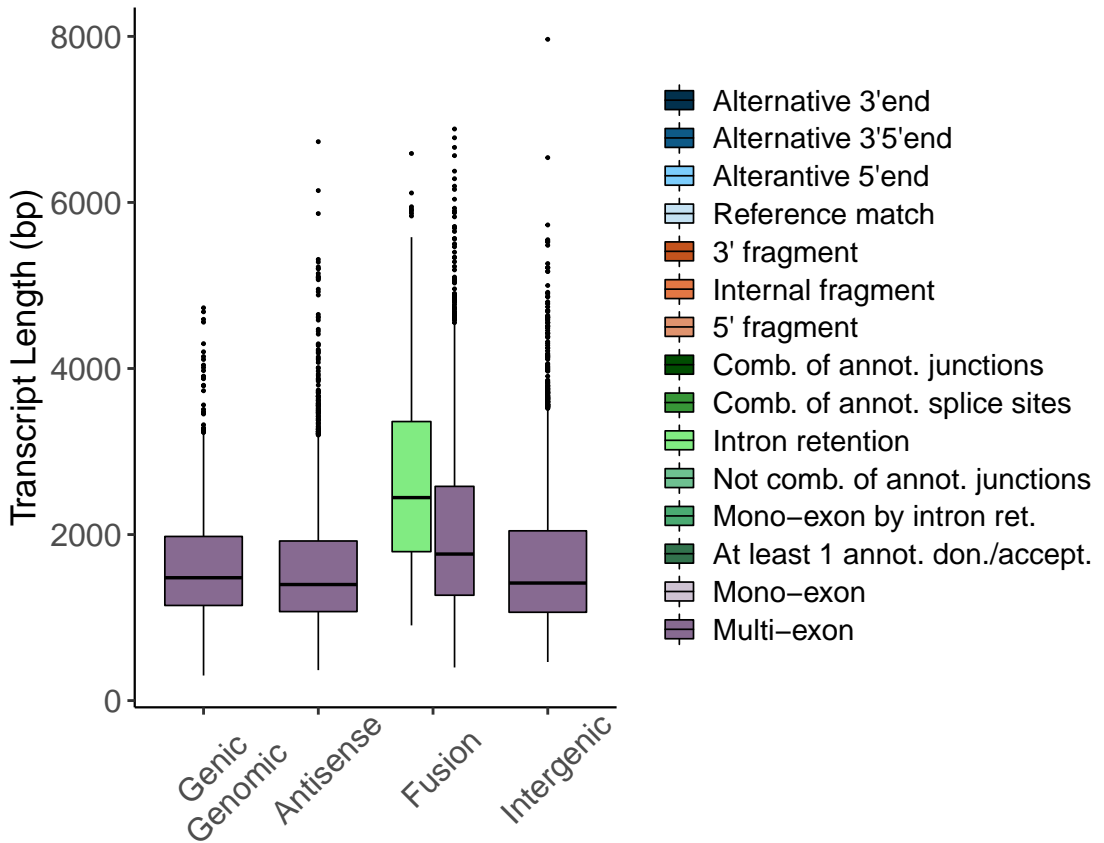
# Transcript Lengths by Subcategory



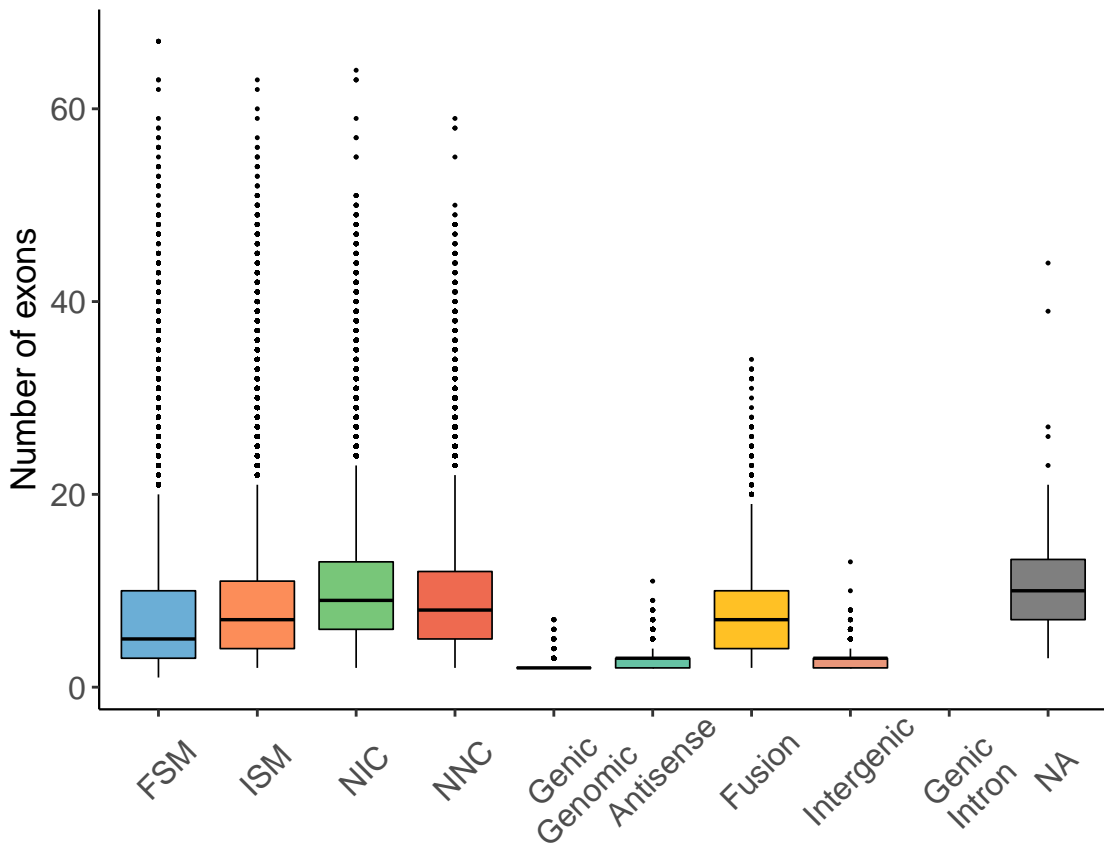
# Transcript Lengths by Subcategory



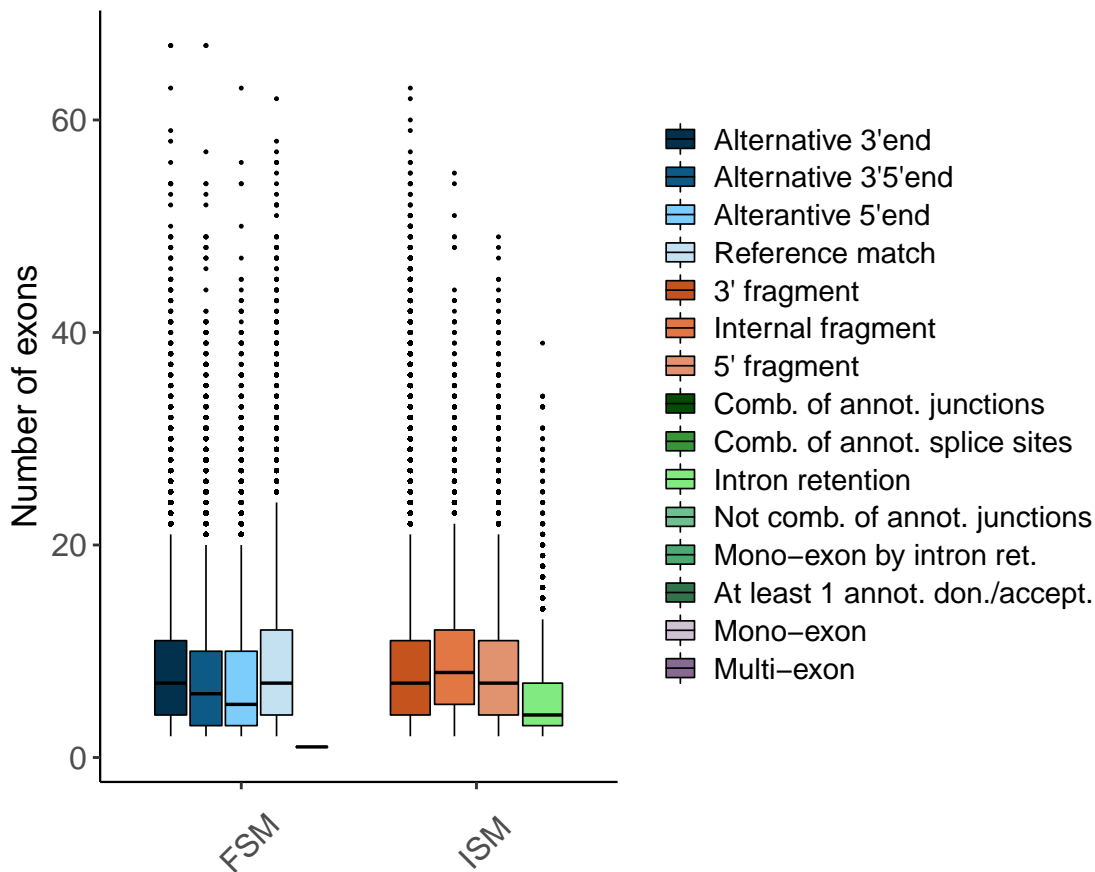
# Transcript Lengths by Subcategory



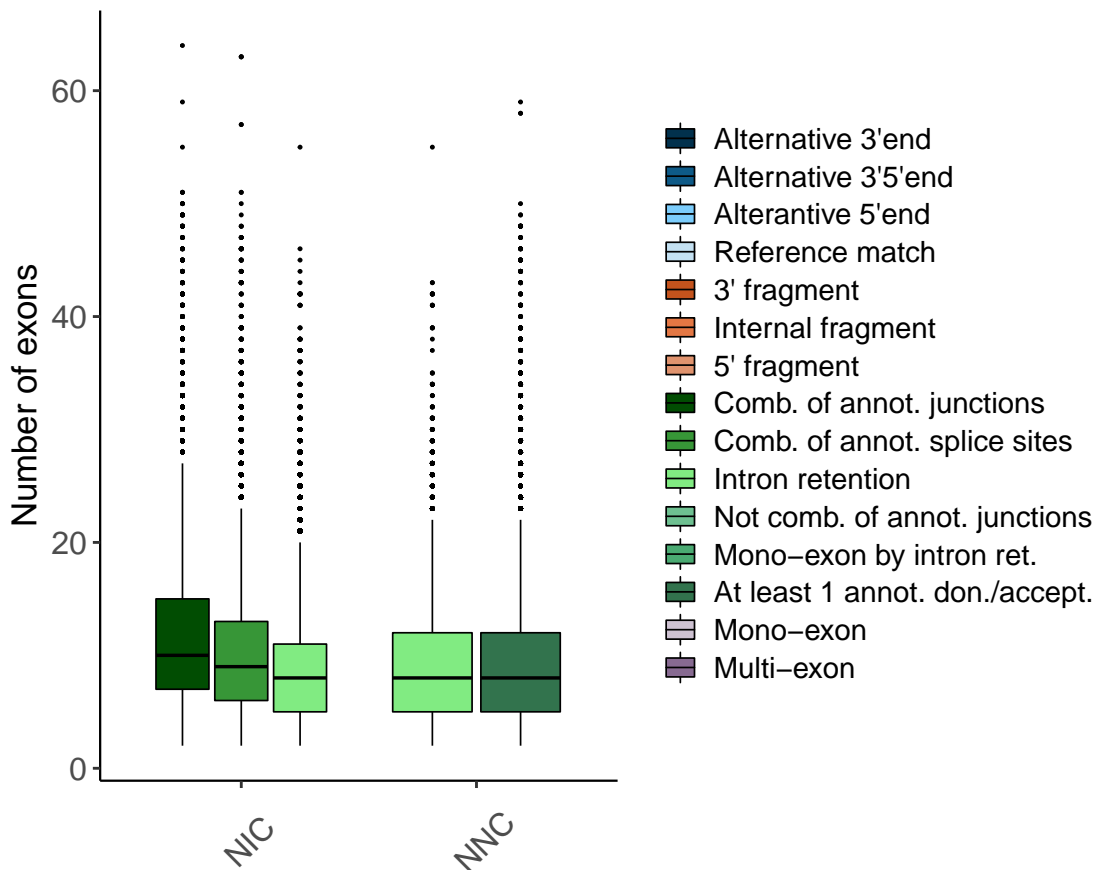
# Exon Counts by Structural Classification



## Exon Counts by Subcategory

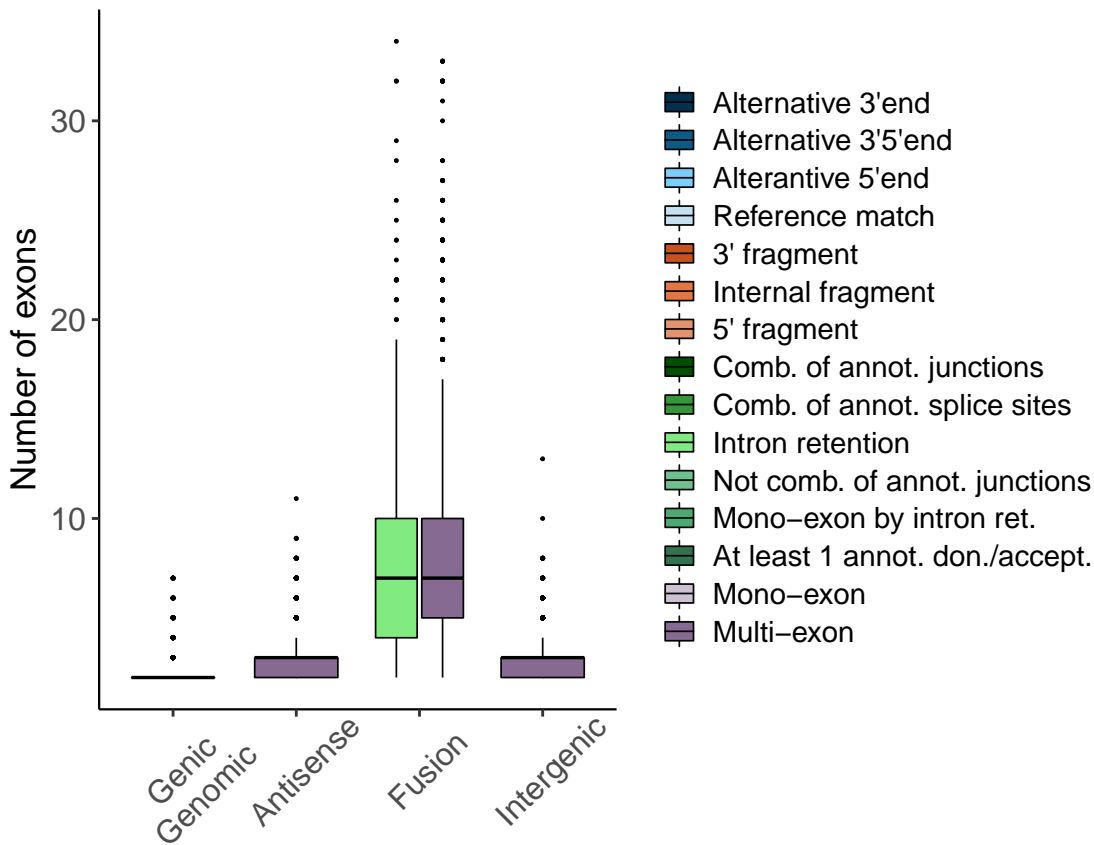


## Exon Counts by Subcategory



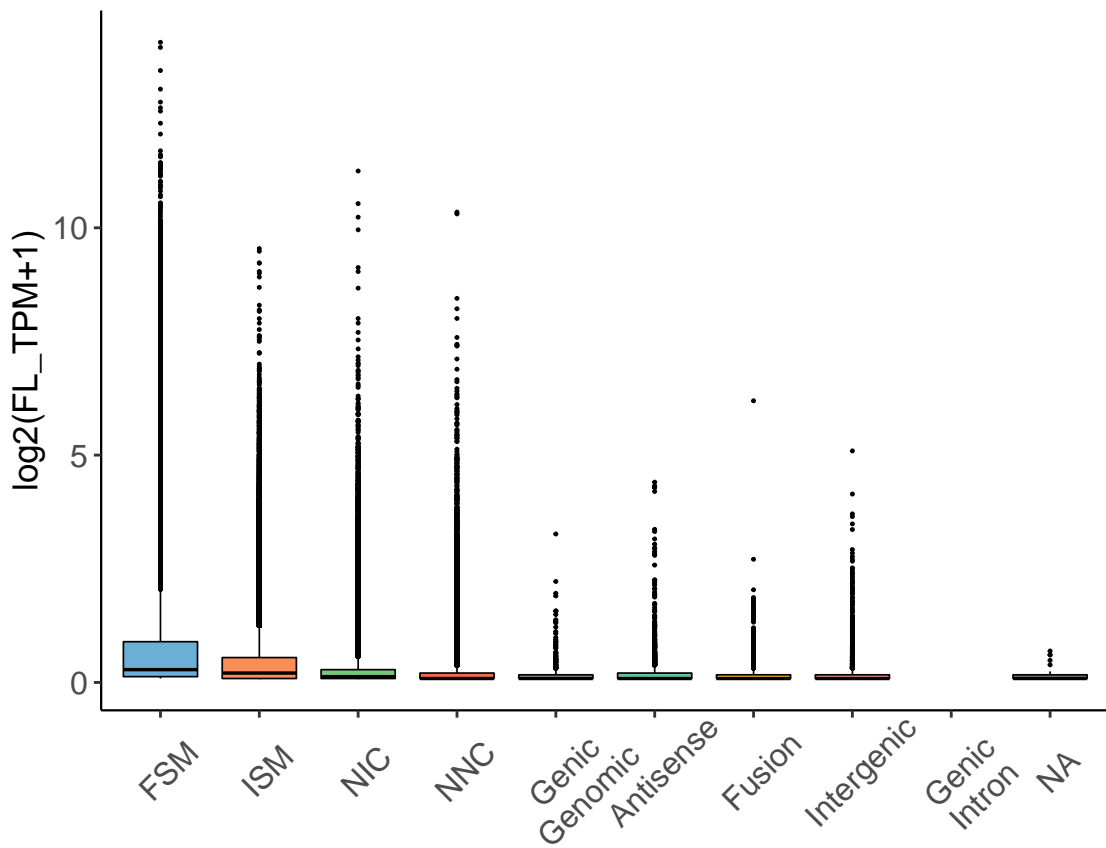


# Exon Counts by Subcategory



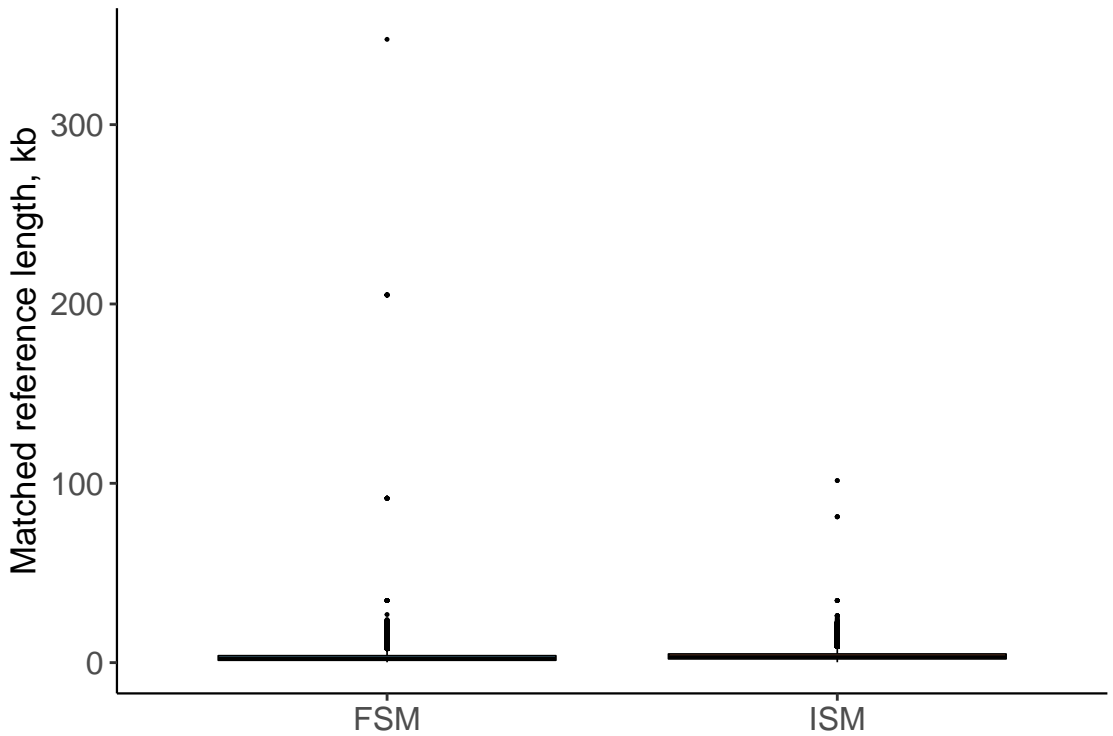
- Alternative 3'end
- Alternative 3'5'end
- Alterantive 5'end
- Reference match
- 3' fragment
- Internal fragment
- 5' fragment
- Comb. of annot. junctions
- Comb. of annot. splice sites
- Intron retention
- Not comb. of annot. junctions
- Mono-exon by intron ret.
- At least 1 annot. don./accept.
- Mono-exon
- Multi-exon

## Long Reads Count by Structural Category



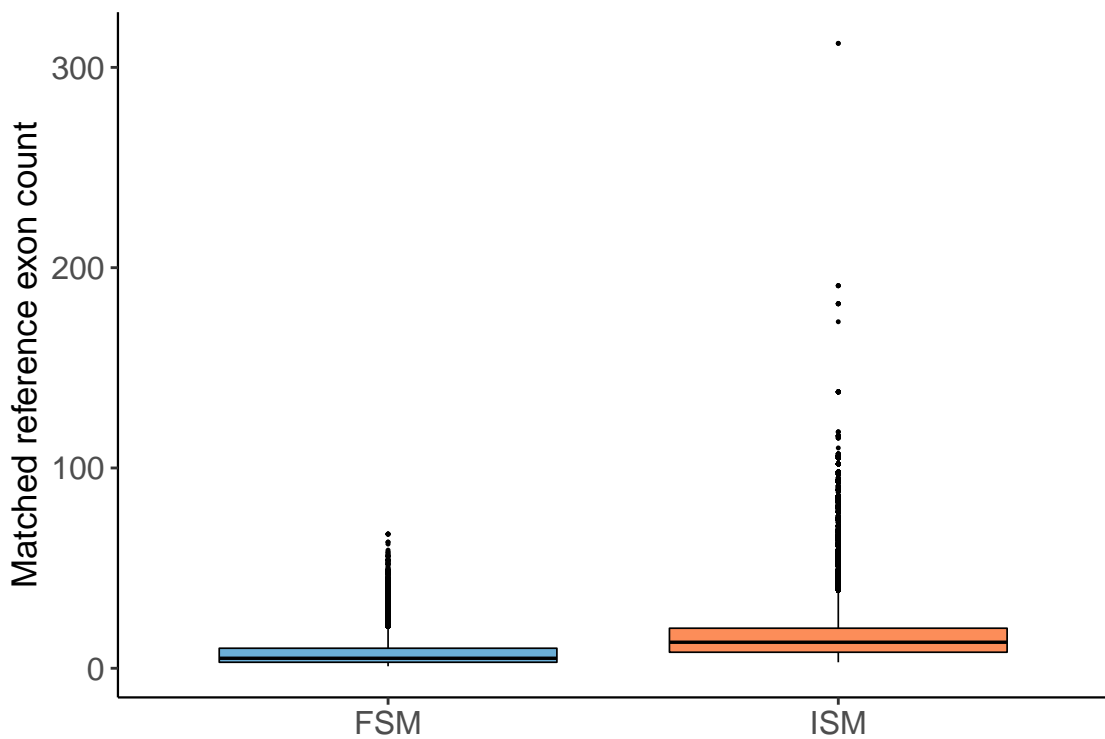
# Length Distribution of Matched Reference Transcripts

Applicable Only to FSM and ISM Categories



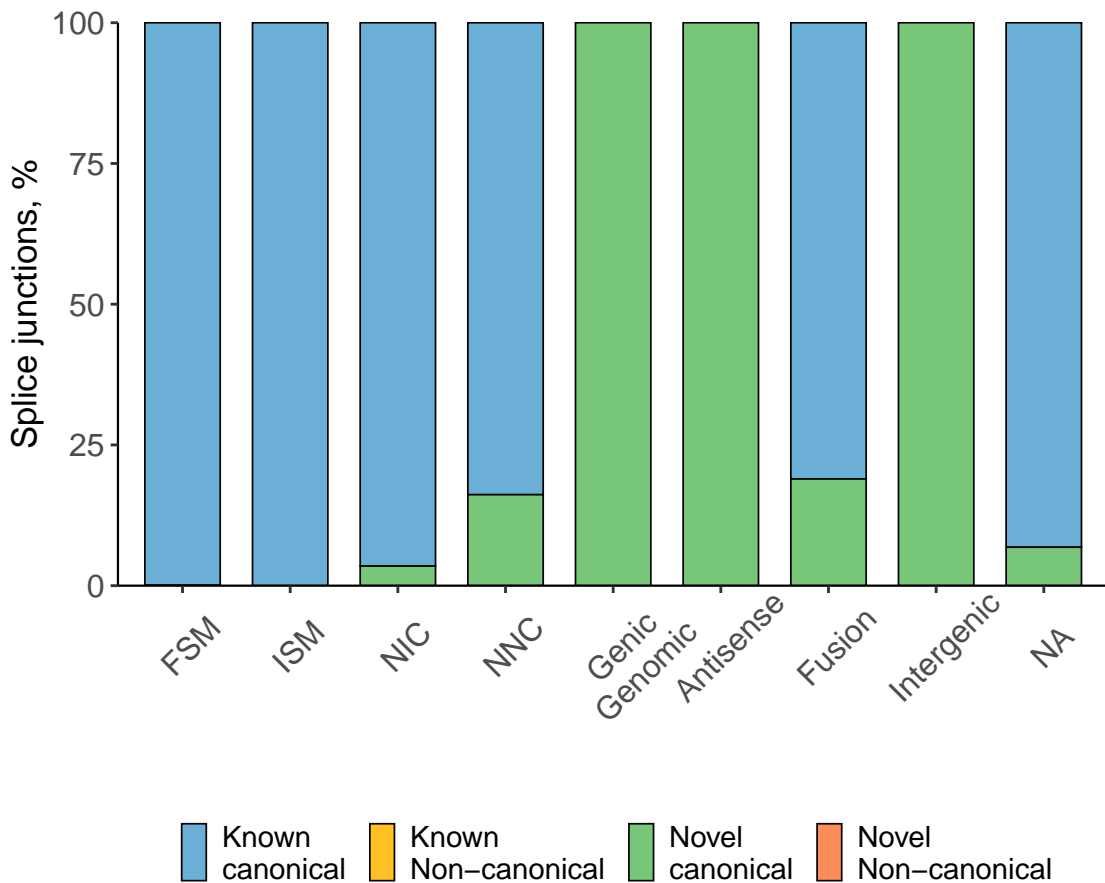
# Exon Count Distribution of Matched Reference Transcripts

Applicable Only to FSM and ISM Categories

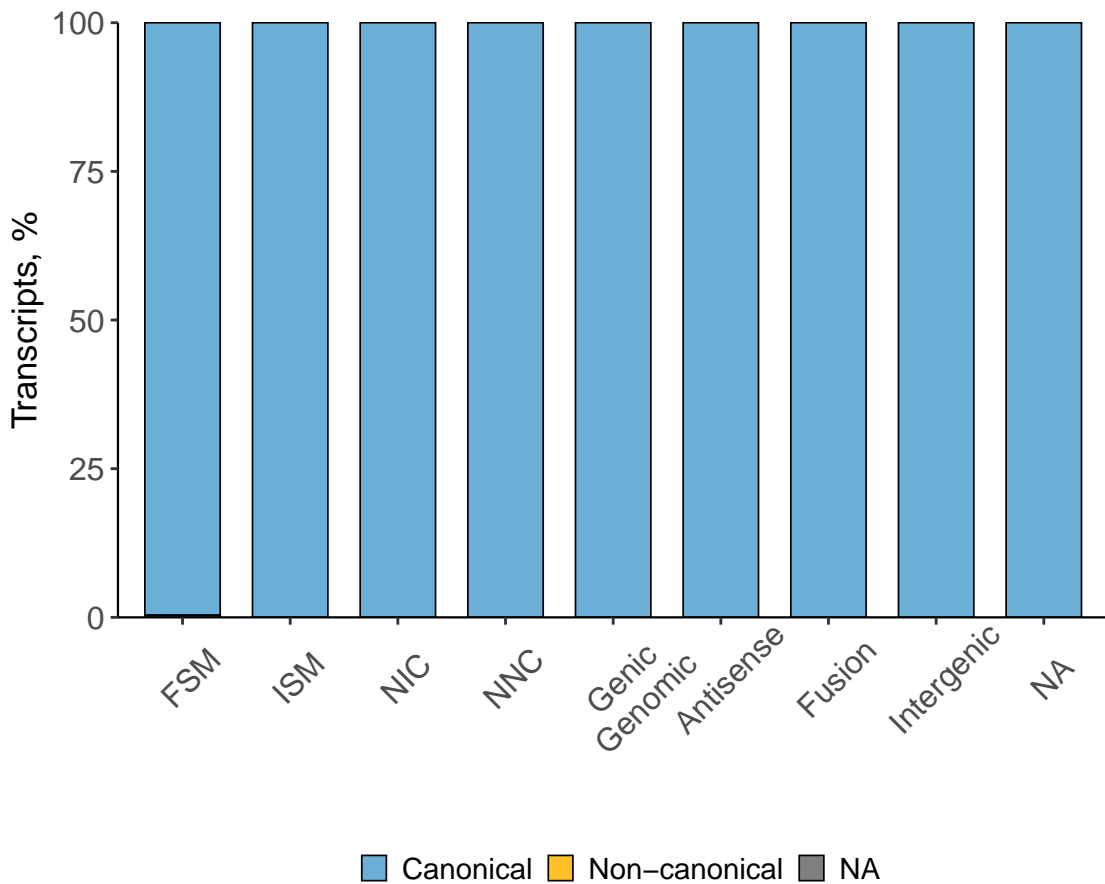


## *Splice Junction Characterization*

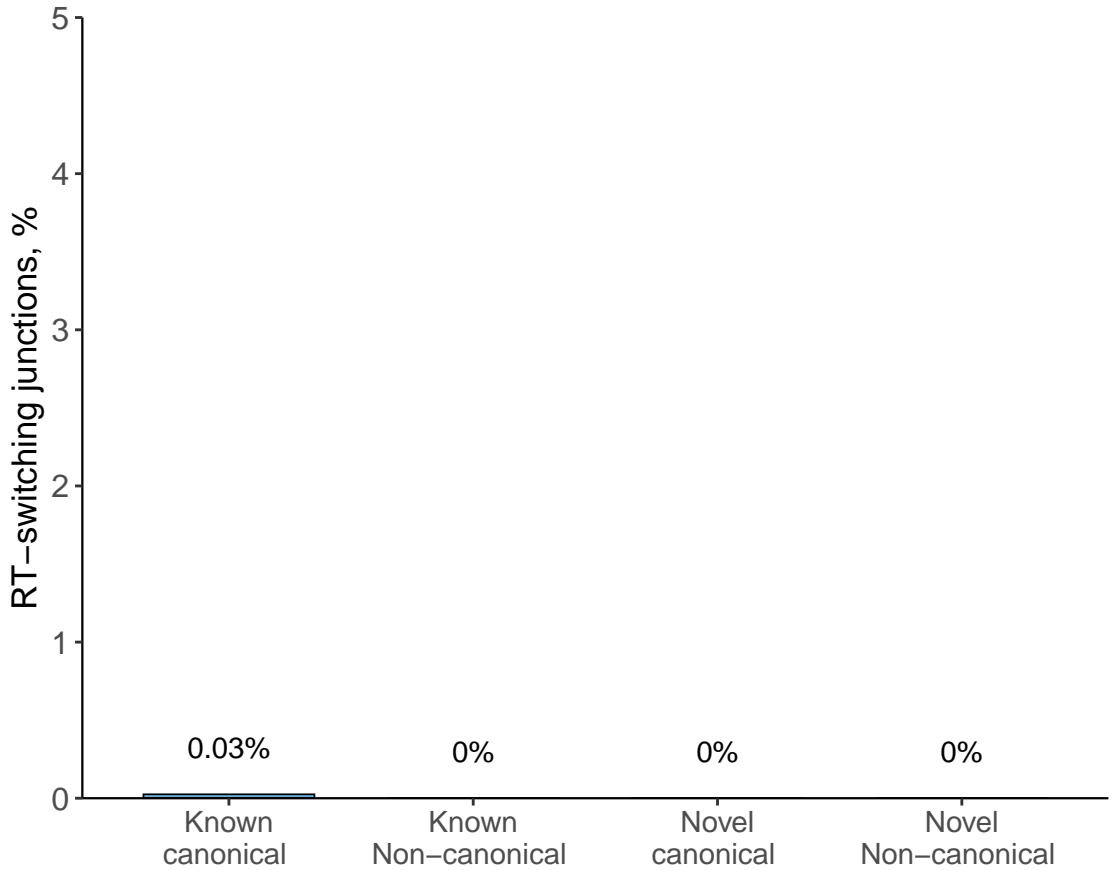
# Distribution of Splice Junctions by Structural Classification



## Distribution of Transcripts by Splice Junctions

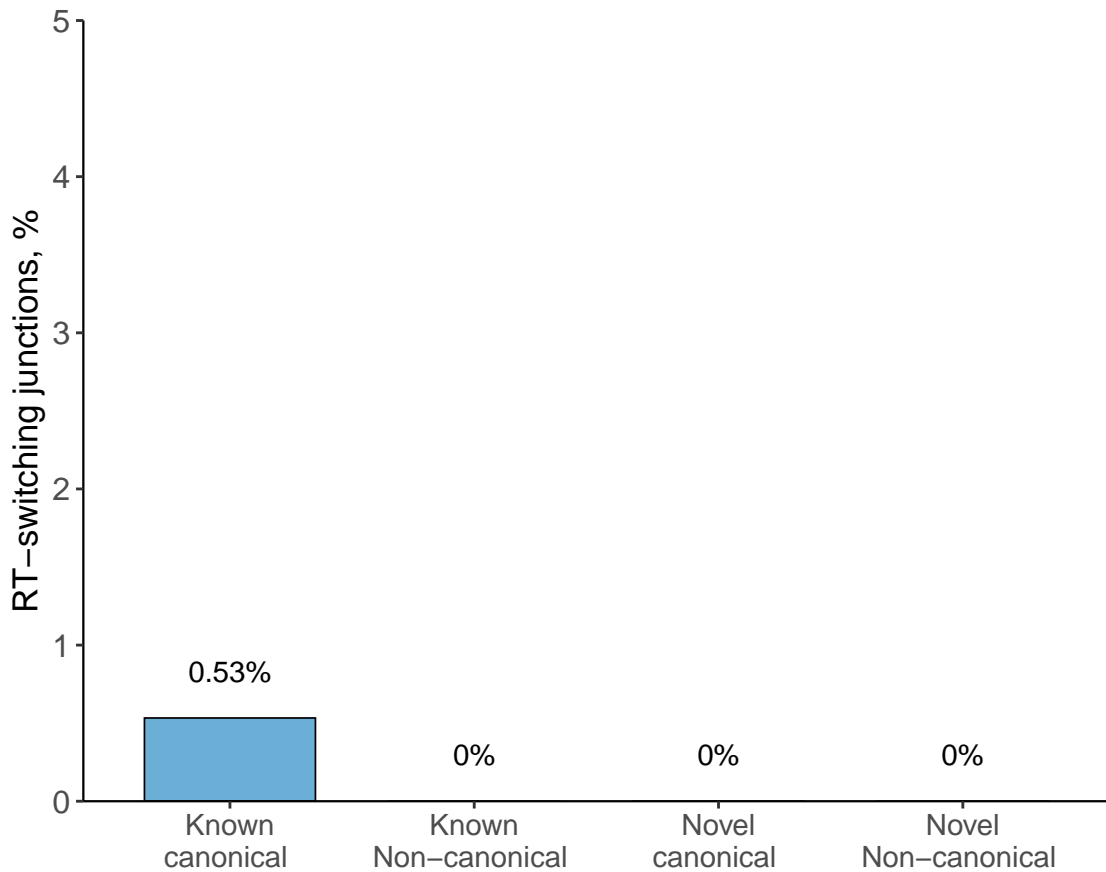


## RT-Switching All Junctions





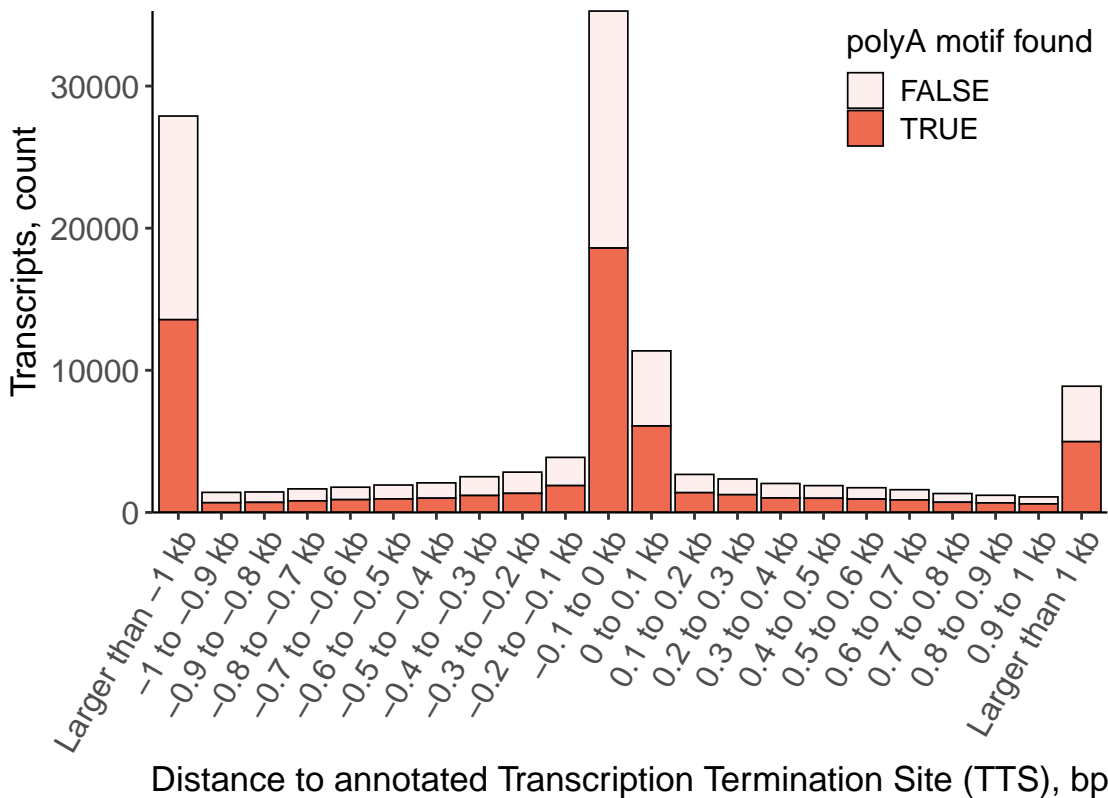
## Unique Junctions RT-switching



*Comparison With Annotated TSS and TTS*

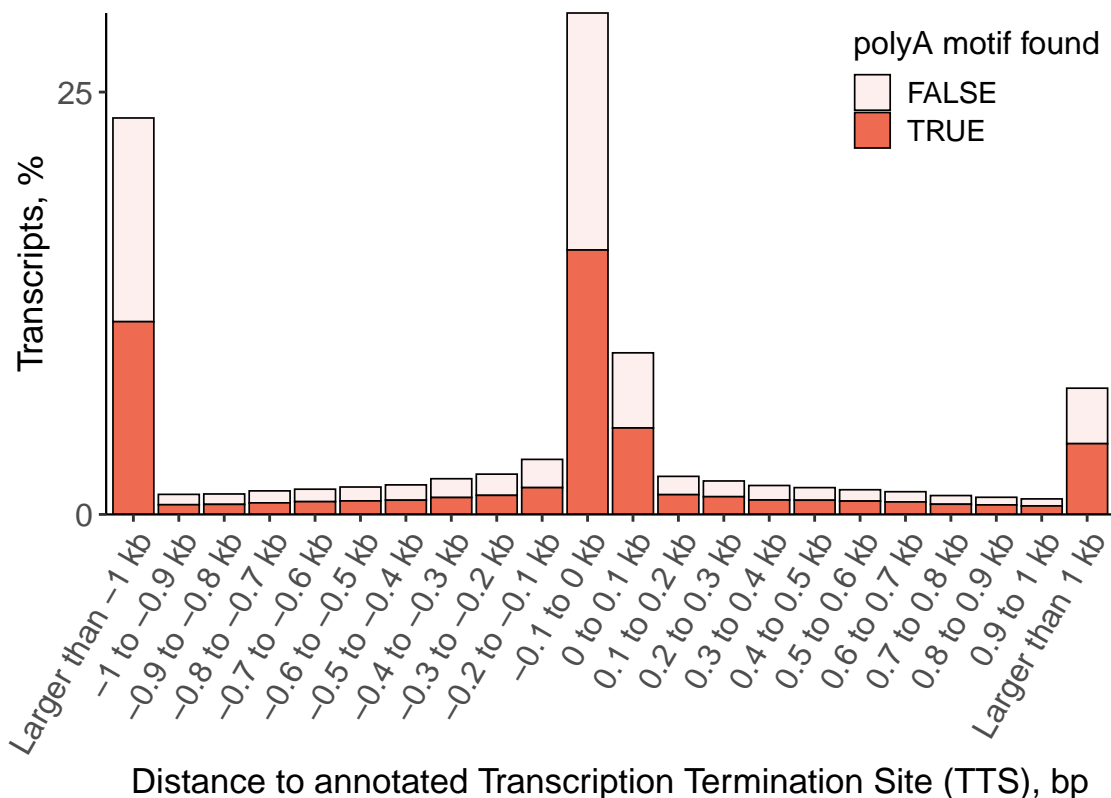
# Distance to annotated Transcription Termination Site (TTS) FSM

Negative values indicate upstream of annotated termination site



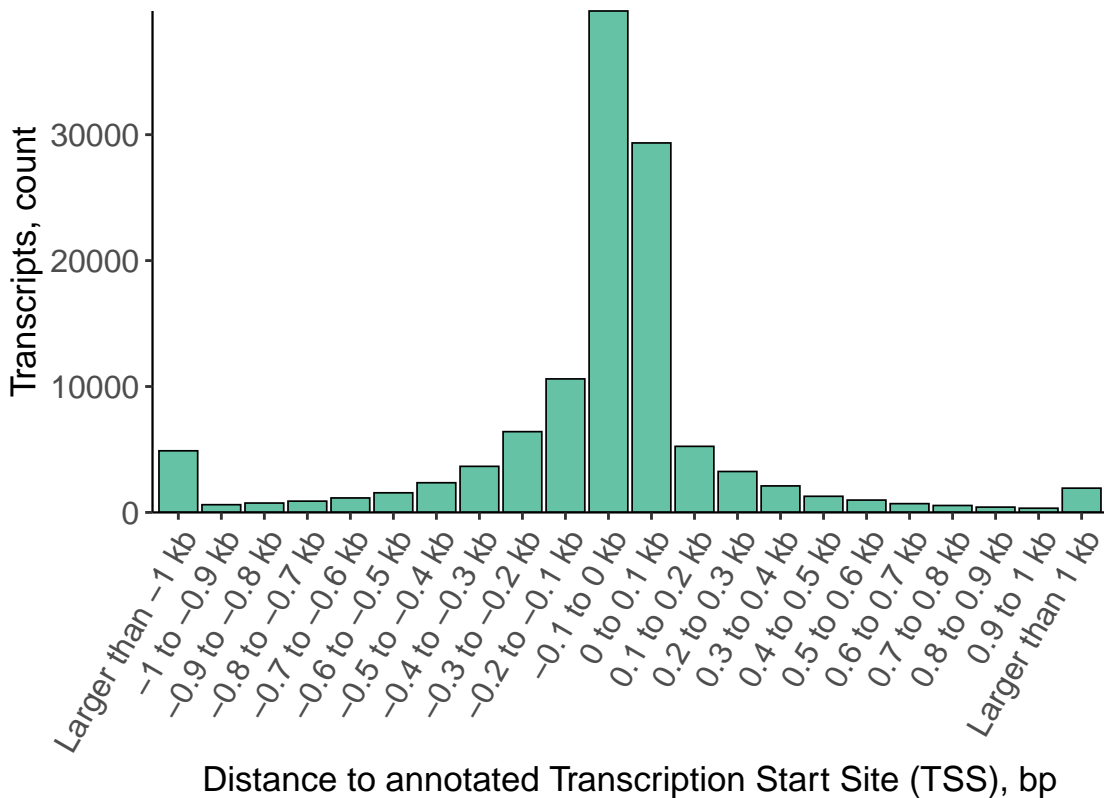
# Distance to annotated Transcription Termination Site (TTS) FSM

Negative values indicate upstream of annotated termination site



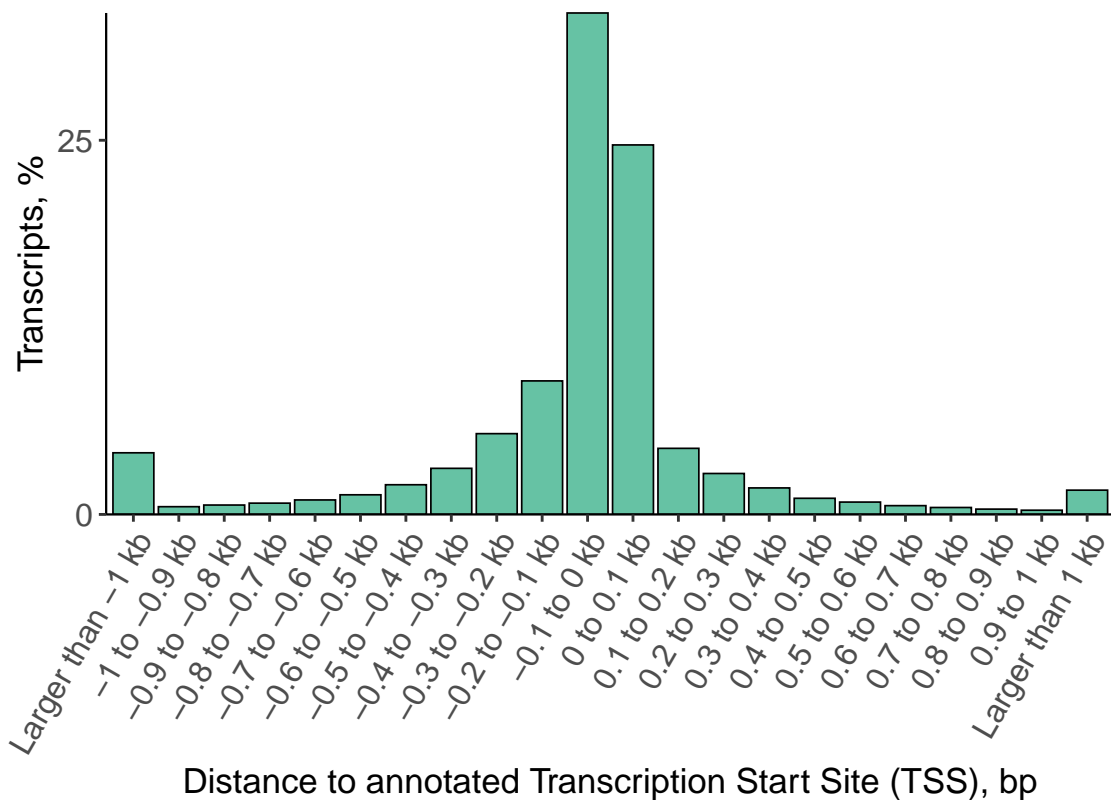
# Distance to Annotated Transcription Start Site for FSM

Negative values indicate downstream of annotated TSS



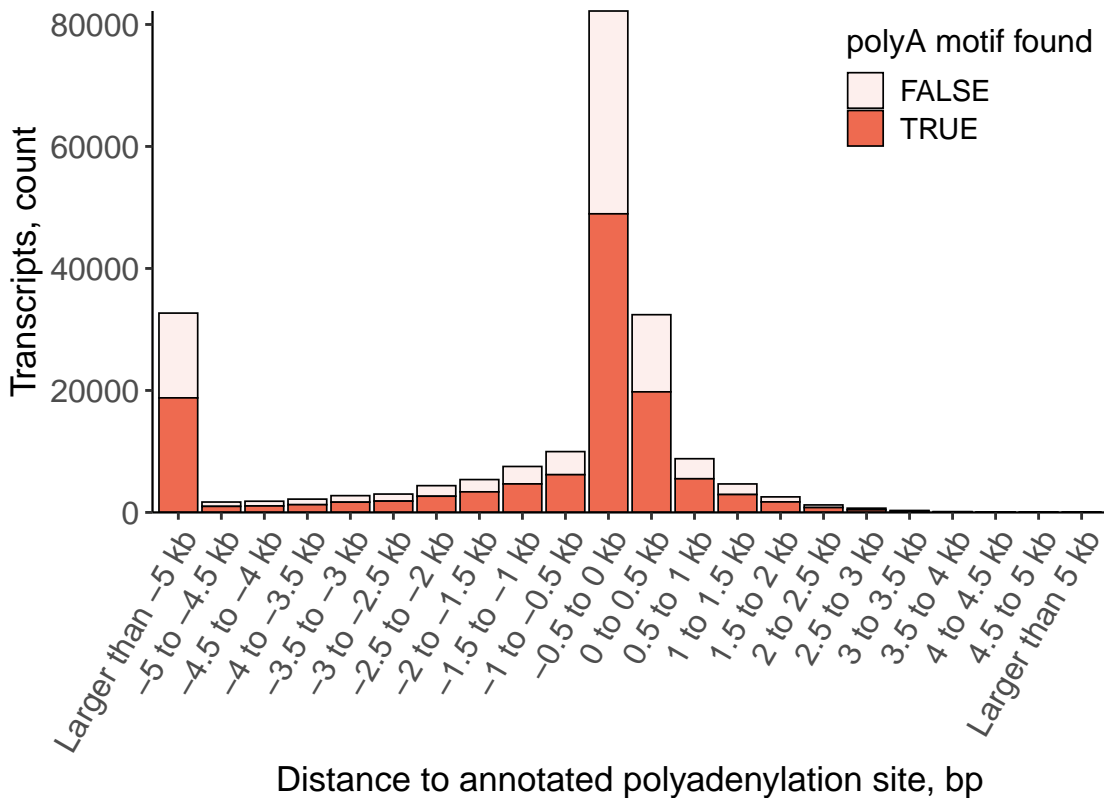
# Distance to Annotated Transcription Start Site for FSM

Negative values indicate downstream of annotated TSS



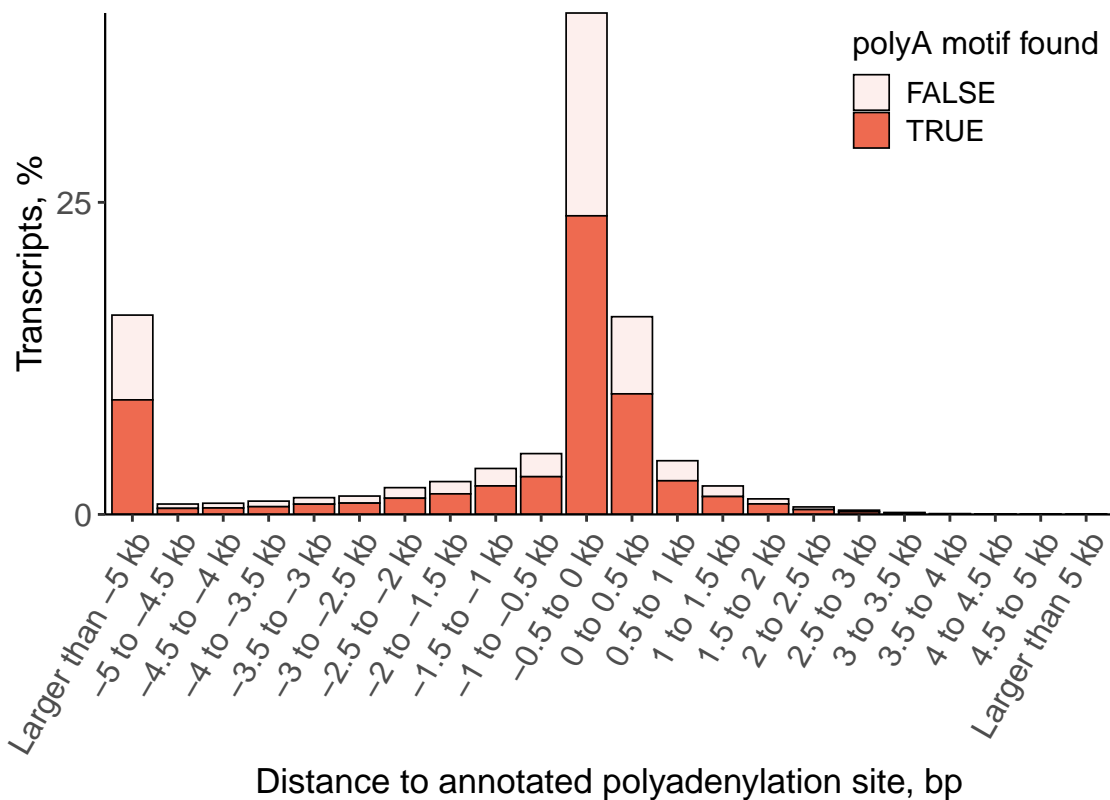
# Distance to Annotated Polyadenylation Site for ISM

Negative values indicate upstream of annotated polyA site



# Distance to Annotated Polyadenylation Site for ISM

Negative values indicate upstream of annotated polyA site

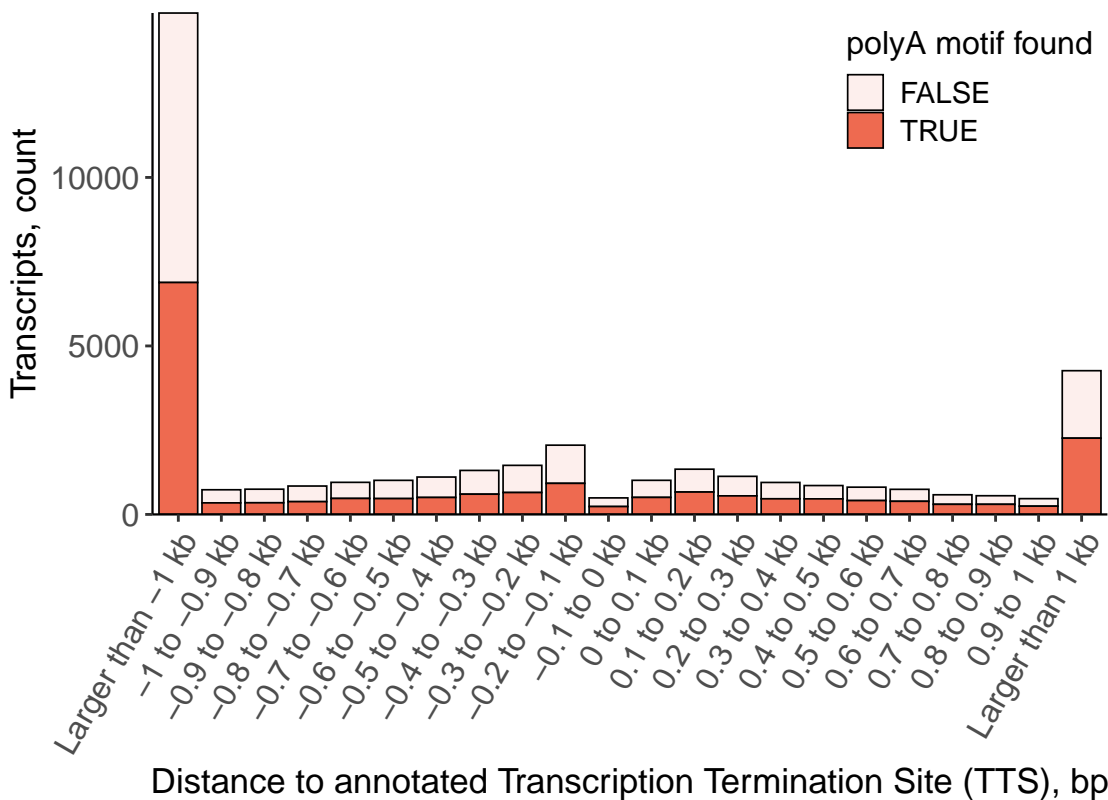




*Comparison With Annotated TSS and TTS  
by Subcategories*

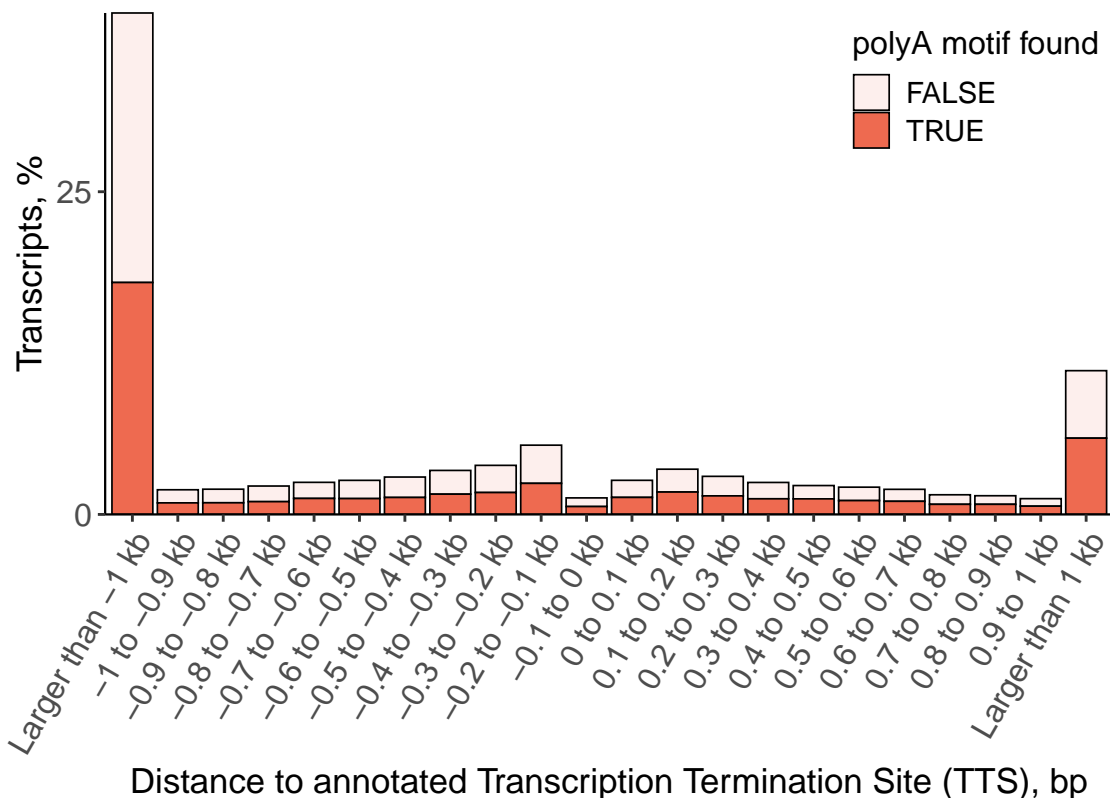
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 3'End

Negative values indicate upstream of annotated termination site



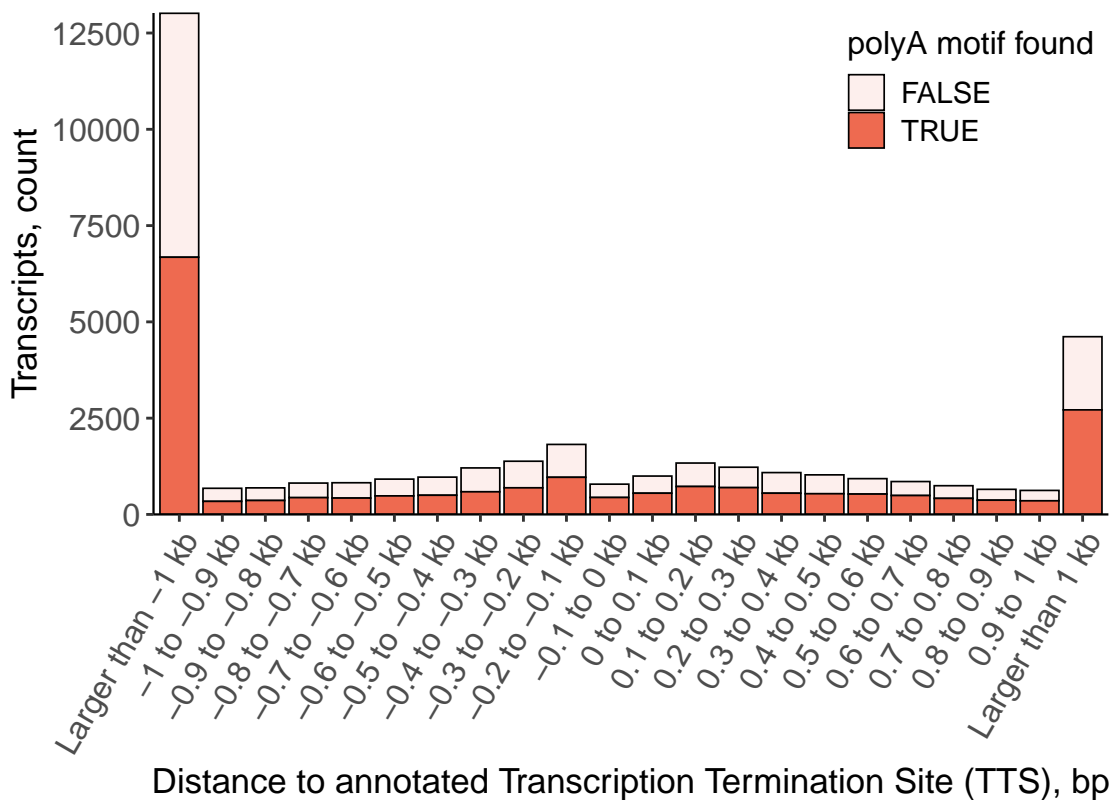
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 3'End

Negative values indicate upstream of annotated termination site



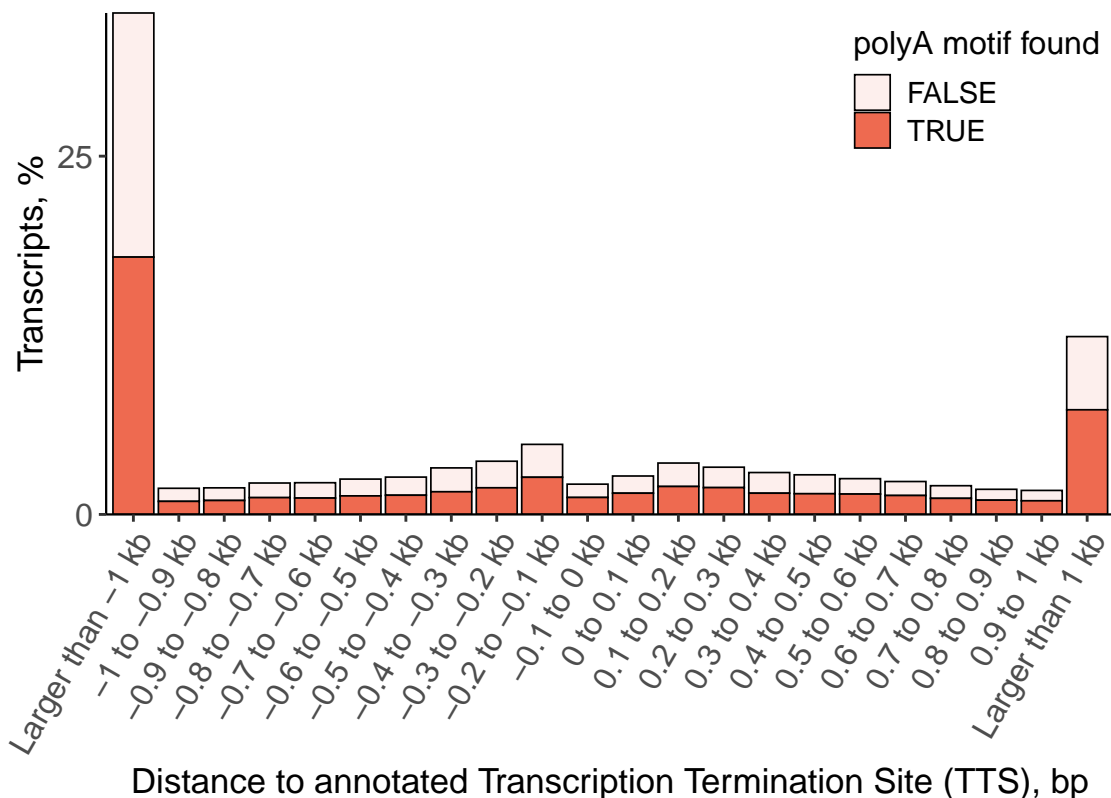
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 3'5'End

Negative values indicate upstream of annotated termination site



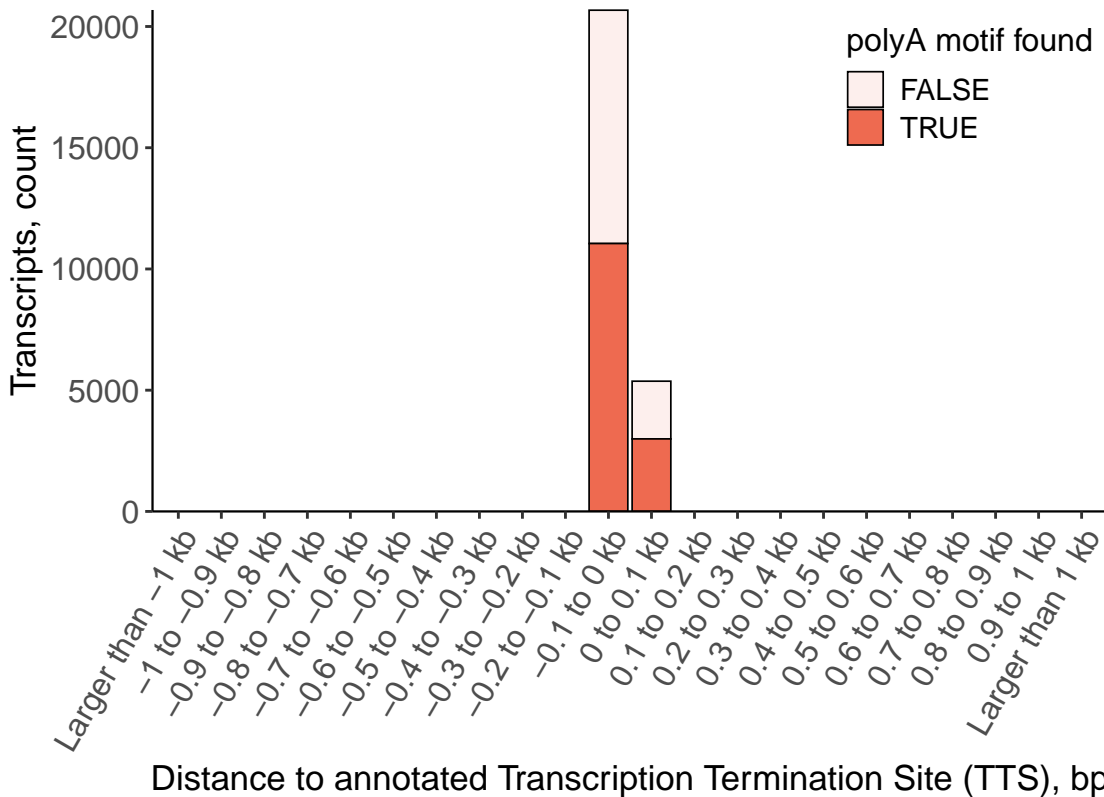
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 3'5'End

Negative values indicate upstream of annotated termination site



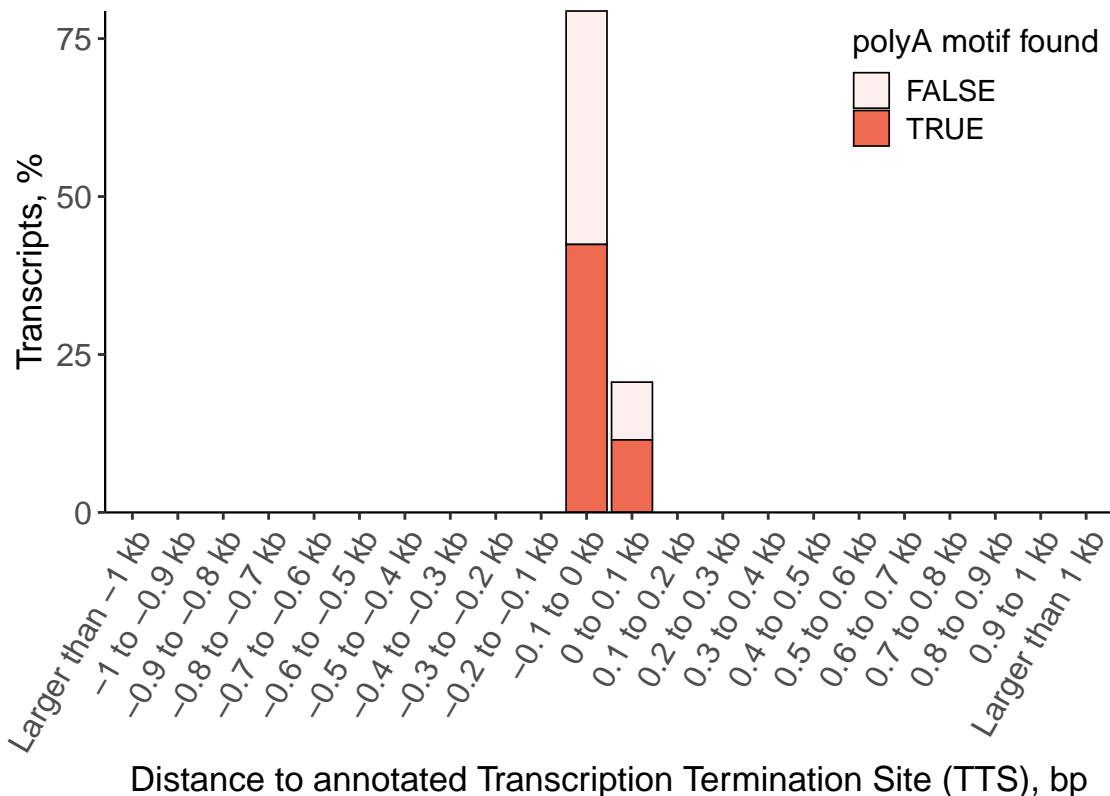
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 5'End

Negative values indicate upstream of annotated termination site



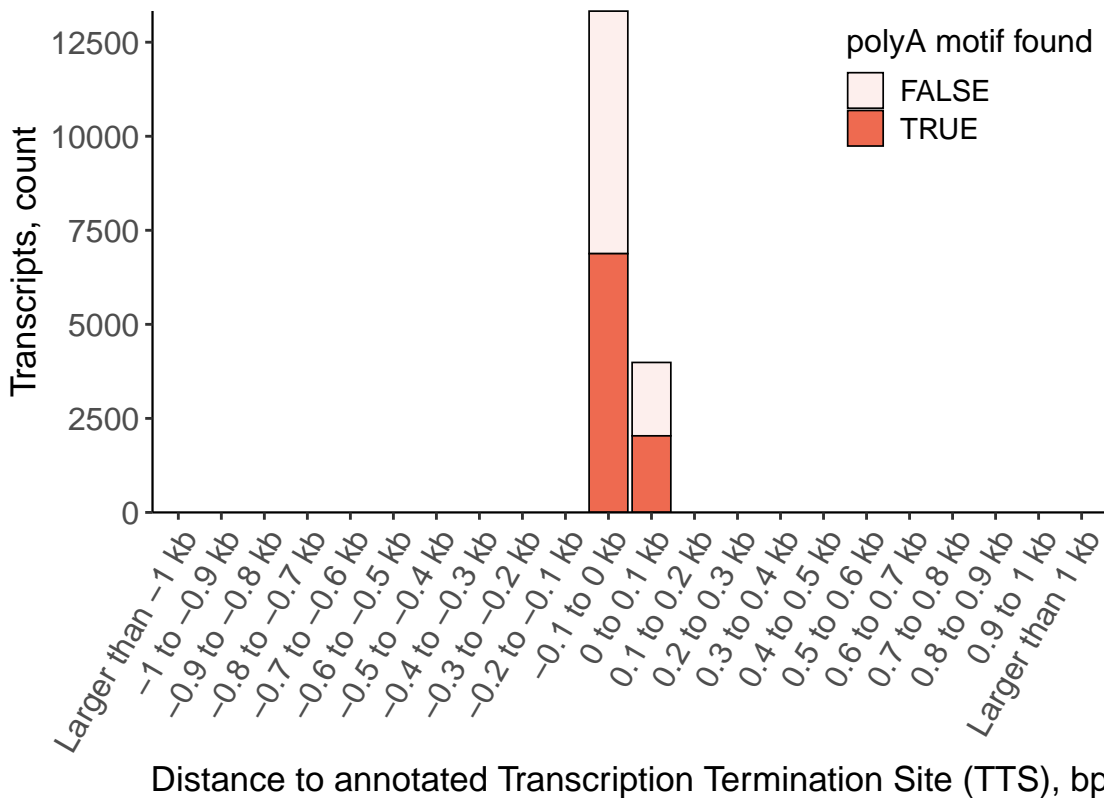
# Distance to annotated Transcription Termination Site (TTS) FSM Alternative 5'End

Negative values indicate upstream of annotated termination site



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

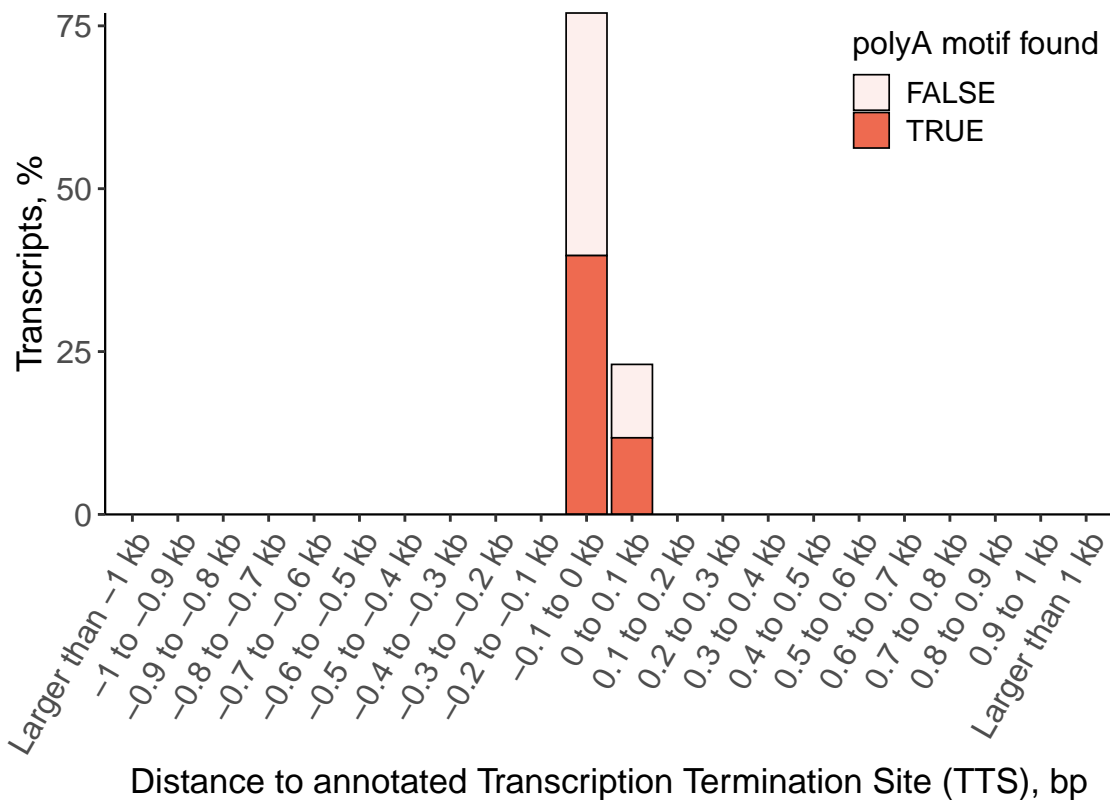
Negative values indicate upstream of annotated termination site





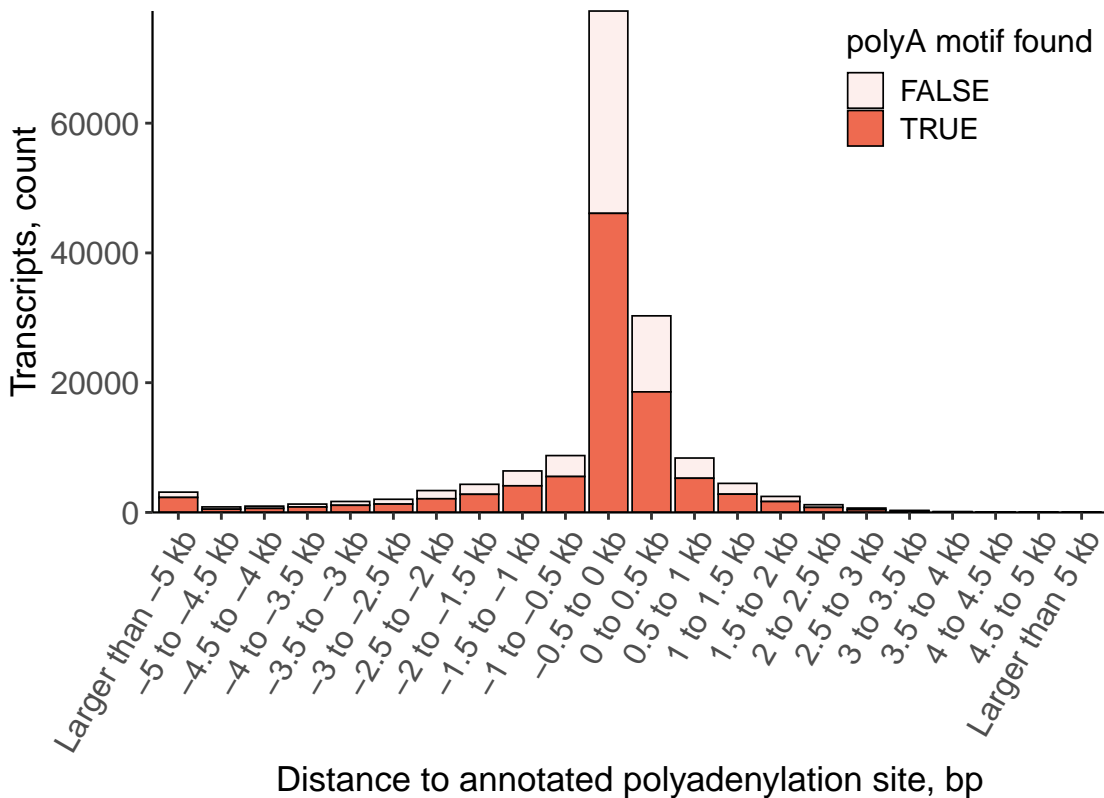
# Distance to annotated Transcription Termination Site (TTS) FSM Reference Match

Negative values indicate upstream of annotated termination site



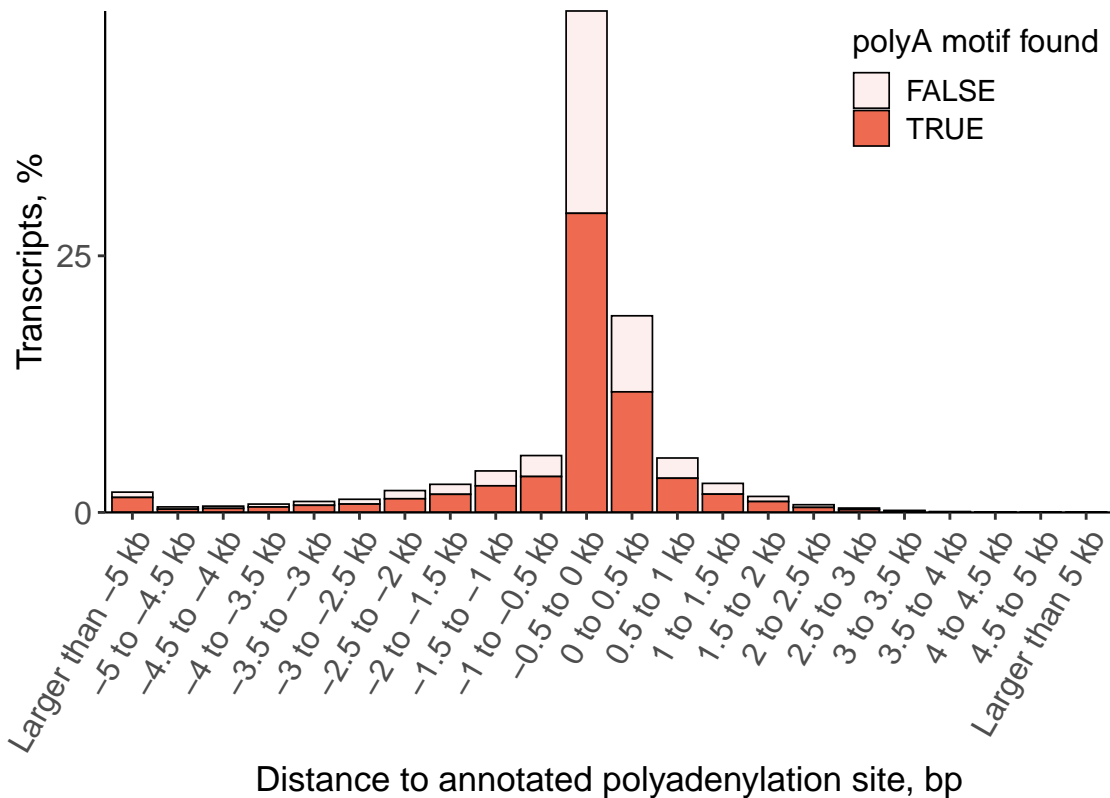
# Distance to Annotated Polyadenylation Site for ISM 3' Fragment

Negative values indicate upstream of annotated polyA site



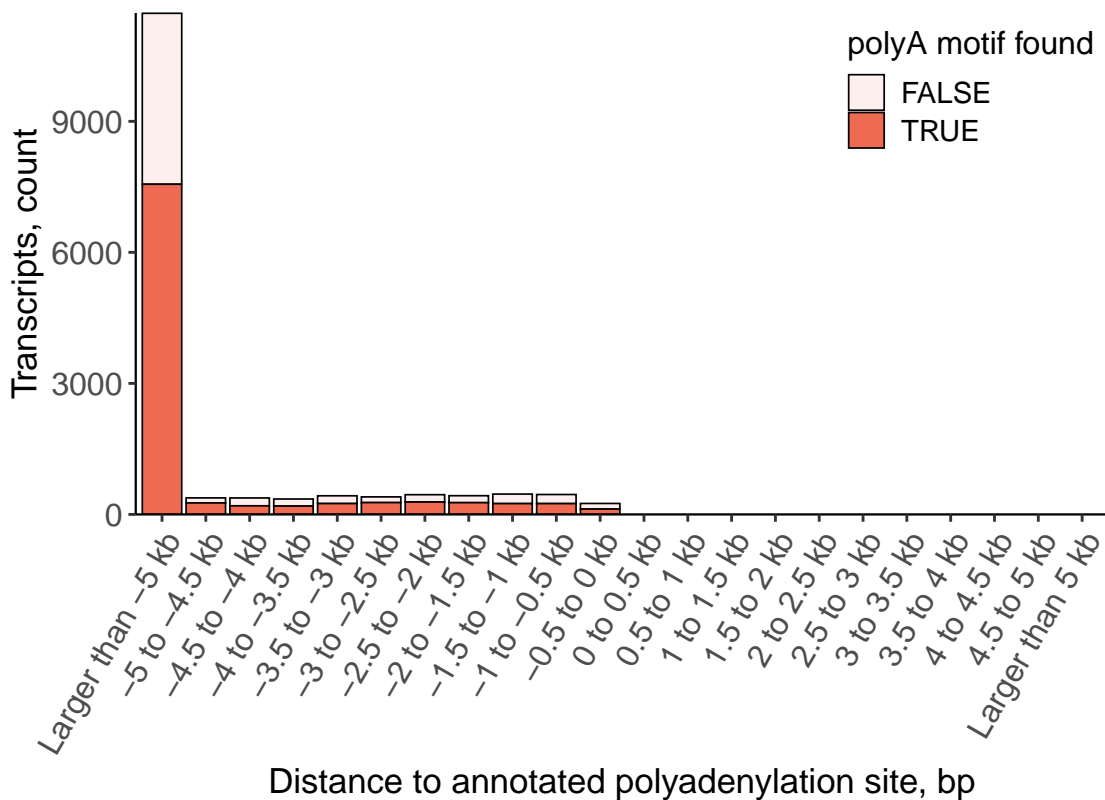
# Distance to Annotated Polyadenylation Site for ISM 3' Fragment

Negative values indicate upstream of annotated polyA site



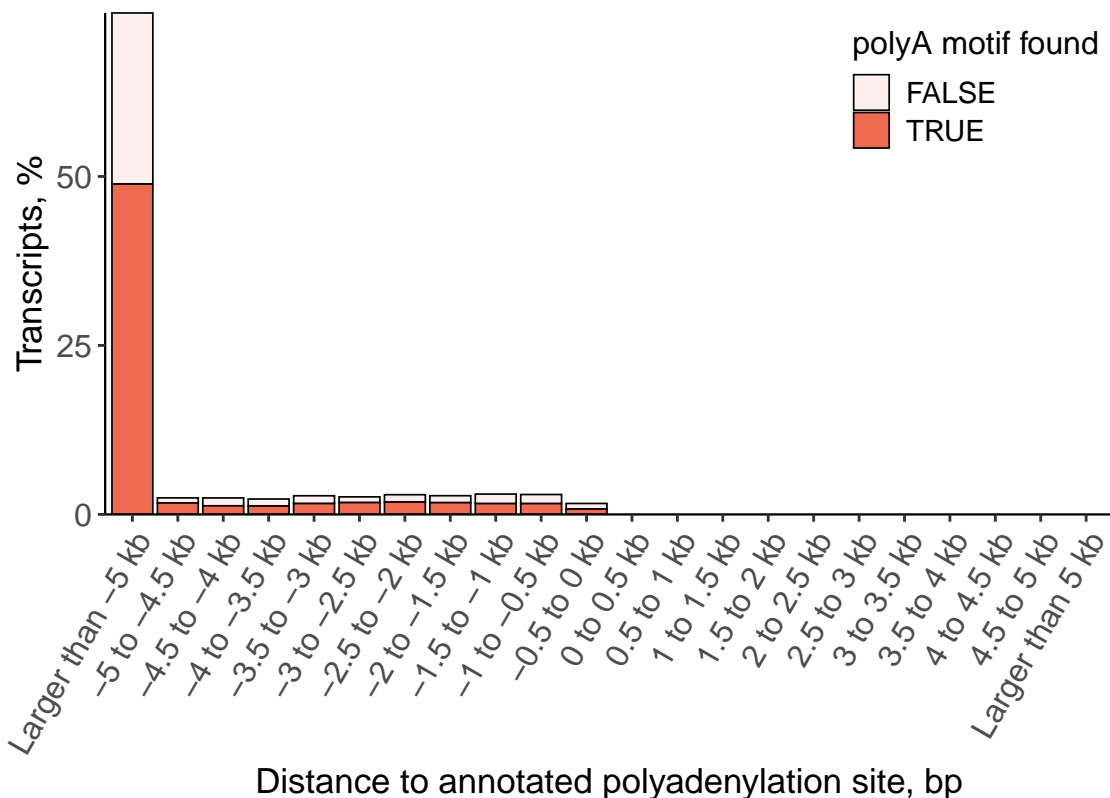
# Distance to Annotated Polyadenylation Site for ISM Internal Fragment

Negative values indicate upstream of annotated polyA site



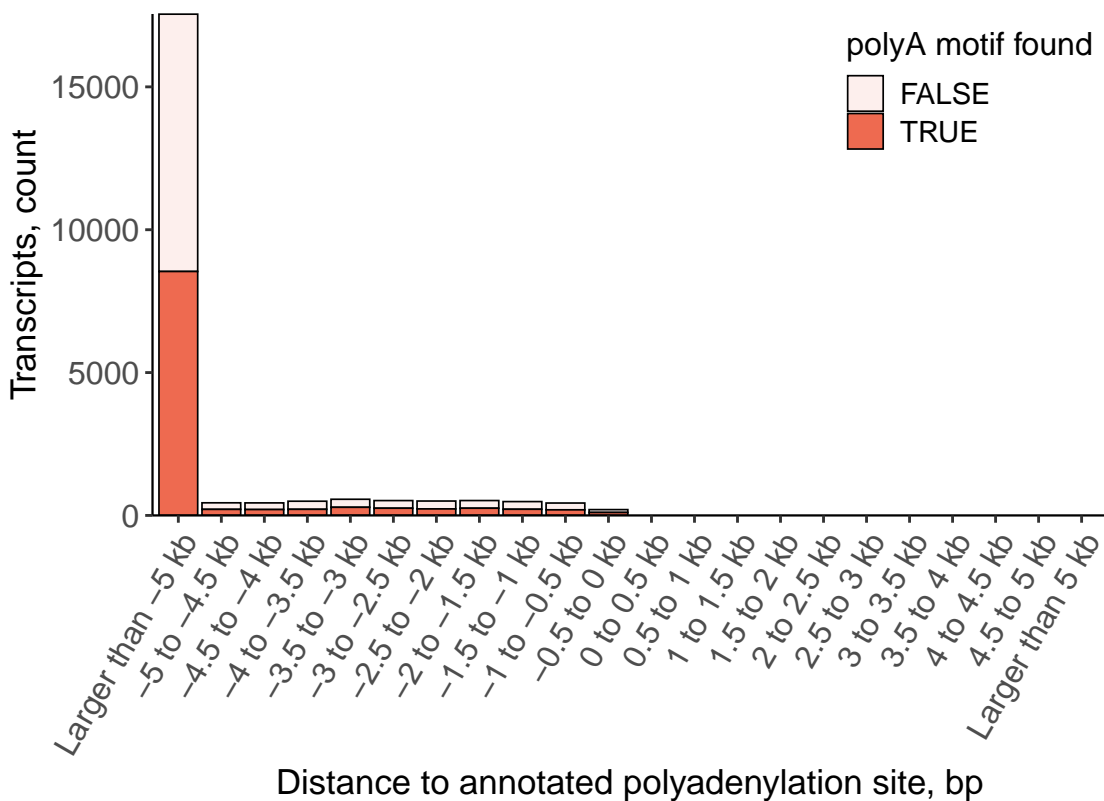
# Distance to Annotated Polyadenylation Site for ISM Internal Fragment

Negative values indicate upstream of annotated polyA site



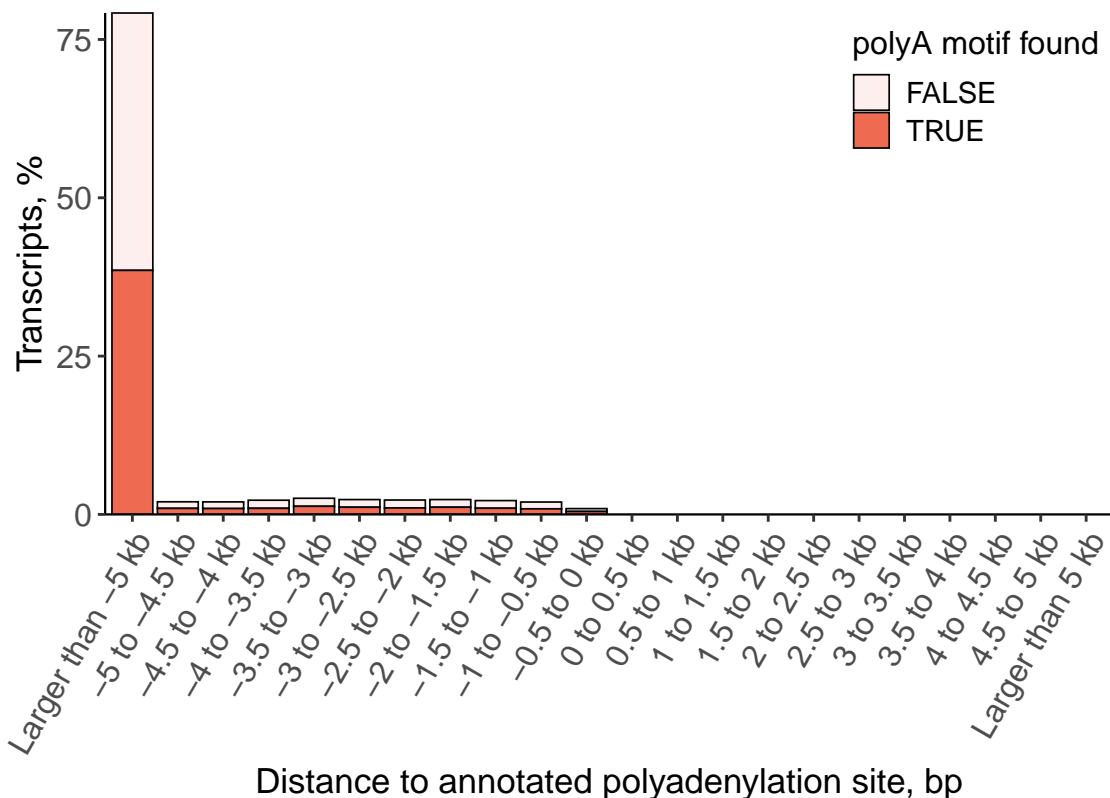
# Distance to Annotated Polyadenylation Site for ISM A5' Fragment

Negative values indicate upstream of annotated polyA site



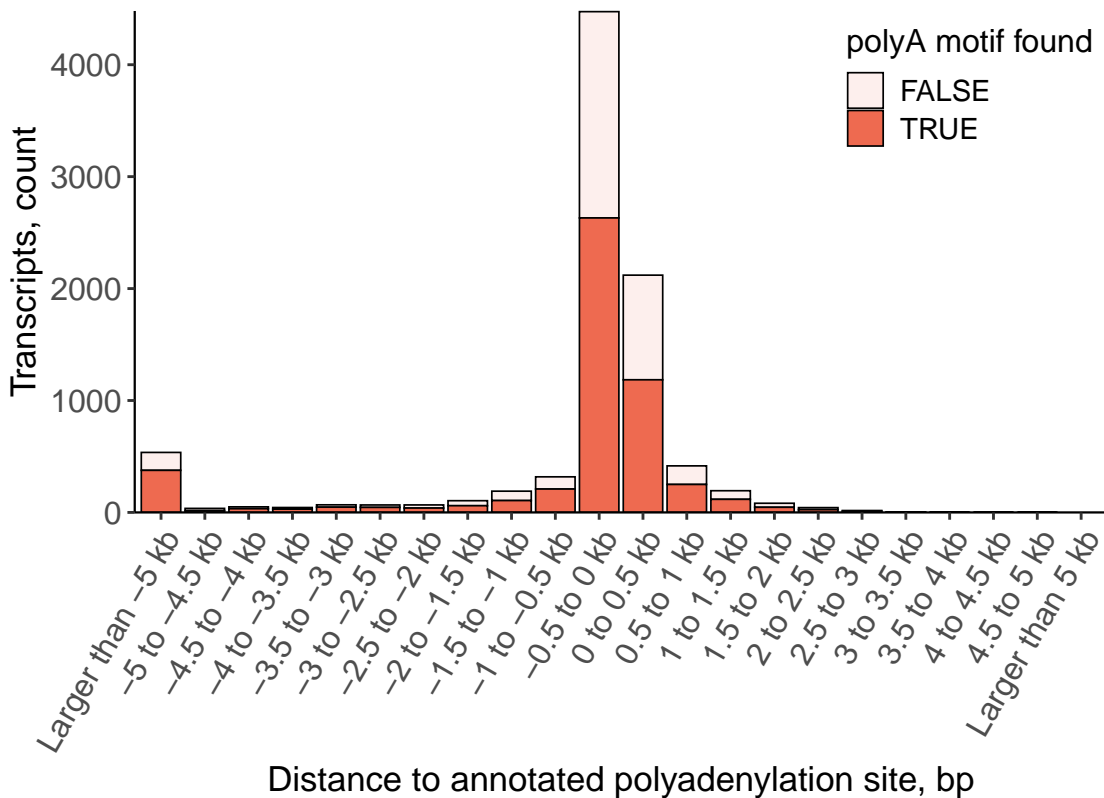
# Distance to Annotated Polyadenylation Site for ISM A5' Fragment

Negative values indicate upstream of annotated polyA site



# Distance to Annotated Polyadenylation Site for ISM Intron Retention

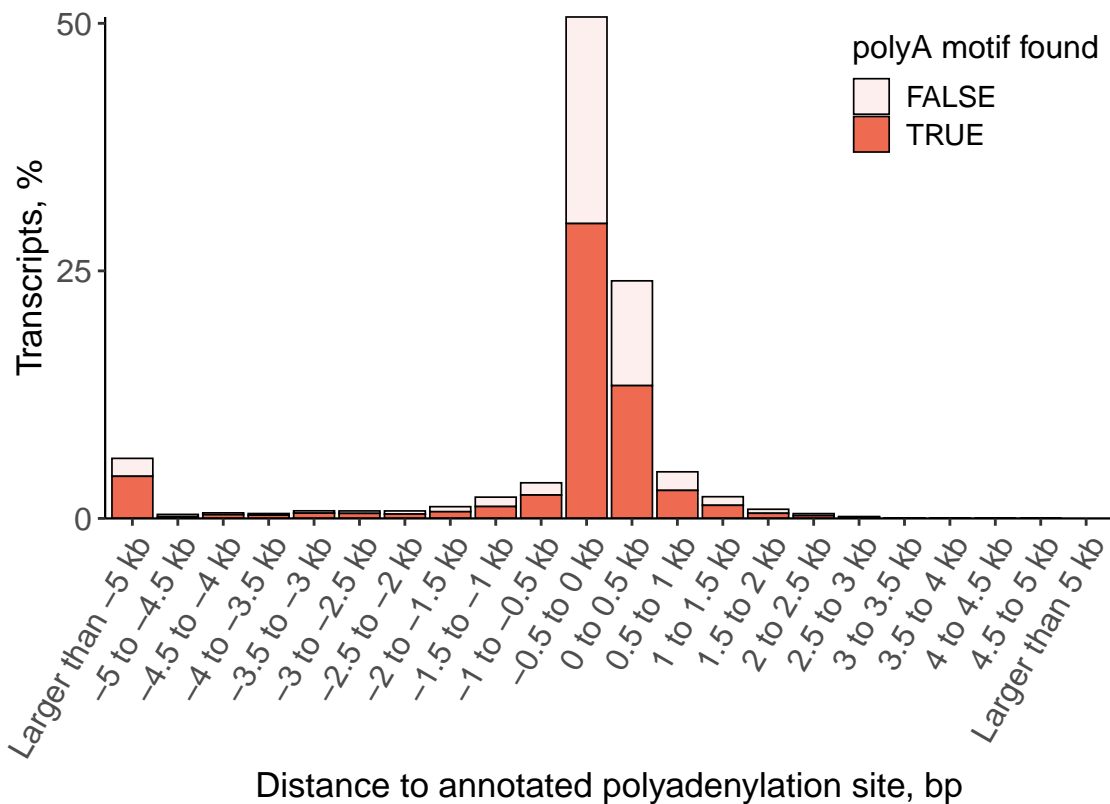
Negative values indicate upstream of annotated polyA site





# Distance to Annotated Polyadenylation Site for ISM Intron Retention

Negative values indicate upstream of annotated polyA site



## *PolyA Distance Analysis*



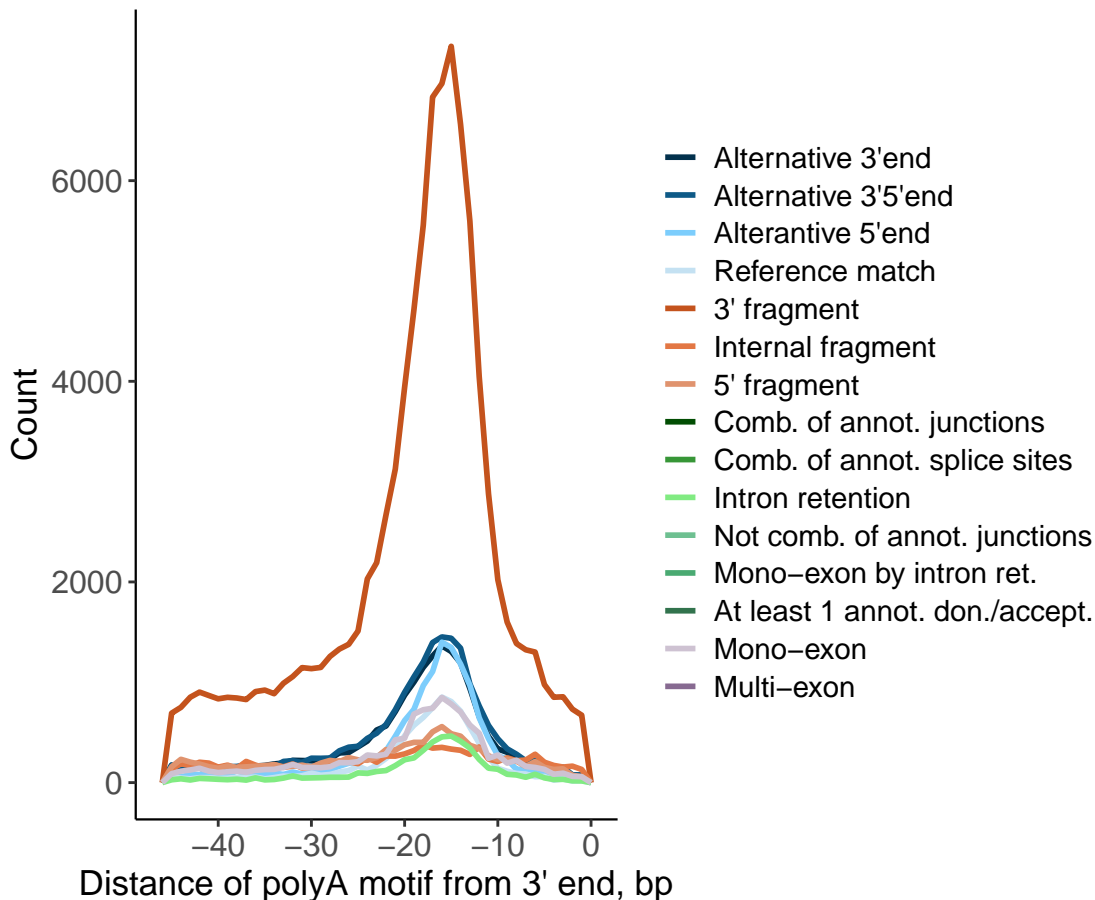
## Frequency of PolyA Motifs

Motif	Count	%
AATAAA	152377	49.5
ATTAAA	43997	14.3
TATAAA	12496	4.1
AAAAAG	11458	3.7
AGTAAA	11458	3.7
AAGAAA	11445	3.7
GGGGCT	9293	3.0
TTTAAA	8372	2.7
AATATA	6907	2.2
CATAAA	6883	2.2
GATAAA	6781	2.2
AATACA	6680	2.2
AAAACA	6285	2.0
AATGAA	5997	1.9
AATAGA	3717	1.2
ACTAAA	3442	1.1

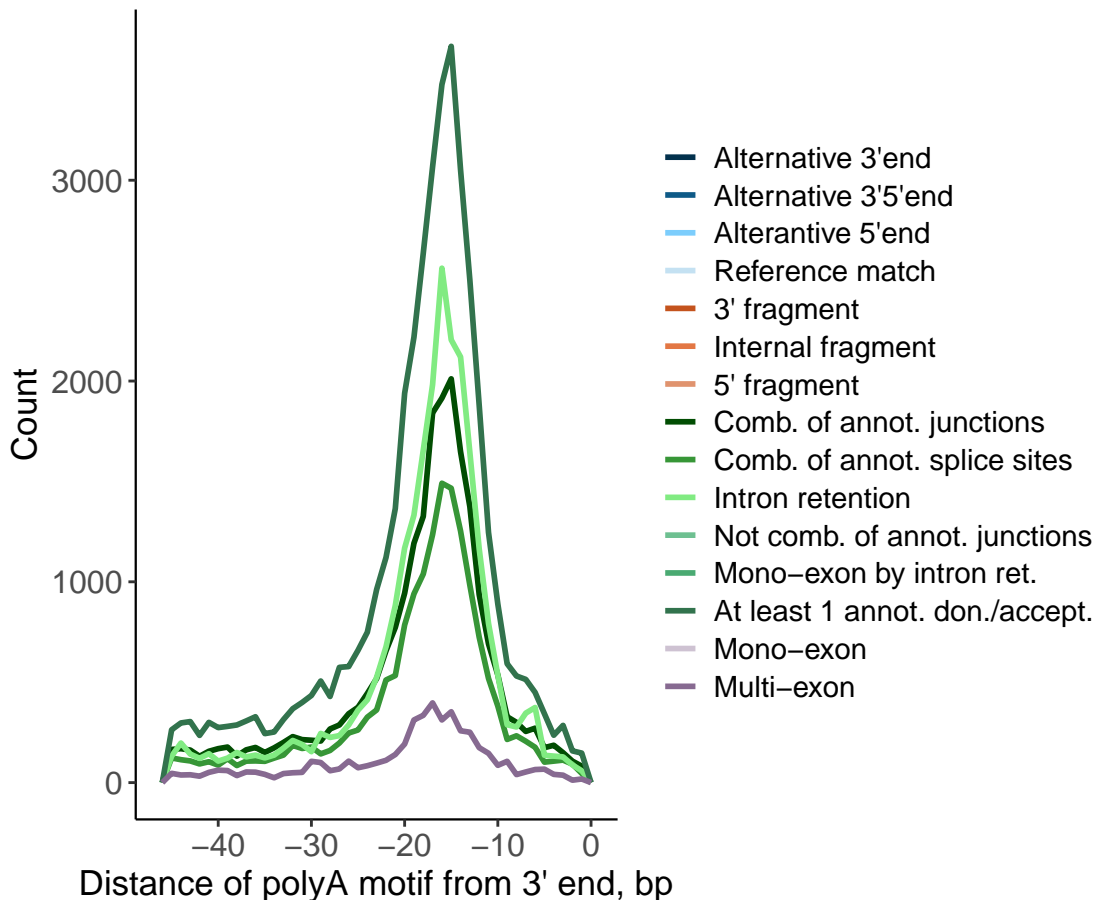
## Number of polyA Motifs Detected

Category	Count	polyA Detected	%
FSM	138528	73009	53
ISM	204661	123098	60
NIC	114899	59060	51
NNC	88389	47218	53
Genic Genomic	704	382	54
Antisense	2035	1197	59
Fusion	2792	1635	59
Intergenic	3041	1947	64
NA	64	42	66

## 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	38302	18359	48
Alternative 3'5'end	37189	19838	53
Alterantive 5'end	26044	14038	54
Reference match	17313	8915	51
3' fragment	158204	97262	61
Internal fragment	15466	9903	64
5' fragment	22156	10703	48
Comb. of annot. junctions	45083	22911	51
Comb. of annot. splice sites	31903	16791	53
Intron retention	58152	30440	52
At least 1 annot. don./accept.	77556	41657	54
Mono-exon	19680	11859	60
Multi-exon	8012	4879	61
NA	53	33	62

## *Frequency of PolyA Motifs*

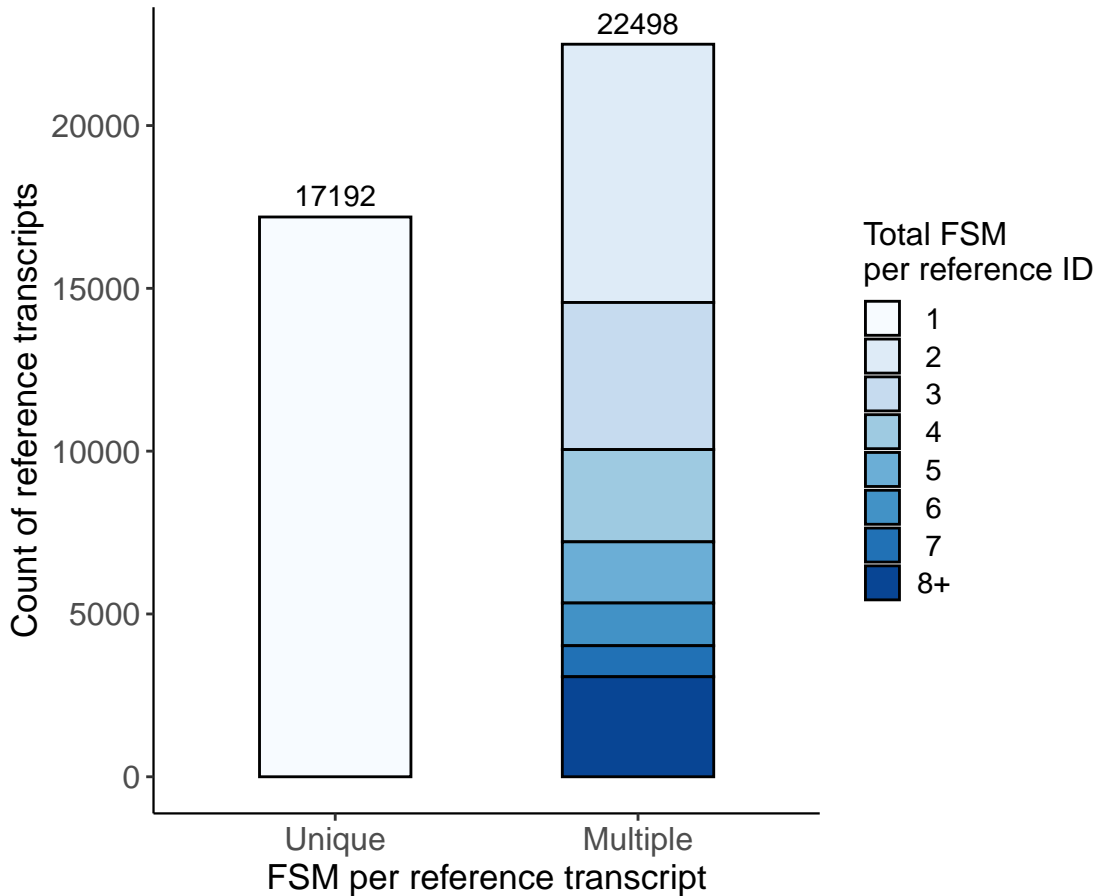
<b>Motif</b>	<b>Count</b>	<b>%</b>
AATAAA	152377	49.5
ATTAAA	43997	14.3
TATAAA	12496	4.1
AAAAAG	11458	3.7
AGTAAA	11458	3.7
AAGAAA	11445	3.7
GGGGCT	9293	3.0
TTTAAA	8372	2.7
AATATA	6907	2.2
CATAAA	6883	2.2
GATAAA	6781	2.2
AATACA	6680	2.2
AAAACA	6285	2.0
AATGAA	5997	1.9
AATAGA	3717	1.2
ACTAAA	3442	1.1



## *Redundancy Analysis*

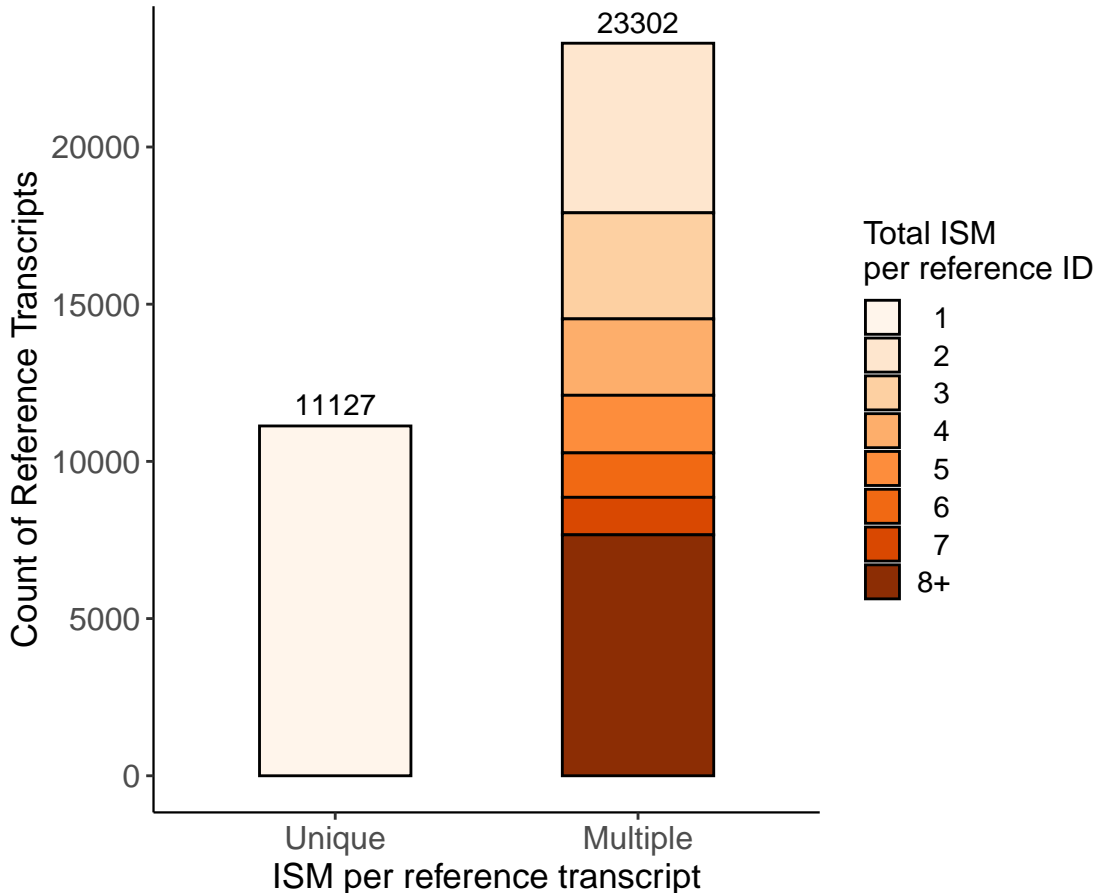
# Reference Transcript Redundancy

Only FSM

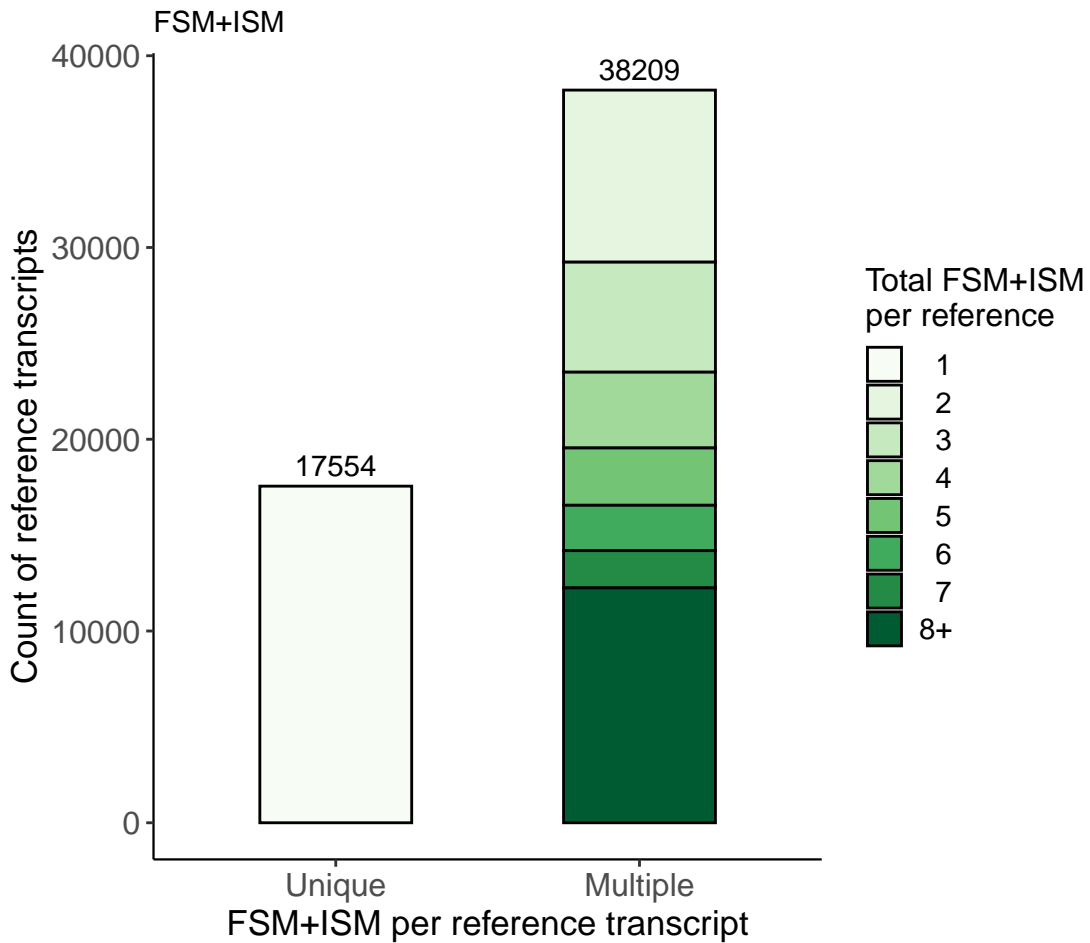


# Reference Transcript Redundancy

Only ISM

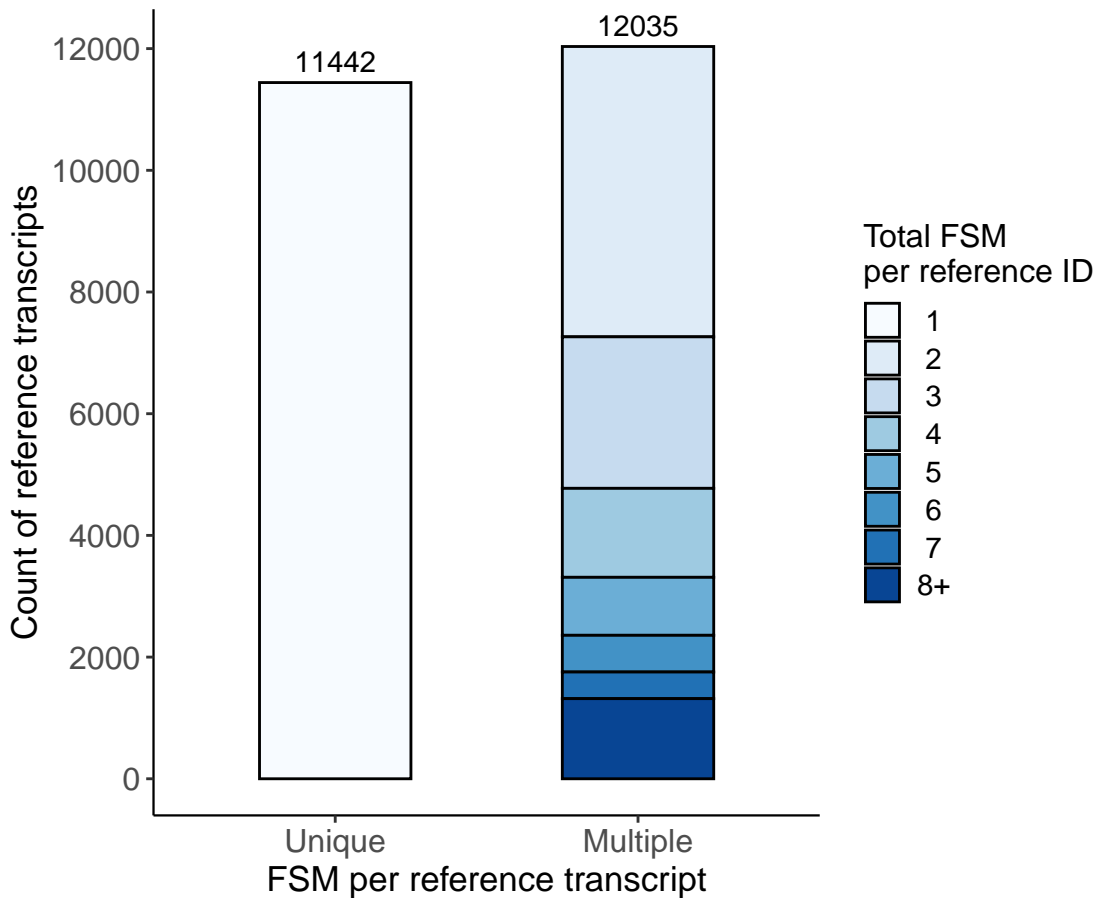


# Reference Transcript Redundancy



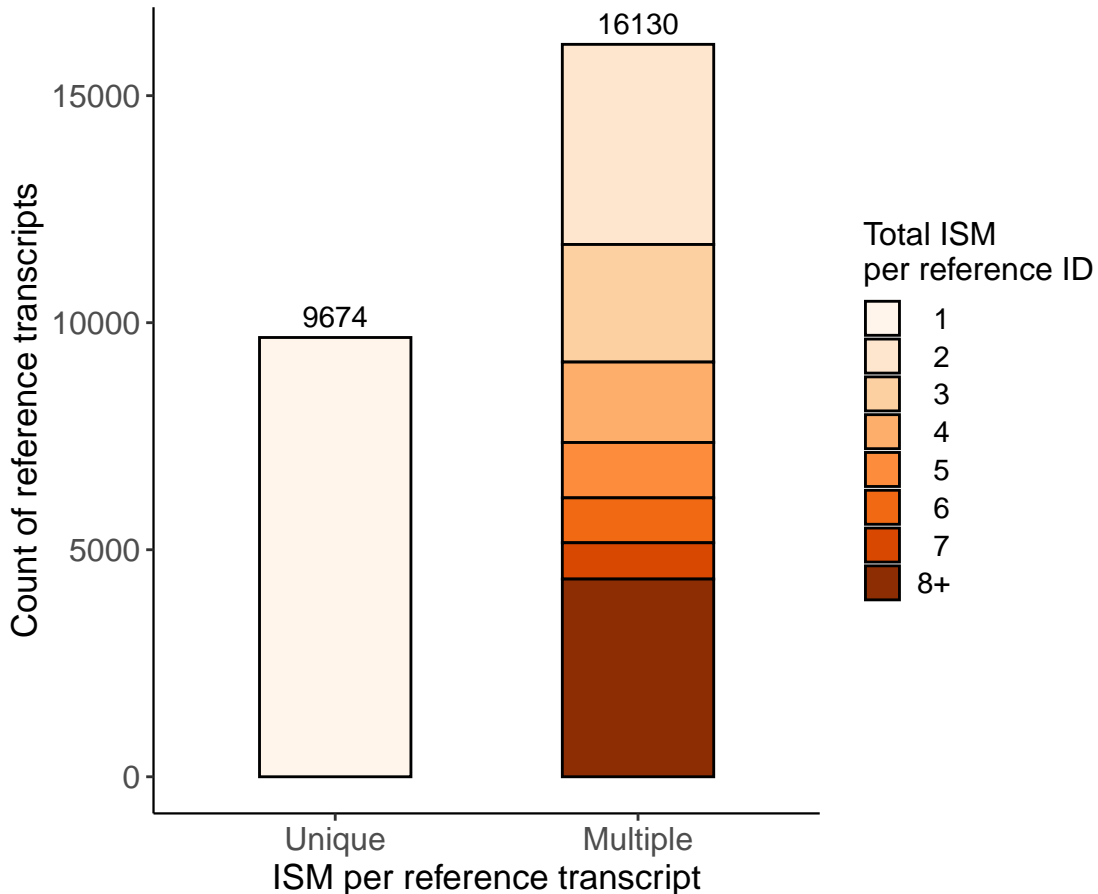
# Reference Transcript Redundancy

Only FSM with a polyA motif found



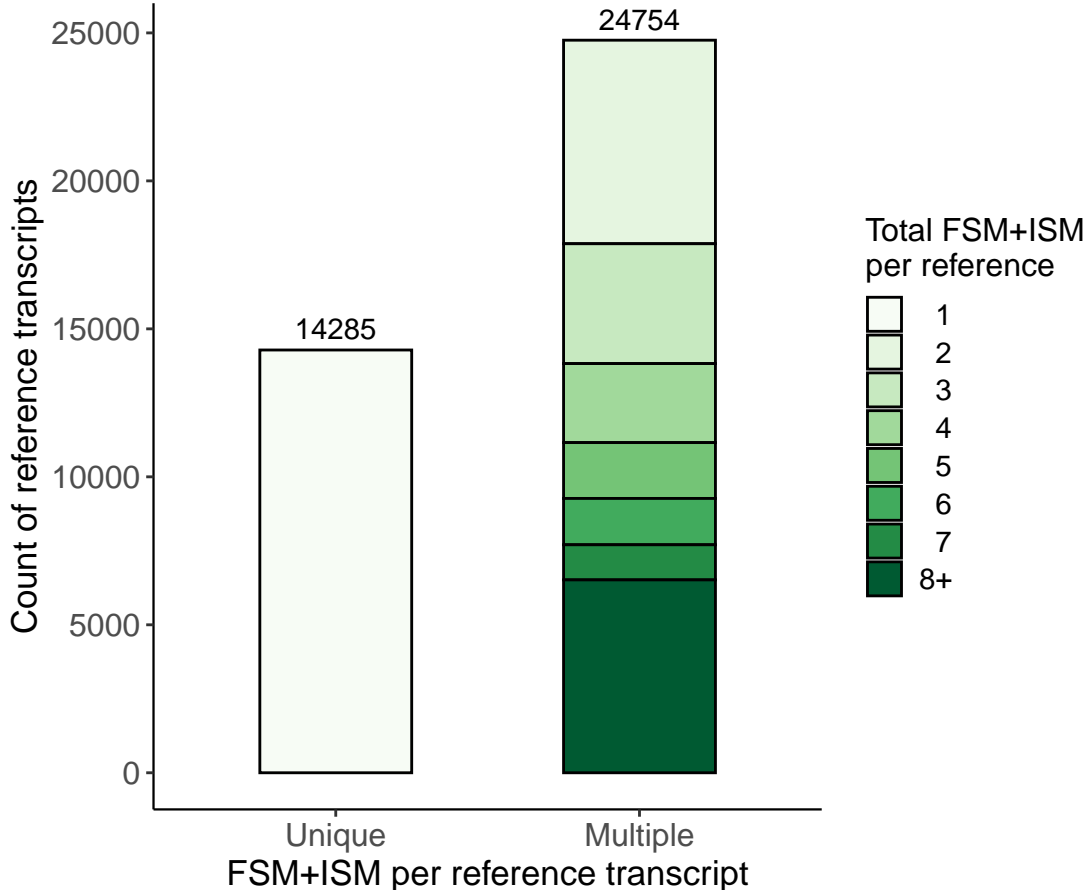
# Reference Transcript Redundancy

Only ISM with a polyA motif found



# Reference Transcript Redundancy

FSM+ISM with a polyA motif found

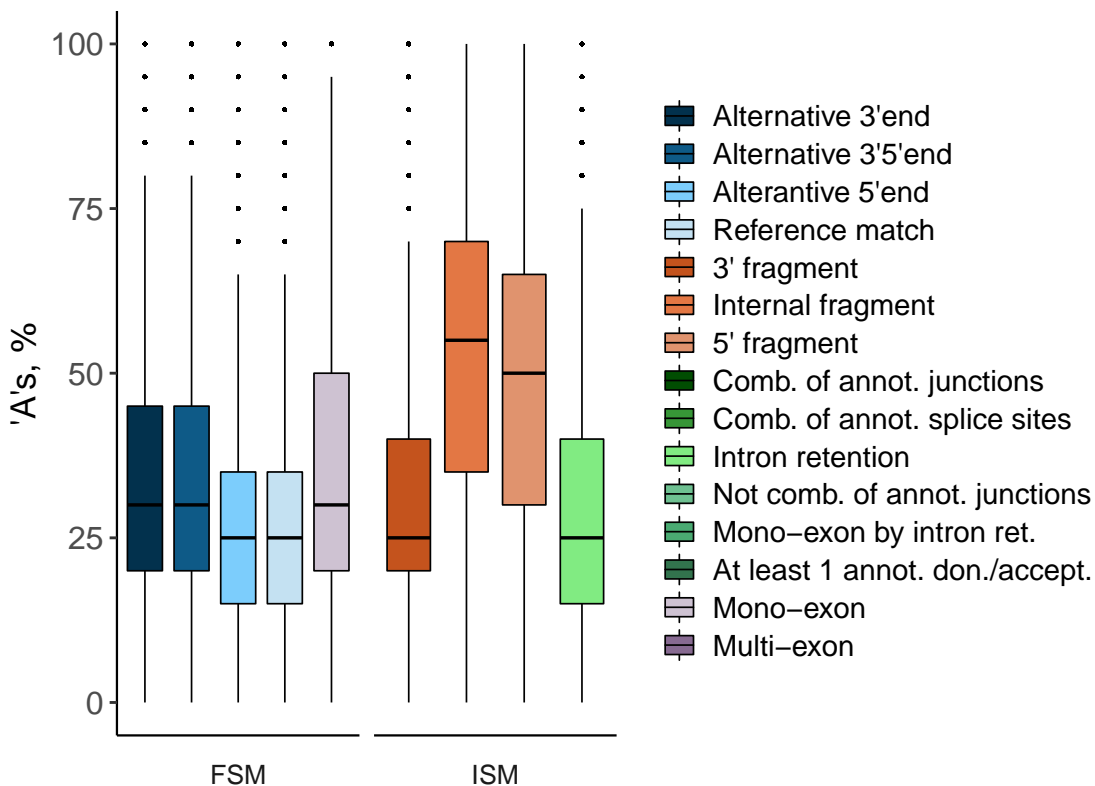


*Intra-Priming Quality Check*



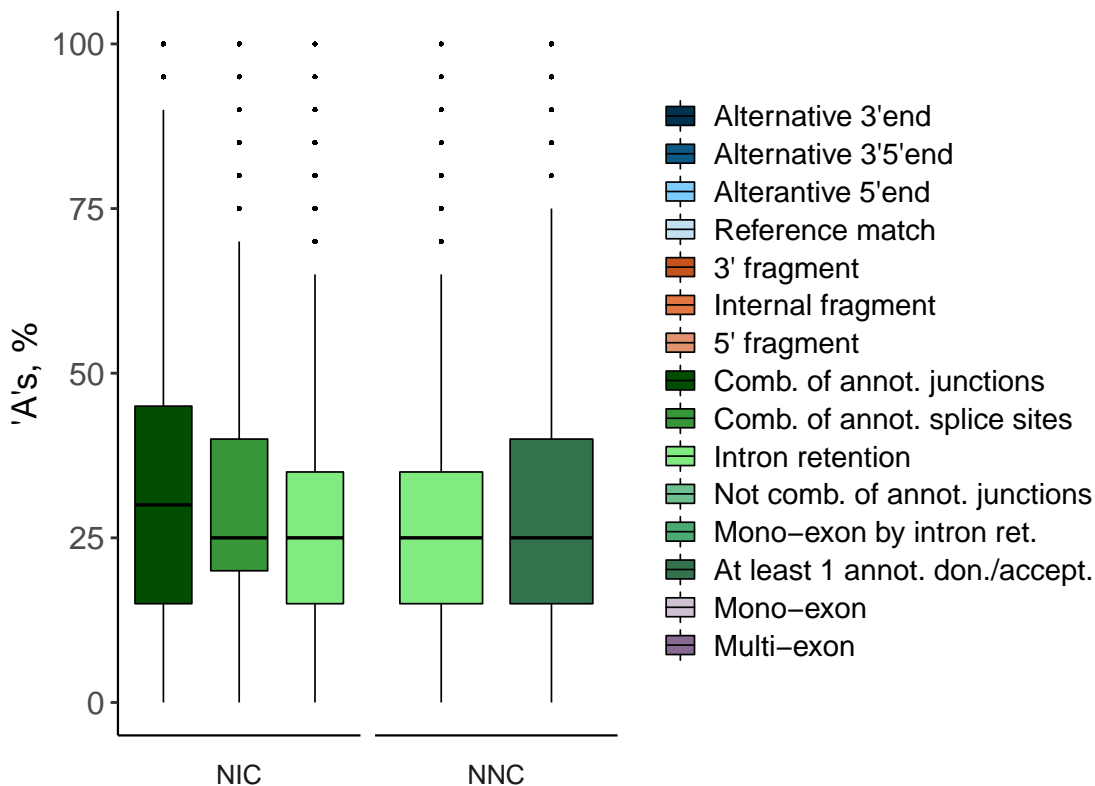
# Possible Intra-Priming by Structural Category

Percent of genomic 'A's in downstream 20 bp



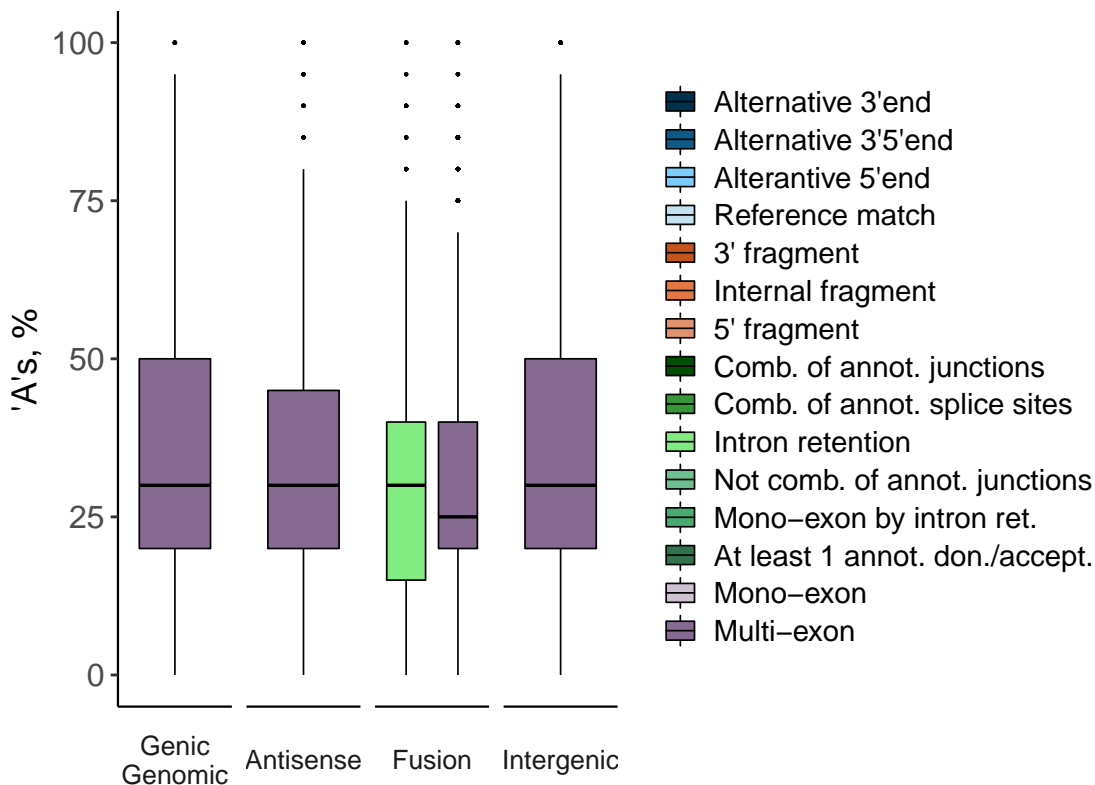
# Possible Intra-Priming by Structural Category

Percent of genomic 'A's in downstream 20 bp



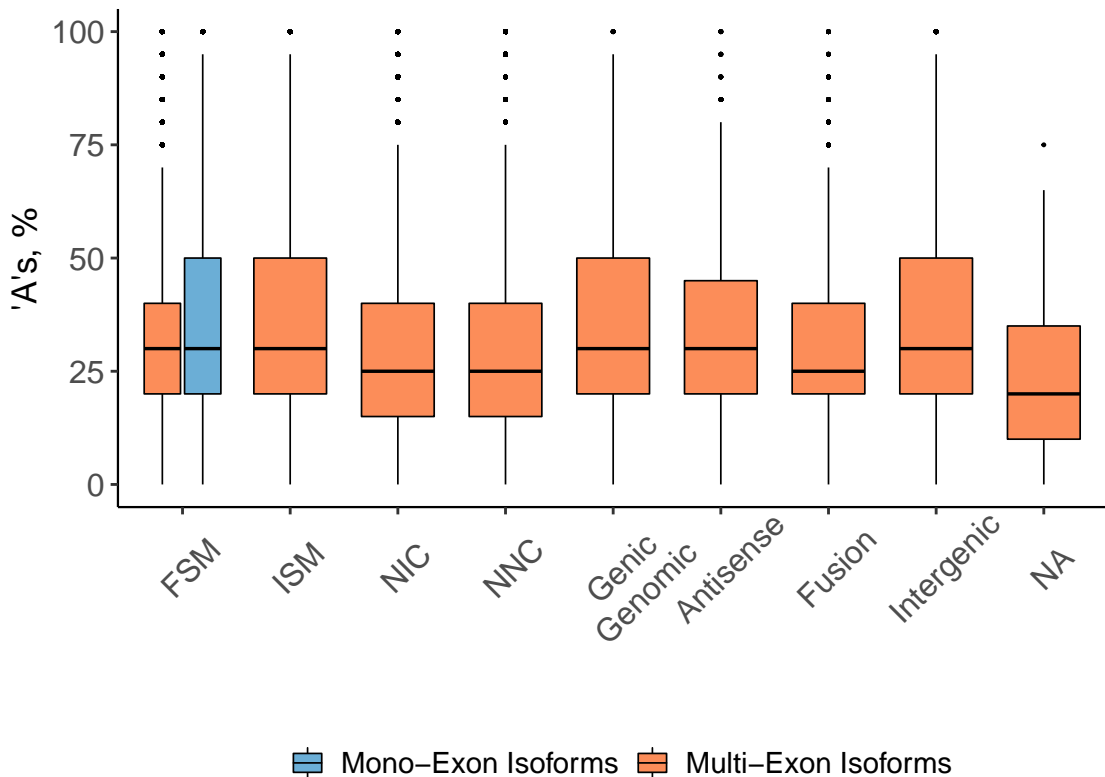
# Possible Intra-Priming by Structural Category

Percent of genomic 'A's in downstream 20 bp



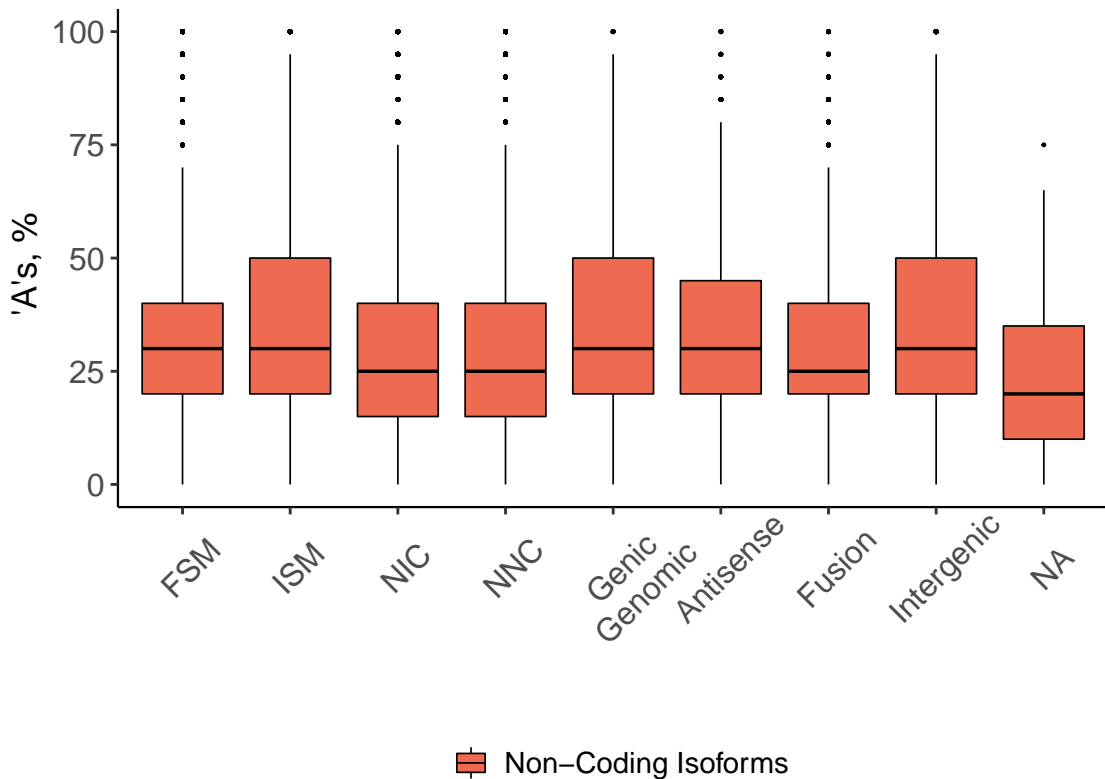
# Mono- vs Multi-Exon Possible Intra-Priming

Percent of genomic 'A's in downstream 20 bp



# Coding vs Non-Coding Possible Intra-Priming

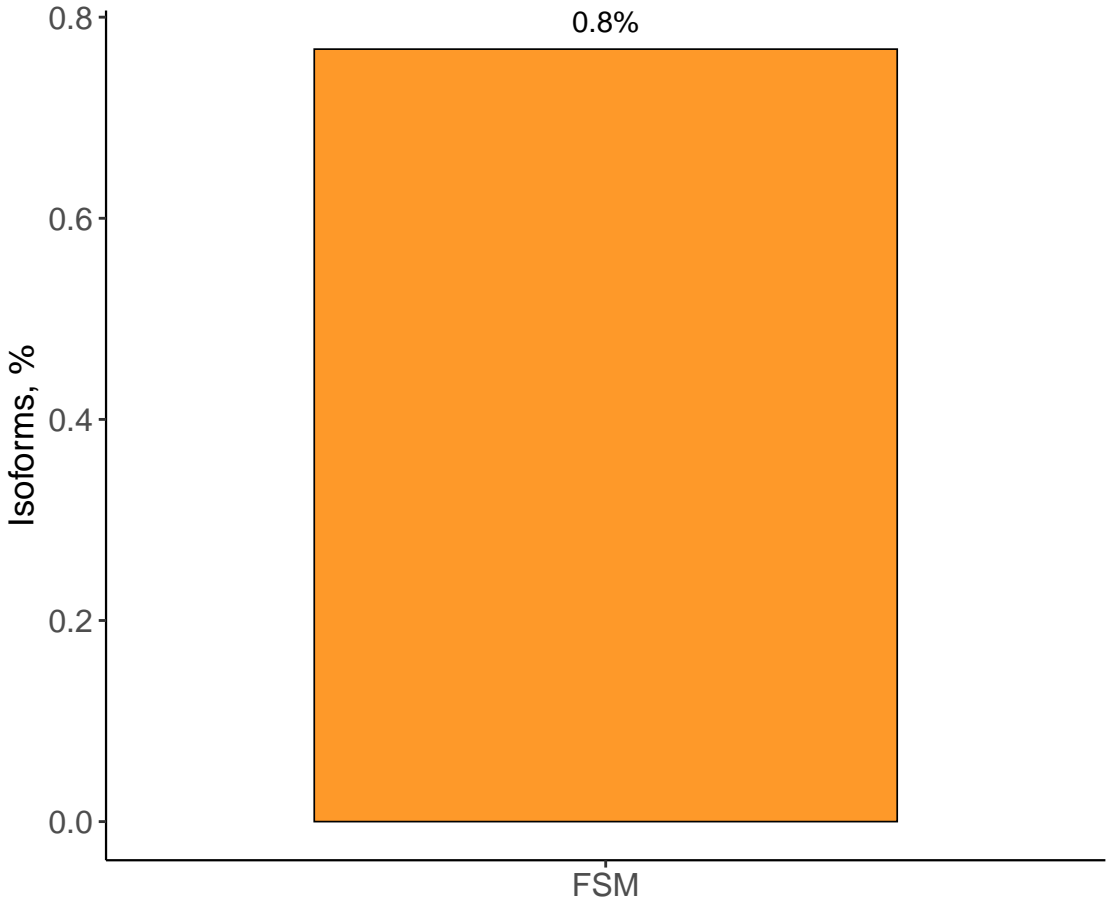
Percent of genomic 'A's in downstream 20 bp



*Features of Bad Quality*

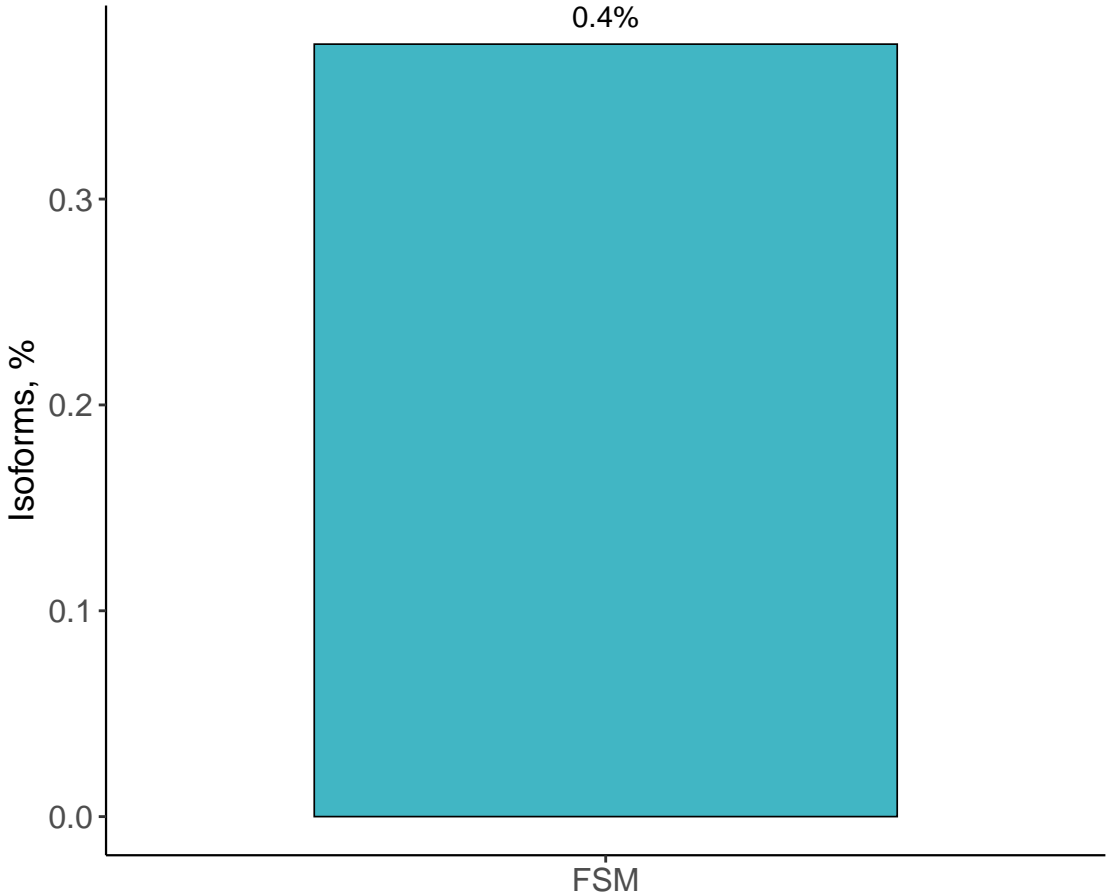
# RT-switching

0.8%



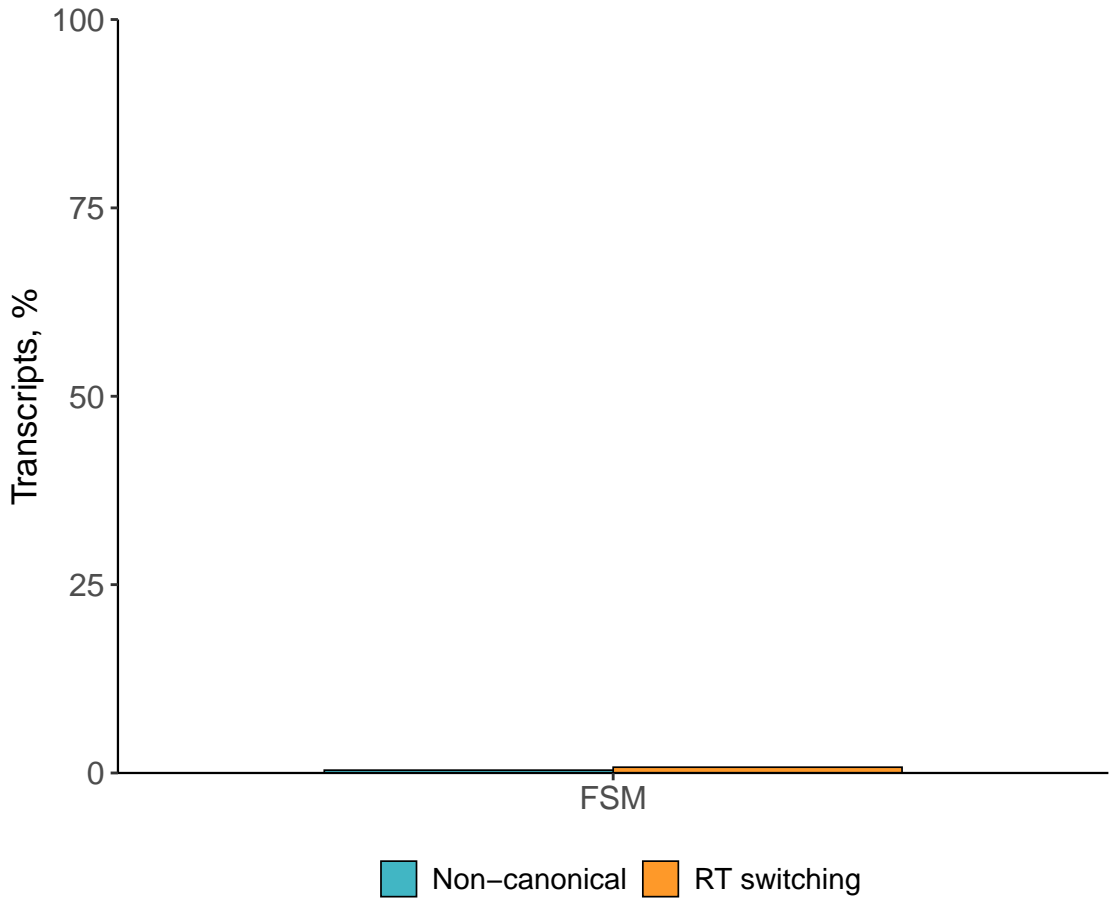
# Non-Canonical Junctions

0.4%



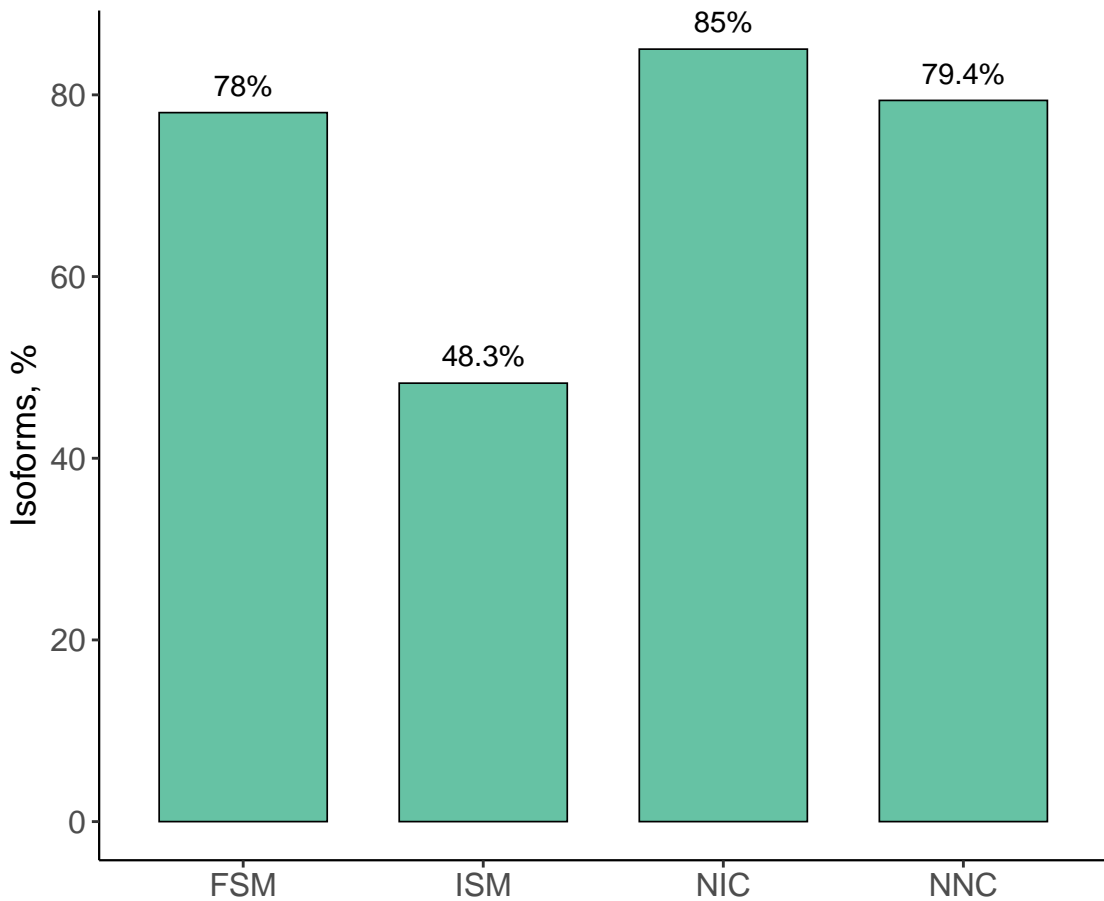


# Quality Control Attributes Across Structural Categories

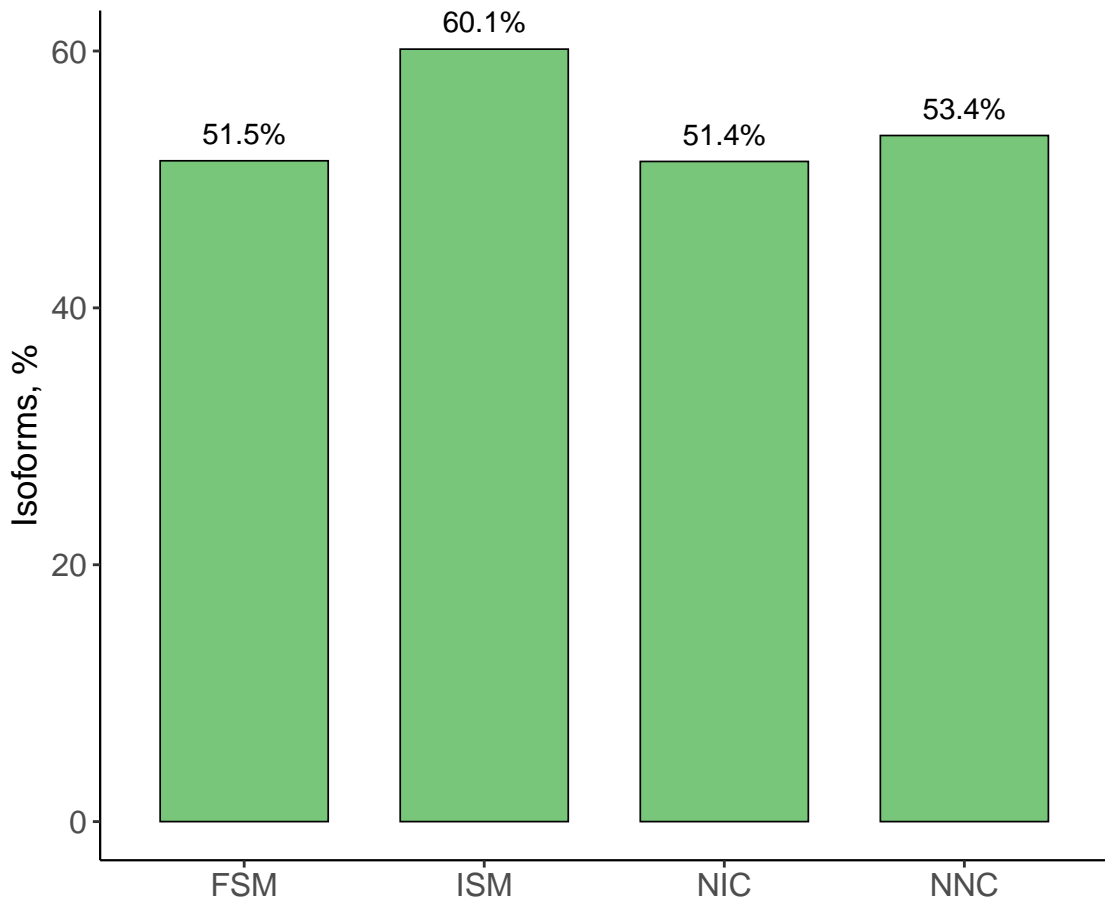


## *Features of Good Quality*

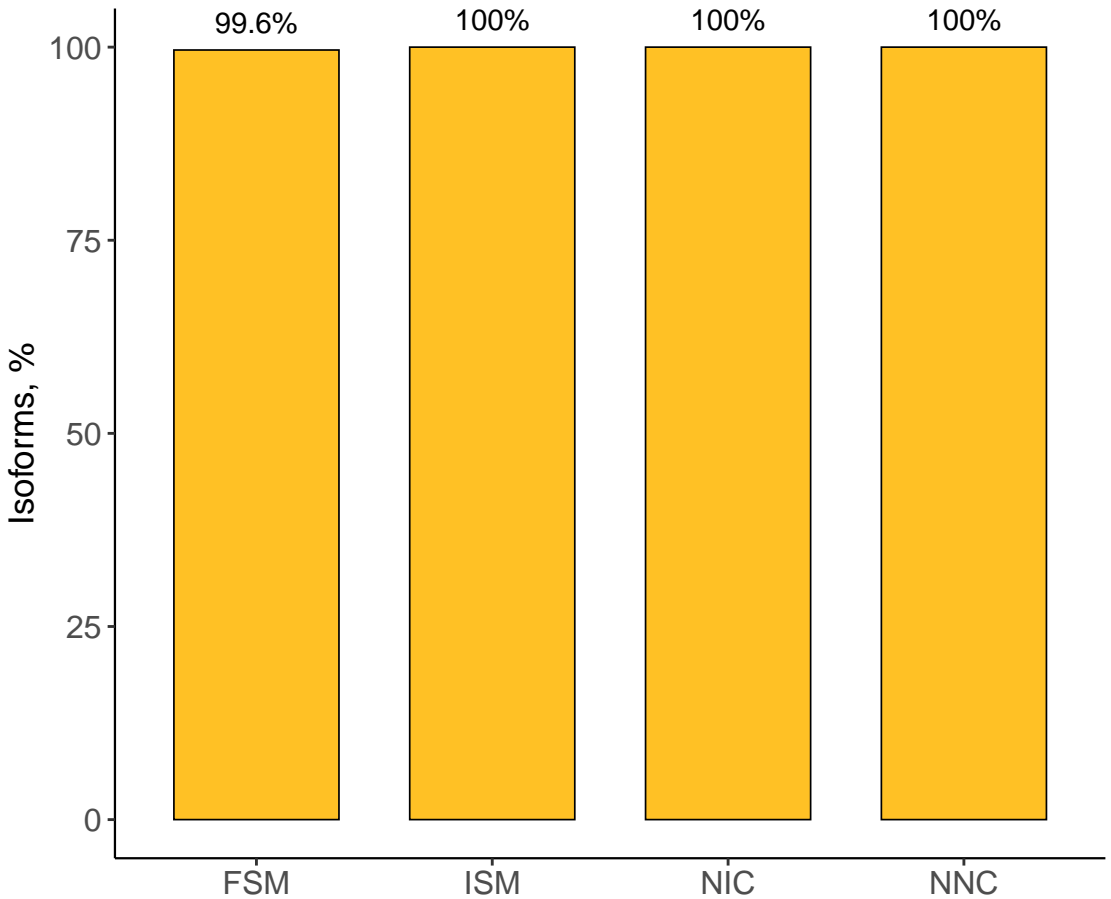
## Annotation Support



## PolyA Support



## All Canonical Junctions



## Good Quality Control Attributes Across Structural Categories

