

# *SQANTI3 filter report*

*Total Genes: 656*

*Total Transcripts: 3925*

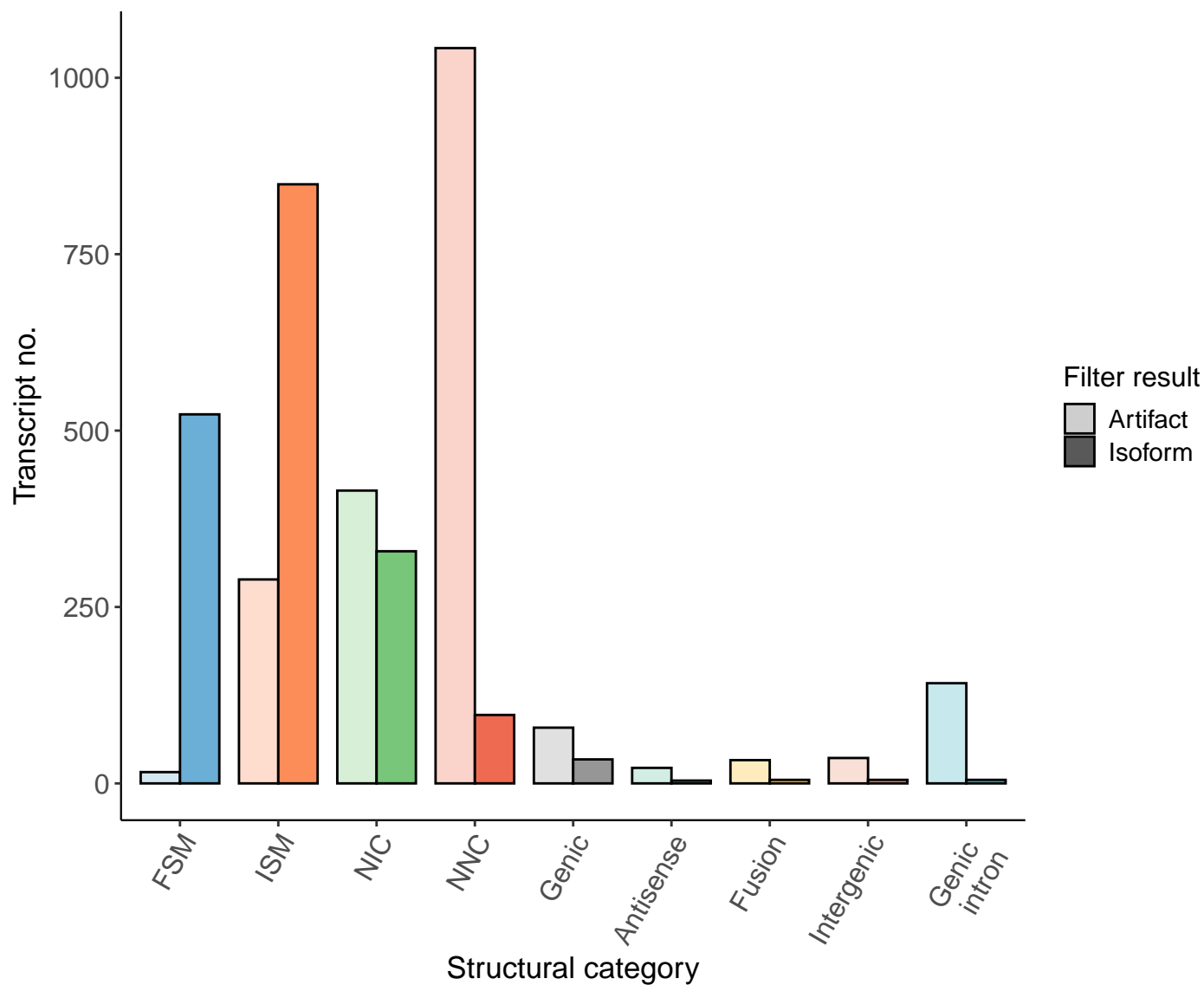
*– Isoforms: 1851 (47%)*

*– Artifacts: 2074 (53%)*

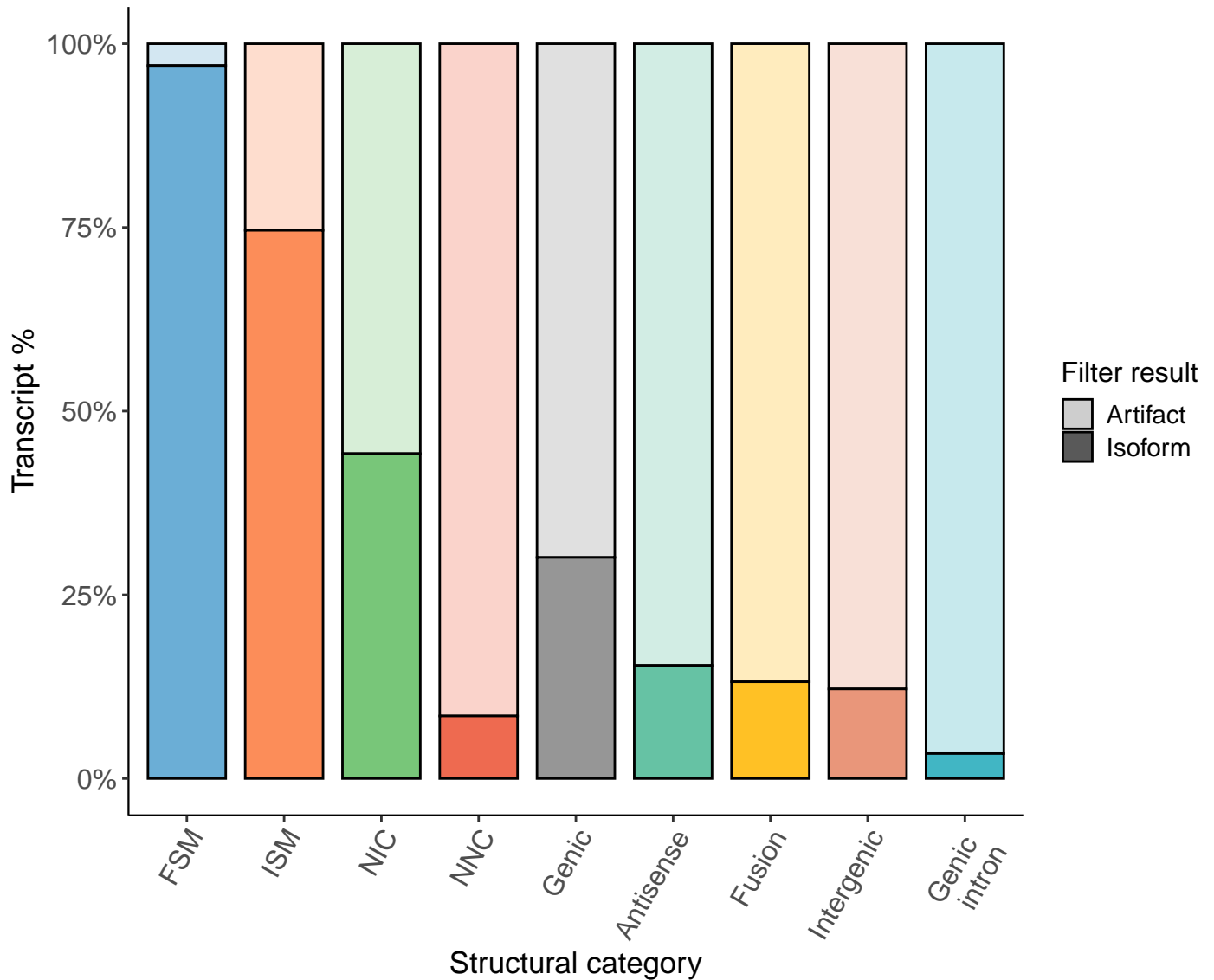
Gene category	Gene no.	No. of genes with artifacts only
Annotated	443	77
Novel	213	199

Structural category	Artifact no.	Isoform no.
FSM	16	523
ISM	289	849
NIC	415	329
NNC	1042	97
Genic	79	34
Antisense	22	4
Fusion	33	5
Intergenic	36	5
Genic intron	142	5

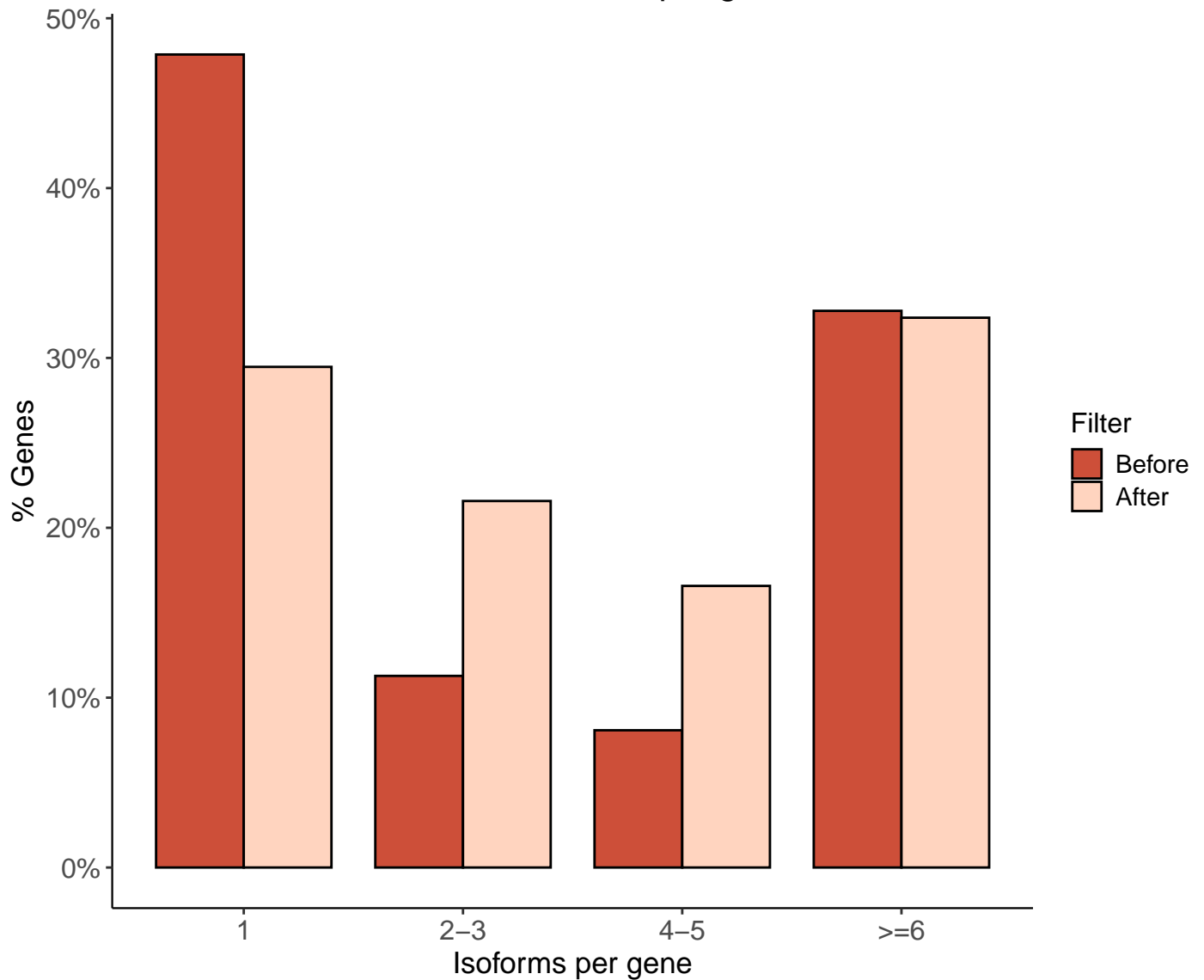
Total isoforms and artifacts by category



% isoforms and artifacts by category

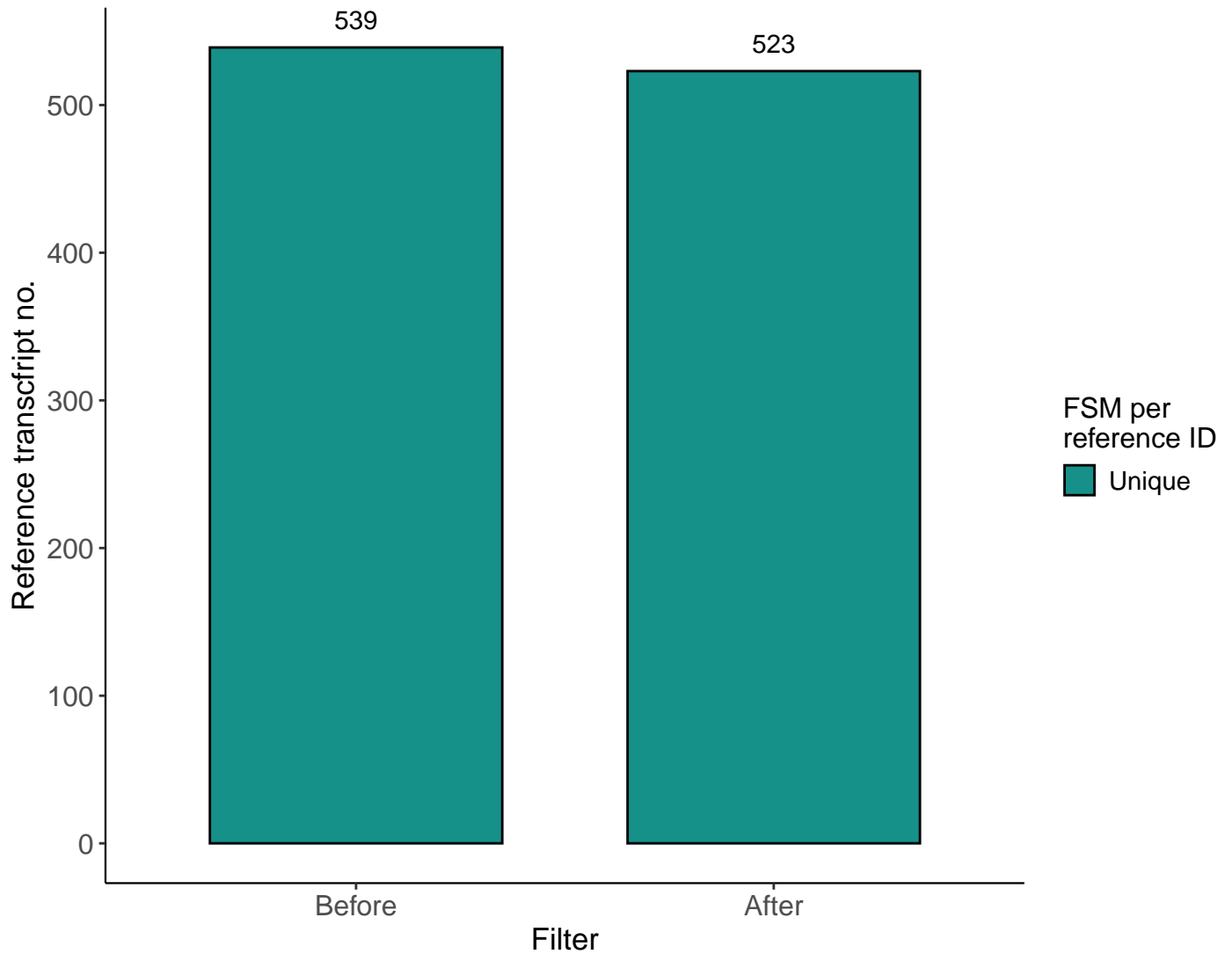


Number of isoforms per gene

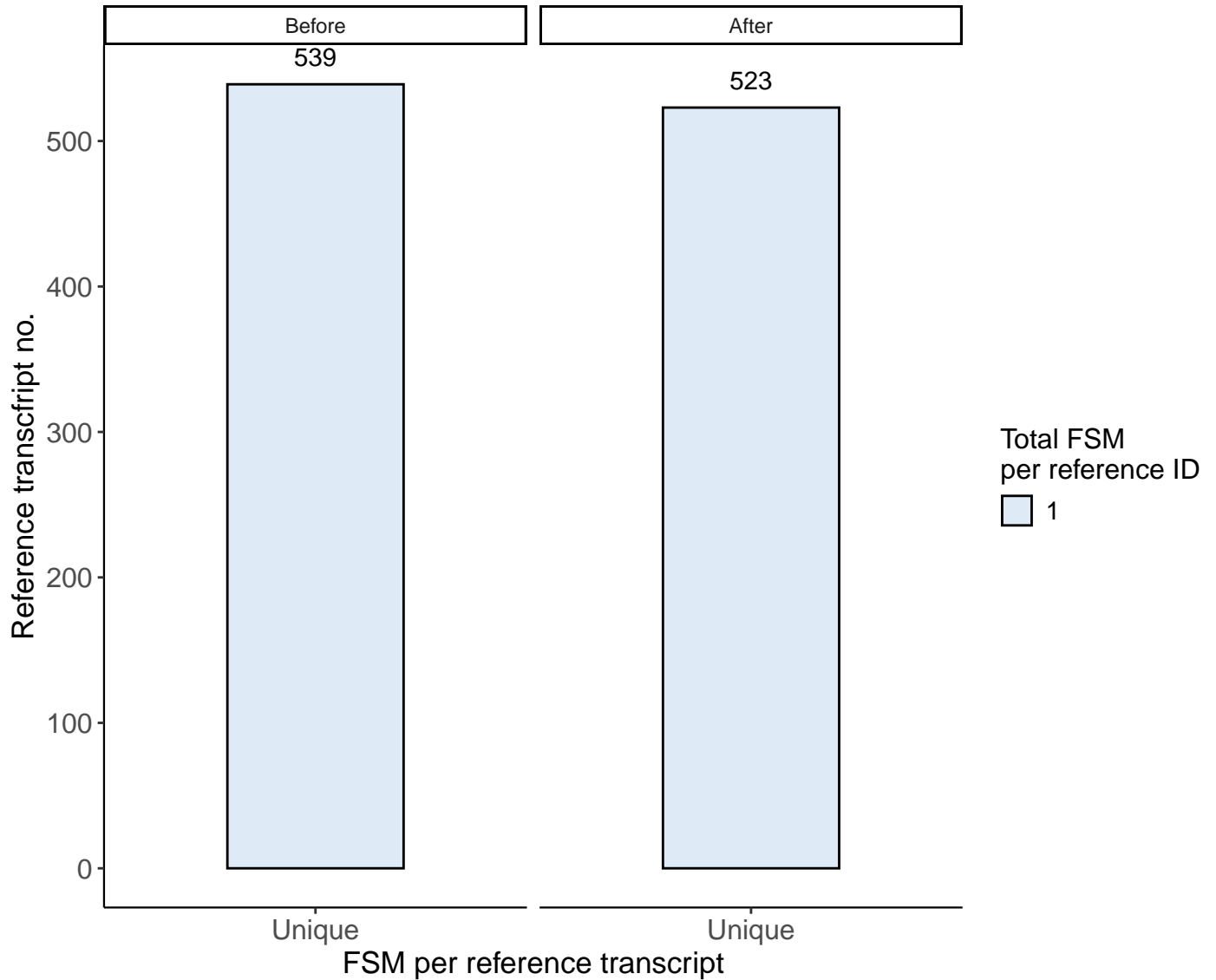


## Reference transcript complexity

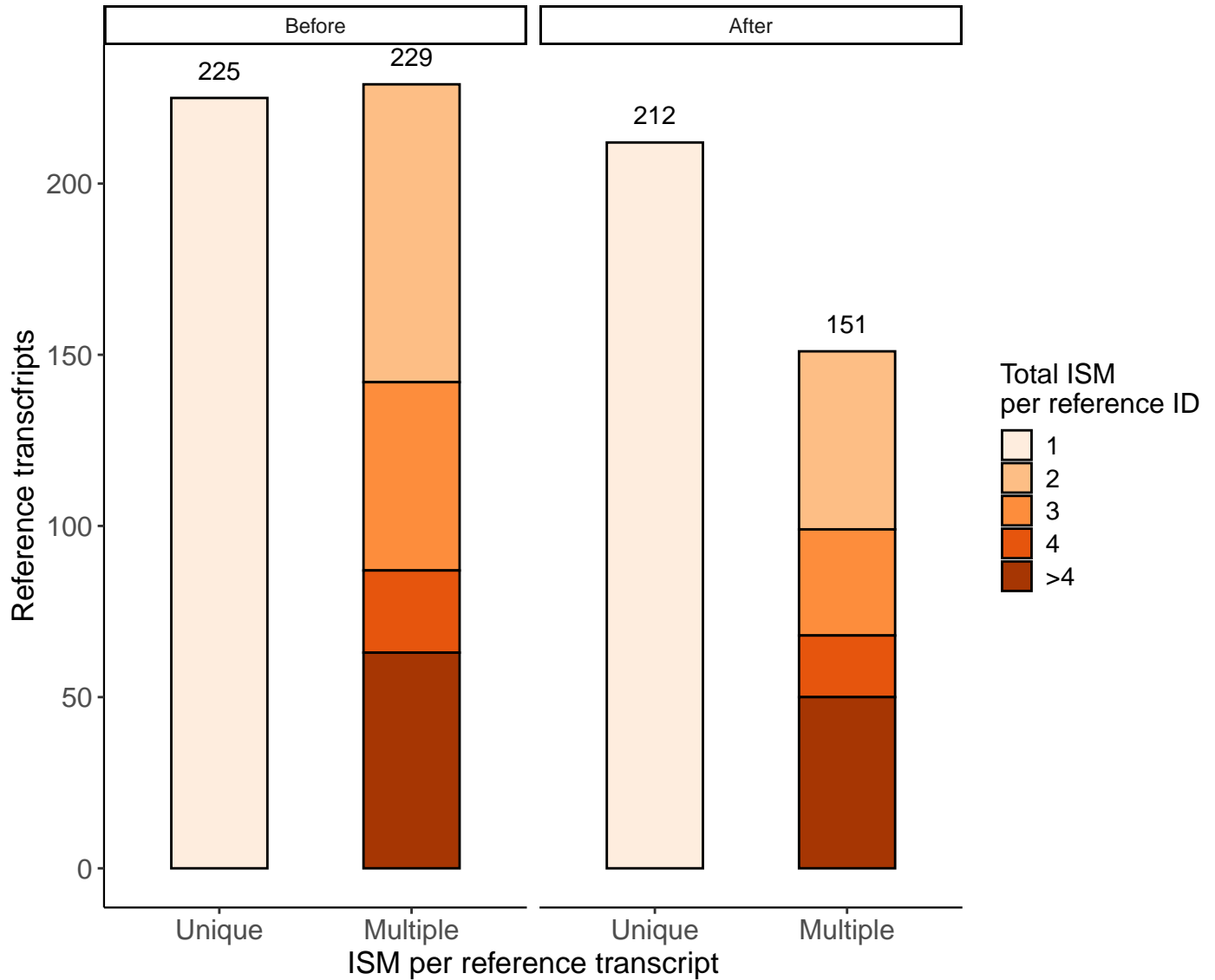
*No. of reference transcripts represented by FSM*



## FSM redundancy

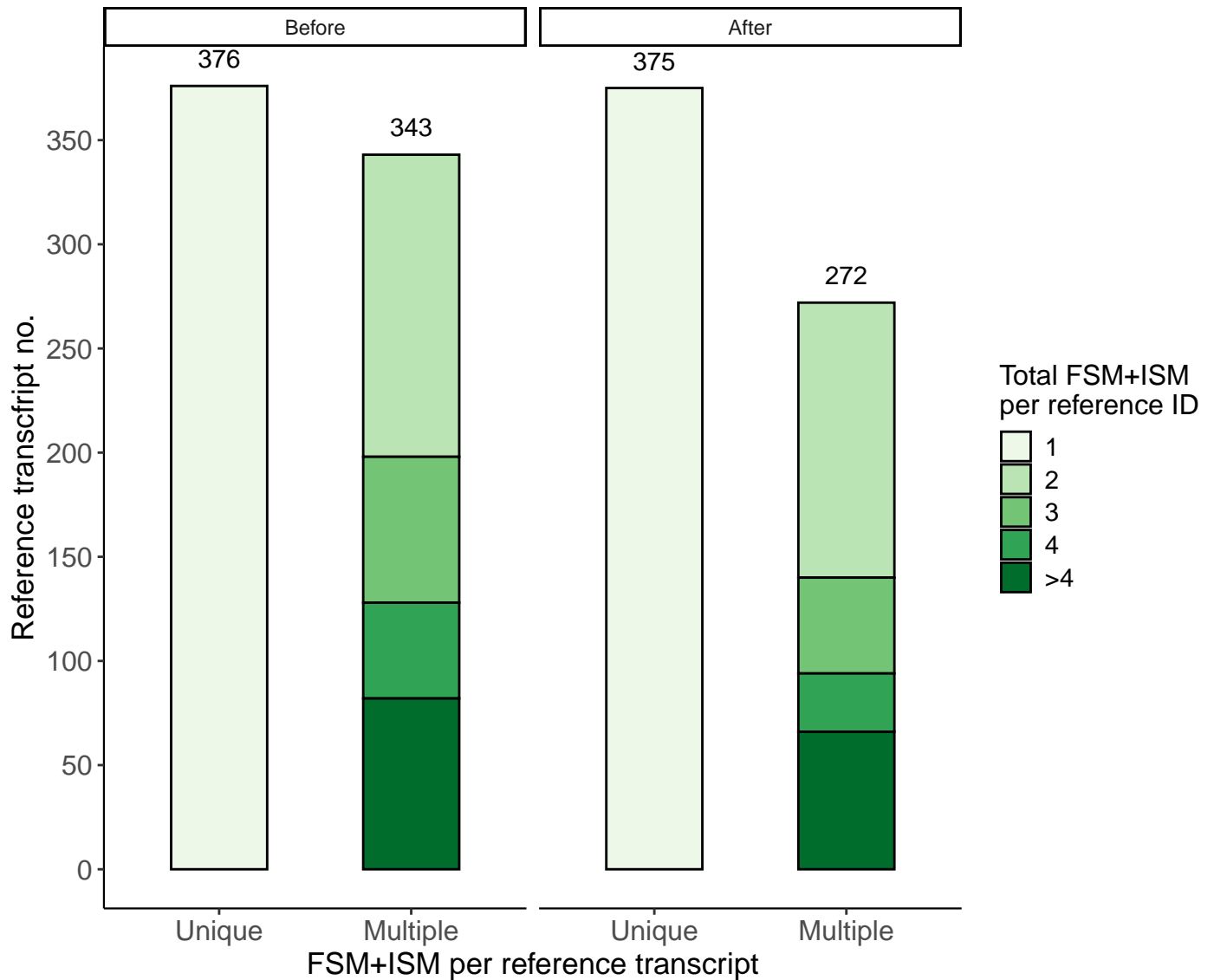


## ISM redundancy





## FSM+ISM redundancy



# *ML classifier performance report*

# Classification model performance on test set

Performance metrics

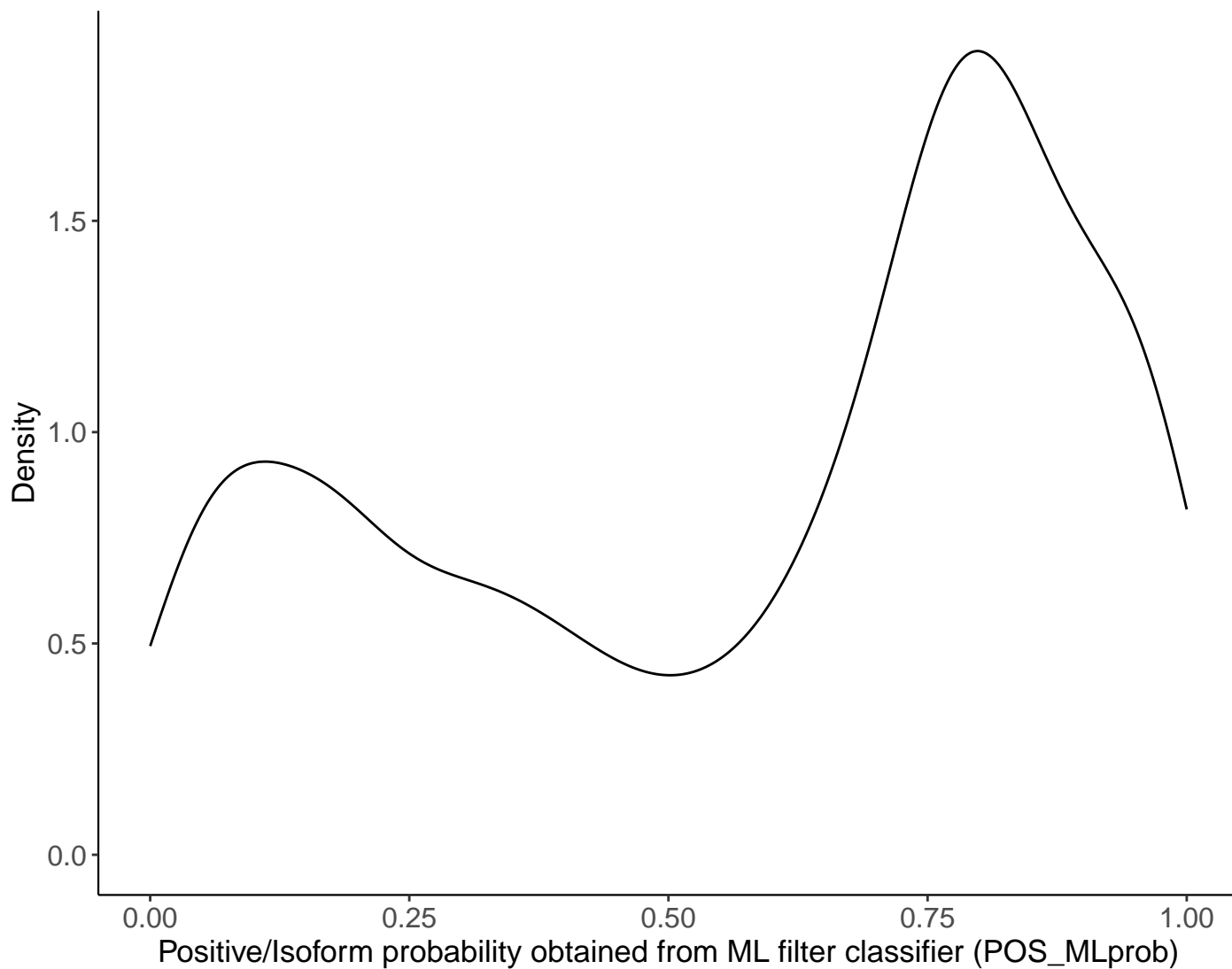
Metric	Value
Accuracy	0.939
Kappa	0.877
AccuracyLower	0.878
AccuracyUpper	0.975
AccuracyNull	0.5
McnemarPValue	0.0233
Sensitivity	0.877
Specificity	1
Pos Pred Value	1
Neg Pred Value	0.891
Precision	1
Recall	0.877
F1	0.935
Prevalence	0.5
Detection Rate	0.439
Detection Prevalence	0.439
Balanced Accuracy	0.939

Confusion matrix

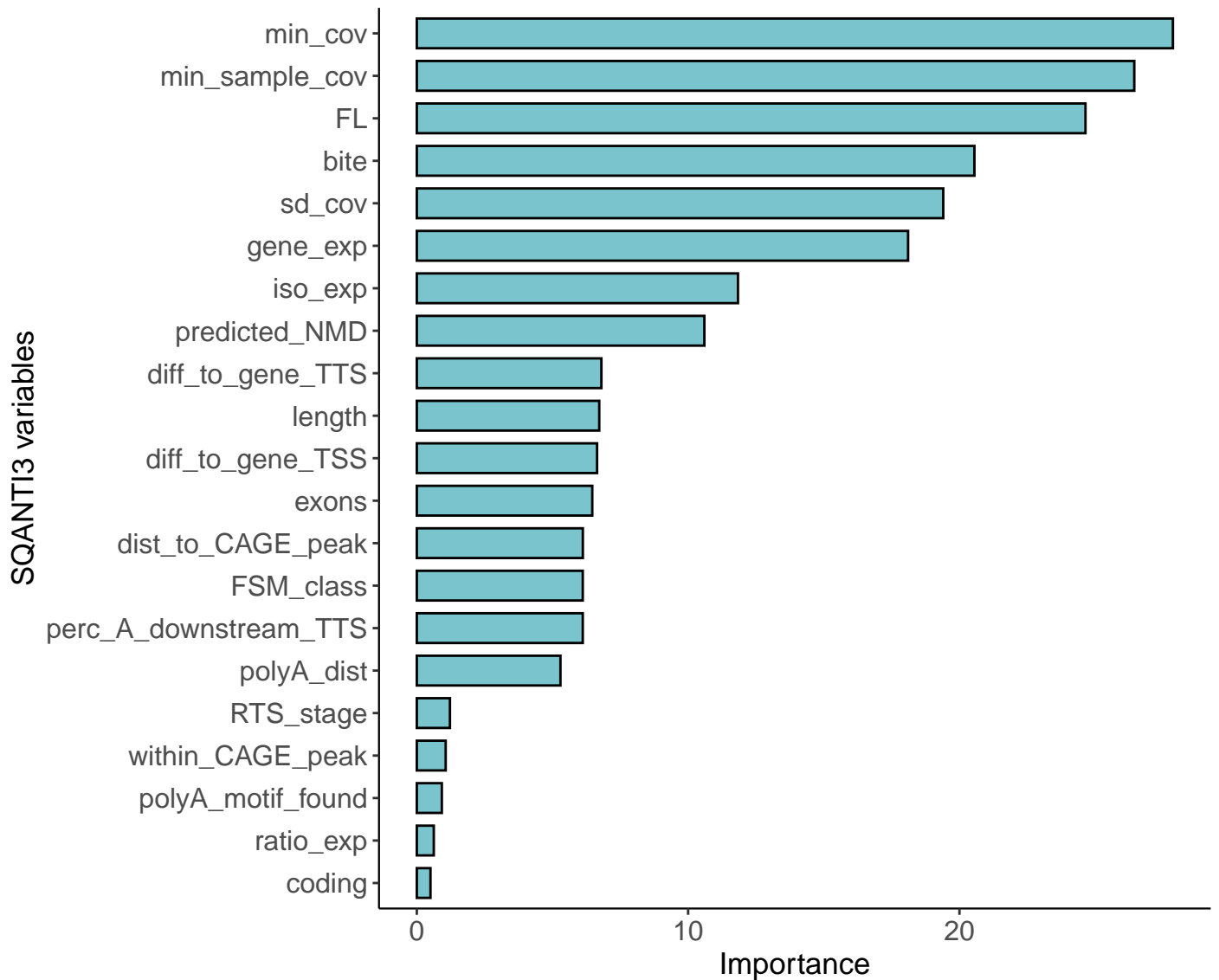
Prediction	Reference	Freq
POS	POS	50
NEG	POS	7
POS	NEG	0
NEG	NEG	57

AccuracyPValue	2.116959e-24
----------------	--------------

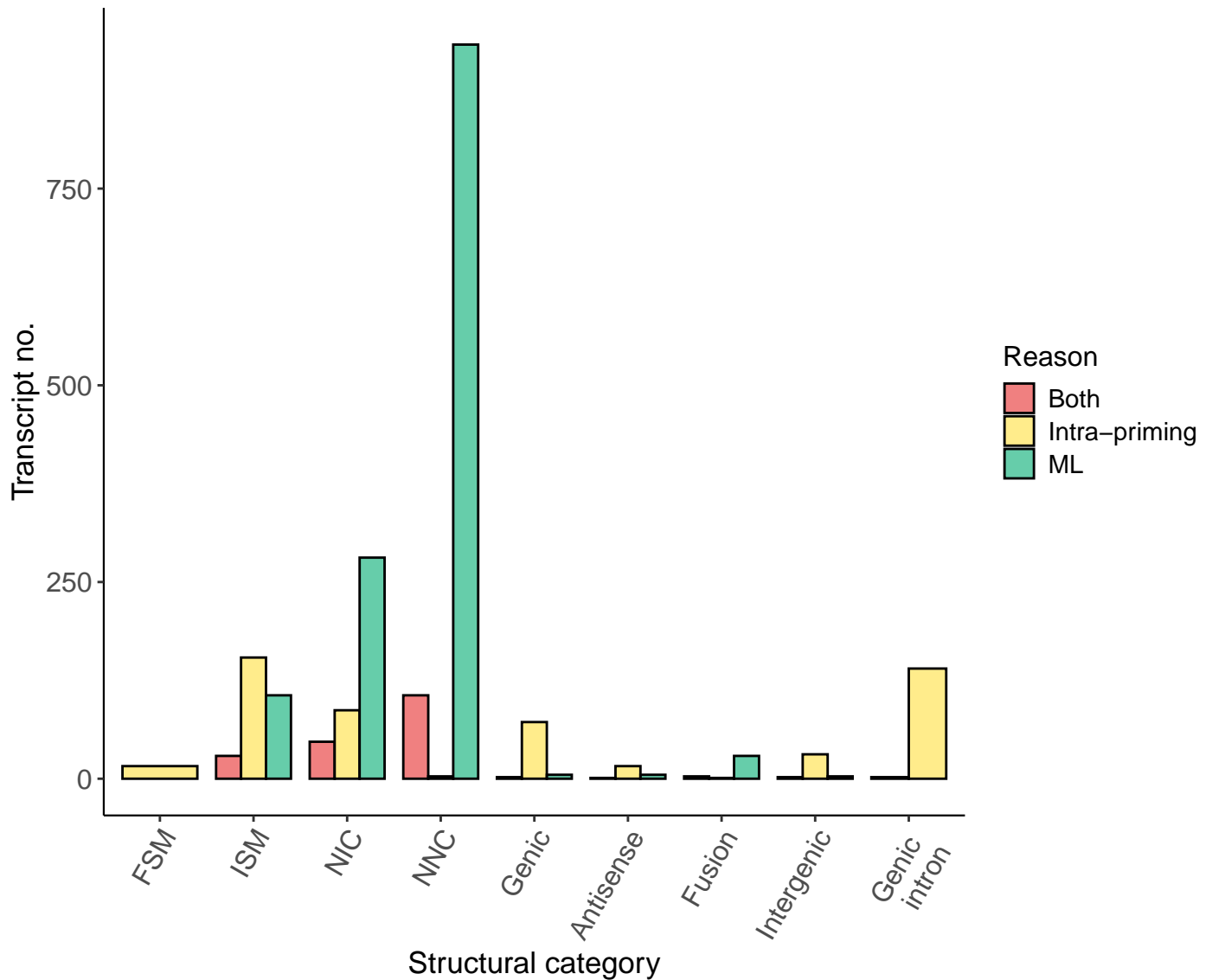
Transcripts classified: 3338



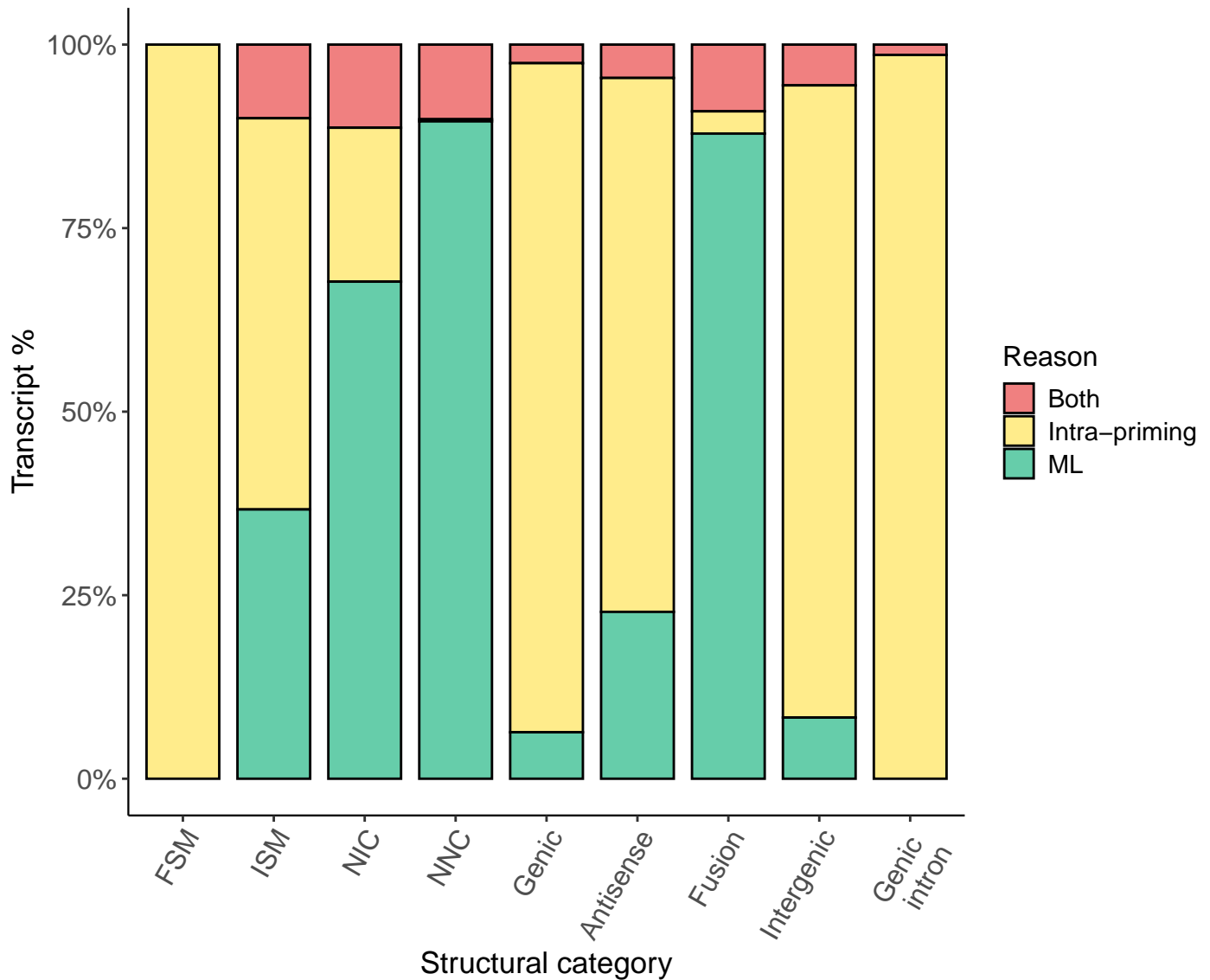
Variable importance in Random Forest classifier



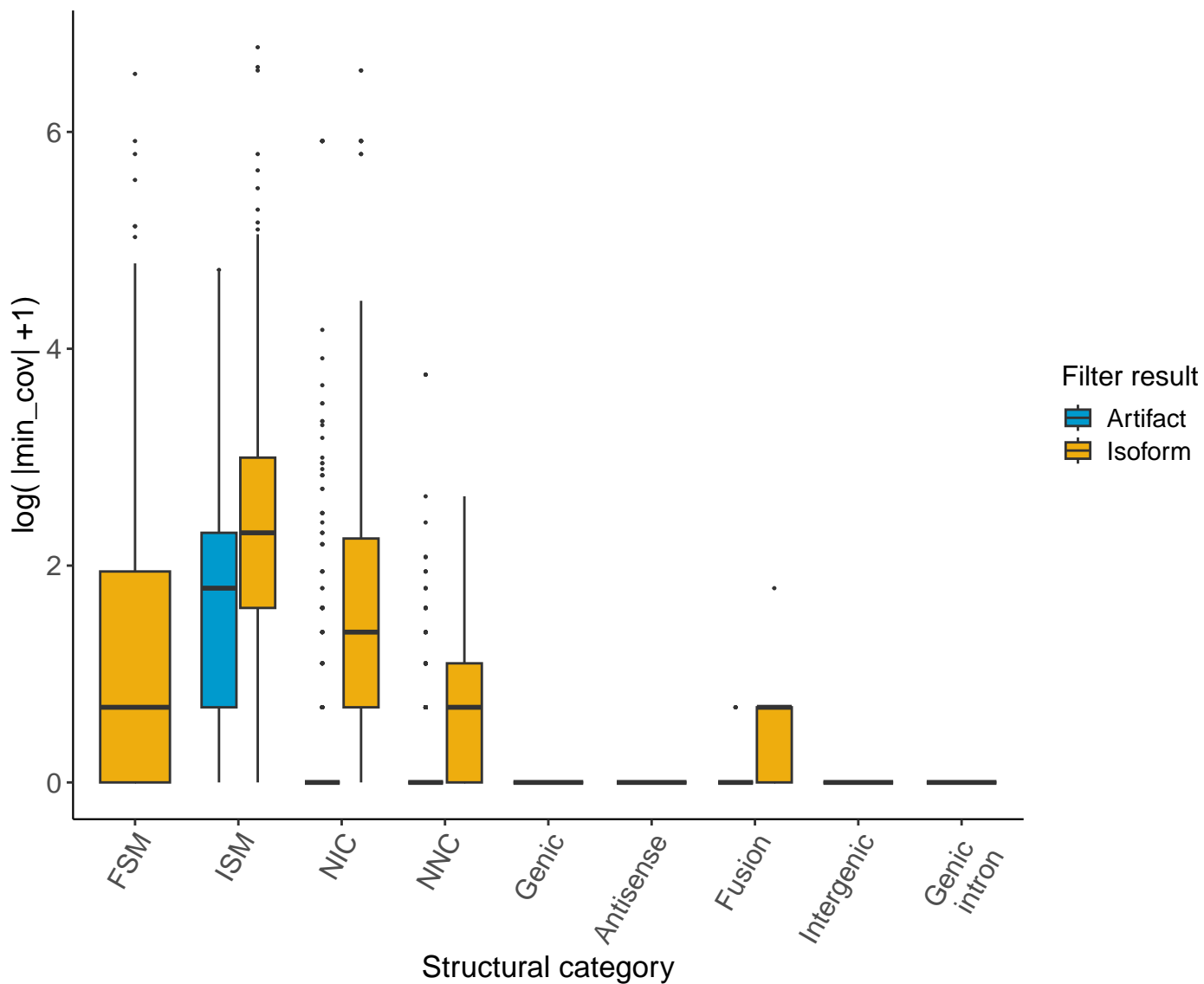
Reason to flag transcripts as artifacts, by category



Reason to flag transcripts as artifacts, by category

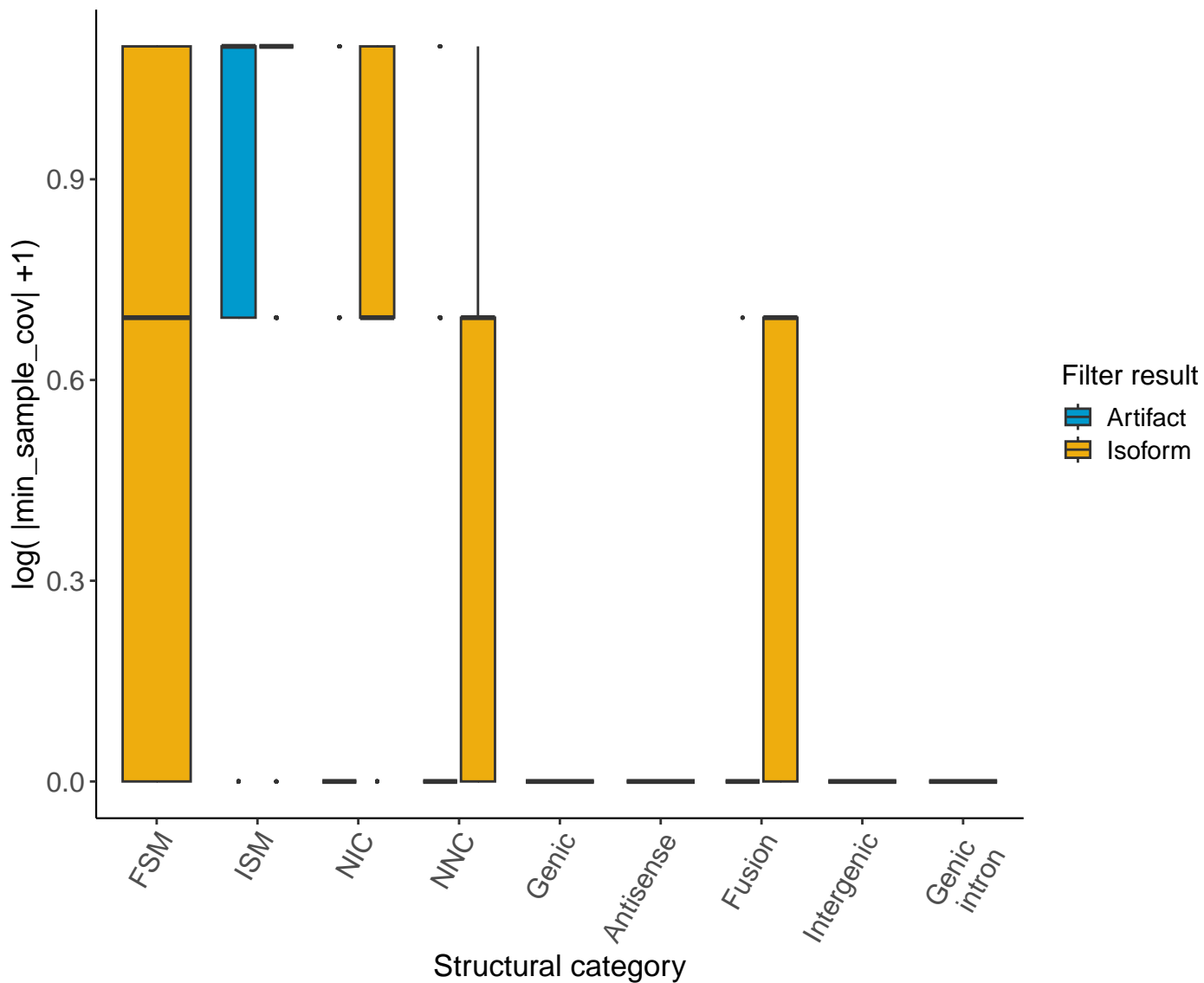


min\_cov – ML importance: 27.87

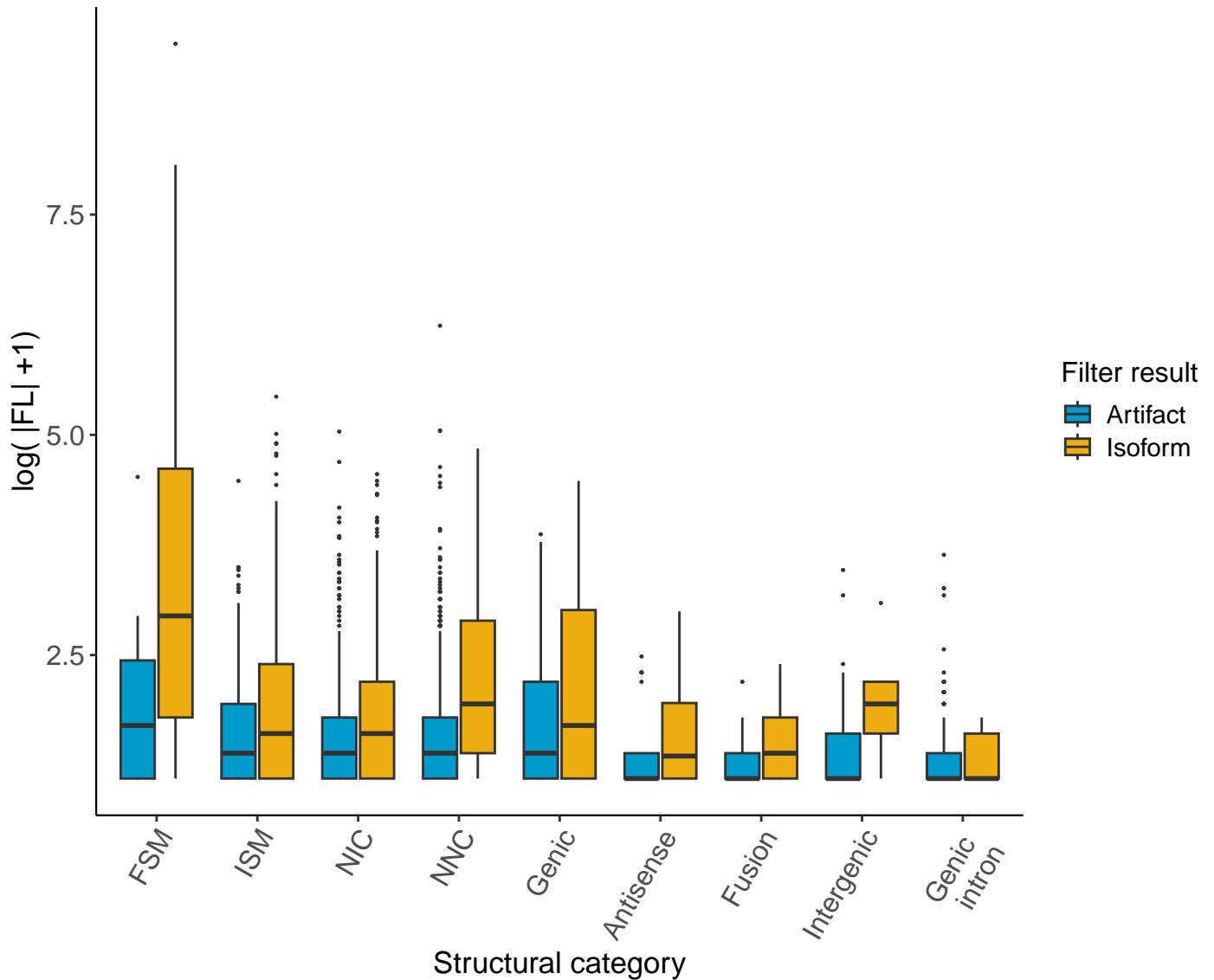




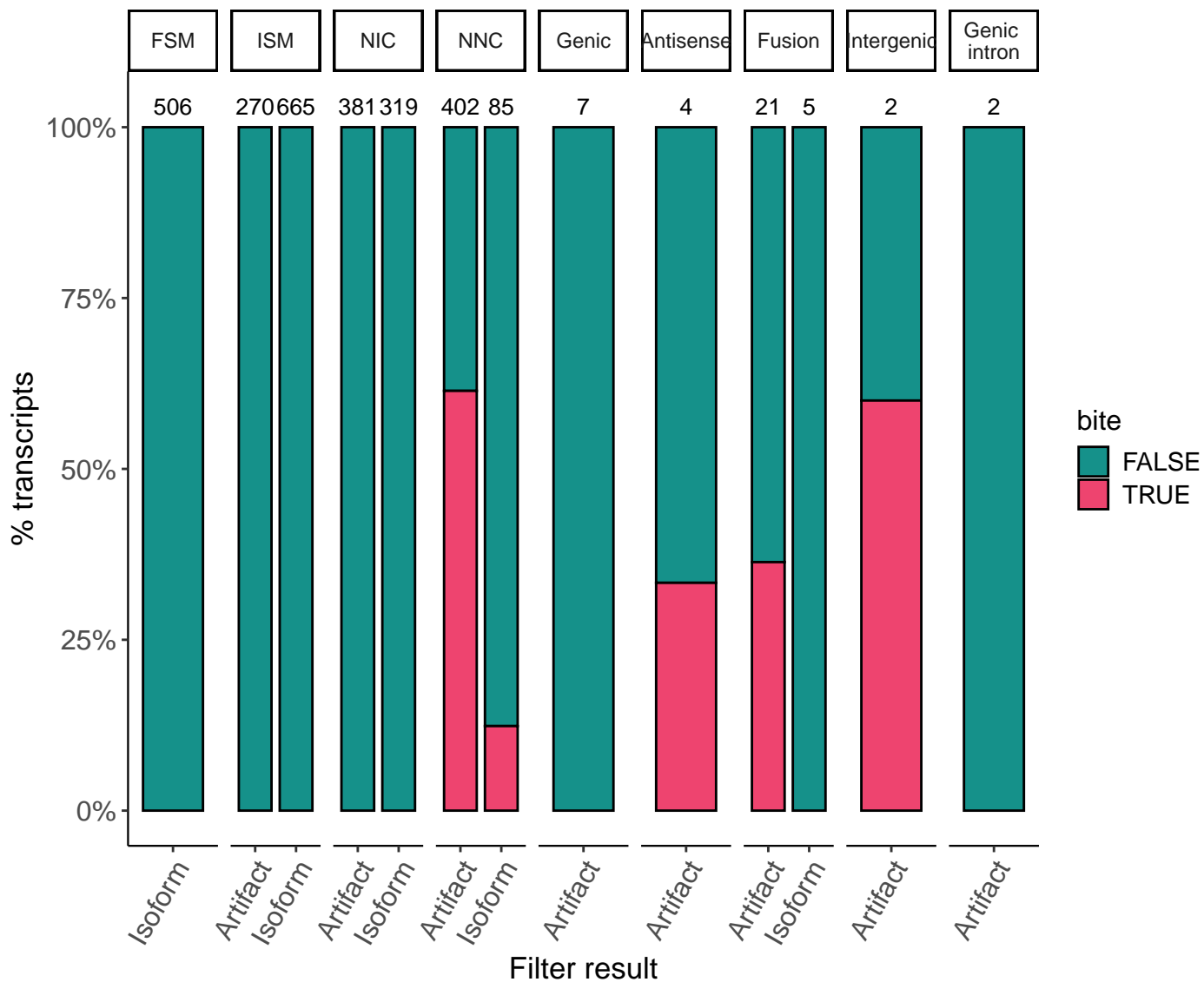
min\_sample\_cov – ML importance: 26.44



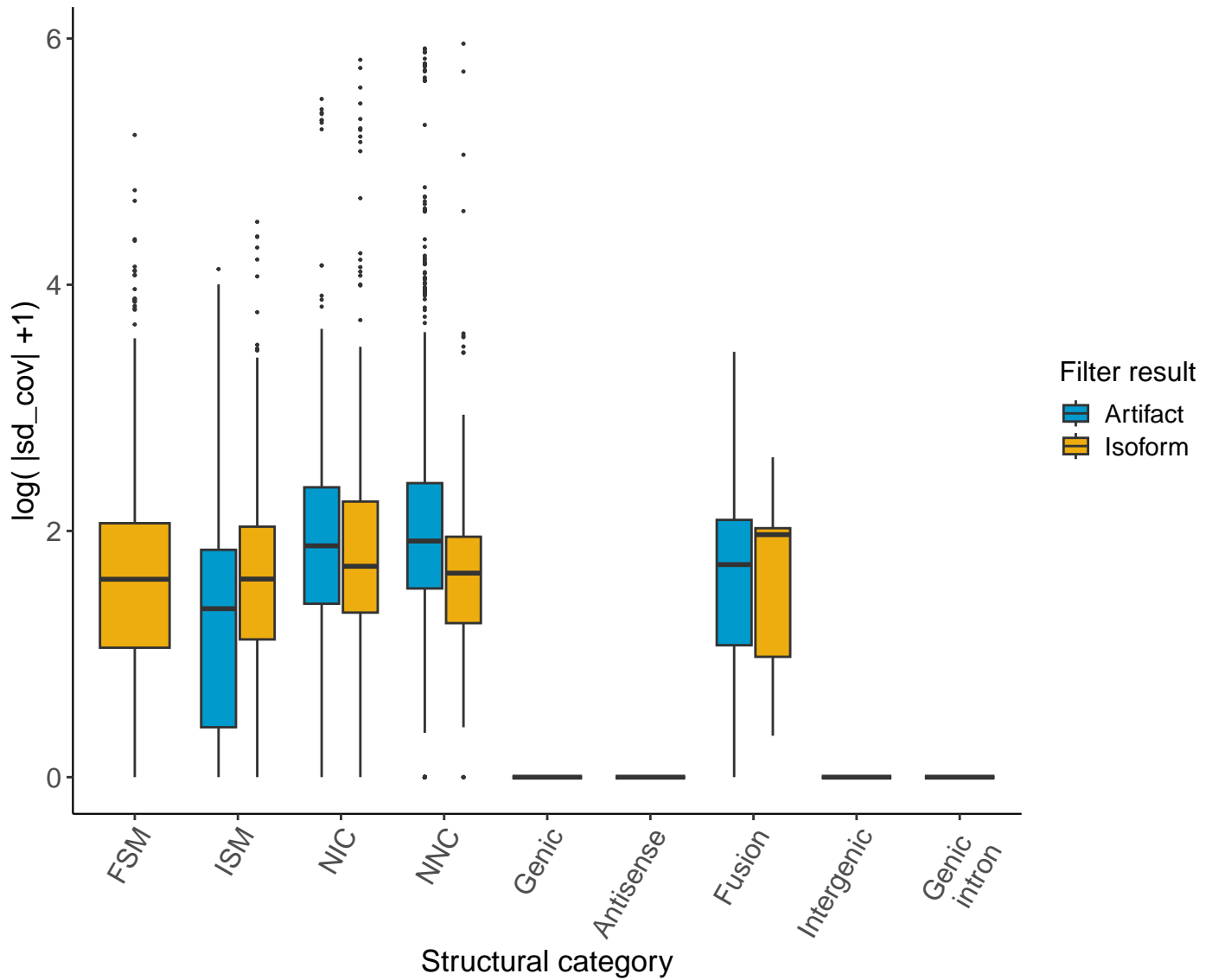
FL – ML importance: 24.64



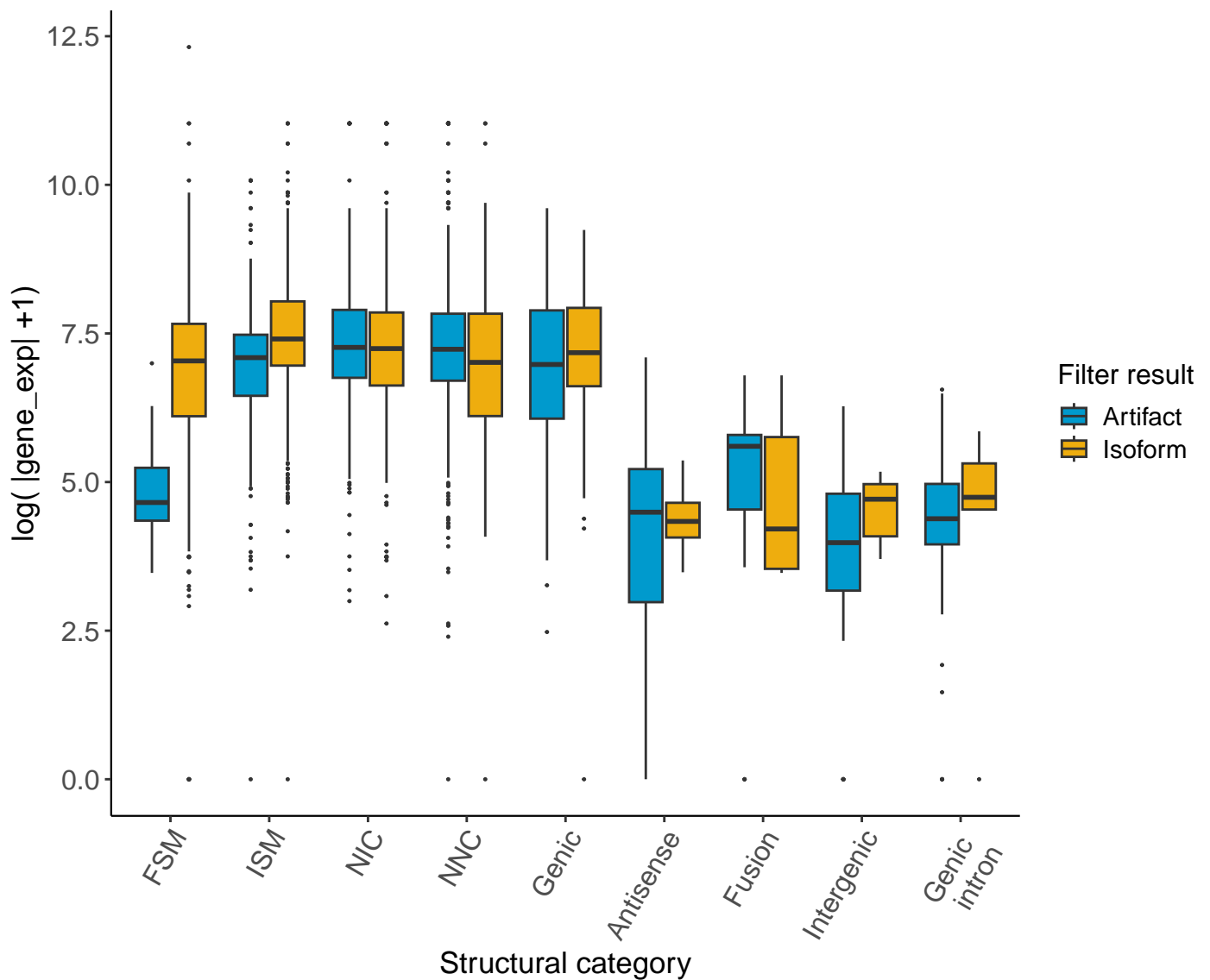
bite – ML importance: 20.55



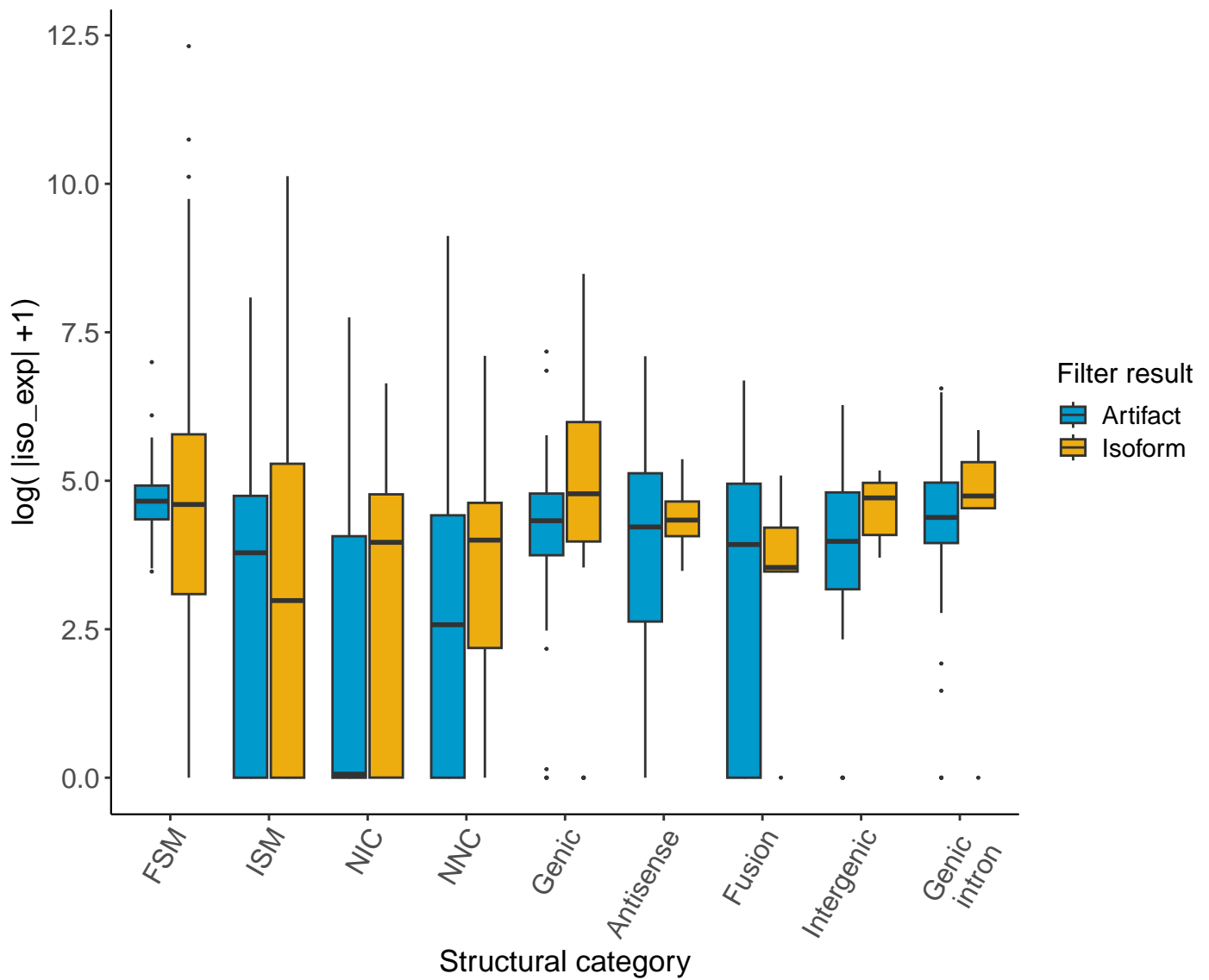
sd\_cov – ML importance: 19.4



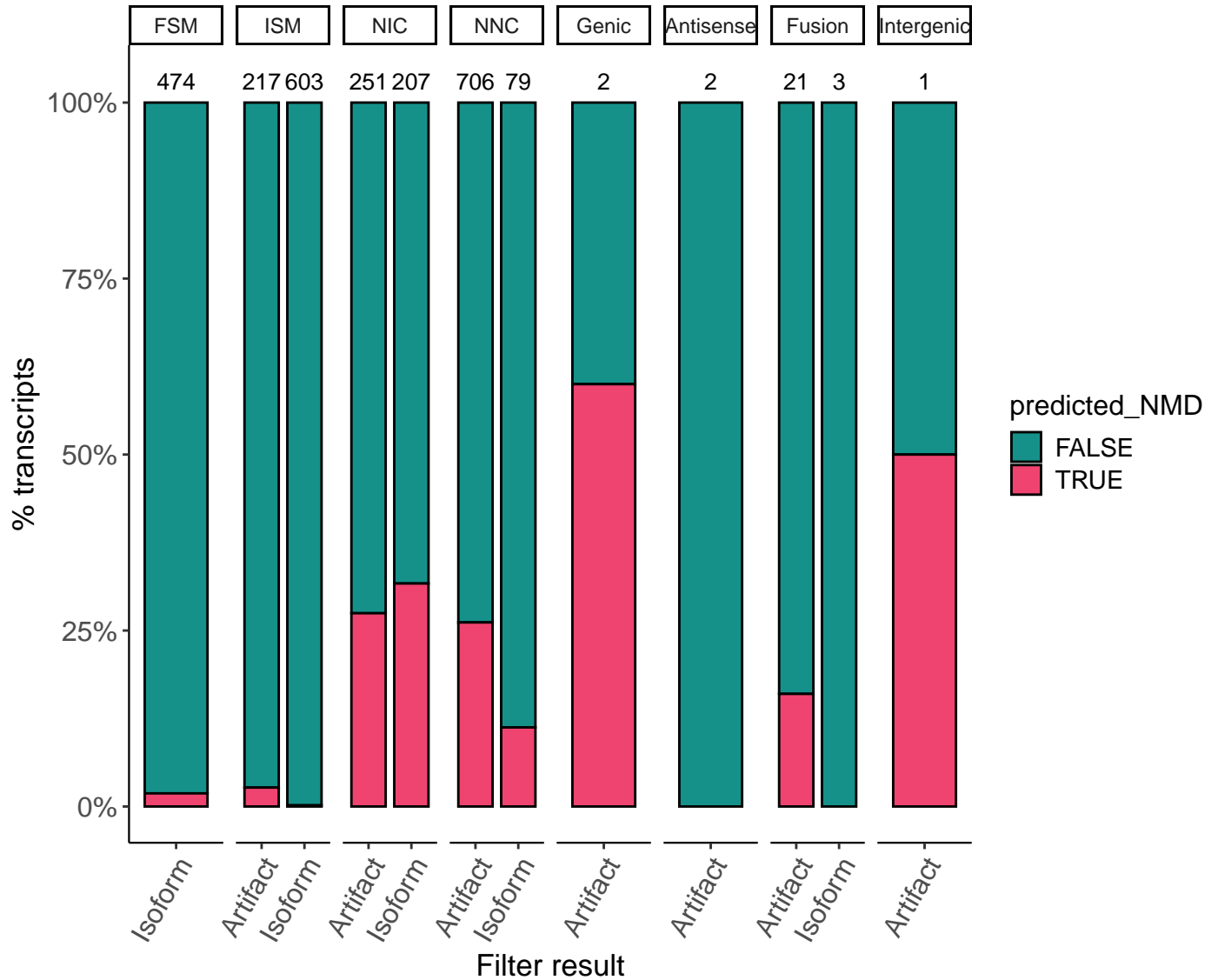
gene\_exp – ML importance: 18.11



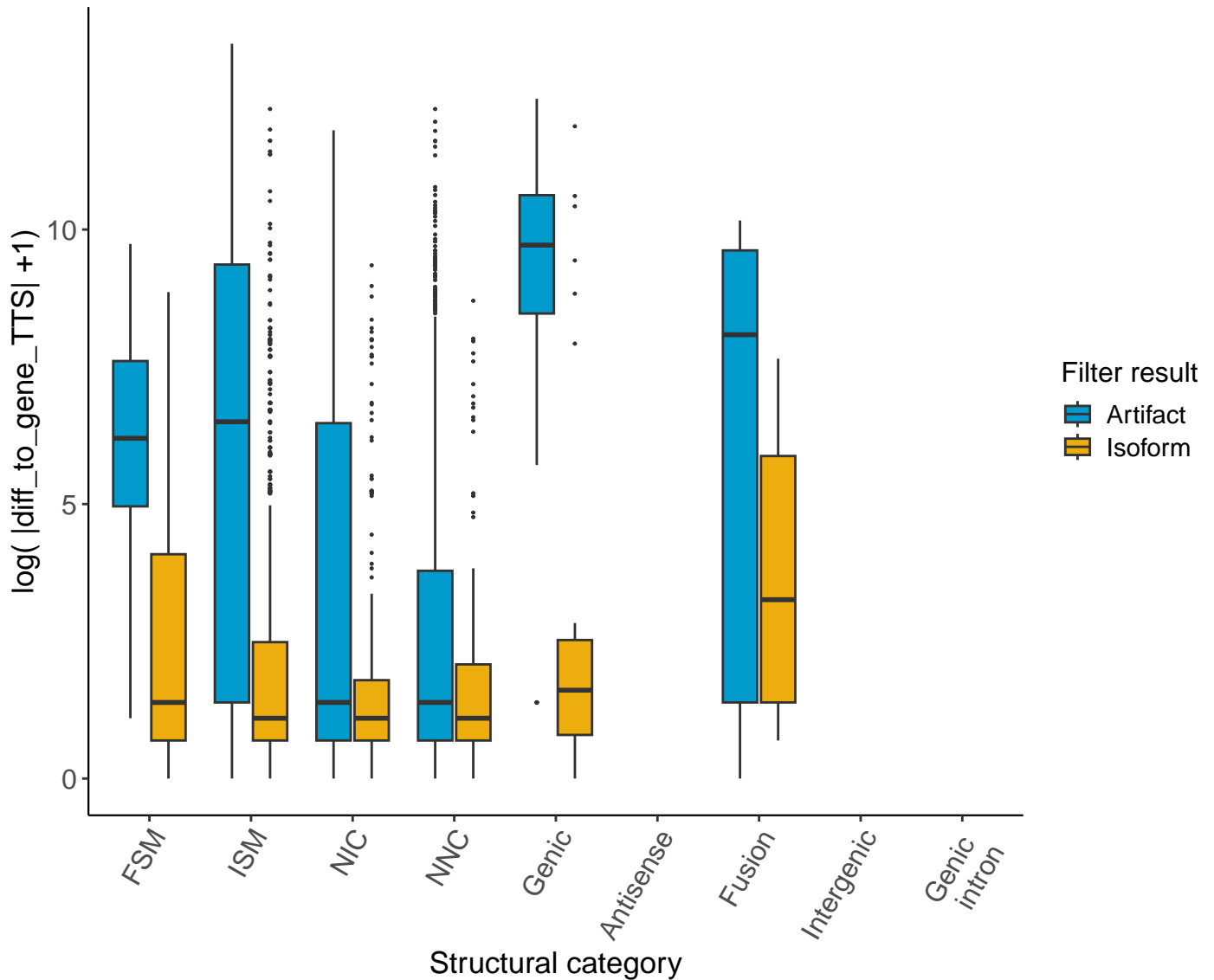
iso\_exp – ML importance: 11.84



predicted\_NMD – ML importance: 10.6

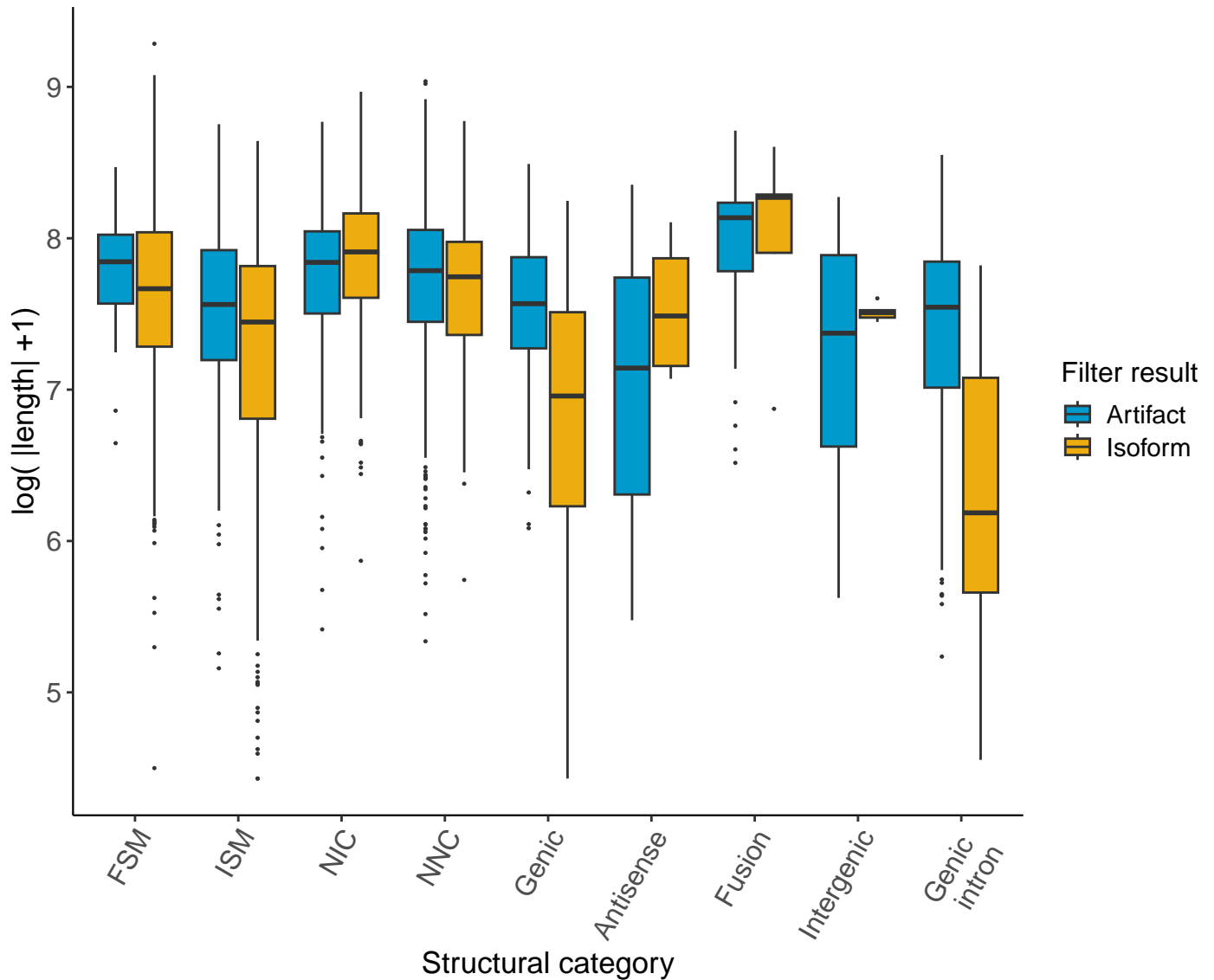


diff\_to\_gene\_TTS – ML importance: 6.8

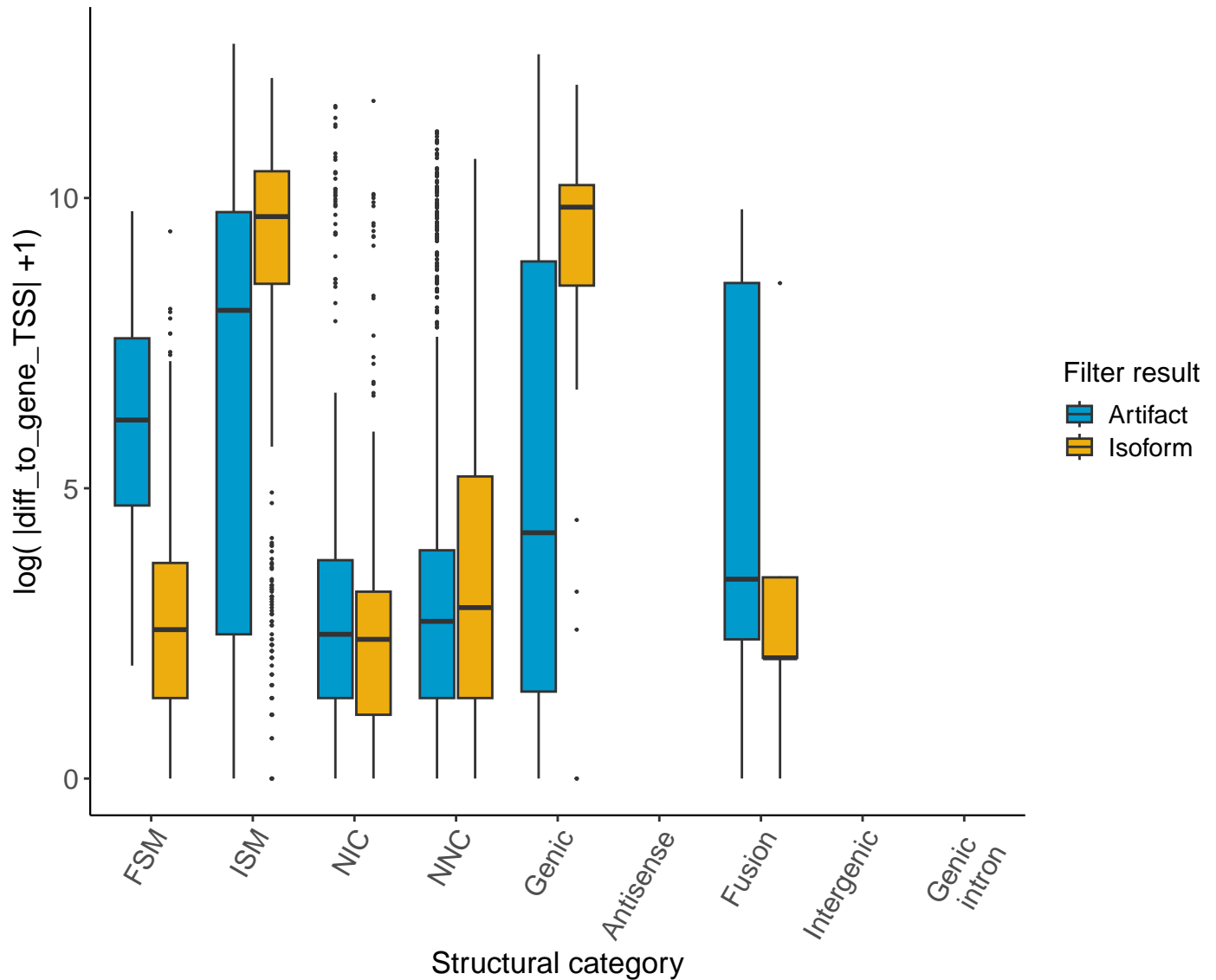




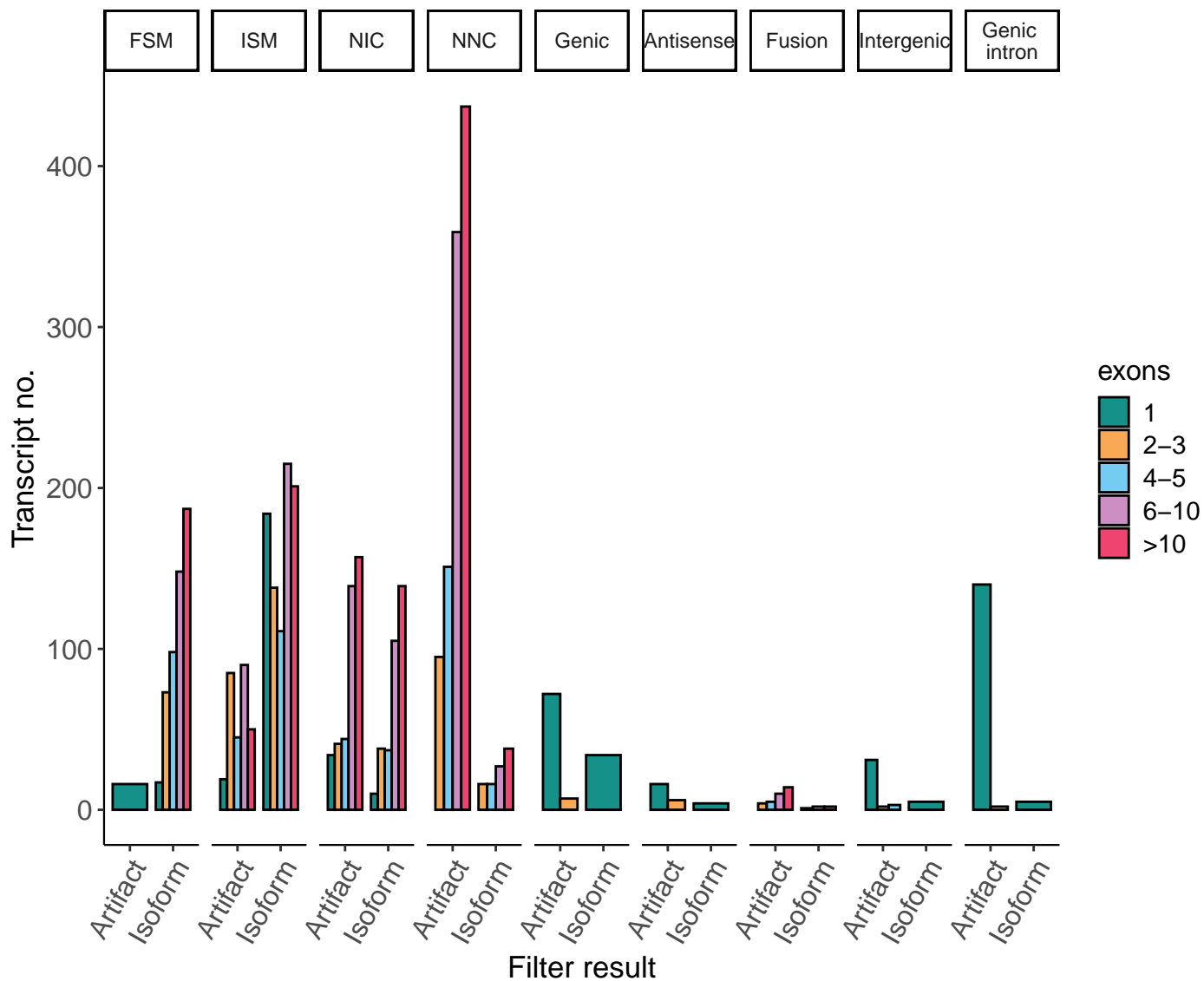
length – ML importance: 6.72



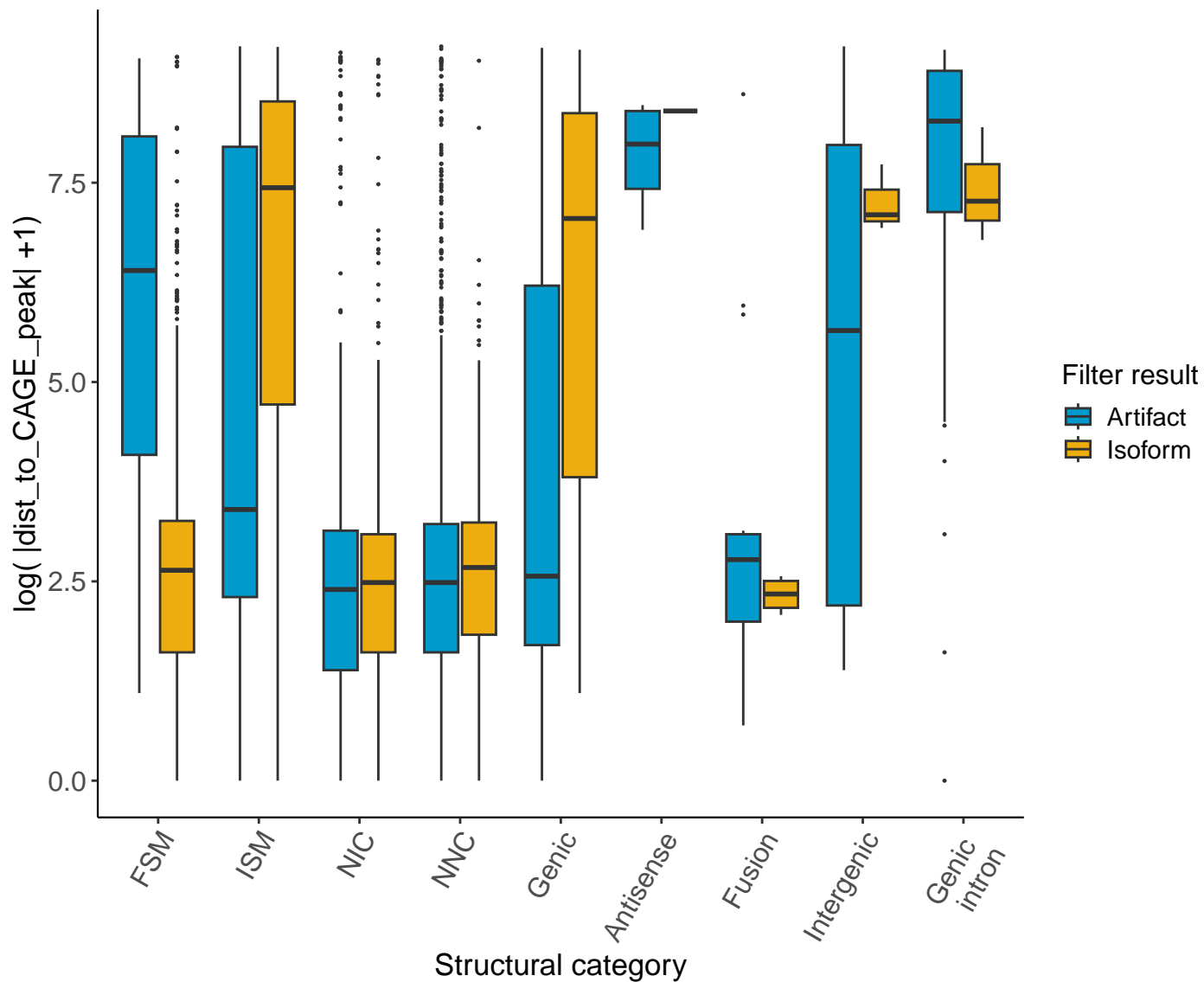
diff\_to\_gene\_TSS – ML importance: 6.64



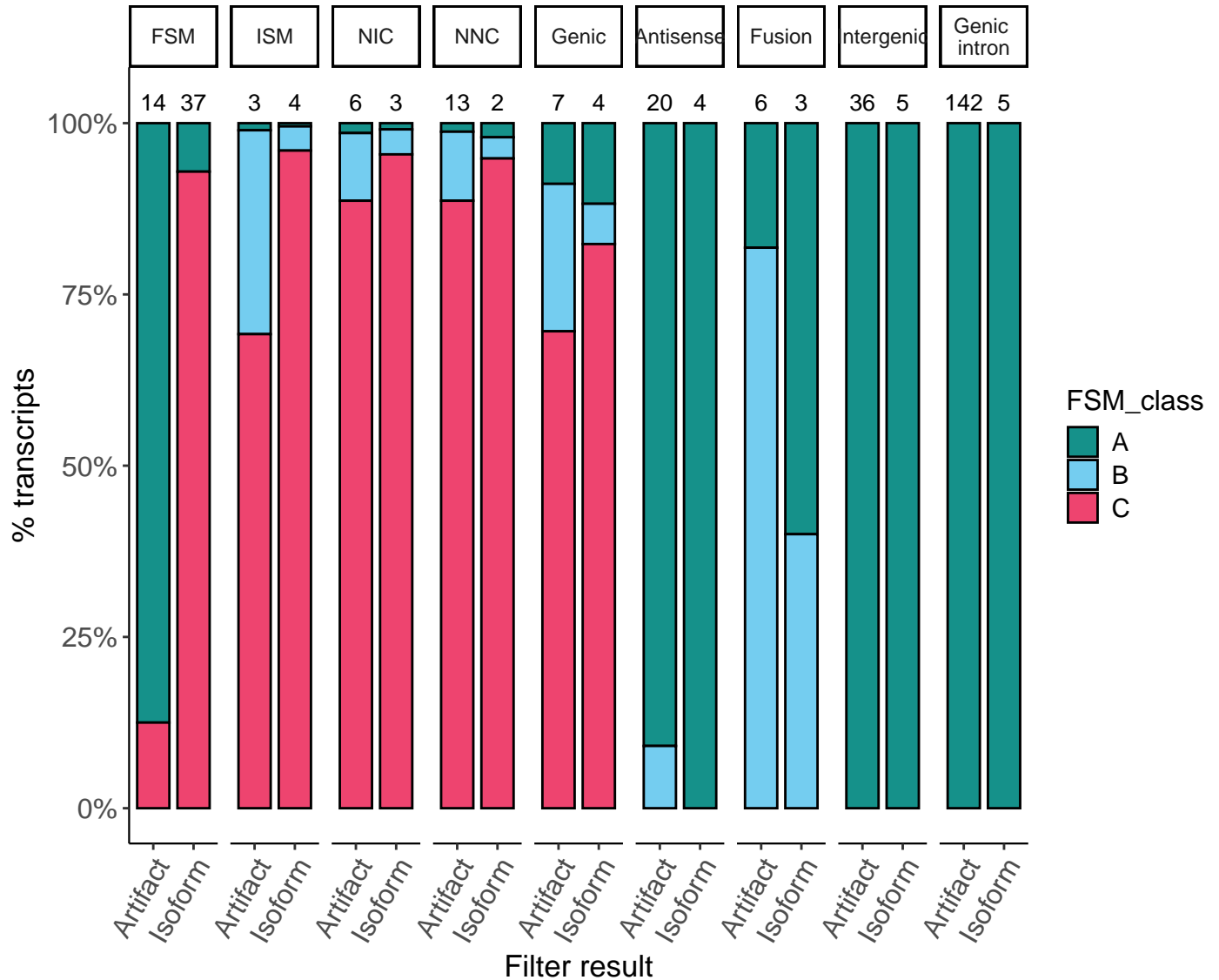
# exons – ML importance: 6.47



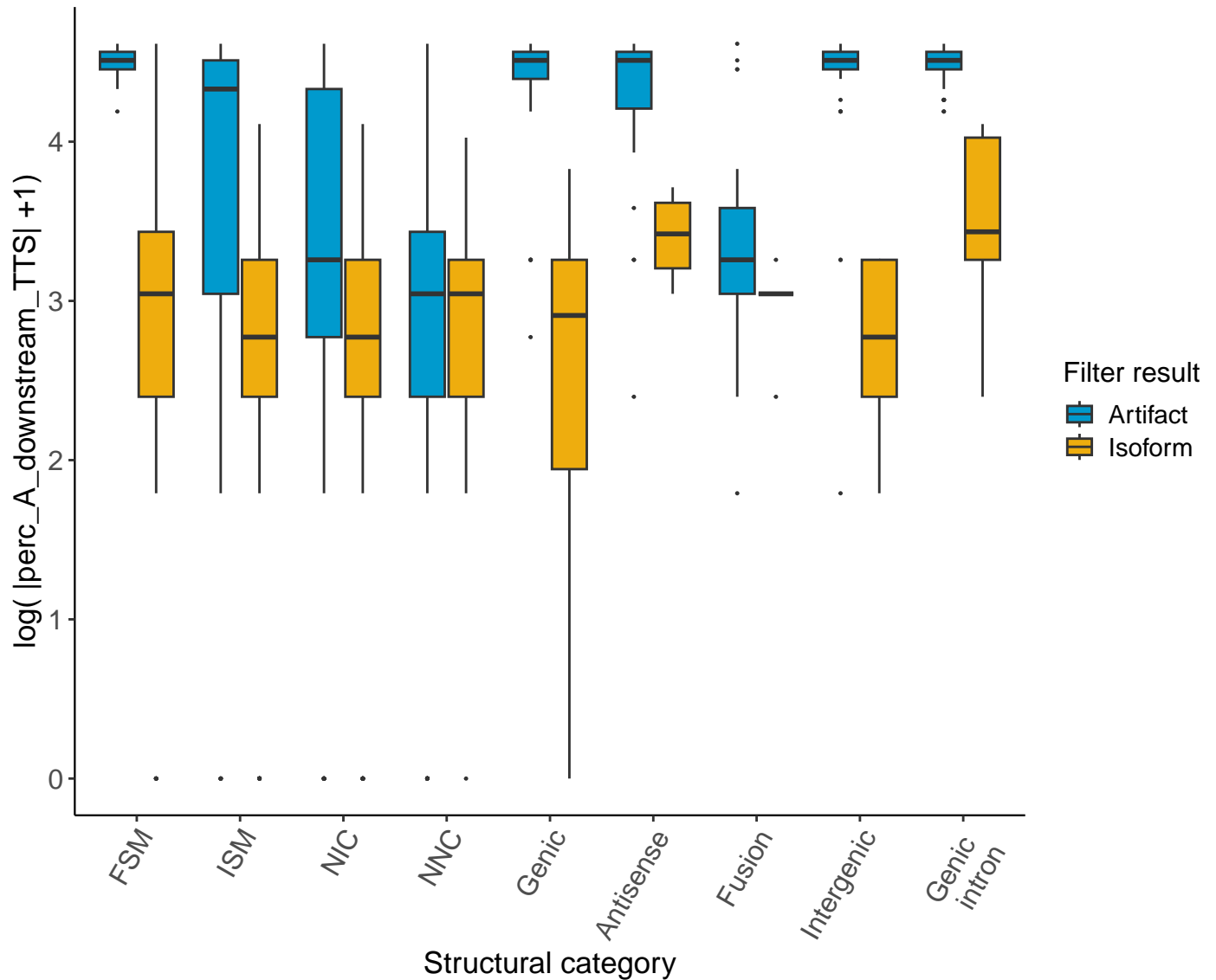
dist\_to\_CAGE\_peak – ML importance: 6.12



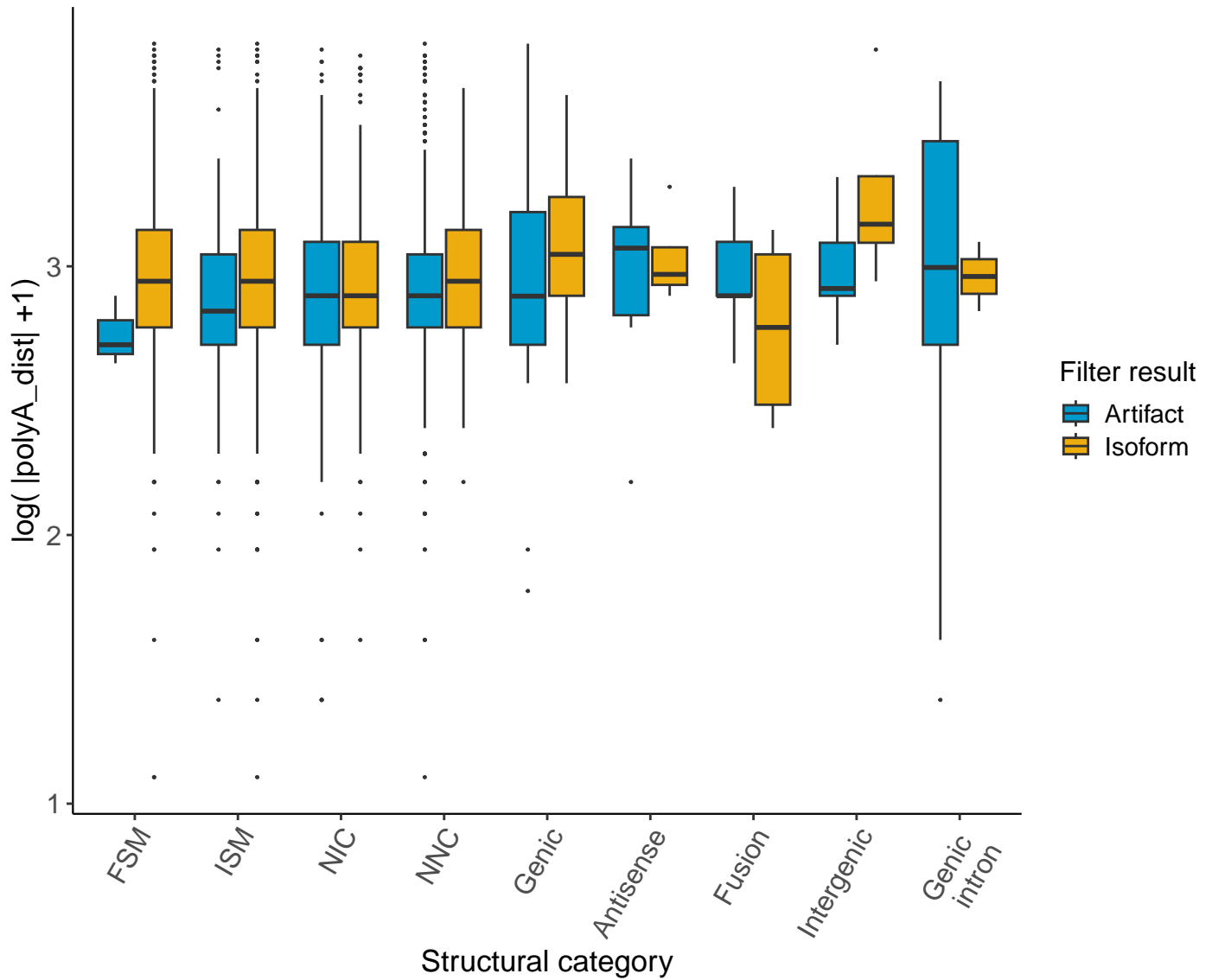
# FSM\_class – ML importance: 6.12



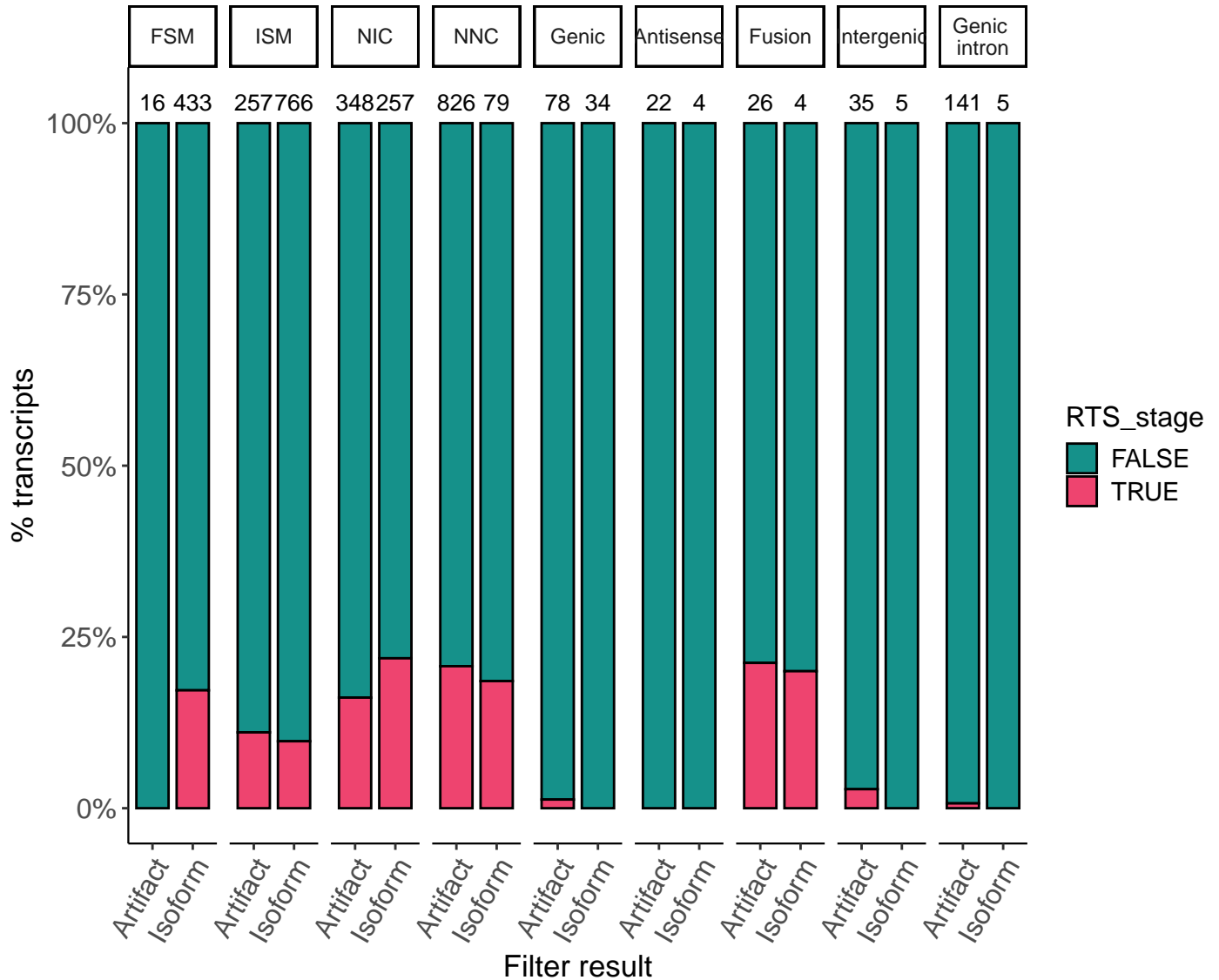
perc\_A\_downstream\_TTS – ML importance: 6.12



polyA\_dist – ML importance: 5.29

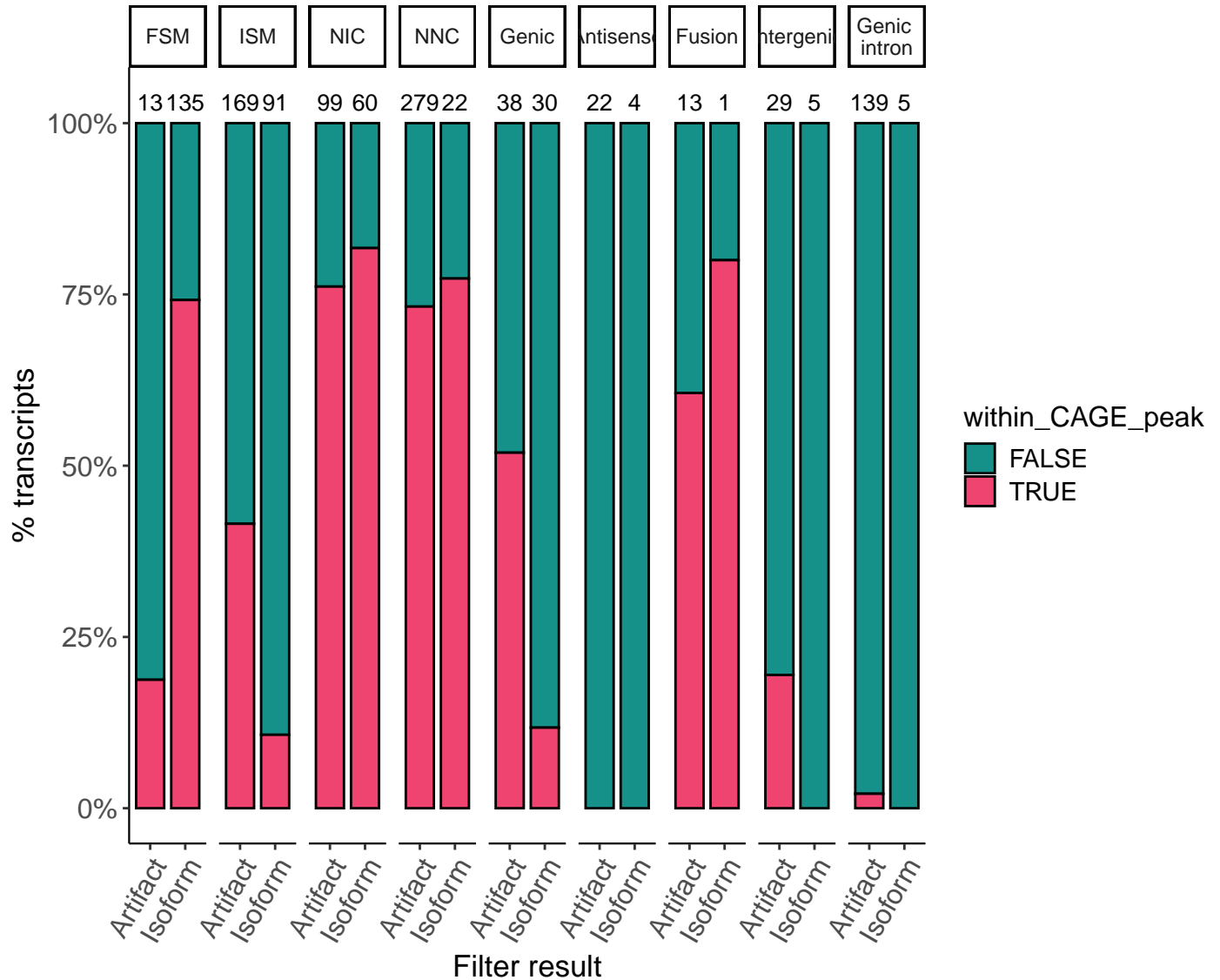


# RTS\_stage – ML importance: 1.22

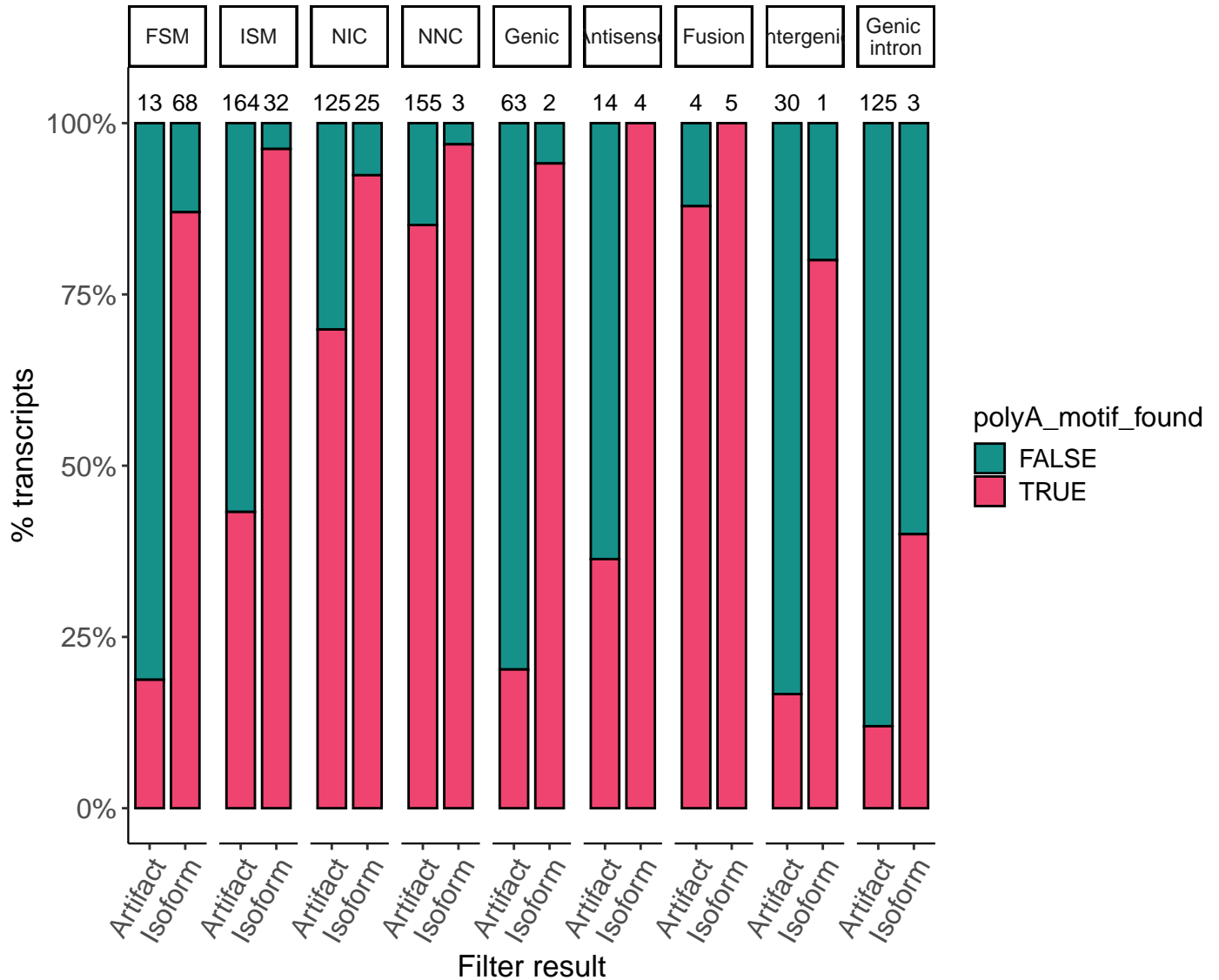




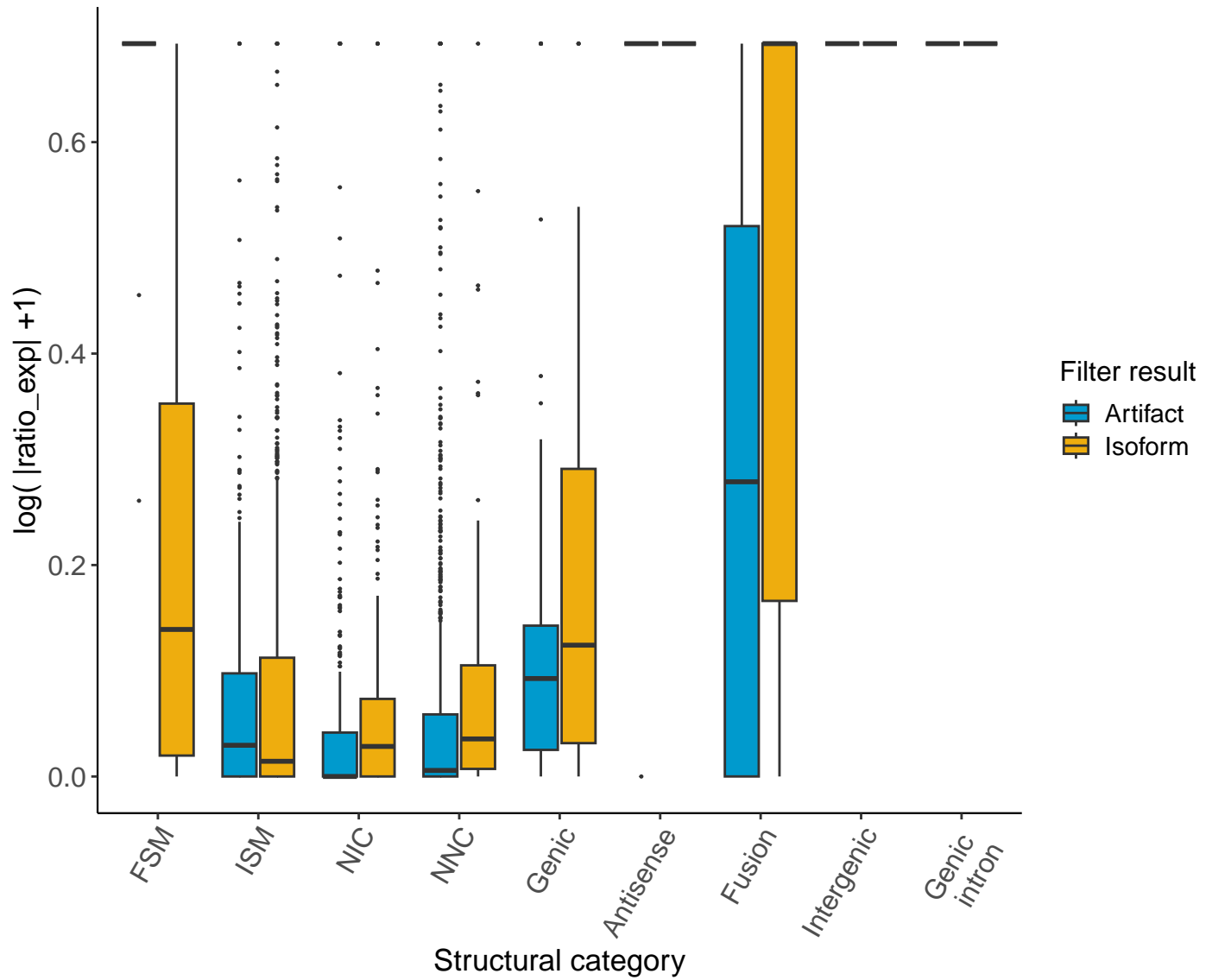
within\_CAGE\_peak – ML importance: 1.06



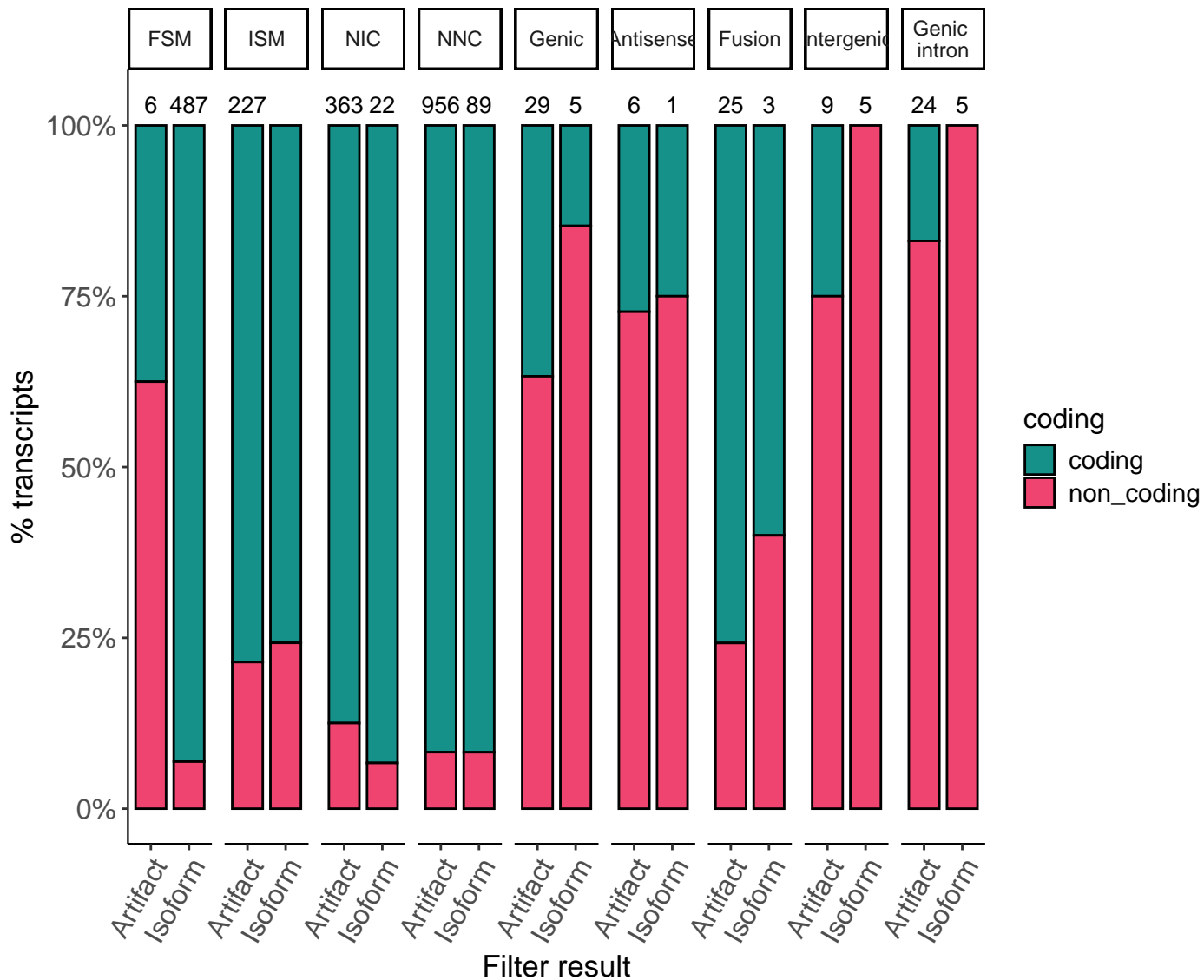
# polyA\_motif\_found – ML importance: 0.92



ratio\_exp – ML importance: 0.63



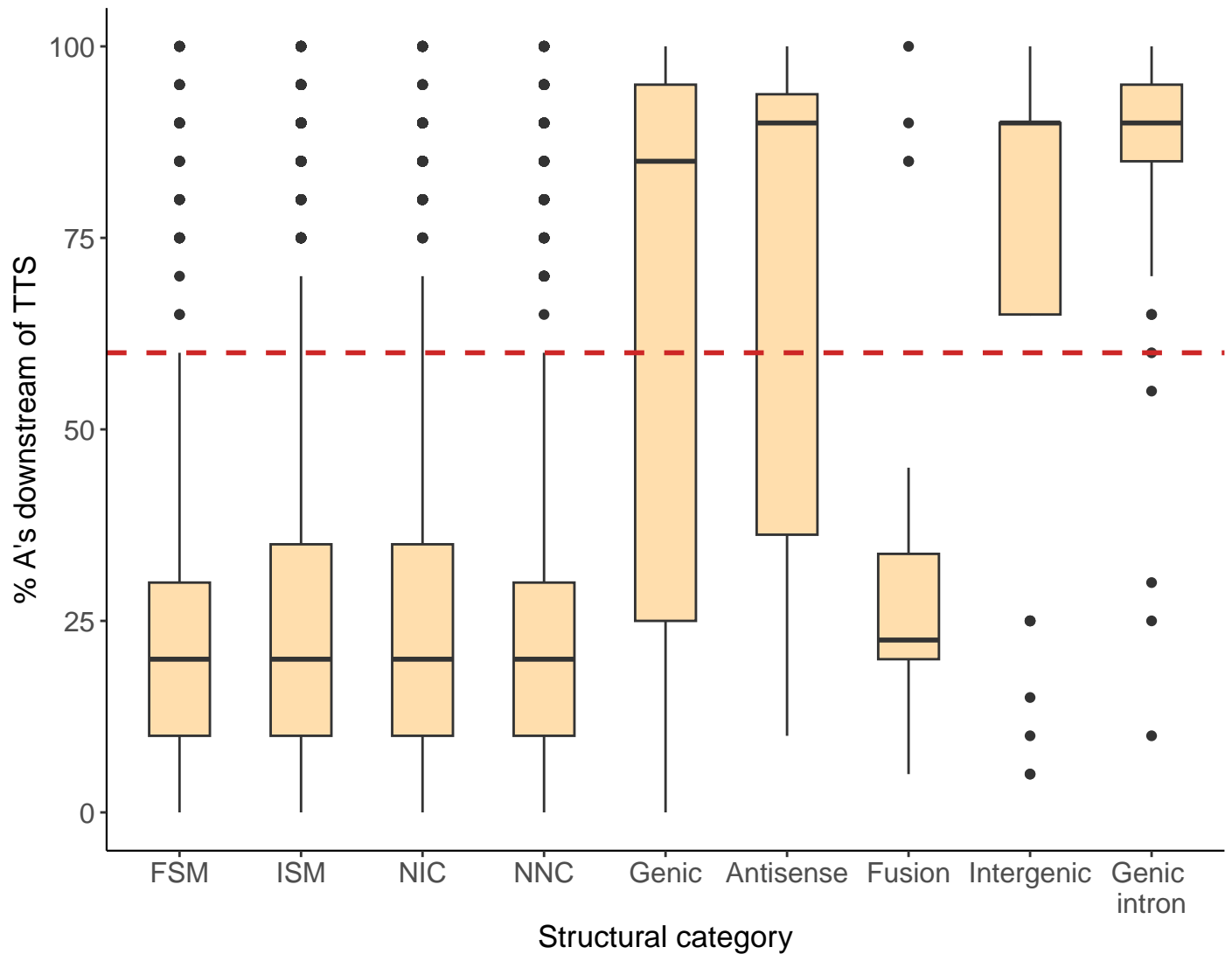
# coding – ML importance: 0.5



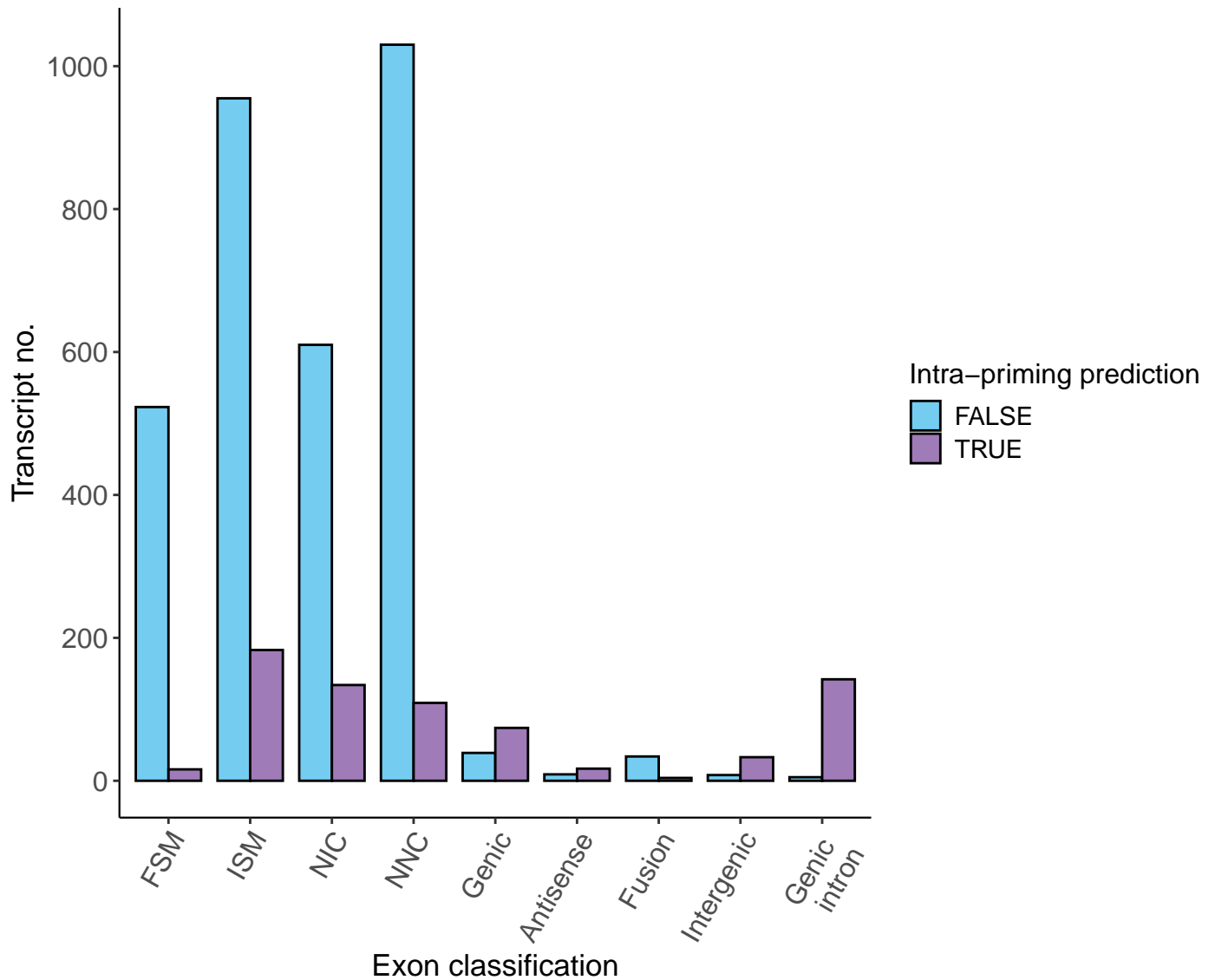
## *Intra-priming filter report*

## A % by category

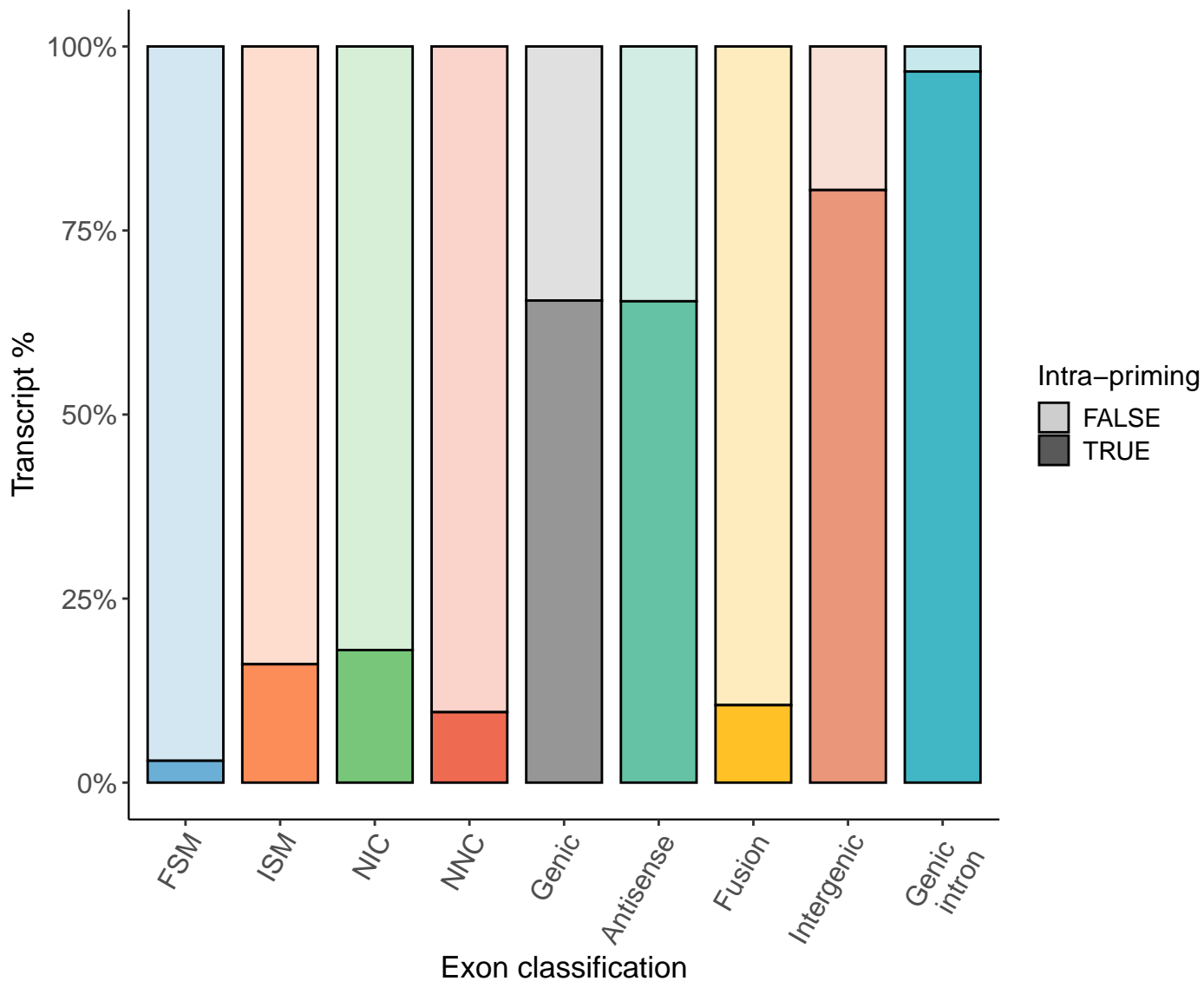
Red line indicates threshold employed in ML filter



Isoforms flagged as intra-priming, by category

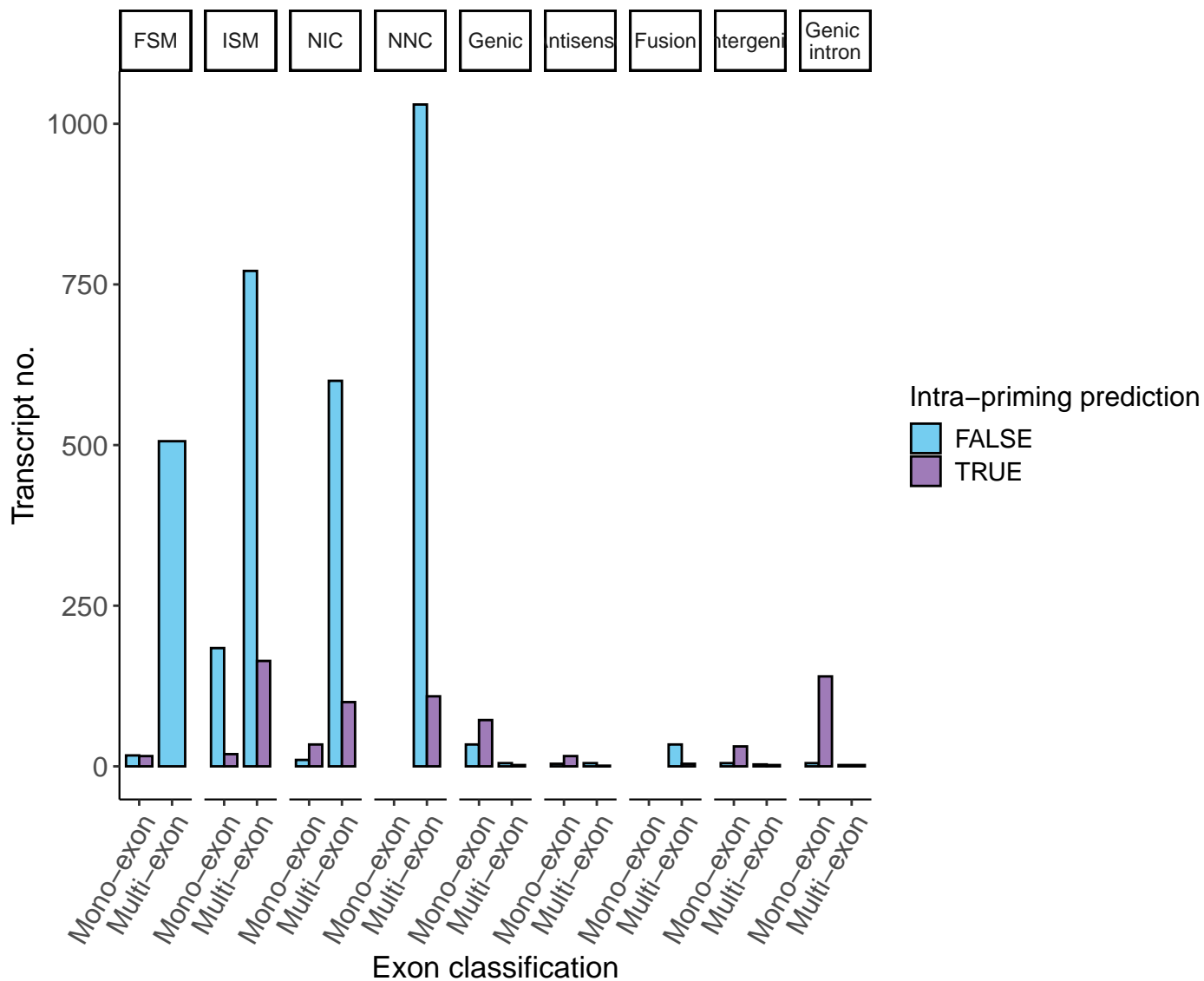


Isoforms flagged as intra-priming, by category (%)





Isoforms flagged as intra-priming, by exon number



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

