

# *SQANTI3 filter report*

*Total Genes: 650*

*Total Transcripts: 3925*

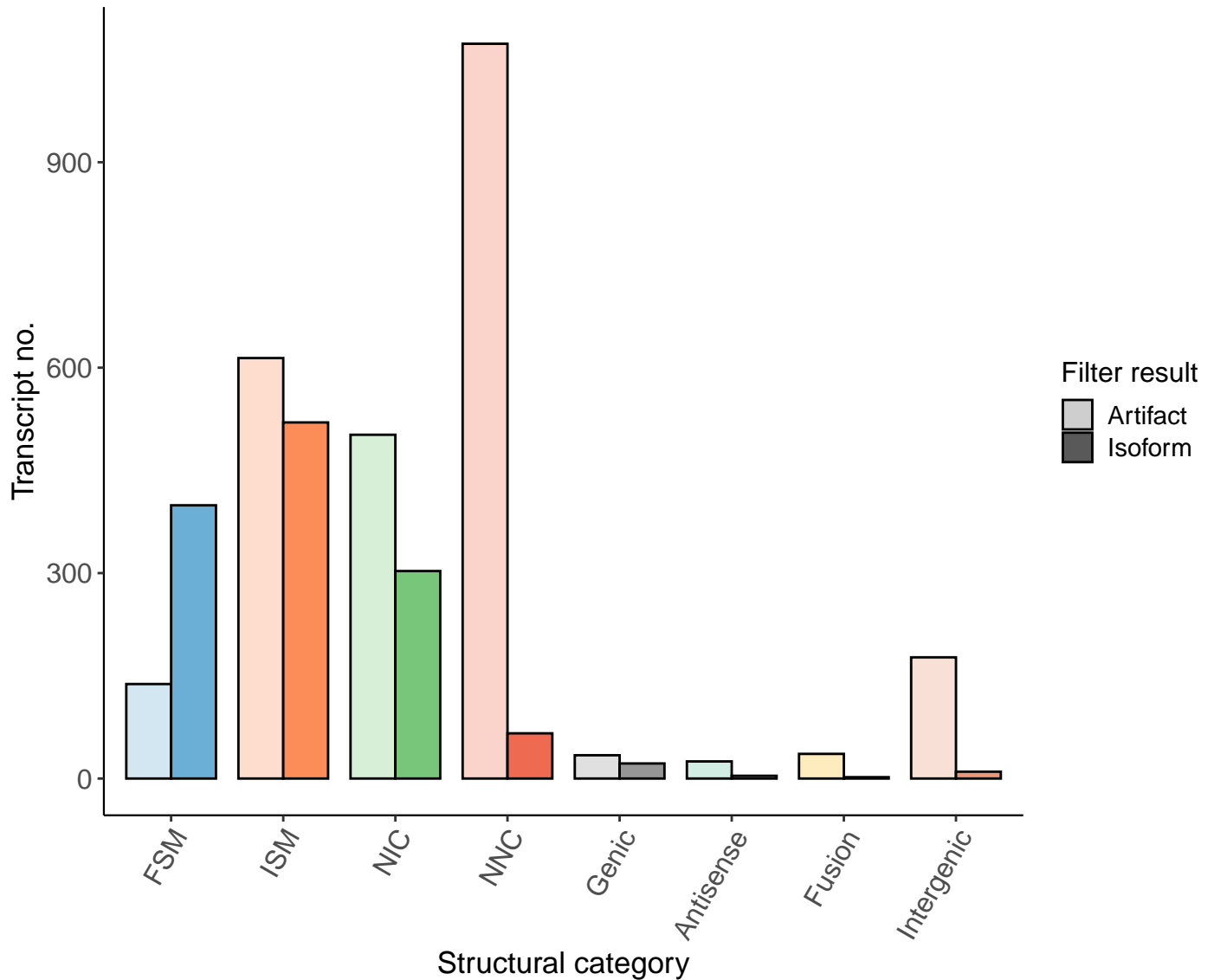
*– Isoforms: 1326 (34%)*

*– Artifacts: 2599 (66%)*

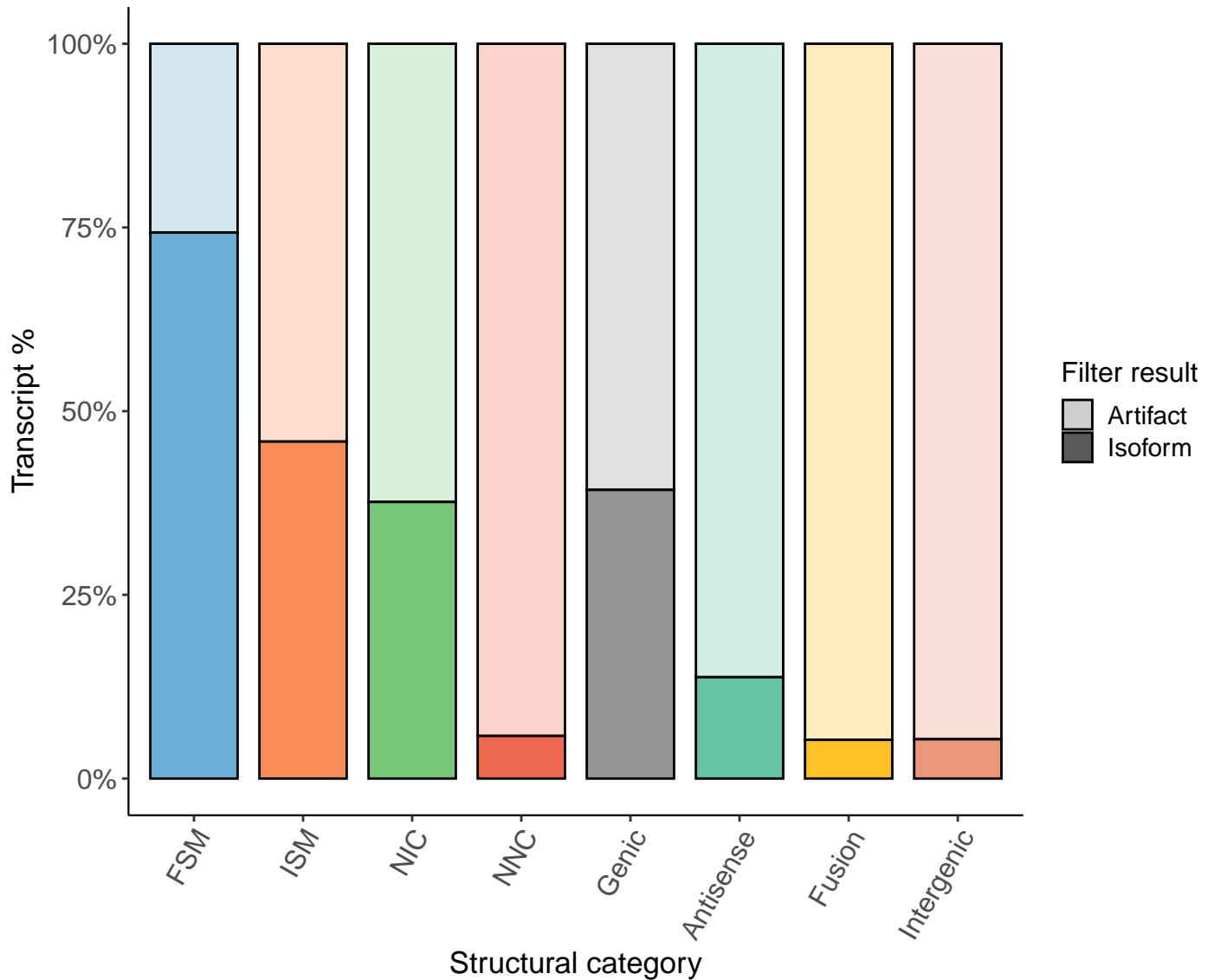
Gene category	Gene no.	No. of genes with artifacts only
Annotated	437	100
Novel	213	199

Structural category	Artifact no.	Isoform no.
FSM	138	399
ISM	614	520
NIC	502	303
NNC	1073	66
Genic	34	22
Antisense	25	4
Fusion	36	2
Intergenic	177	10

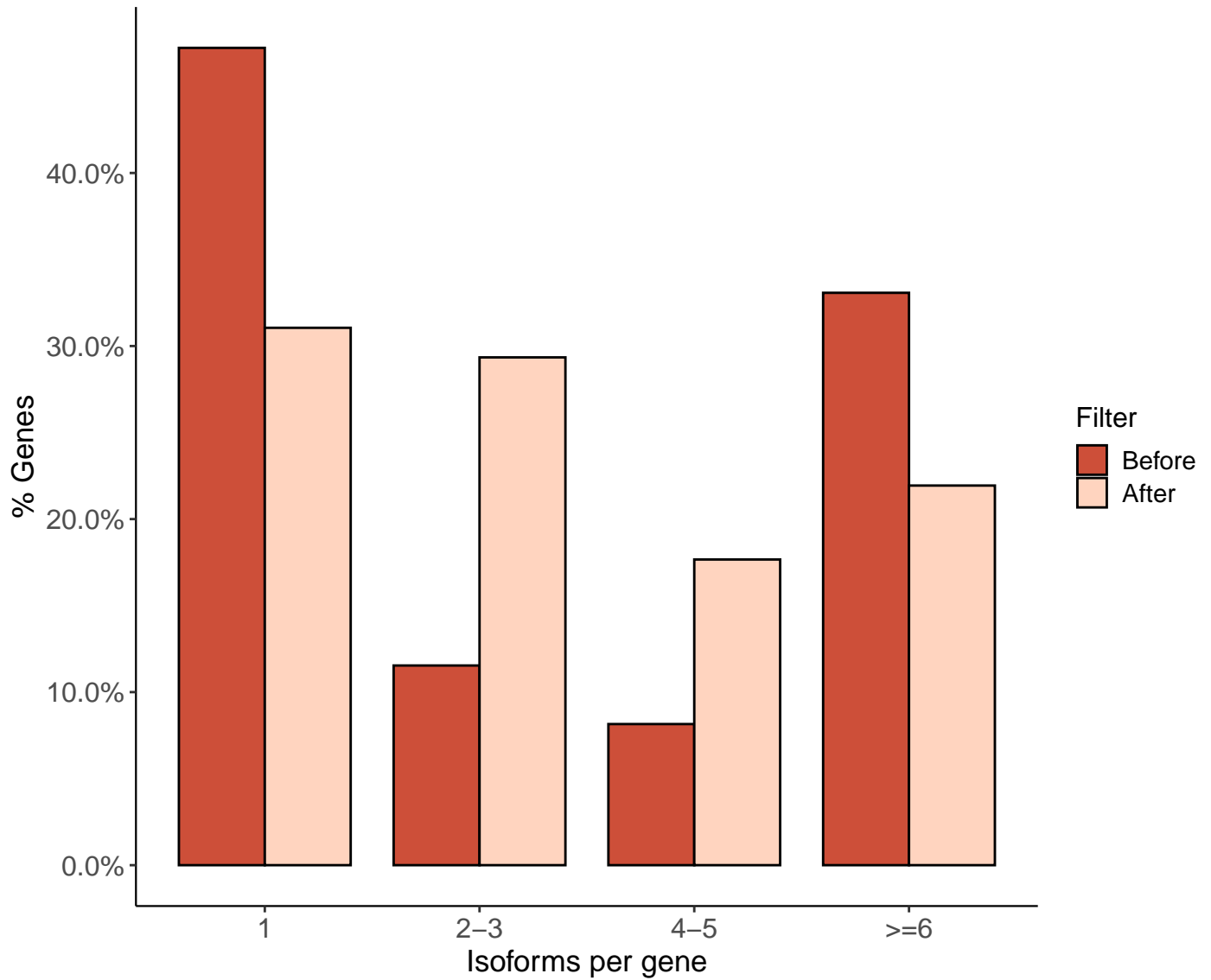
Total isoforms and artifacts by category



% isoforms and artifacts by category

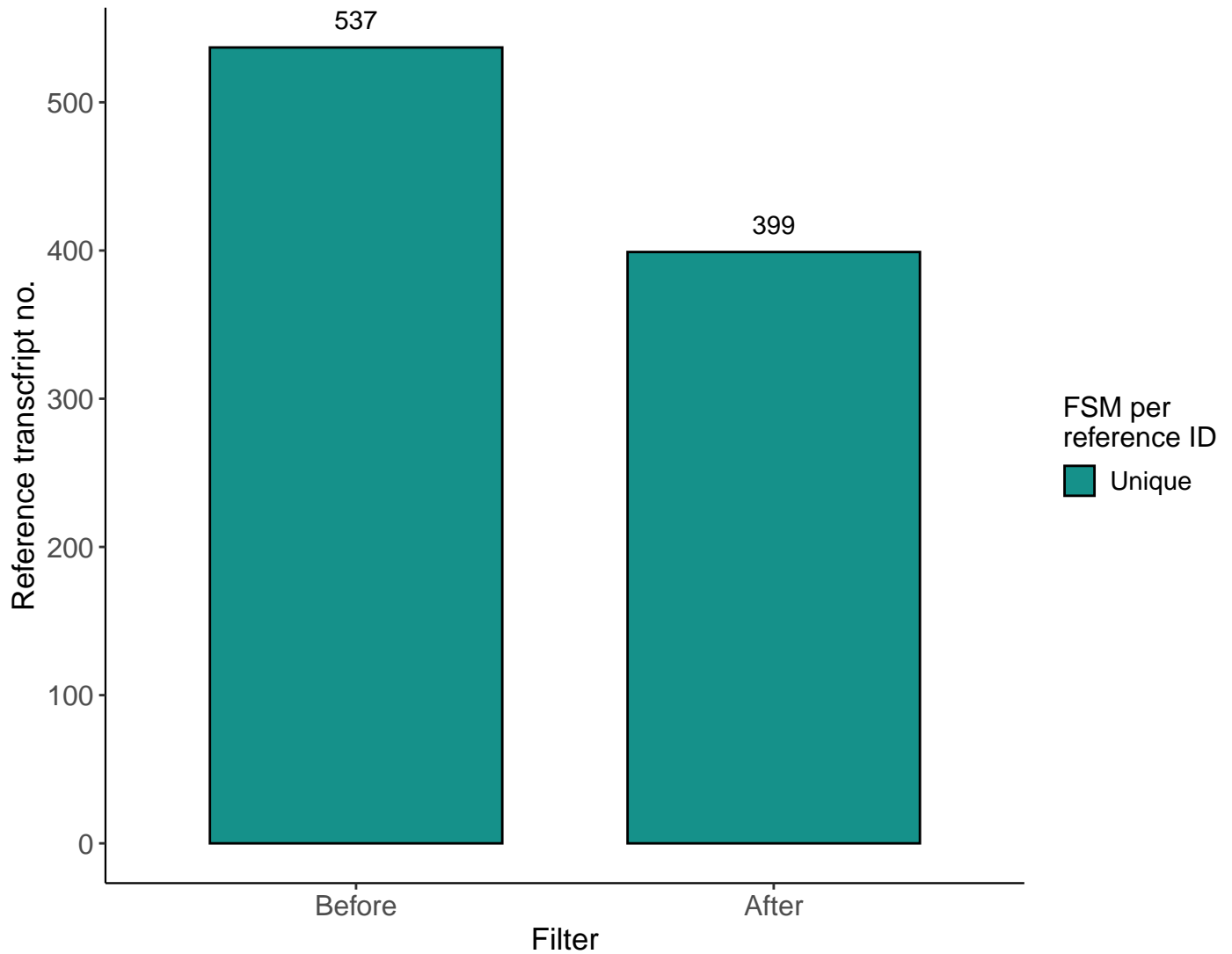


Number of isoforms per gene

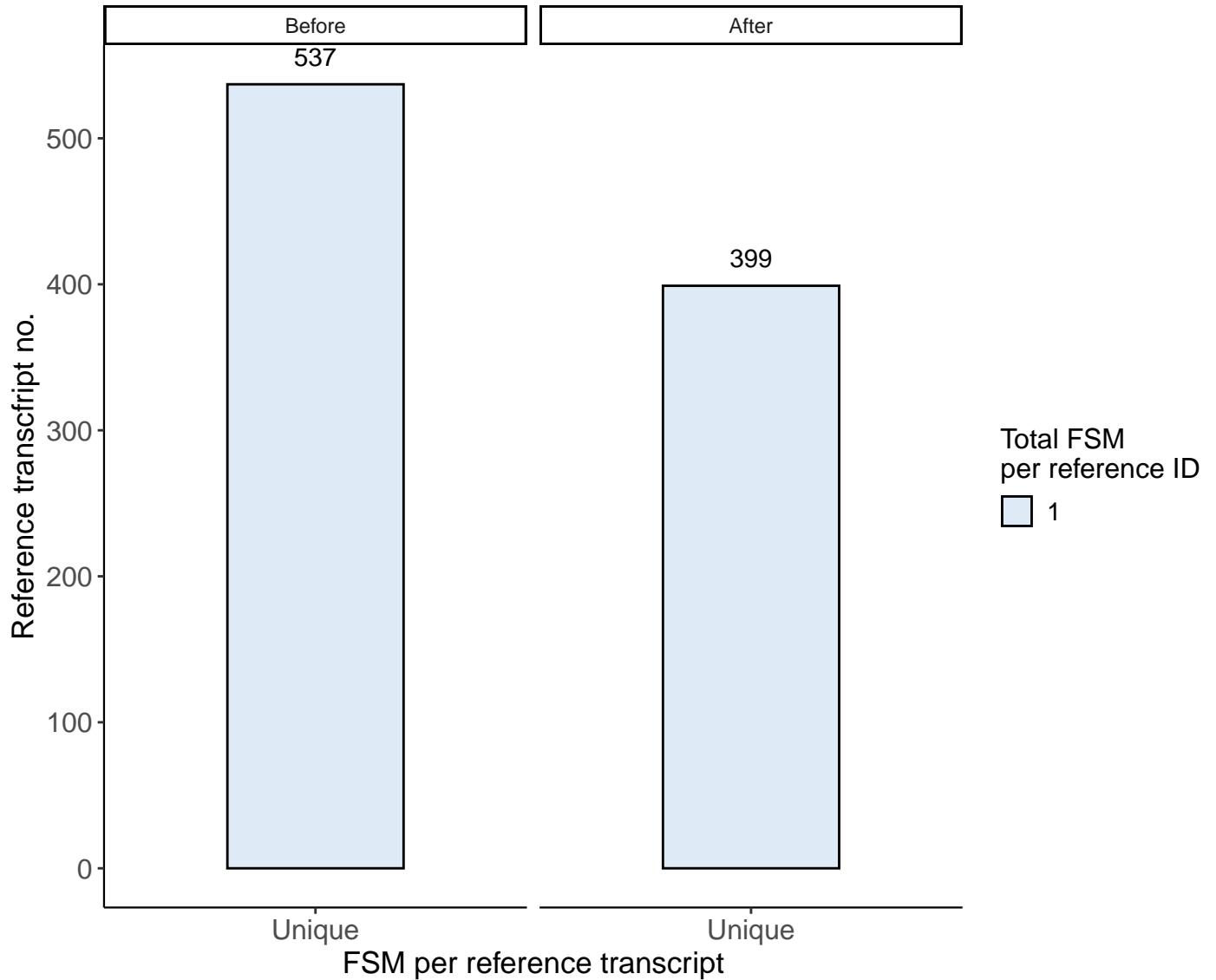


## Reference transcript complexity

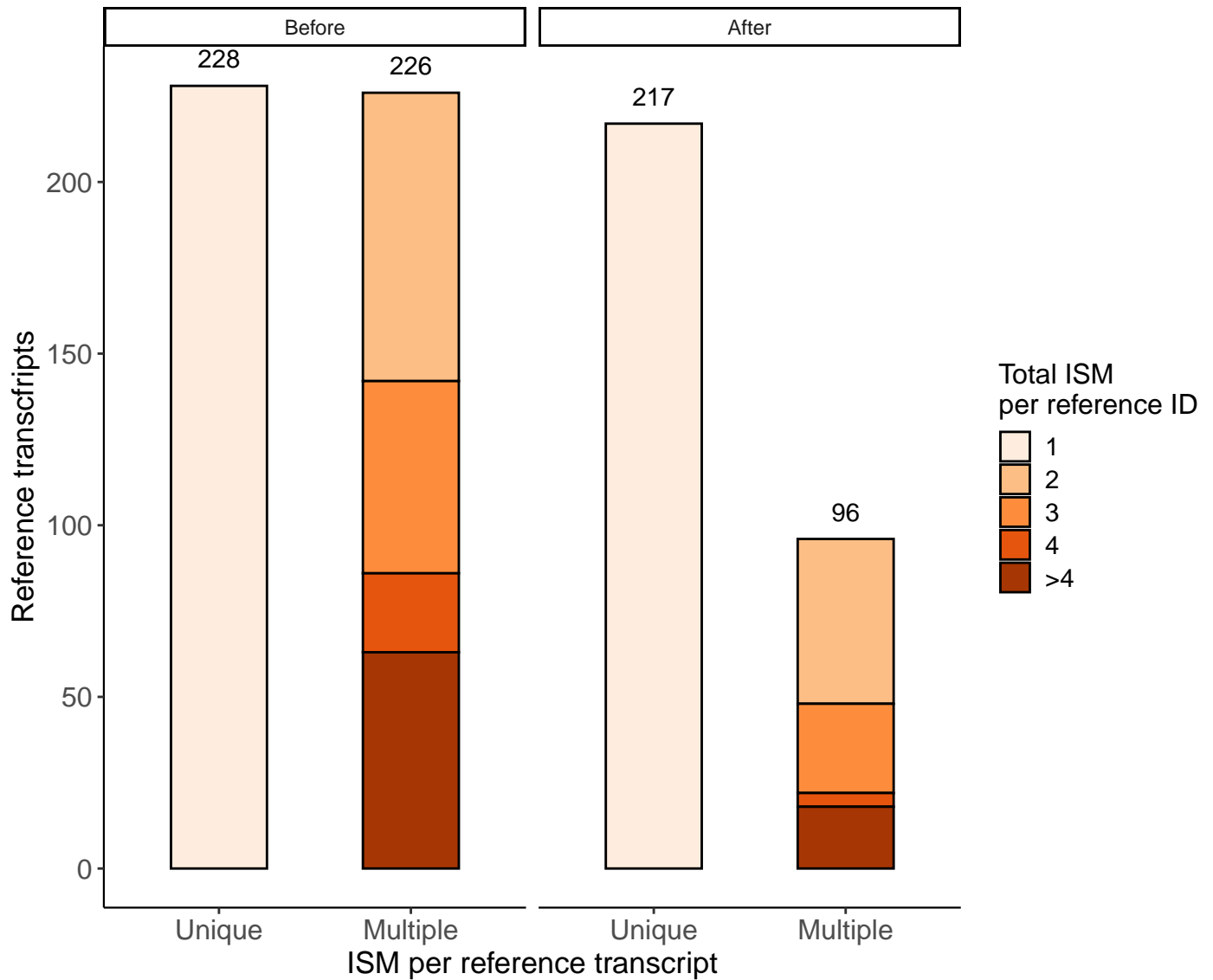
*No. of reference transcripts represented by FSM*



## FSM redundancy

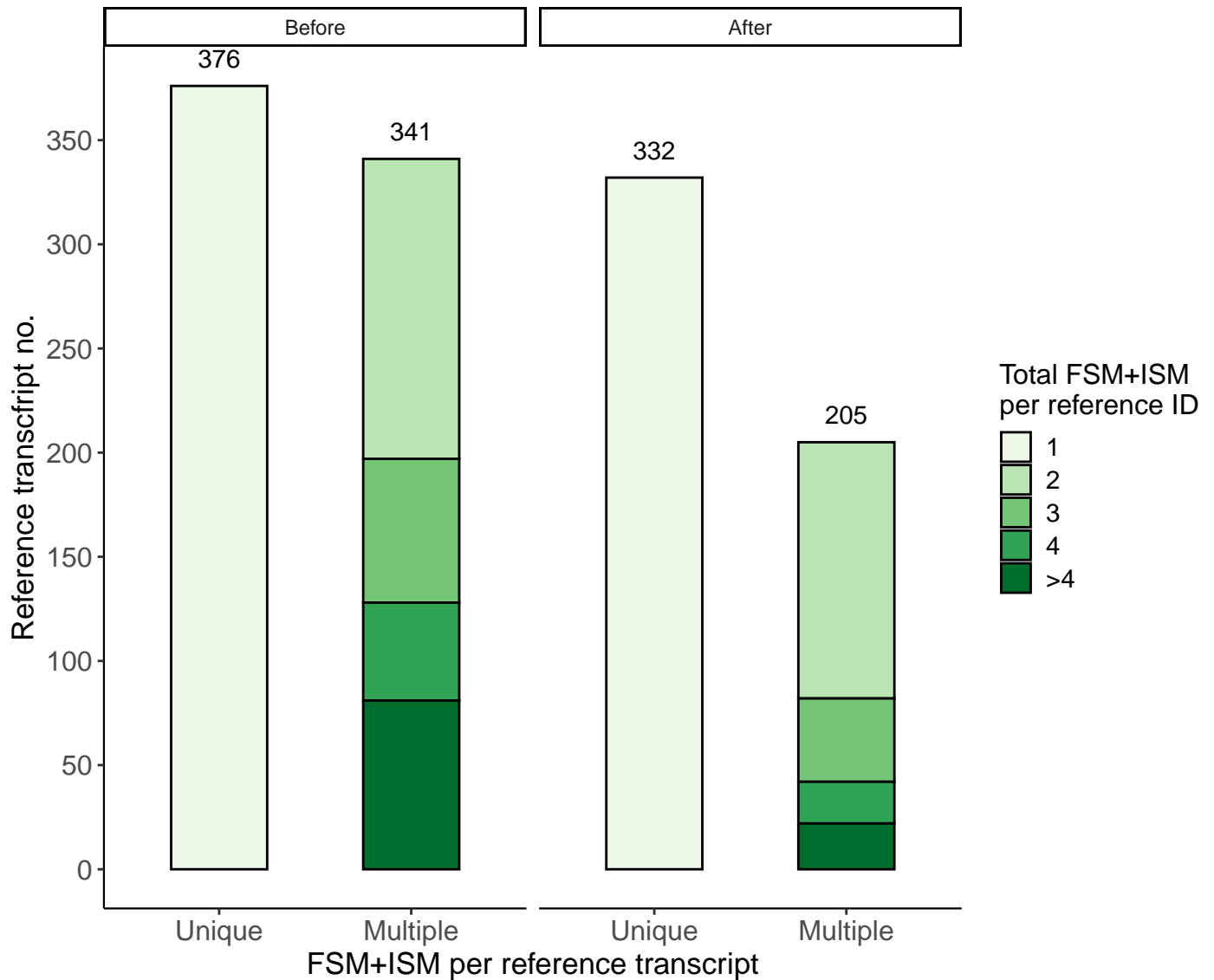


## ISM redundancy

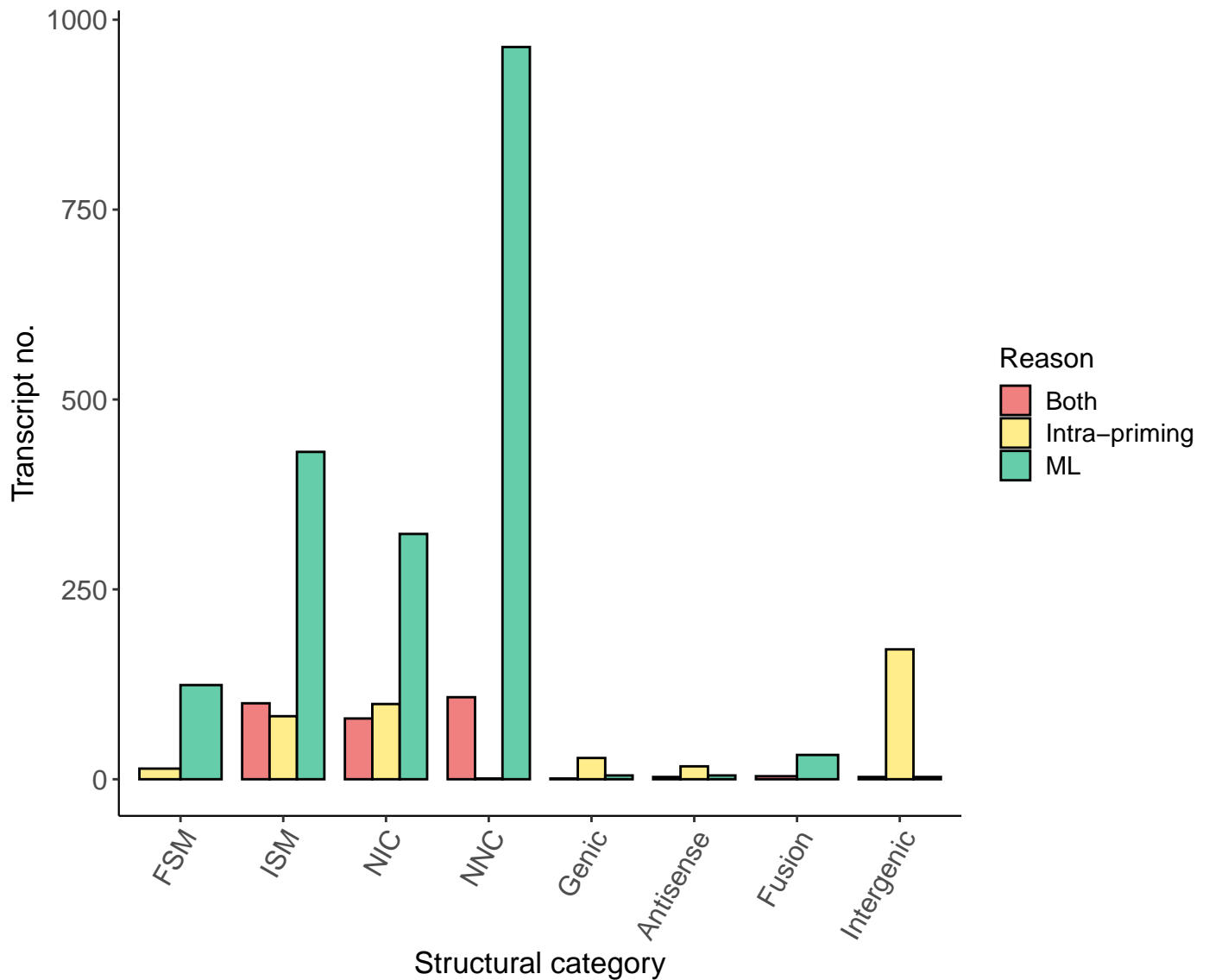




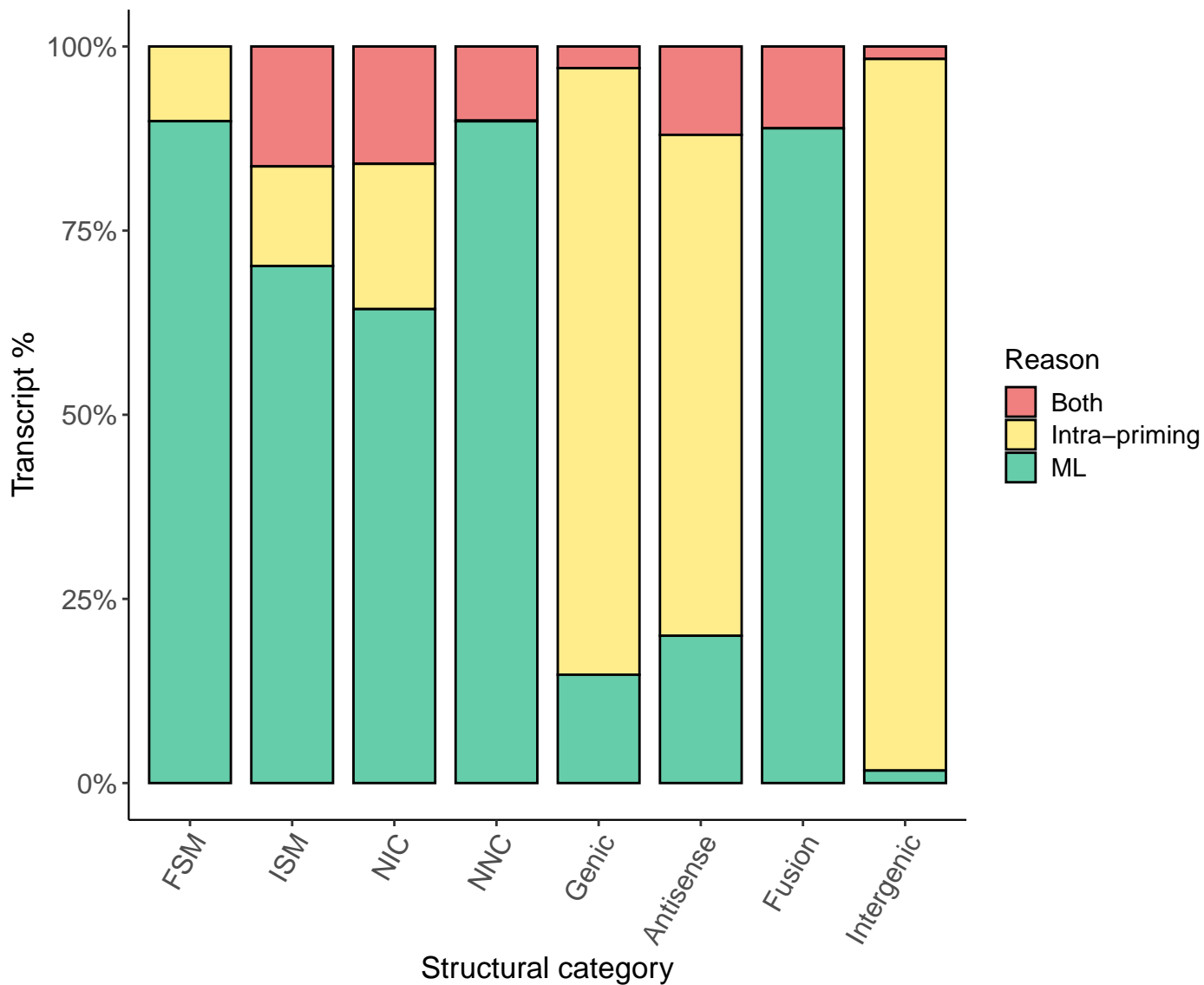
## FSM+ISM redundancy



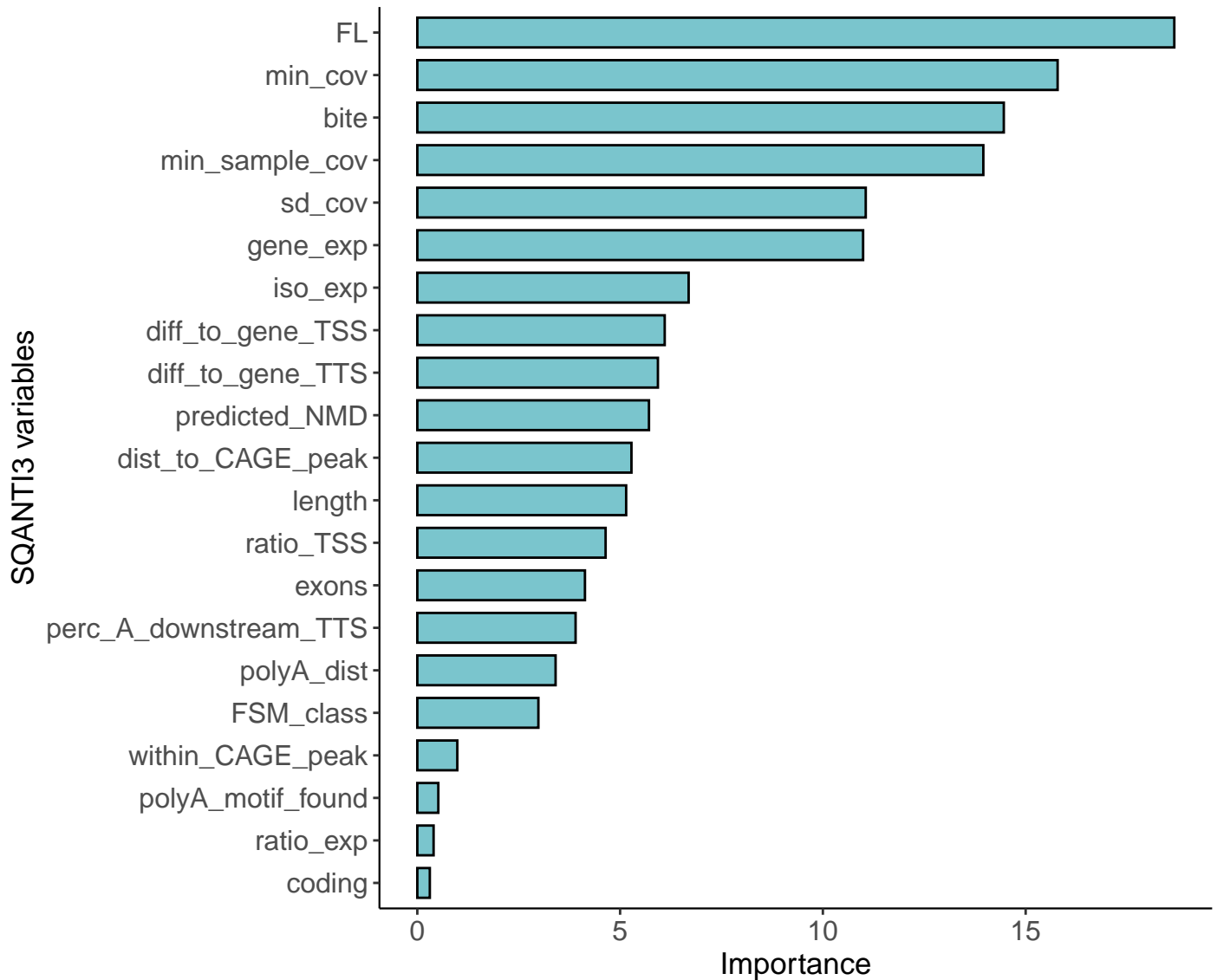
Reason to flag transcripts as artifacts, by category



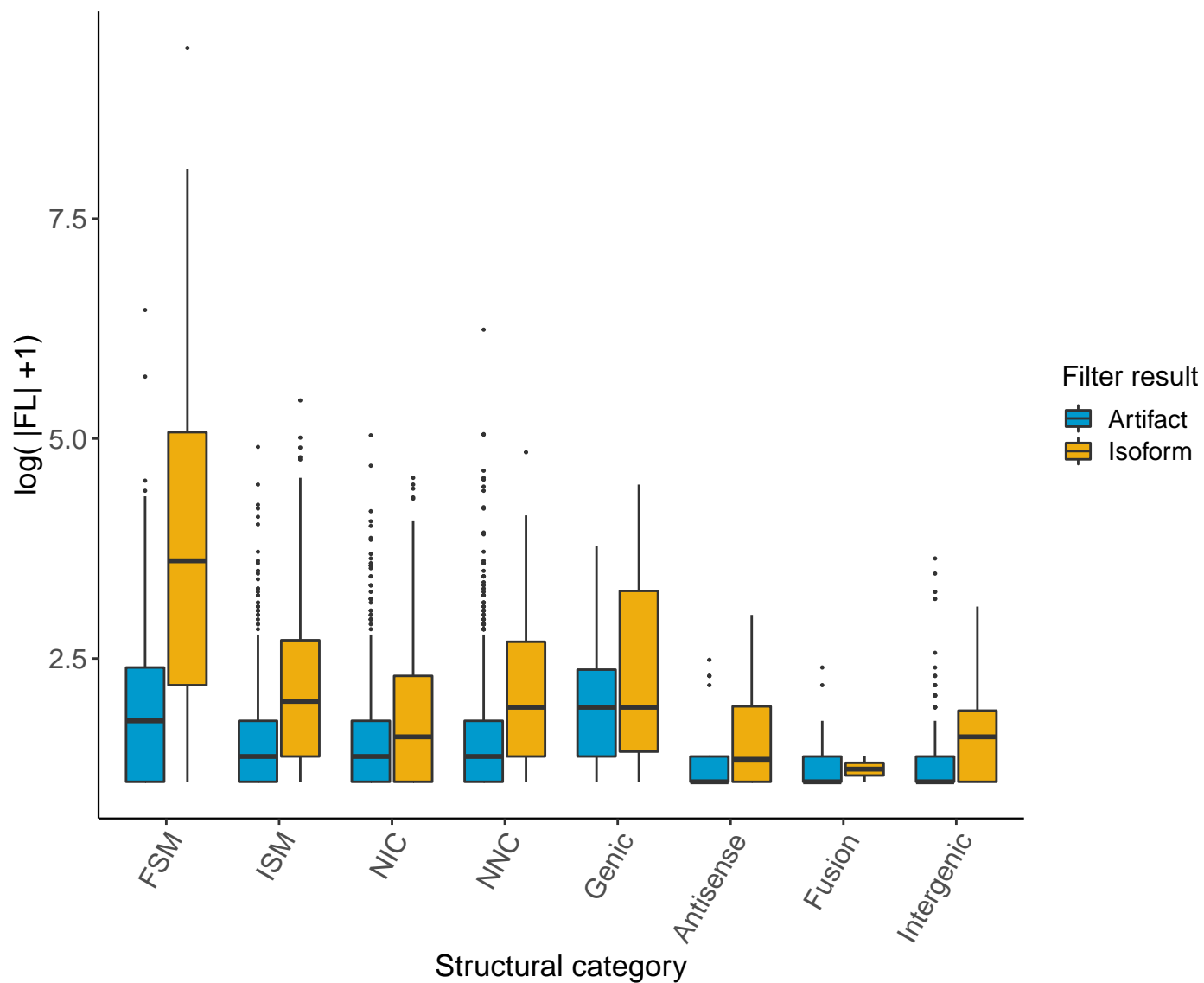
Reason to flag transcripts as artifacts, by category



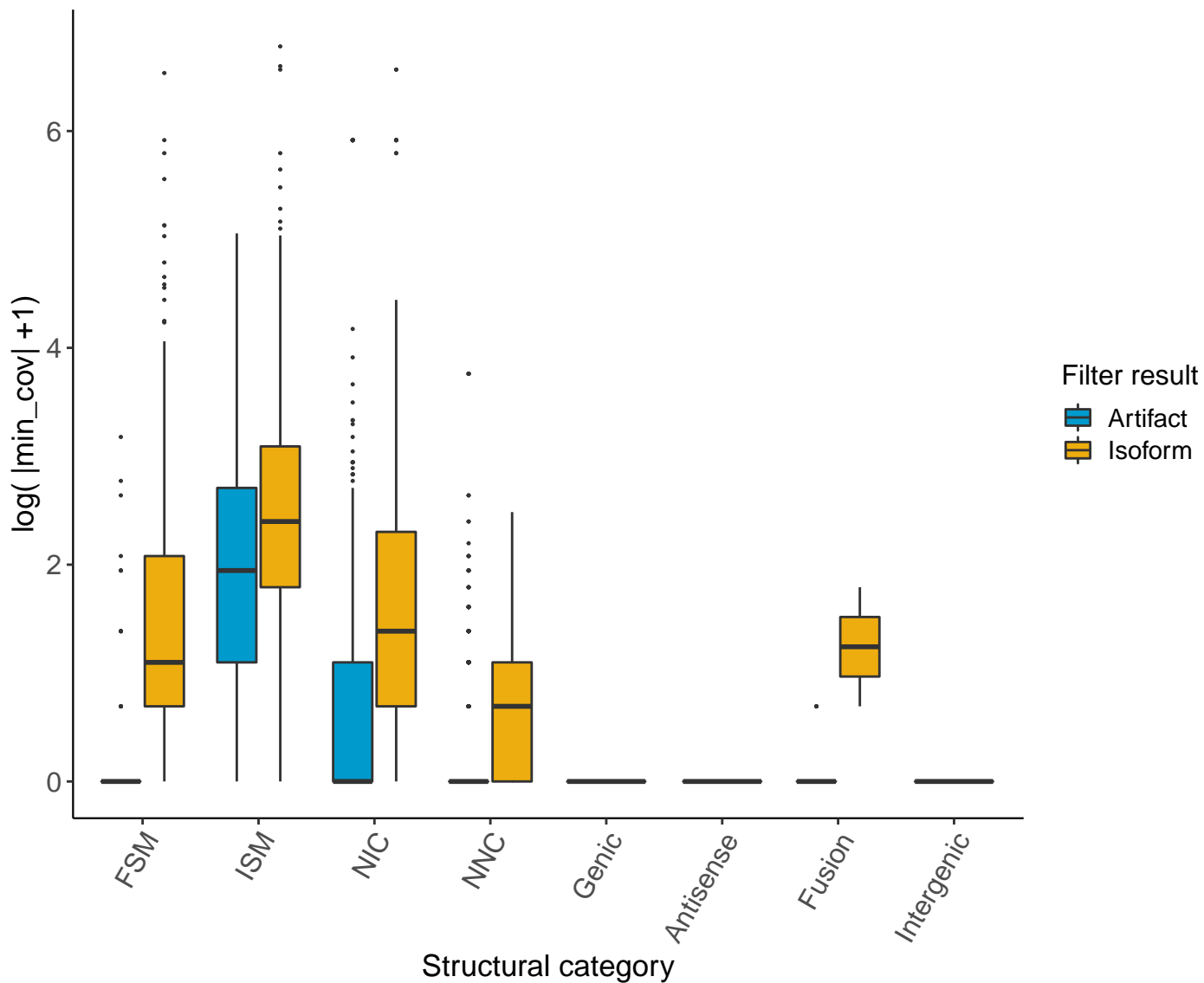
Variable importance in Random Forest classifier



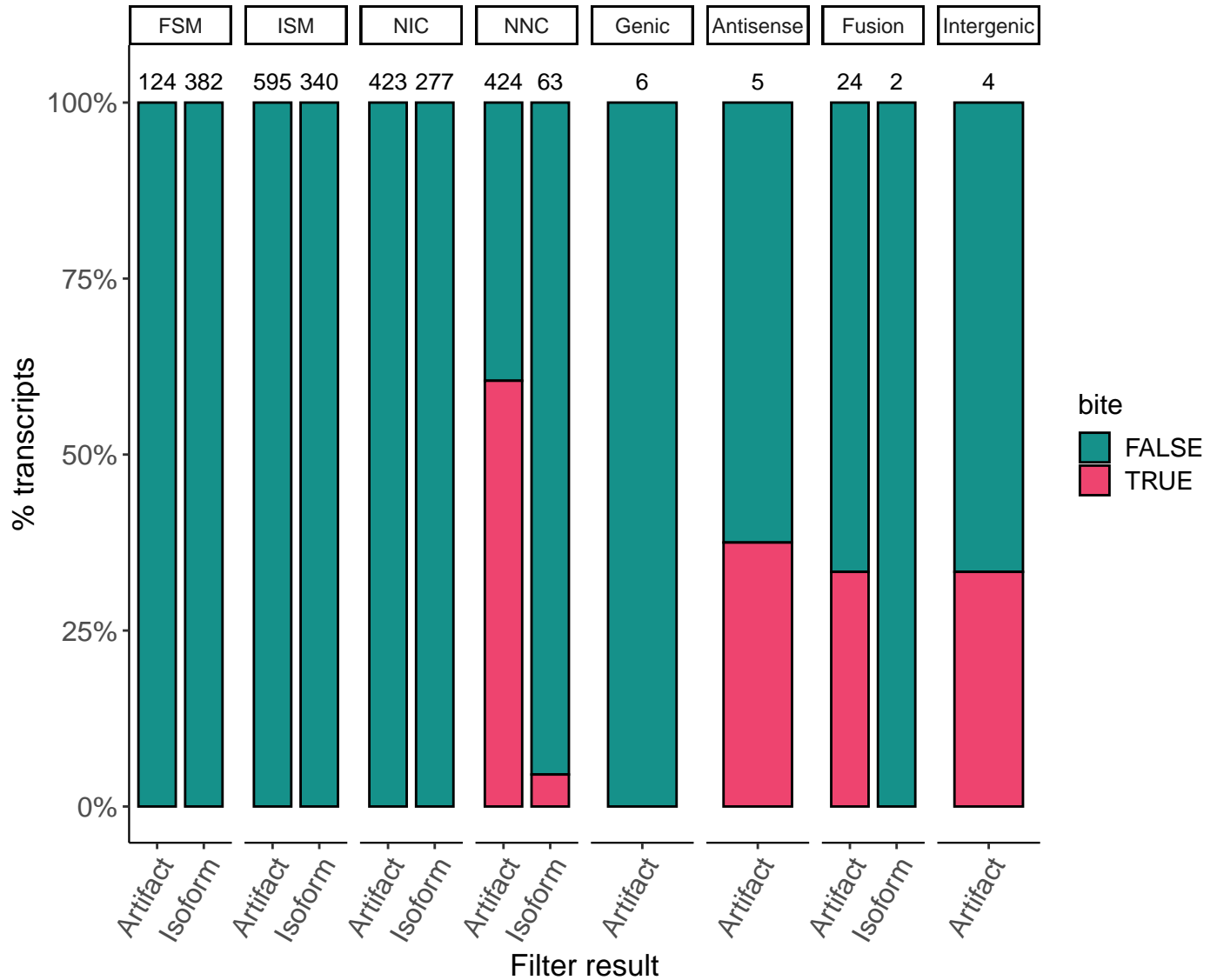
FL – ML importance: 18.67



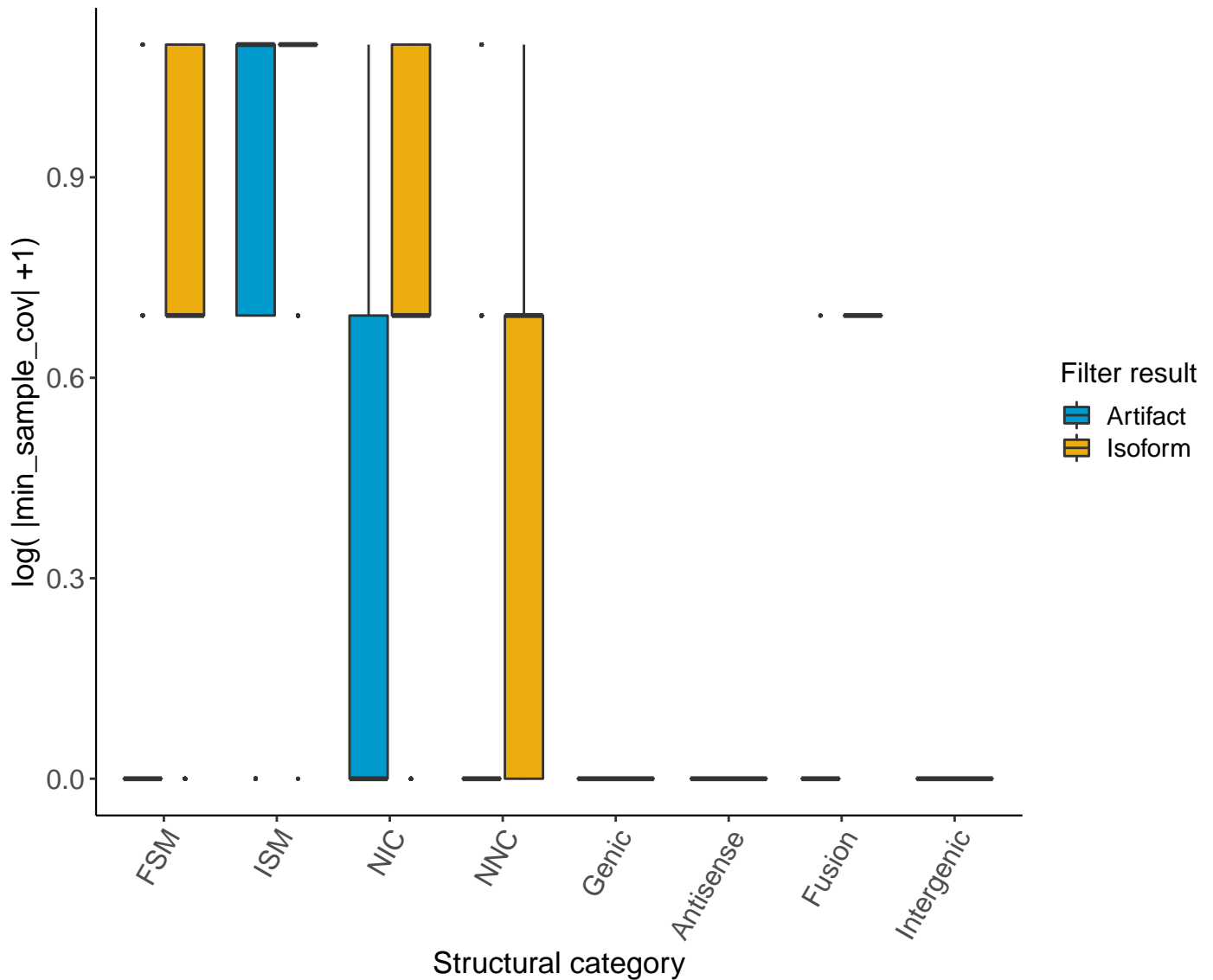
min\_cov – ML importance: 15.79



bite – ML importance: 14.46

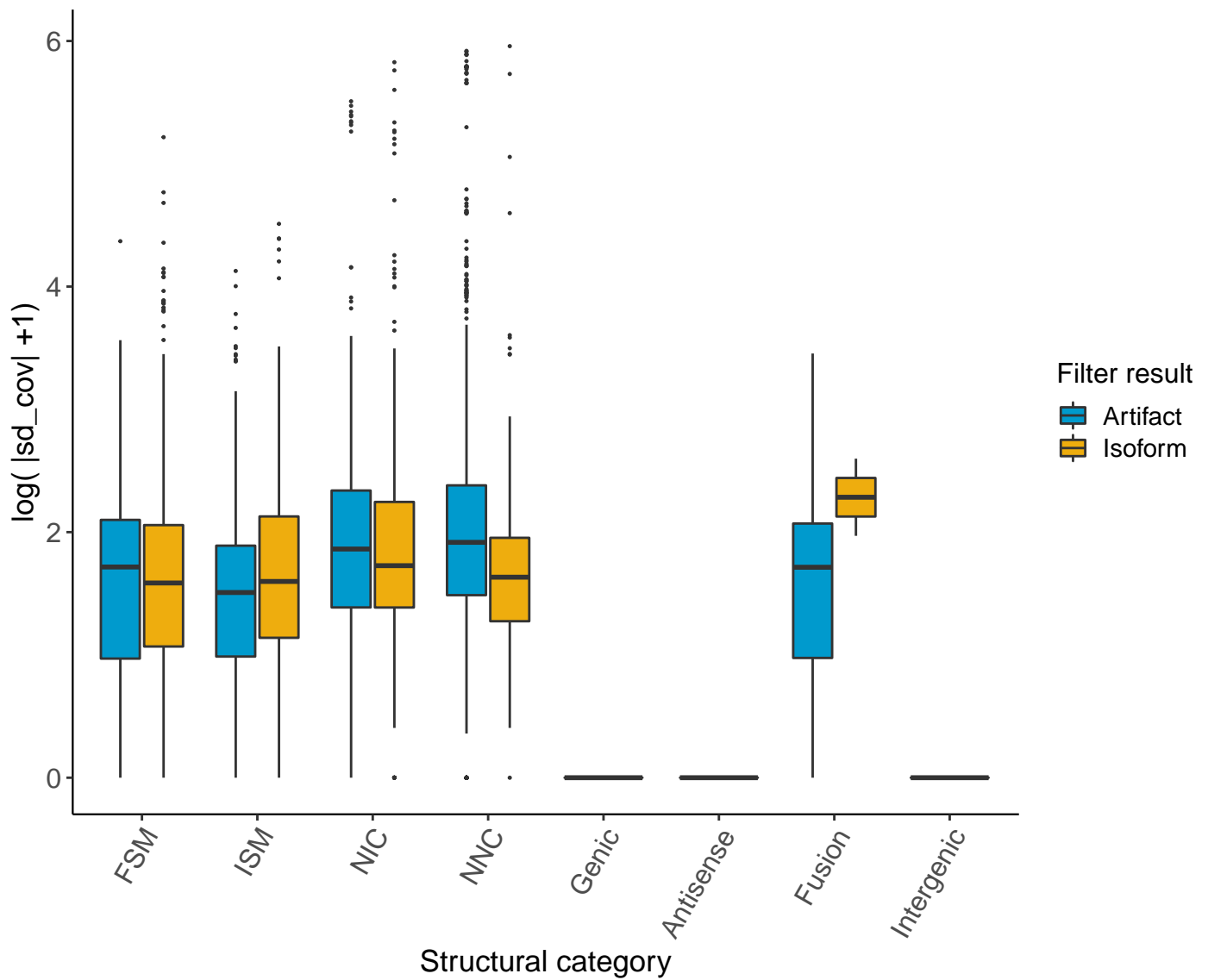


min\_sample\_cov – ML importance: 13.96

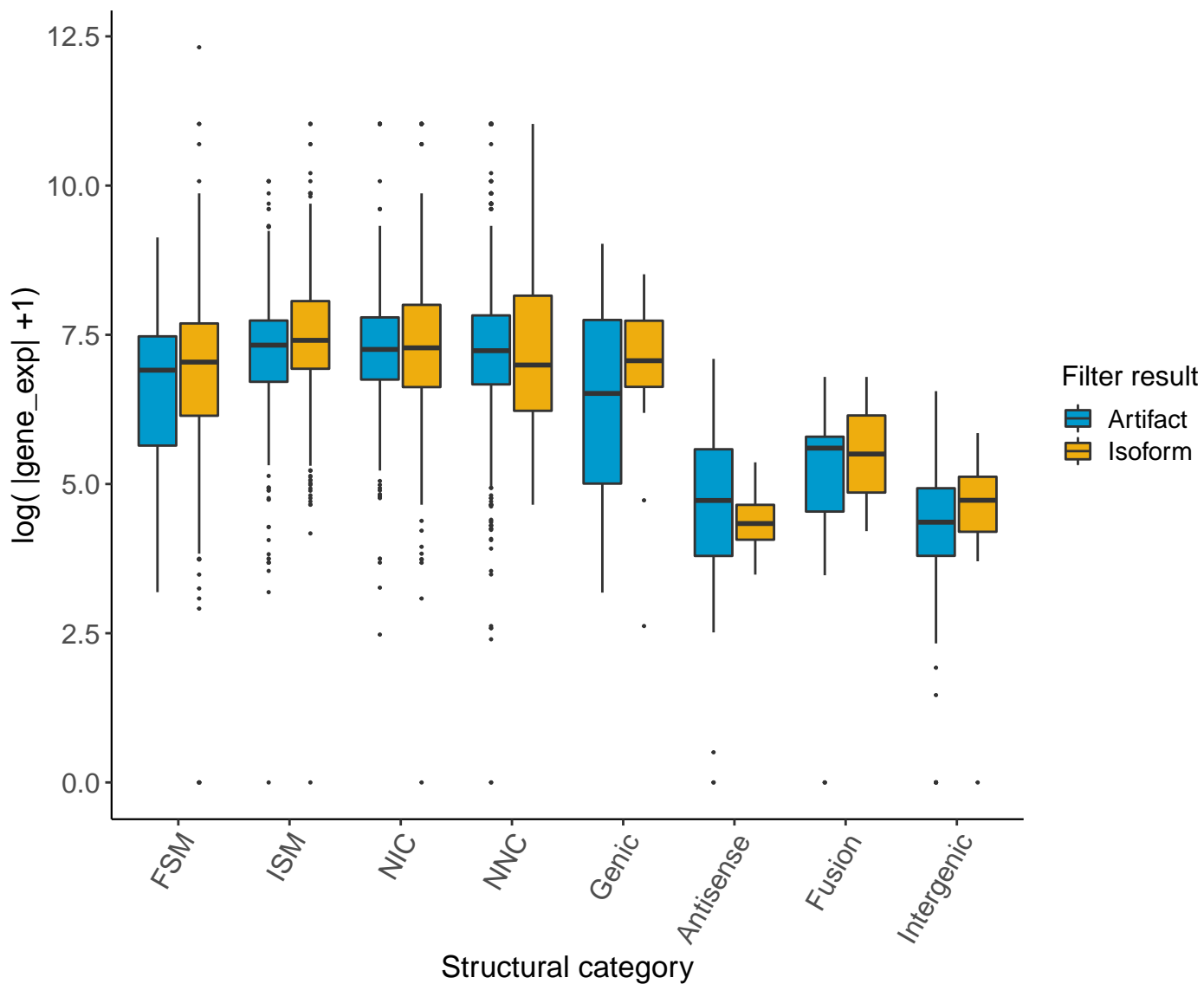




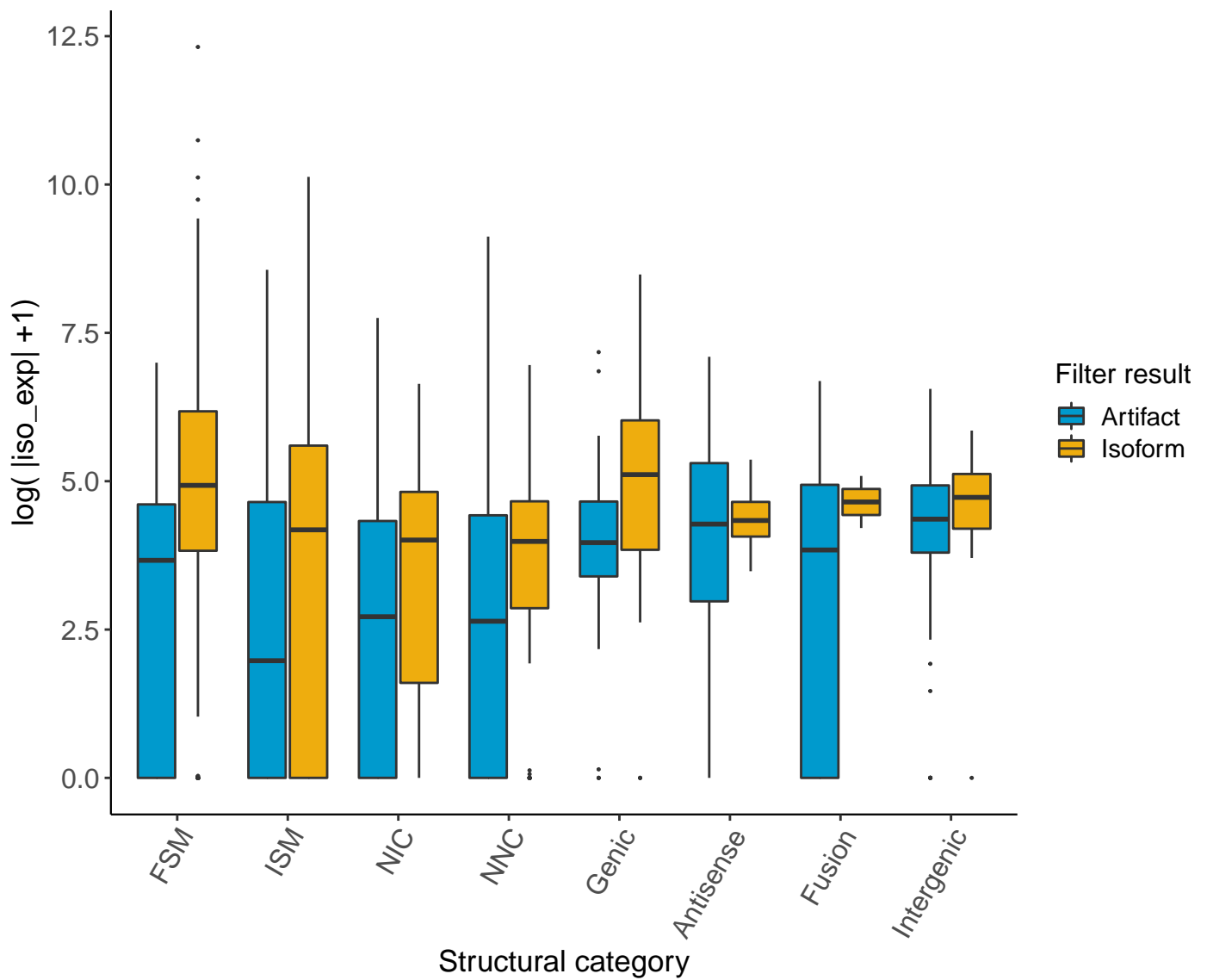
sd\_cov – ML importance: 11.06



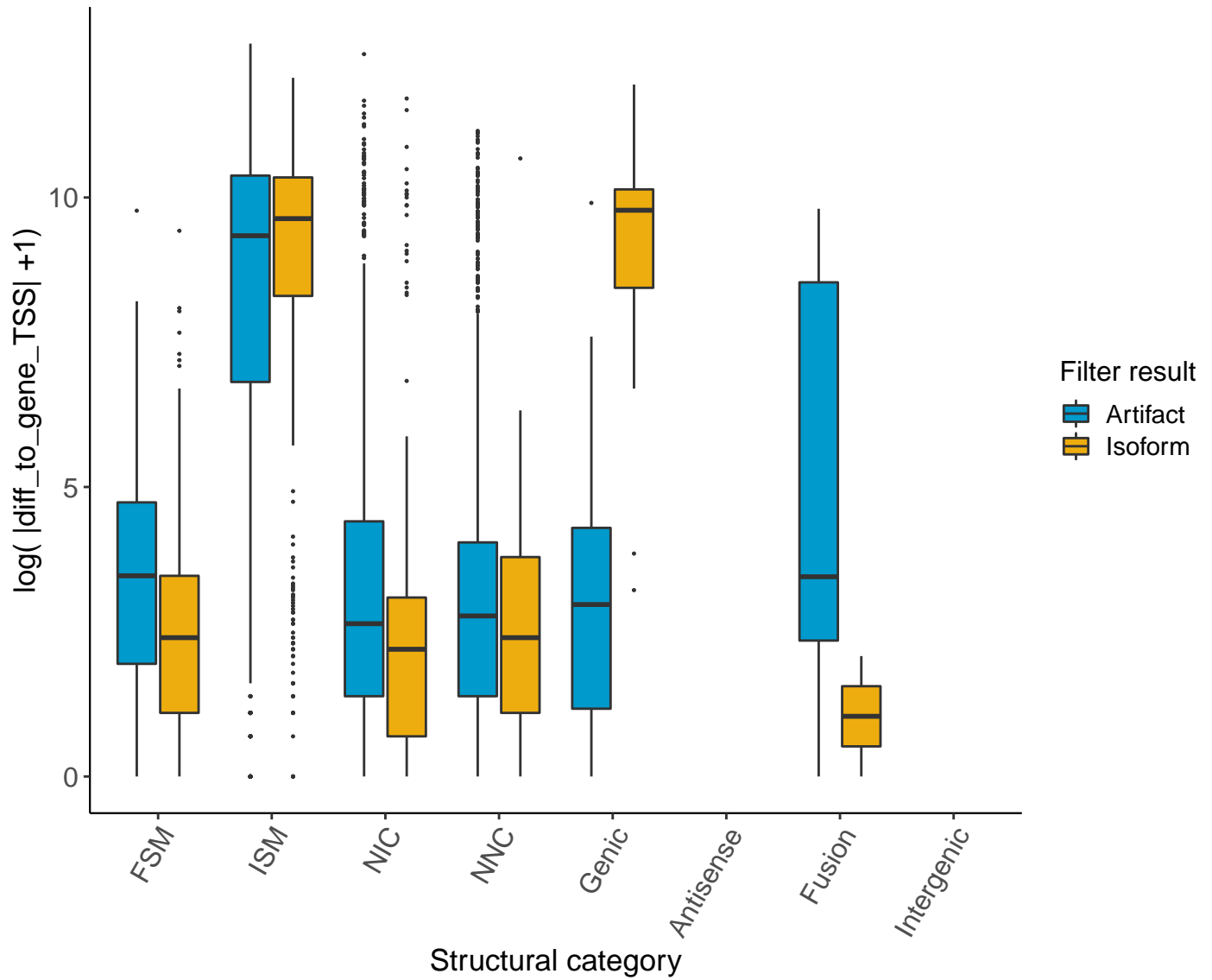
gene\_exp – ML importance: 10.99



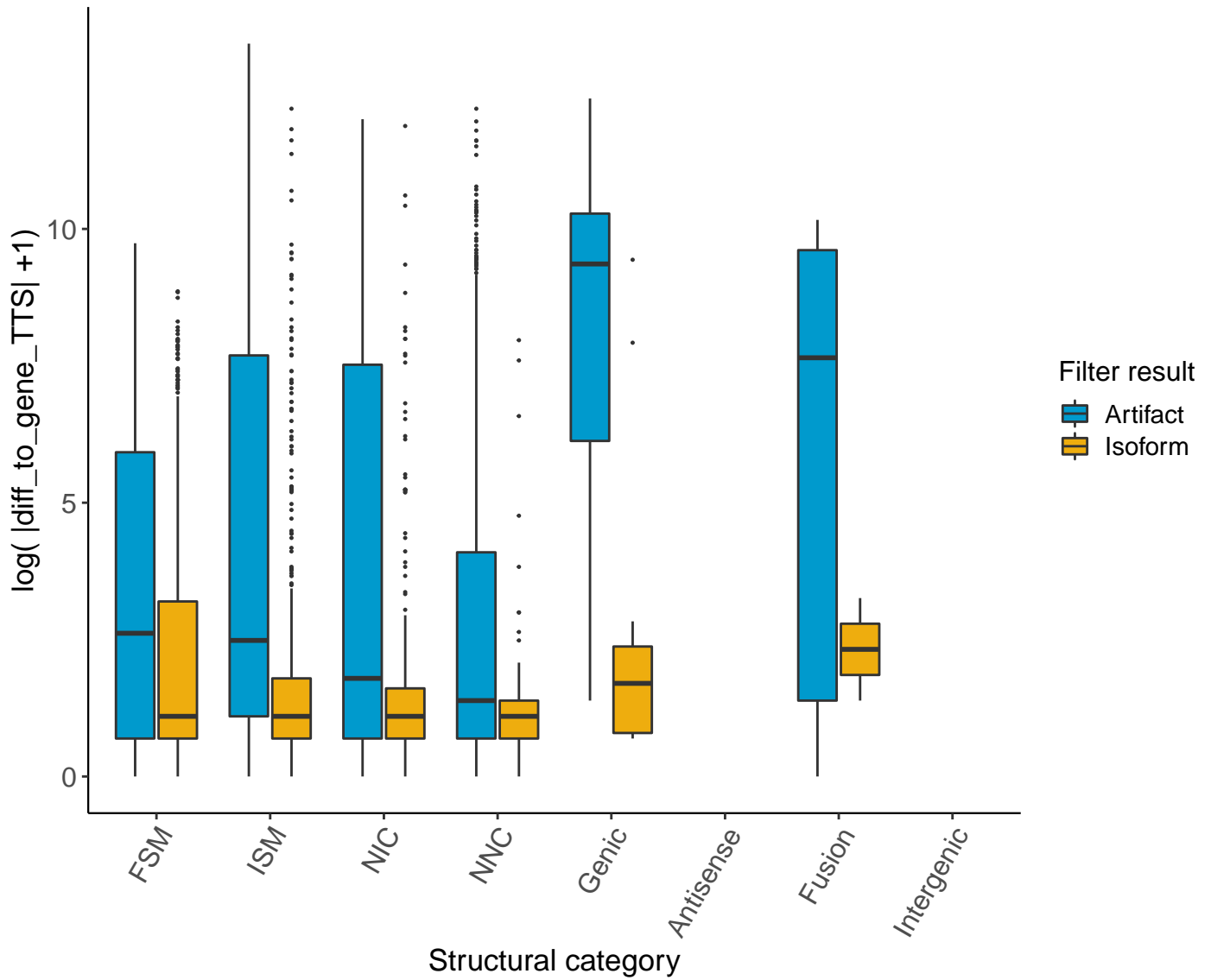
iso\_exp – ML importance: 6.69



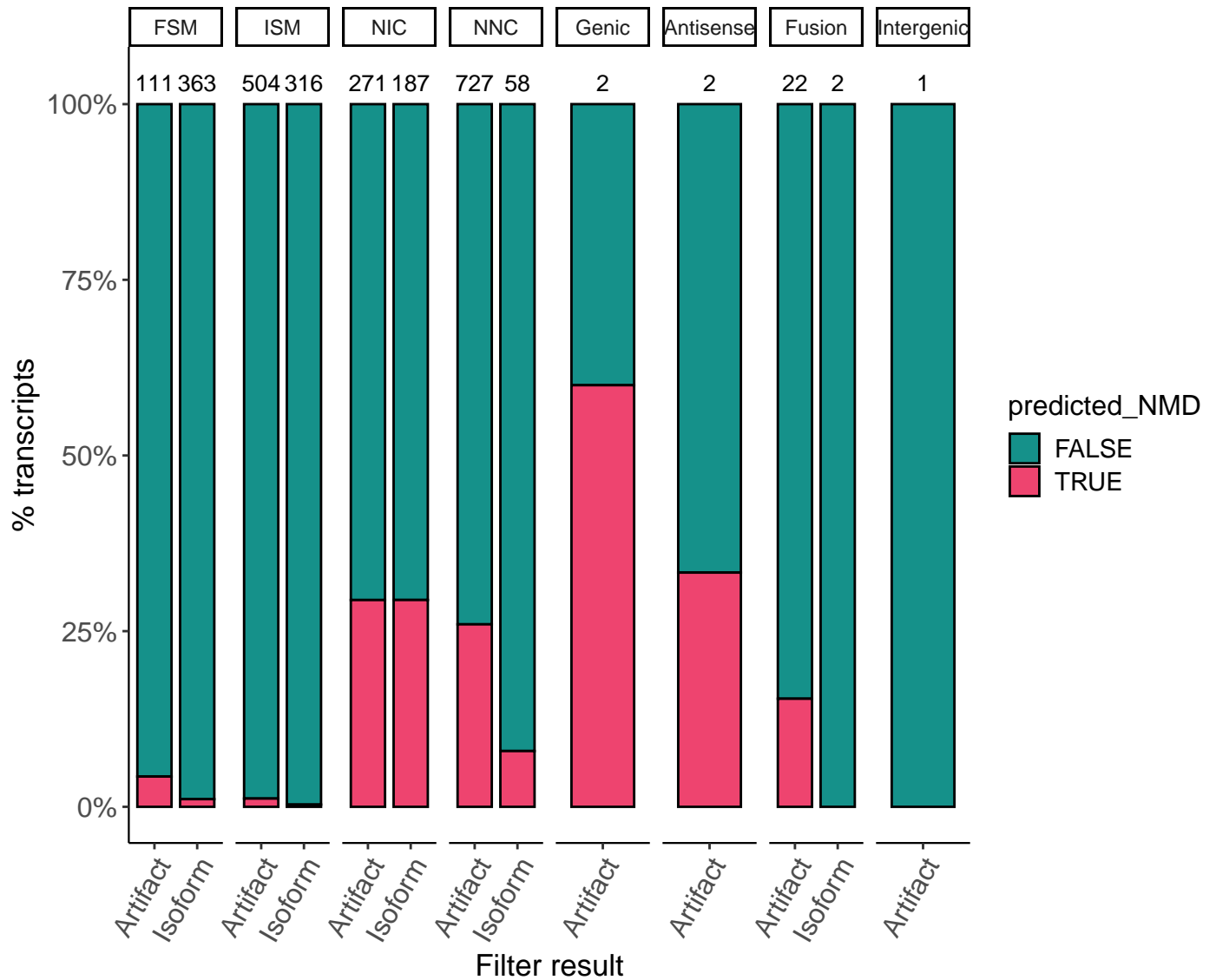
diff\_to\_gene\_TSS – ML importance: 6.1



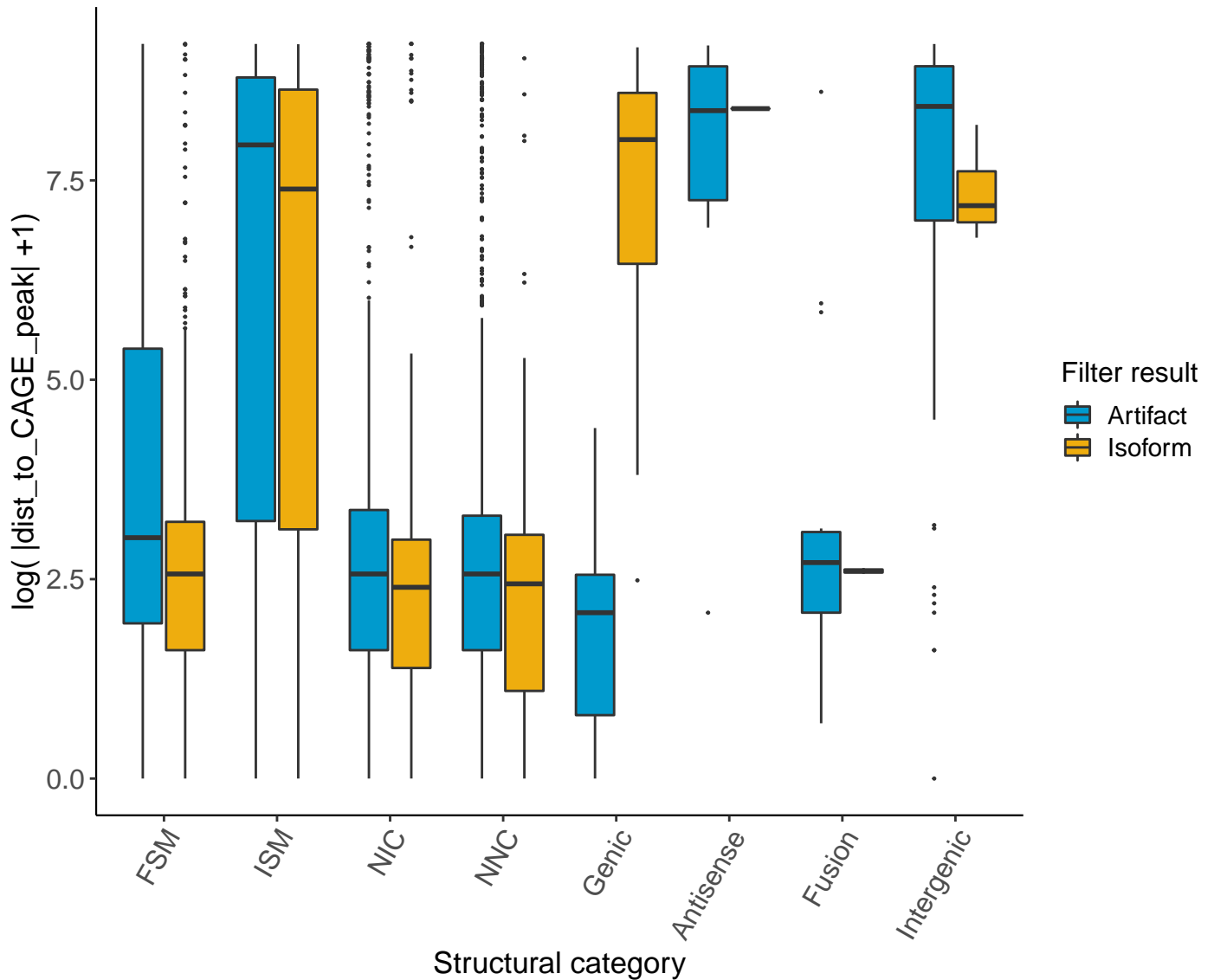
diff\_to\_gene\_TTS – ML importance: 5.93



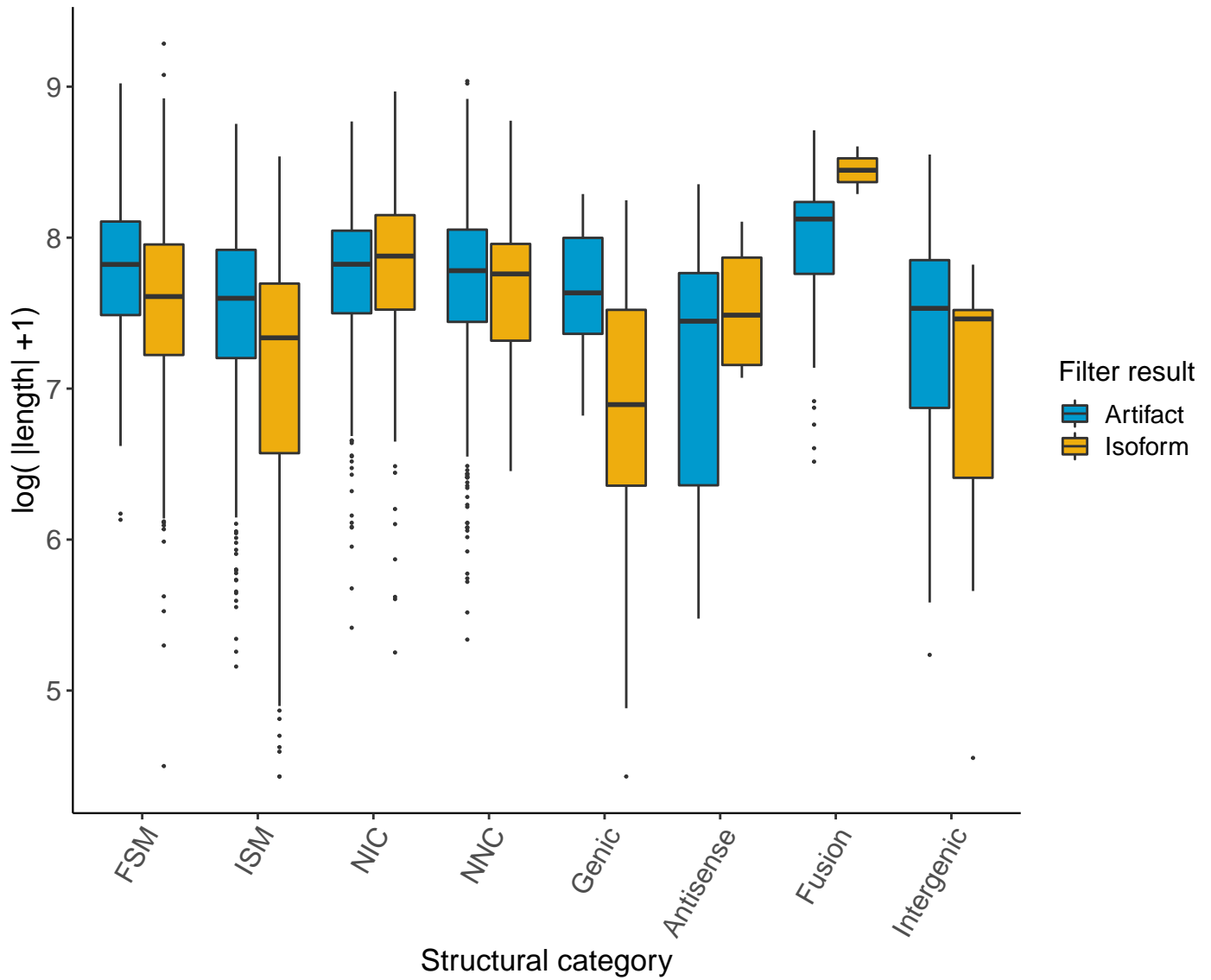
predicted\_NMD – ML importance: 5.71



dist\_to\_CAGE\_peak – ML importance: 5.28

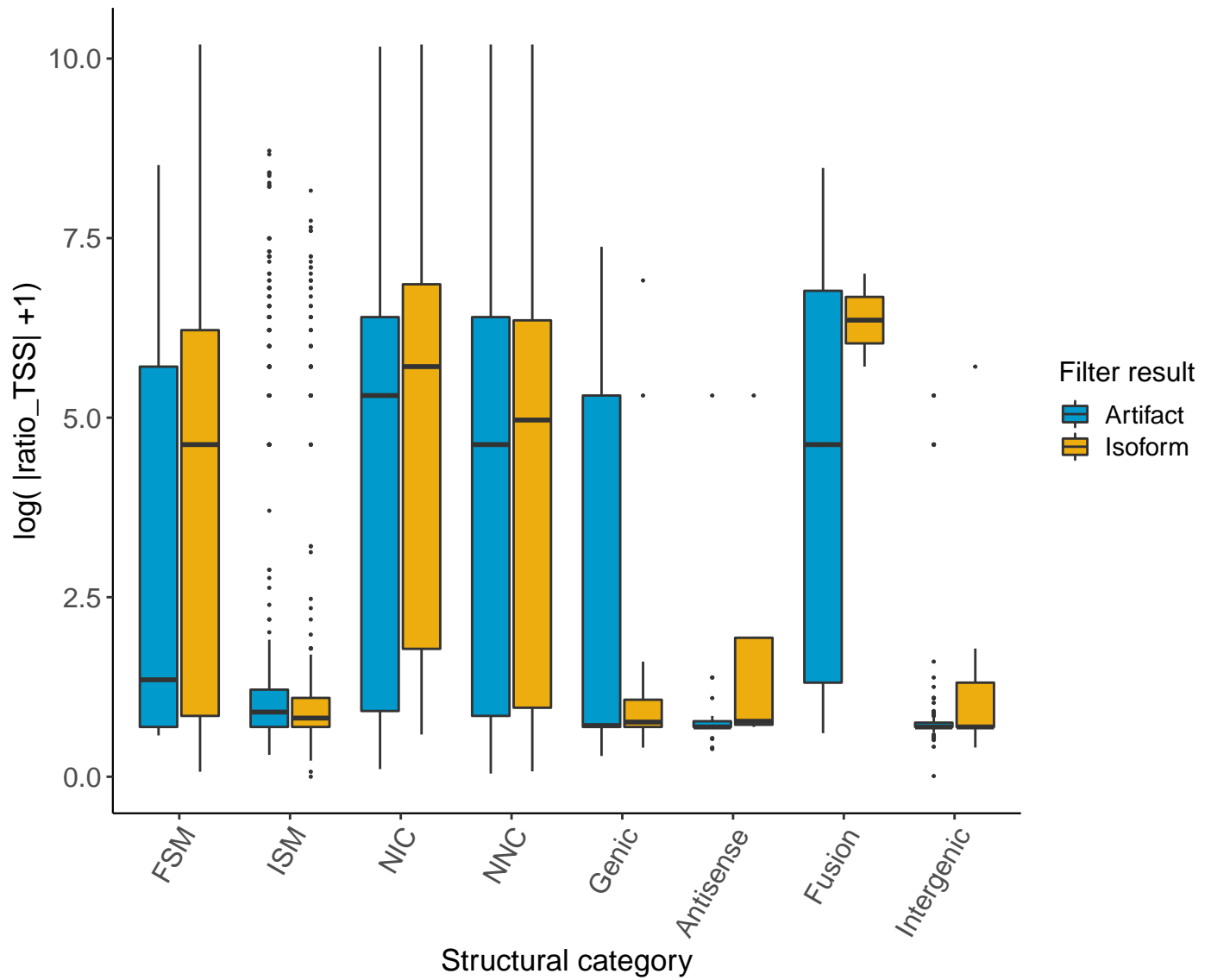


length – ML importance: 5.15

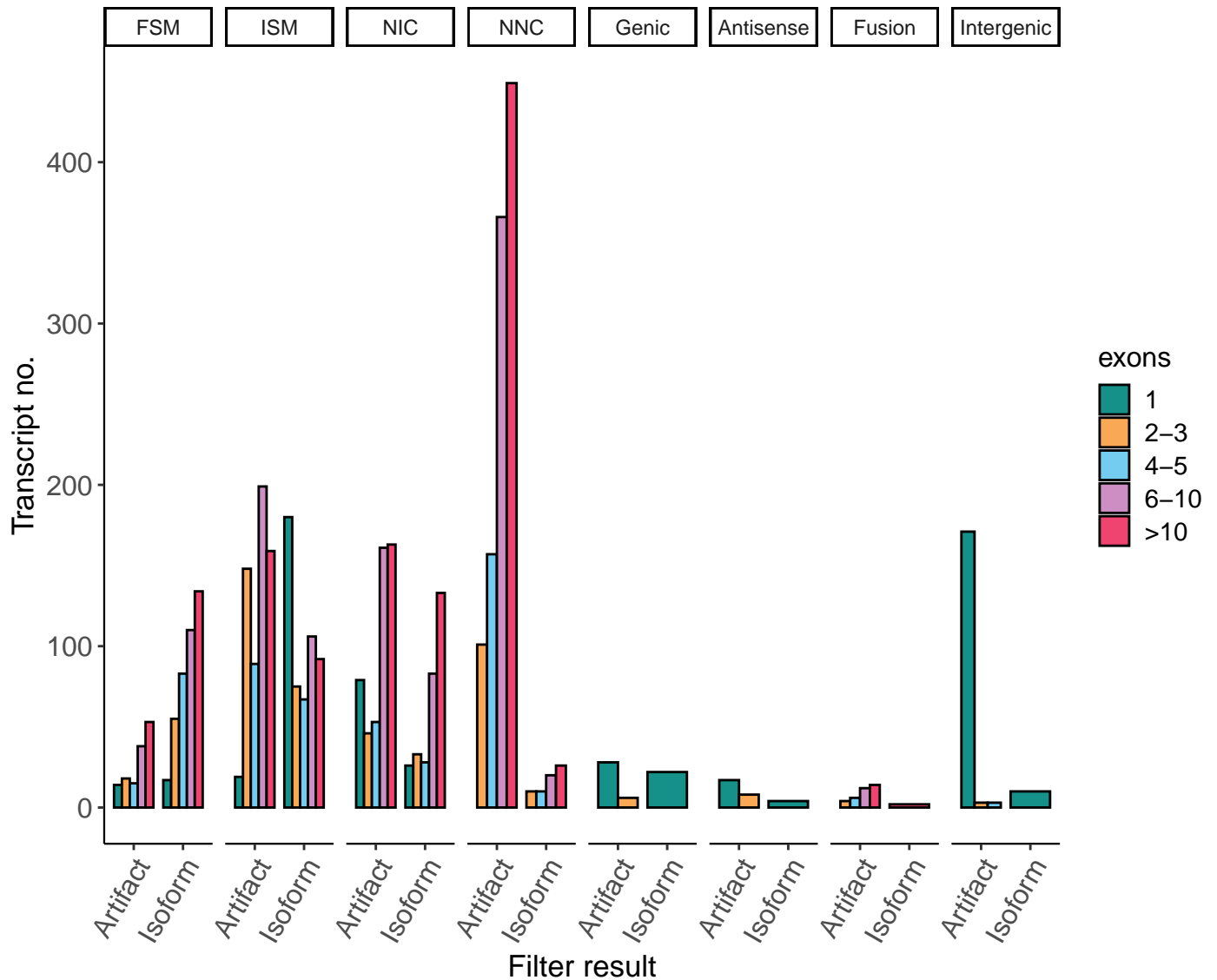




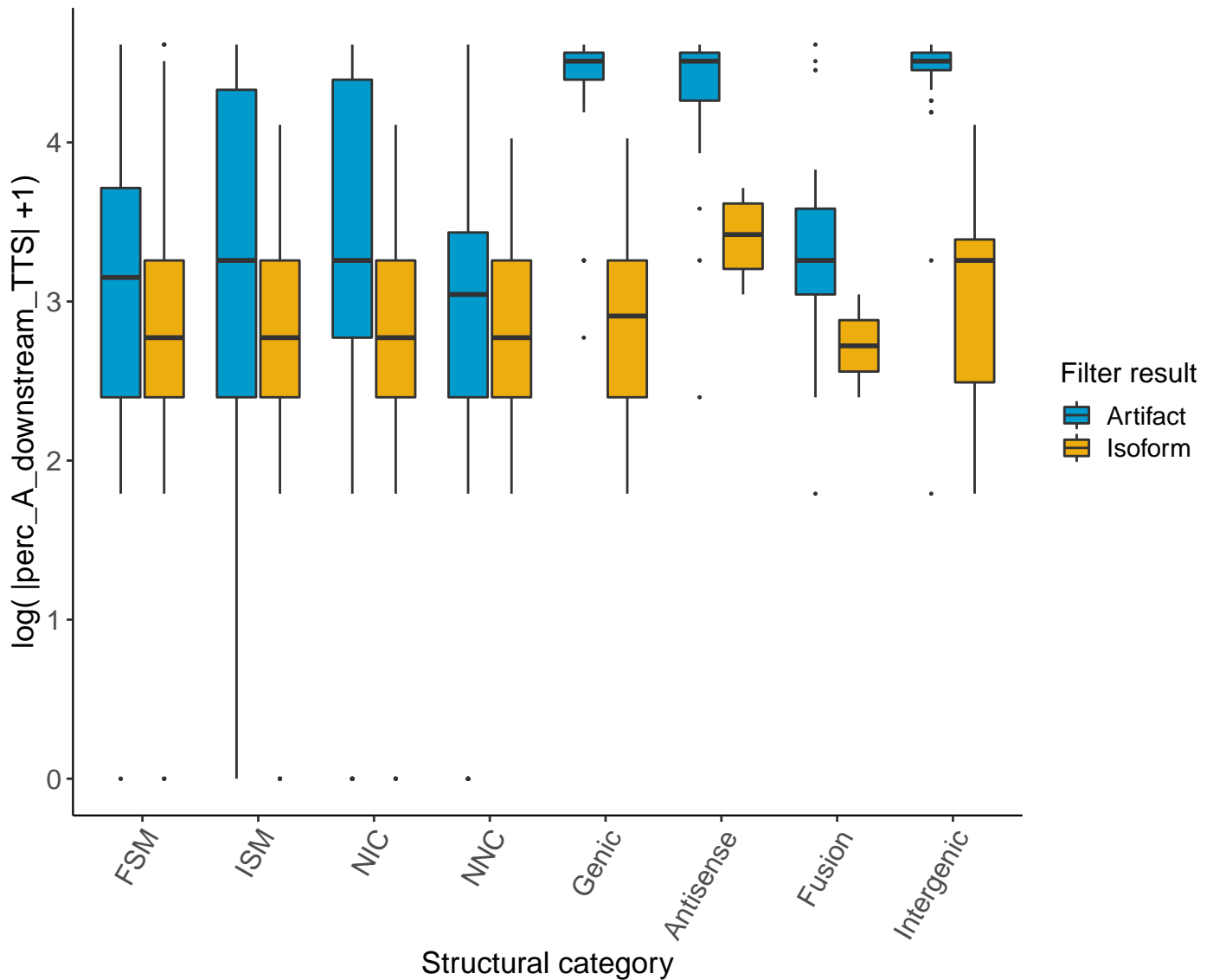
ratio\_TSS – ML importance: 4.64



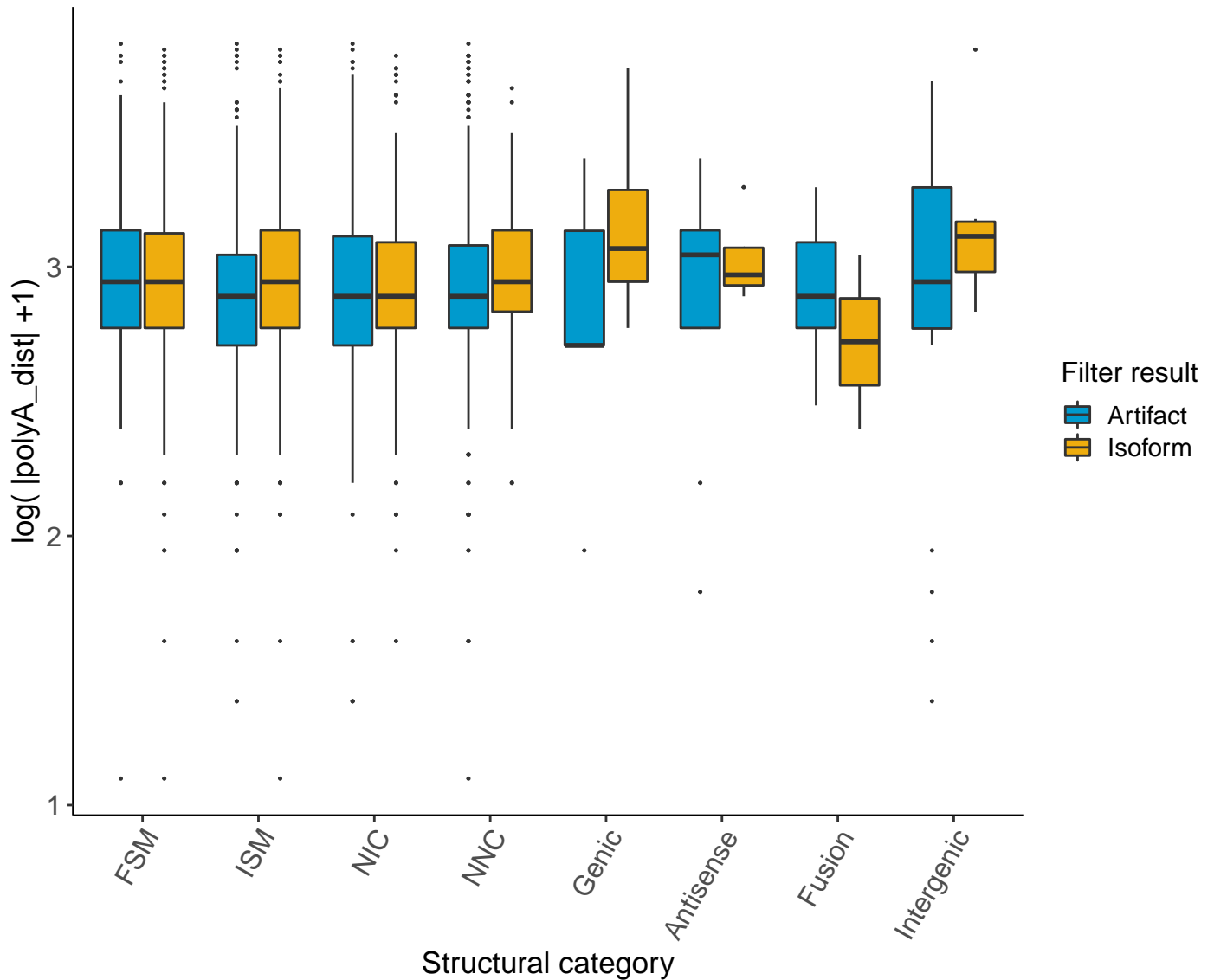
# exons – ML importance: 4.13



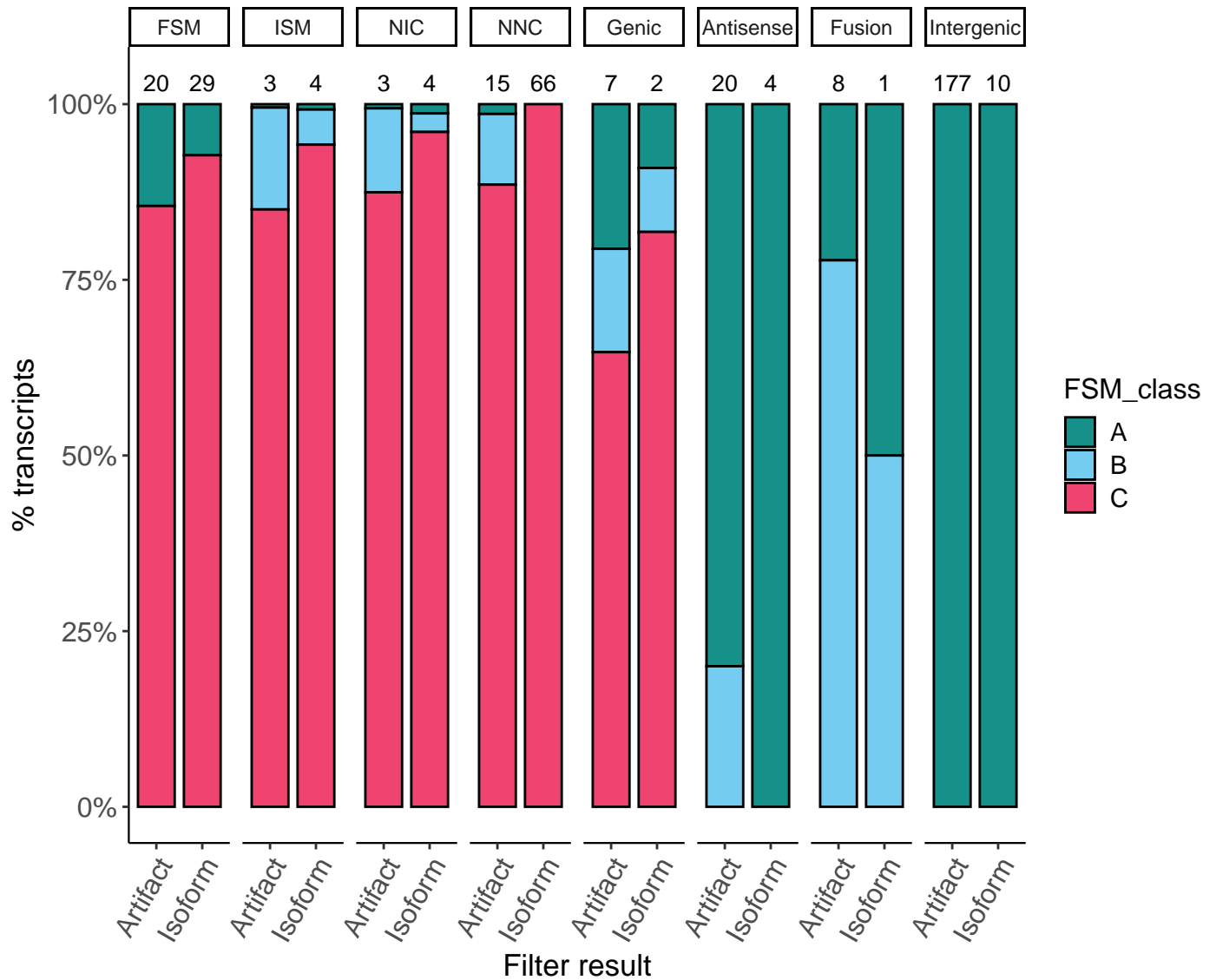
perc\_A\_downstream\_TTS – ML importance: 3.9



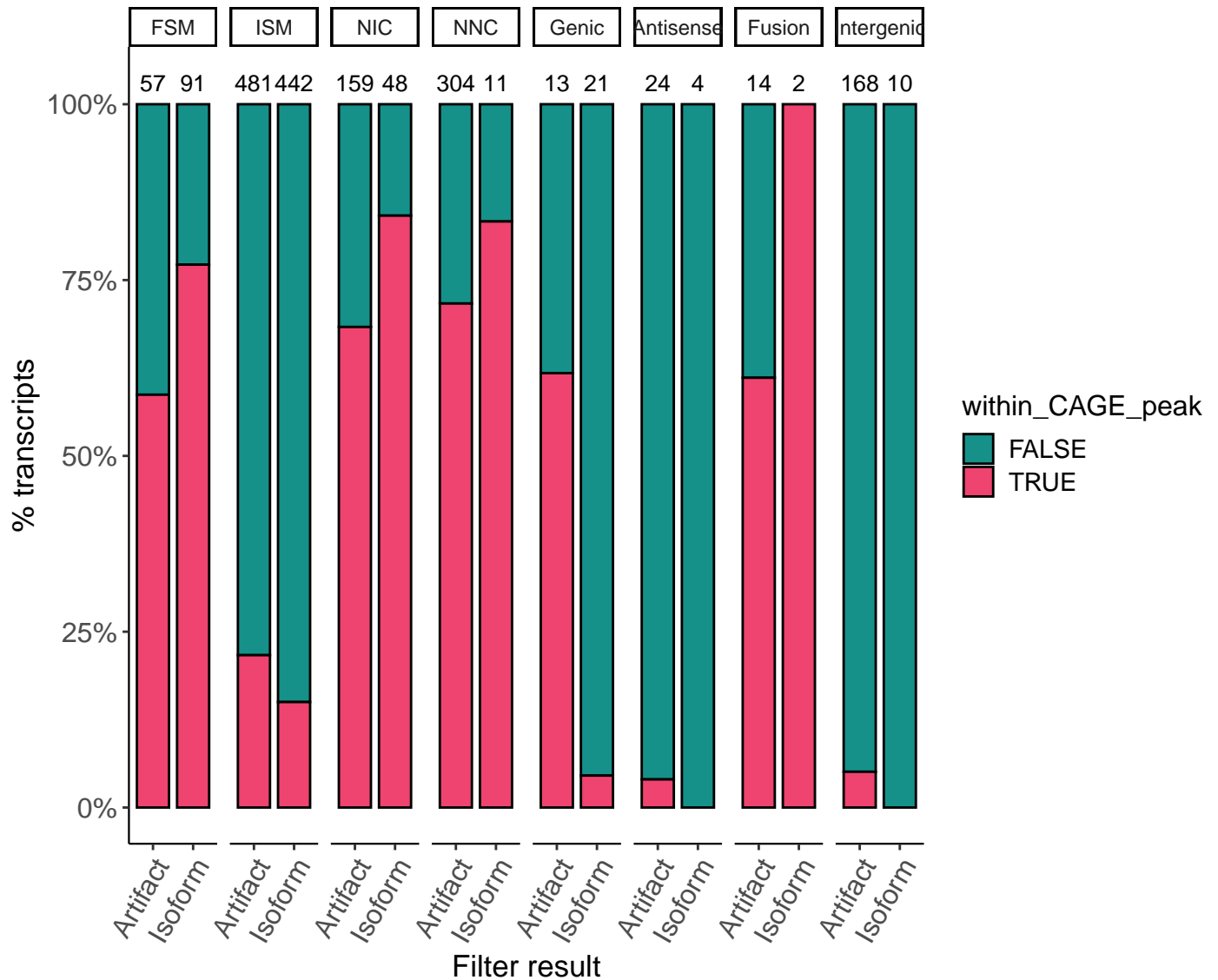
polyA\_dist – ML importance: 3.41



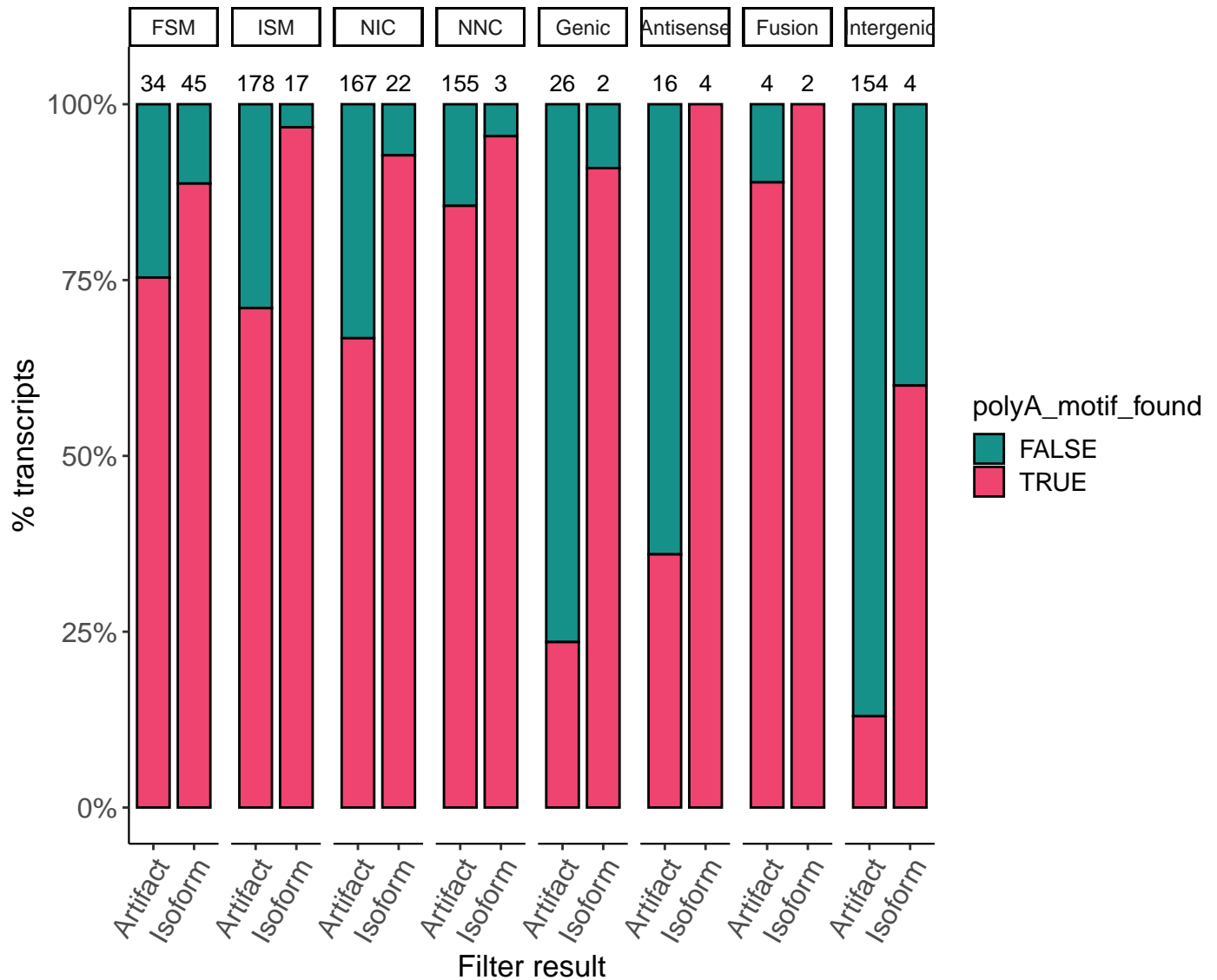
# FSM\_class – ML importance: 2.99



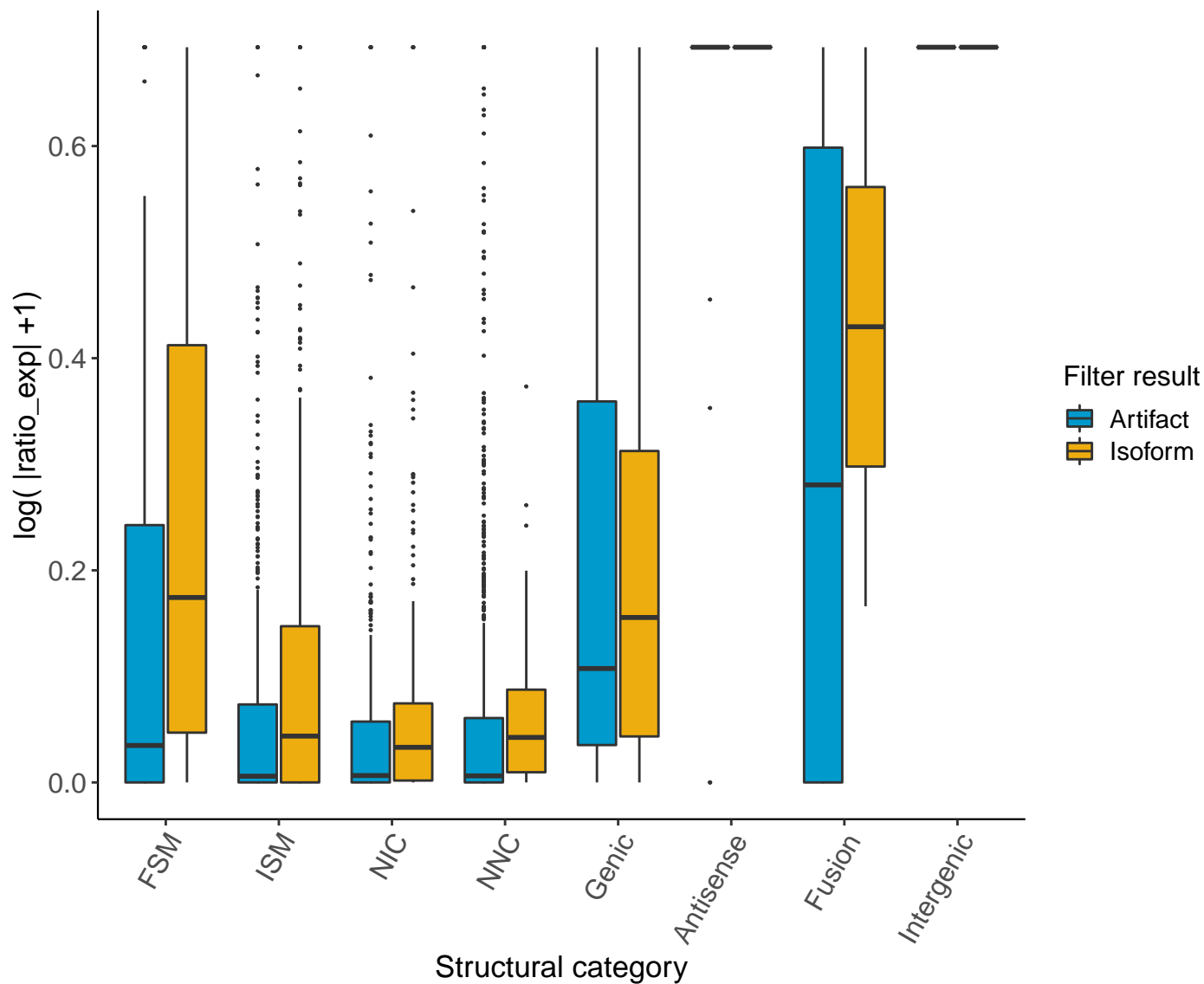
within\_CAGE\_peak – ML importance: 0.99



# polyA\_motif\_found – ML importance: 0.52

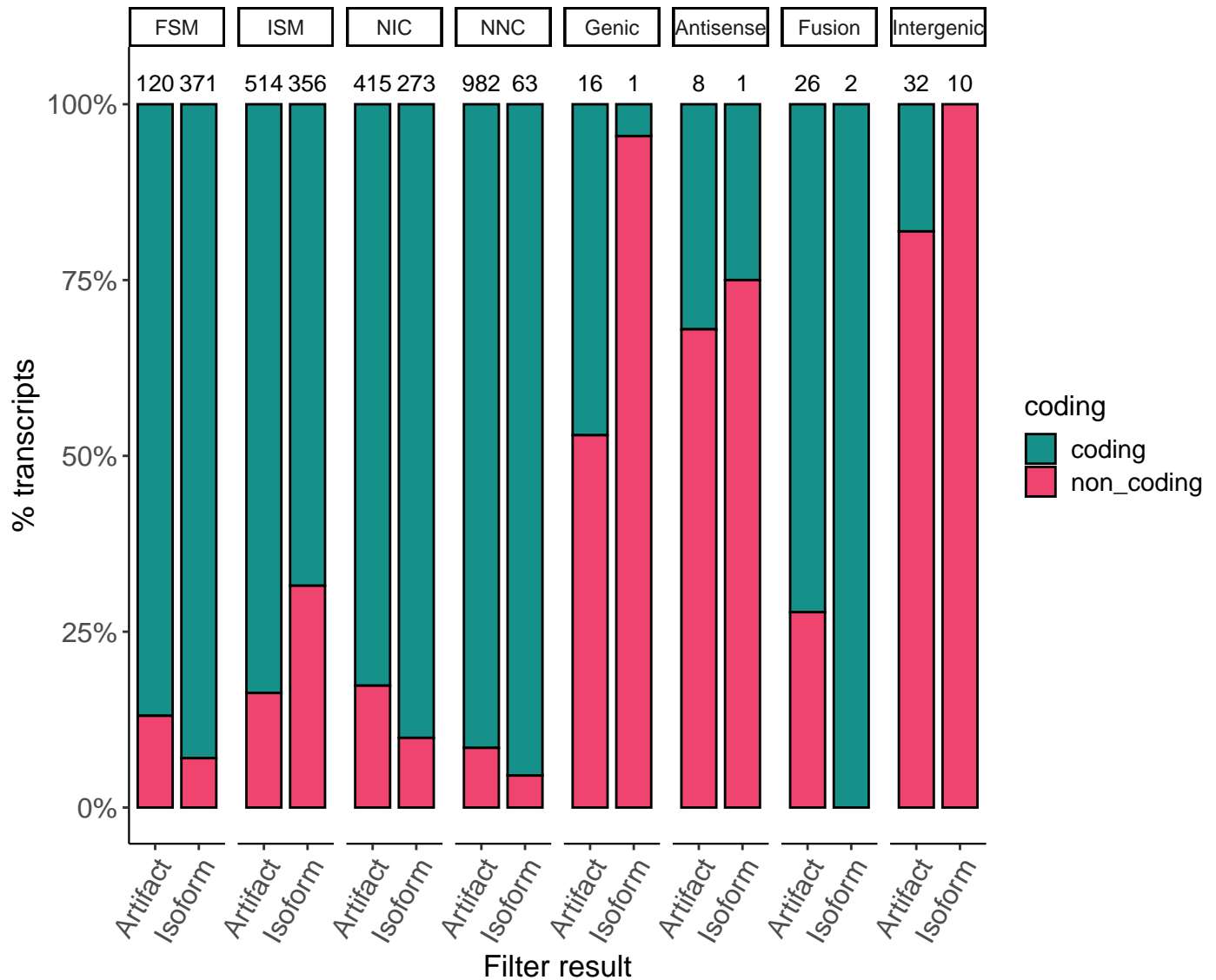


ratio\_exp – ML importance: 0.4



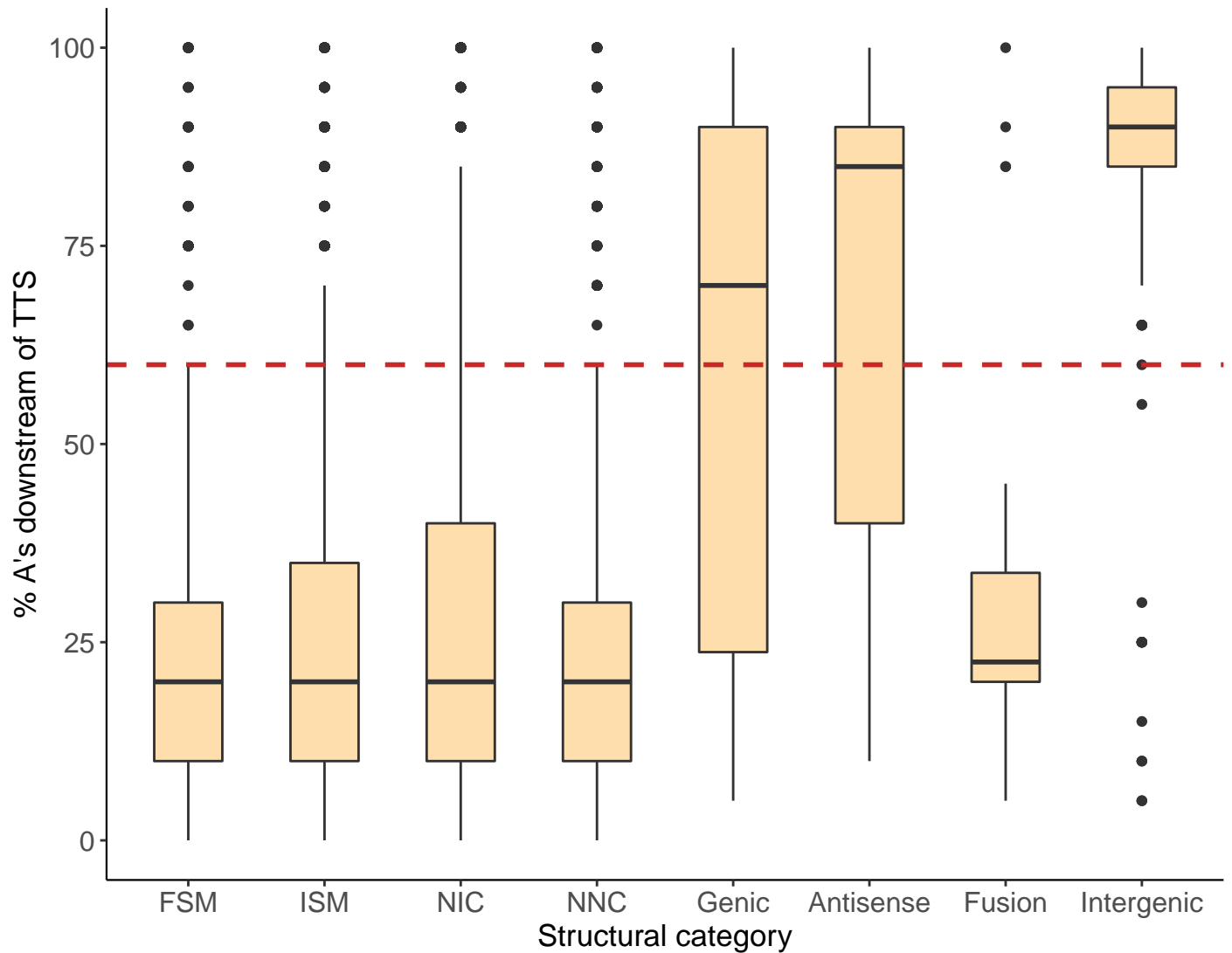


coding – ML importance: 0.31

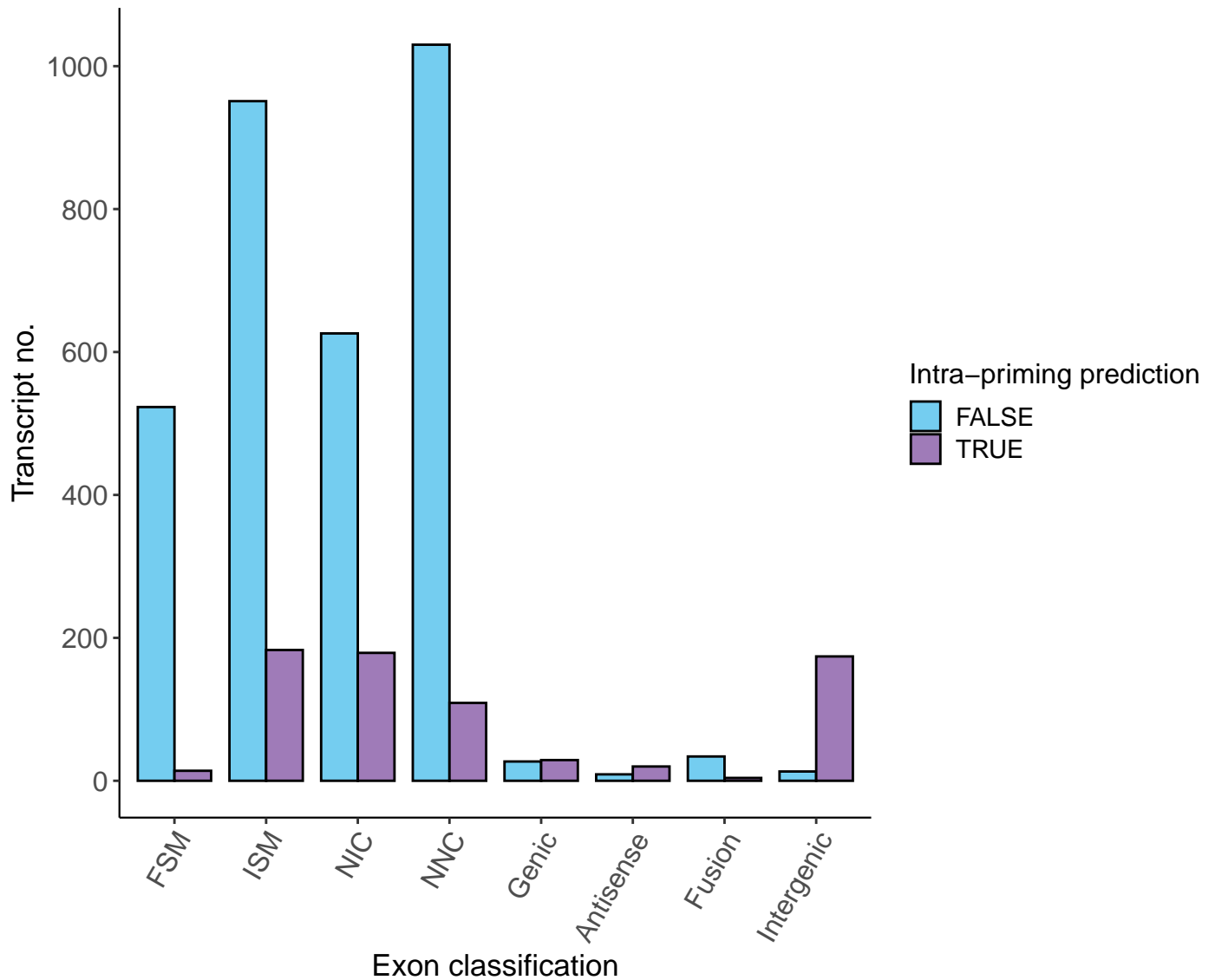


## A % by category

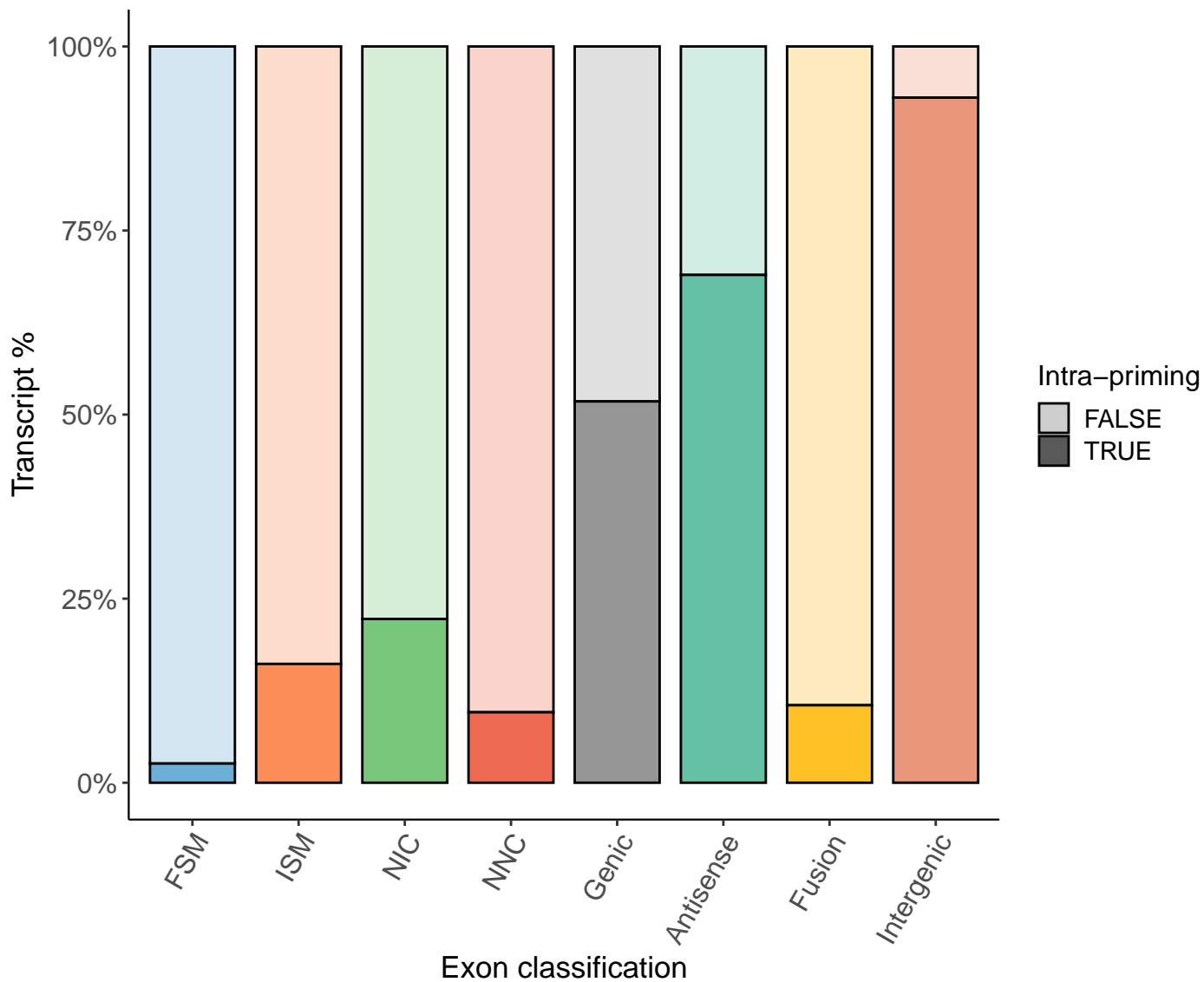
Red line indicates threshold employed in ML filter



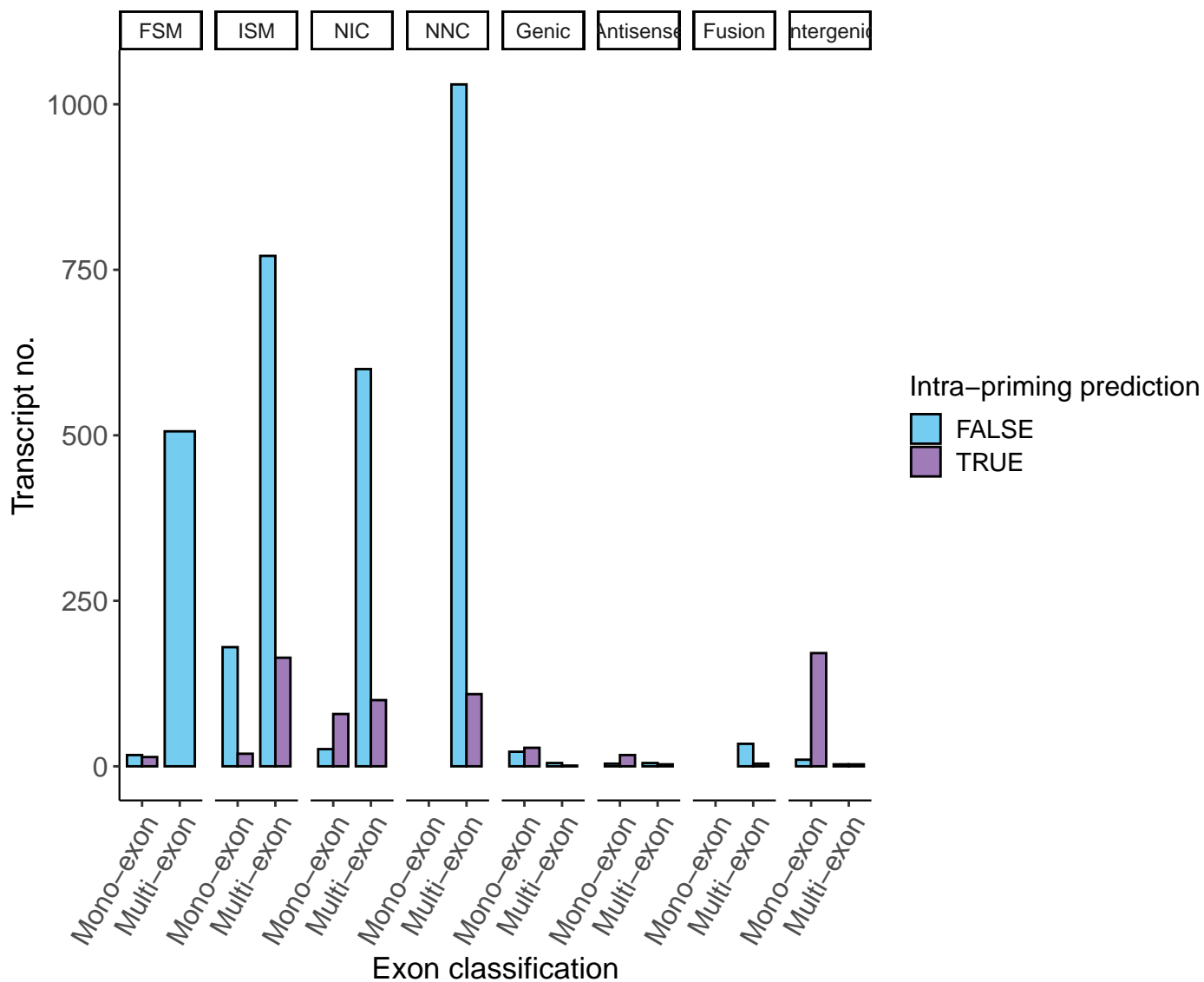
Isoforms flagged as intra-priming, by category



Isoforms flagged as intra-priming, by category (%)



Isoforms flagged as intra-priming, by exon number



Isoforms flagged as intra-priming, by exon number (%)

