

SQANTI3 filter report

Total Genes: 652

Total Transcripts: 3925

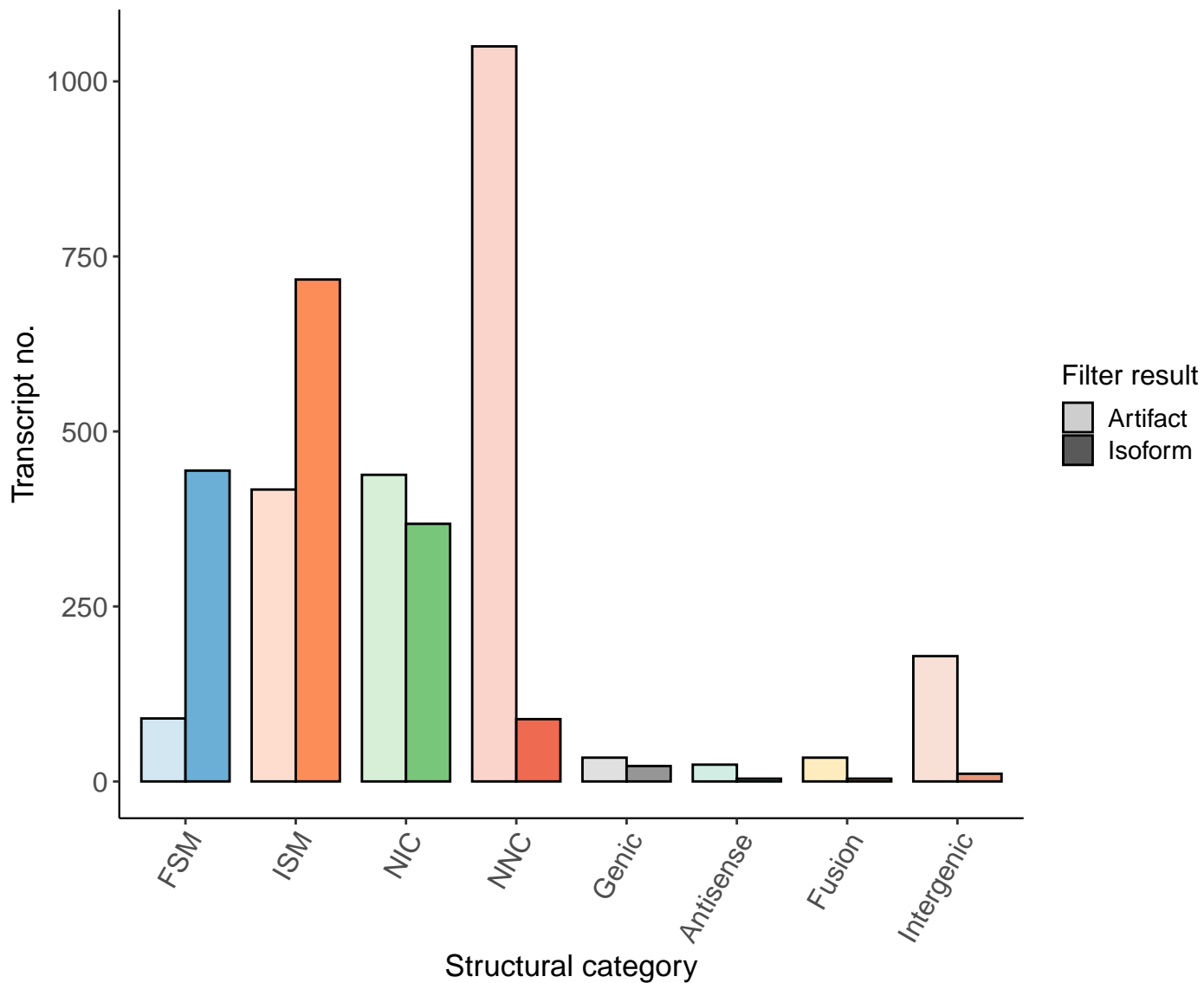
– Isoforms: 1659 (42%)

– Artifacts: 2266 (58%)

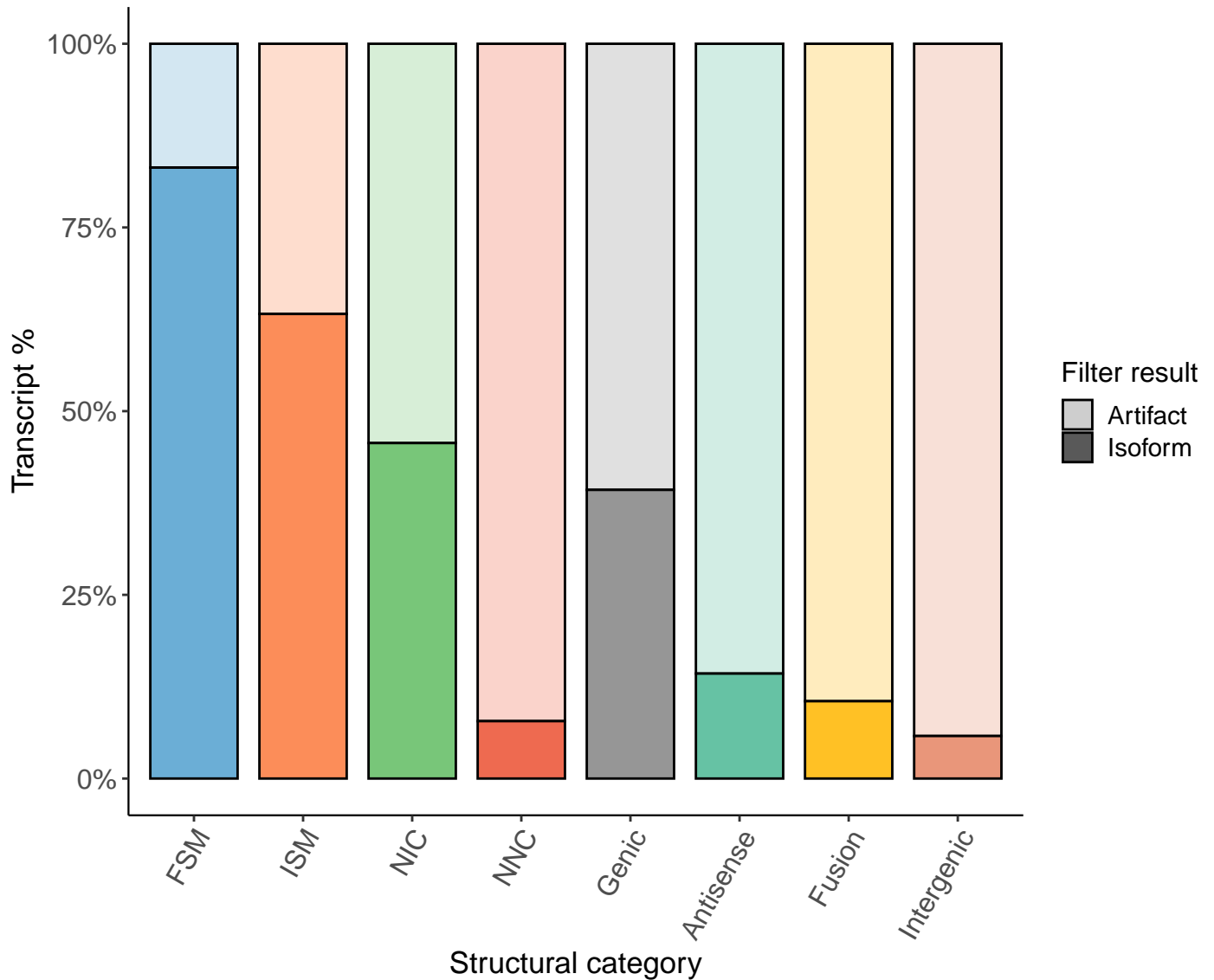
Gene category	Gene no.	No. of genes with artifacts only
Annotated	436	77
Novel	216	201

Structural category	Artifact no.	Isoform no.
FSM	90	444
ISM	417	717
NIC	438	368
NNC	1050	89
Genic	34	22
Antisense	24	4
Fusion	34	4
Intergenic	179	11

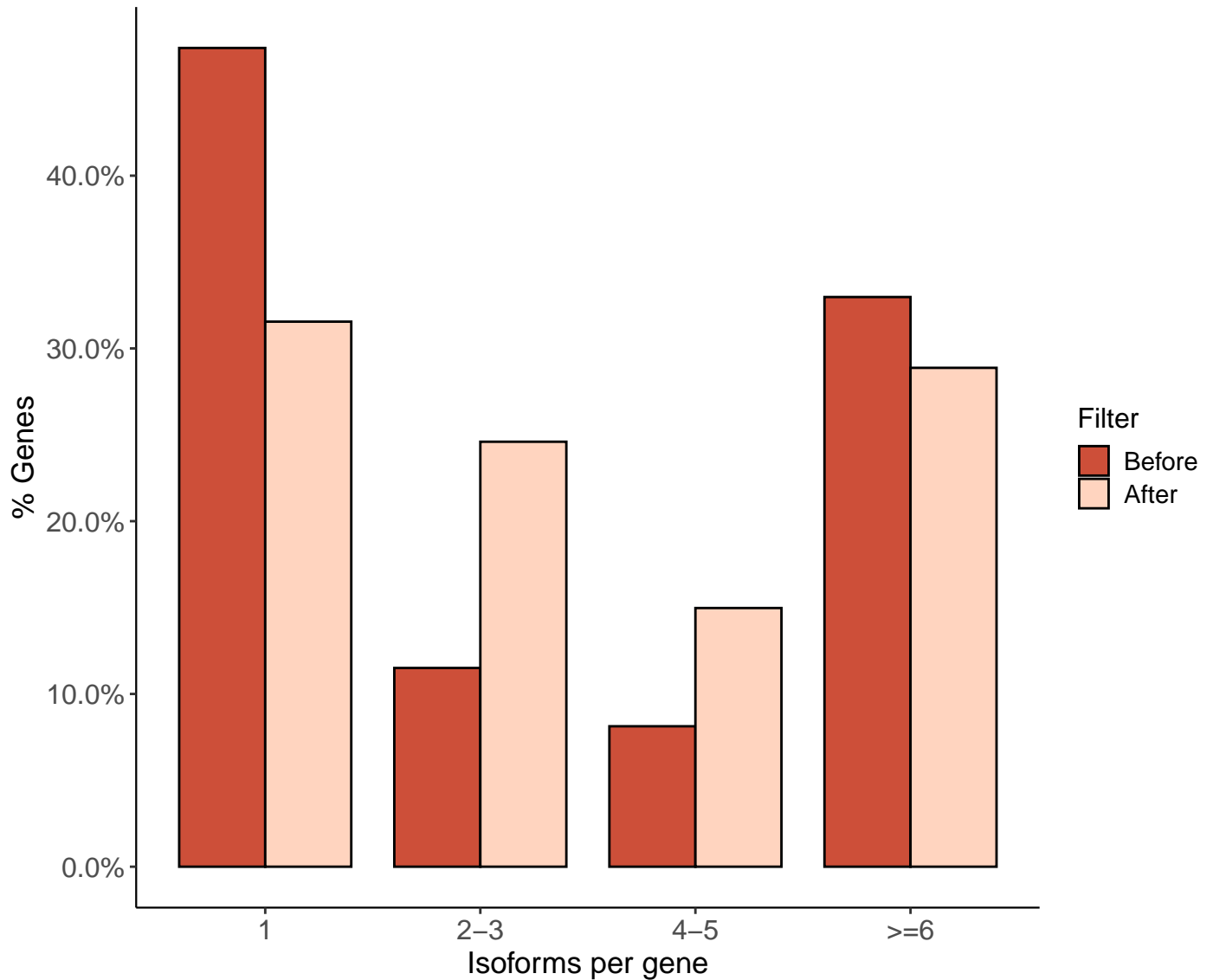
Total isoforms and artifacts by category



% isoforms and artifacts by category

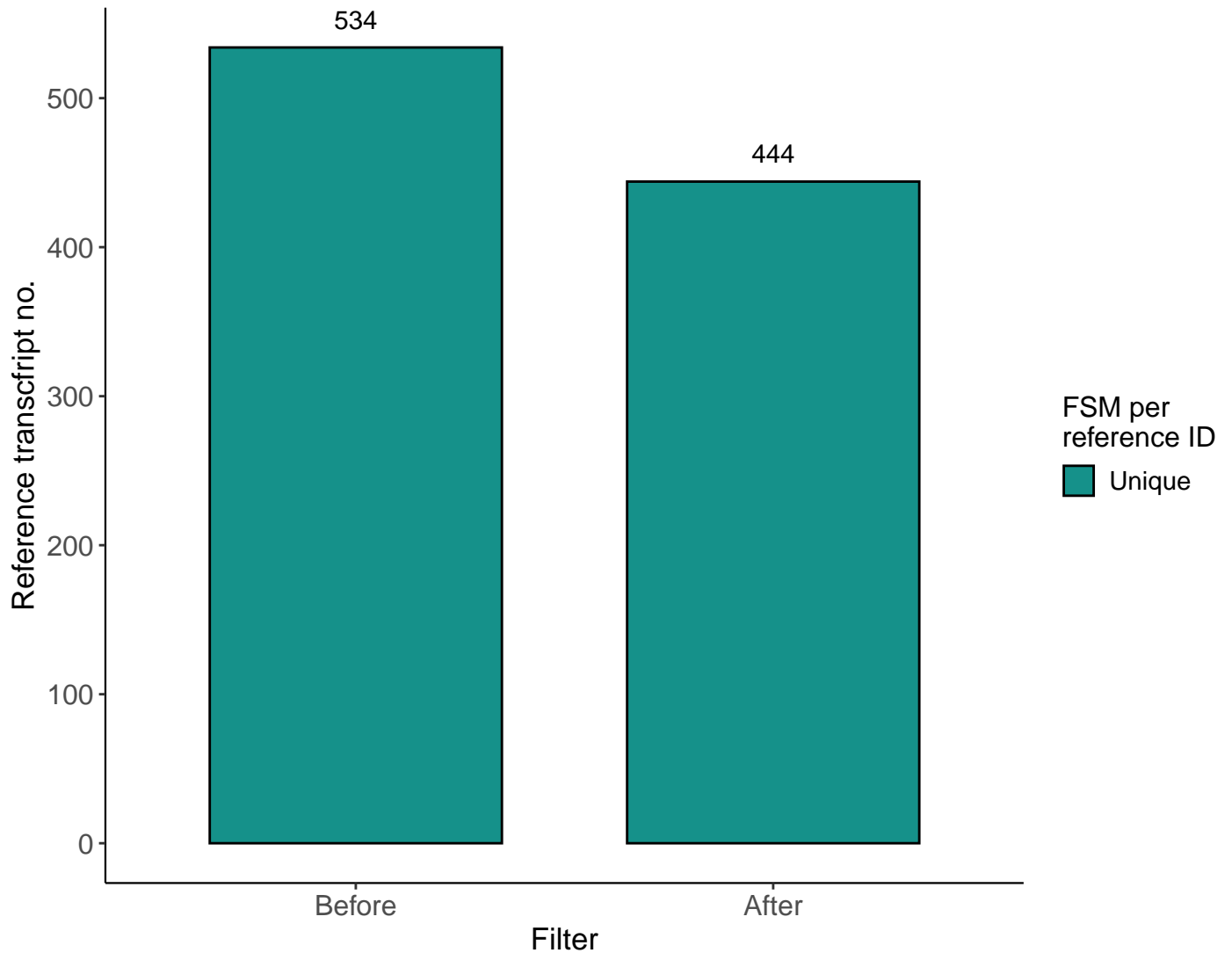


Number of isoforms per gene

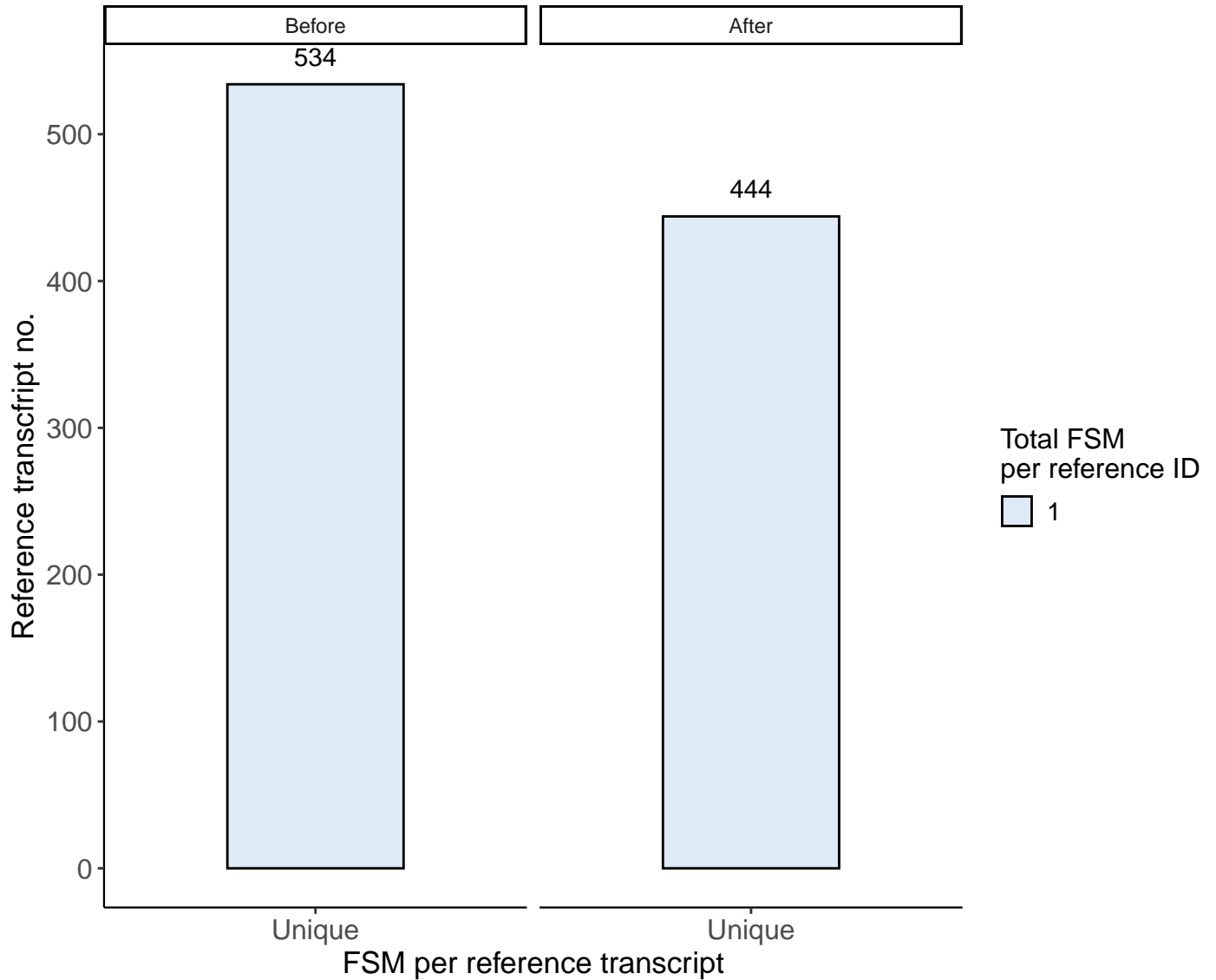


Reference transcript complexity

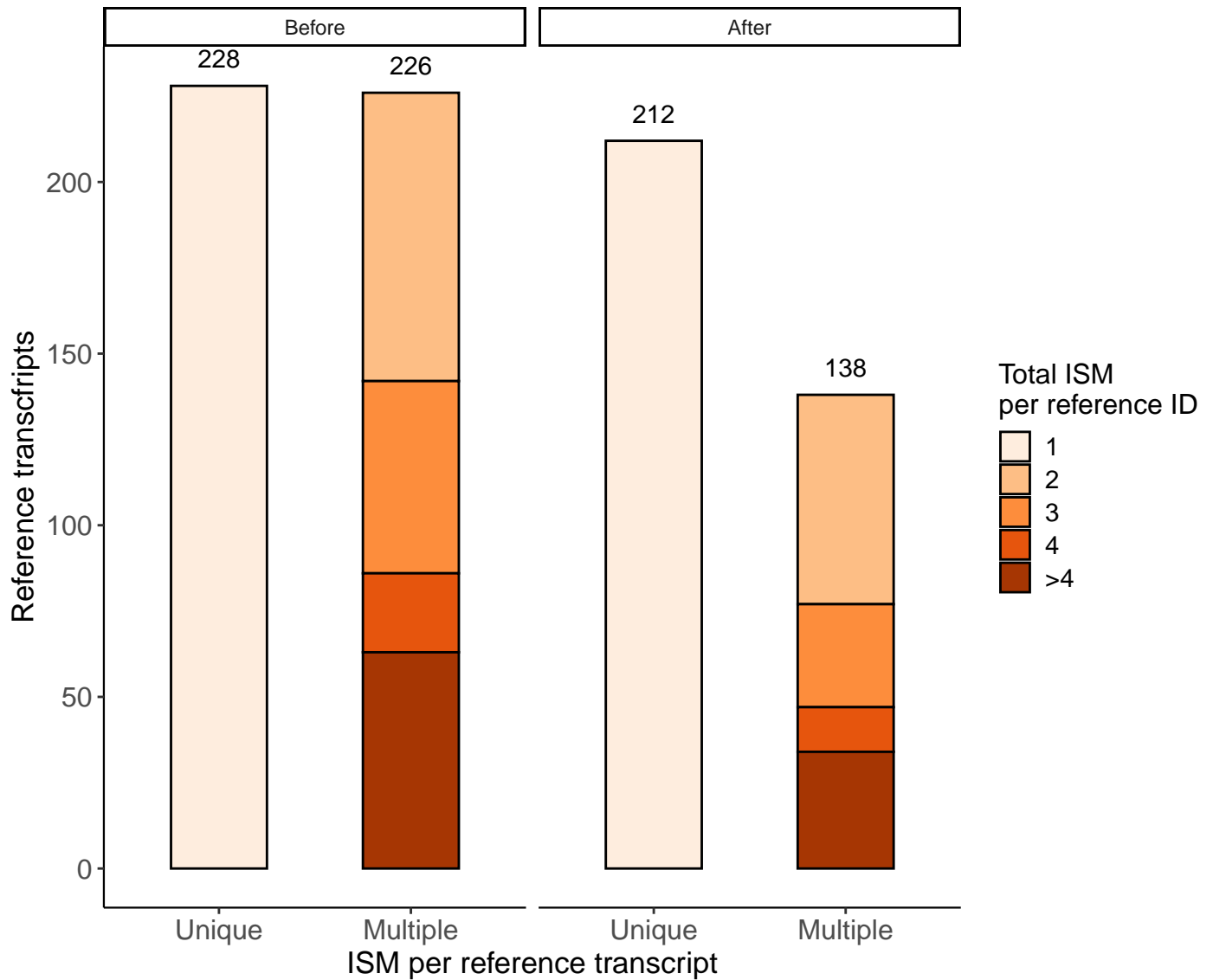
No. of reference transcripts represented by FSM



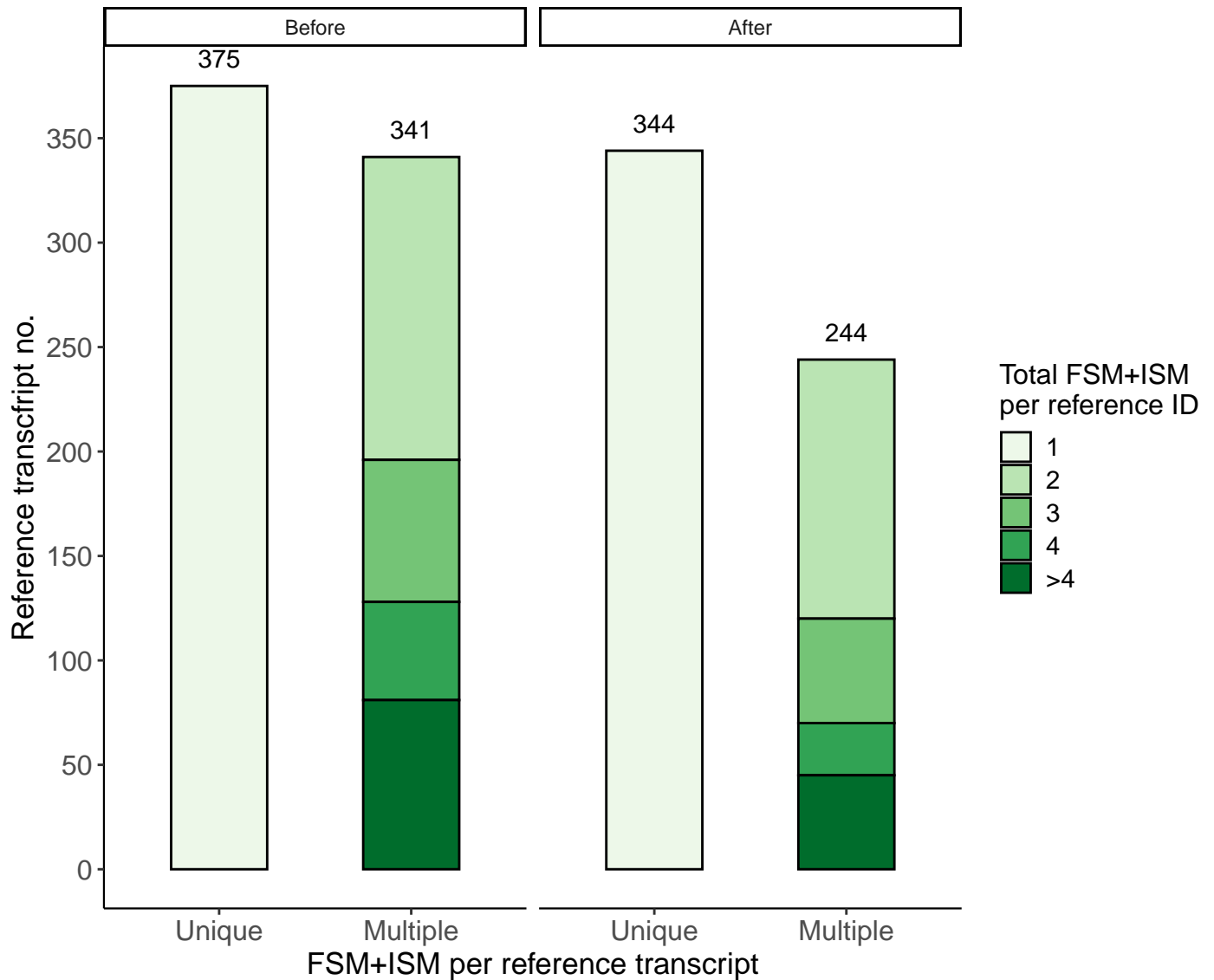
FSM redundancy



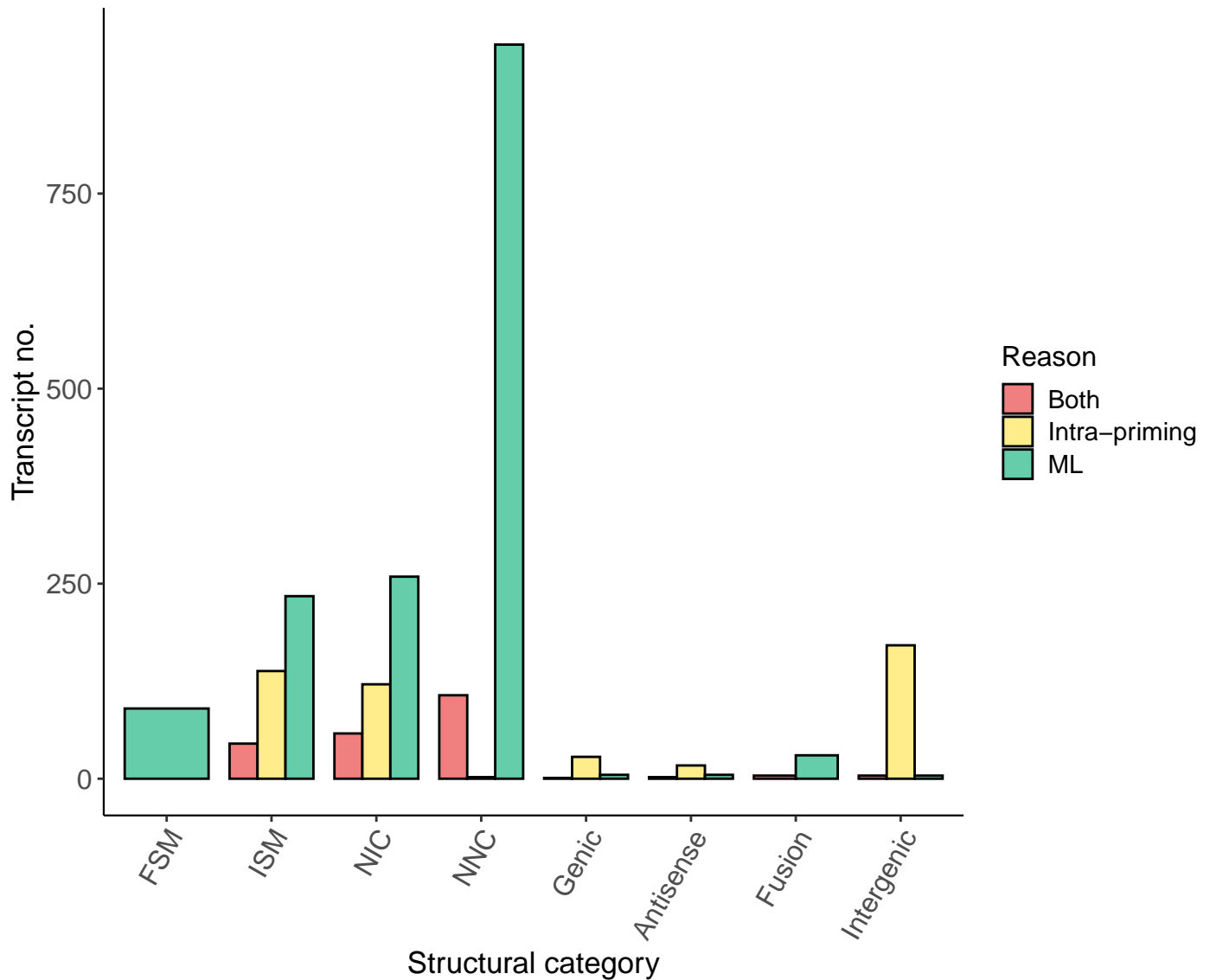
ISM redundancy



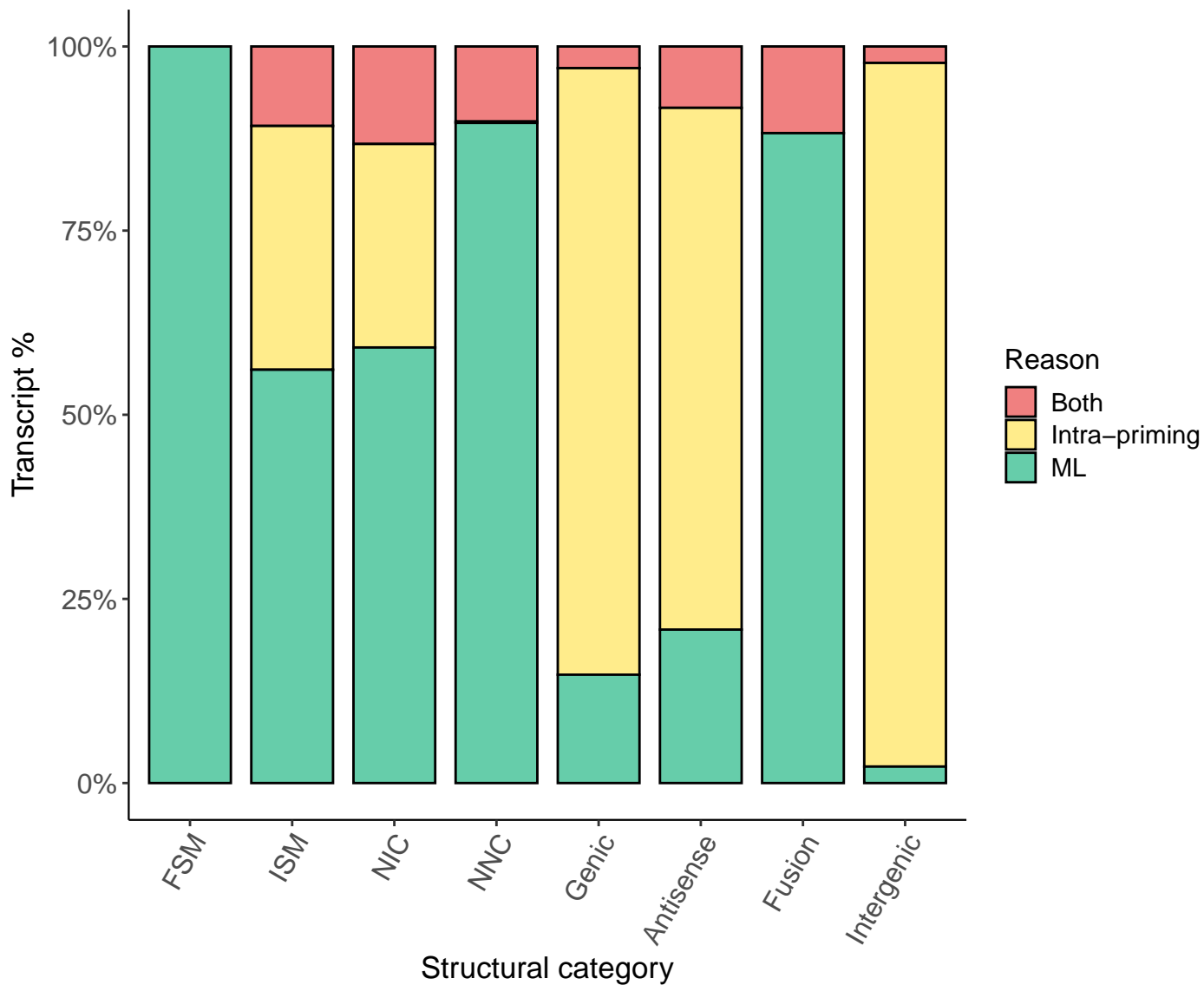
FSM+ISM redundancy



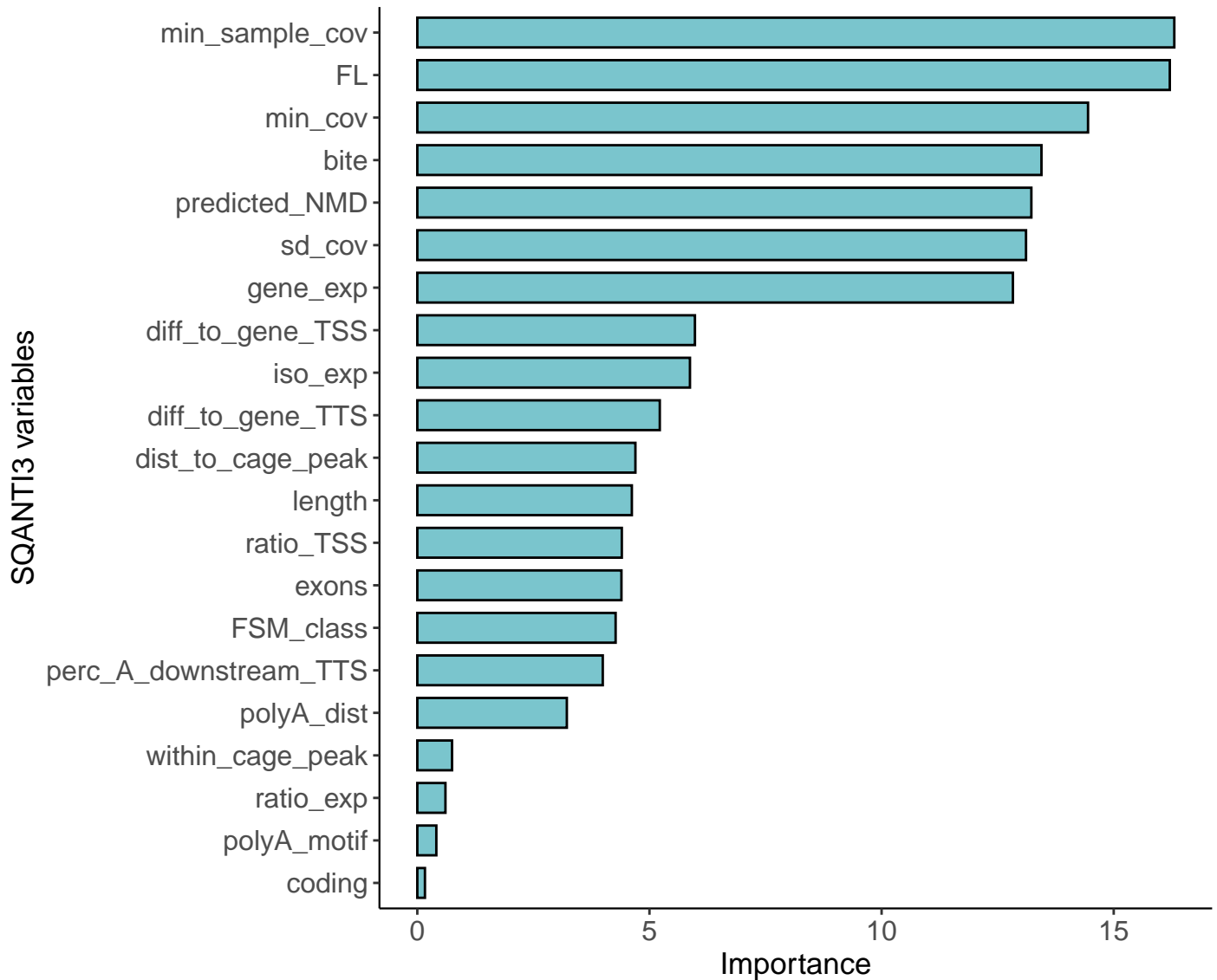
Reason to flag transcripts as artifacts, by category



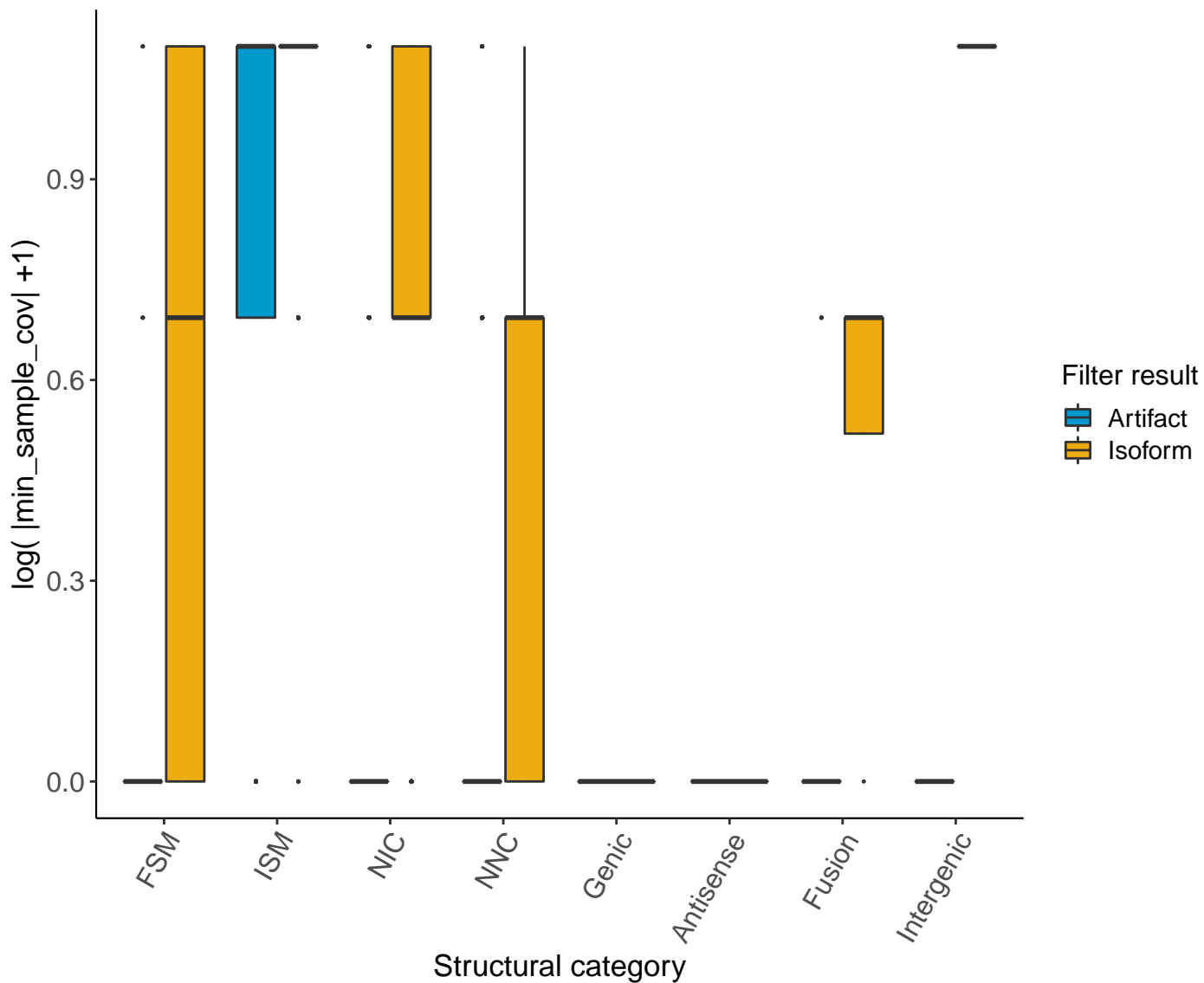
Reason to flag transcripts as artifacts, by category



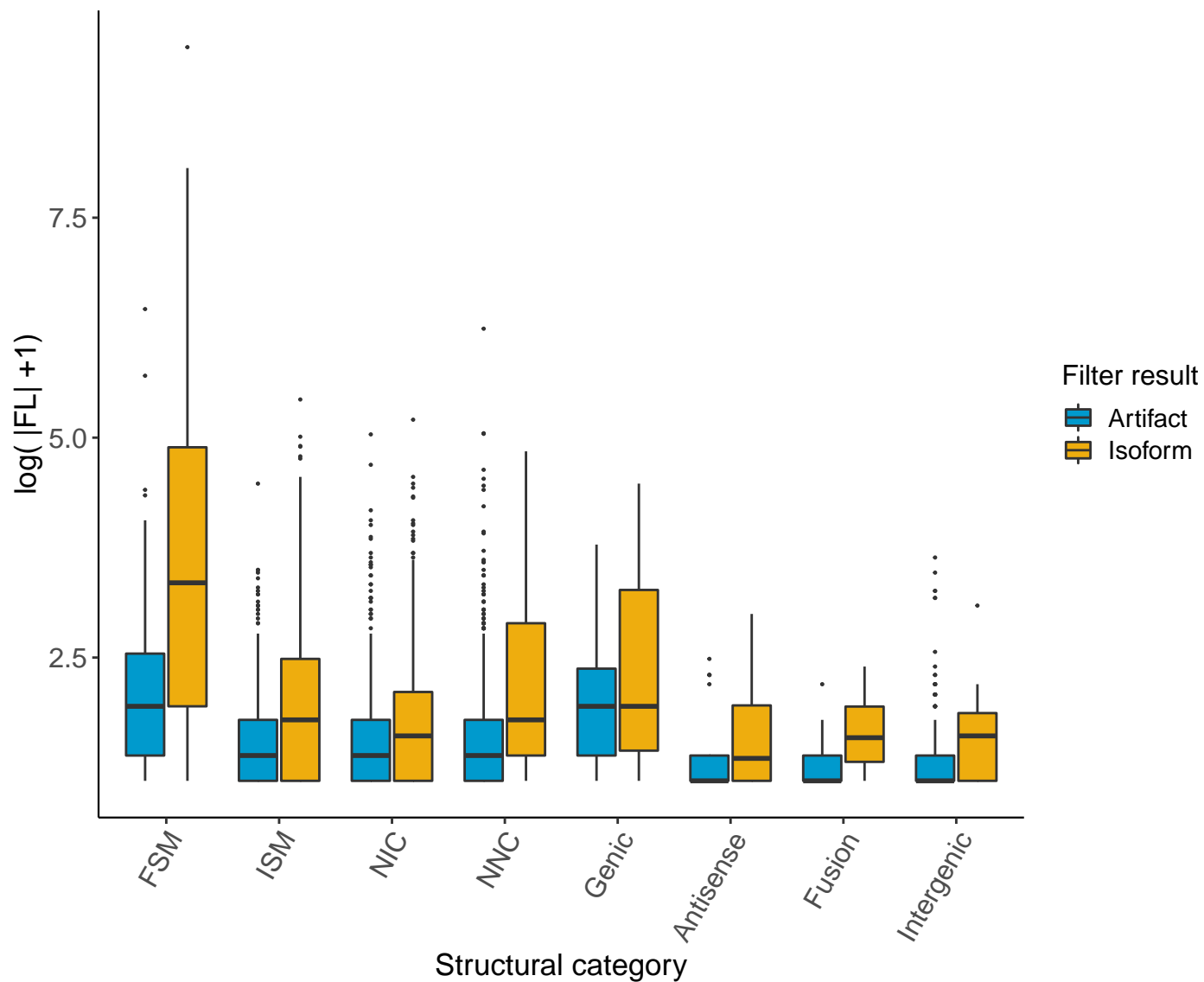
Variable importance in Random Forest classifier



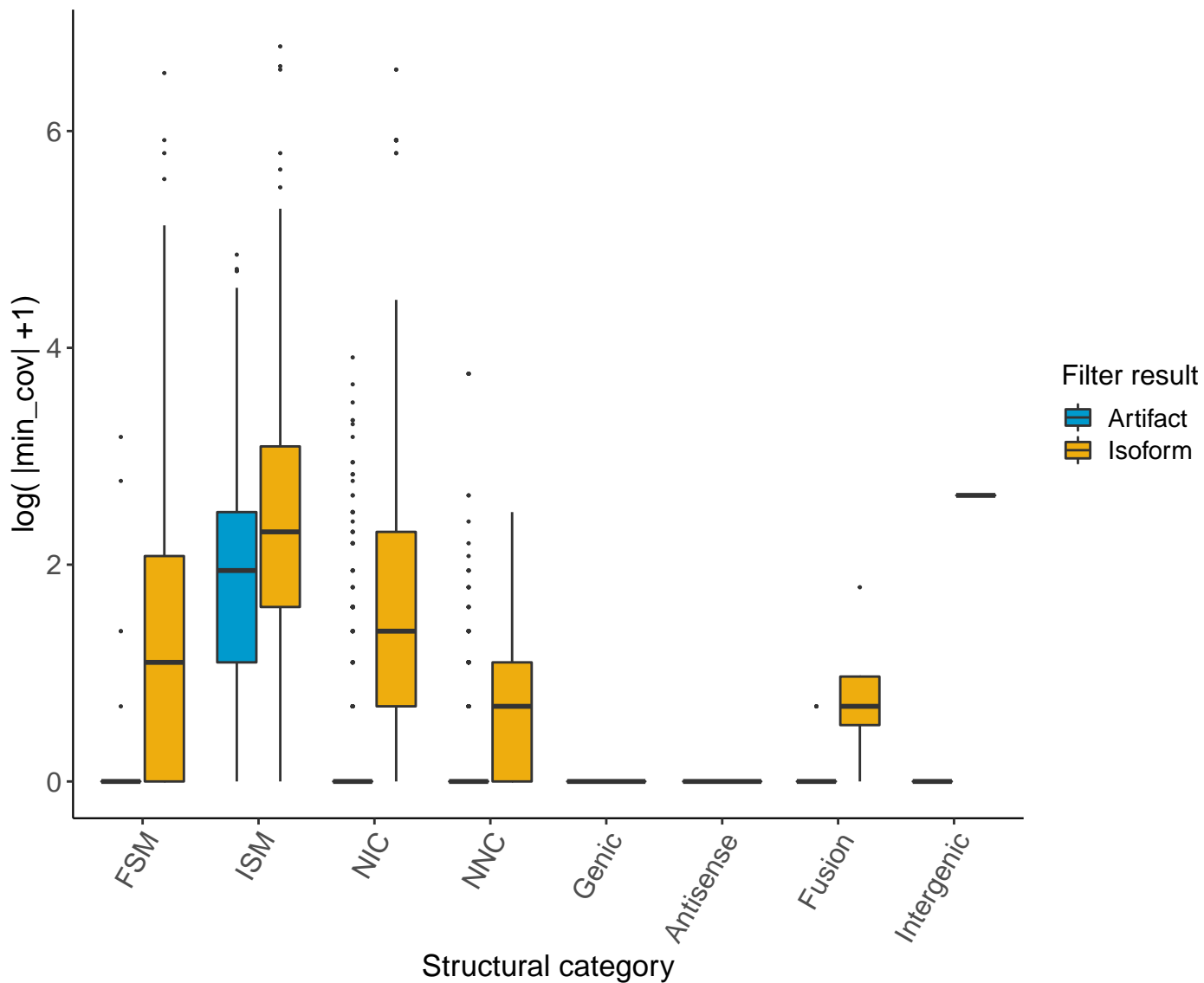
min_sample_cov – ML importance: 16.3



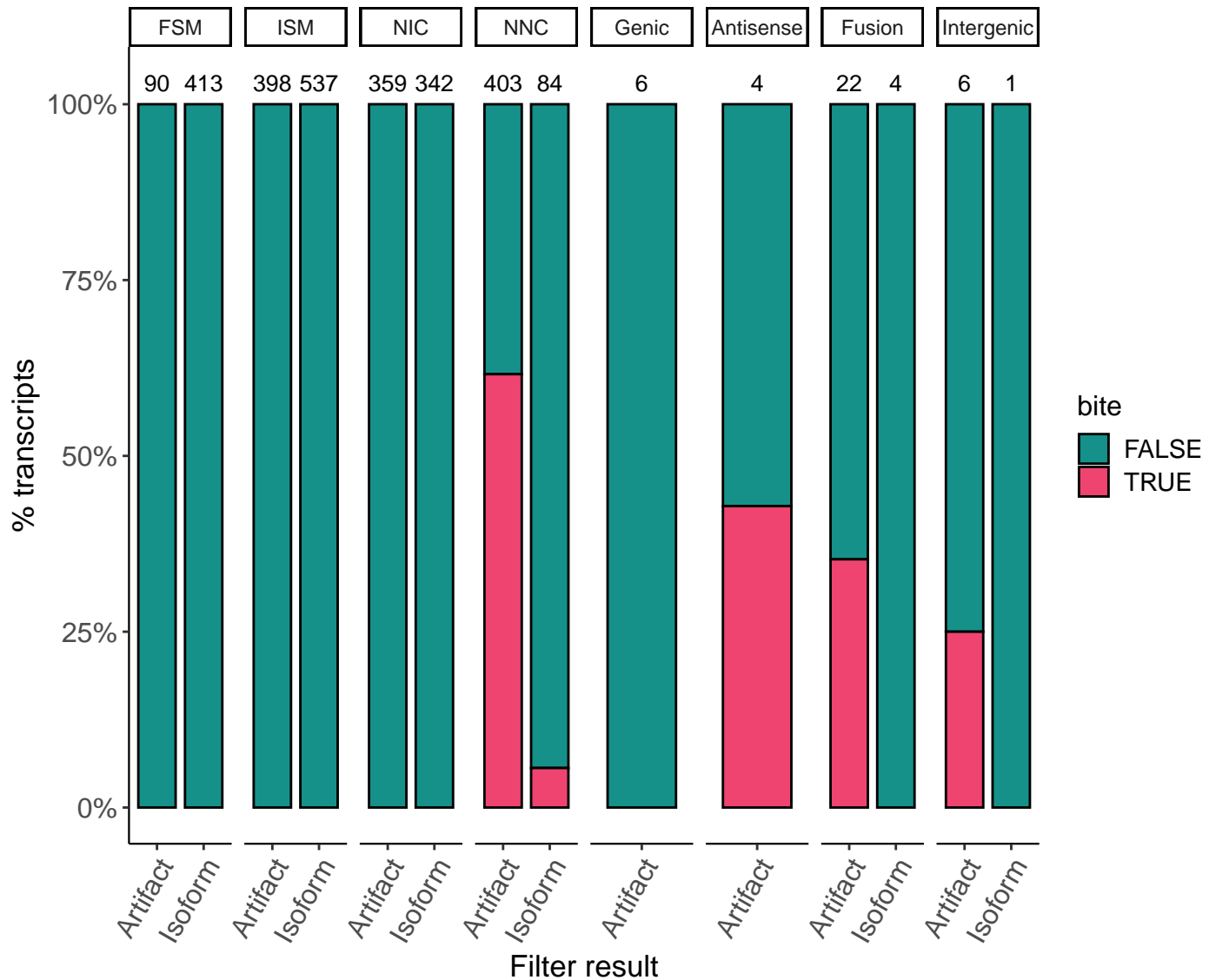
FL – ML importance: 16.21



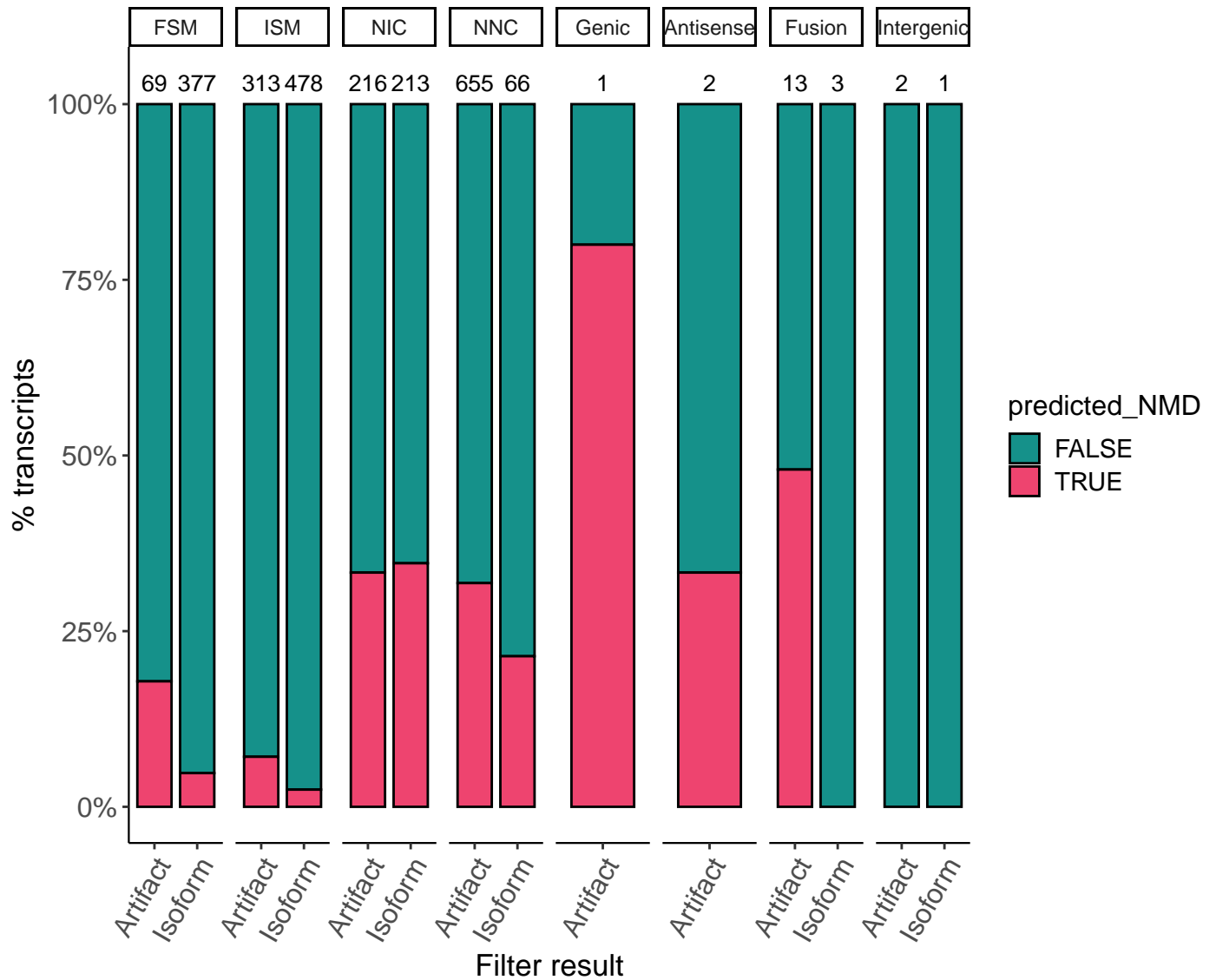
min_cov – ML importance: 14.45



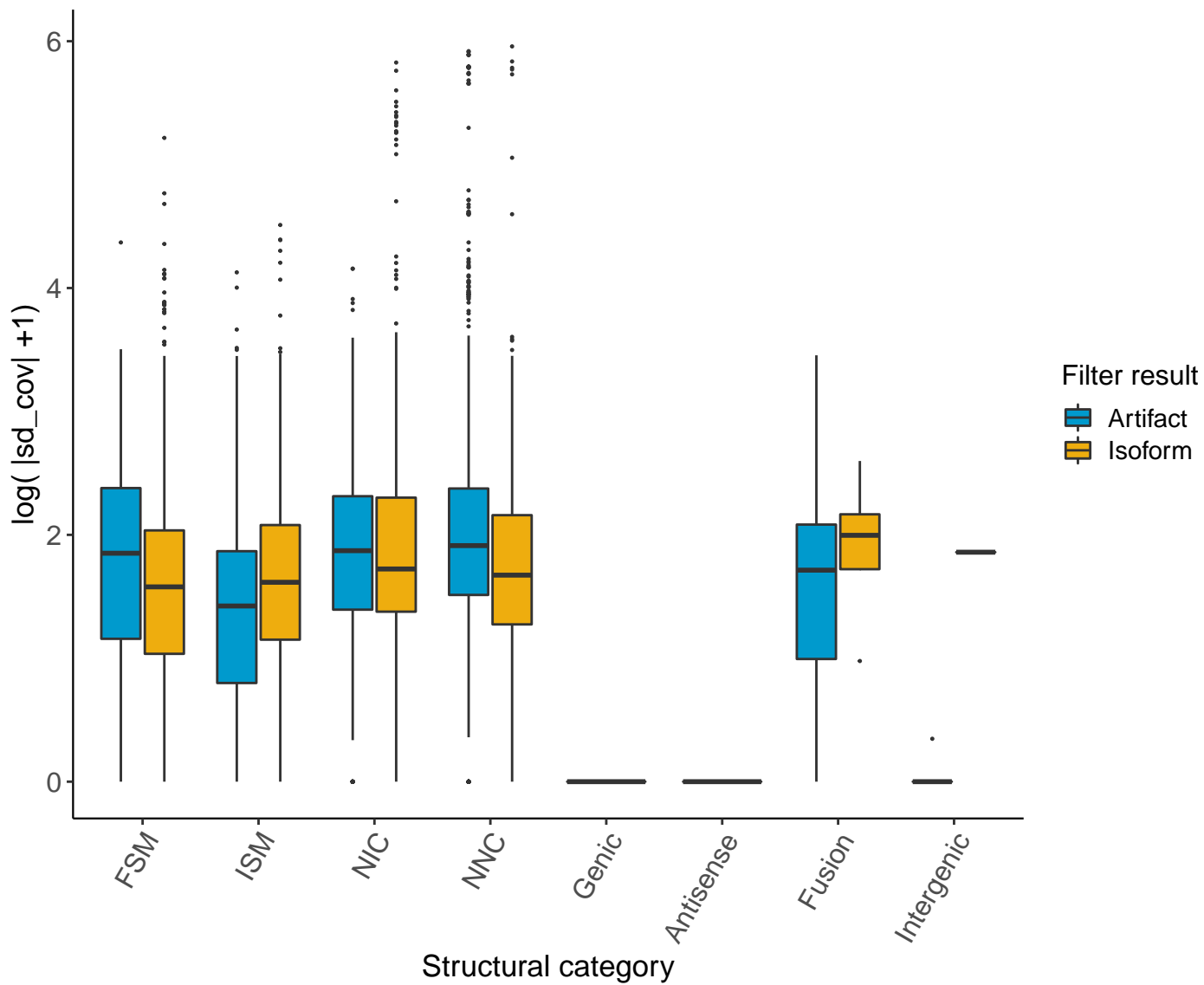
bite – ML importance: 13.44



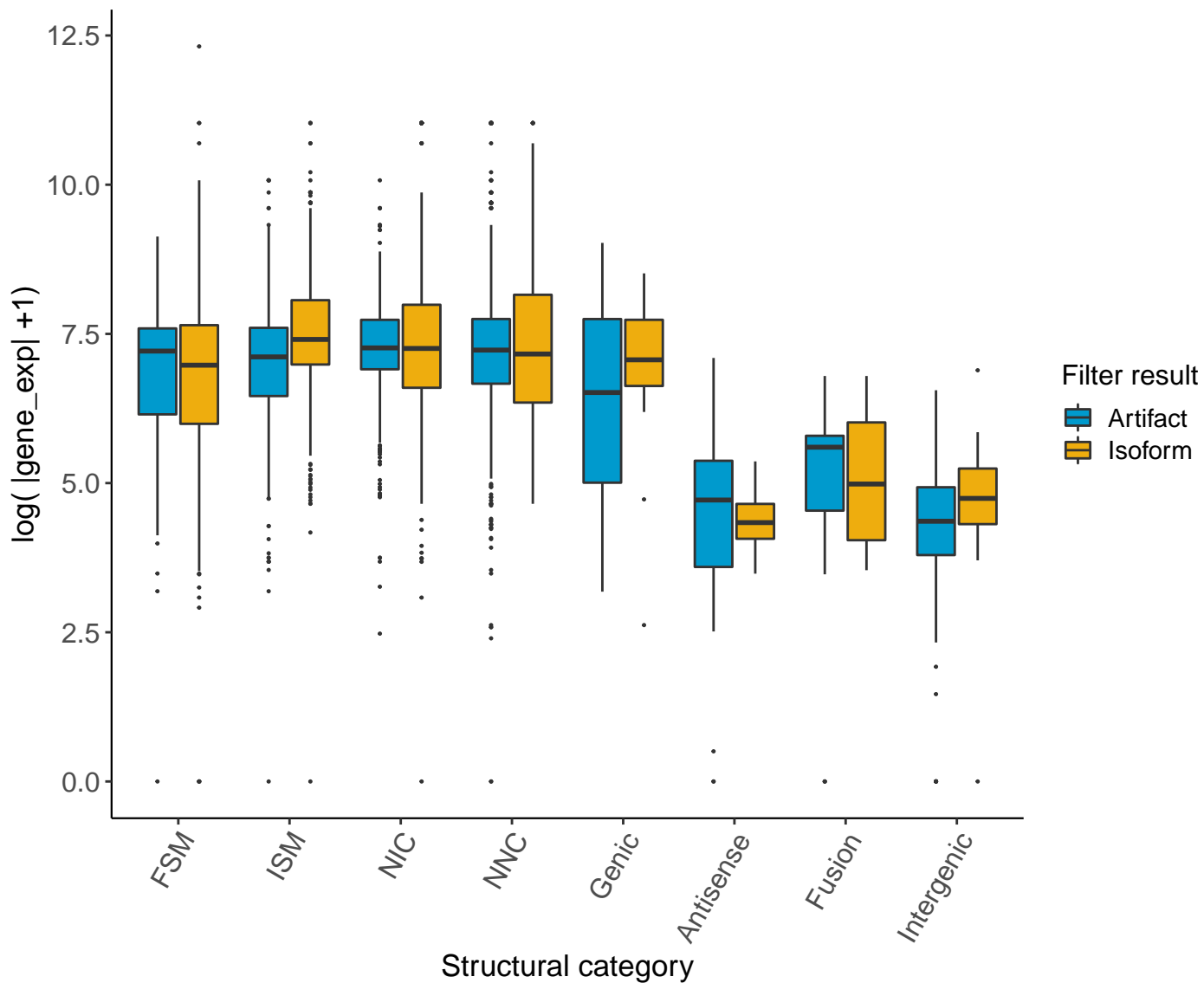
predicted_NMD – ML importance: 13.23



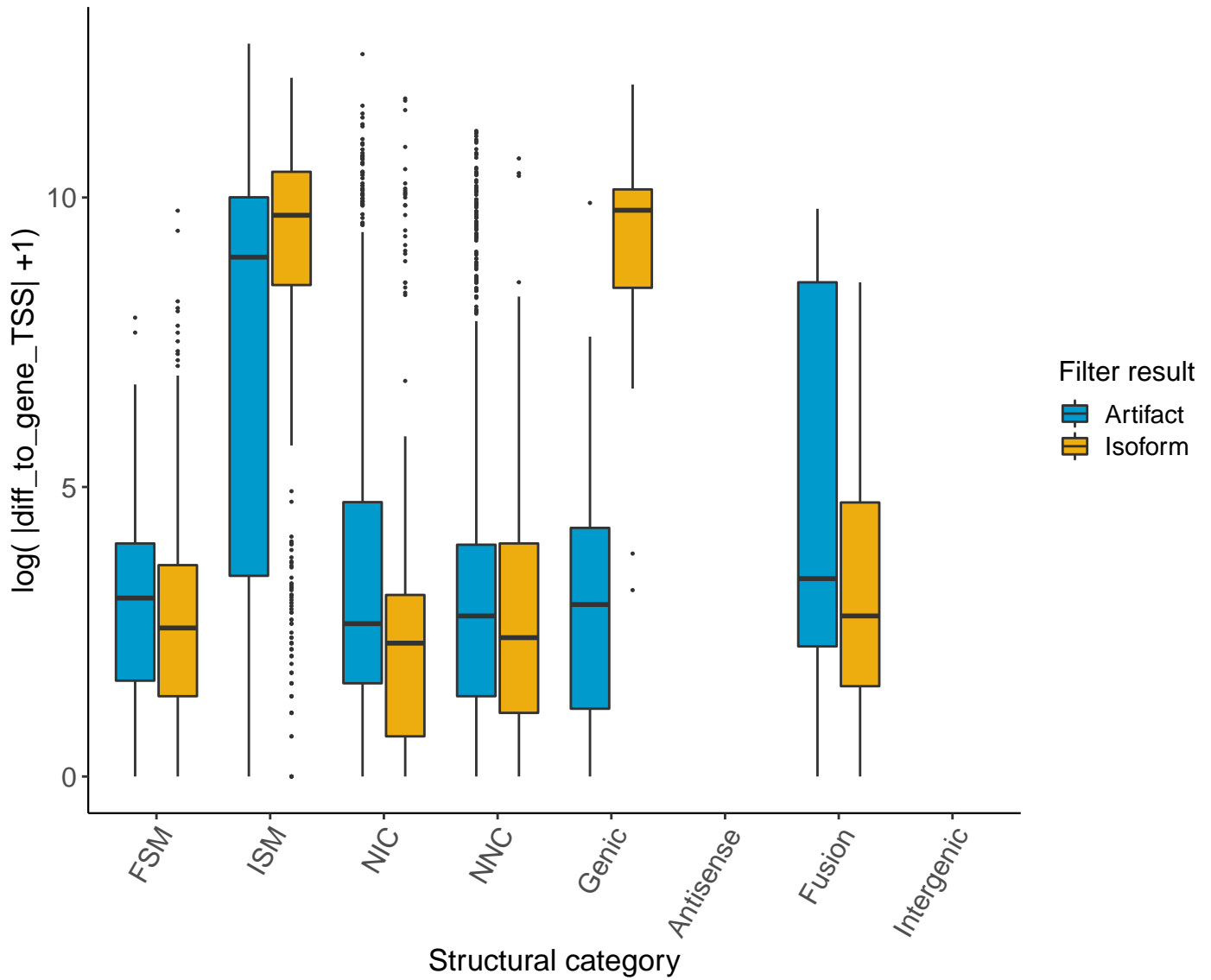
sd_cov – ML importance: 13.11



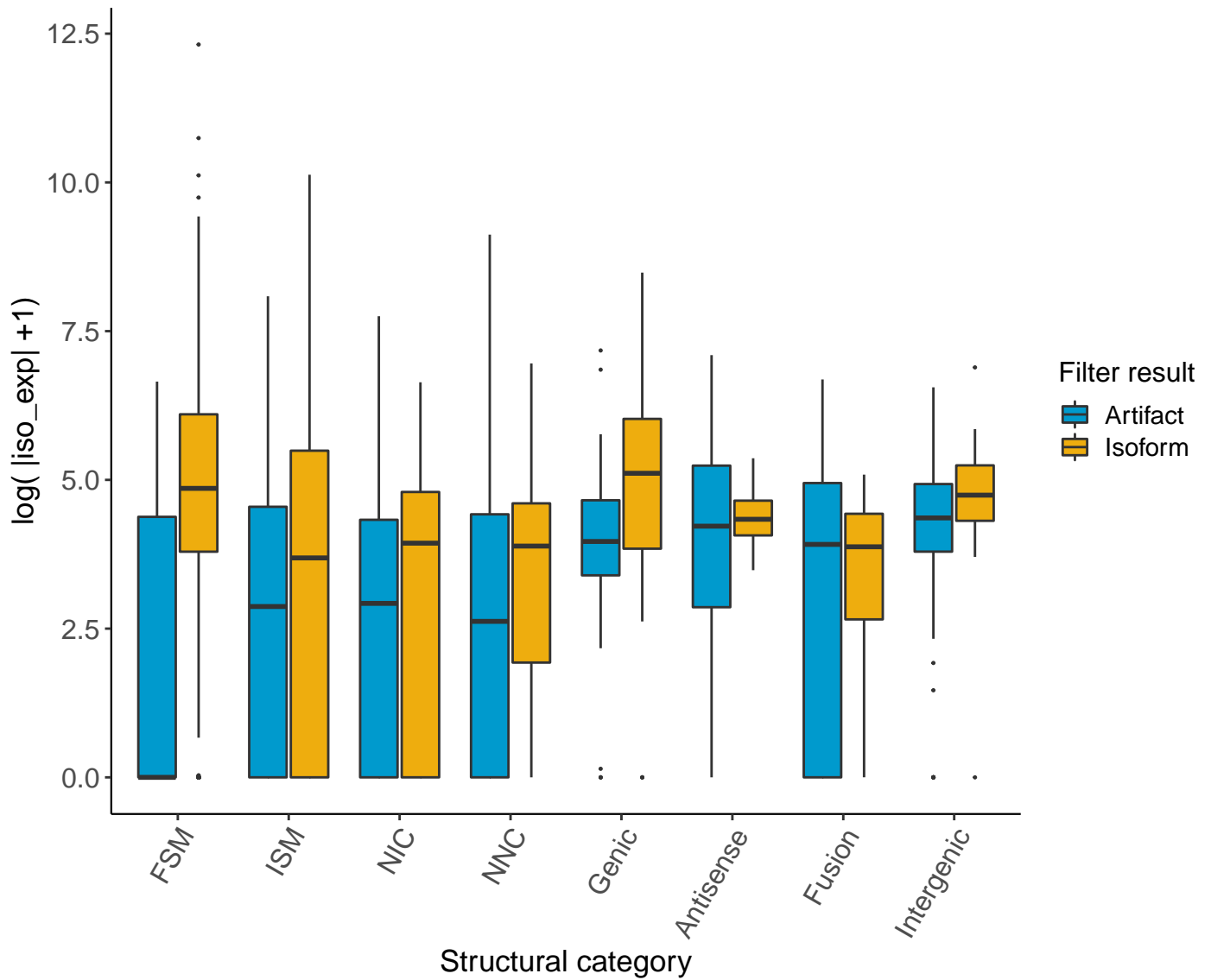
gene_exp – ML importance: 12.83



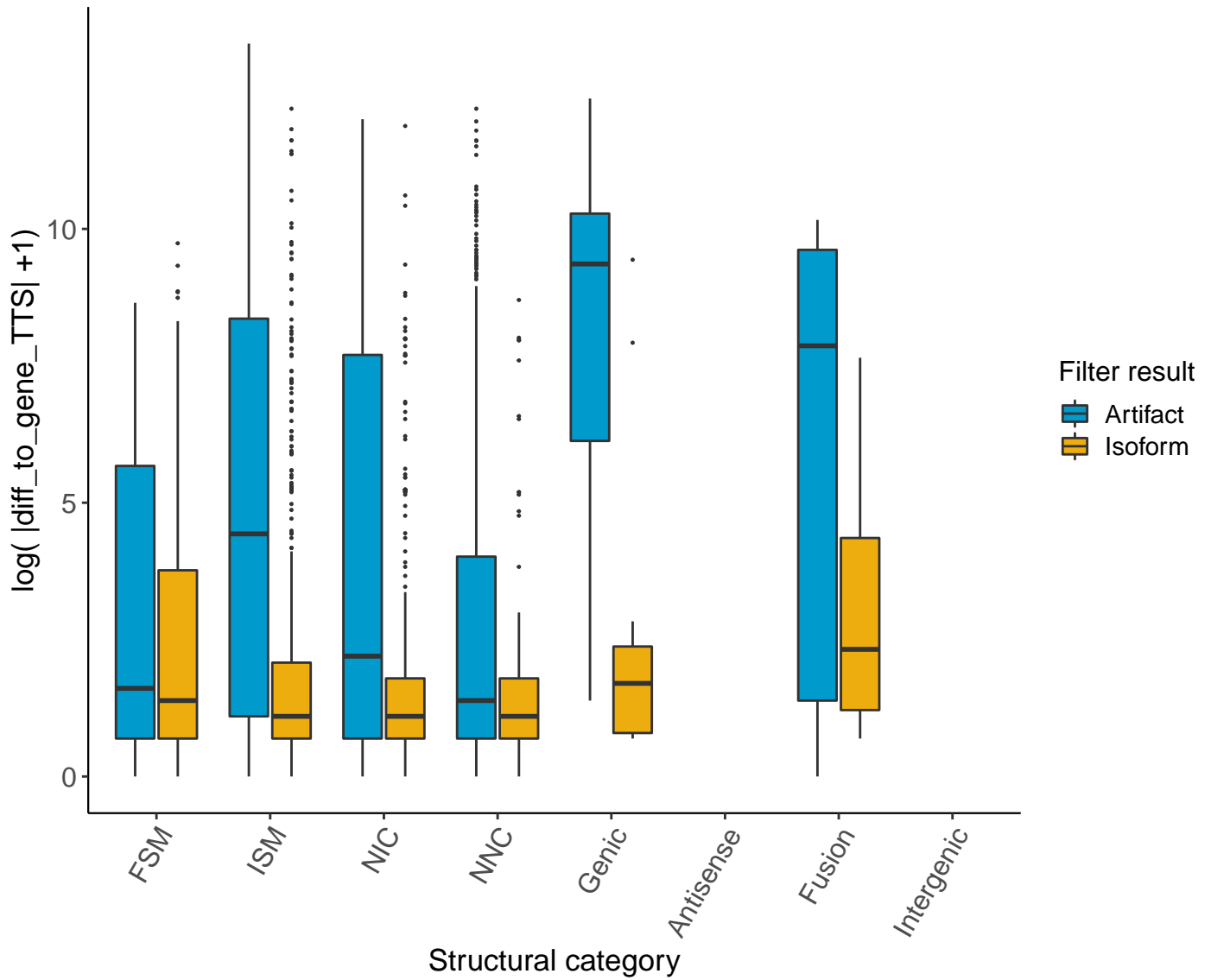
diff_to_gene_TSS – ML importance: 5.98



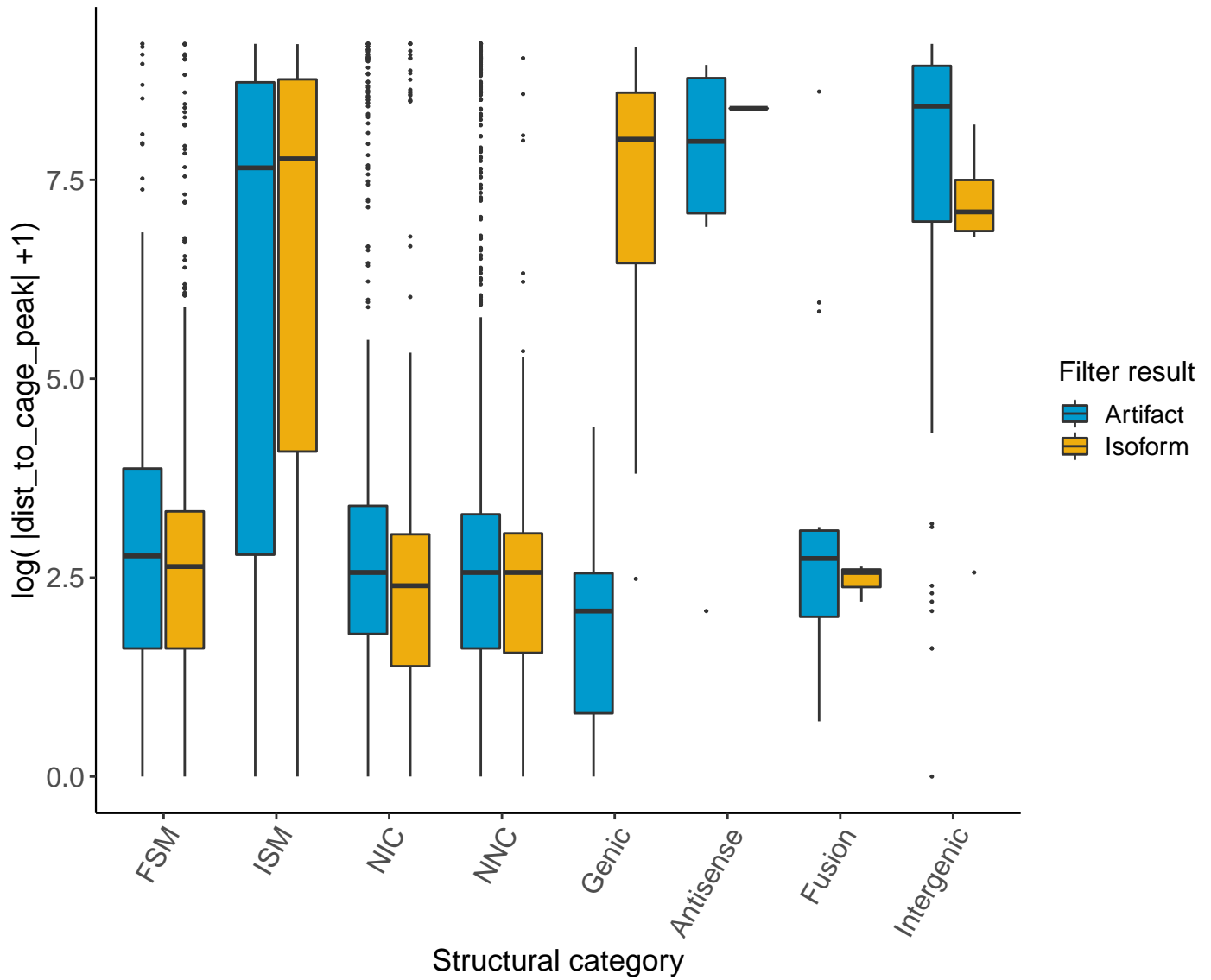
iso_exp – ML importance: 5.87



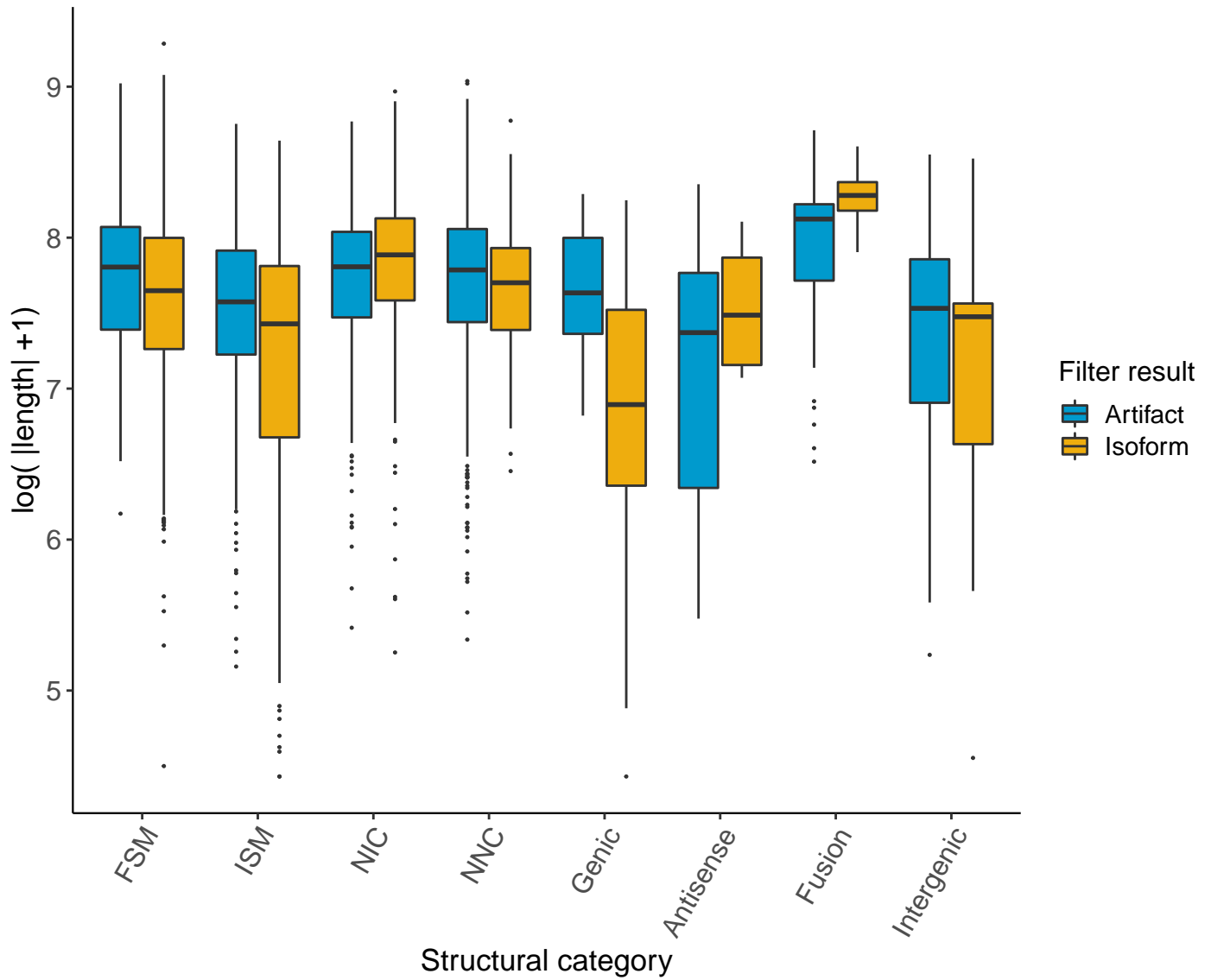
diff_to_gene_TTS – ML importance: 5.22



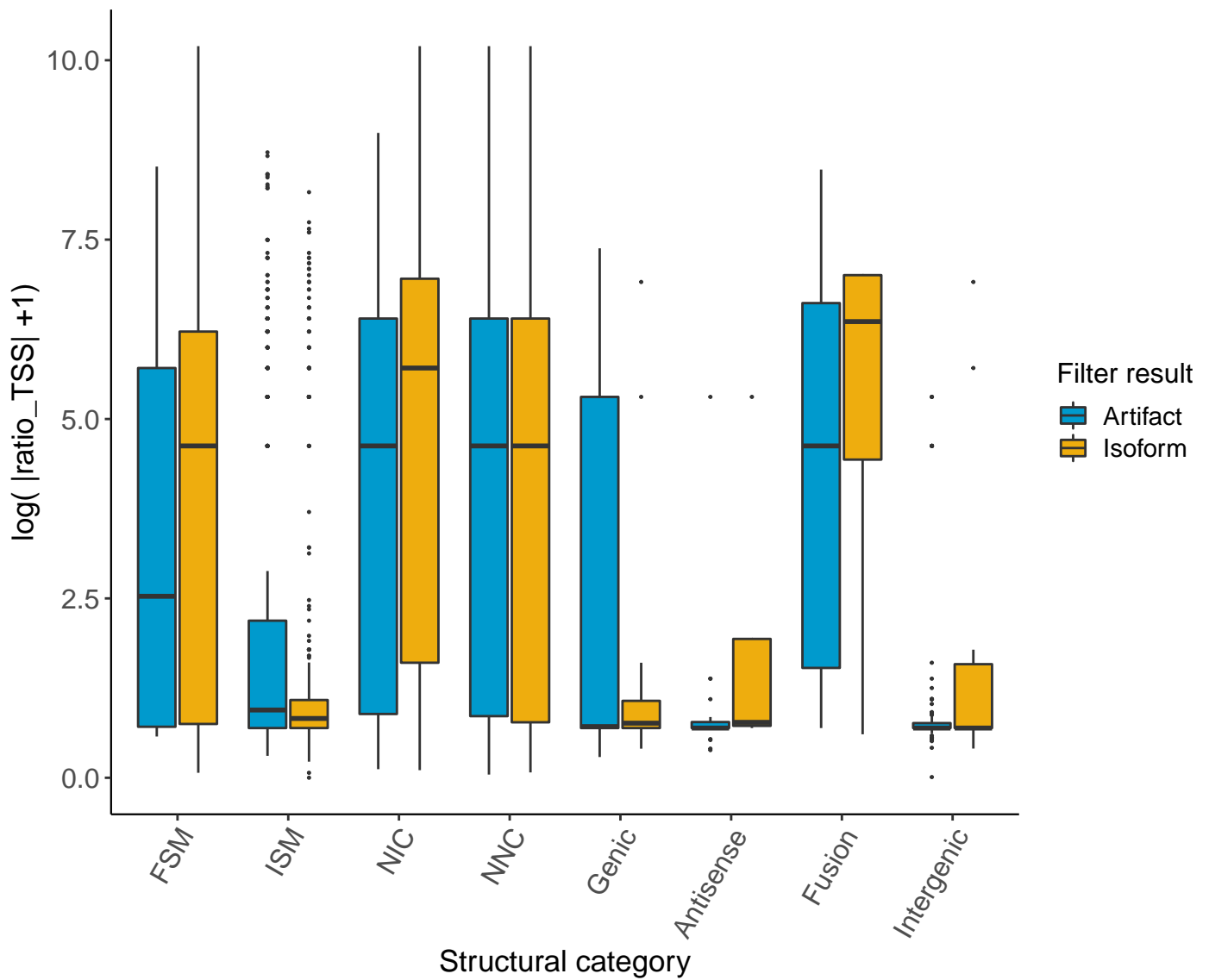
dist_to_cage_peak – ML importance: 4.7



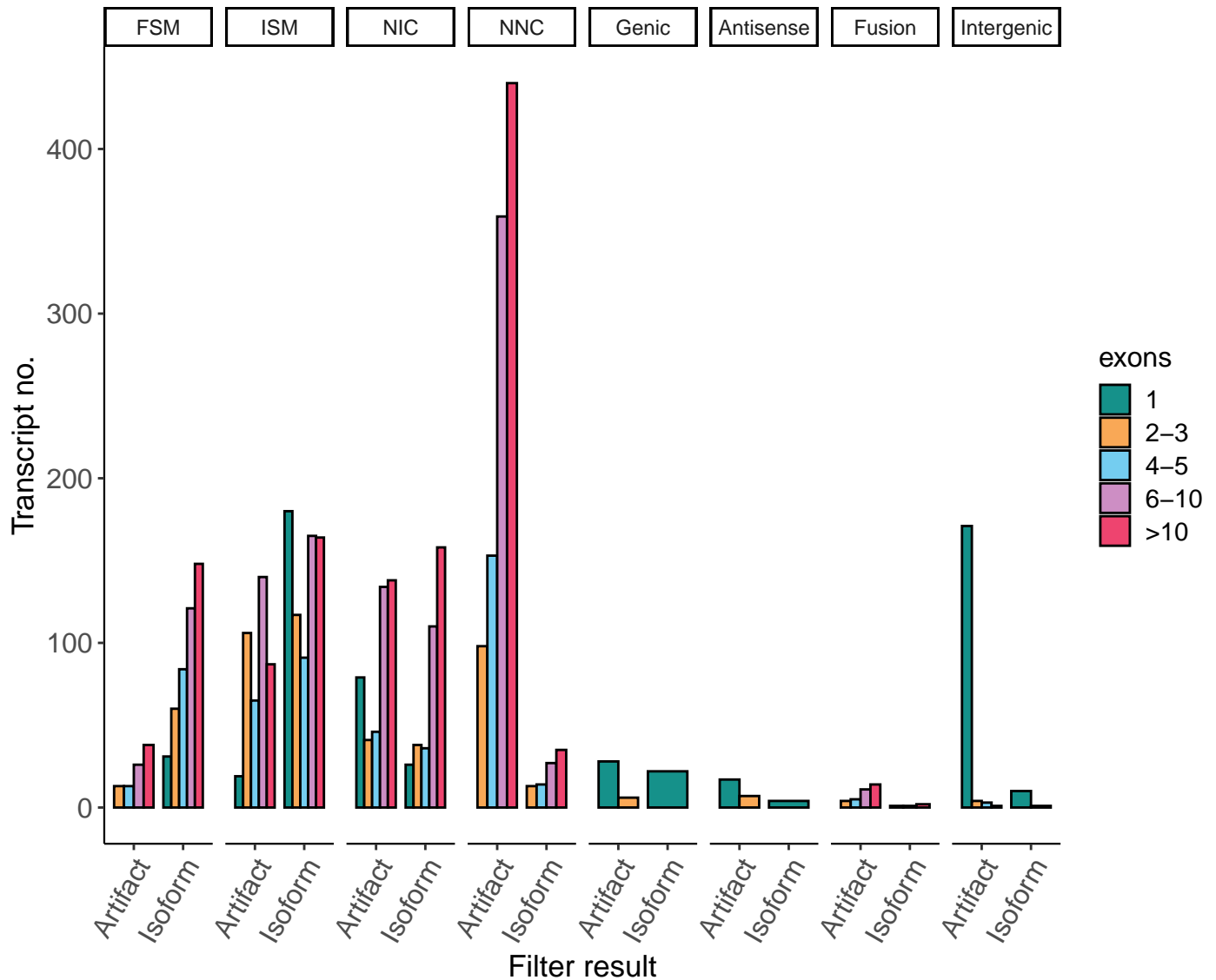
length – ML importance: 4.62



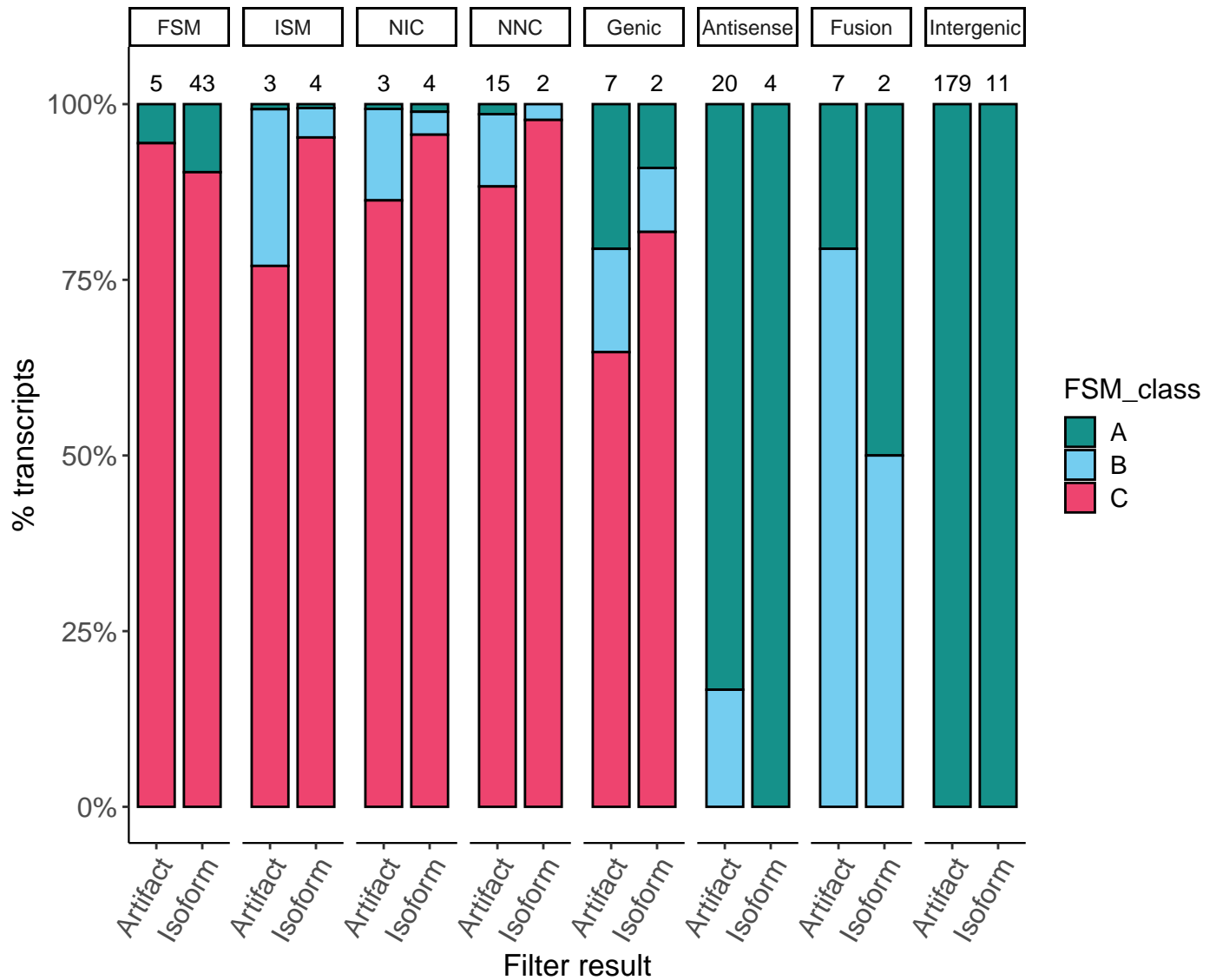
ratio_TSS – ML importance: 4.4



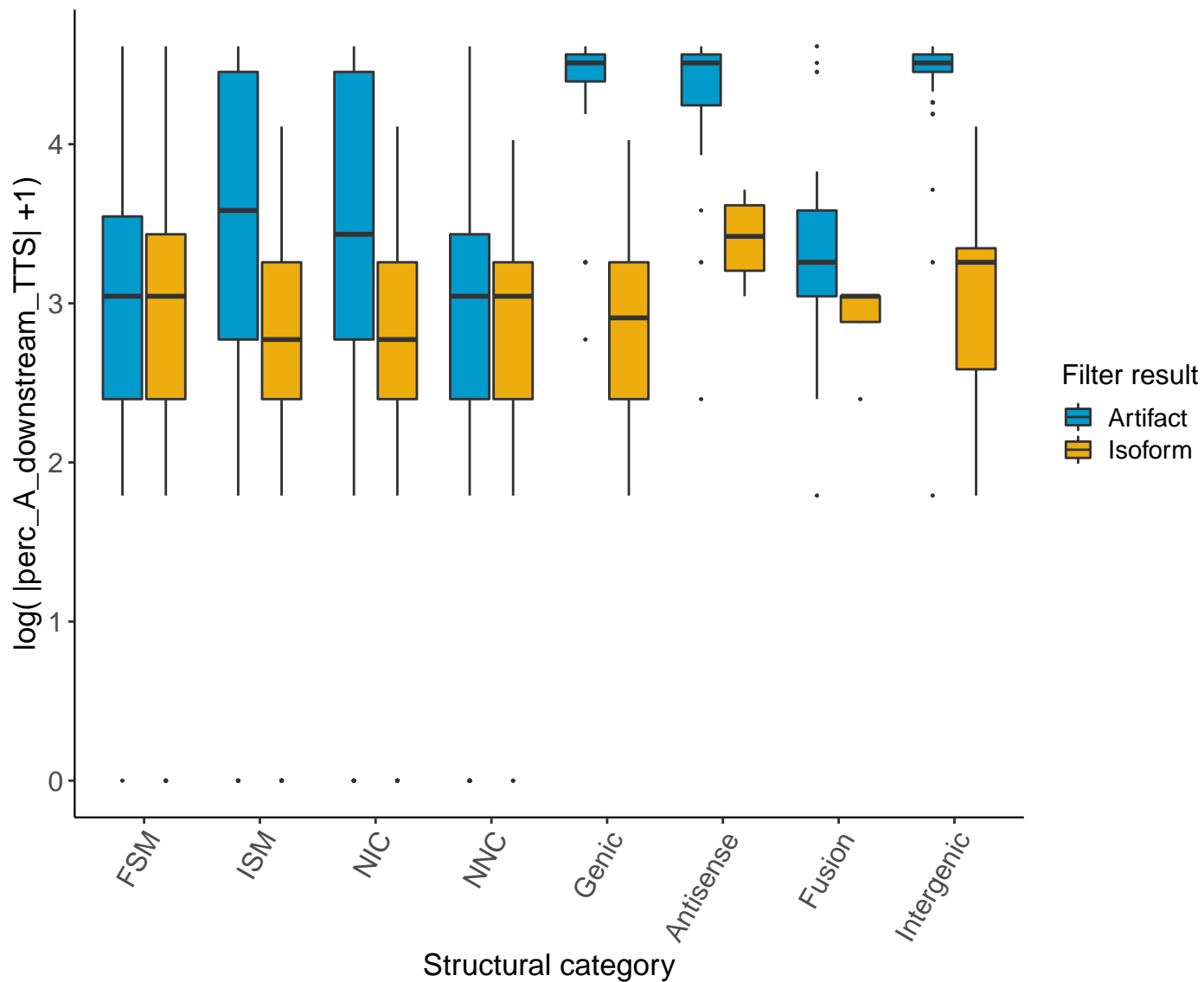
exons – ML importance: 4.4



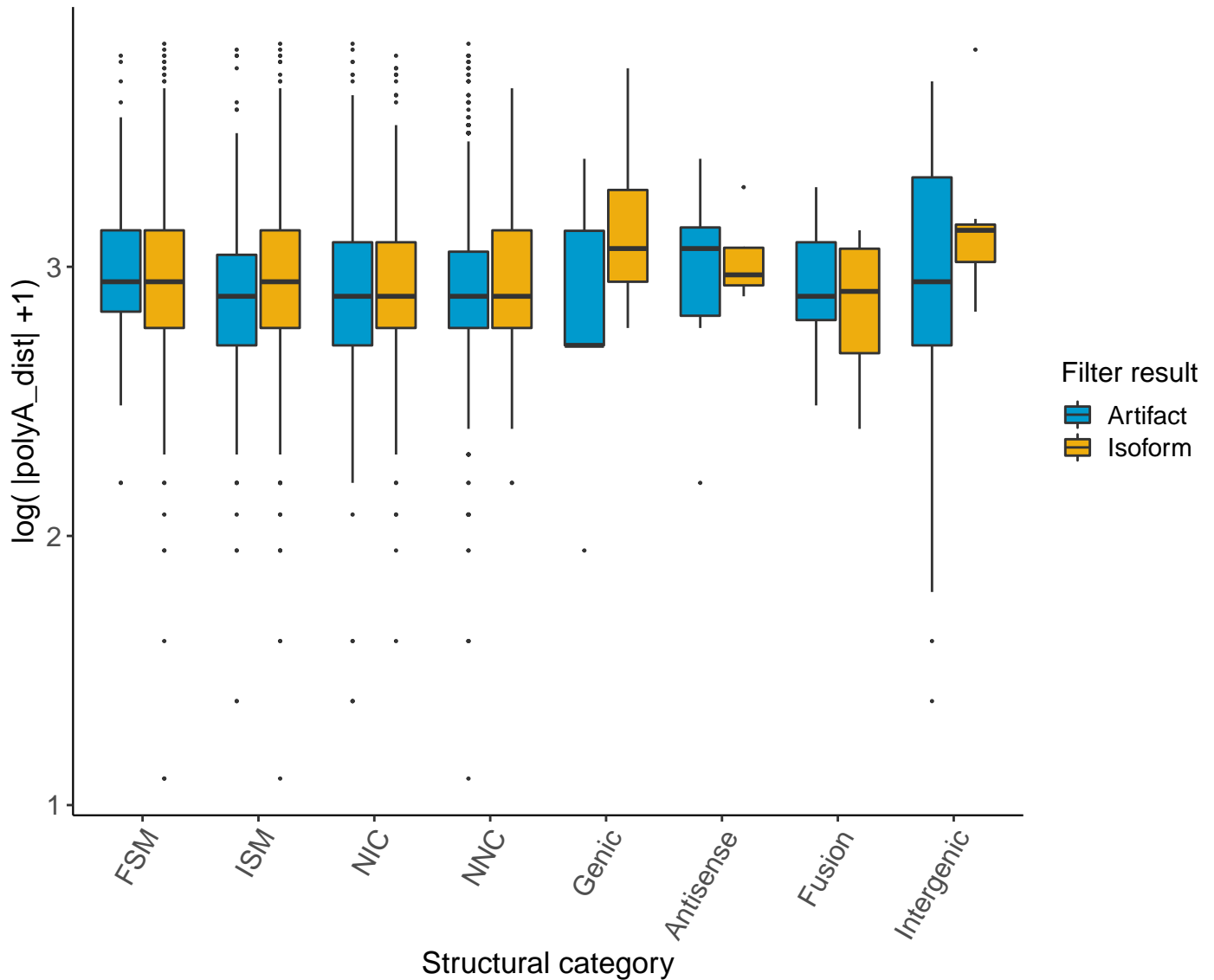
FSM_class – ML importance: 4.27



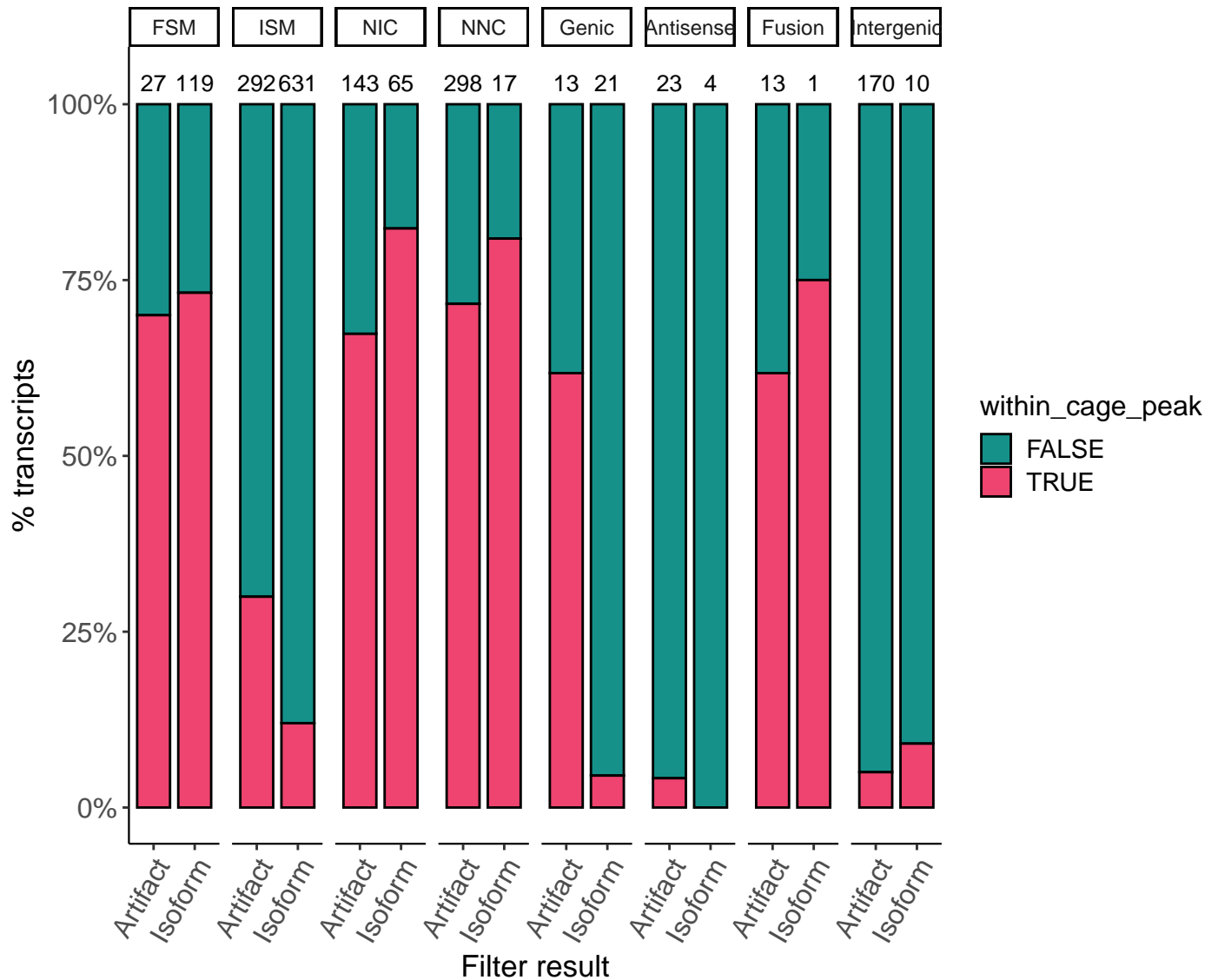
perc_A_downstream_TTS – ML importance: 3.99



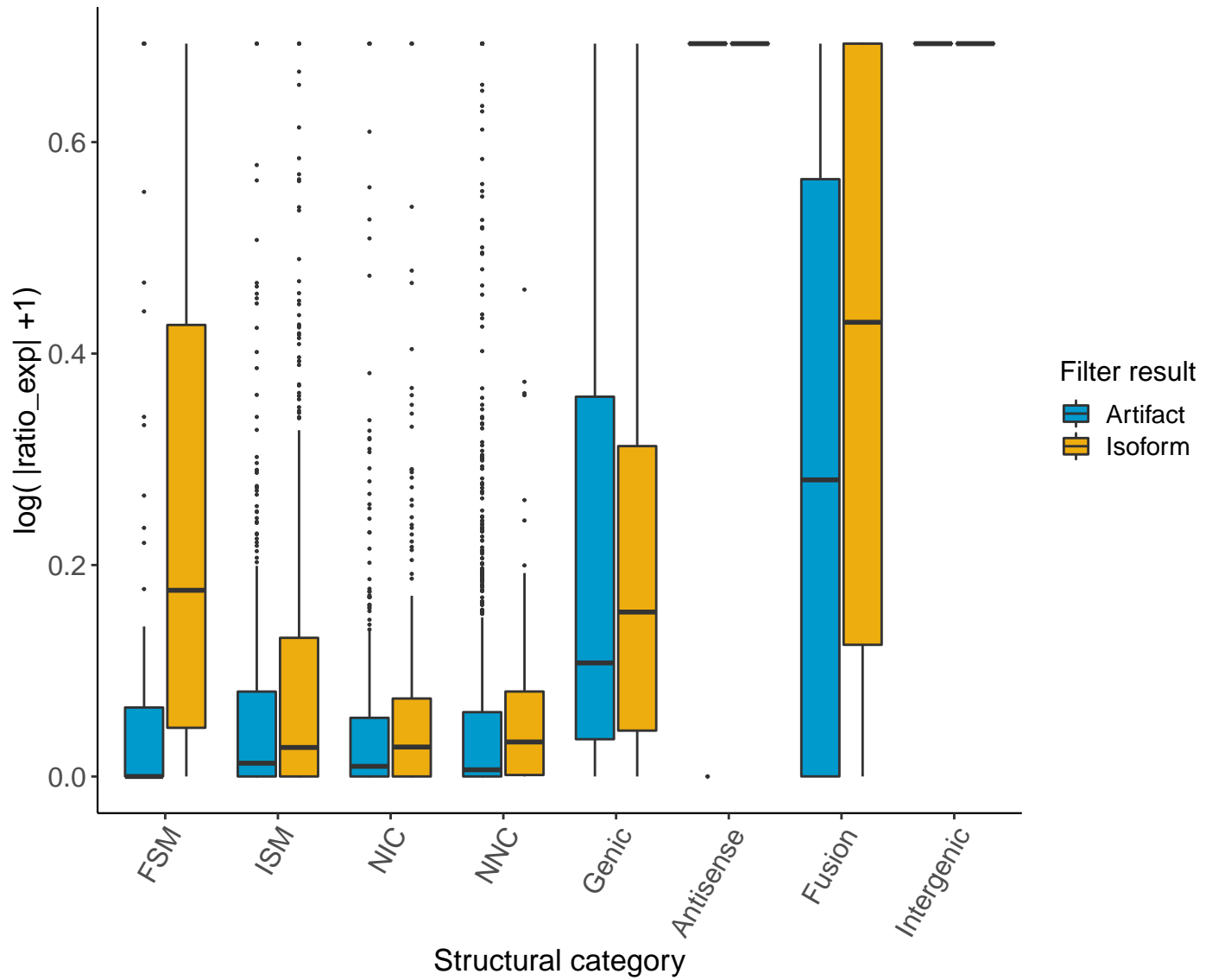
polyA_dist – ML importance: 3.22



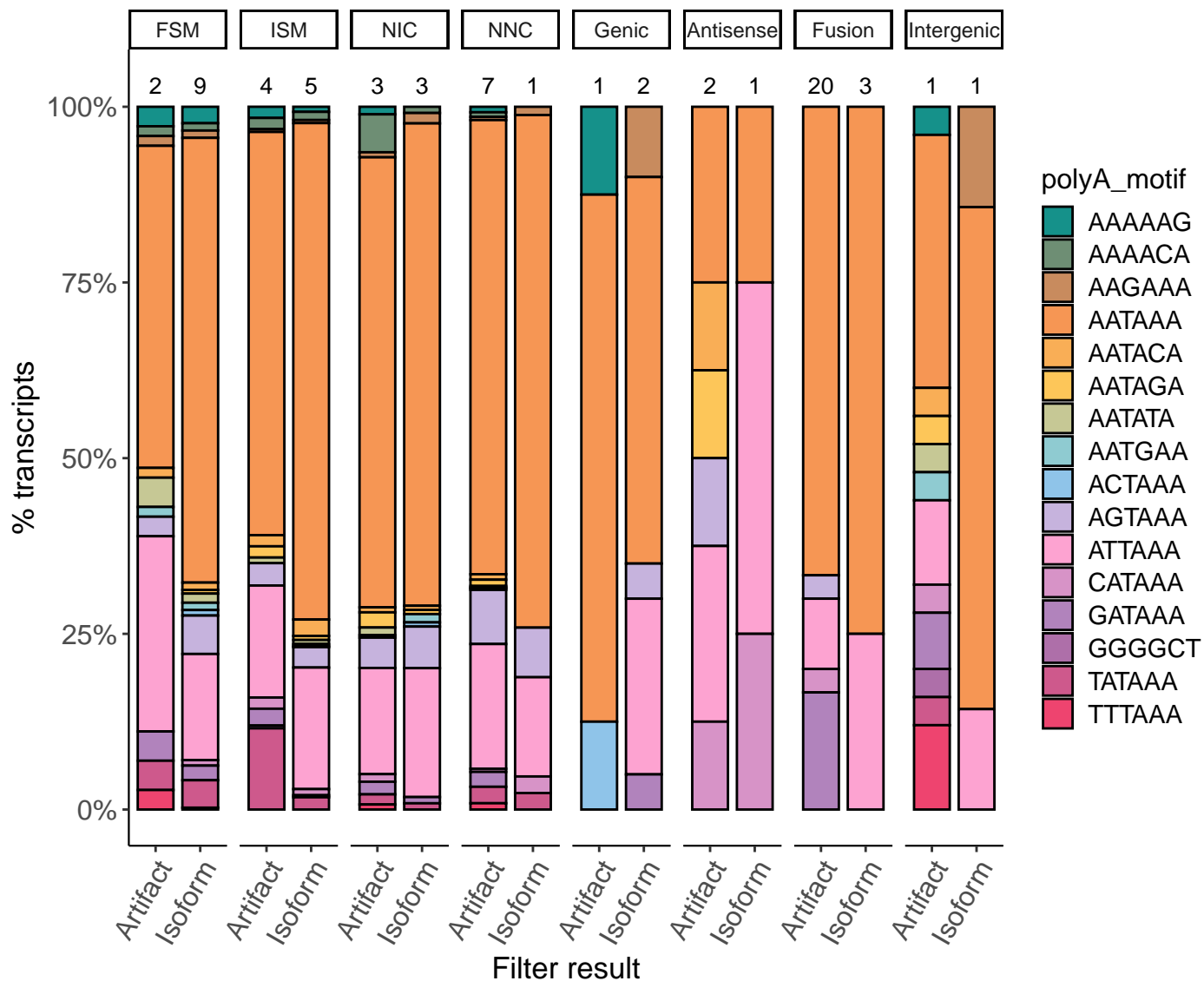
within_cage_peak – ML importance: 0.75



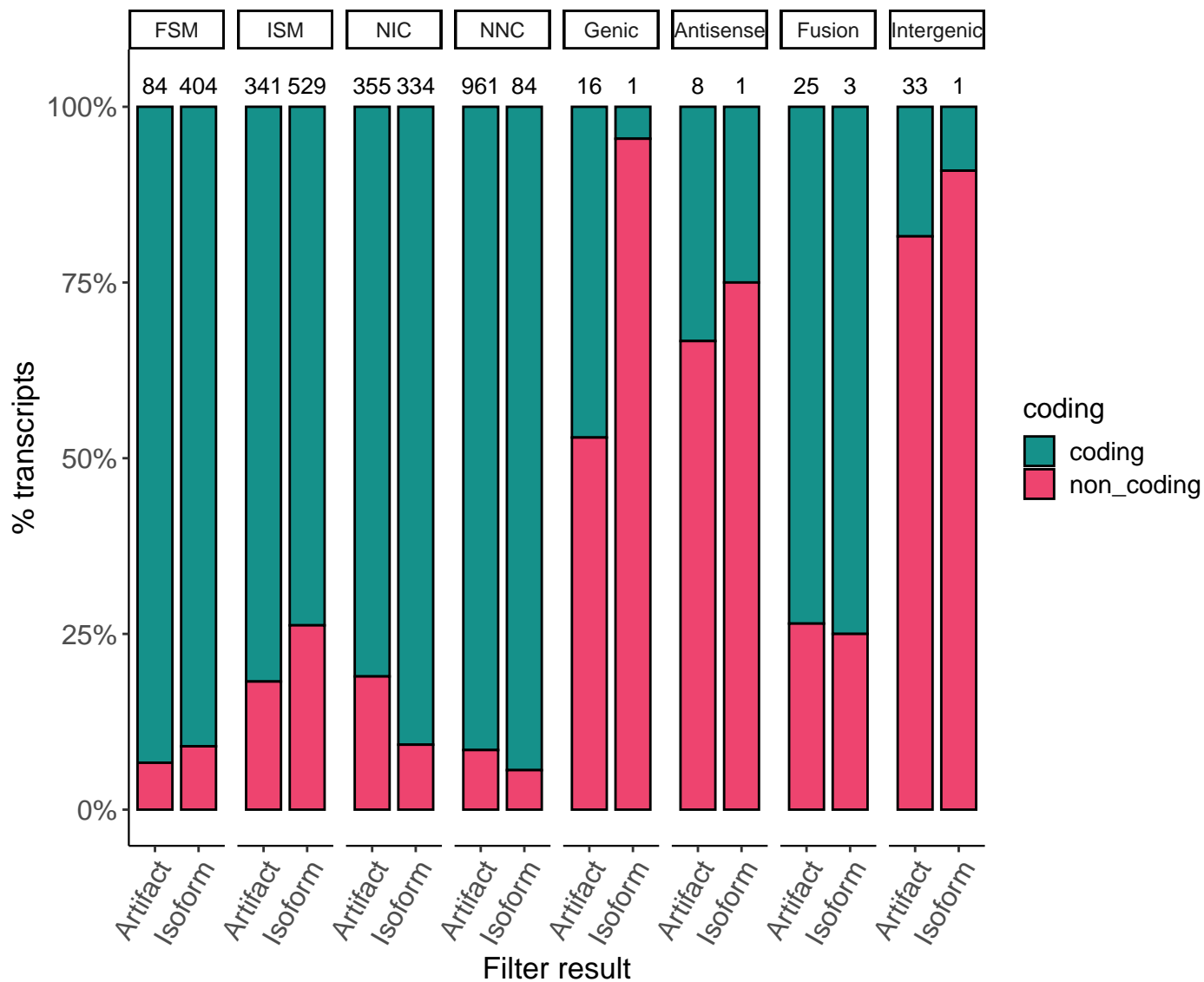
ratio_exp – ML importance: 0.61



polyA_motif – ML importance: 0.41

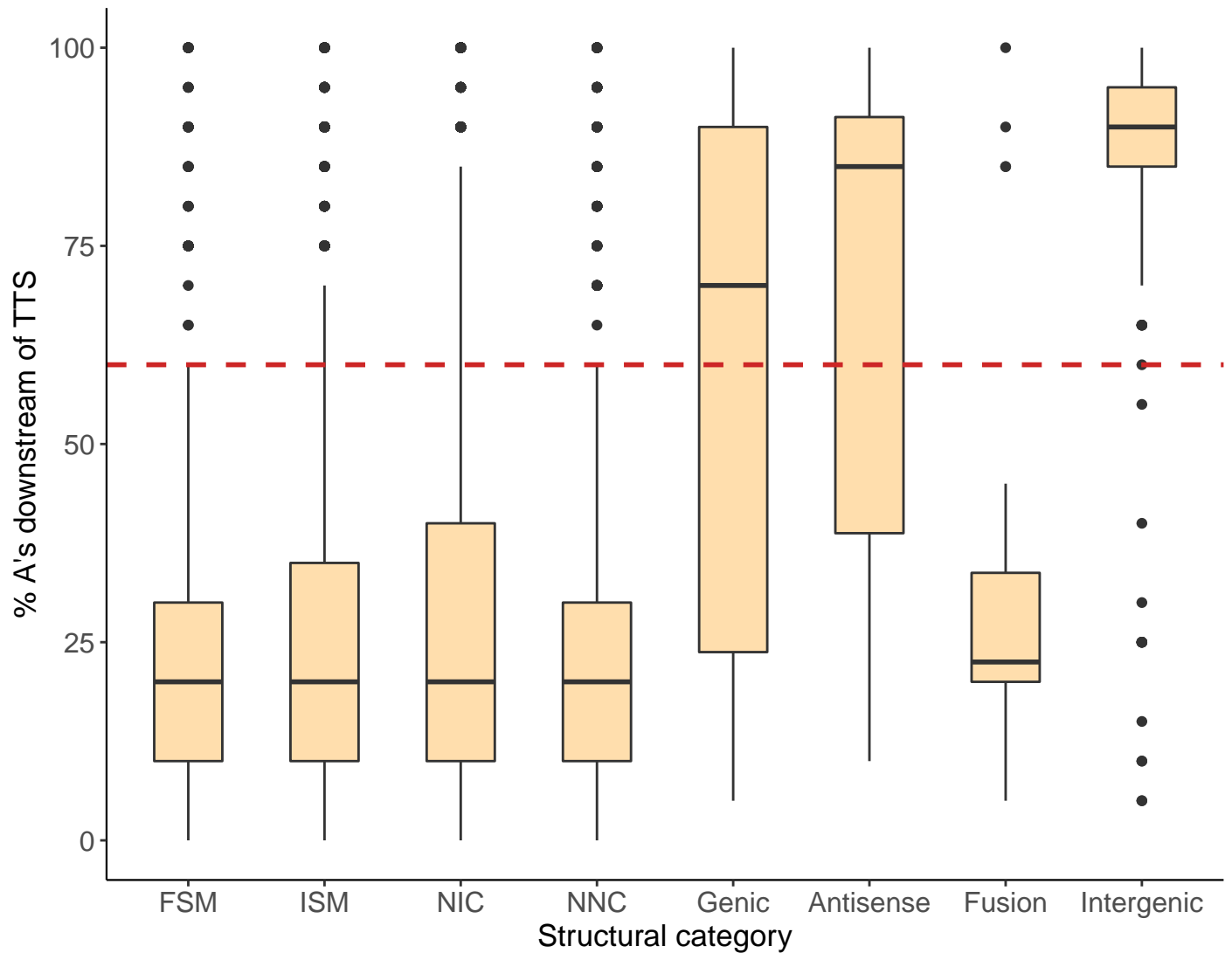


coding – ML importance: 0.16

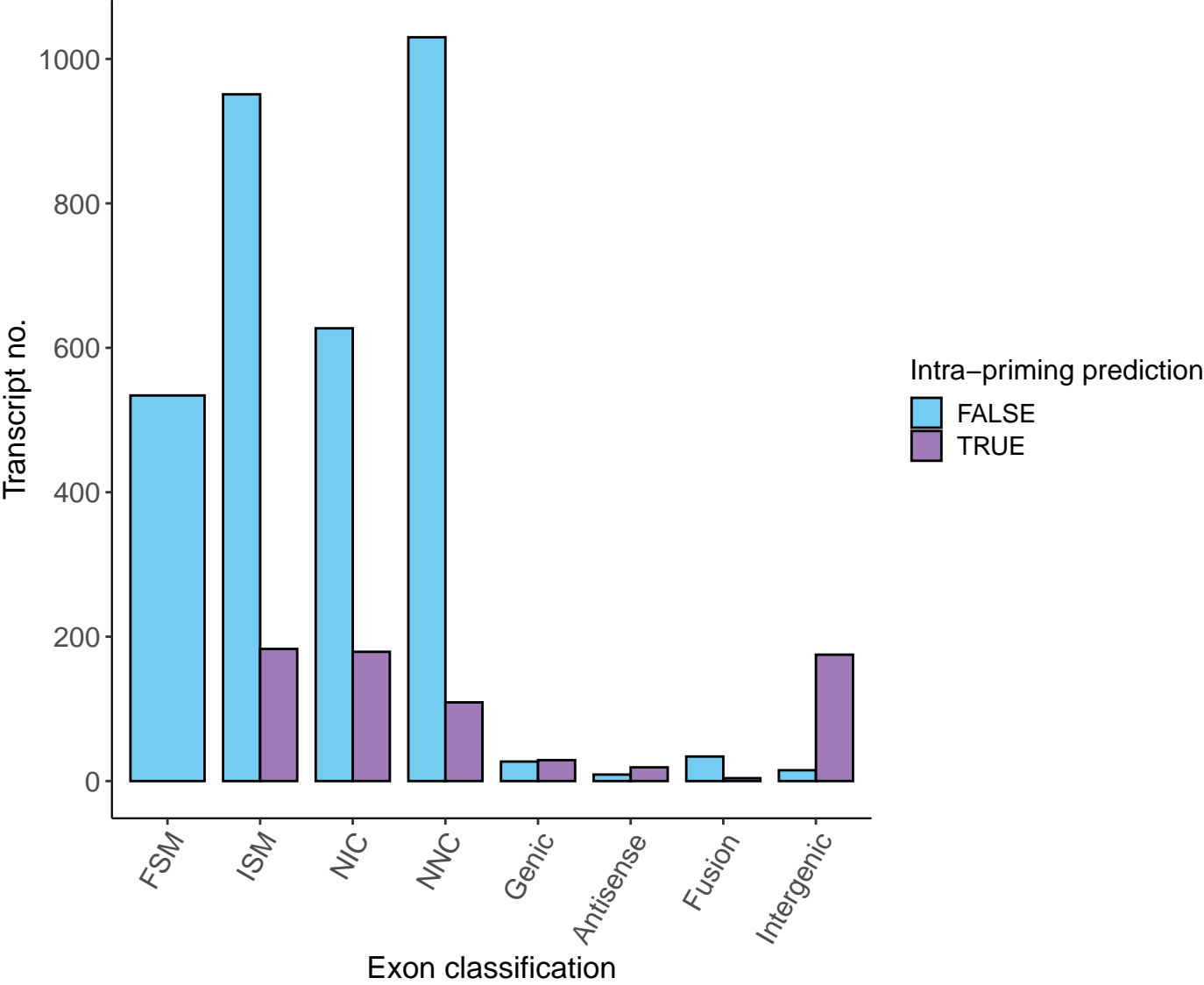


A % by category

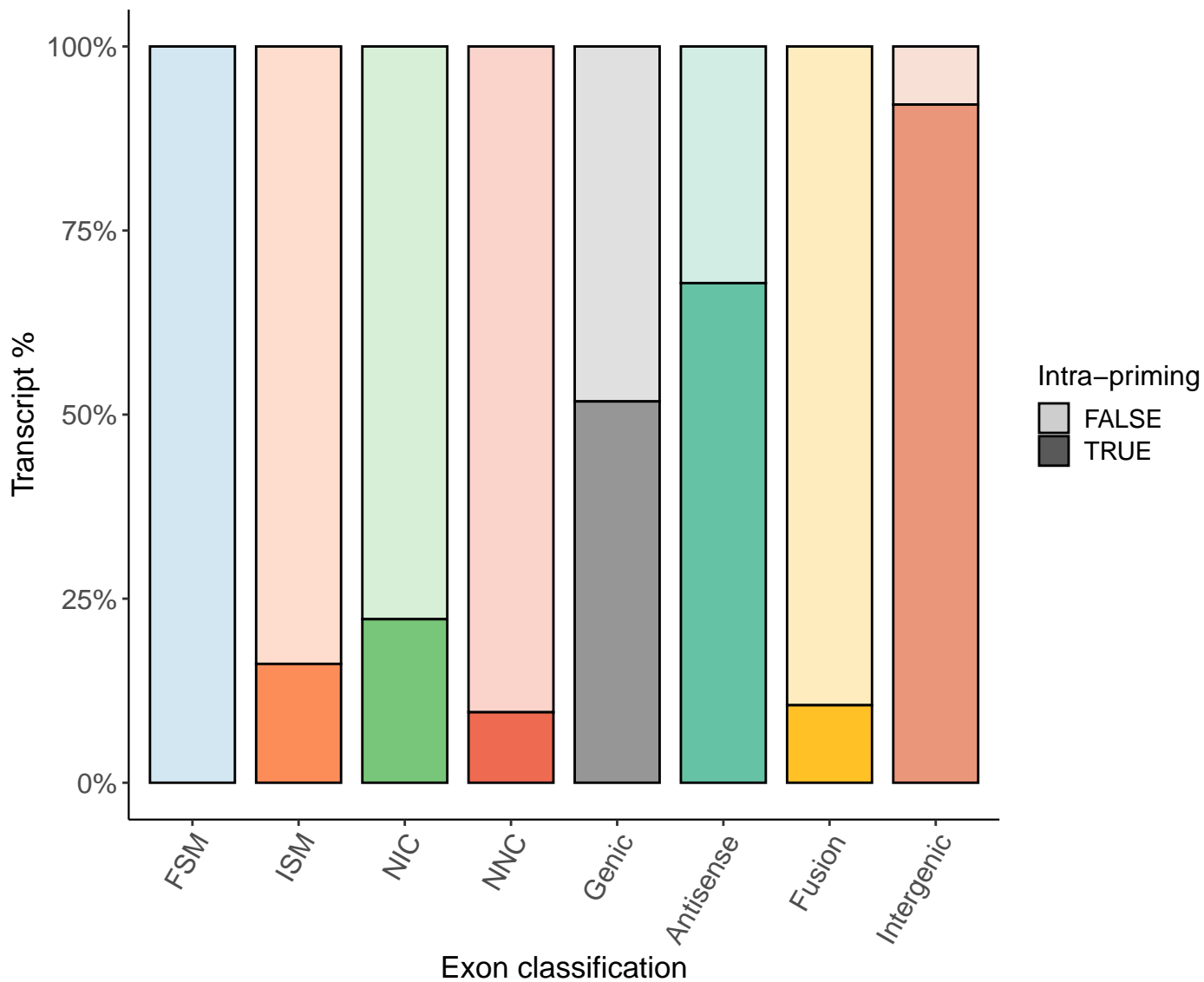
Red line indicates threshold employed in ML filter



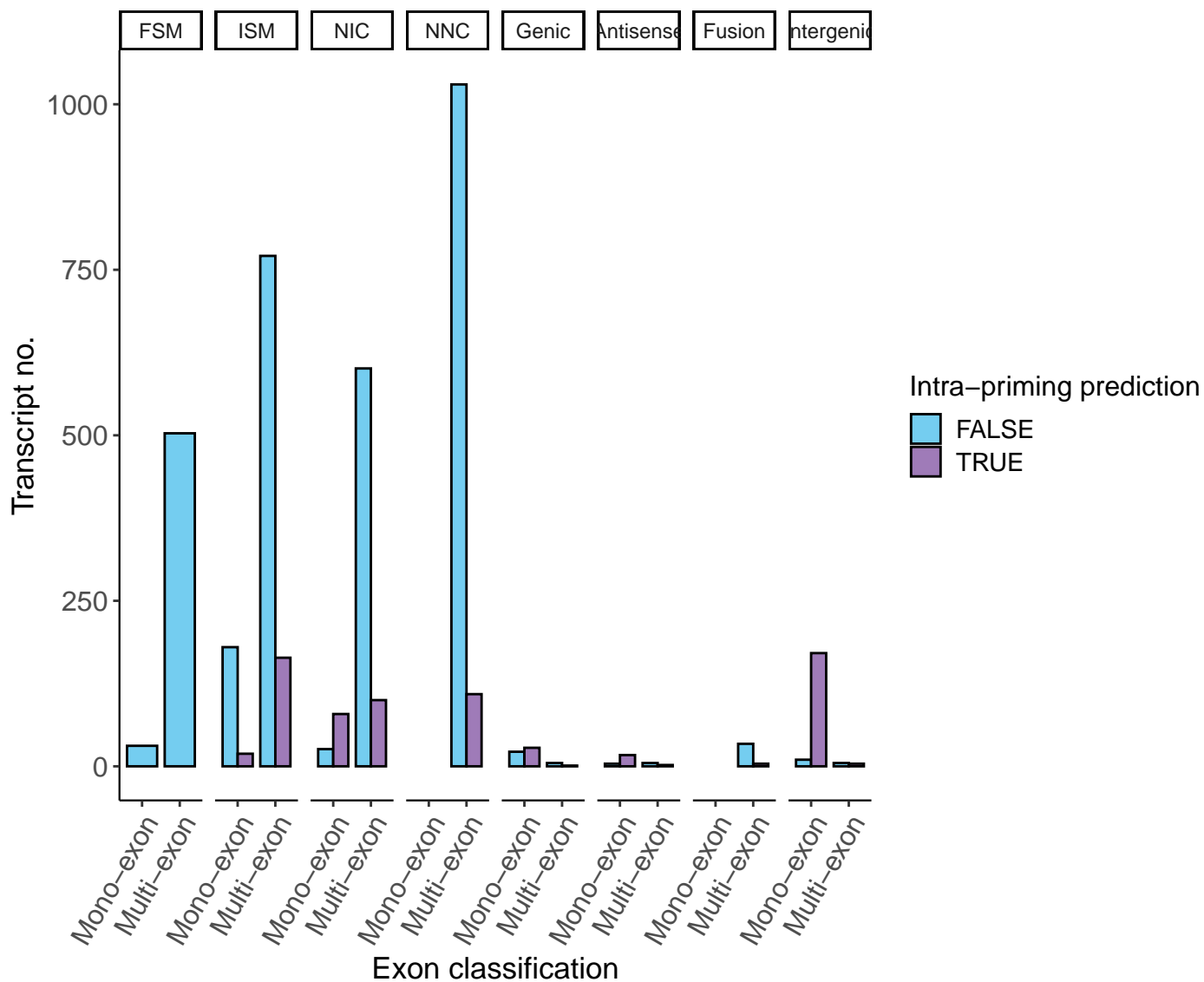
Isoforms flagged as intra-priming, by category



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



Isoforms flagged as intra-priming, by exon number



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

