

Comparative evaluation of RNA-Seq library preparation methods for strand-specificity and low input

Dimitra Sarantopoulou¹, Soon Yew Tang², Emanuela Ricciotti², Nicholas F. Lahens¹, Damien Lekkas², Jonathan Schug³, Xiaofeng S. Guo⁴, Georgios K. Paschos², Garret A. FitzGerald^{1,2}, Allan I. Pack⁴, Gregory R. Grant^{1,5,*}

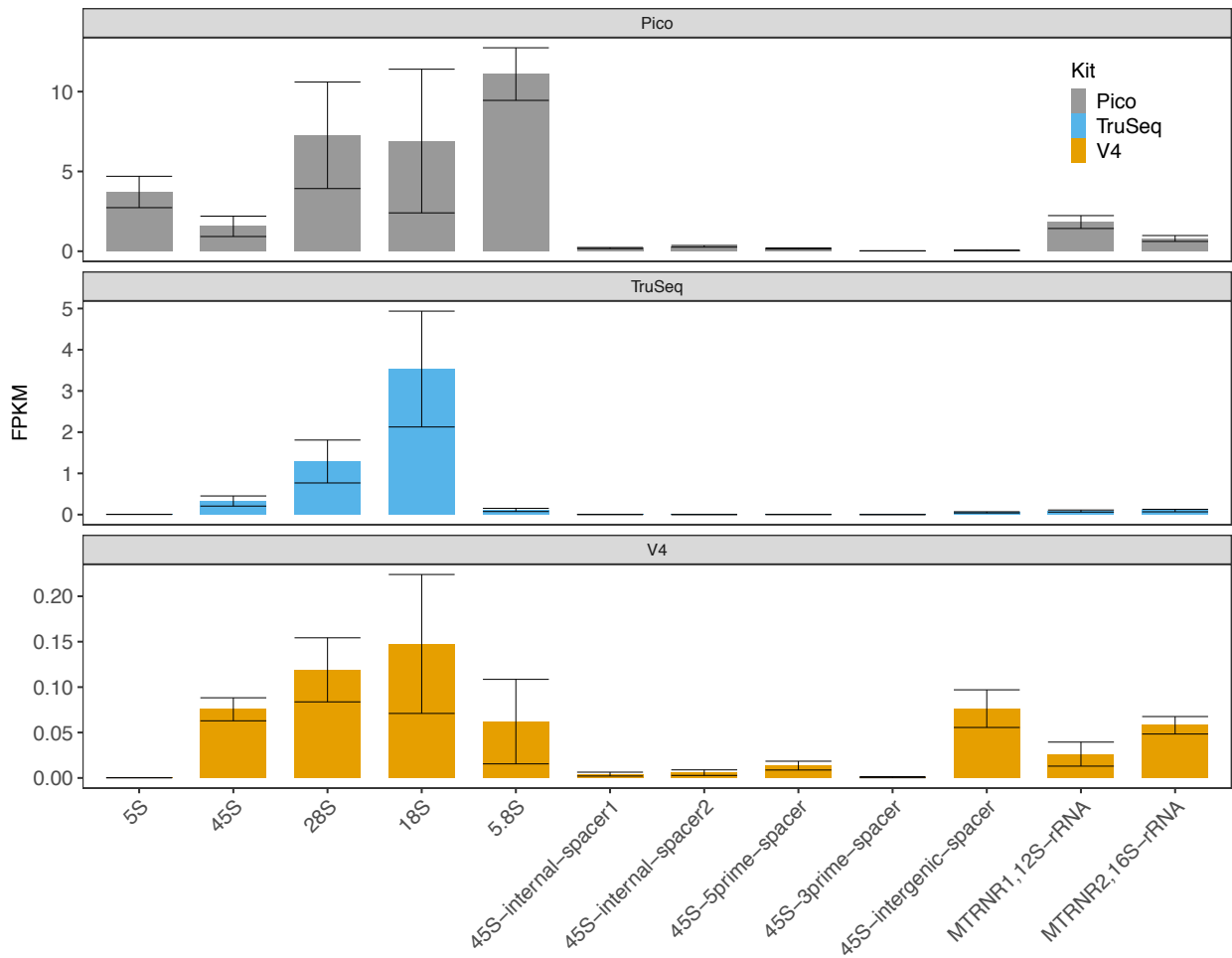
1. Institute for Translational Medicine and Therapeutics 2. Department of Systems Pharmacology and Translational Therapeutics 3. Next Generation Sequencing core 4. Center for Sleep and Circadian Neurobiology 5. Department of Genetics; All affiliations at University of Pennsylvania, Philadelphia, PA, USA

Supplemental Material – Table of contents

Supplemental Fig. S1 – Distribution of the retained ribosomal signal	2
Supplemental Fig. S2 - Retained intron signal in Pico	3
Supplemental Fig. S3 - Concordance of gene expression profiles	4
Supplemental Fig. S4 - Effects on differential expression	5

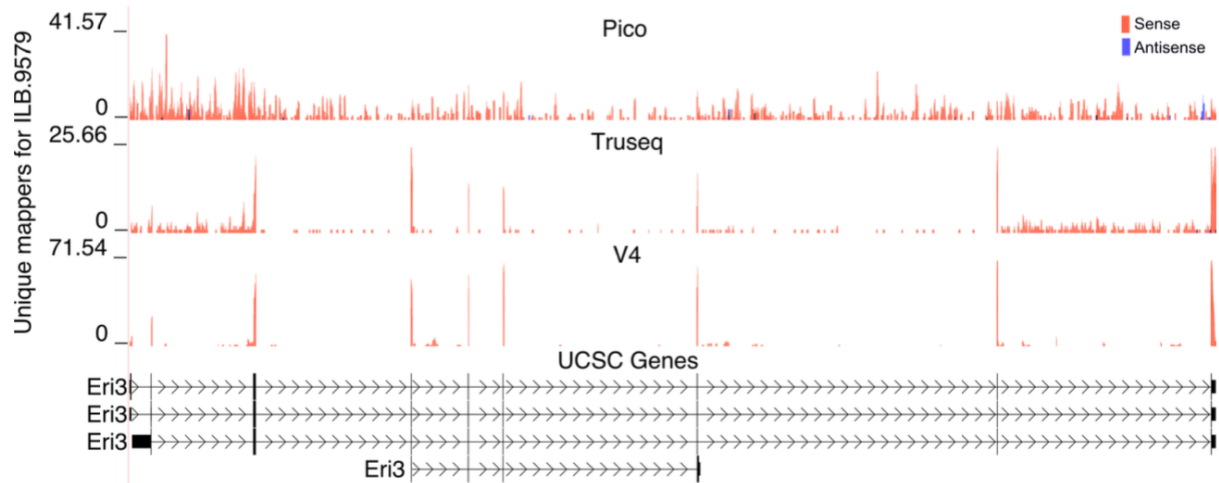
Supplemental Fig. S1 – Distribution of the retained ribosomal signal

The distribution of the FPKM normalized ribosomal retention among the different ribosomal RNA subunits (5S, 45S, 18S, 28S, 5.8S, 45S spacers, and mitochondrial rRNAs).



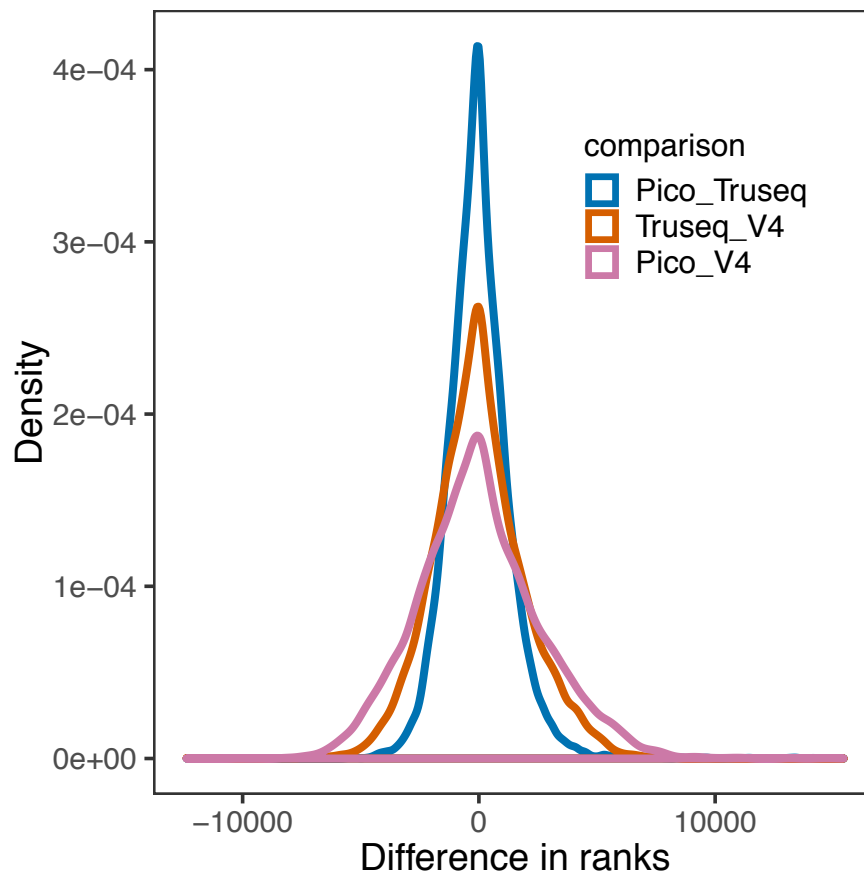
Supplemental Fig. S2 - Retained intron signal in Pico

Example of *Eri3* gene body coverage in one of the IL-1 β treated samples (ILB.9579) illustrates high intron signal retained in Pico.



Supplemental Fig. S3 - Concordance of gene expression profiles

The concordance of gene expression distribution between two kits is shown by the difference in ranks of two kits, sorting by average gene expression. The narrower the gene expression distributions are, the more concordant two kits are to each other. The X-axis represents the value of the difference in row ranks in linear scale.



Supplemental Fig. S4 - Effects on differential expression

a) As a control, a second round of TruSeq library prep, sequencing, and analysis were performed on the same samples. This heatmap displays the ratio of the number of DE genes found by both TruSeq datasets, to the number of genes found DE in at least one, at varying q -value cutoffs. We observe high overlap of DEGs at all significant cutoffs, which suggests that TruSeq identifies a similar set of DEGs in different sequencing runs. b) Absolute value of the $\log_2$ adjusted fold-change values for the set of non-DEGs at q -value > 0.3 . c) Coefficient of variation within each condition for the set of non-DEGs at q -value > 0.3 and mean expression per condition > 2 . The CVs of both conditions are summarized per kit. b-c) The error bars represent the 95% confidence intervals around the mean fold-change and CV values. d) Heatmap of the ratio of the number of DE genes found by both kits, to the number of genes found DE by at least one of the two kits, at varying q -value cutoffs, for the repeated DE analysis with DESeq2.

