Ribolog benchmark

Hossein Asgharian; Sohit Miglani; Hani Goodarzi

Sep. 2022

In this document, we compare the capabilities and performance metrics of Ribolog with four alternative methods: Xtail, RiboDiff, Riborex and Anota2seq.

1. Catalogue of functionalities

Table 1 list the main tests and tools offered by Ribolog and whether they are included by four alternative methods for TER testing and one alternative method for stalling bias characterization.

Table 1. Main capabilities of Ribolog compared with five other ribo-seq packages.

Function	Ribolog	Xatil	Riborex	RiboDiff	Anota2seq	RUST
Stalling bias detection	Yes					Yes
Stalling bias correction	Yes					
Codon-level RPF count tabulation and visualization	Yes					Yes
Replicate-level QC	Yes					
Statistical basis of differential translation test	Logistic regression (binomial)	Negative binomial	DESeq2, EdgeR (Negative binomial); Voom (Weighted log-linear)	Negative binomial	Negative binomial	
Test possible with one sample per condition (non-replicated data)	Yes	Yes				
Allows for complex experimental design (multiple covariates, batch indicators, interaction terms, etc.)	Yes				Yes	
Significance testing with empirical null distribution	Yes					
Meta-analysis	Yes					
Upstream ORF and stop codon readthrough testing	Yes					

This table is not exhaustive. For a more comprehensive review of computational tools for ribosome profiling data visualization, pre-processing and analysis, please see:

Kiniry, S.J., Michel, A.M. and Baranov, P.V., 2020. Computational methods for ribosome profiling data analysis. Wiley Interdisciplinary Reviews: RNA, 11(3), p.e1577.

2. Performance metrics and run time

We used the datasets simulated by the authors of Xtail, provided in the supplementary information of the Xiao et al. 2016 paper. The experimental design included two conditions (control and treatment). We ran Ribolog, Xtail, RiboDiff, Riborex and Anota2seq on the 2 vs. 2 and the 3 vs. 3 replicated datasets. Unless explicitly mentioned, all reported results pertain to the 3 vs. 3 comparison.

2.1. p-values and TERs

Figure 1 shows that all tested methods generate the expected shape for p-value distribution with a uniform base and a peak towards zero - enriched in differentially translated genes. Tables 2 and 3 demonstrate generally high concordance among all methods except for anota2seq.

Figure 1. Distribution of unadjusted p-values for 3 replicates

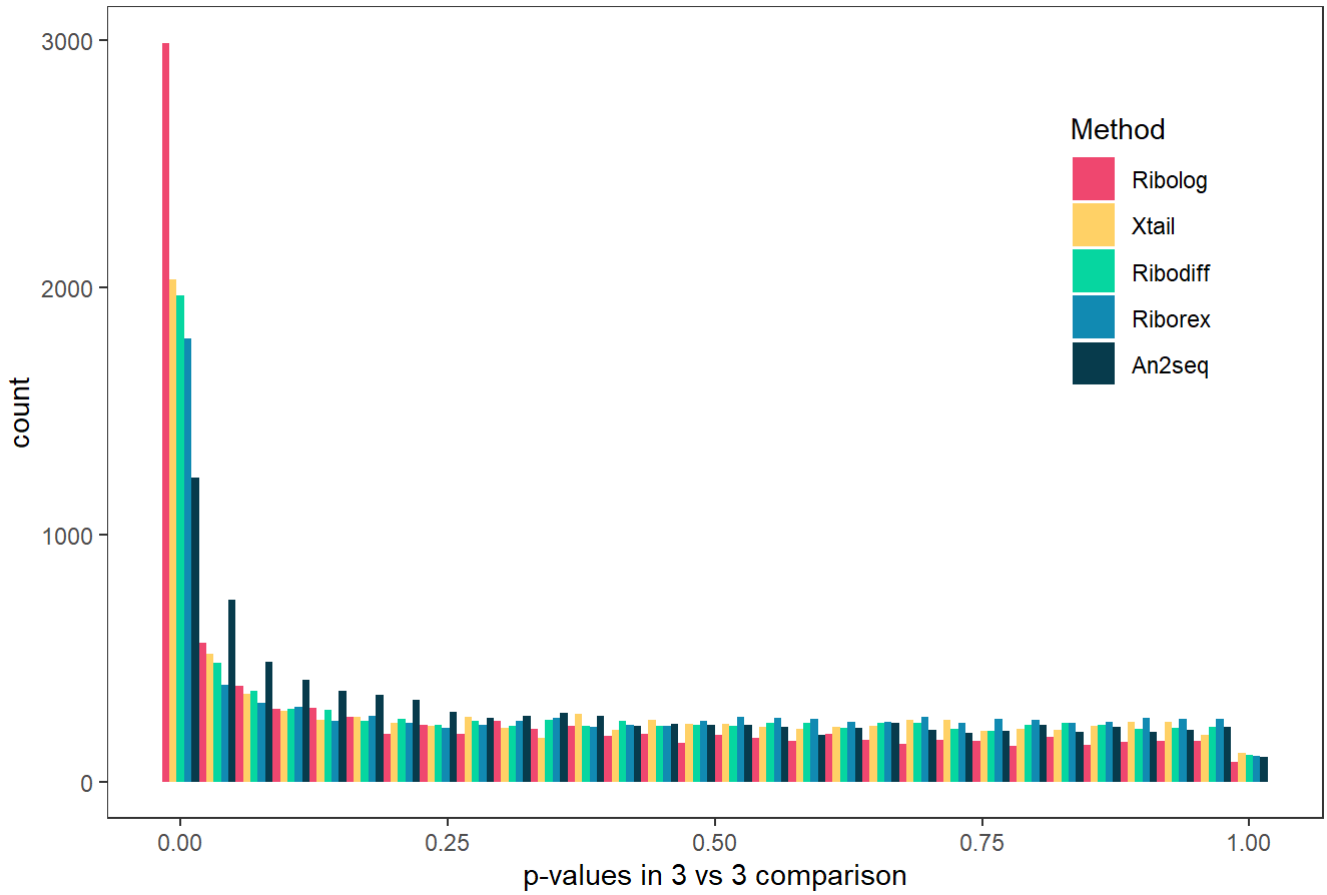


Table 2. Correlation matrix of logTER values by the 5 methods

	Ribolog	Xtail	Riborex	Ribodiff	An2seq
Ribolog	NA	0.9988168	0.9976505	0.9999996	0.4042822
Xtail	0.9988168	NA	0.9996866	0.9988500	0.4049293
Riborex	0.9976505	0.9996866	NA	0.9976989	0.4040311
Ribodiff	0.9999996	0.9988500	0.9976989	NA	0.4042891
An2seq	0.4042822	0.4049293	0.4040311	0.4042891	NA

Table 3. Correlation matrix of adjusted p-values by the 5 methods

	Ribolog	Xtail	Riborex	Ribodiff	An2seq
Ribolog	NA	0.9520433	0.9366877	0.9336835	0.5187760

	Ribolog	Xtail	Riborex	Ribodiff	An2seq
Xtail	0.9520433	NA	0.9896866	0.9852312	0.5606464
Riborex	0.9366877	0.9896866	NA	0.9959878	0.5761905
Ribodiff	0.9336835	0.9852312	0.9959878	NA	0.5626667
An2seq	0.5187760	0.5606464	0.5761905	0.5626667	NA

2.2. Precision-recall and ROC curves

Table 4. Receiver operating characteristic curve (ROC) and precision-recall curve AUC for the 5 methods

Method	ROC_AUC	PR_AUC
An2seq	0.7805404	0.9080005
Ribodiff	0.9541419	0.9835671
Ribolog	0.9606422	0.9867449
Riborex	0.9565605	0.9845953
Xtail	0.9581297	0.9853938

Figure 2. ROC curve for the 5 methods

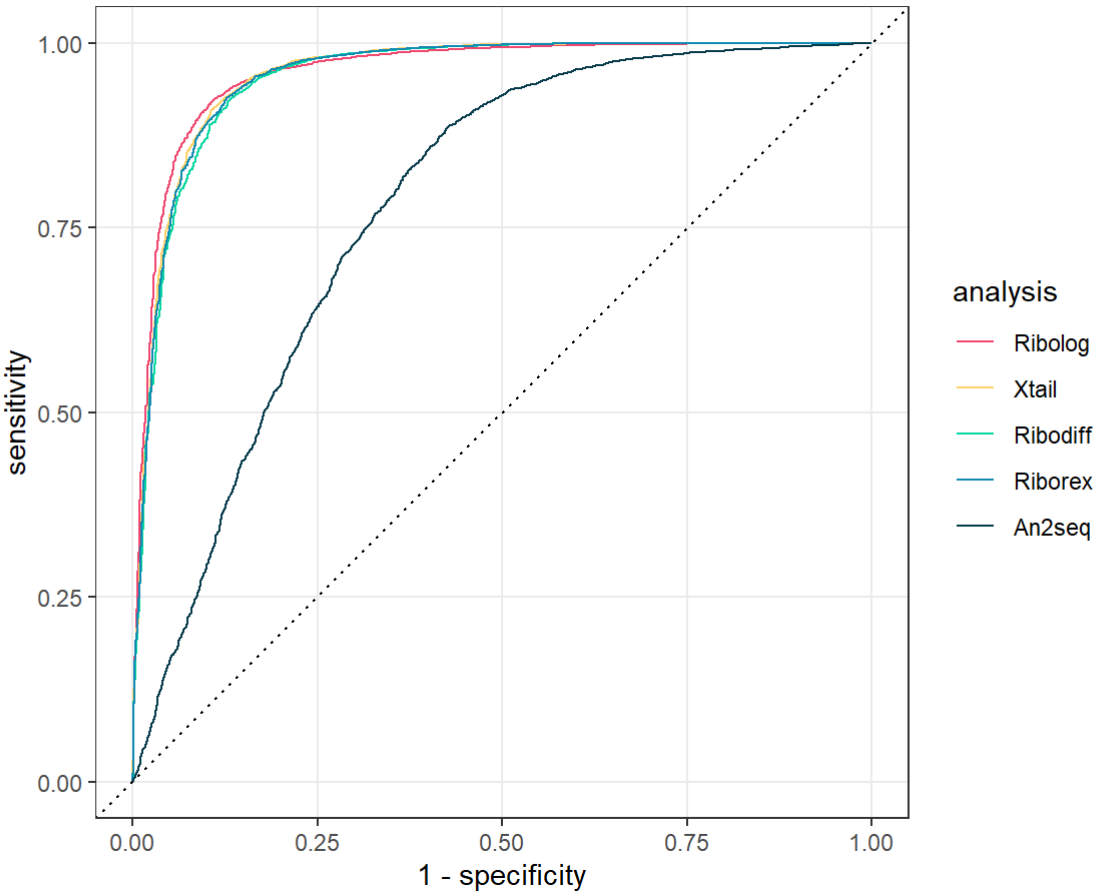
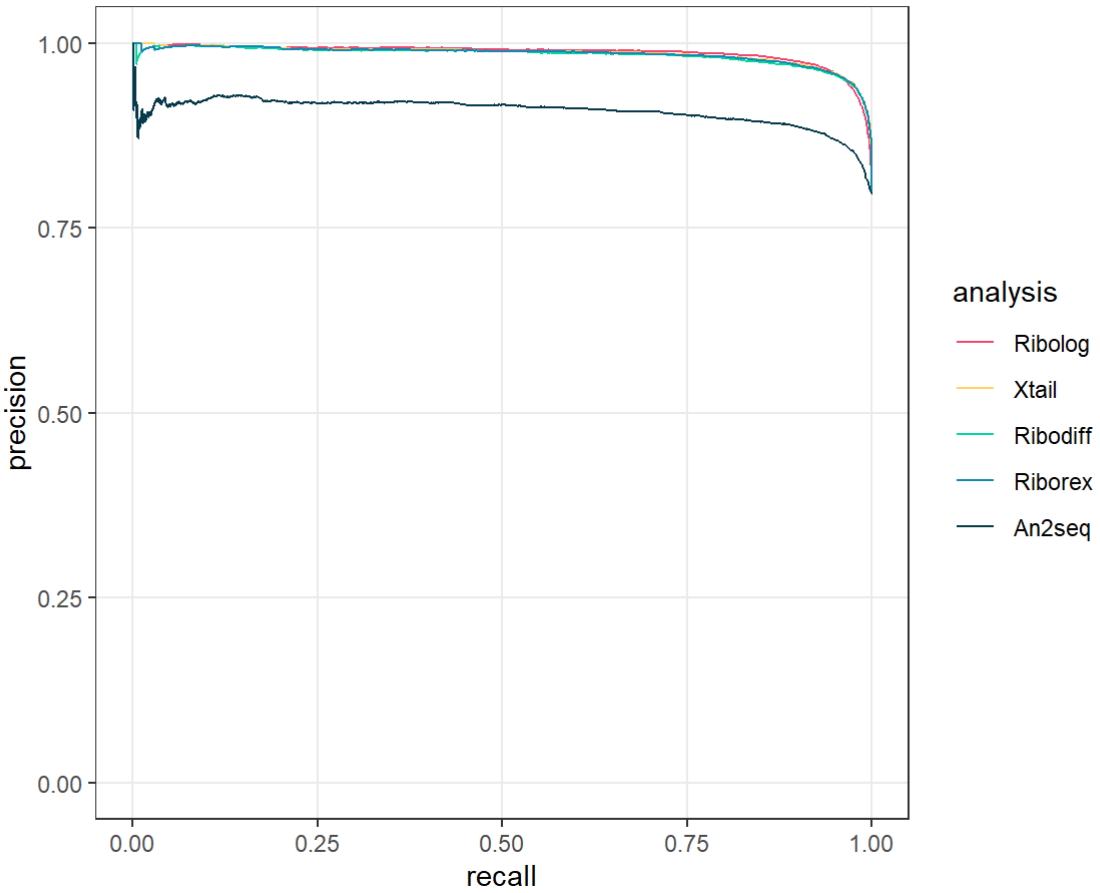


Figure 3. Precision-recall curve for the 5 methods



2.3. Robustness to sequencing coverage variation

We randomly subsampled reads to bring the total read count of each sample to 1X, 1/2X, 1/5X and 1/10X of its original total read count. We ran them through the usual analysis pipeline for each method and recorded the performance metrics. The 3 vs 3 tests were run for all methods, and the 2 vs 2 for Ribolog and Xtail. Figures 4-7 show that by sacrificing ~5% specificity, Ribolog retains much of its sensitivity even at extremely low coverages. For example, at 0.1X coverage, sensitivity of 3 vs 3 tests drops from 0.90 to 0.63 for Ribolog, and from 0.83 to 0.29 for Xtail.

Figure 4. Retention of sensitivity against decreasing coverage, 3 vs 3 tests.

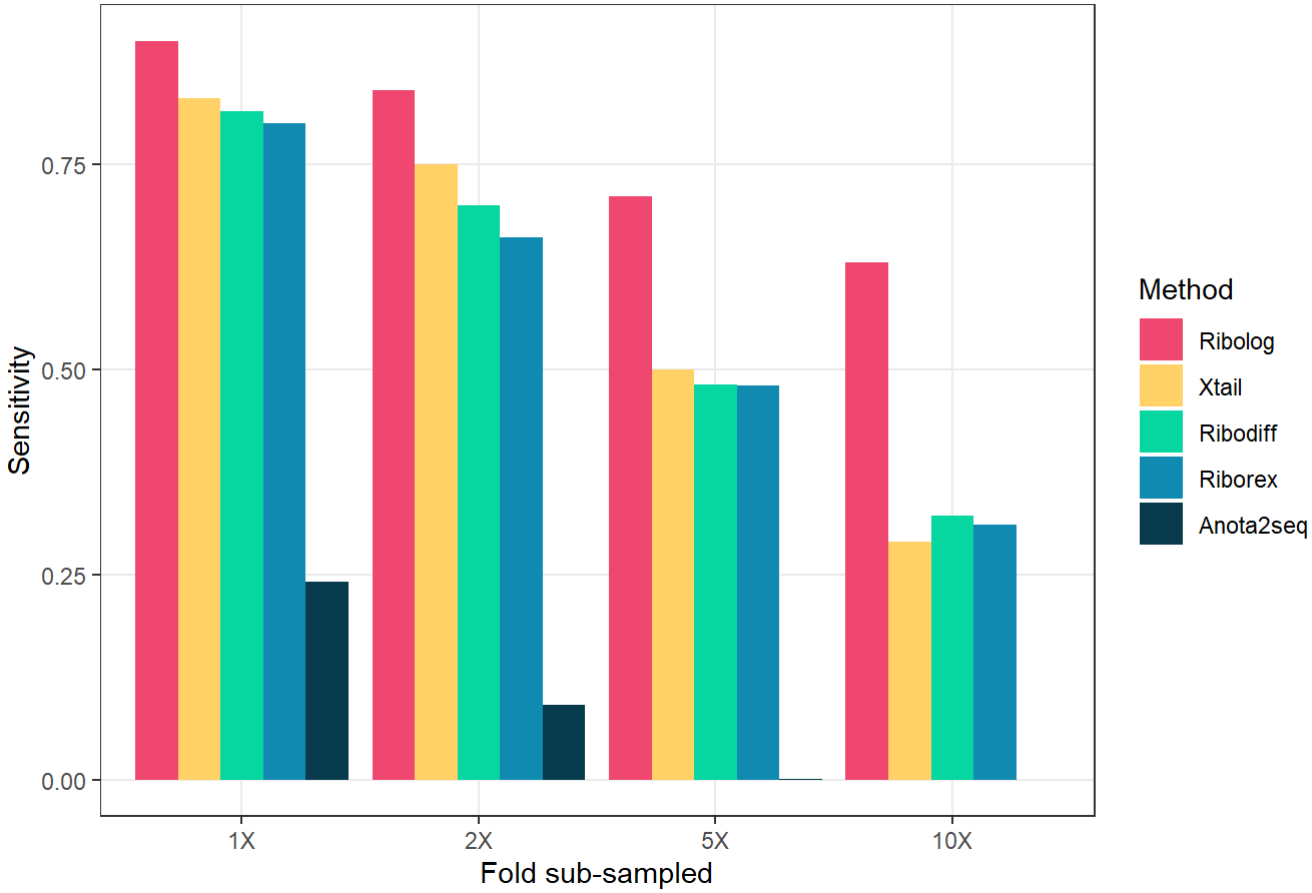


Figure 5. Retention of sensitivity against decreasing coverage, 2 vs 2 tests.

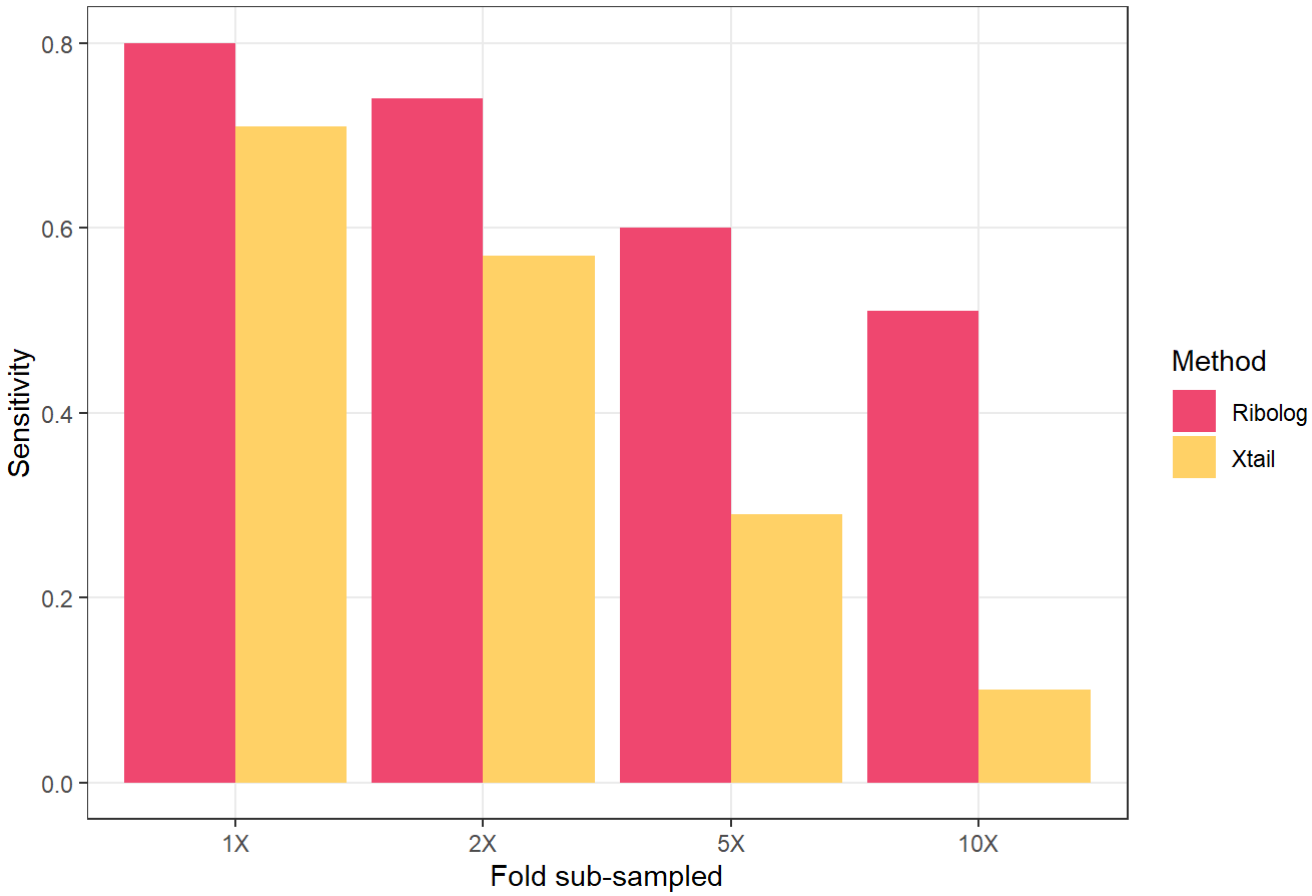


Figure 6. Retention of specificity against decreasing coverage, 3 vs 3 tests.

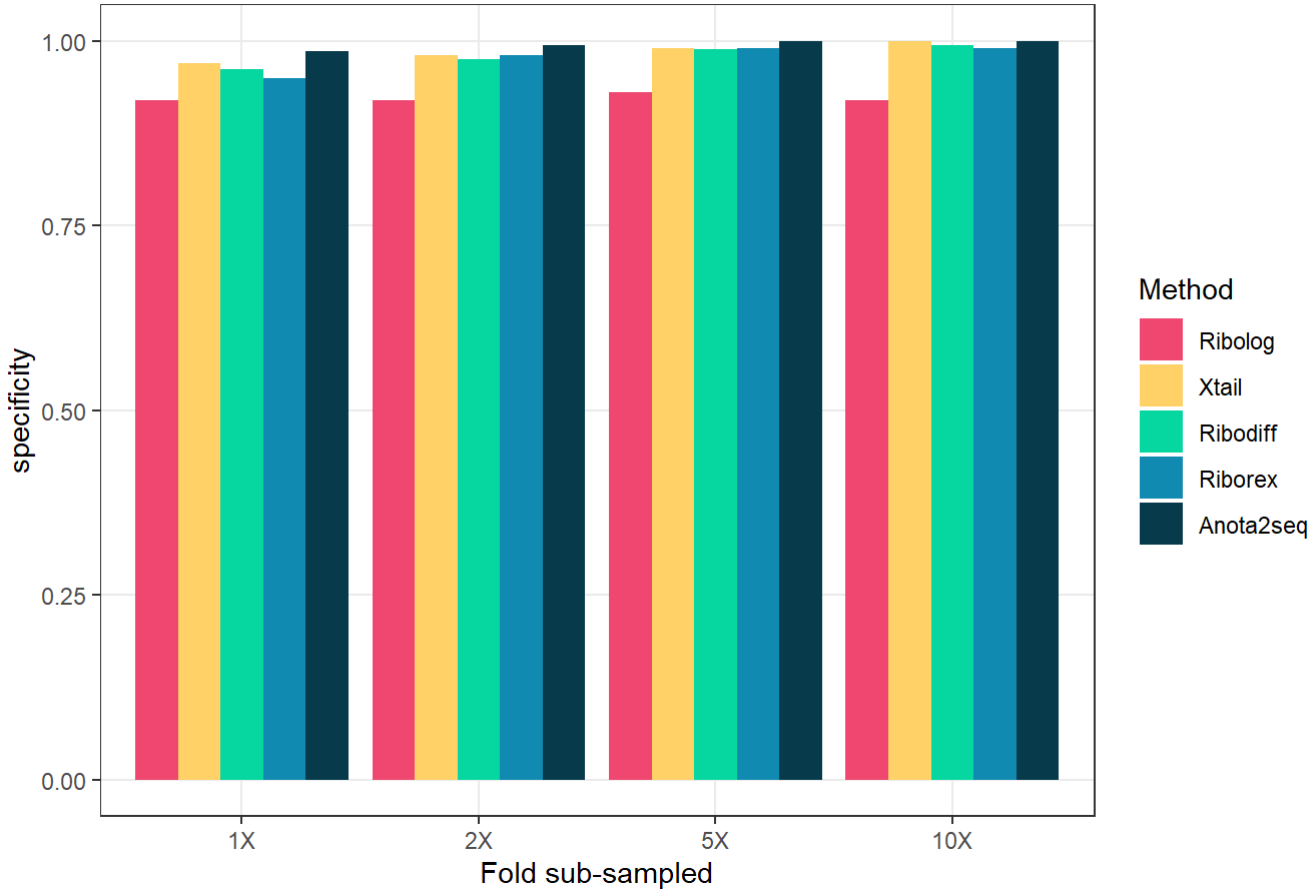
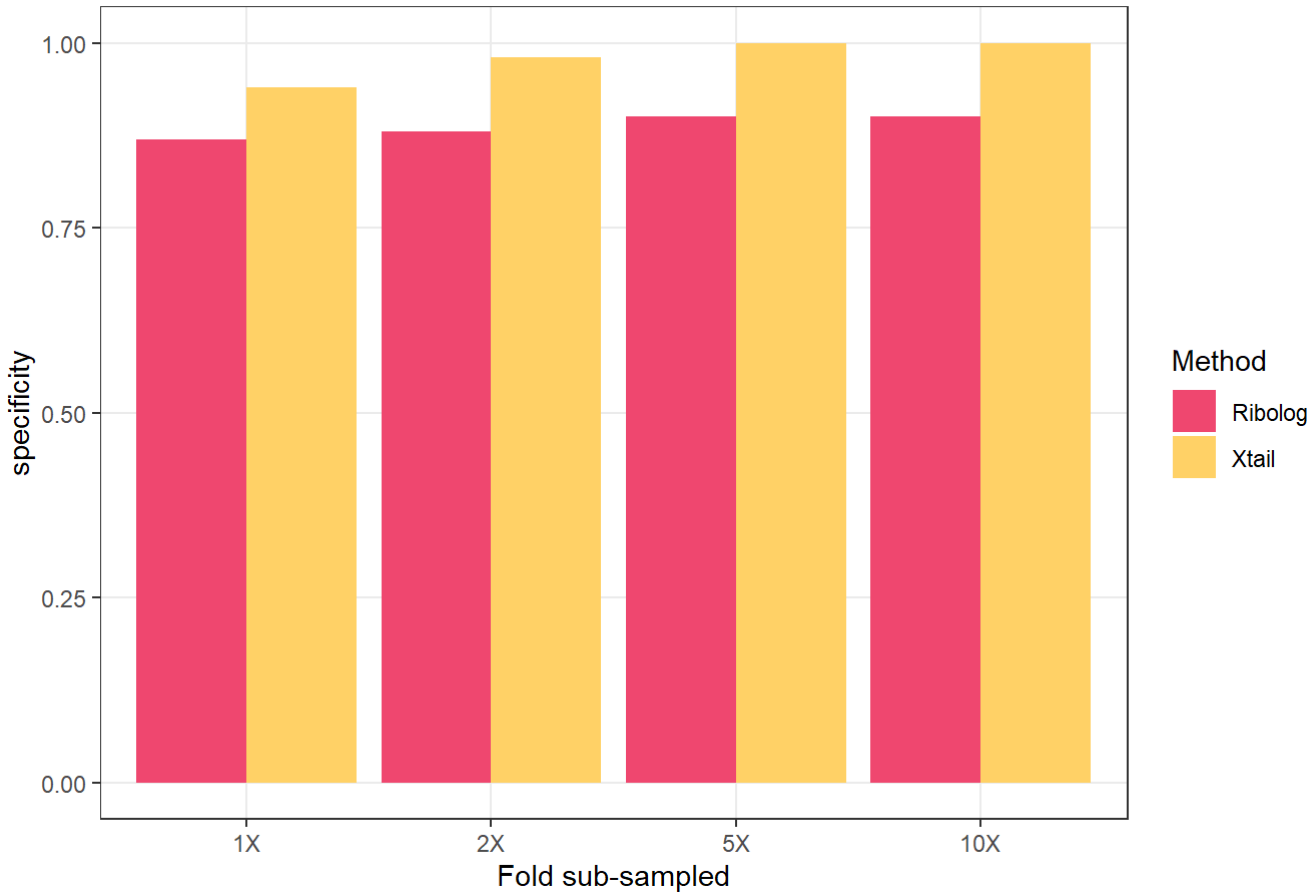


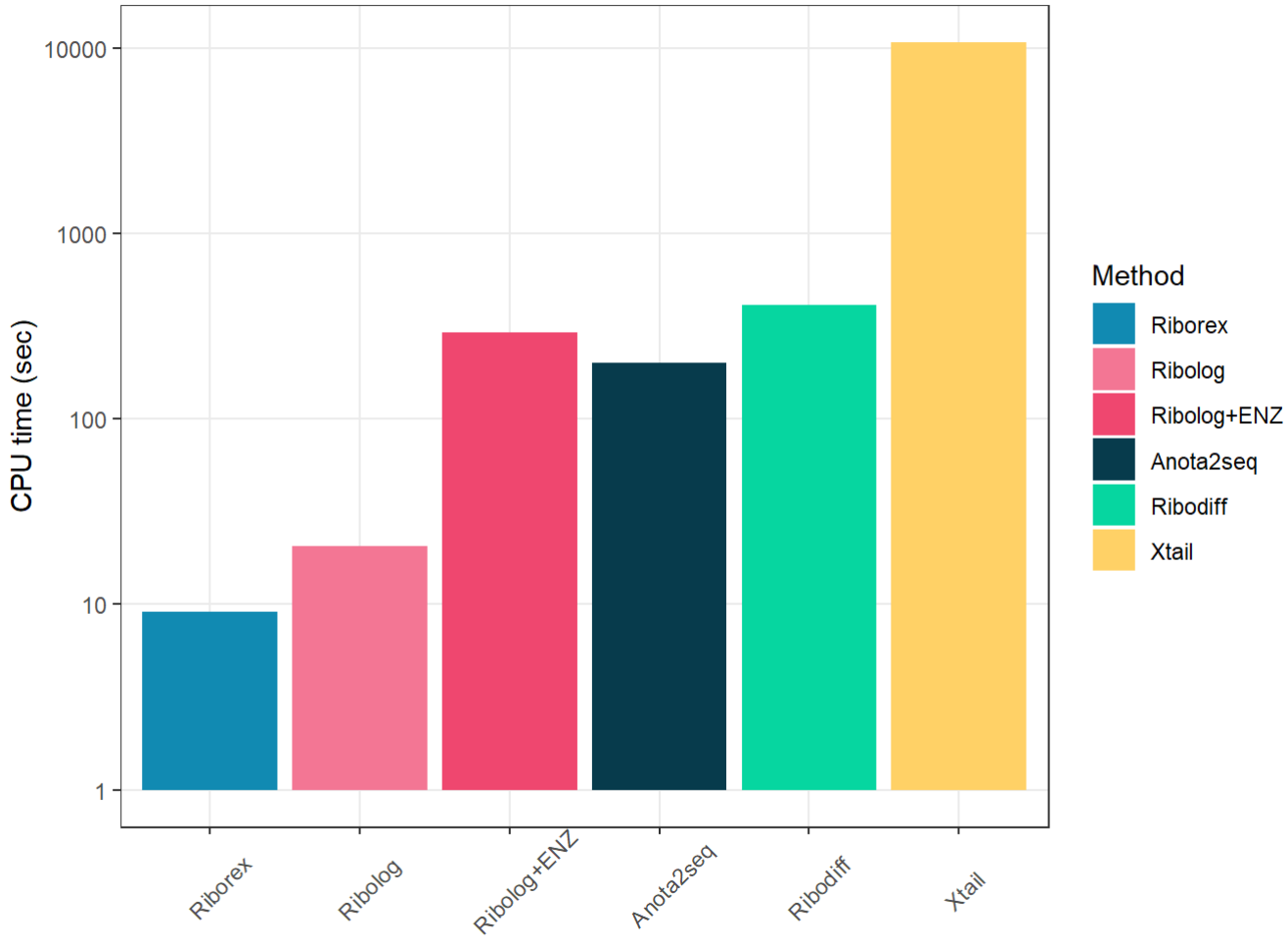
Figure 7. Retention of specificity against decreasing coverage, 2 vs 2 tests.



2.4. Run times

Even with the empirical null testing - which improves specificity tremendously but adds to processing time - Ribolog runs as fast as or faster than other methods, except for Riborex.

Figure 8. Run time of 3 vs 3 tests for Ribolog with and without empirical null testing, compared with four alternative methods. Note: The y axis is on log scale.



2.5. Effect of stalling bias correction with CELP

The CELP (Consistent Excess of Loess Predictions) module of Ribolog detects and corrects stalling bias and other types of local RPF read density nonuniformity in ribo-seq data. To evaluate the impact of this correction, we artificially multiplied the read counts at one codon (“GTC”) by 10 in ~300 genes of three MDA cell line samples. Then, we ran Ribolog and Xtail to compare the two groups (original MDA vs GTC-stalled MDA). Uncorrected, these simulated stalling reads were expected to contribute to the total read count in one group of samples and yield inflated TER estimates. Figures 9 and 10 shows that the application of CELP significantly reduces this bias compared with naive Ribolog (without CELP) and Xtail. Note that CELP was applied to the whole dataset here, without any prior knowledge of which genes, which samples or which codons were affected by the simulated bias.

Figure 9. Density plots of log2 TER
between original MDA vs GTC-stalled MDA samples.

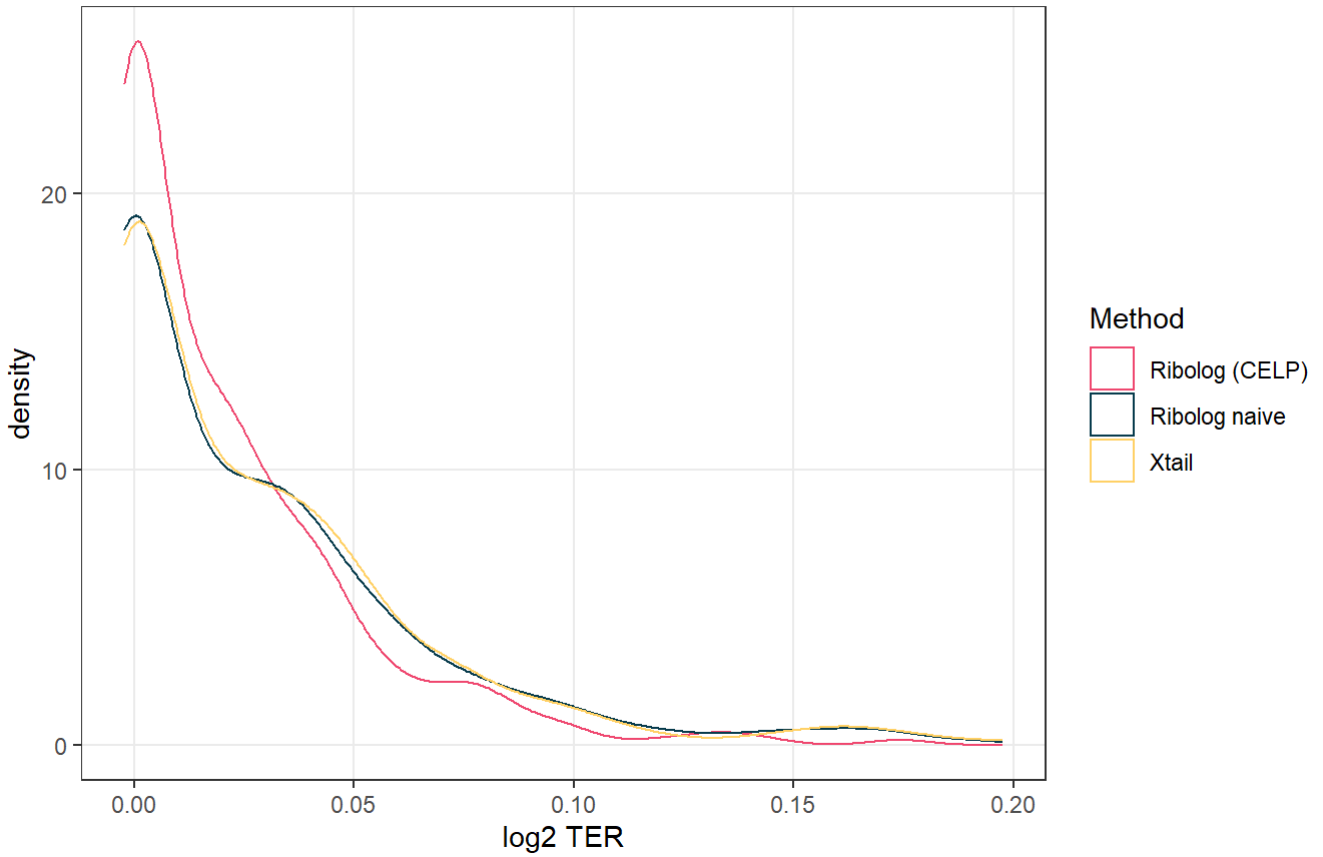


Figure 10. Box plots of log2 TER
between original MDA vs GTC-stalled MDA samples.
The p-values were produced by t tests comparing log2 TERs
from Xtail with those from naive Ribolog or Ribolog with CELP.