

SQANTI3 filter report

Total Genes: 650

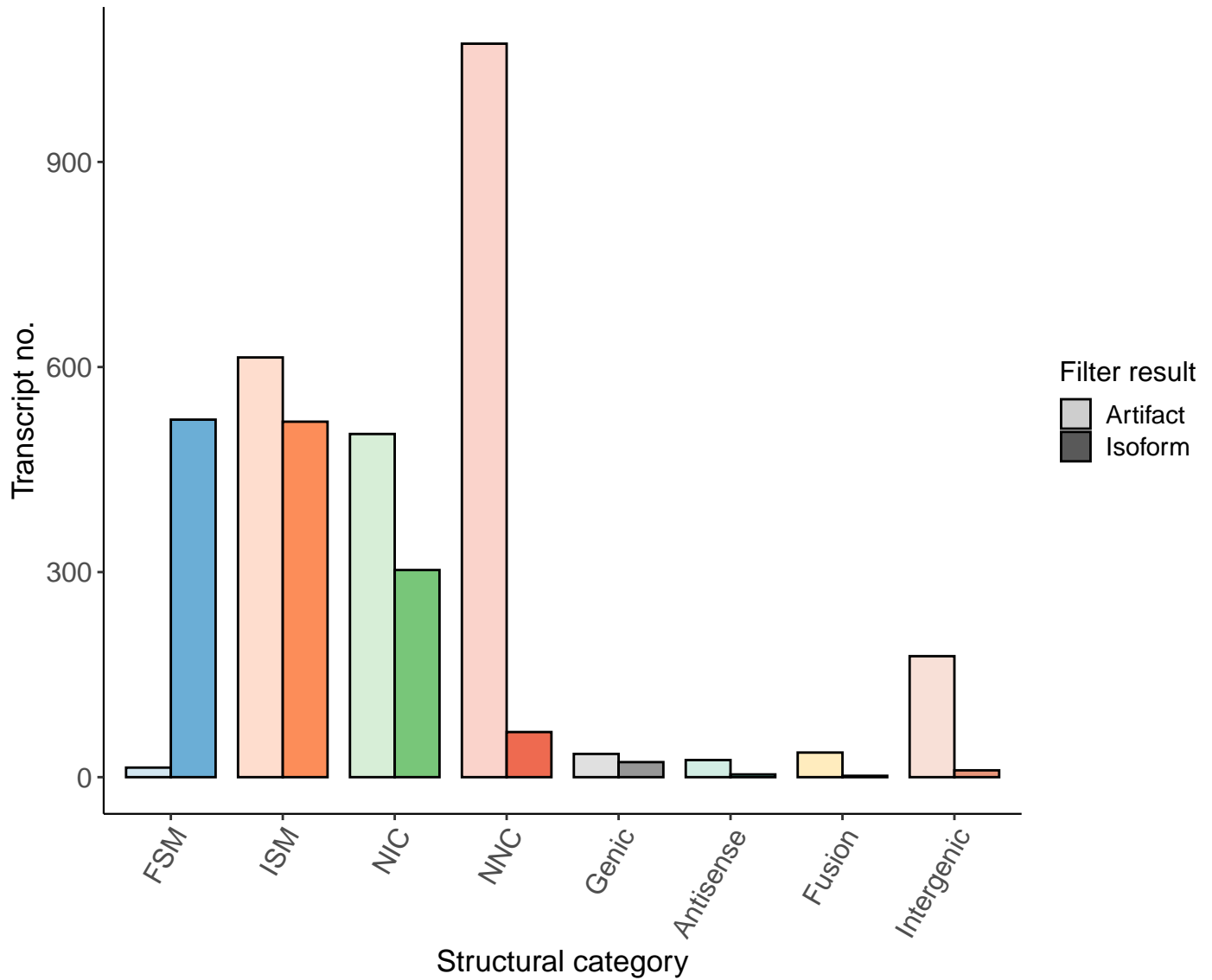
Total Transcripts: 3925

– Isoforms: 1450 (37%)

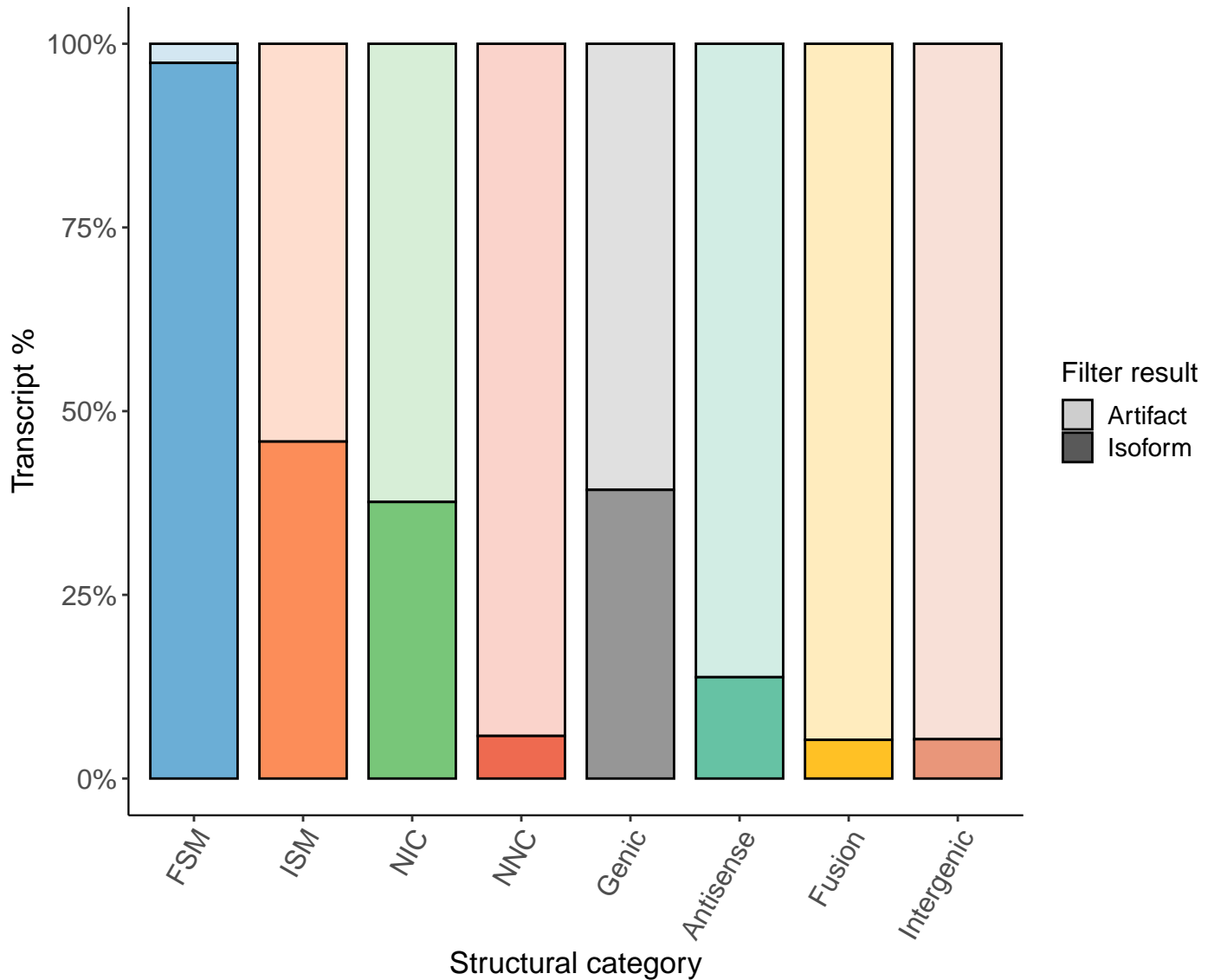
– Artifacts: 2475 (63%)

			Structural category	Artifact no.	Isoform no.
			FSM	14	523
			ISM	614	520
			NIC	502	303
			NNC	1073	66
			Genic	34	22
			Antisense	25	4
			Fusion	36	2
			Intergenic	177	10
Gene category	Gene no.	No. of genes with artifacts only			
Annotated	437	81			
Novel	213	199			

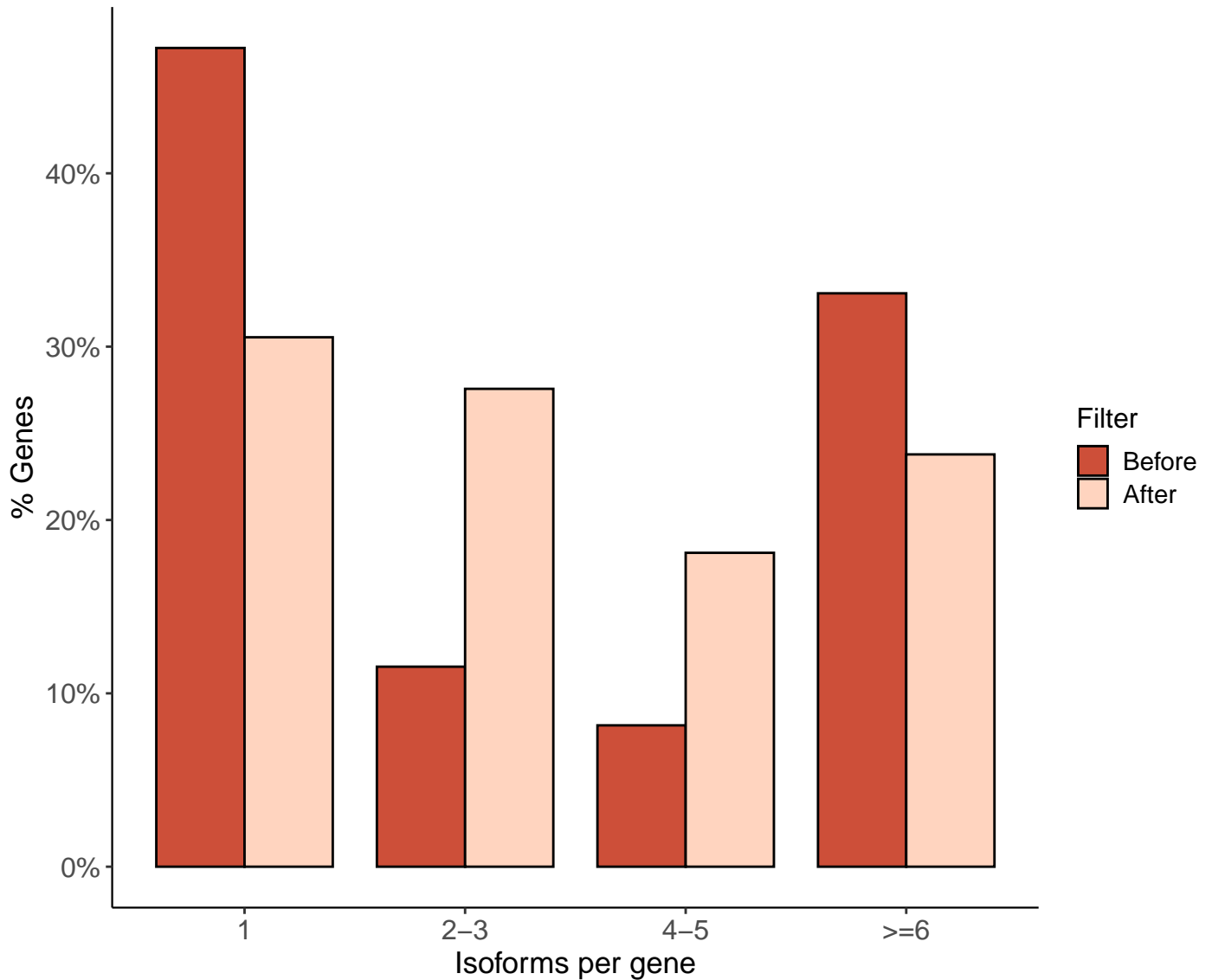
Total isoforms and artifacts by category



% isoforms and artifacts by category

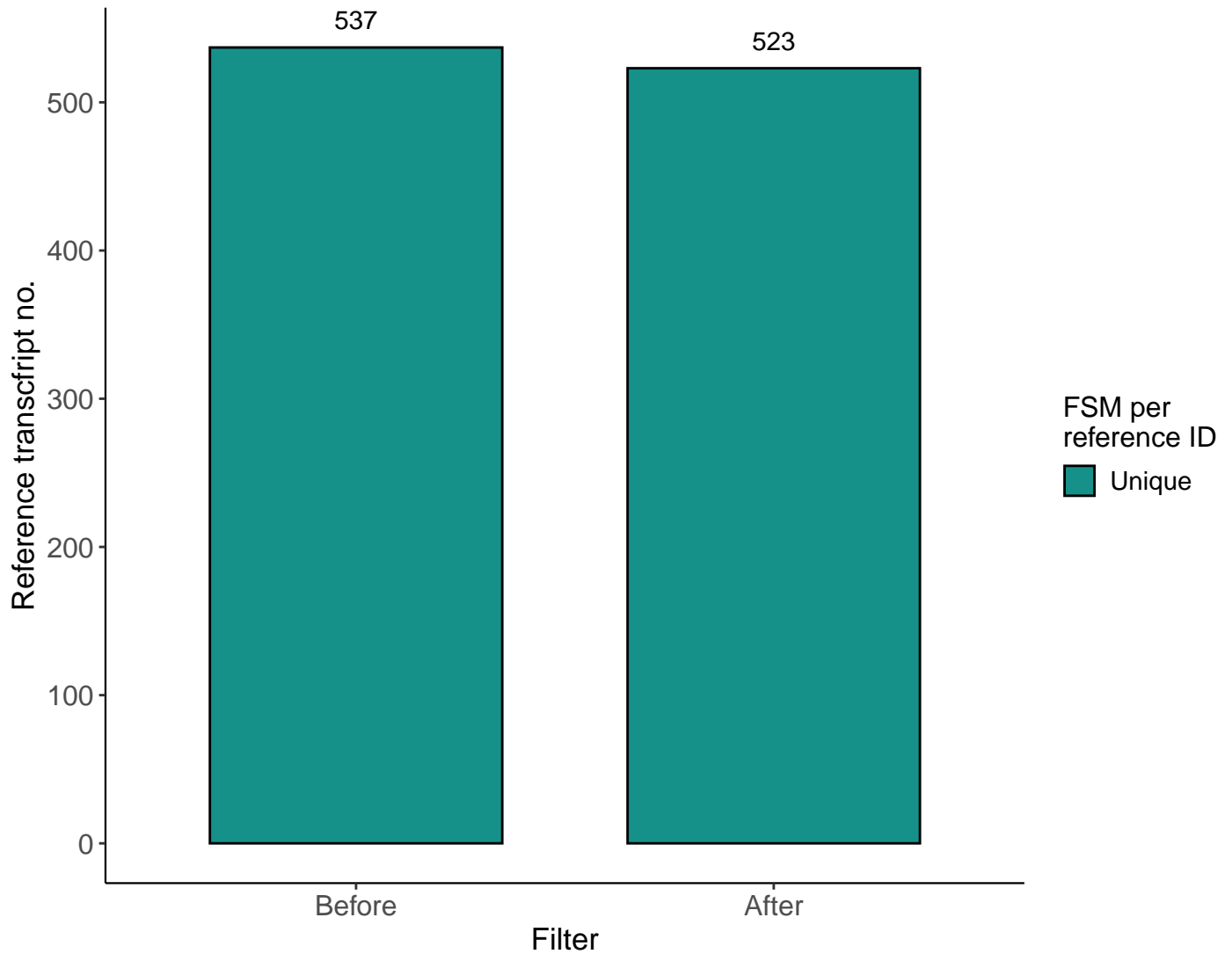


Number of isoforms per gene

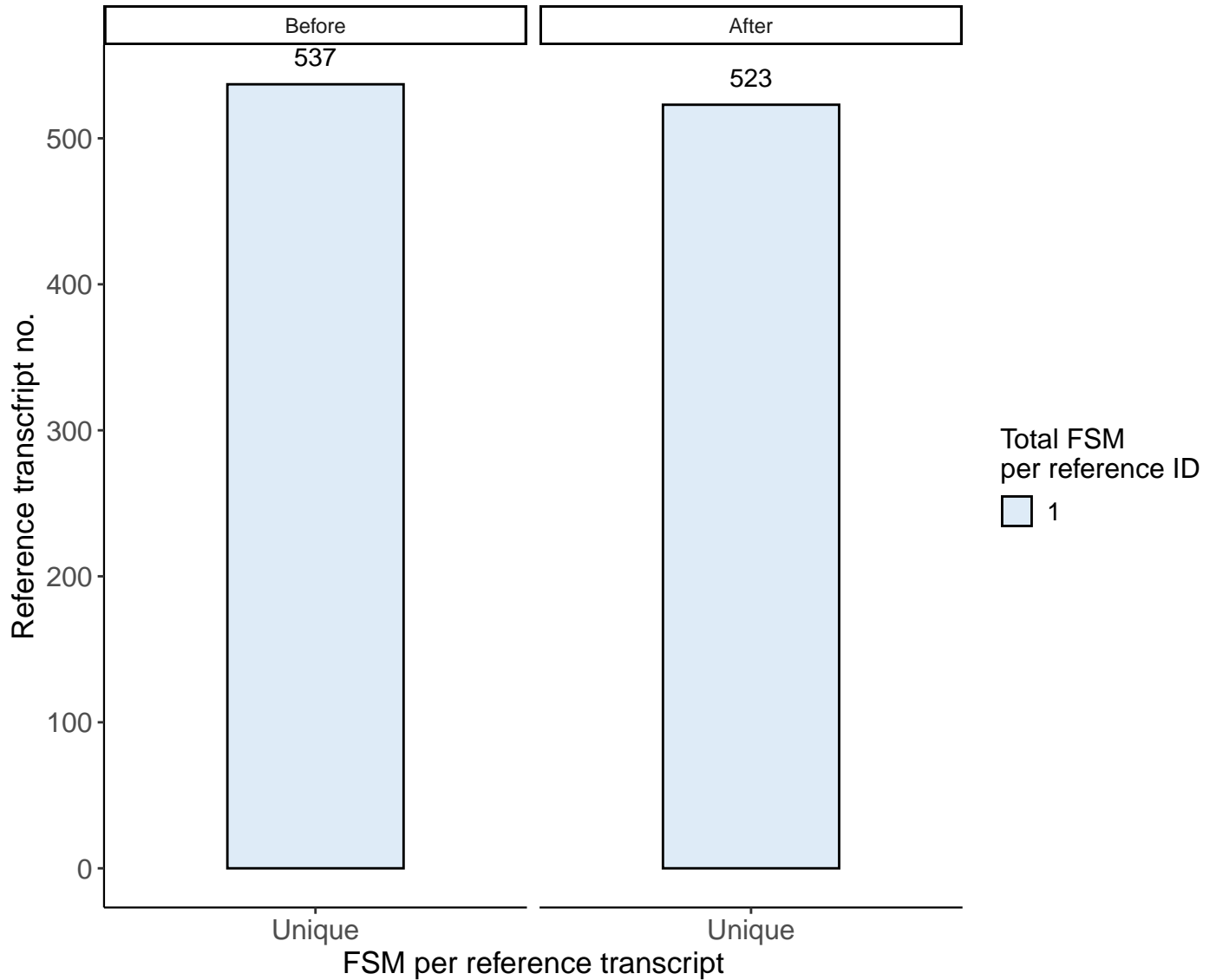


Reference transcript complexity

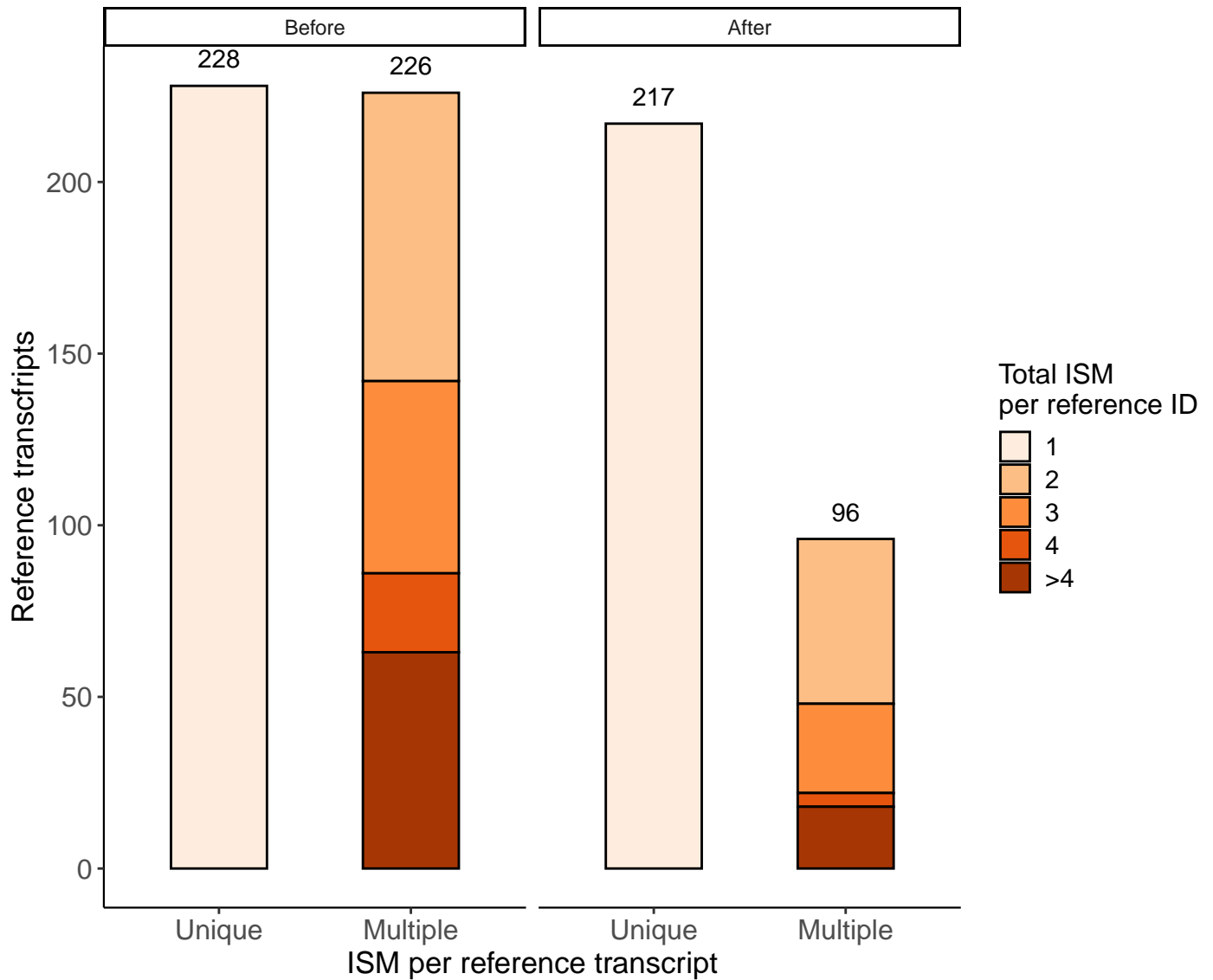
No. of reference transcripts represented by FSM



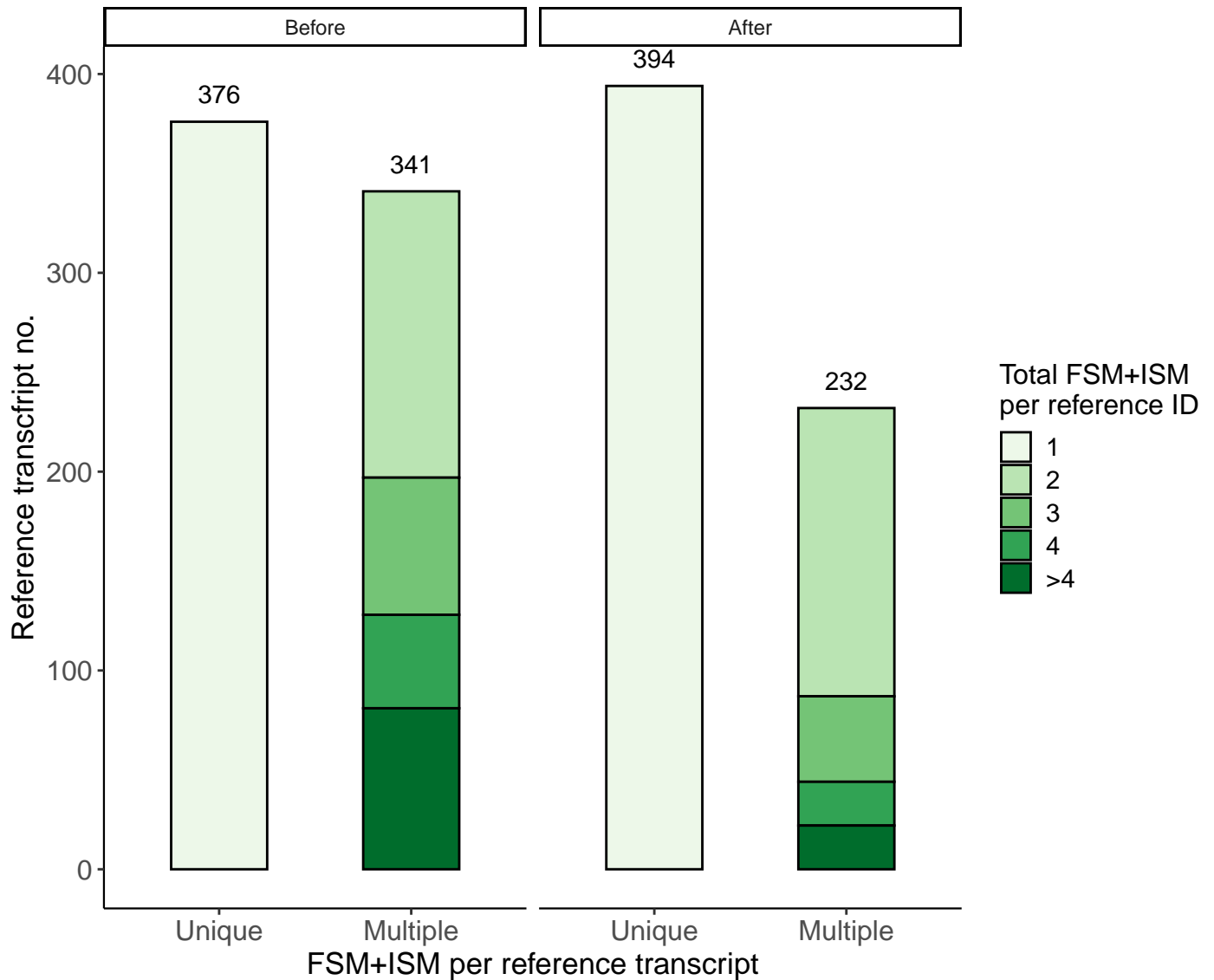
FSM redundancy



ISM redundancy



FSM+ISM redundancy



ML classifier performance report

Classification model performance on test set

Performance metrics

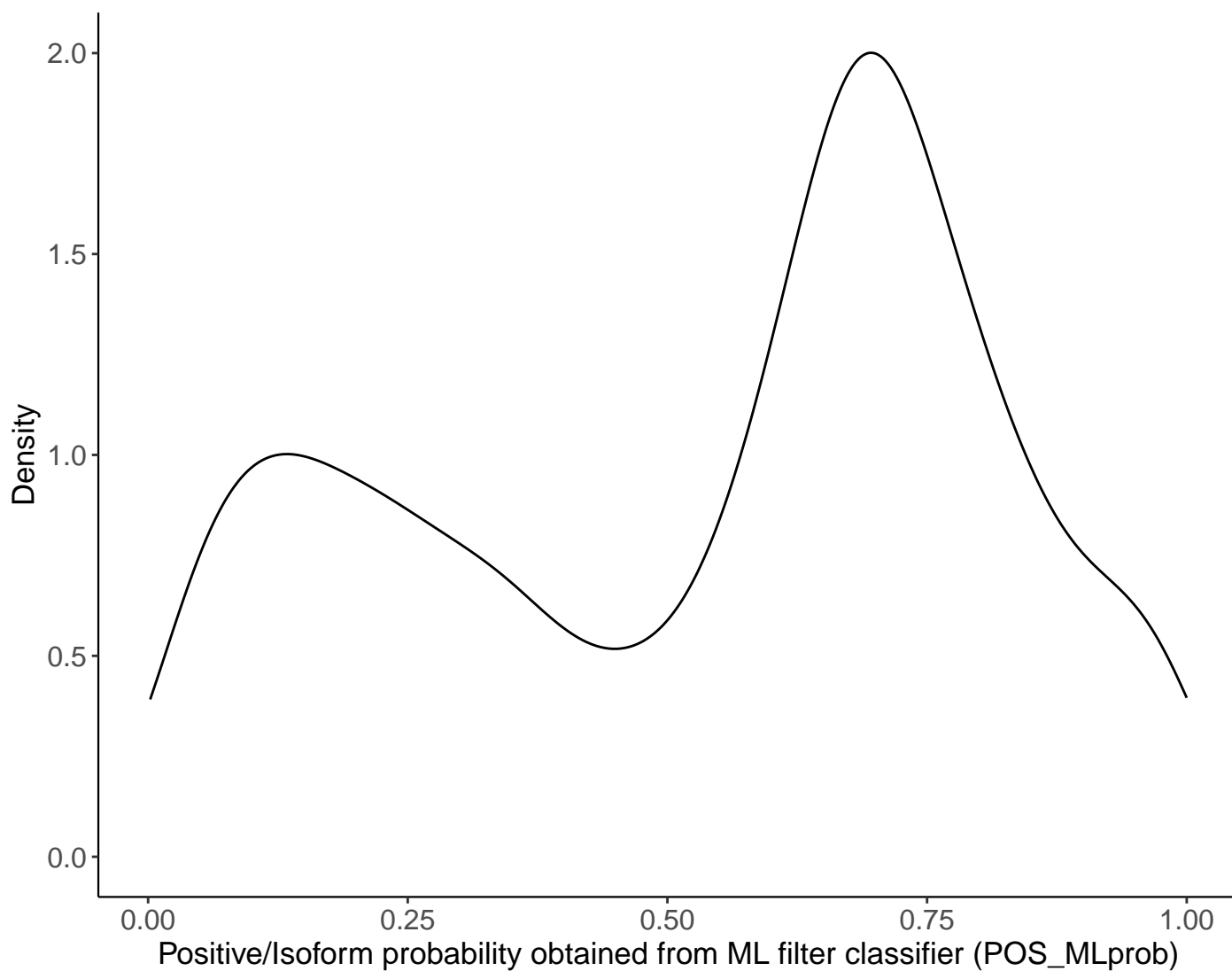
Metric	Value
Accuracy	0.860
Kappa	0.719
AccuracyLower	0.782
AccuracyUpper	0.918
AccuracyNull	0.5
McnemarPValue	0.000 [4m1 [24m [4m7 [24m [4m7 [24m
Sensitivity	0.719
Specificity	1
Pos Pred Value	1
Neg Pred Value	0.781
Precision	1
Recall	0.719
F1	0.837
Prevalence	0.5
Detection Rate	0.360
Detection Prevalence	0.360
balanced Accuracy	0.860

Confusion matrix

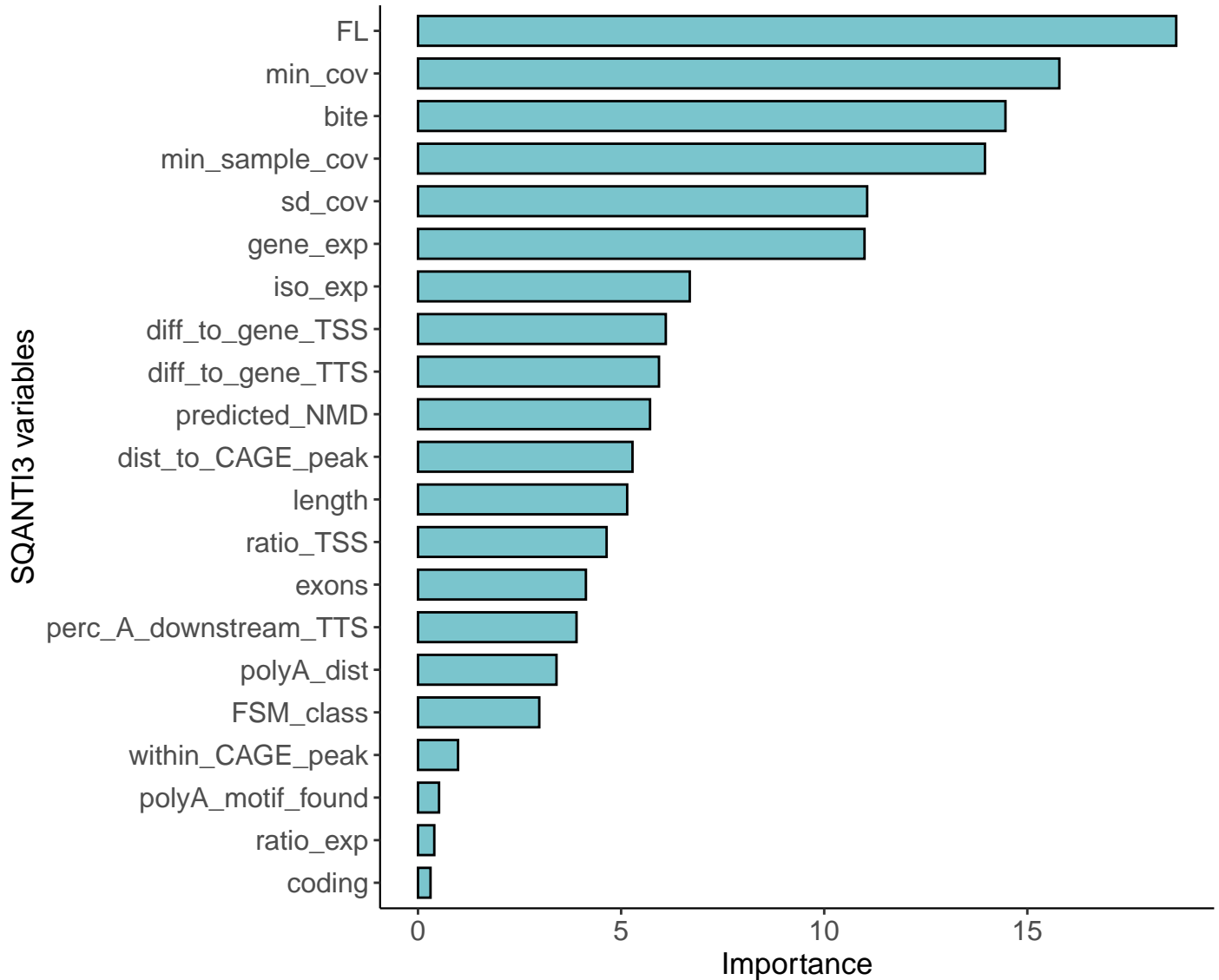
Prediction	Reference	Freq
POS	POS	41
NEG	POS	16
POS	NEG	0
NEG	NEG	57

AccuracyPValue	7.387082e-16
----------------	--------------

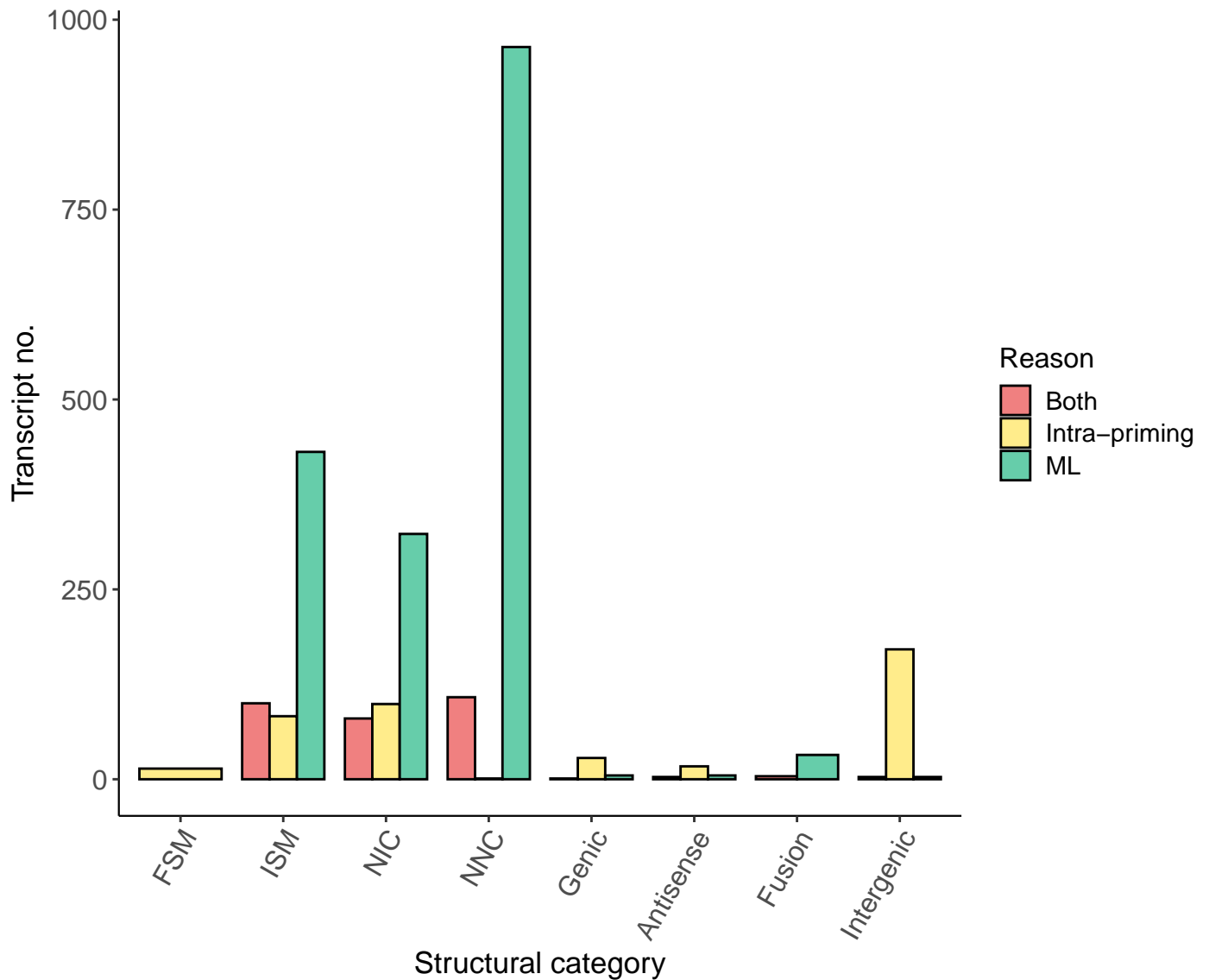
Transcripts classified: 3338



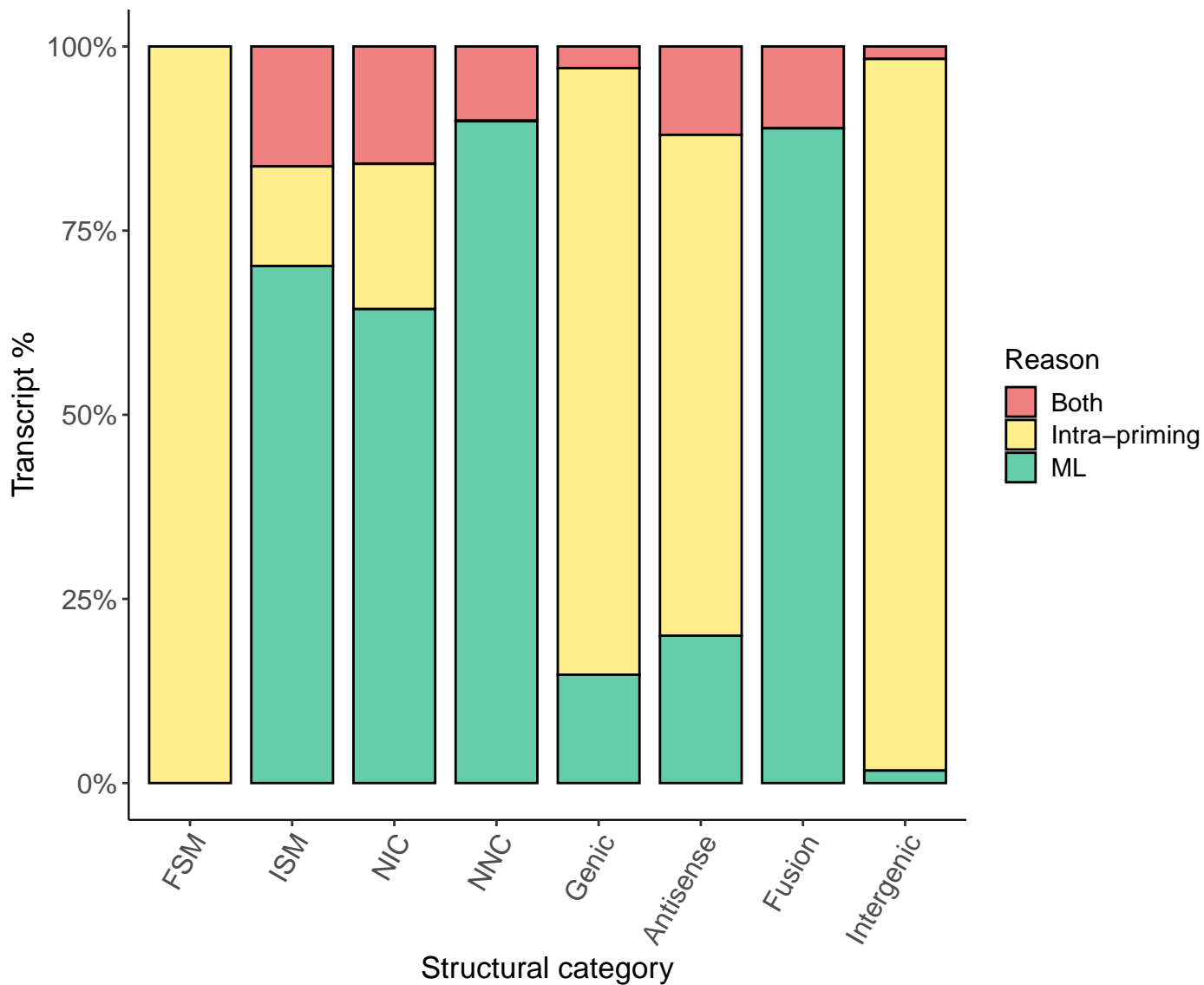
Variable importance in Random Forest classifier



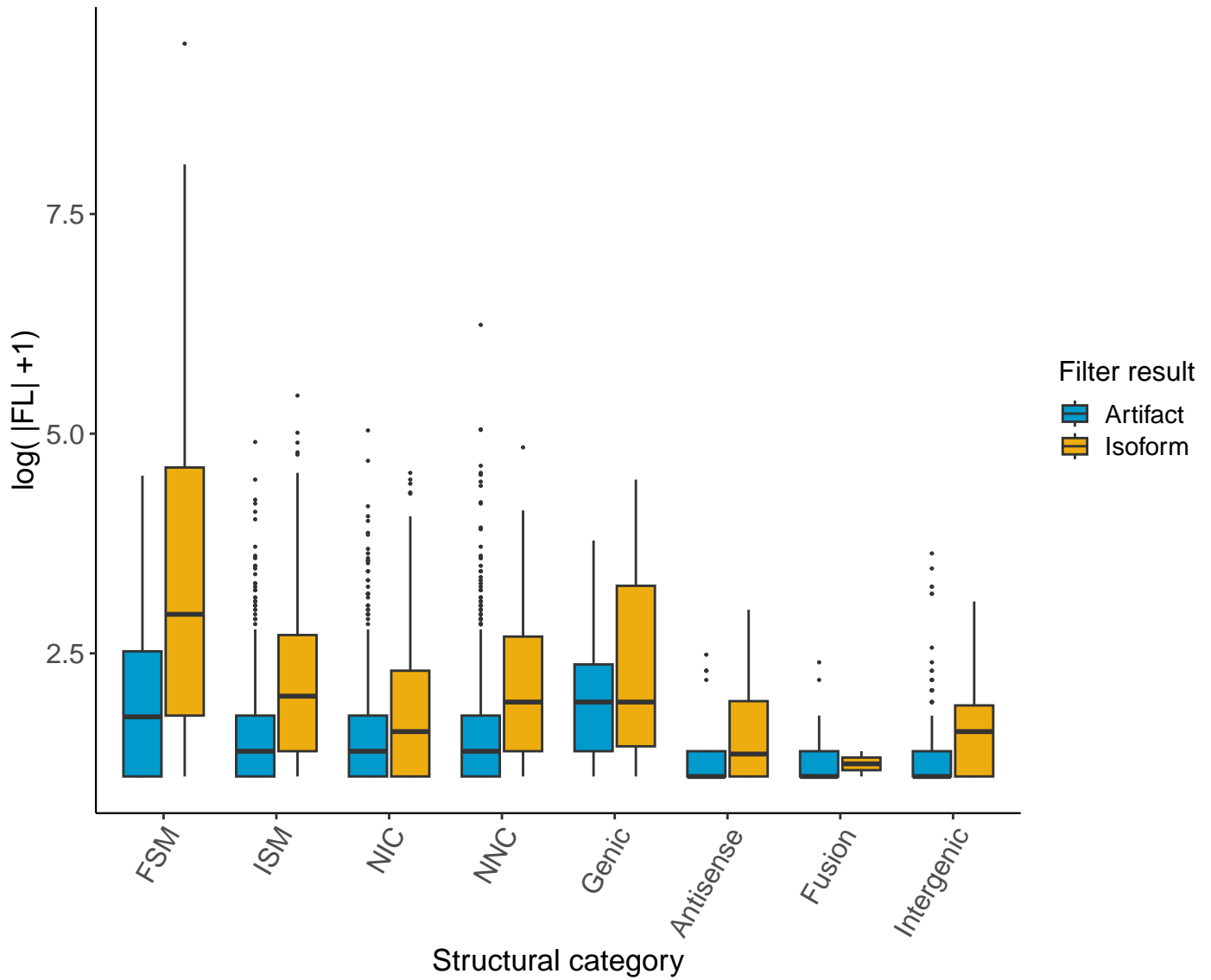
Reason to flag transcripts as artifacts, by category



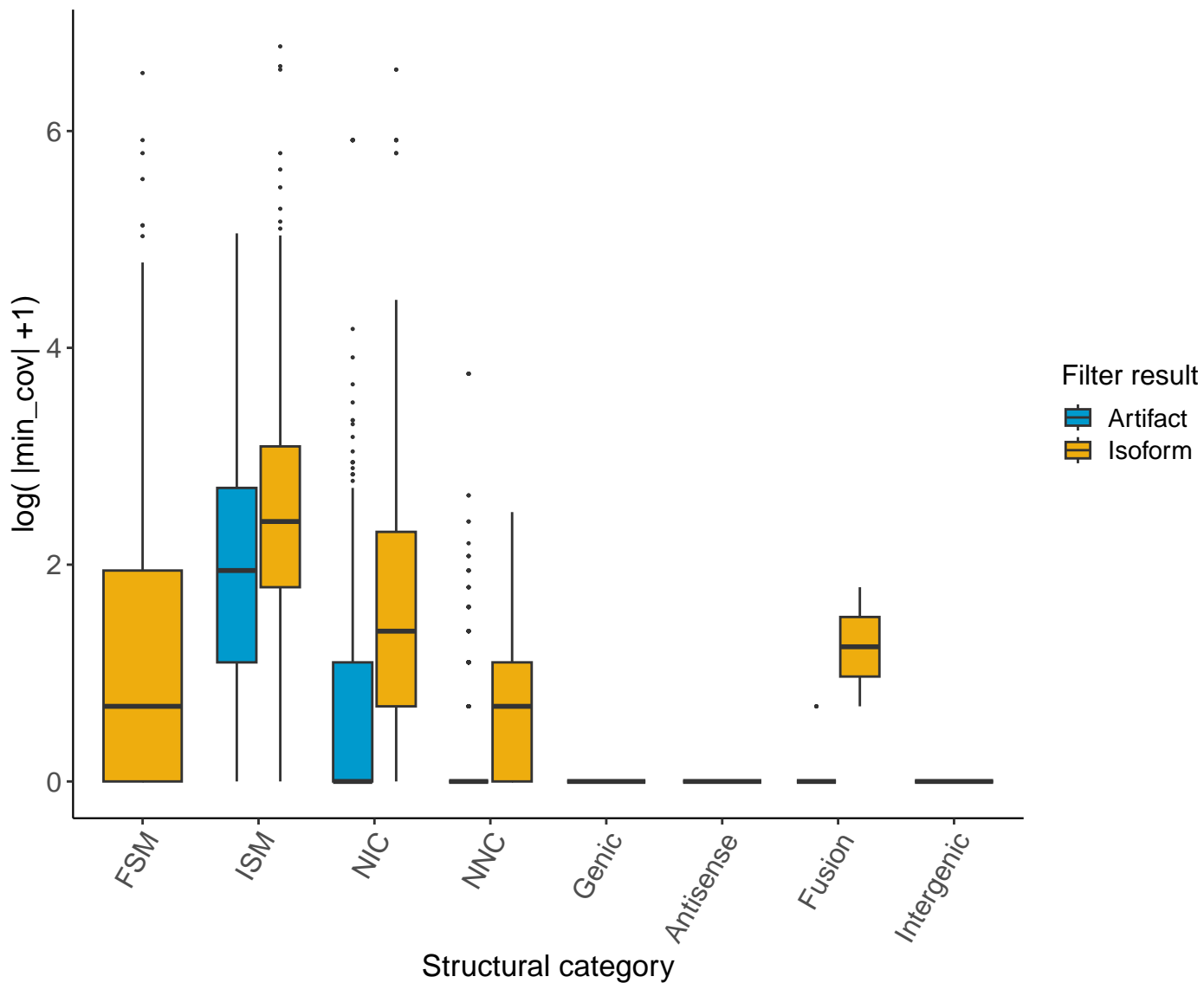
Reason to flag transcripts as artifacts, by category



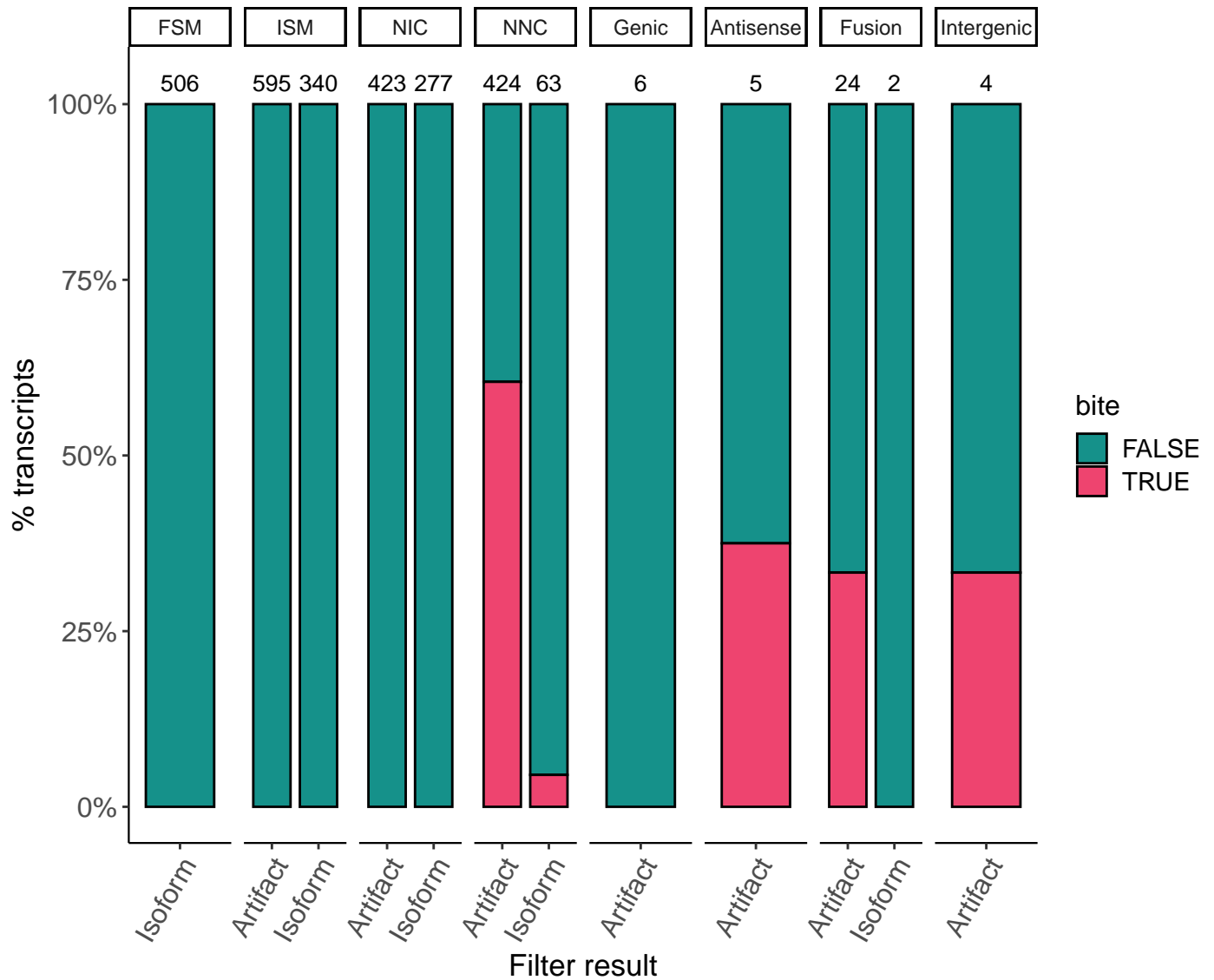
FL – ML importance: 18.67



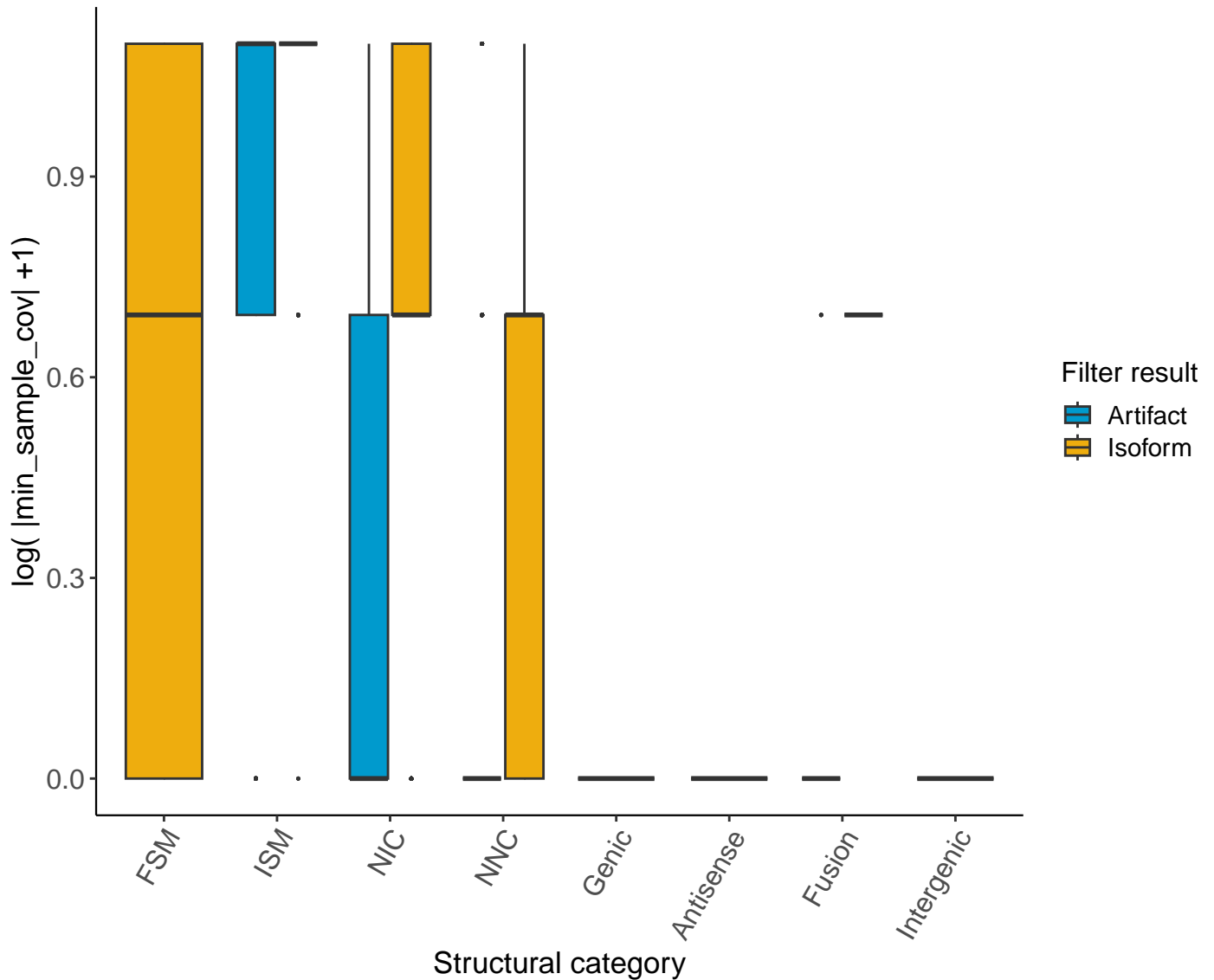
min_cov – ML importance: 15.79



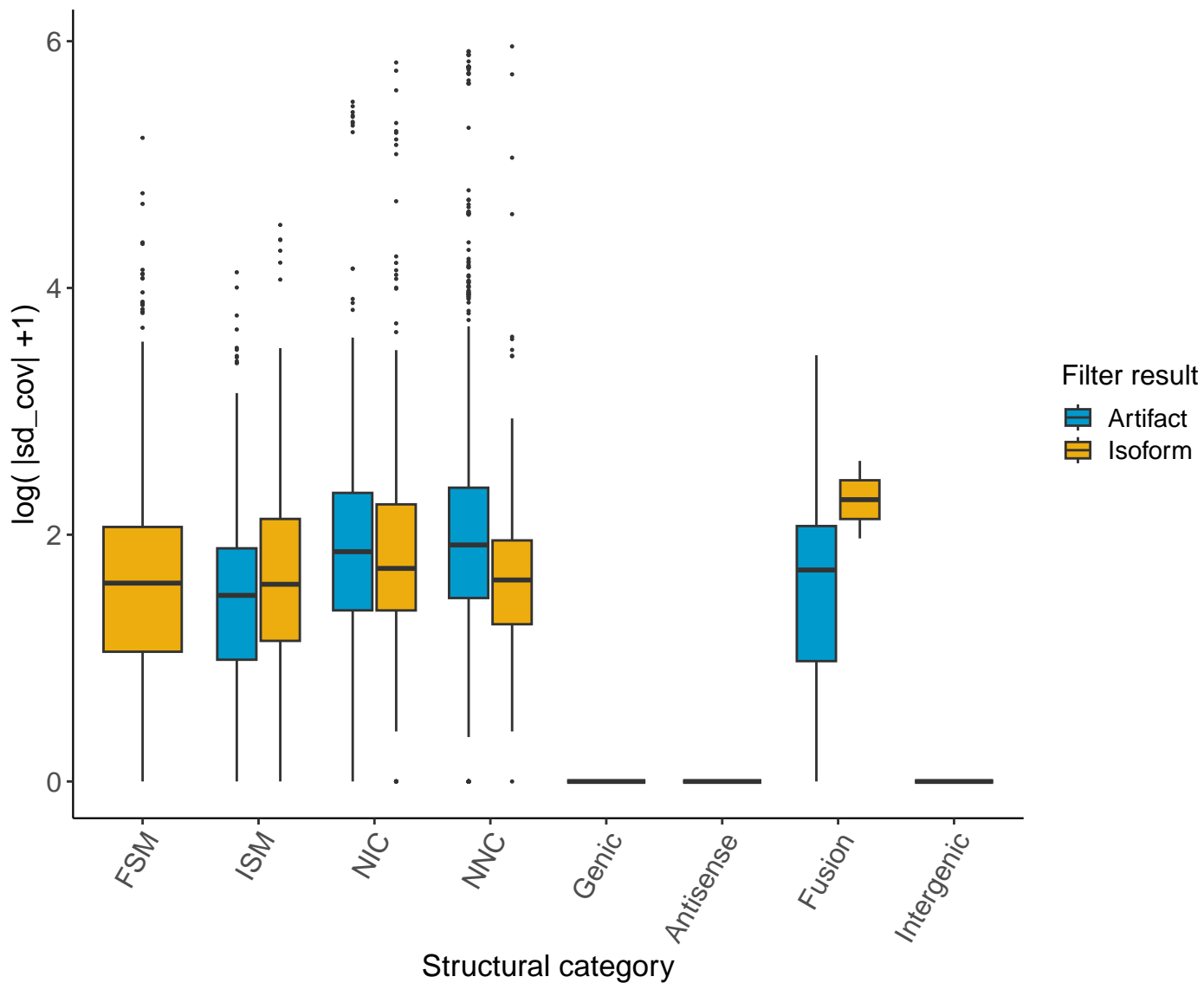
bite – ML importance: 14.46



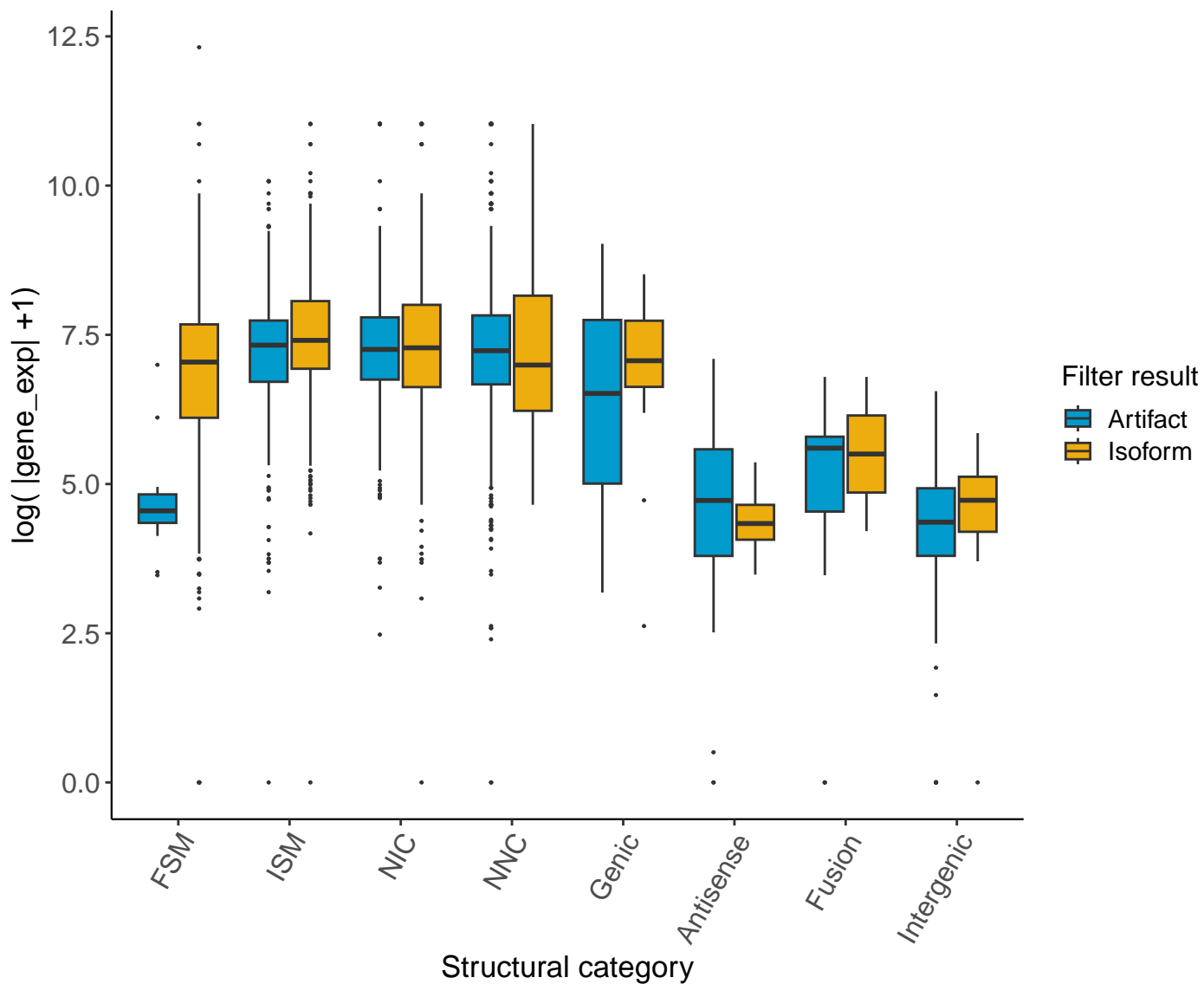
min_sample_cov – ML importance: 13.96



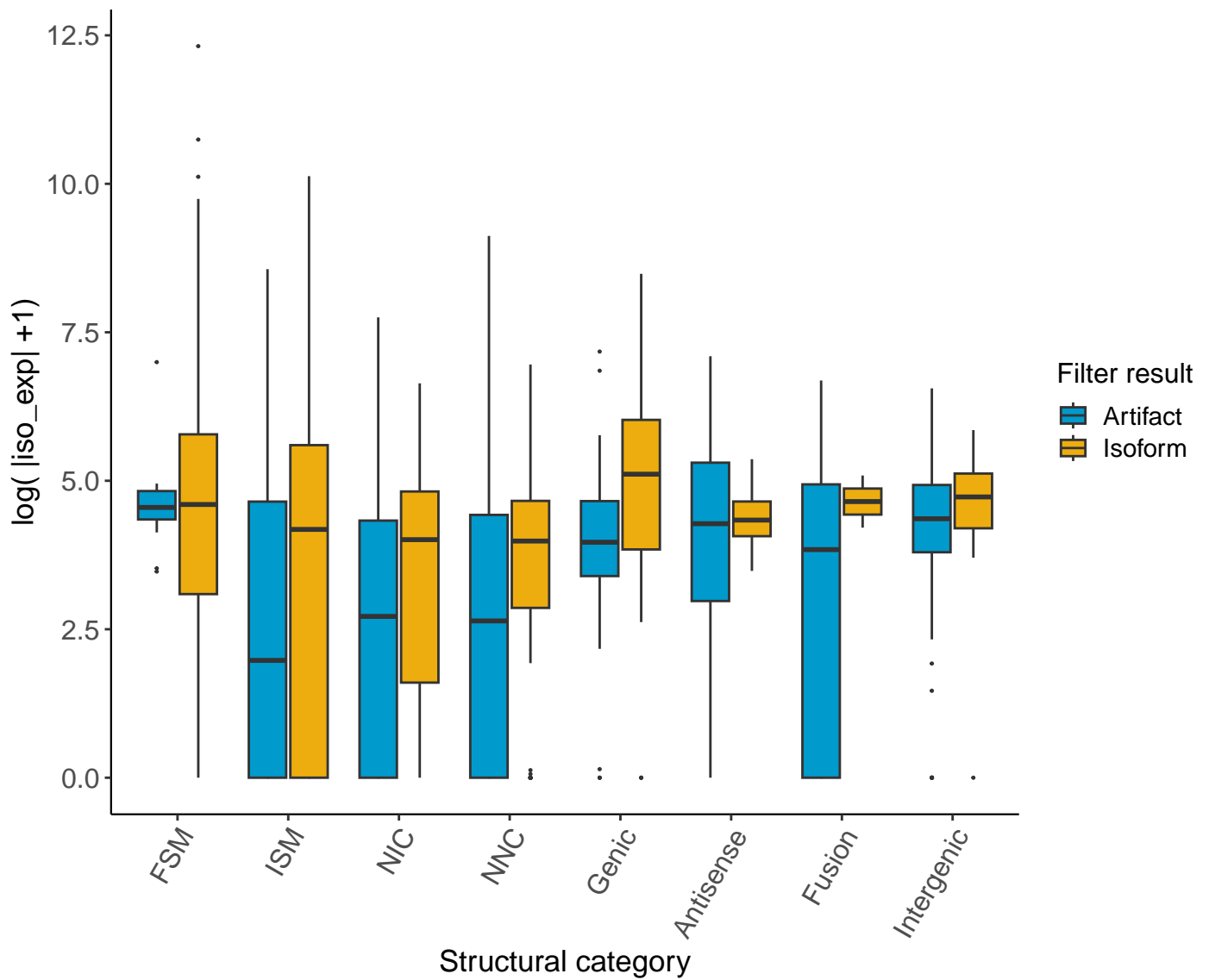
sd_cov – ML importance: 11.06



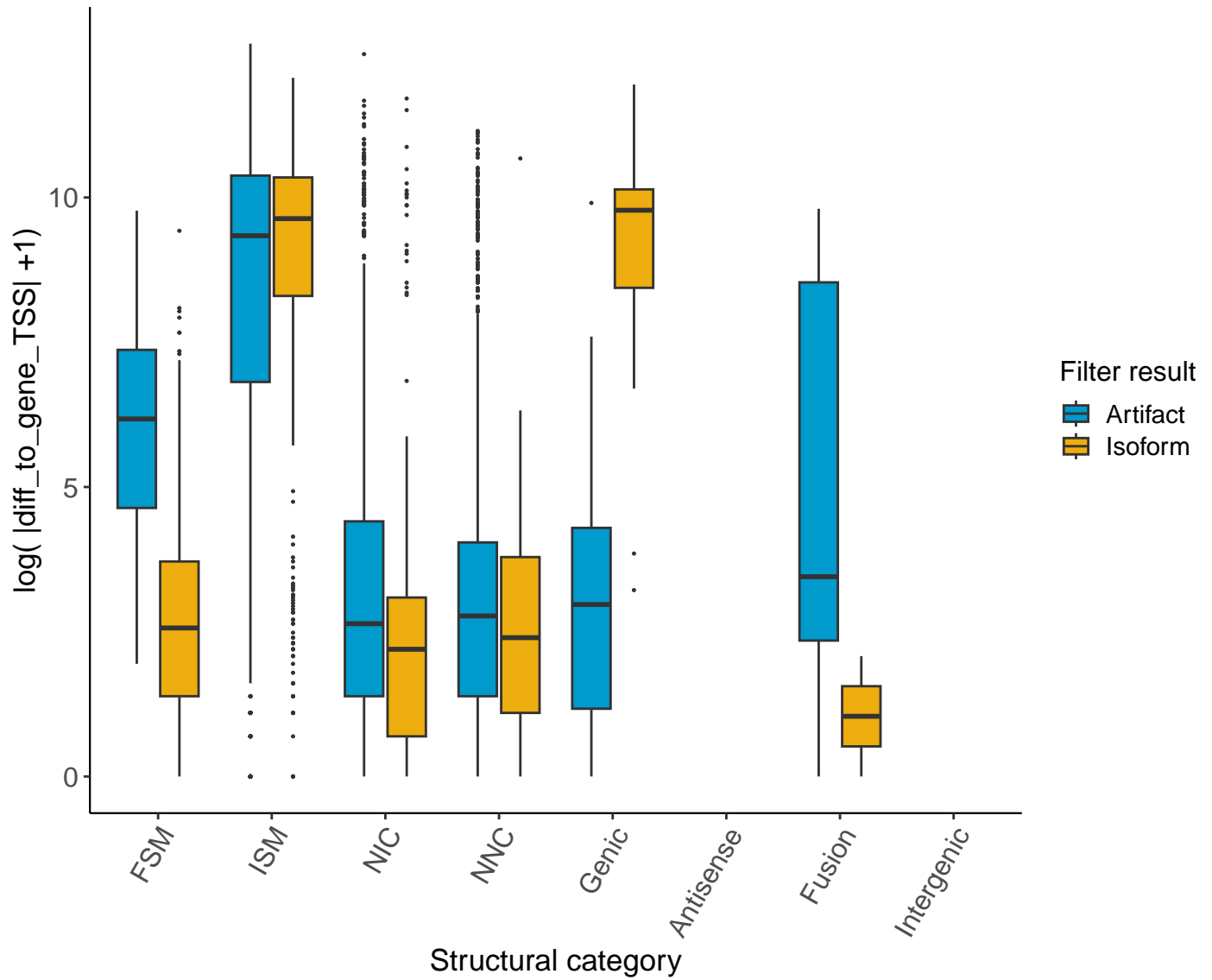
gene_exp – ML importance: 10.99



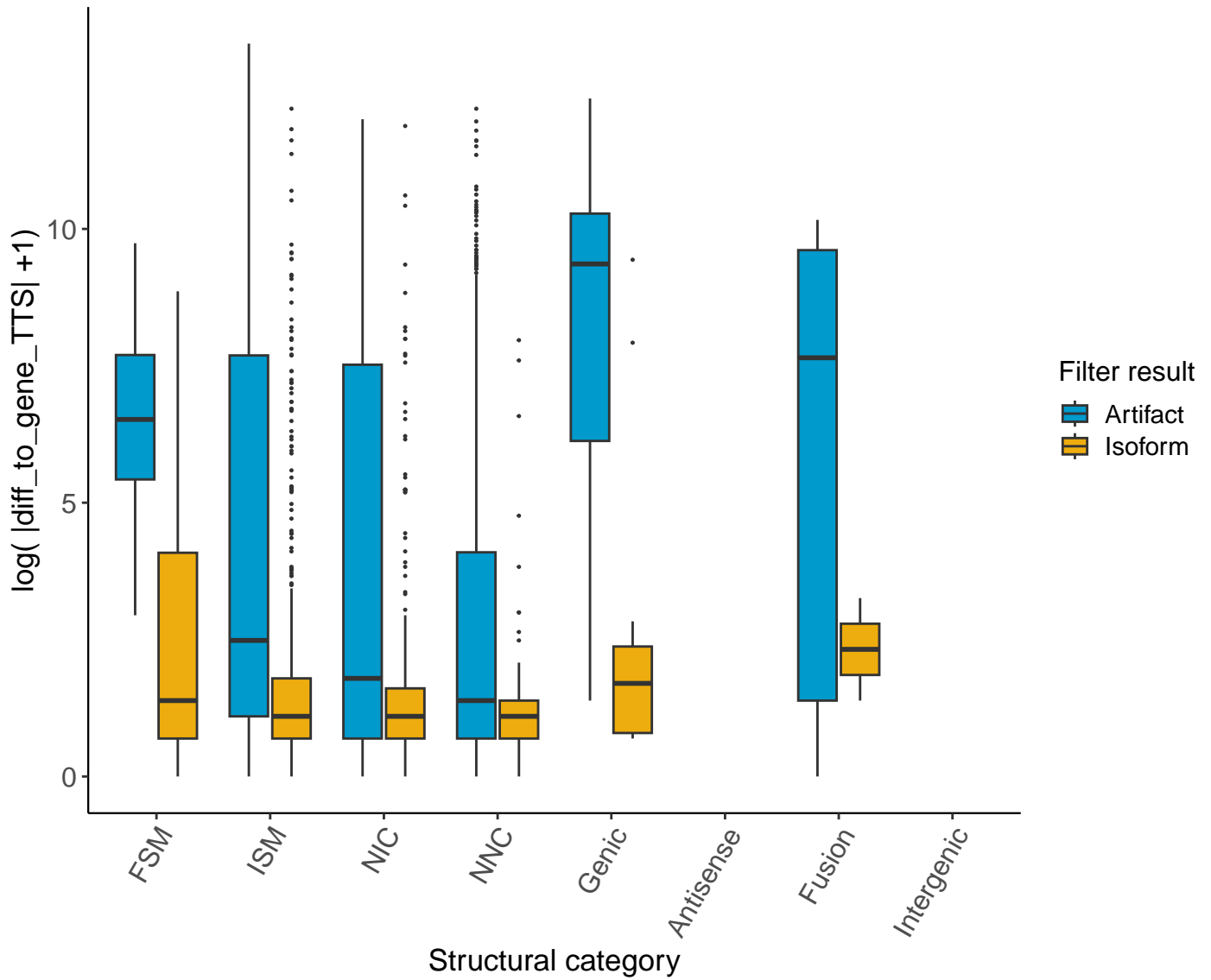
iso_exp – ML importance: 6.69



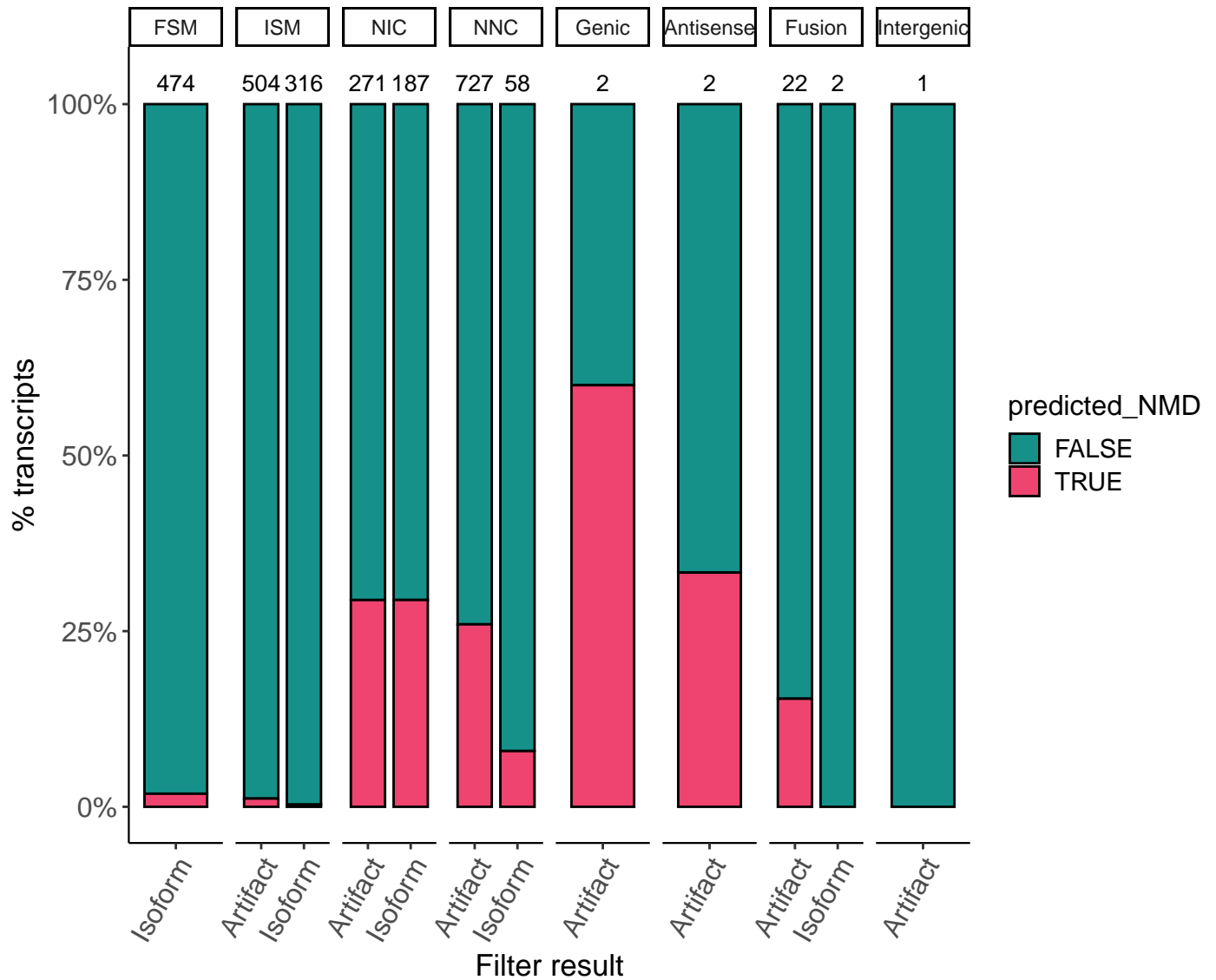
diff_to_gene_TSS – ML importance: 6.1



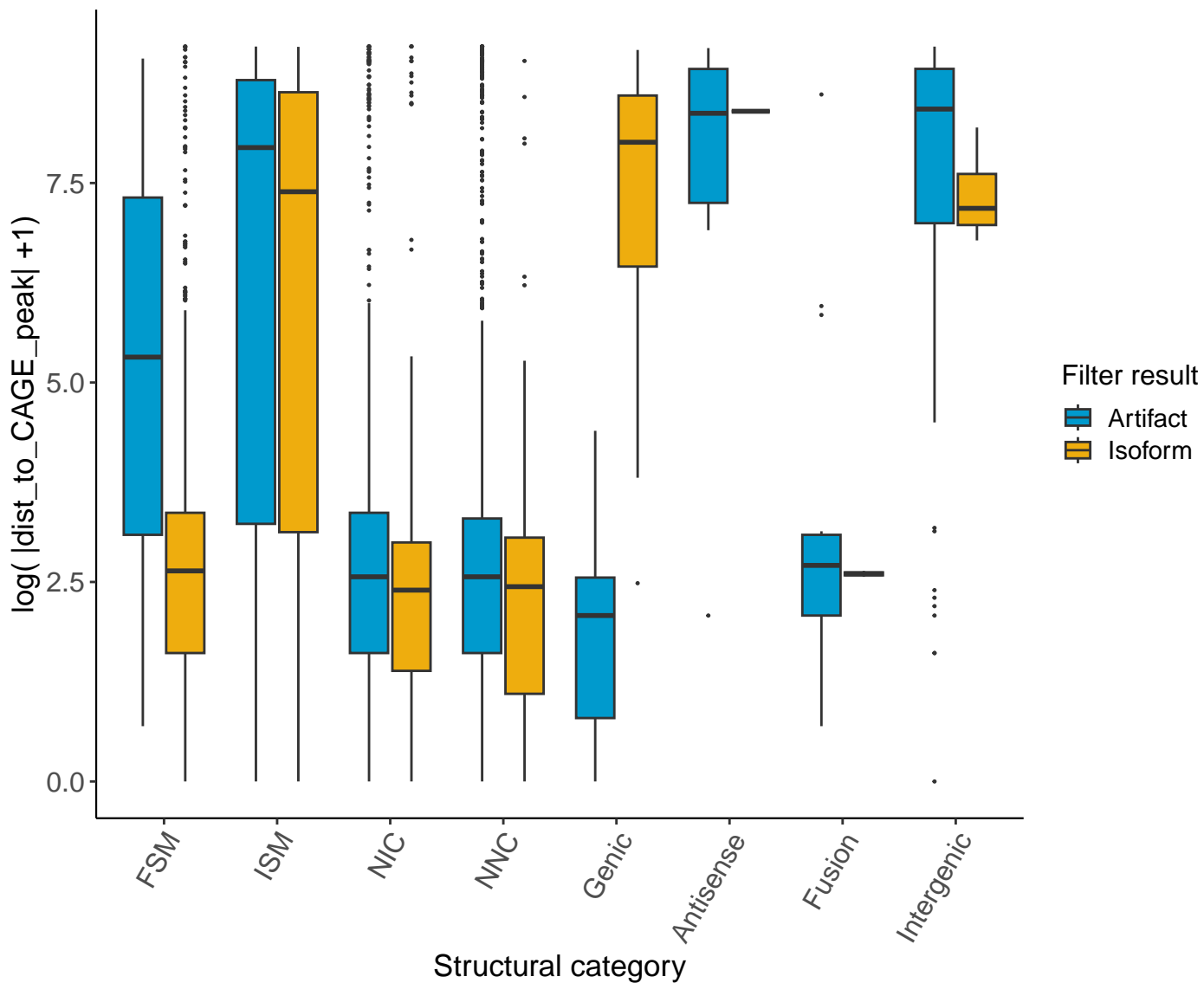
diff_to_gene_TTS – ML importance: 5.93



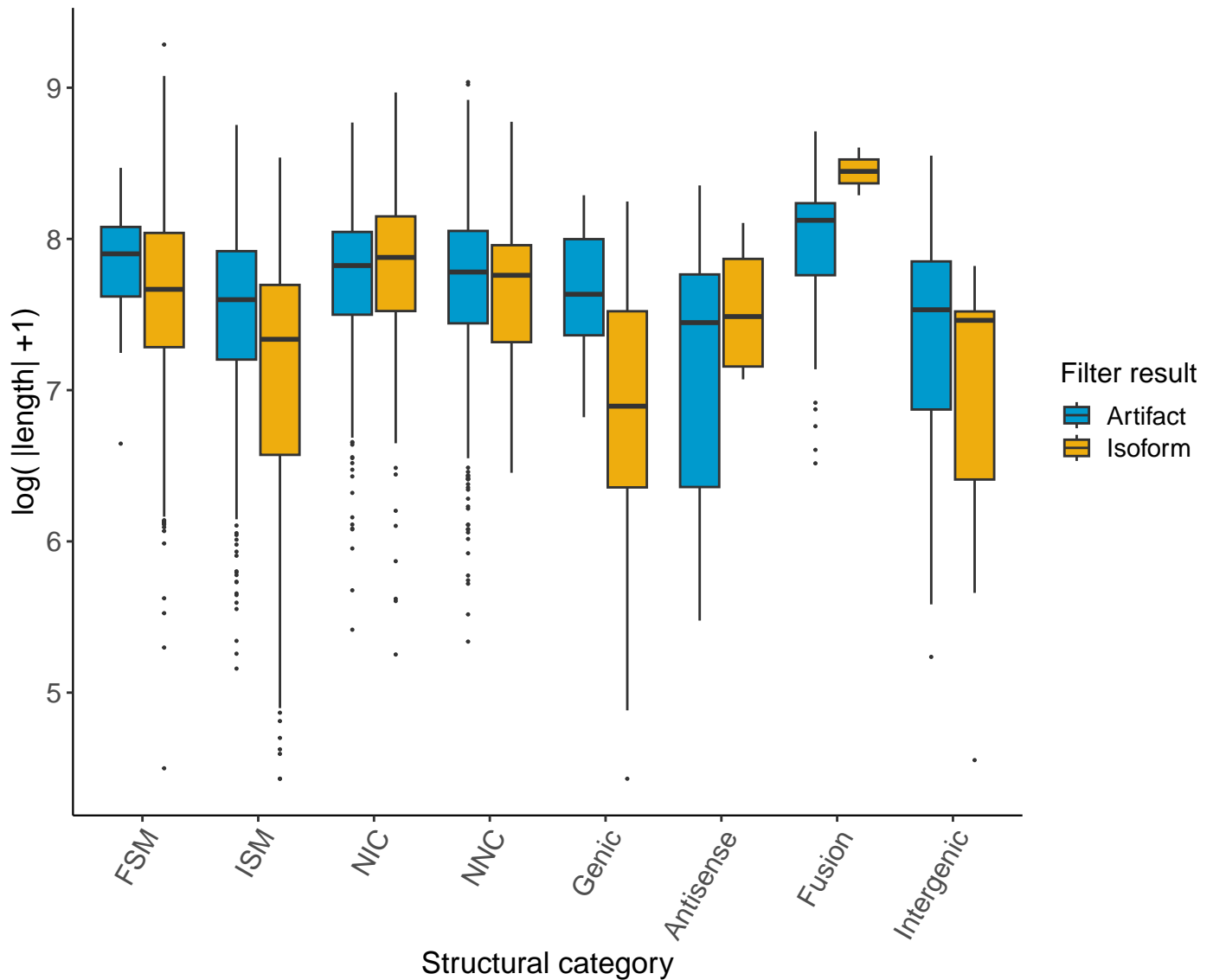
predicted_NMD – ML importance: 5.71



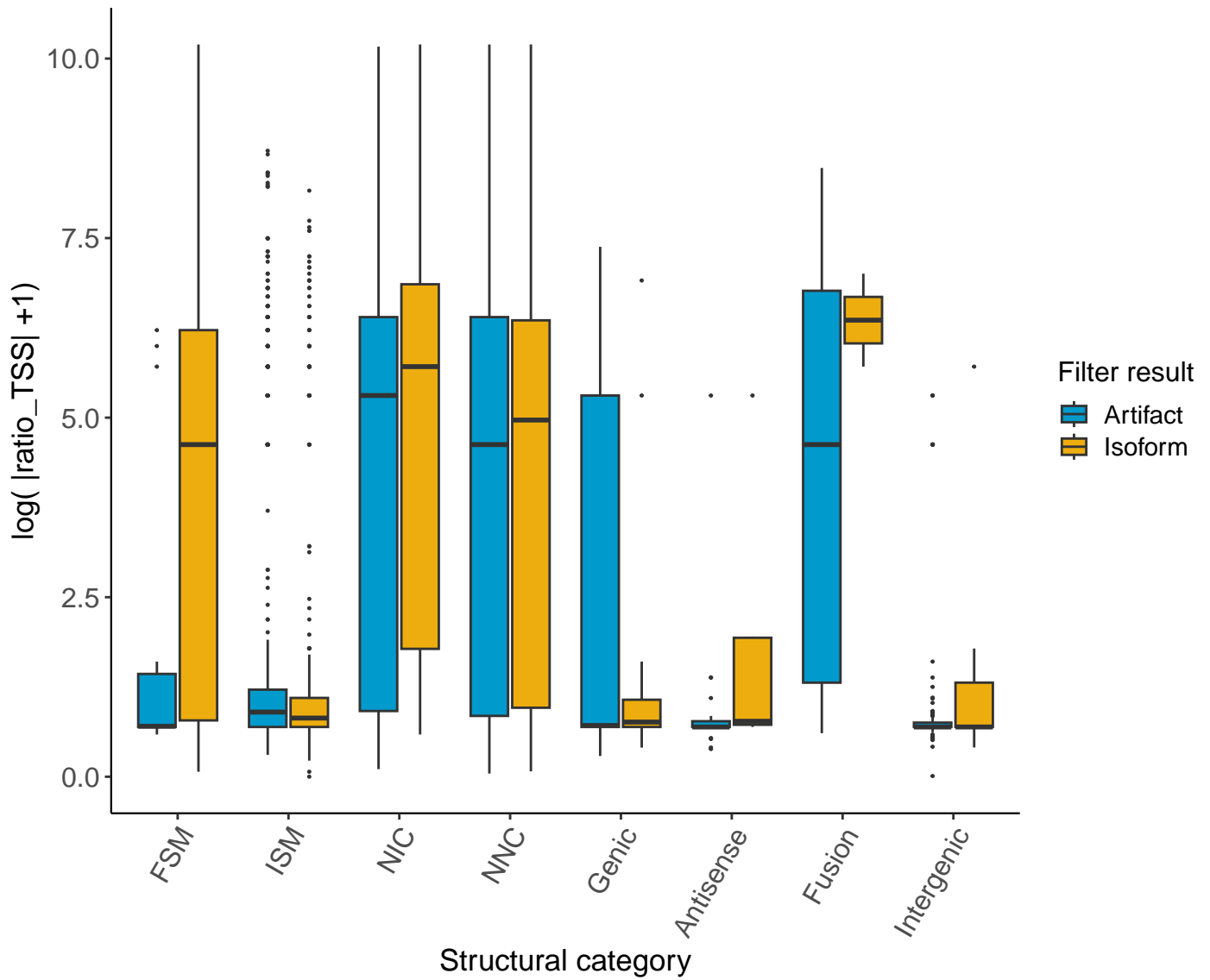
dist_to_CAGE_peak – ML importance: 5.28



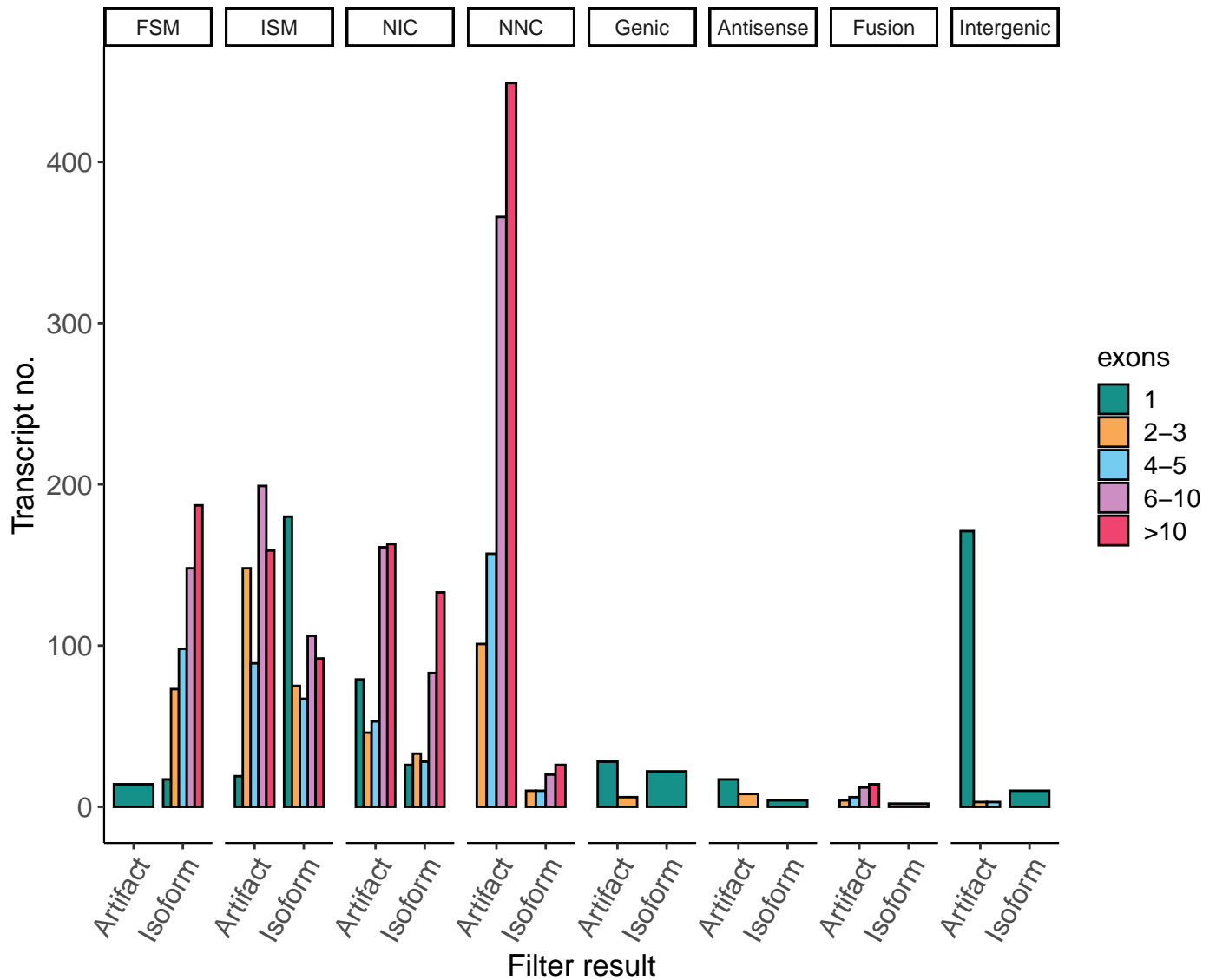
length – ML importance: 5.15



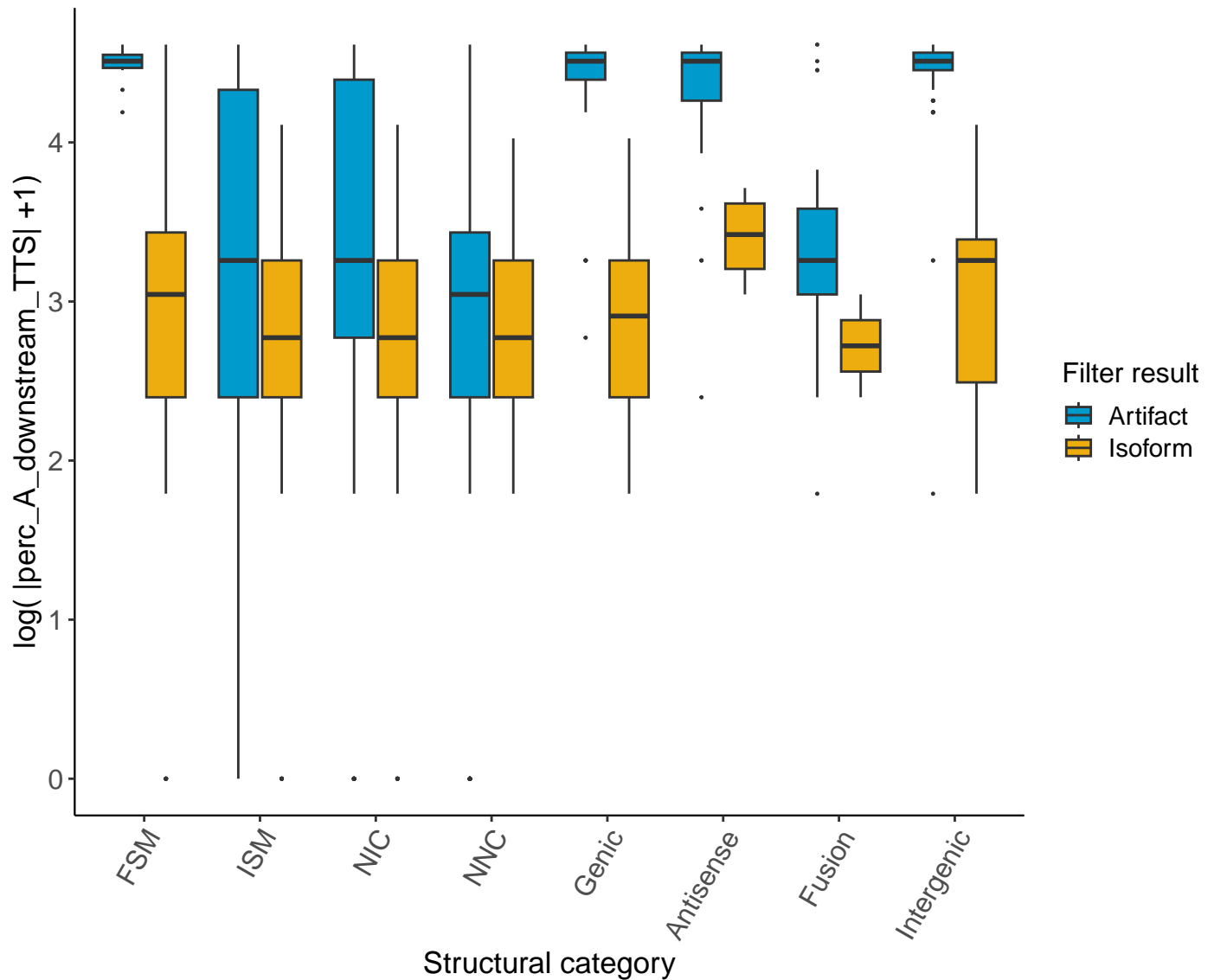
ratio_TSS – ML importance: 4.64



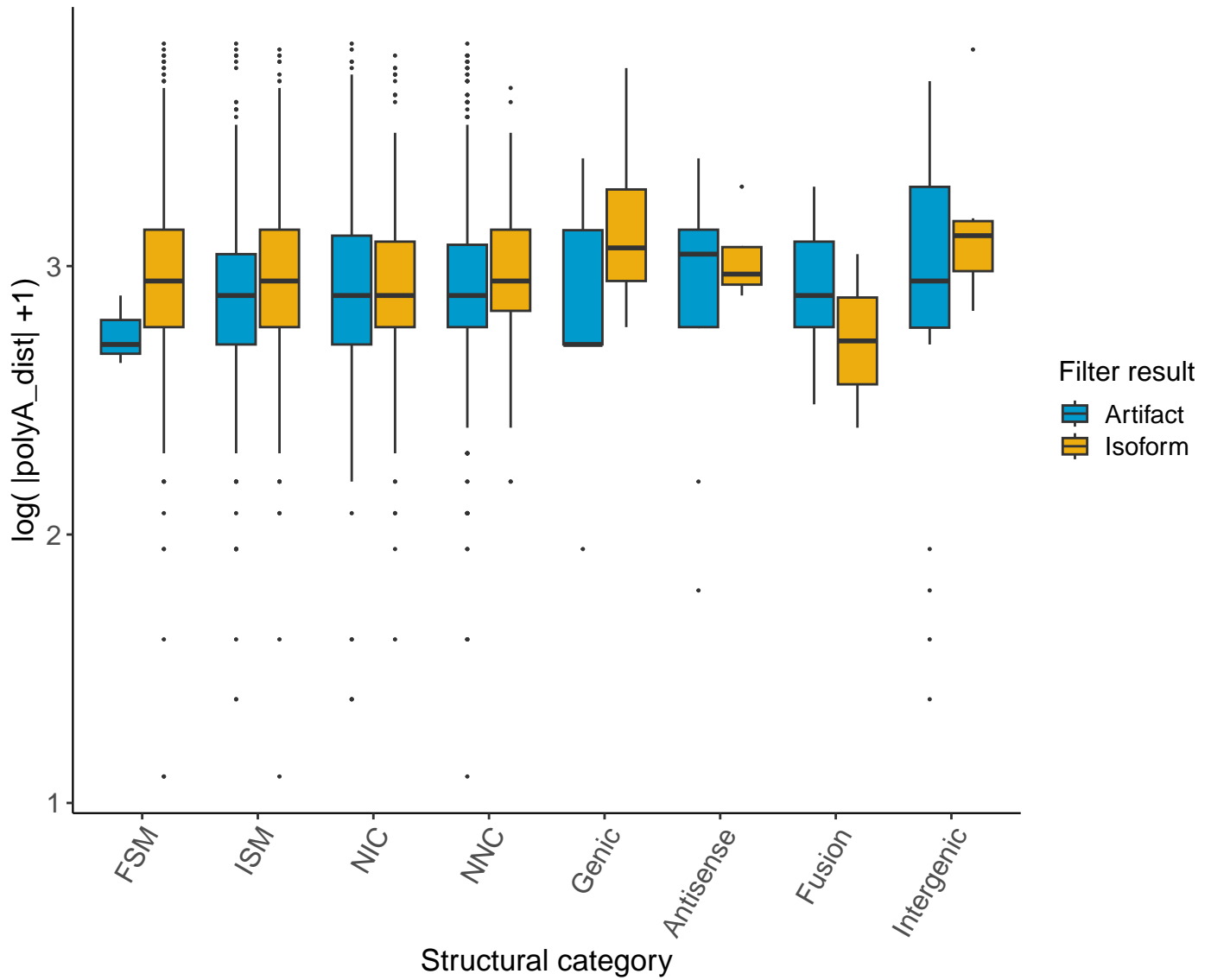
exons – ML importance: 4.13



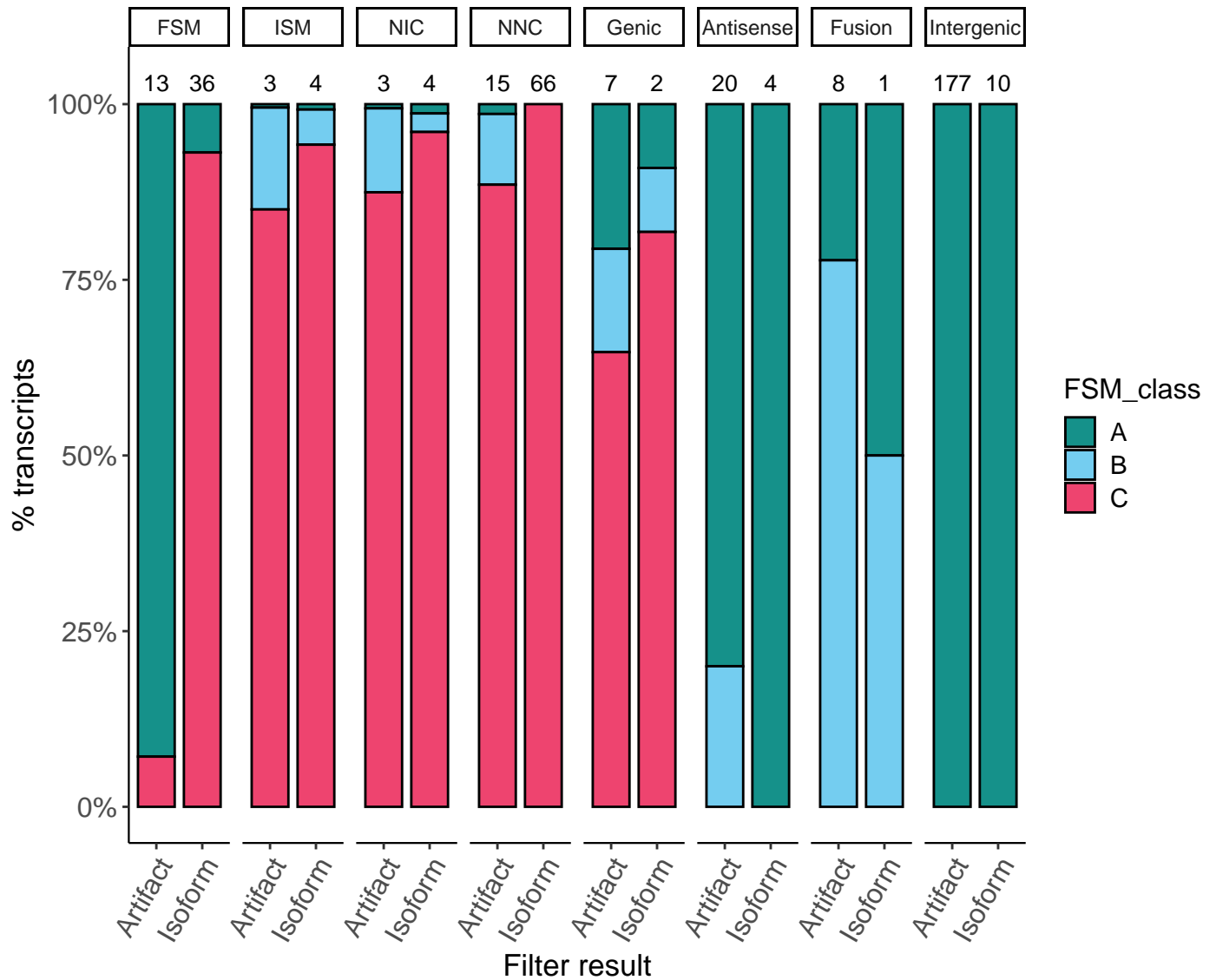
perc_A_downstream_TTS – ML importance: 3.9



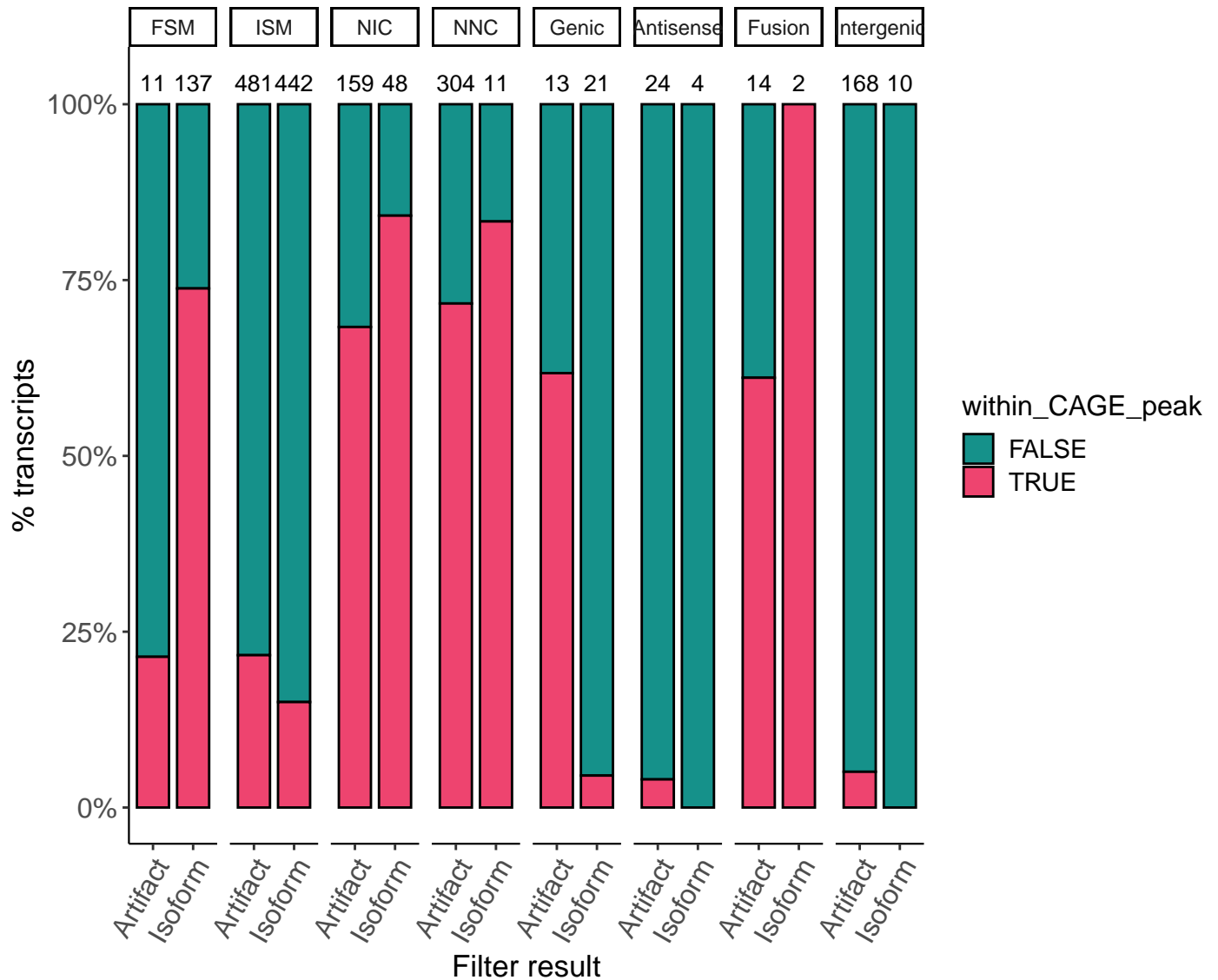
polyA_dist – ML importance: 3.41



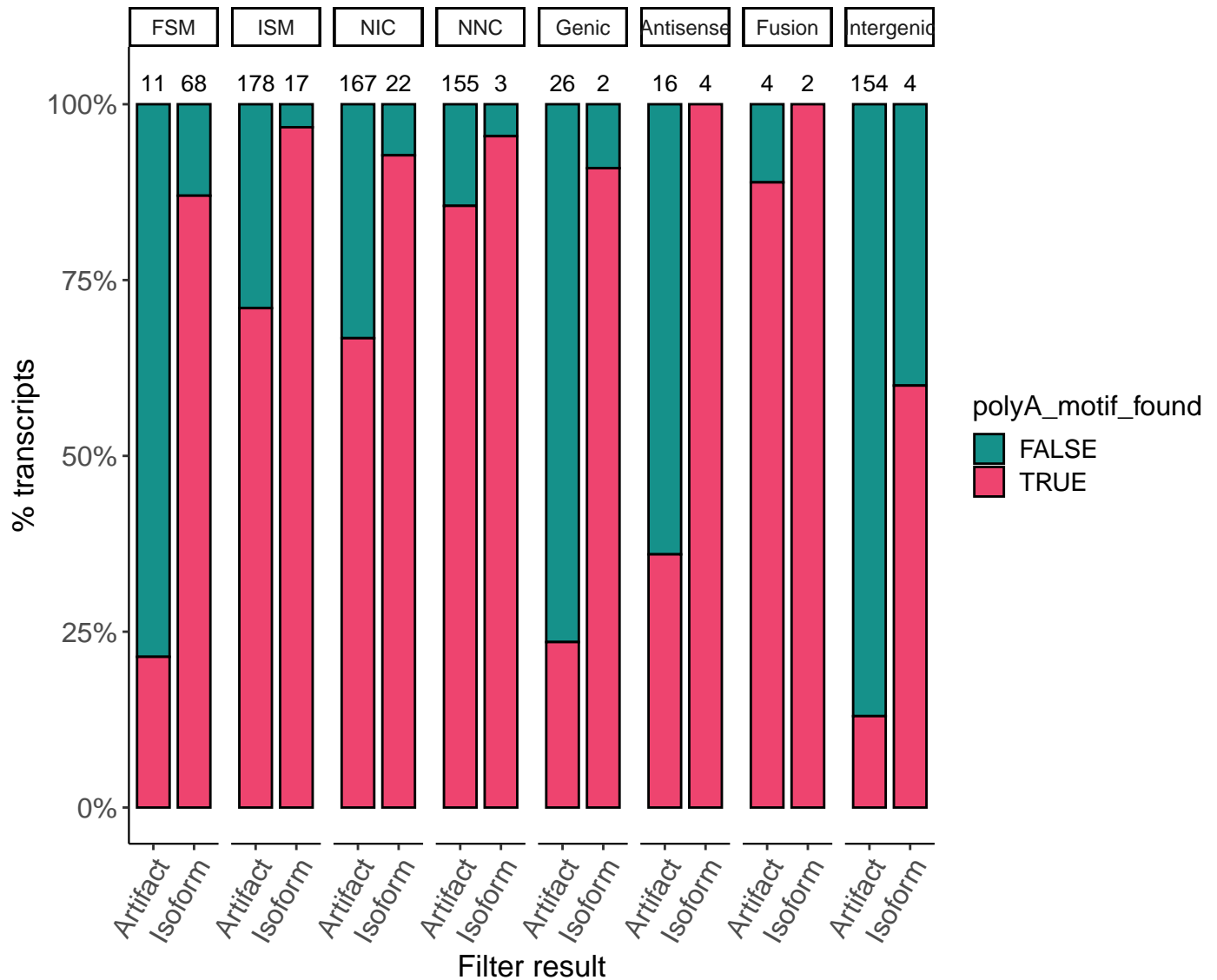
FSM_class – ML importance: 2.99



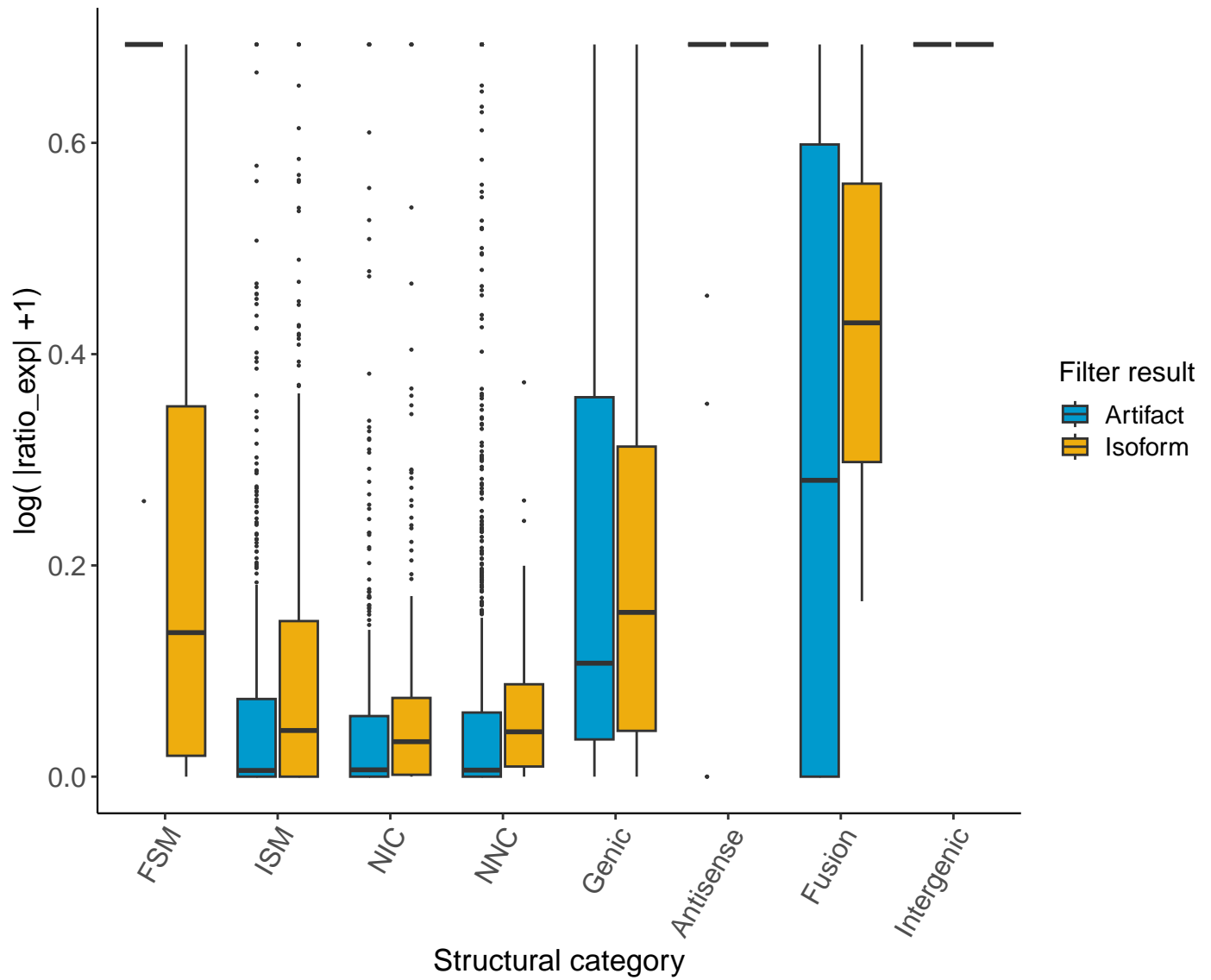
within_CAGE_peak – ML importance: 0.99



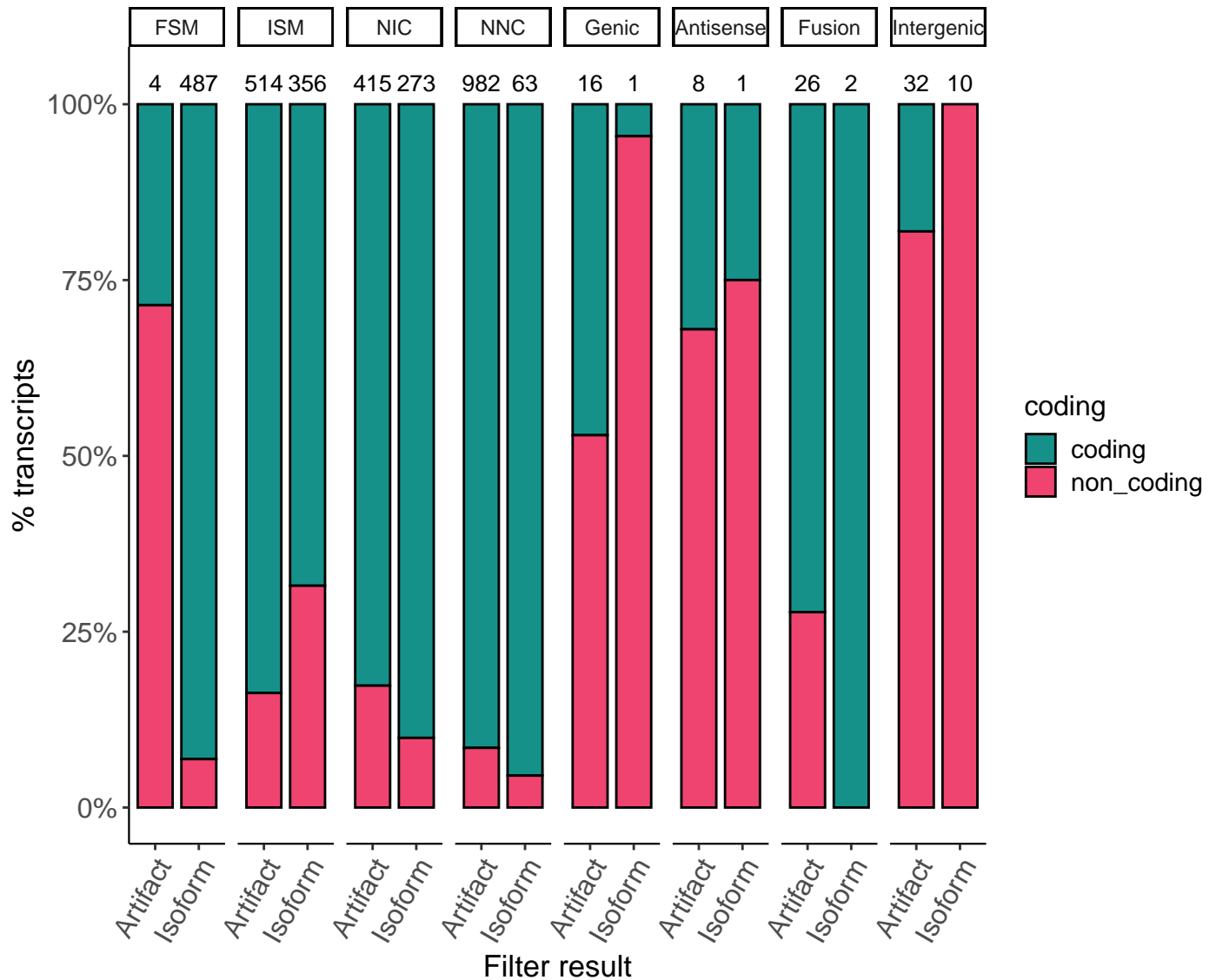
polyA_motif_found – ML importance: 0.52



ratio_exp – ML importance: 0.4



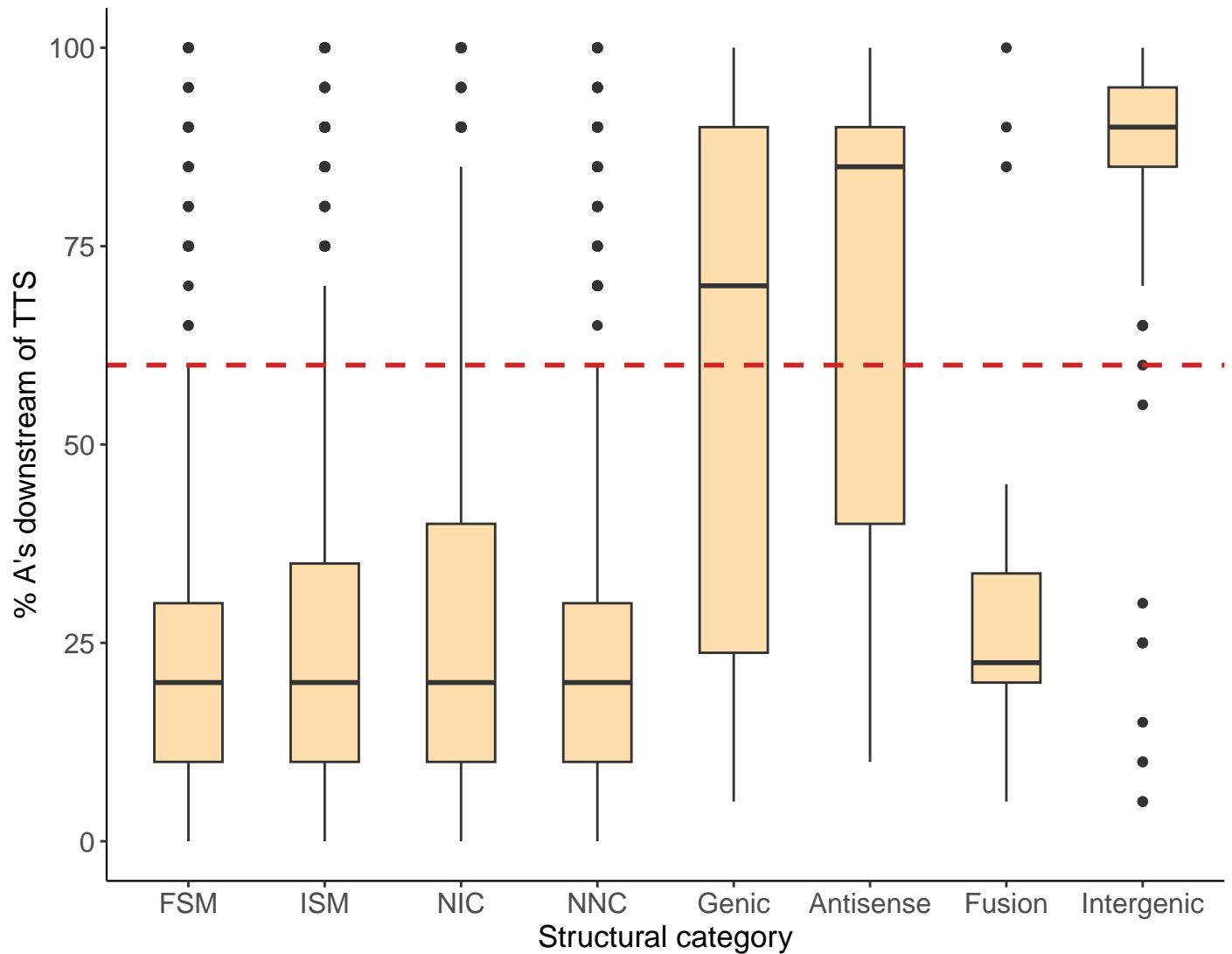
coding – ML importance: 0.31



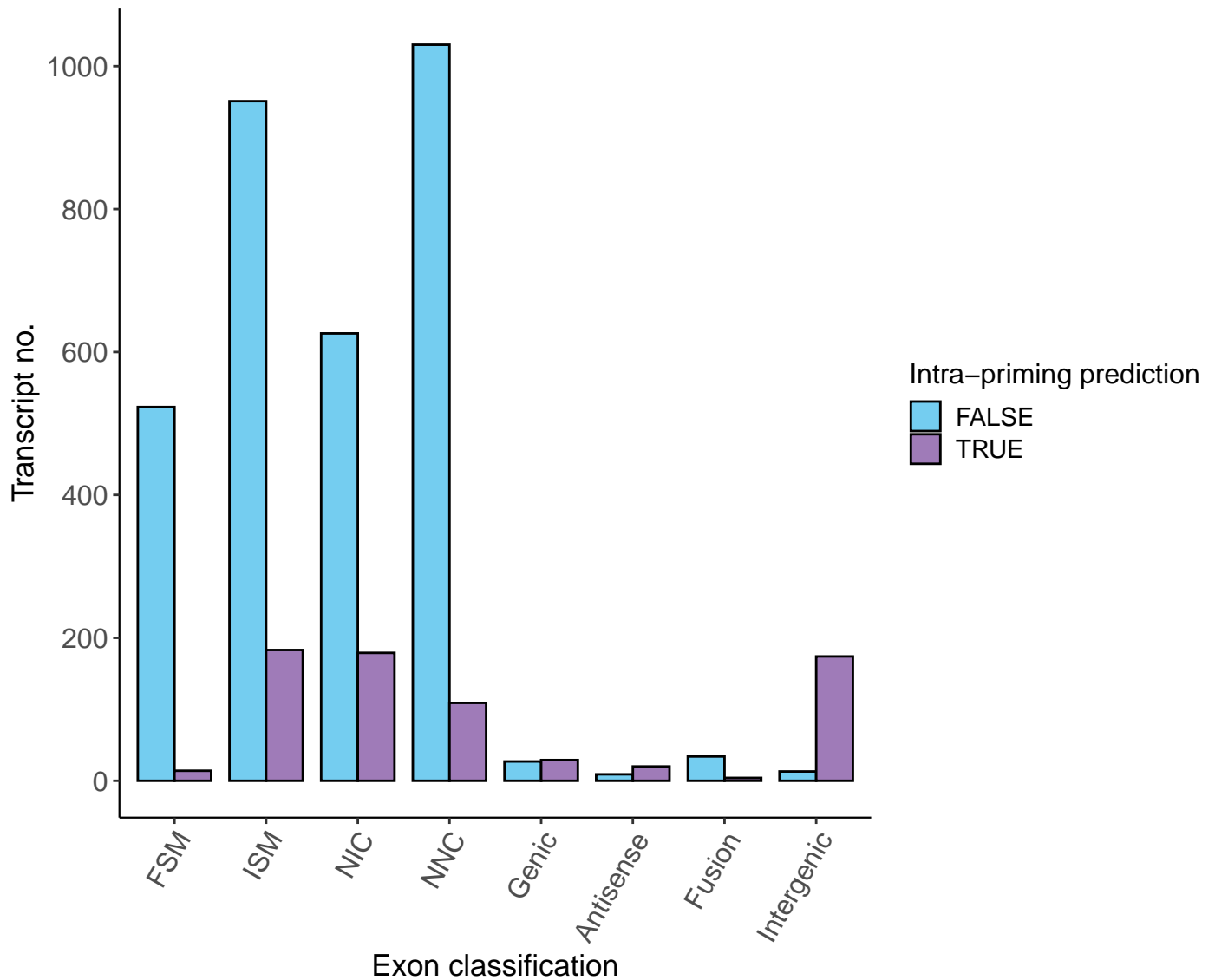
Intra-primming filter report

A % by category

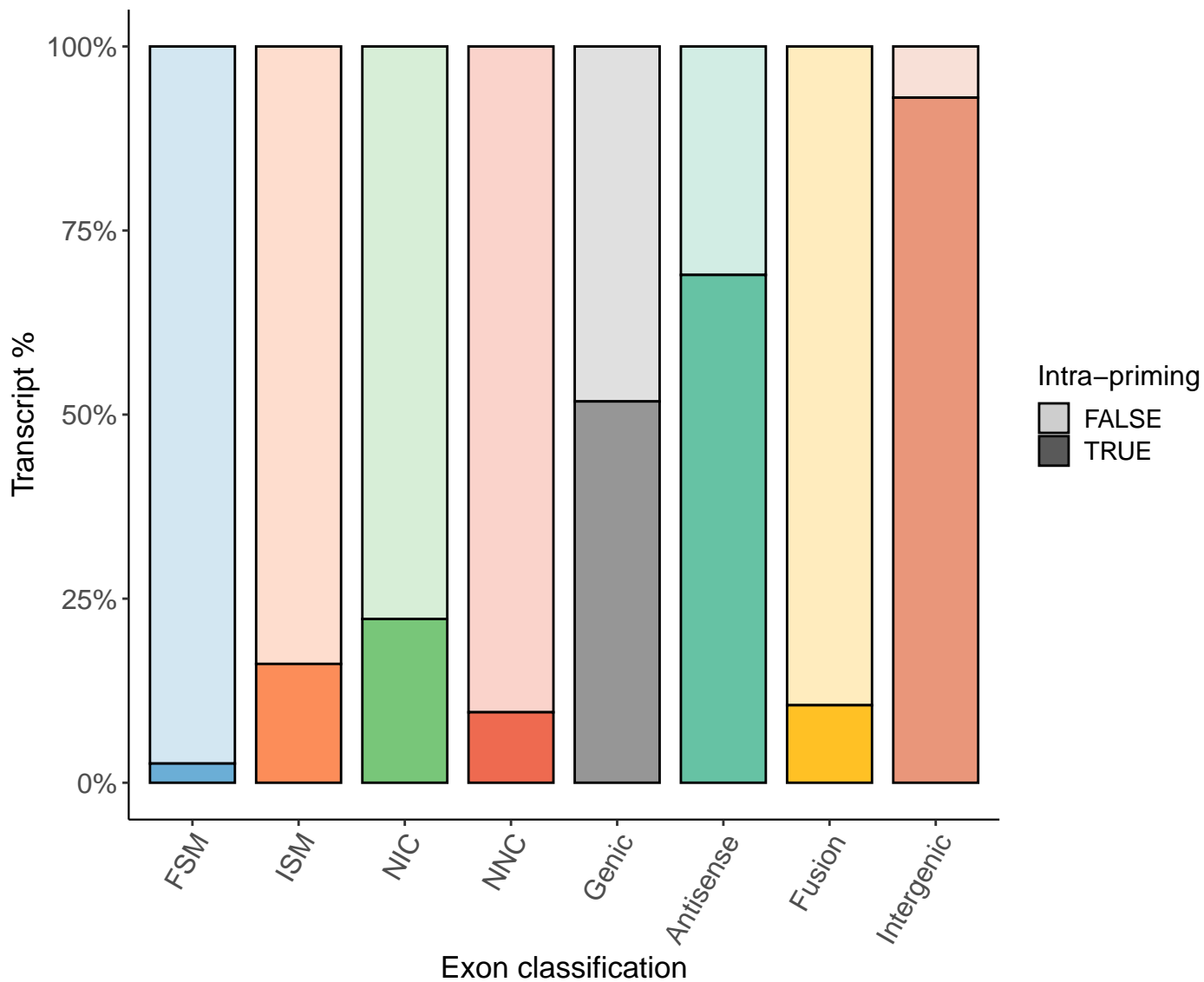
Red line indicates threshold employed in ML filter



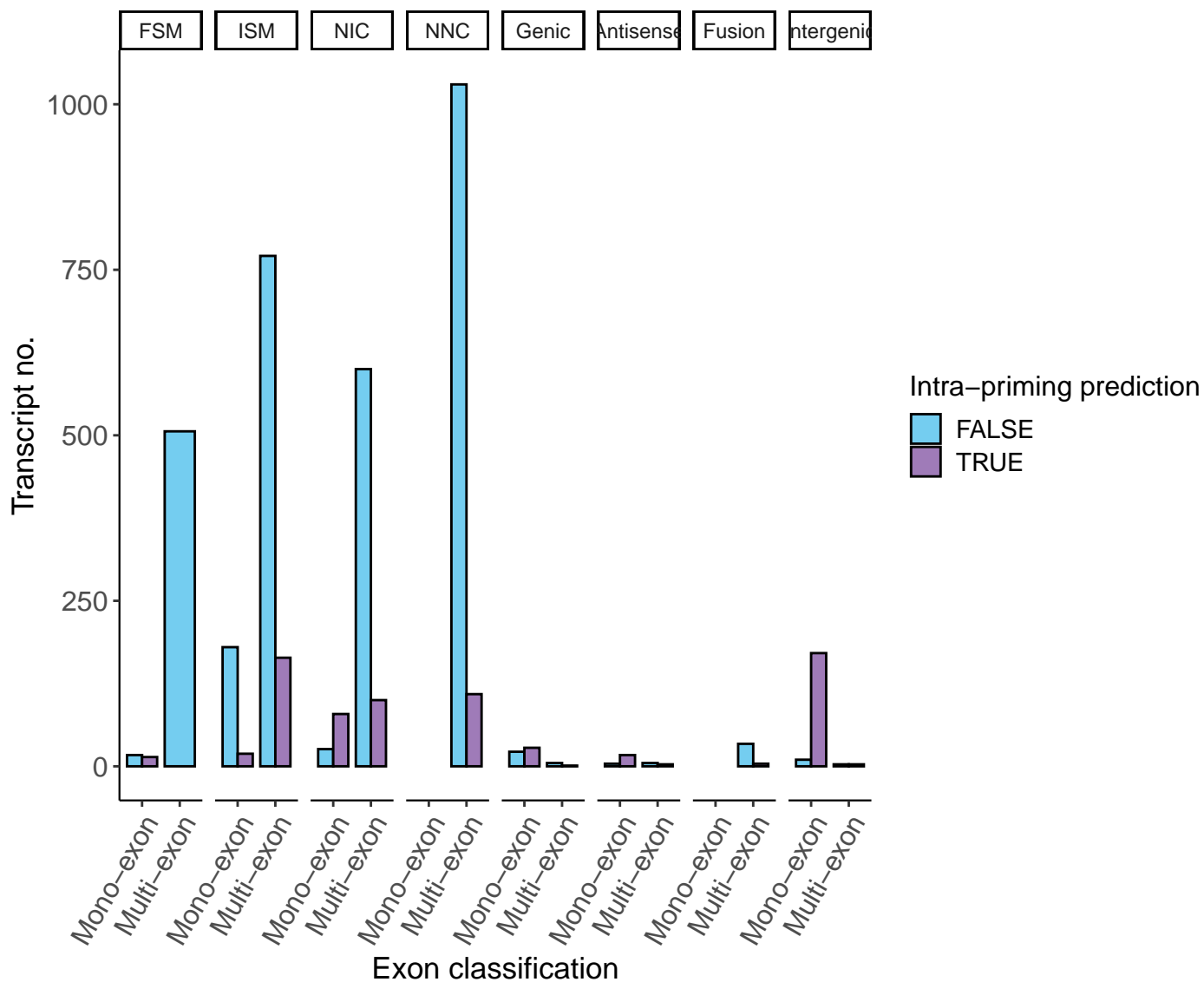
Isoforms flagged as intra-priming, by category



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



Isoforms flagged as intra-priming, by exon number



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

