SQANTI3 filter report

Total Genes: 651

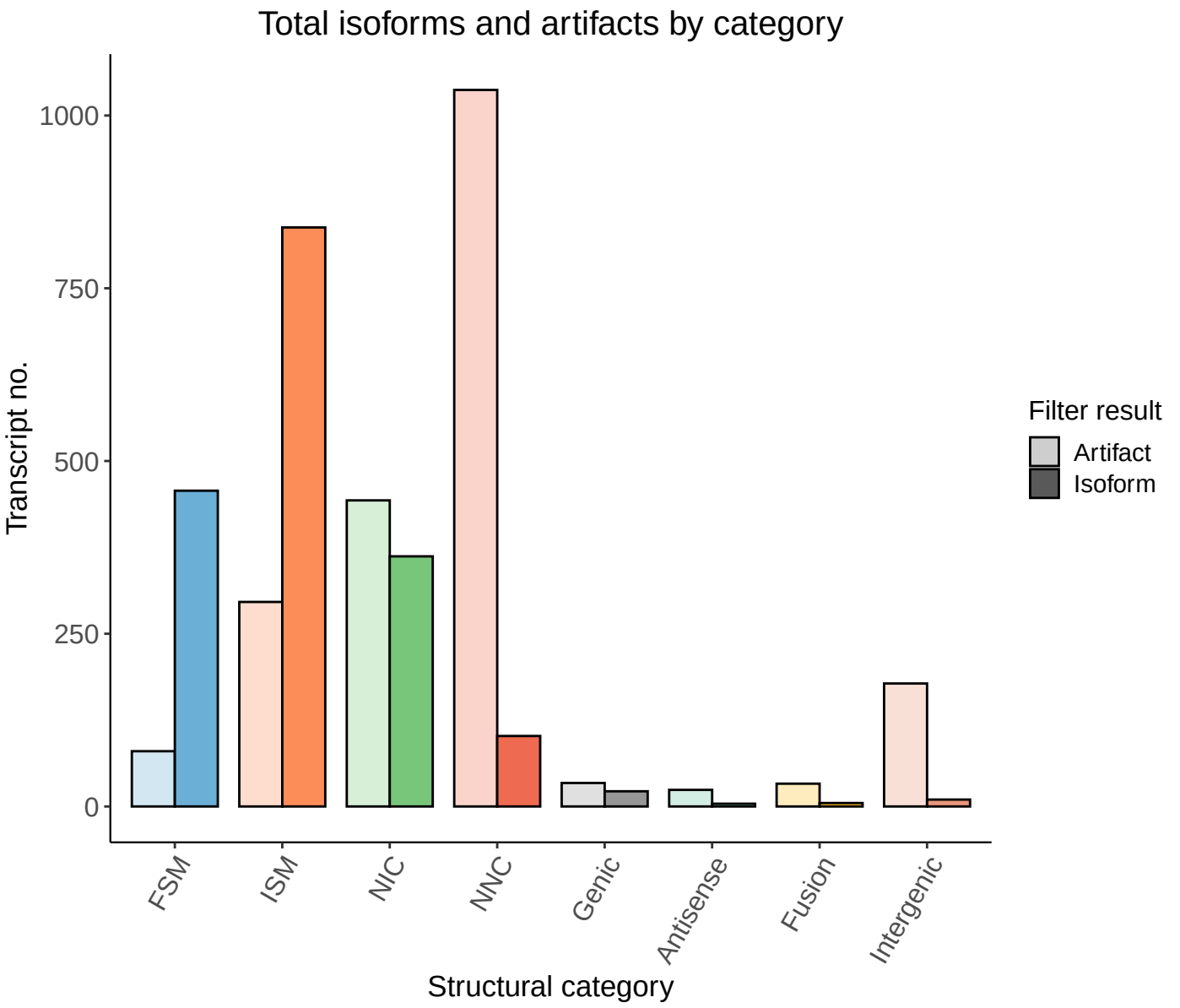
Total Transcripts: 3925

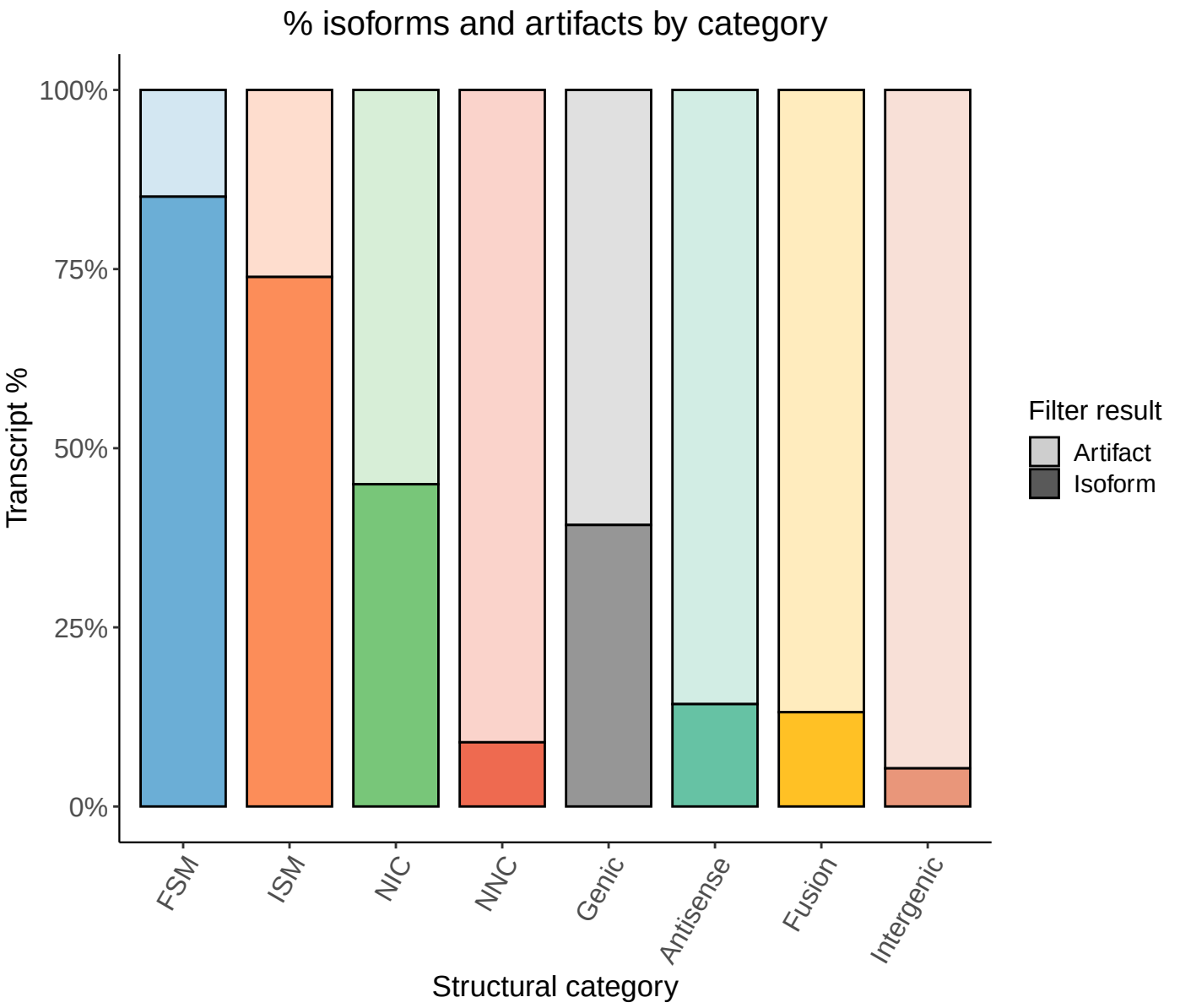
− Isoforms: 1800 (46%)

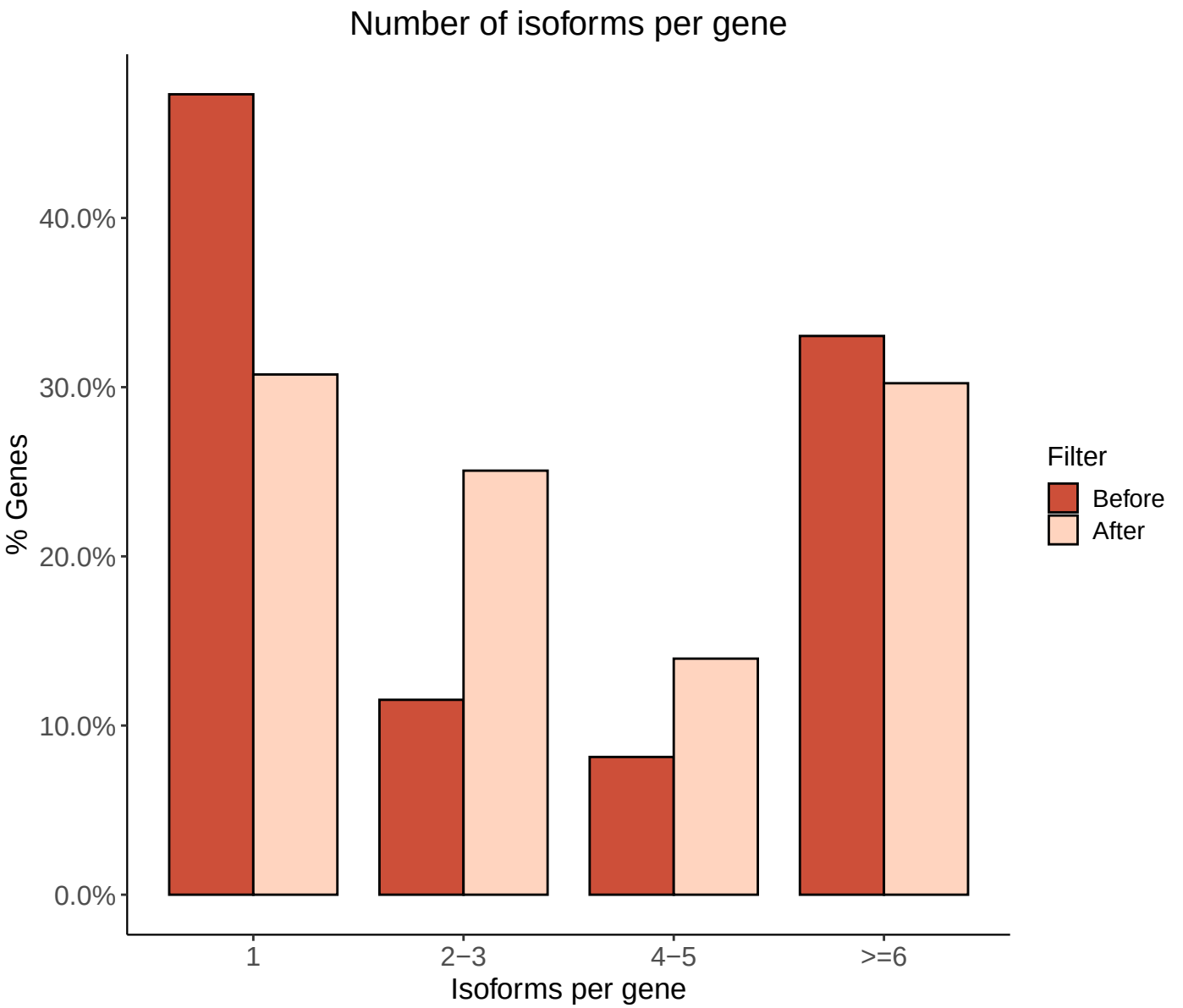
− Artifacts: 2125 (54%)

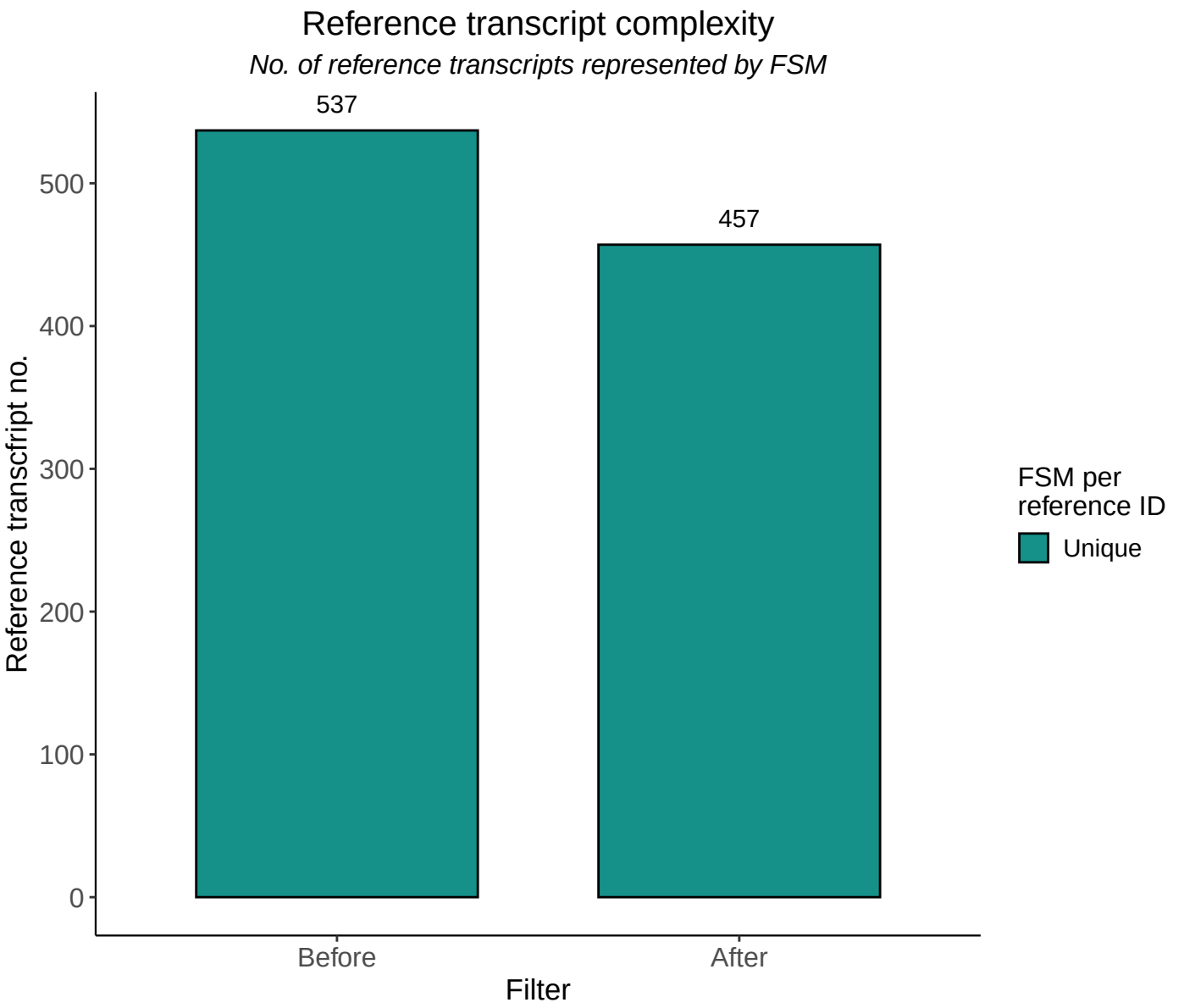
Gene category	Gene no.	No. of genes with artifacts only
Annotated	437	64
Novel	214	200

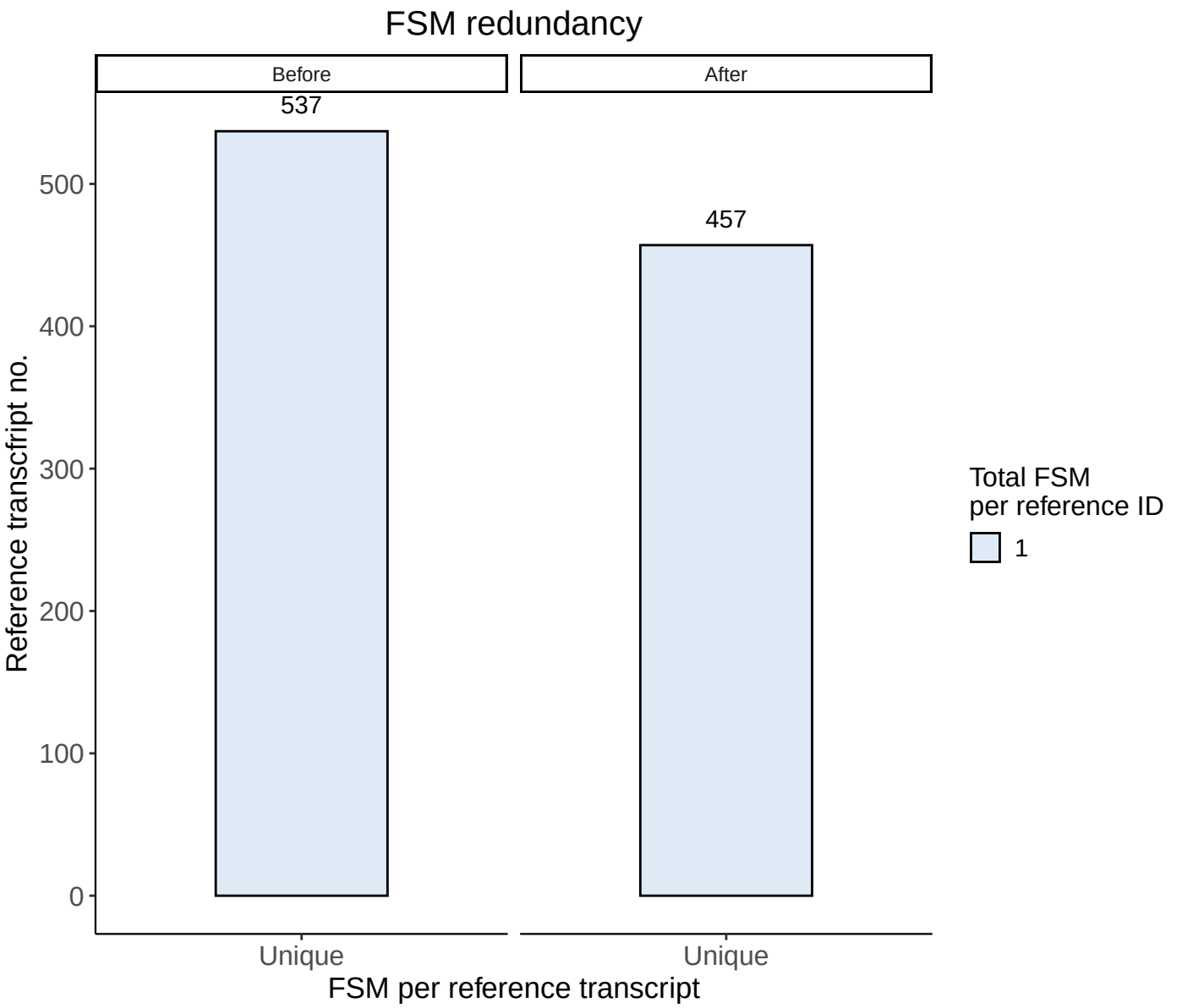
Structural category	Artifact no.	Isoform no.
FSM	80	457
ISM	296	838
NIC	443	362
NNC	1037	102
Genic	34	22
Antisense	24	4
Fusion	33	5
Intergenic	178	10

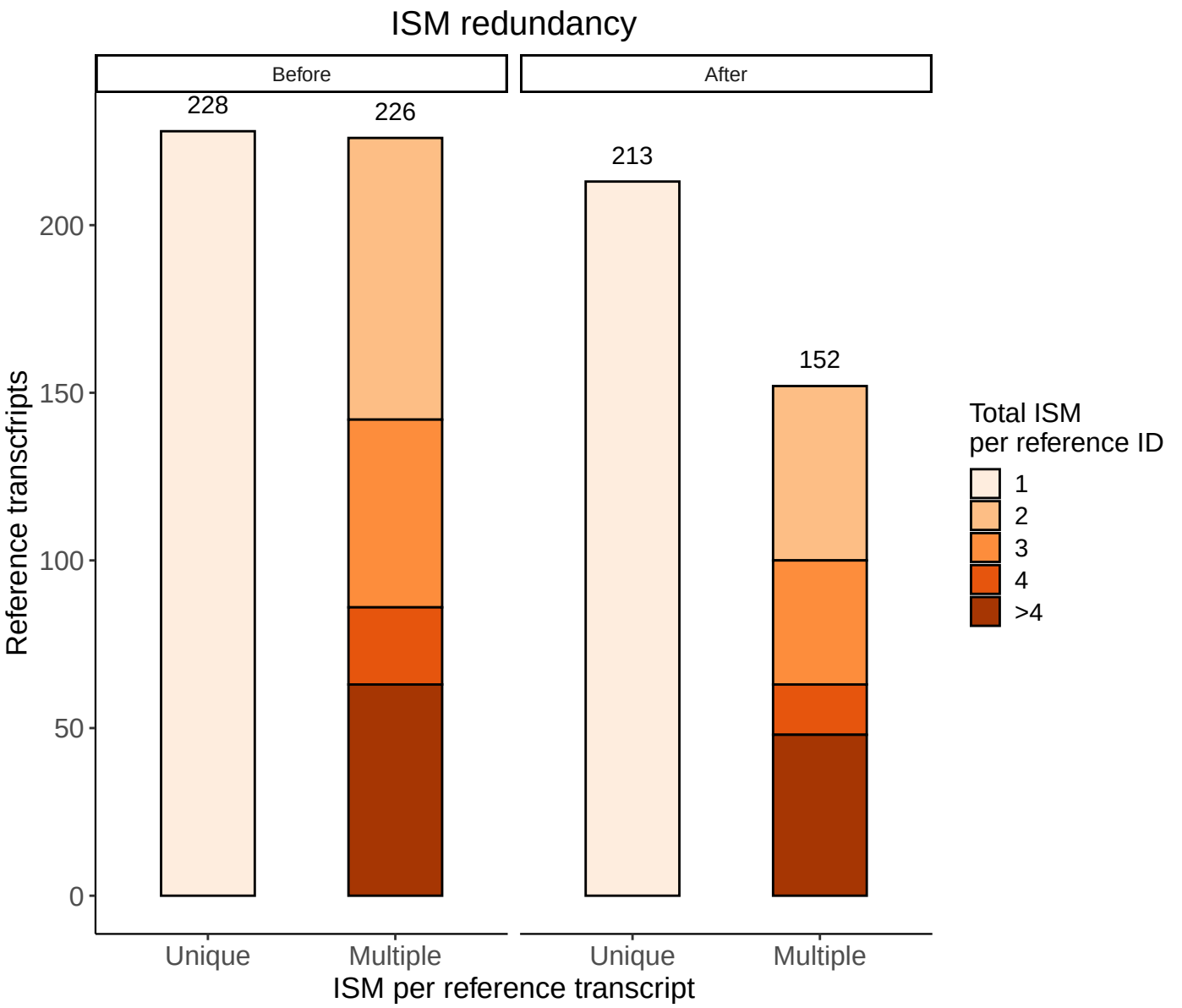

Total isoforms and artifacts by category
Transcript no.
1000
750
500
250
0
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

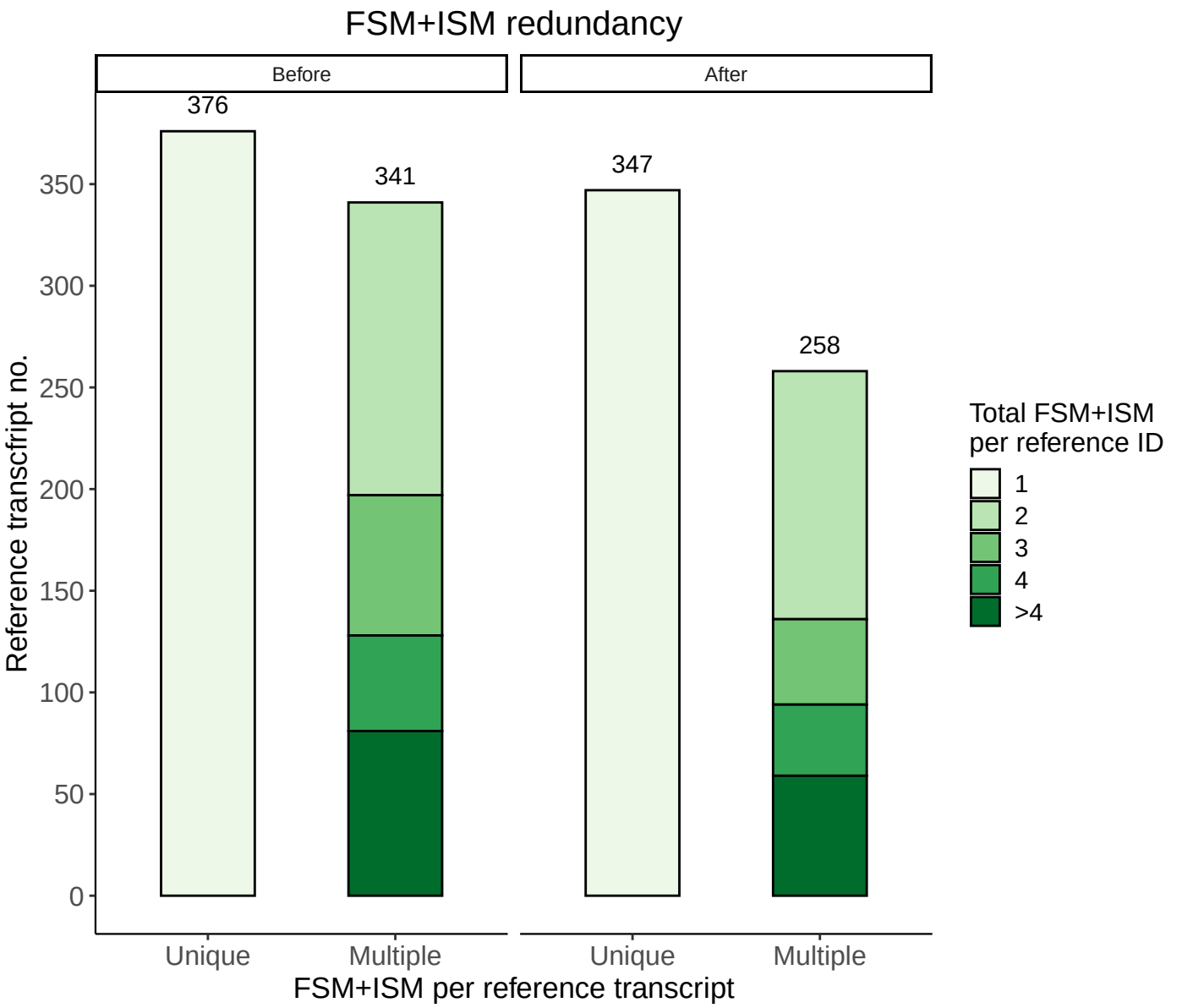

% isoforms and artifacts by category
100%
75%
50%
25%
0%
Transcript %
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

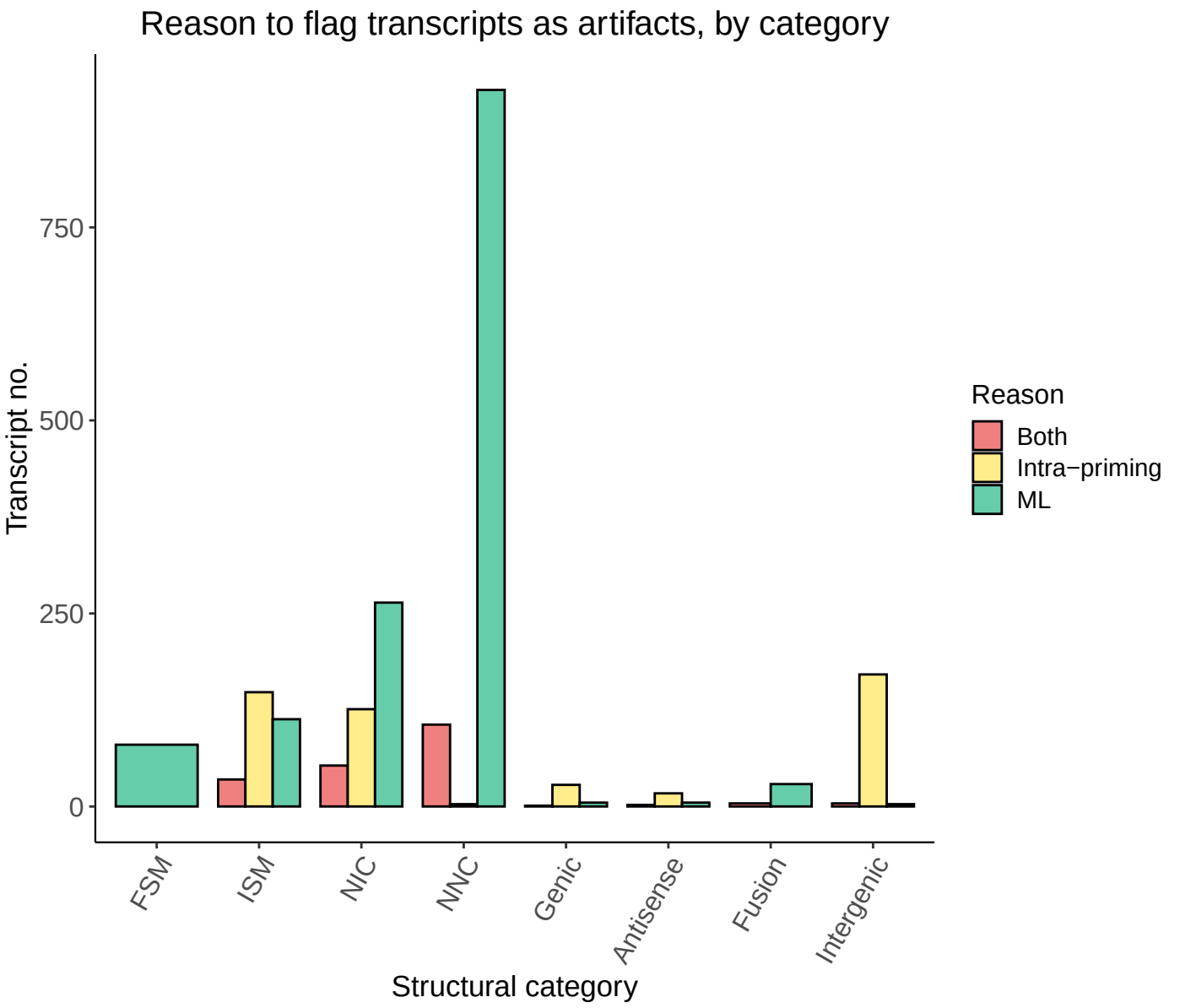

Number of isoforms per gene
40.0%
30.0%
20.0%
10.0%
0.0%
% Genes
1
2−3
4−5
>=6
Isoforms per gene
Filter
Before
After

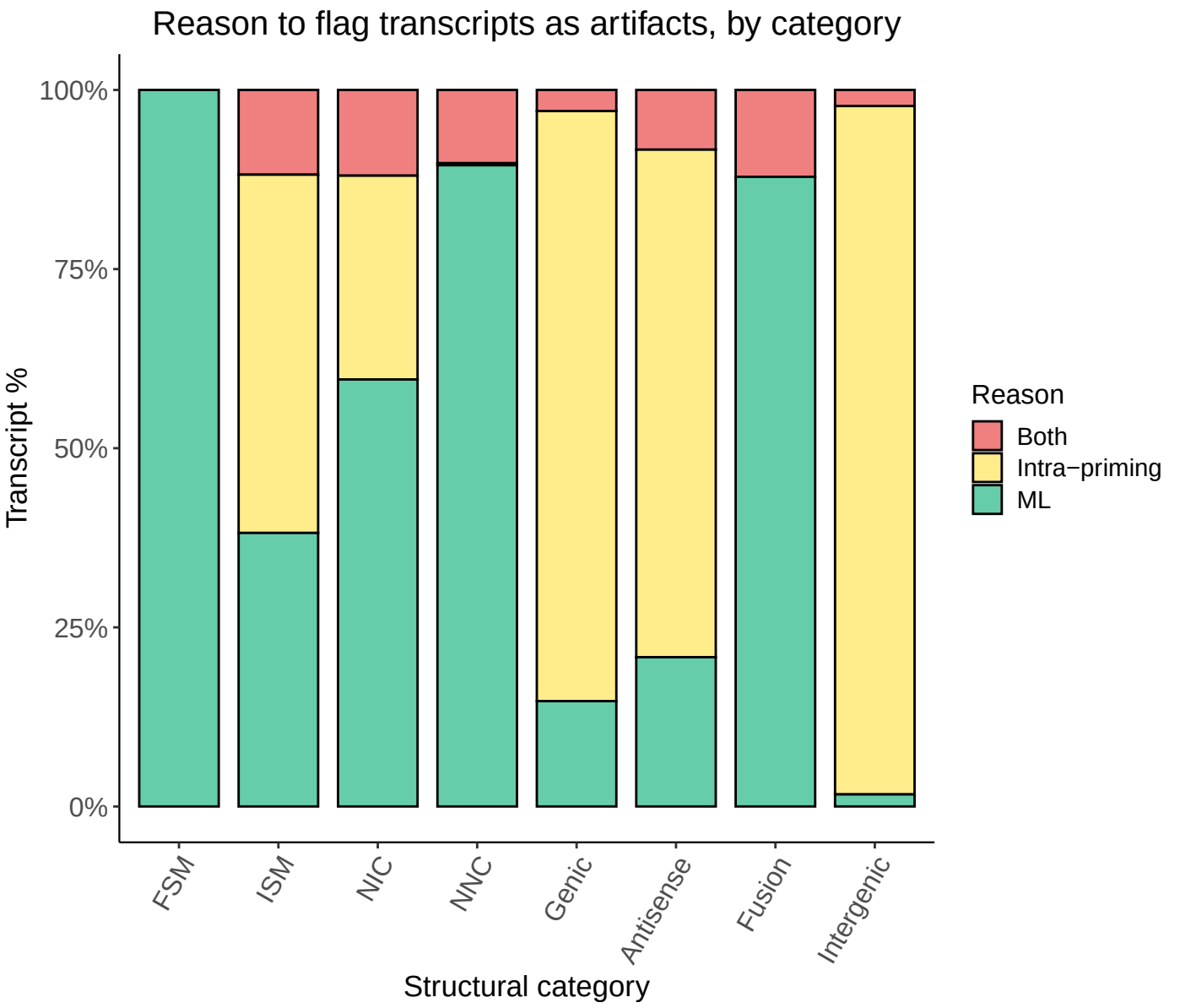

Reference transcript complexity
No. of reference transcripts represented by FSM
537
457
500
400
300
200
100
0
Reference transcfript no.
Before
After
Filter
FSM per reference ID
Unique

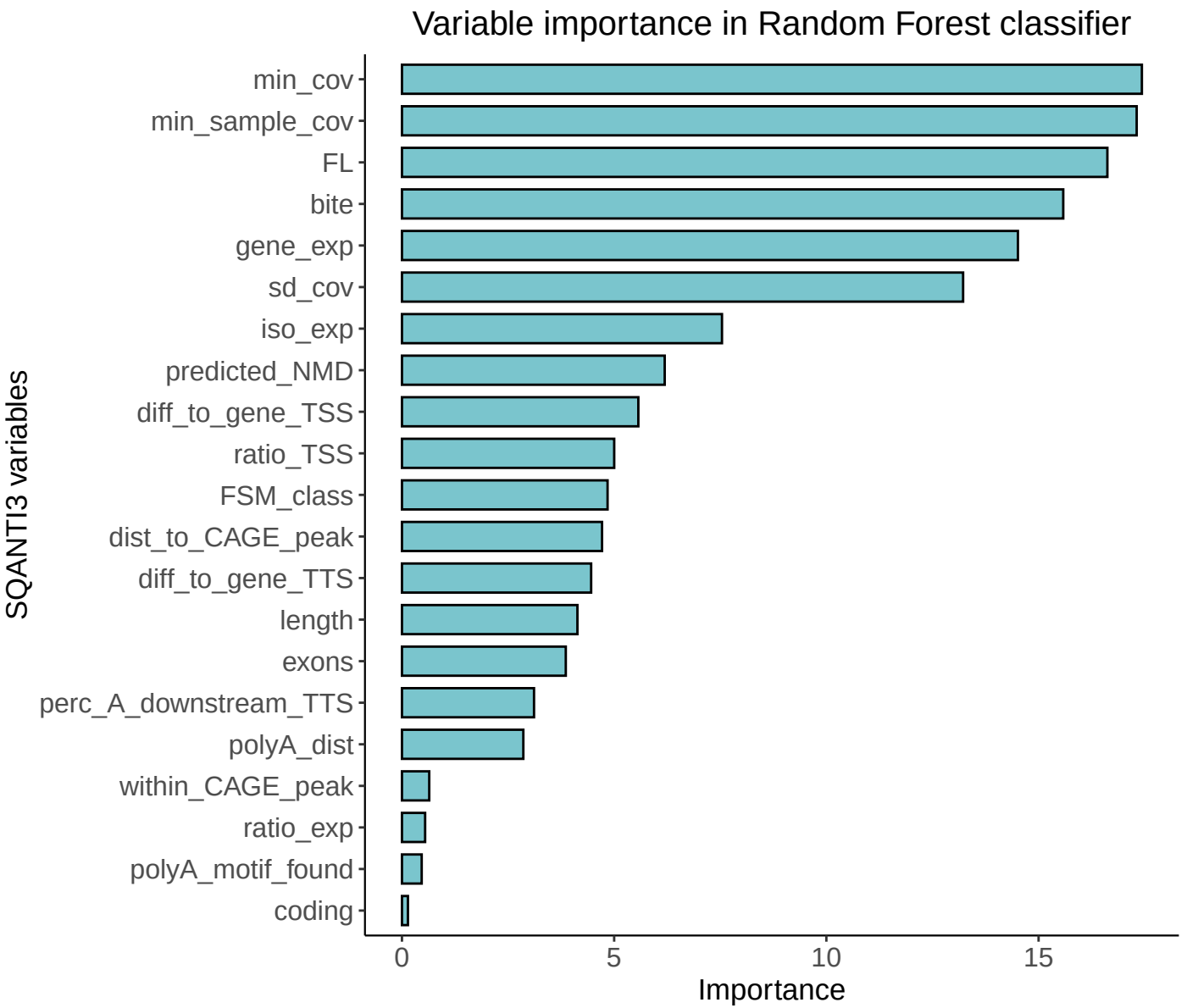

FSM redundancy
Before
After
537
457
500
400
300
200
100
0
Reference transcfript no.
Unique
Unique
FSM per reference transcript
Total FSM
per reference ID
1

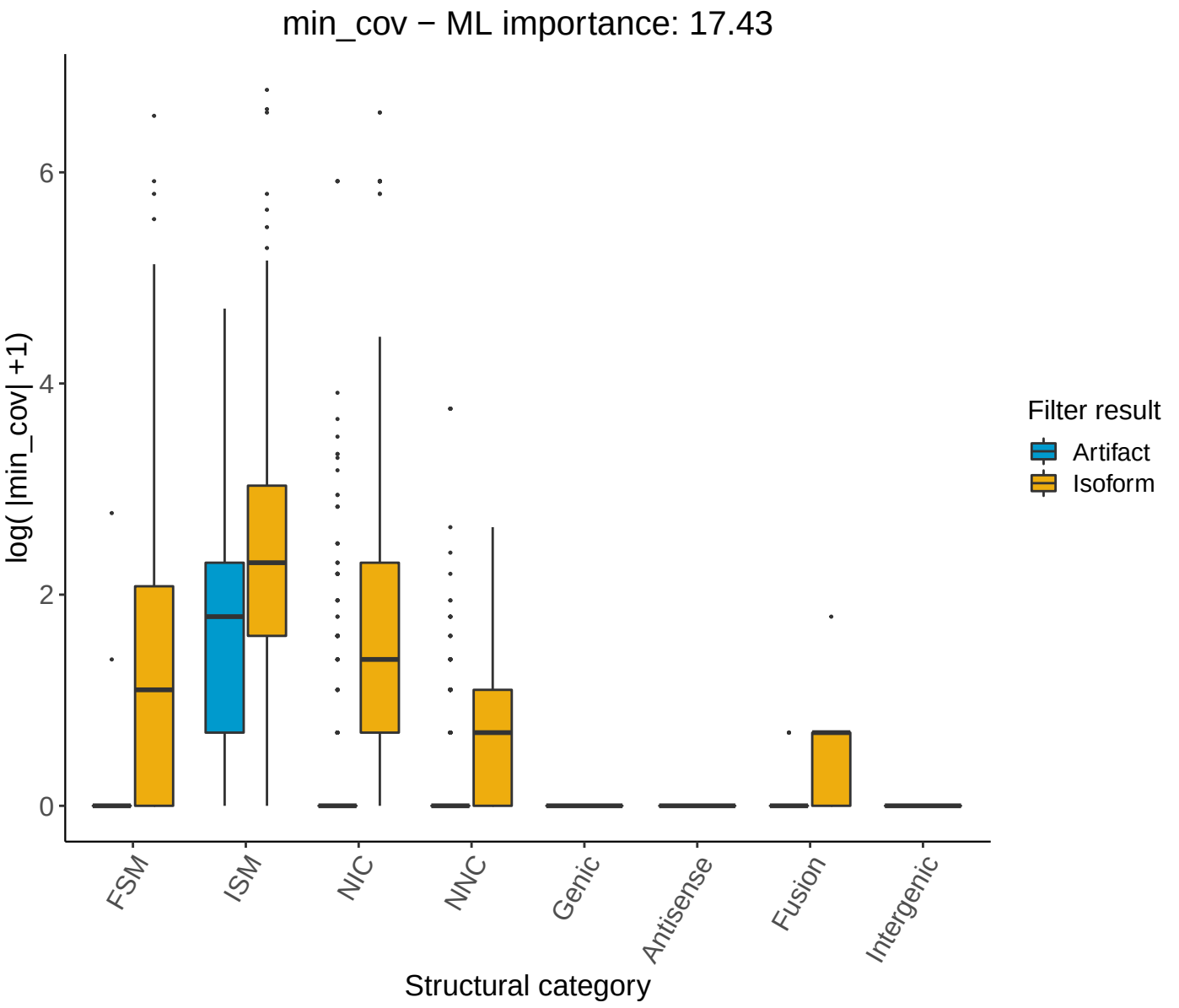

ISM redundancy
Before
After
228
226
213
152
200
150
100
50
0
Reference transcfripts
Unique
Multiple
Unique
Multiple
ISM per reference transcript
Total ISM
per reference ID
1
2
3
4
>4

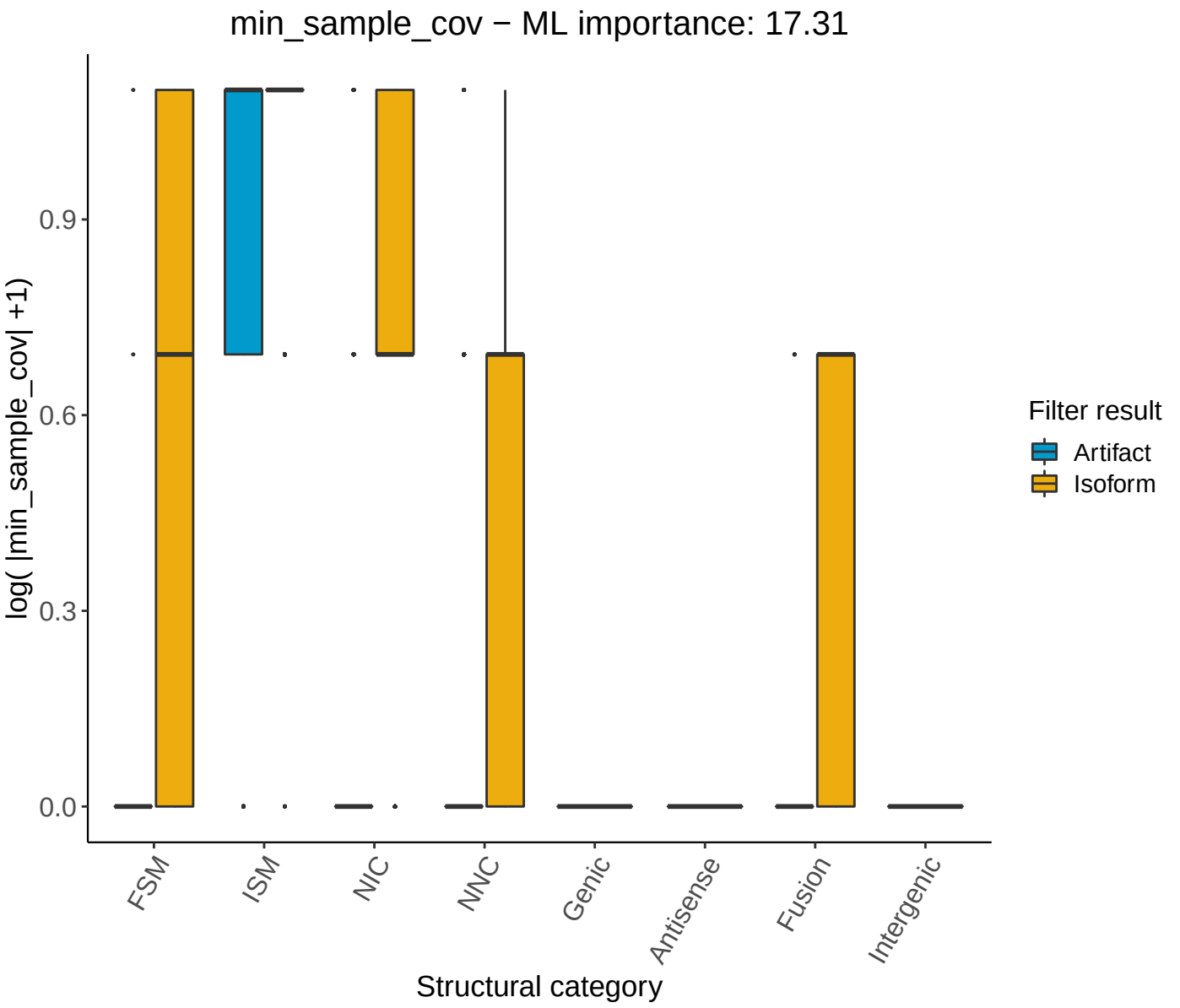

FSM+ISM redundancy
Before
After
376
341
347
258
350
300
250
200
150
100
50
0
Reference transcfript no.
Unique
Multiple
Unique
Multiple
FSM+ISM per reference transcript
Total FSM+ISM per reference ID
1
2
3
4
>4

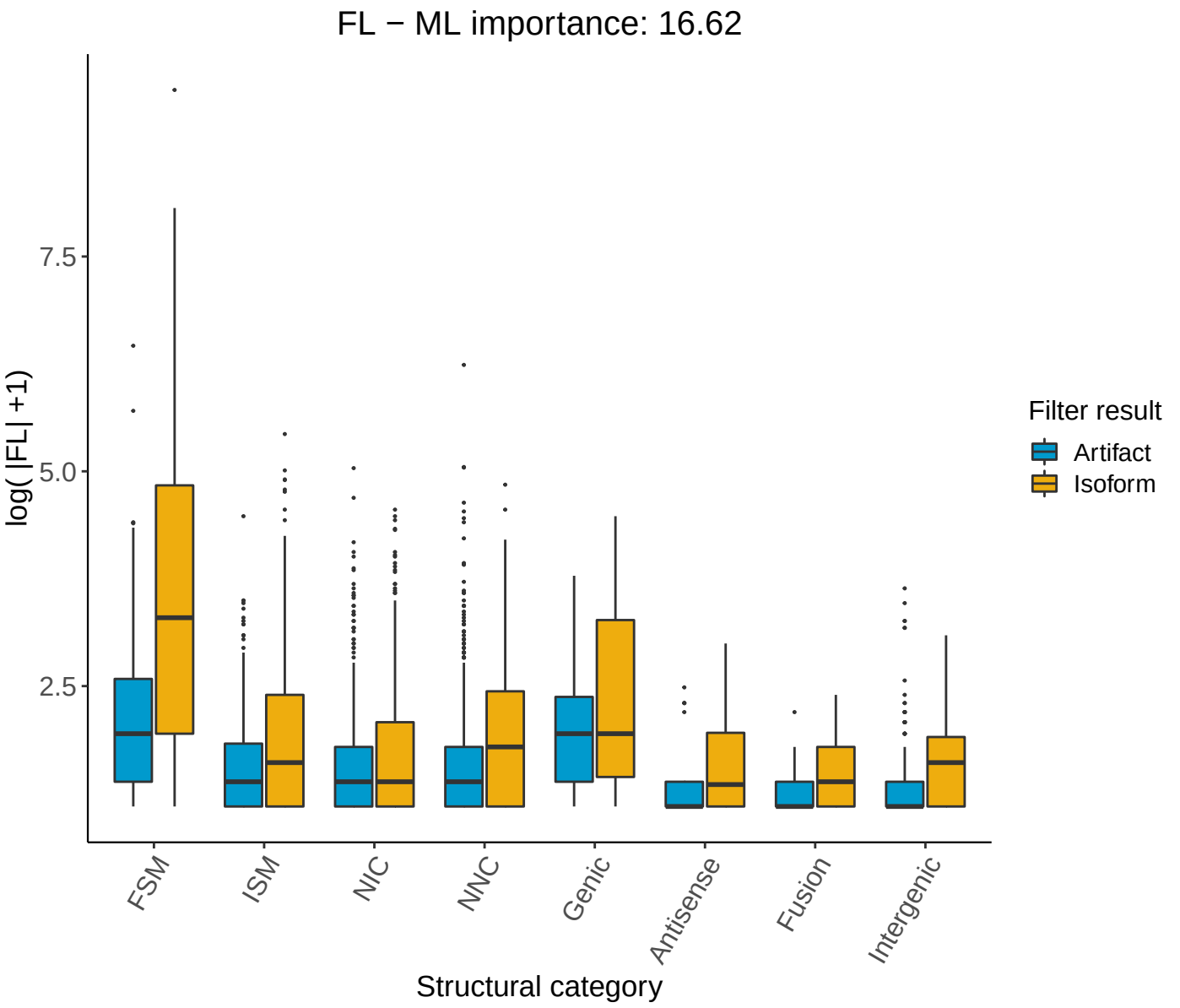

Reason to flag transcripts as artifacts, by category
Transcript no.
750
500
250
0
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Reason
Both
Intra−priming
ML

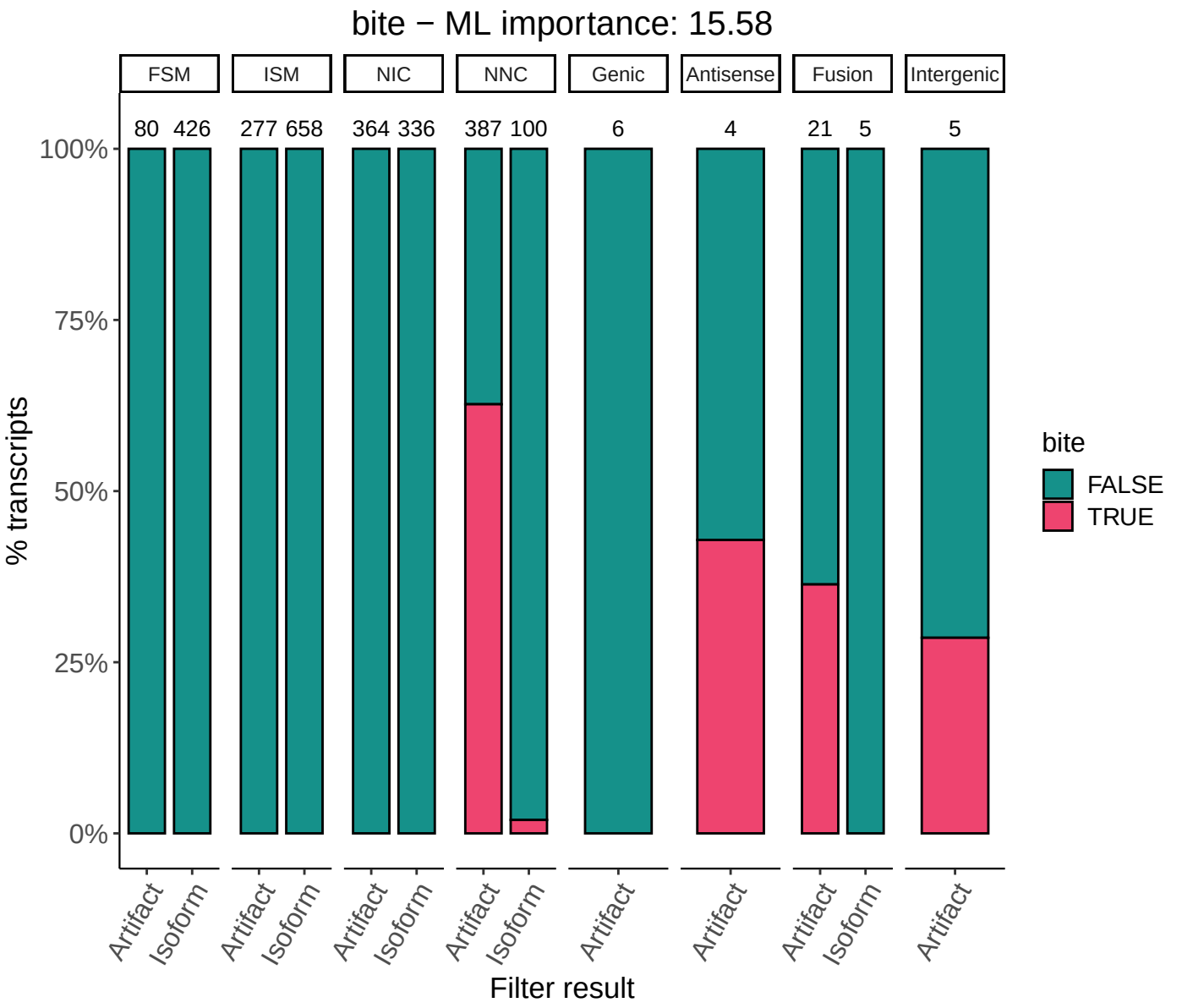

Reason to flag transcripts as artifacts, by category
100%
75%
50%
25%
0%
Transcript %
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Reason
Both
Intra−priming
ML

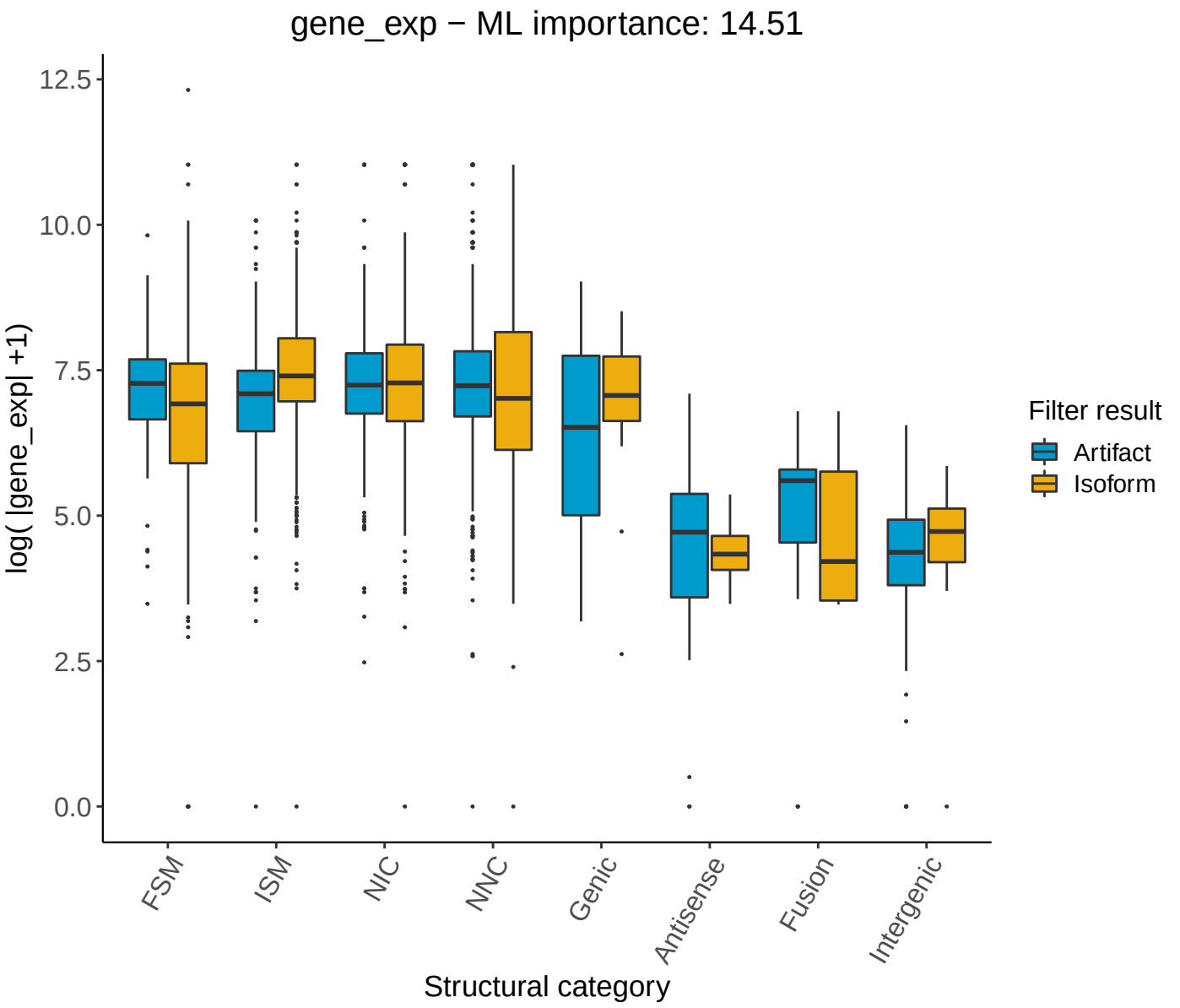

Variable importance in Random Forest classifier
SQANTI3 variables
min_cov
min_sample_cov
FL
bite
gene_exp
sd_cov
iso_exp
predicted_NMD
diff_to_gene_TSS
ratio_TSS
FSM_class
dist_to_CAGE_peak
diff_to_gene_TTS
length
exons
perc_A_downstream_TTS
polyA_dist
within_CAGE_peak
ratio_exp
polyA_motif_found
coding
0
5
10
15
Importance

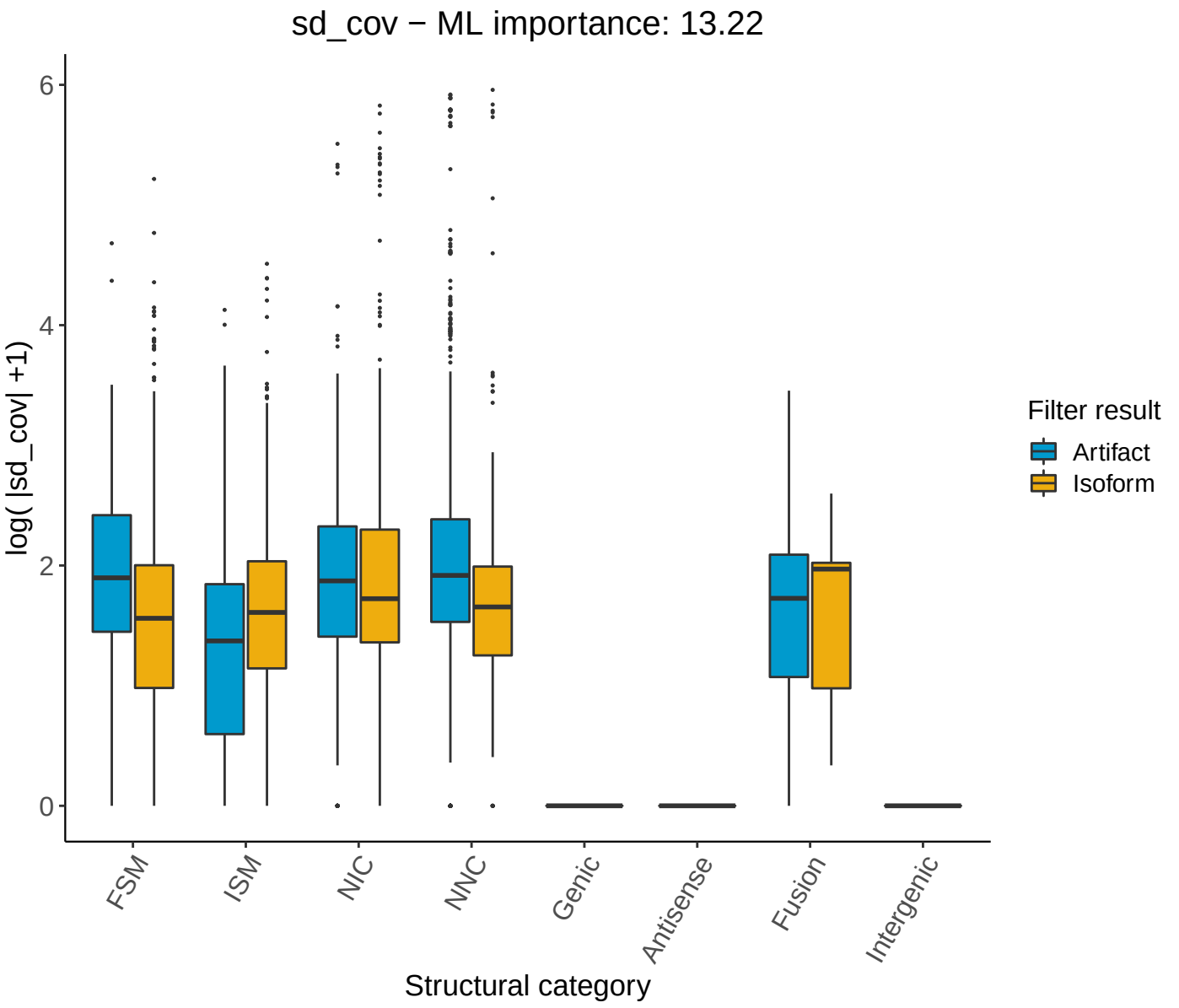

min_cov − ML importance: 17.43
log(|min_cov| +1)
0
2
4
6
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

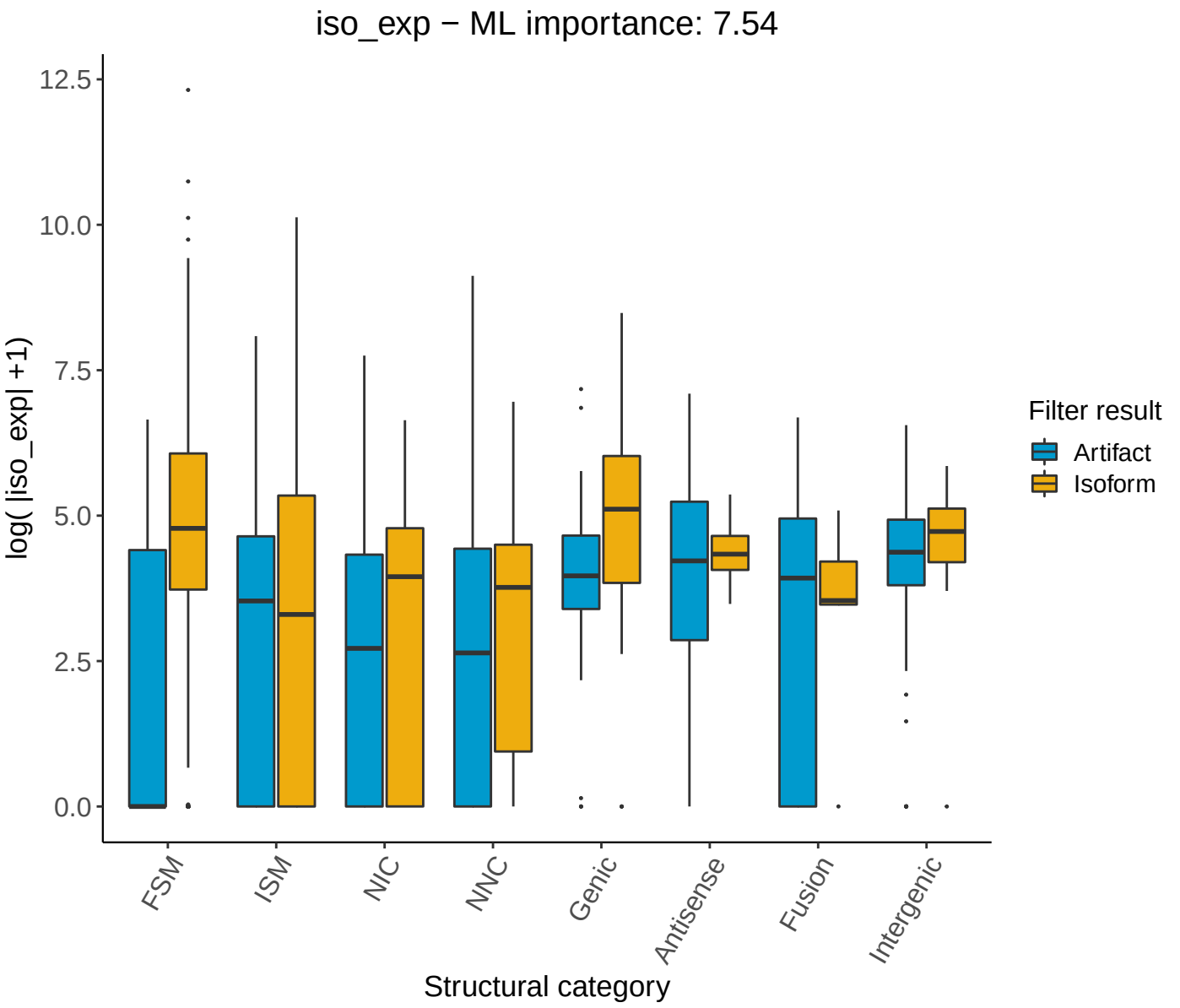

min_sample_cov − ML importance: 17.31
log(|min_sample_cov| +1)
0.0
0.3
0.6
0.9
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

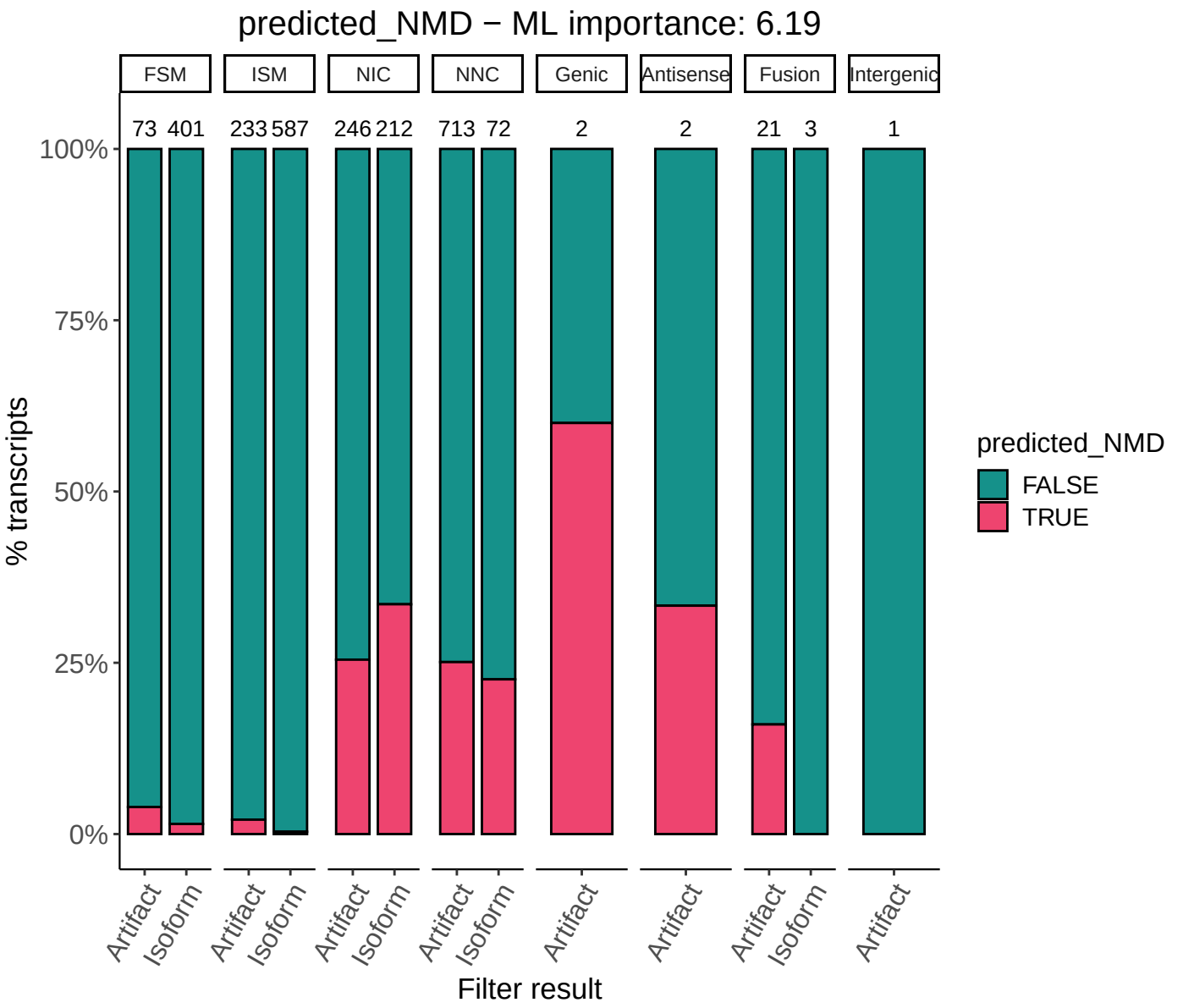

FL − ML importance: 16.62
log(|FL| +1)
7.5
5.0
2.5
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

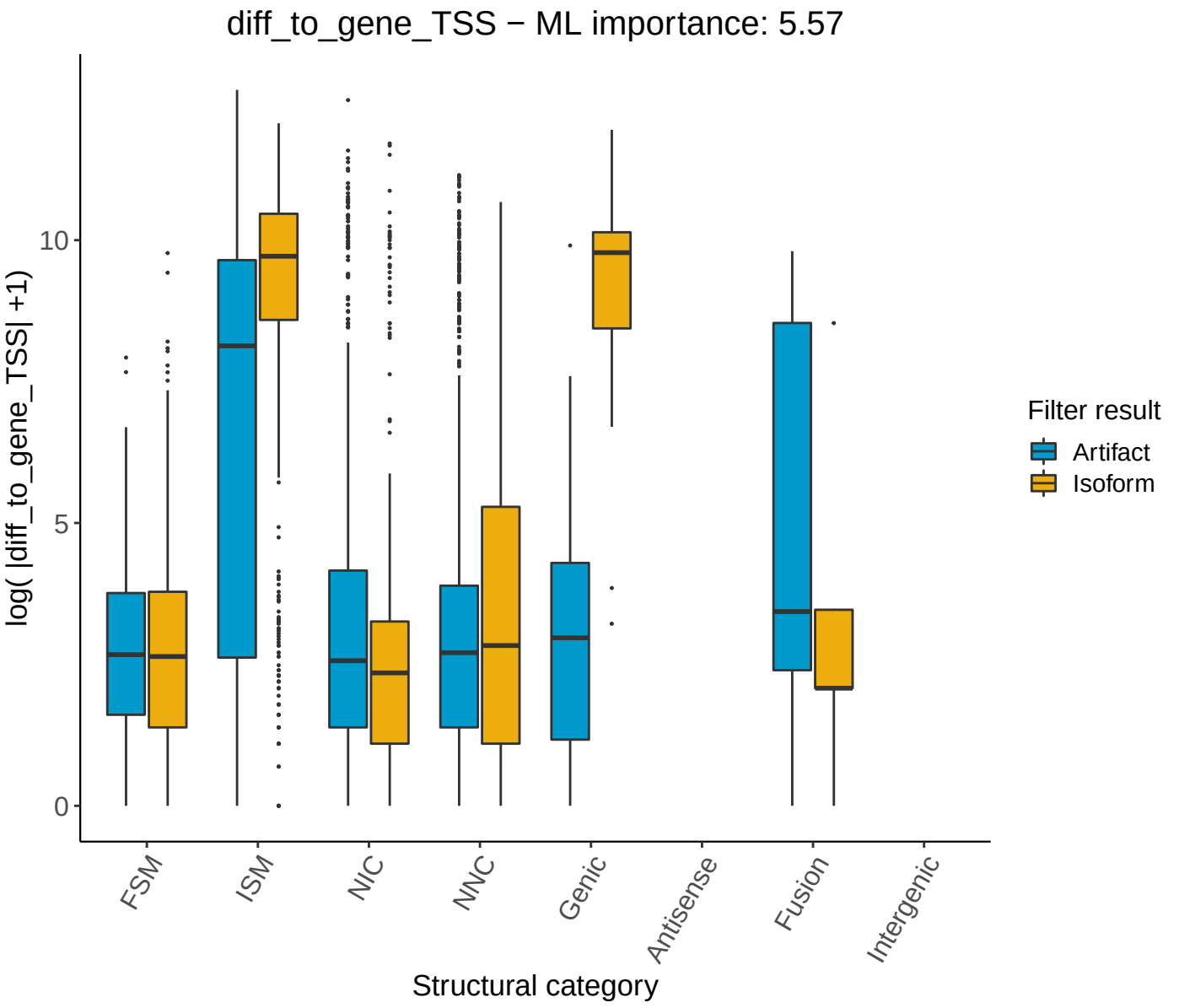

bite − ML importance: 15.58
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
80 426
277 658
364 336
387 100
6
4
21 5
5
100%
75%
50%
25%
0%
% transcripts
Artifact
Isoform
Filter result
bite
FALSE
TRUE

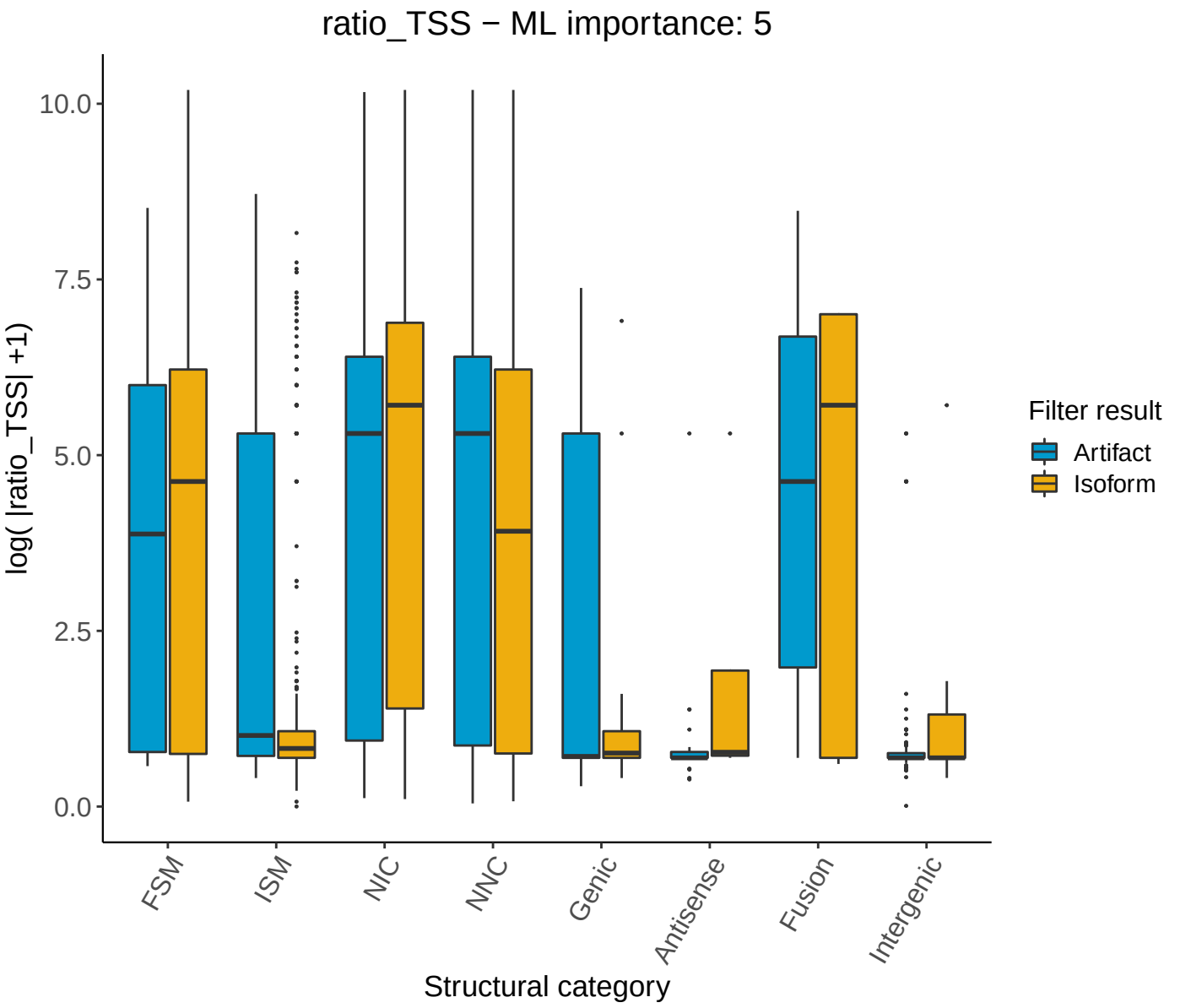

gene_exp − ML importance: 14.51
log(|gene_exp| +1)
12.5
10.0
7.5
5.0
2.5
0.0
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

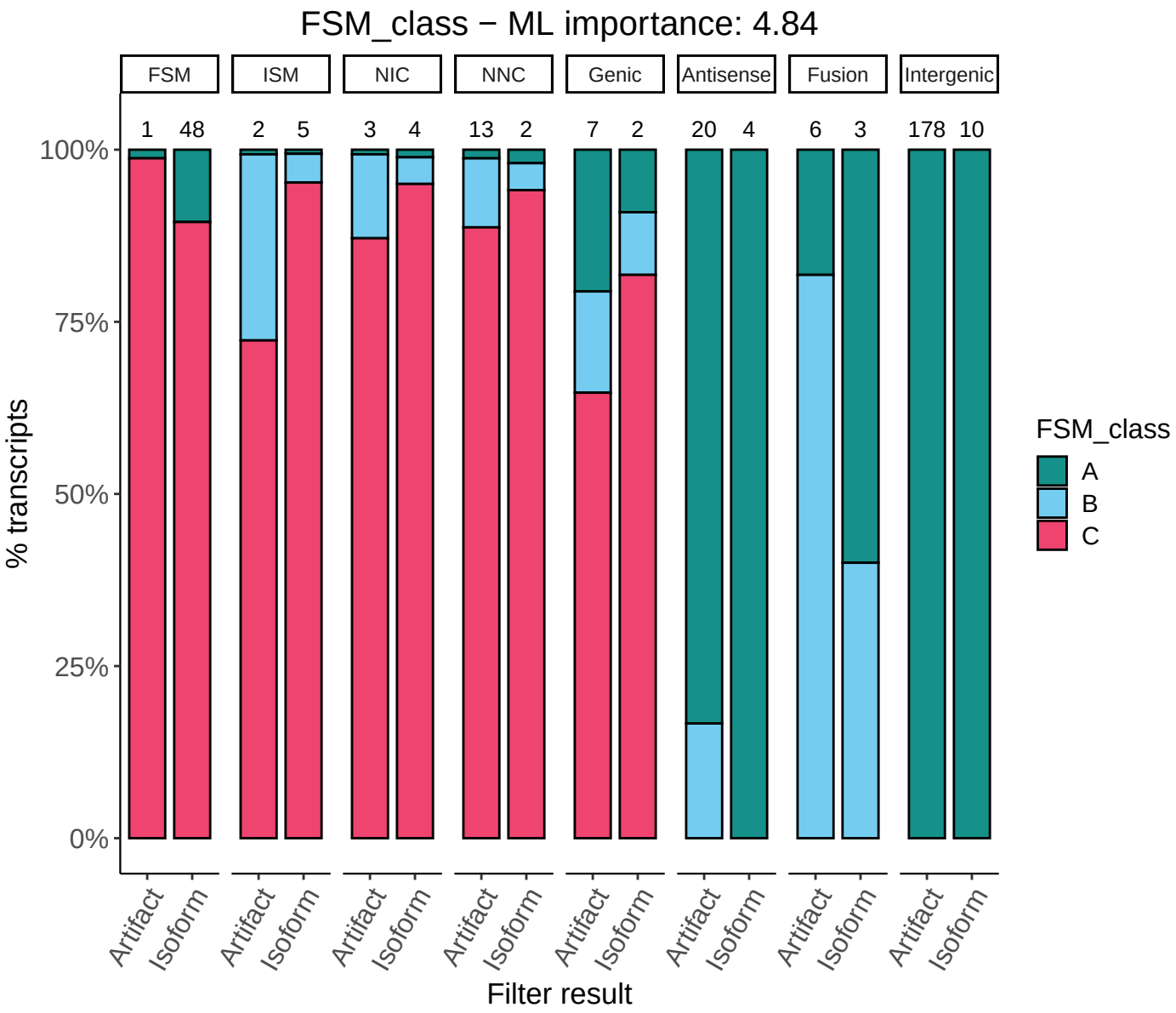

sd_cov − ML importance: 13.22
log(|sd_cov| +1)
6
4
2
0
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

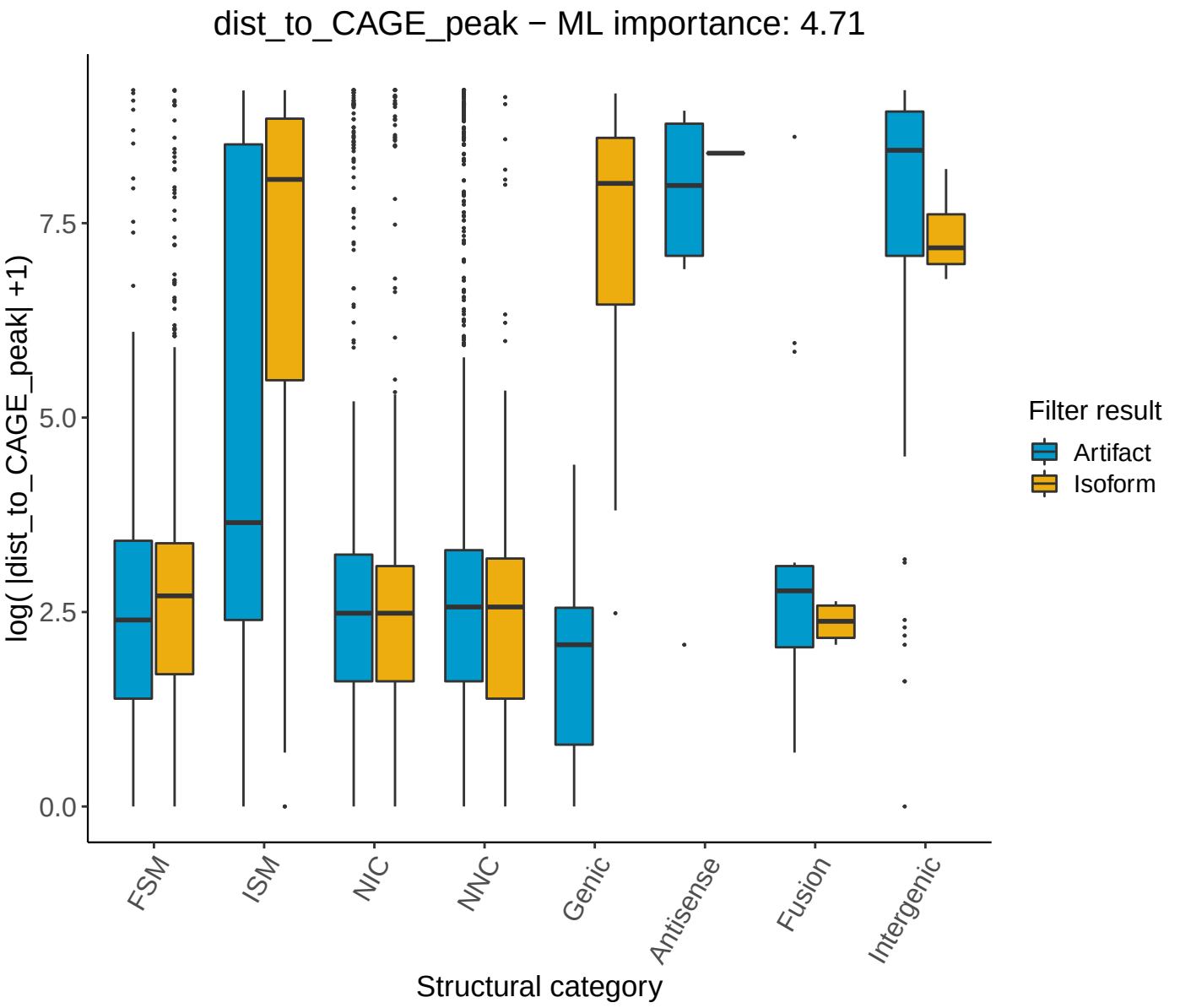

iso_exp − ML importance: 7.54
log(|iso_exp| +1)
12.5
10.0
7.5
5.0
2.5
0.0
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

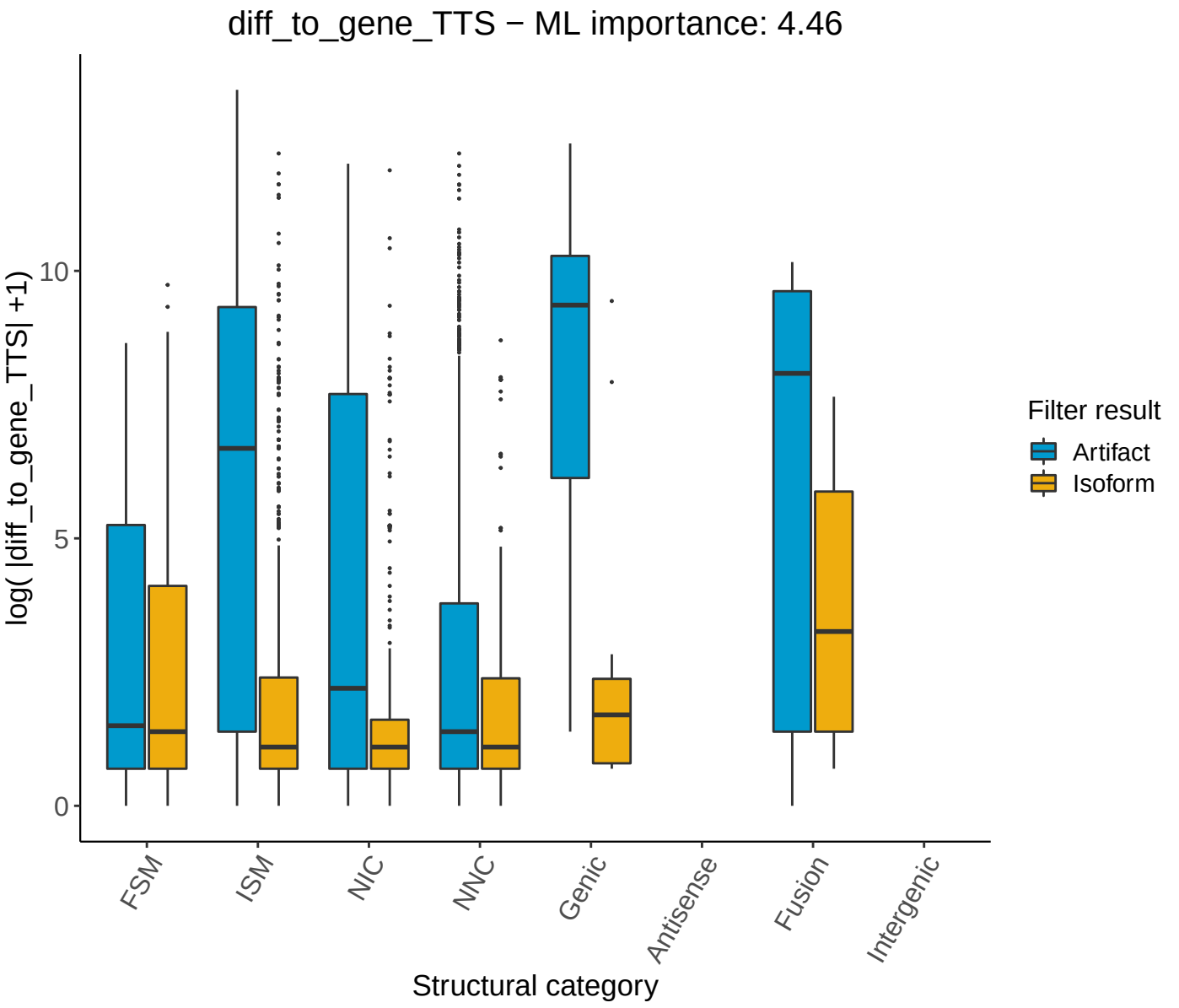

predicted_NMD − ML importance: 6.19
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
73 401
233 587
246 212
713 72
2
2
21 3
1
100%
75%
50%
25%
0%
% transcripts
Artifact
Isoform
Filter result
predicted_NMD
FALSE
TRUE

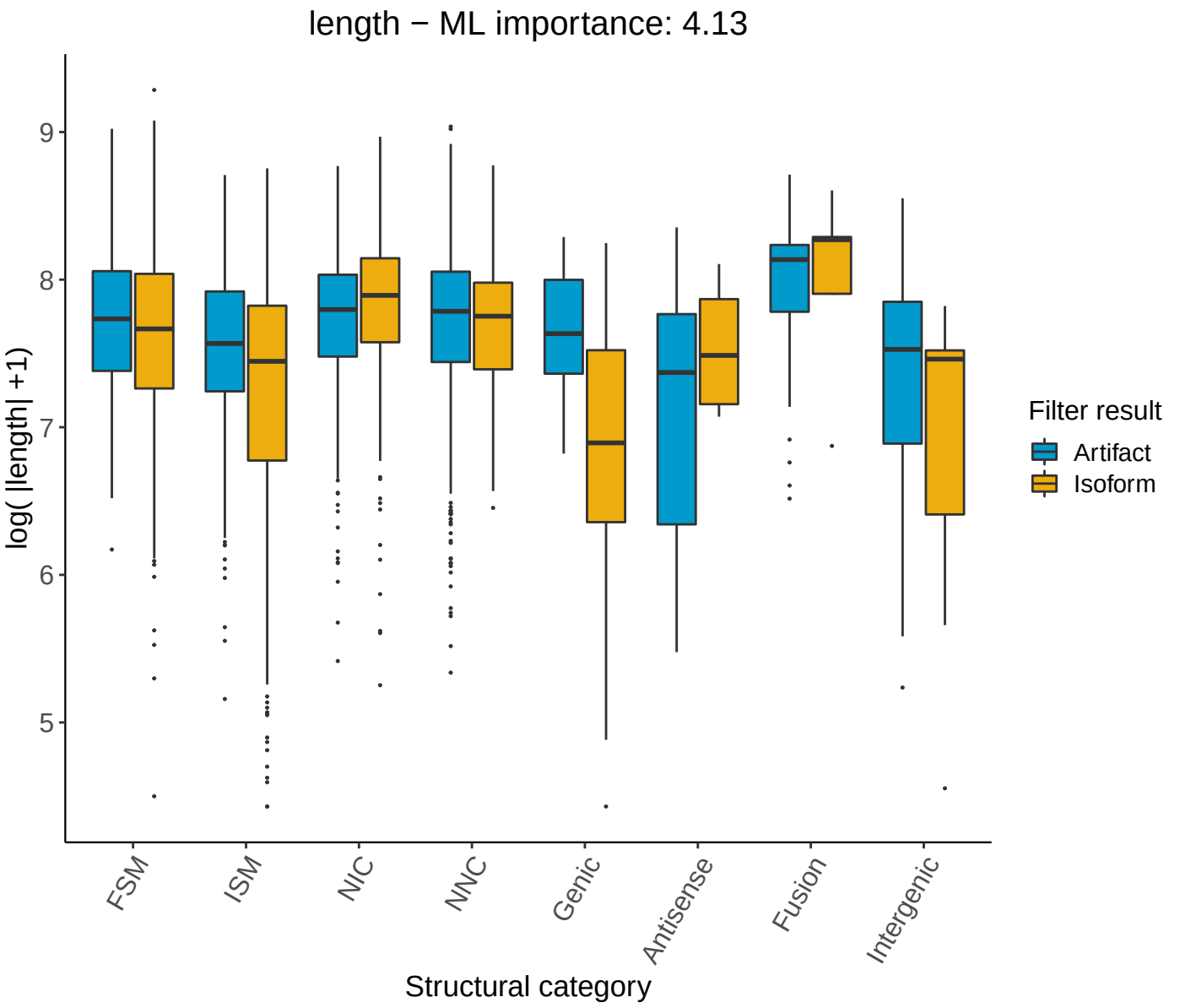

diff_to_gene_TSS − ML importance: 5.57
log(|diff_to_gene_TSS| +1)
10
5
0
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

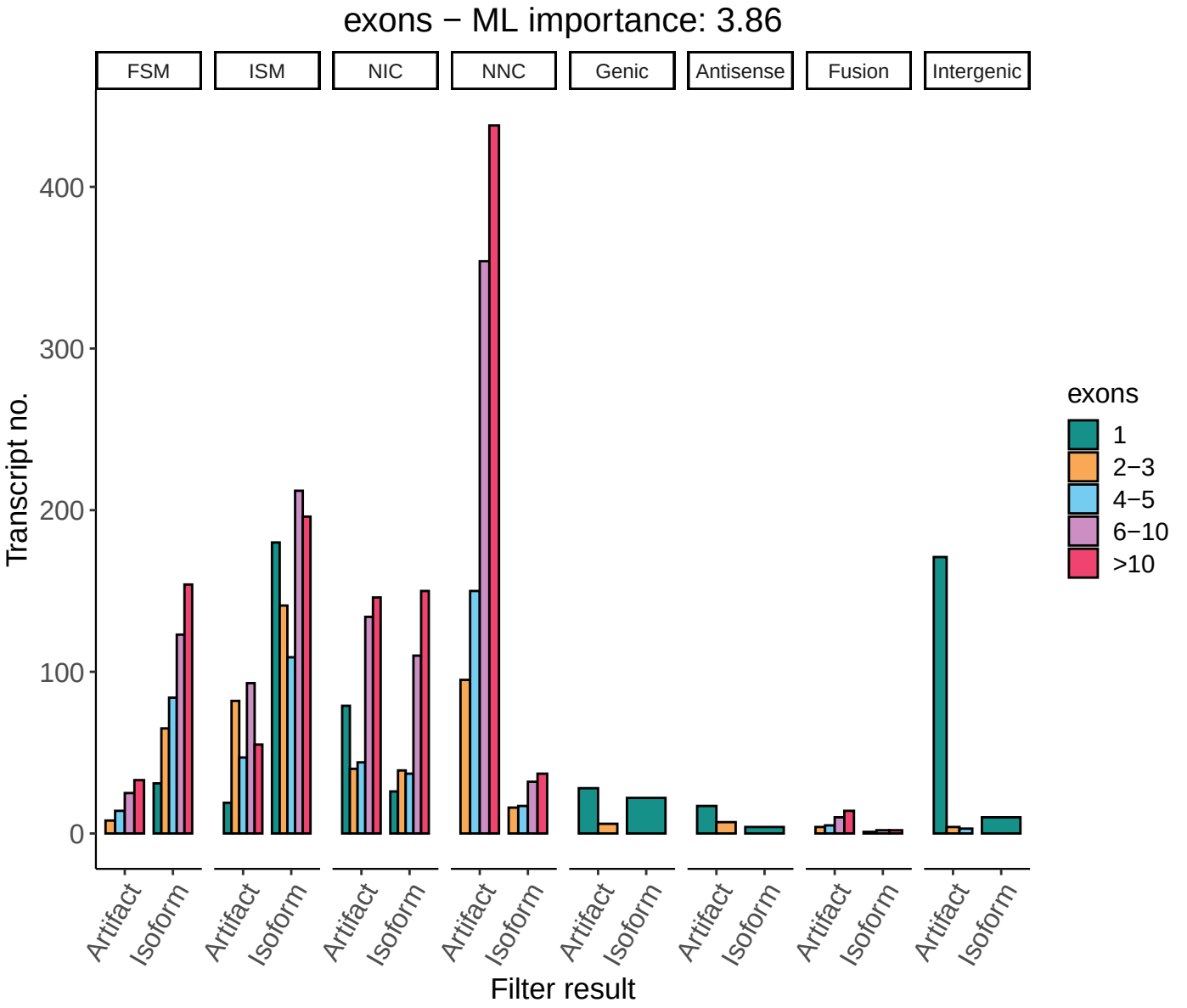

ratio_TSS − ML importance: 5
log(|ratio_TSS| +1)
10.0
7.5
5.0
2.5
0.0
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

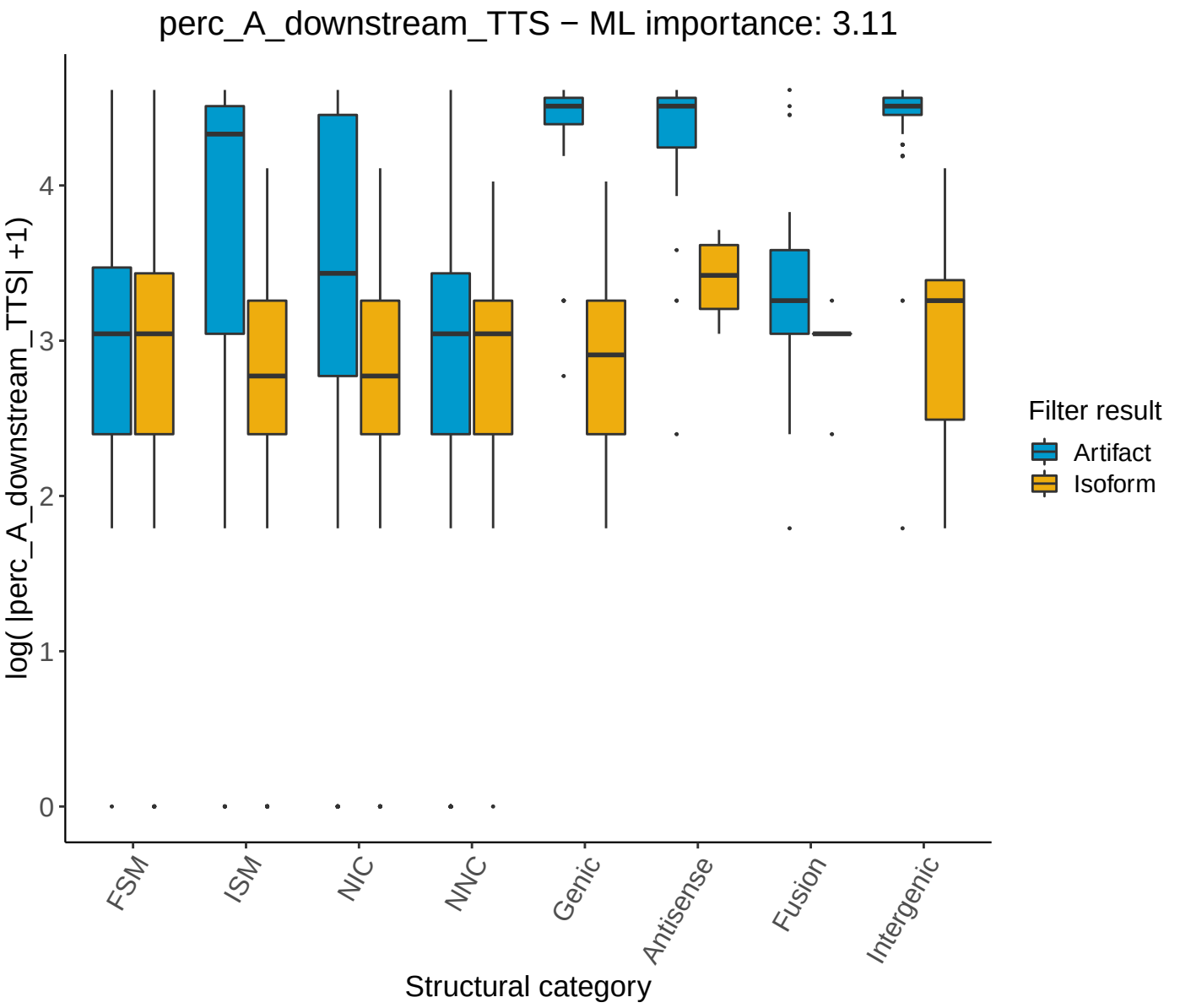

FSM_class − ML importance: 4.84
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
1 48
2 5
3 4
13 2
7 2
20 4
6 3
178 10
100%
75%
50%
25%
0%
% transcripts
Artifact
Isoform
Filter result
FSM_class
A
B
C

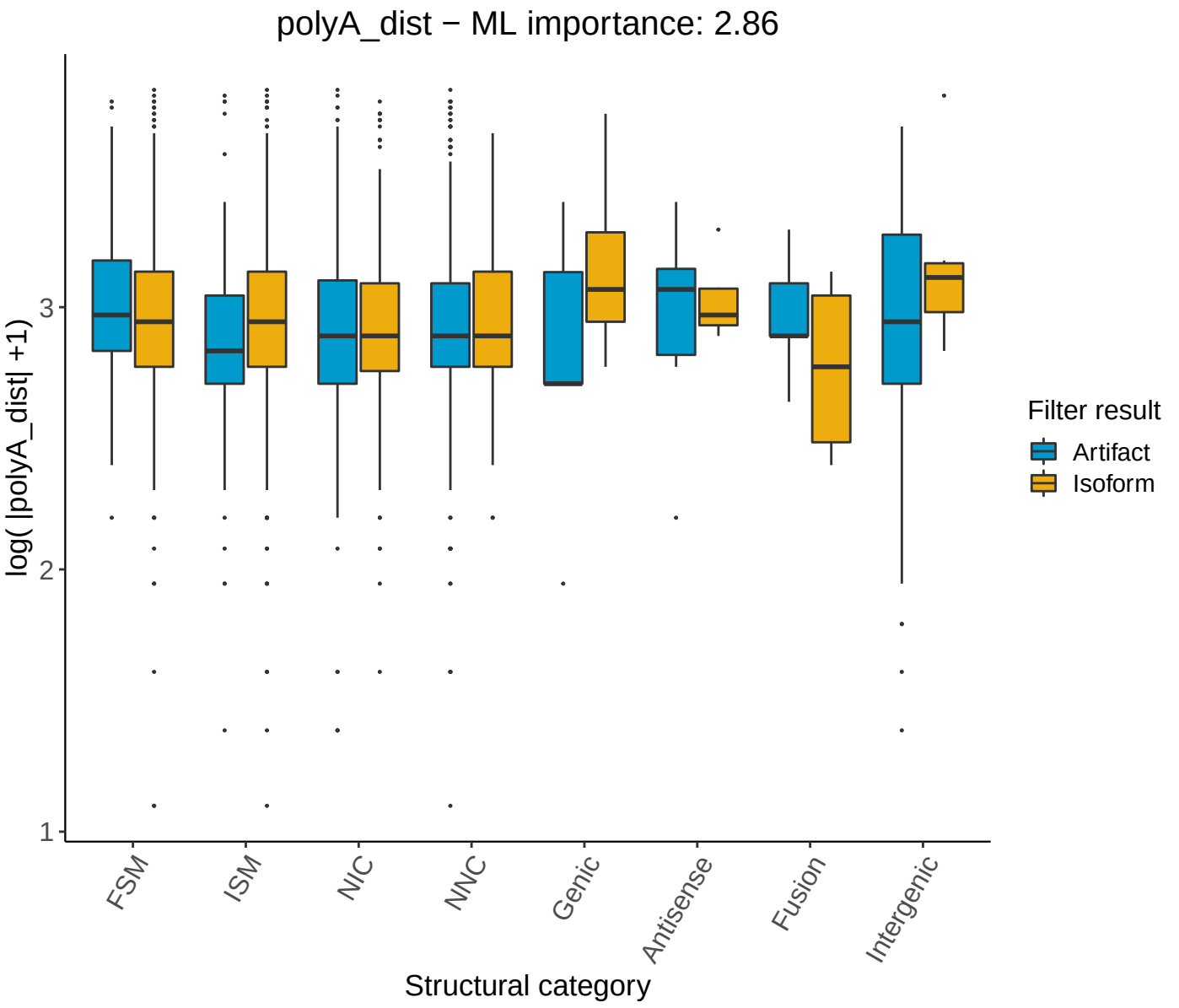

dist_to_CAGE_peak − ML importance: 4.71
log(|dist_to_CAGE_peak| +1)
7.5
5.0
2.5
0.0
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

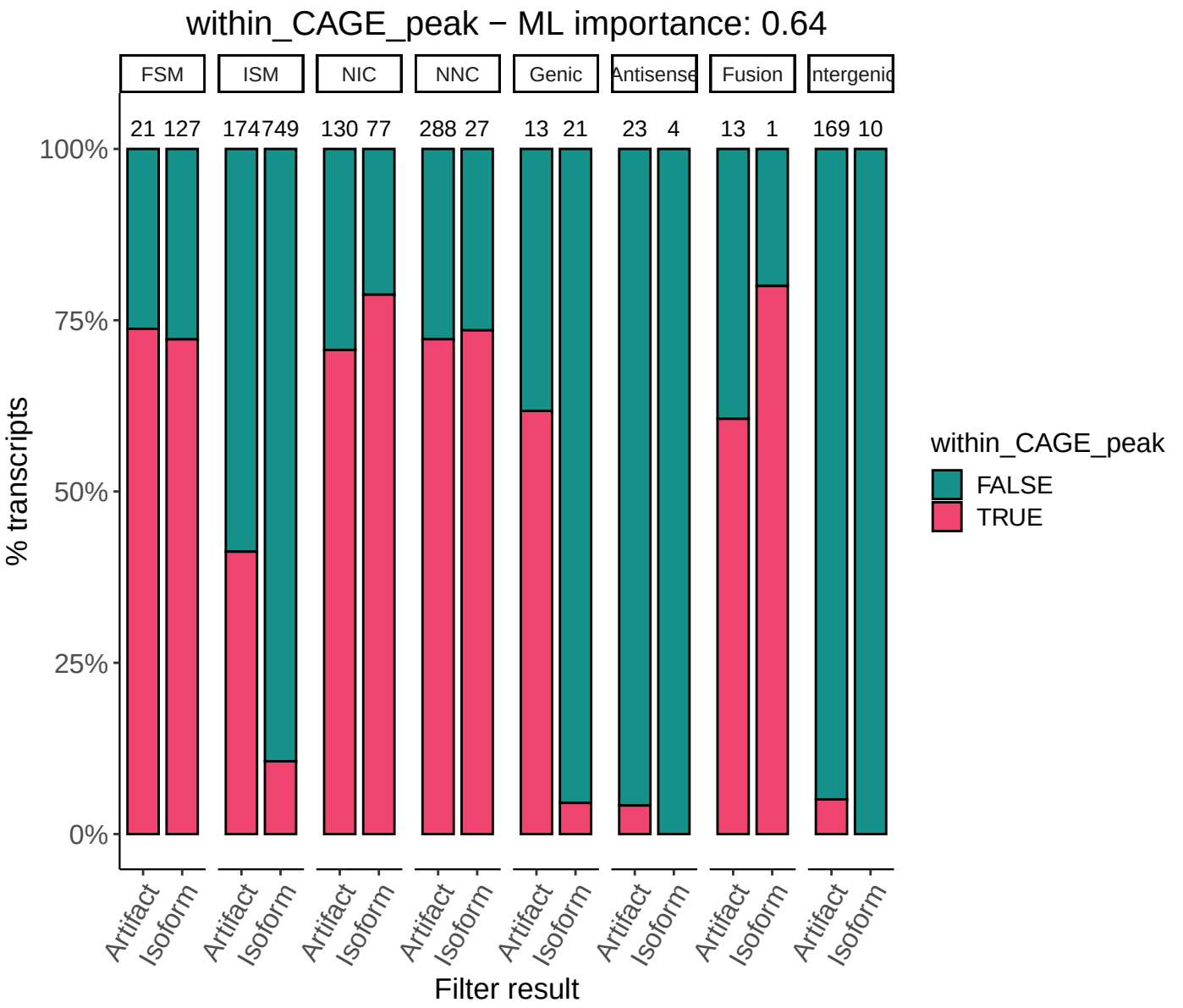

diff_to_gene_TTS − ML importance: 4.46
log(|diff_to_gene_TTS| +1)
10
5
0
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

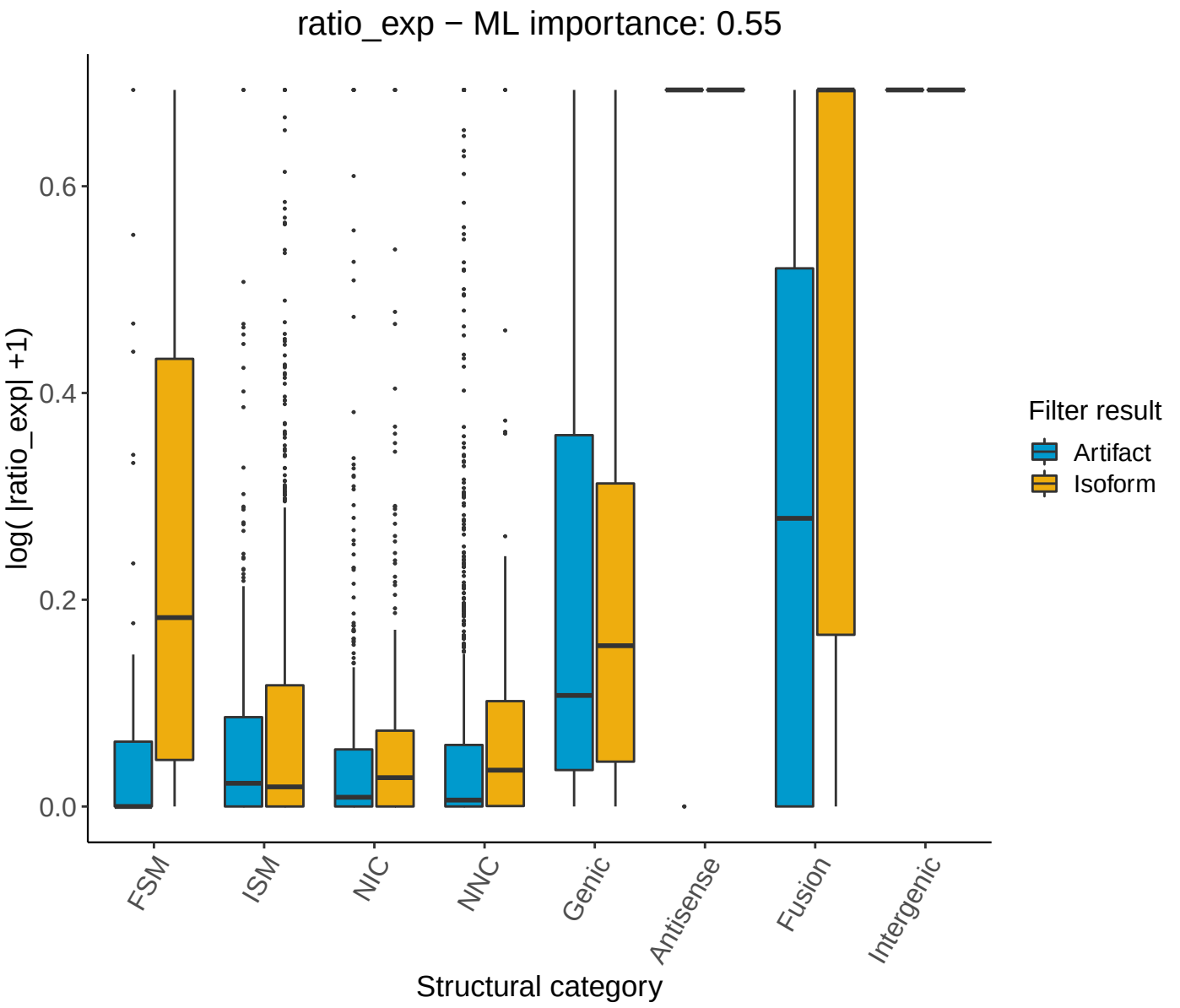

length − ML importance: 4.13
log(|length| +1)
9
8
7
6
5
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

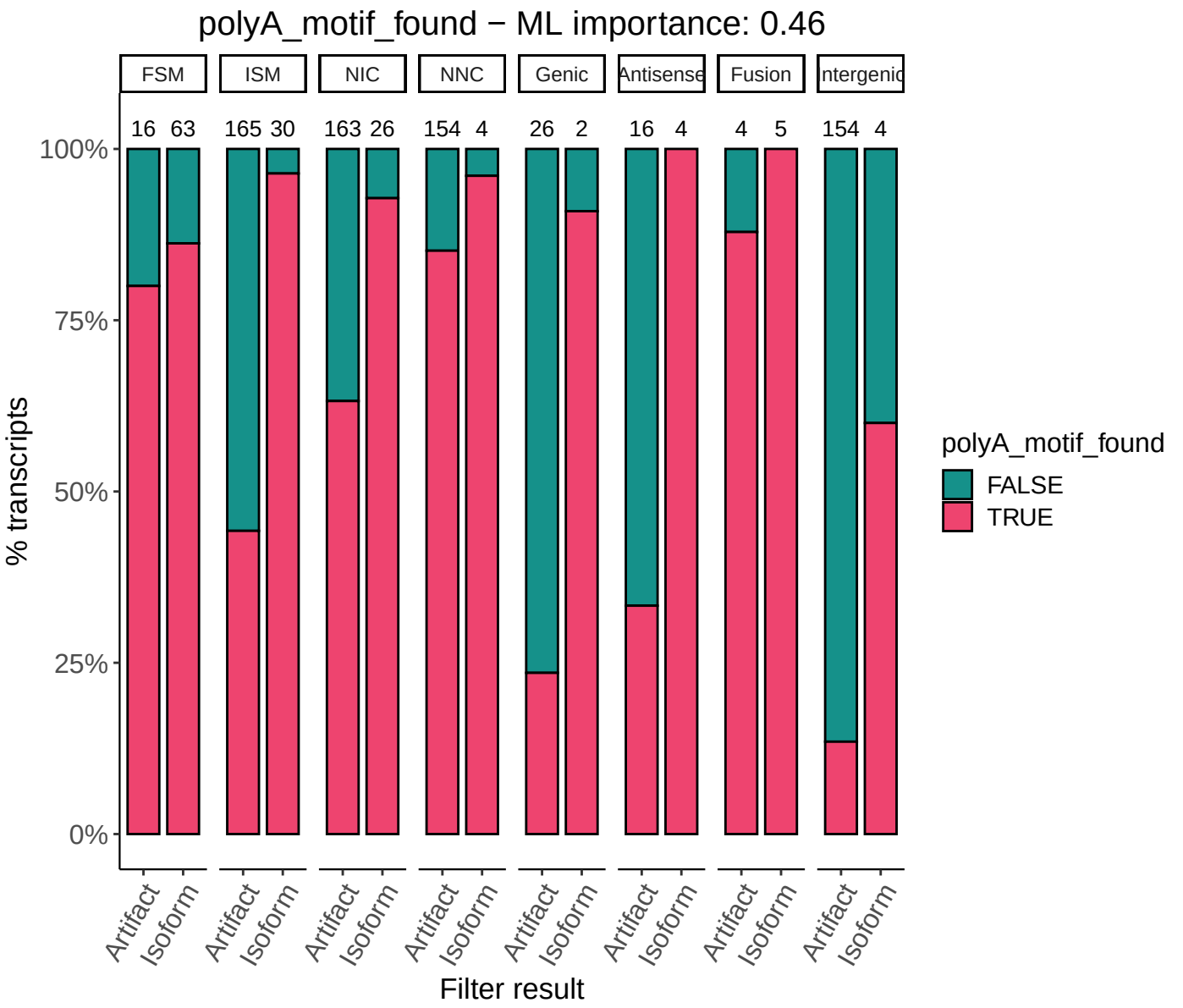

exons − ML importance: 3.86
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Transcript no.
0
100
200
300
400
Artifact
Isoform
Filter result
exons
1
2−3
4−5
6−10
>10

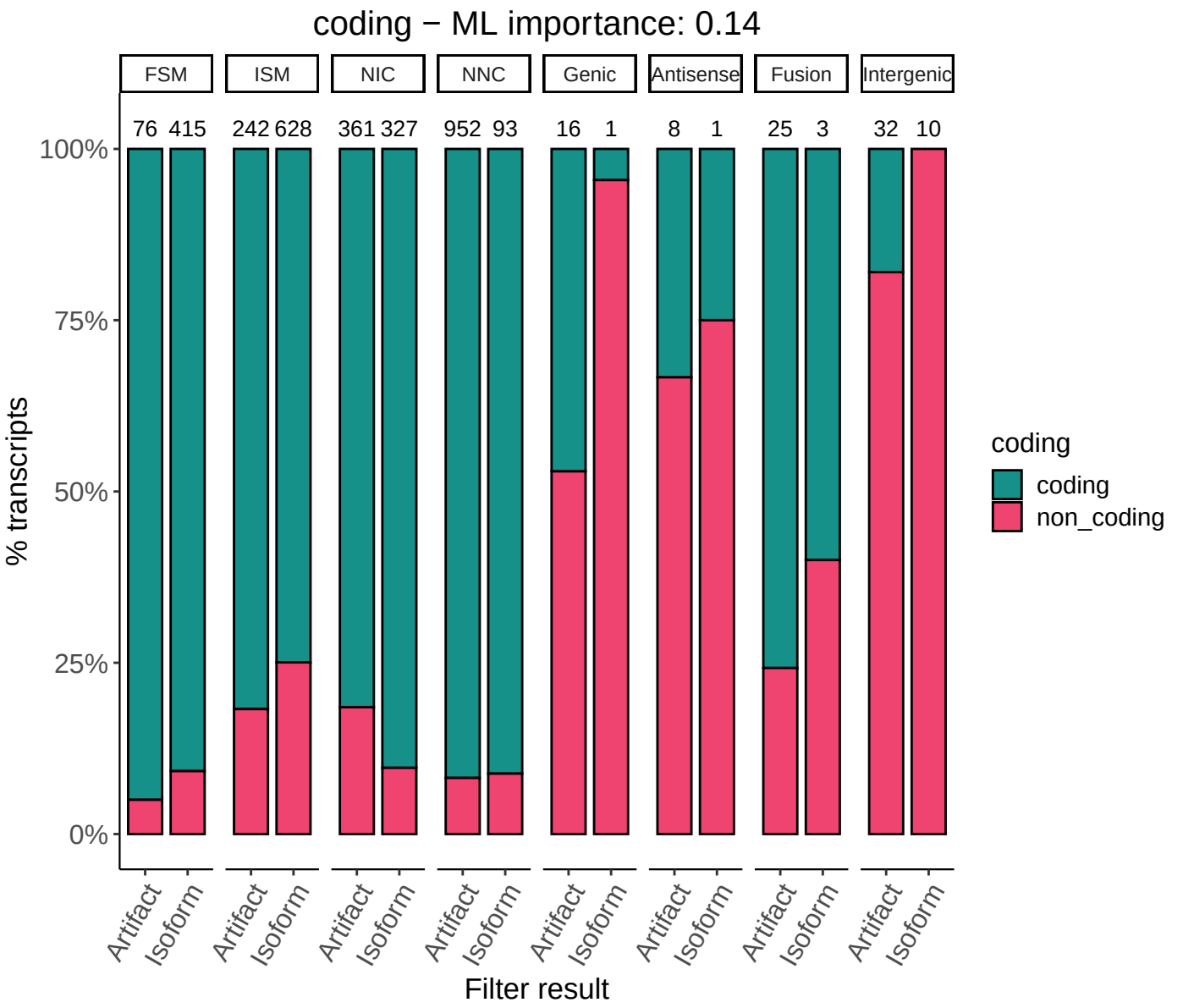

perc_A_downstream_TTS − ML importance: 3.11
log(|perc_A_downstream_TTS| +1)
0
1
2
3
4
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

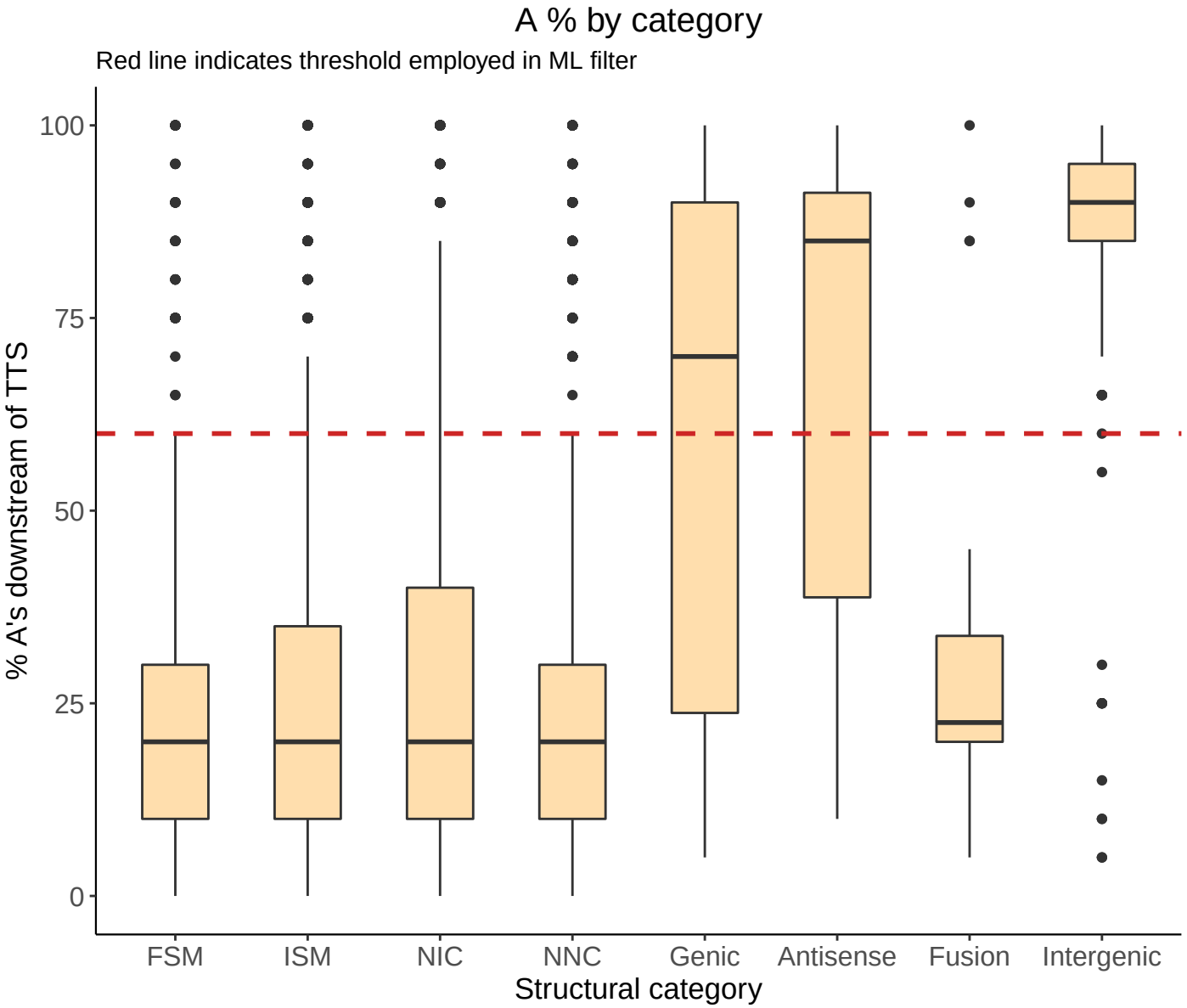

polyA_dist − ML importance: 2.86
log(|polyA_dist| +1)
3
2
1
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

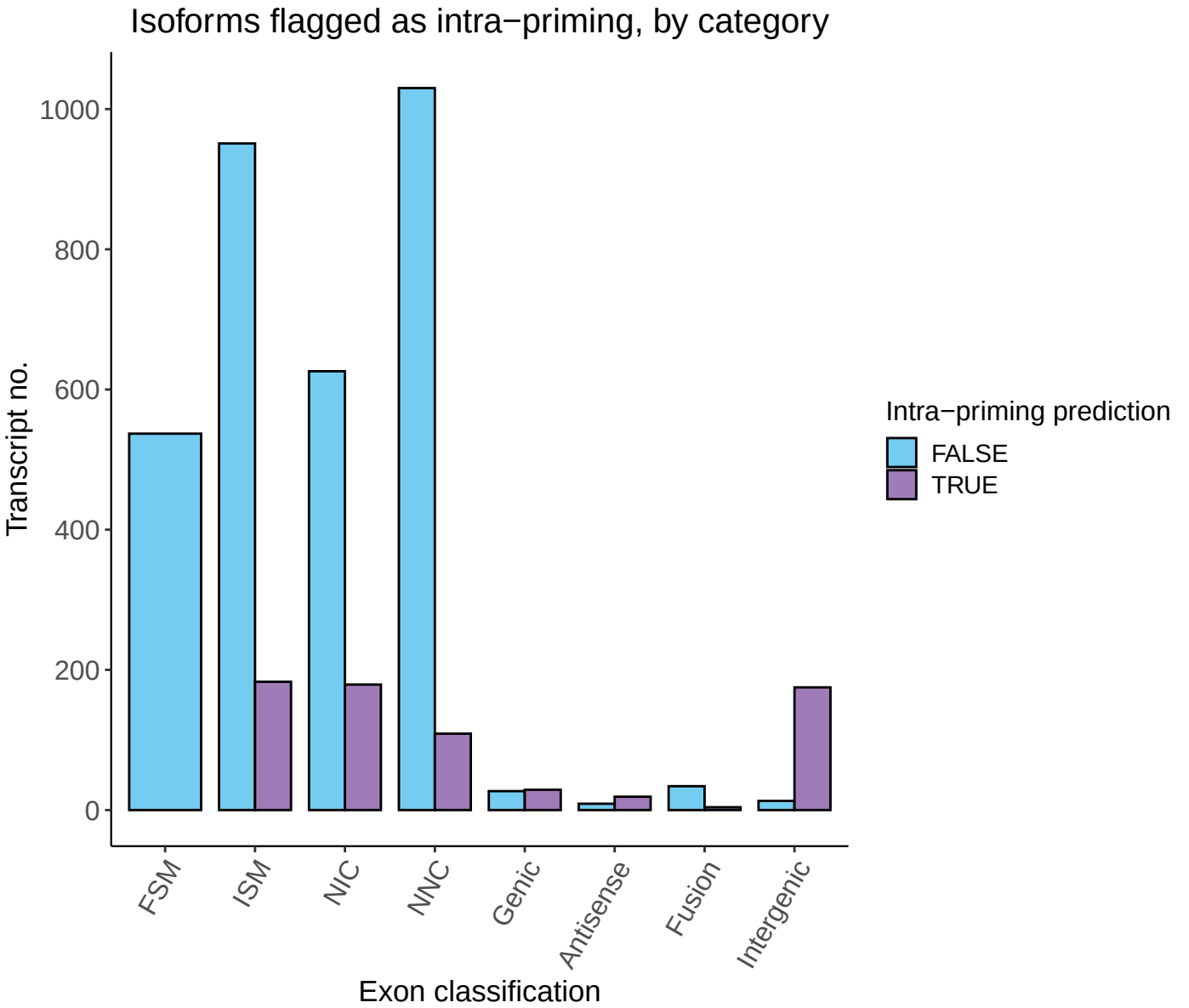

within_CAGE_peak − ML importance: 0.64
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
ntergenic
21 127
174 749
130 77
288 27
13 21
23 4
13 1
169 10
100%
75%
50%
25%
0%
% transcripts
Artifact
Isoform
Filter result
within_CAGE_peak
FALSE
TRUE

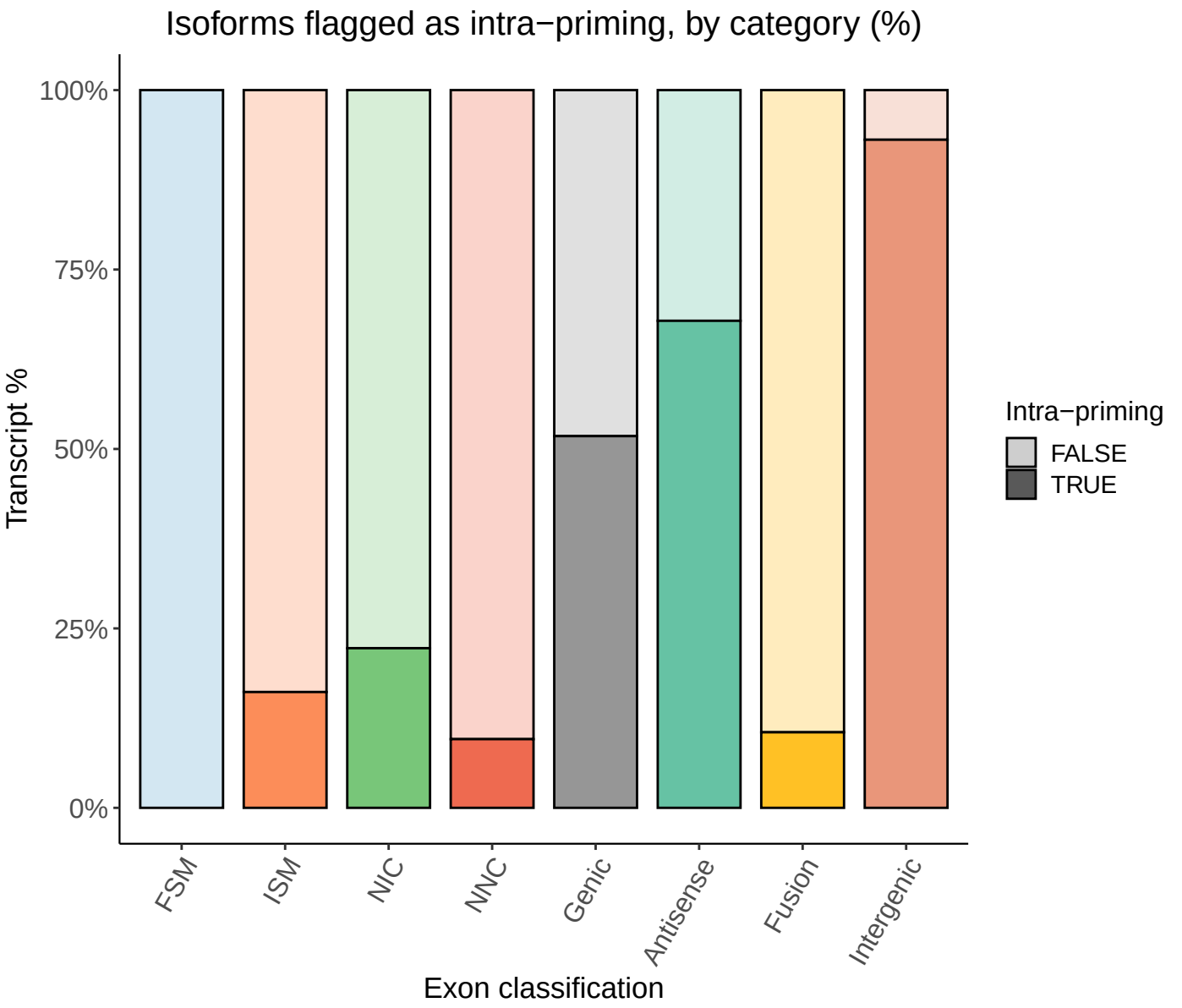

ratio_exp − ML importance: 0.55
log(|ratio_exp| +1)
0.0
0.2
0.4
0.6
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category
Filter result
Artifact
Isoform

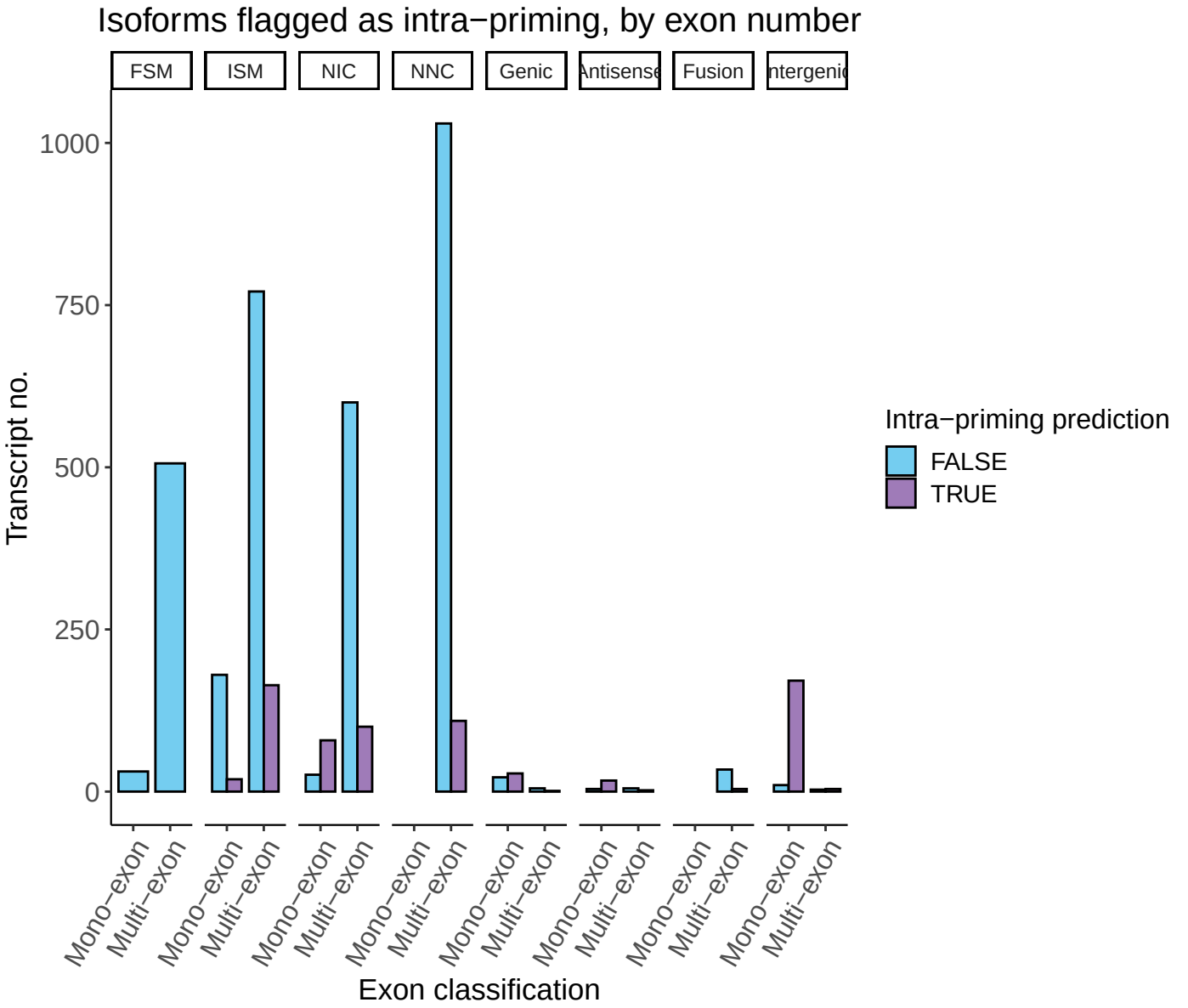

polyA_motif_found − ML importance: 0.46
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
16 63
165 30
163 26
154 4
26 2
16 4
4 5
154 4
100%
75%
50%
25%
0%
% transcripts
Artifact
Isoform
Filter result
polyA_motif_found
FALSE
TRUE

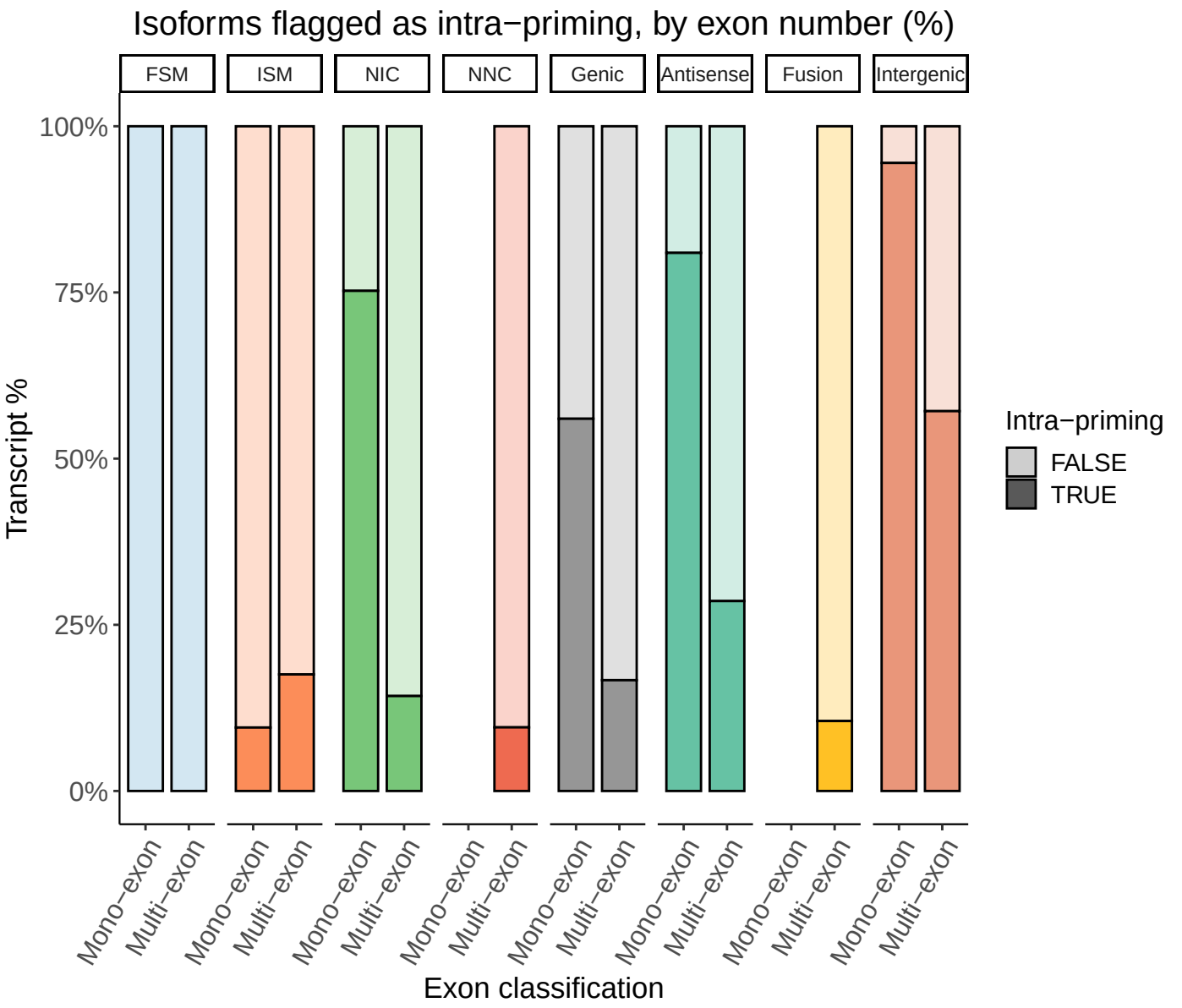

coding − ML importance: 0.14
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
76 415
242 628
361 327
952 93
16 1
8 1
25 3
32 10
100%
75%
50%
25%
0%
% transcripts
Artifact
Isoform
Filter result
coding
coding
non_coding


A % by category
Red line indicates threshold employed in ML filter
% A's downstream of TTS
100
75
50
25
0
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Structural category


Isoforms flagged as intra−priming, by category
Transcript no.
1000
800
600
400
200
0
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Exon classification
Intra−priming prediction
FALSE
TRUE


Isoforms flagged as intra−priming, by category (%)
100%
75%
50%
25%
0%
Transcript %
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
Exon classification
Intra−priming
FALSE
TRUE


Isoforms flagged as intra−priming, by exon number
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
1000
750
500
250
0
Transcript no.
Mono−exon
Multi−exon
Exon classification
Intra−priming prediction
FALSE
TRUE


Isoforms flagged as intra−priming, by exon number (%)
FSM
ISM
NIC
NNC
Genic
Antisense
Fusion
Intergenic
100%
75%
50%
25%
0%
Transcript %
Mono−exon
Multi−exon
Exon classification
Intra−priming
FALSE
TRUE