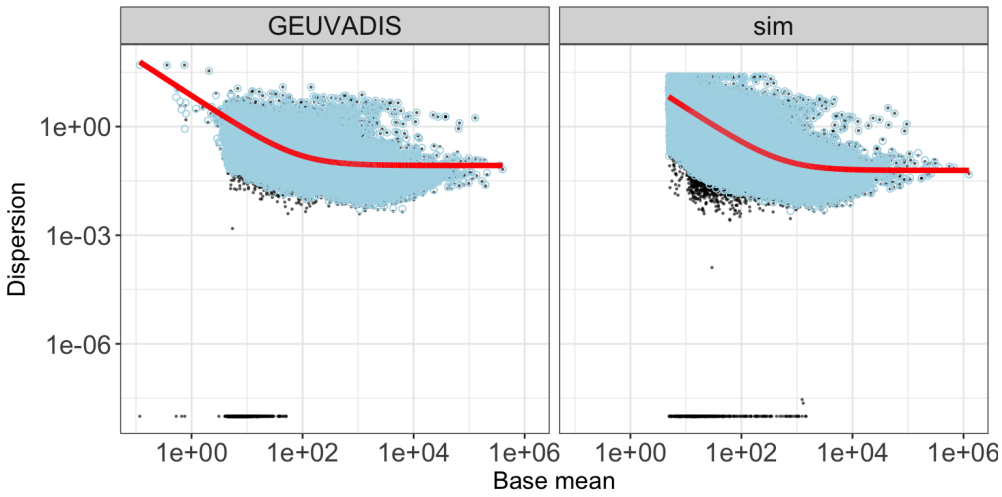
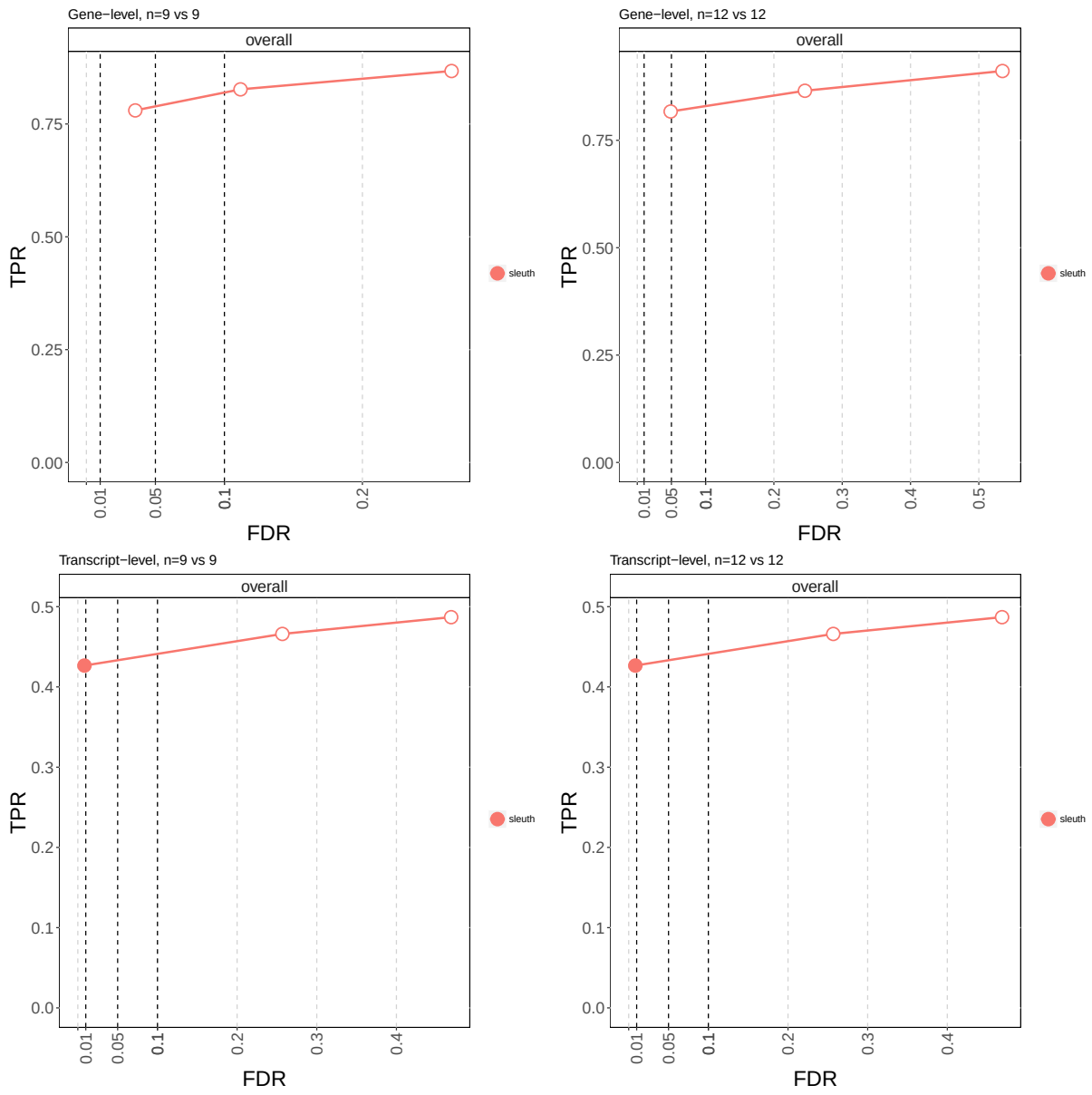
Supplementary Figures

Michael I. Love, Charlotte Soneson, Rob Patro

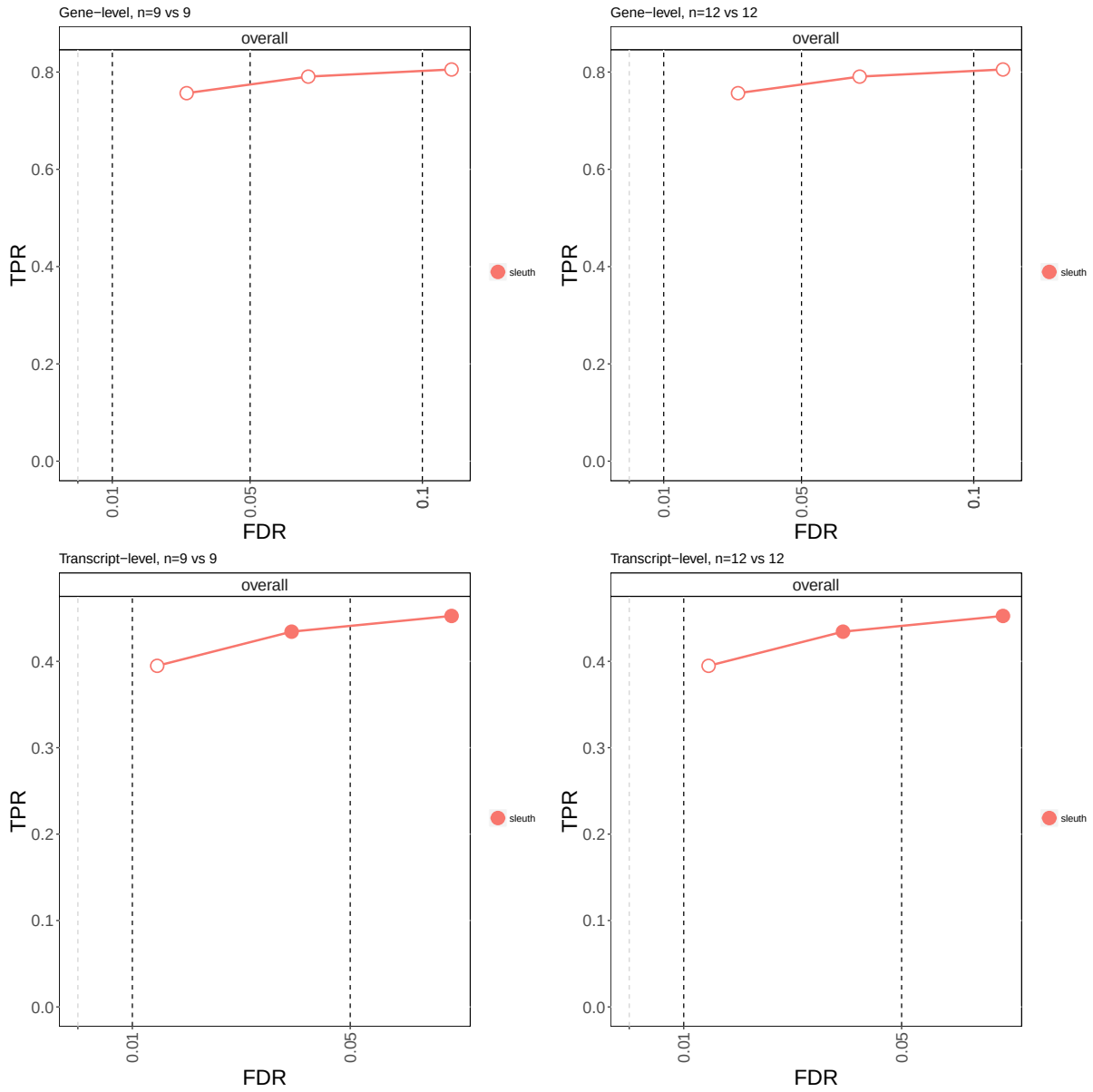
June 13, 2018



Supplementary Figure 1: Dispersion-over-mean comparison plot produced by *countsimQC*. The left panel shows *DESeq2* estimates of dispersion per gene over the mean of normalized counts from the GEUVADIS project, provided by the Recount2 project ($n = 458$ non-duplicated samples). The right panel shows estimates of dispersion per transcript over the mean of normalized counts for *Salmon* estimated transcript counts for the simulated dataset (the 12 vs 12 comparison), showing only the transcripts where the mean of counts over samples was greater than 5. Black points indicate maximum likelihood estimates (Cox-Reid adjusted), blue points indicate posterior estimates, and the red line indicates the parametric trend line. Points at the bottom of the plot indicate maximum likelihood estimates of 10^{-8}. The design formula included sequencing center and population for GEUVADIS, and the condition variable for the simulated dataset. The simulation dataset was constructed by drawing mean and dispersions parameters from the joint distribution of the estimates from the GEUVADIS project. The full *countsimQC* report can be found at https://github.com/mikelove/swimdown/tree/master/countsimqc.



Supplementary Figure 2: We performed additional experiments to assess the false discovery rate control for *sleuth* at per-group sample size of 9 (left column) and 12 (right column), at the gene-level (top row) and the transcript-level (bottom row). To determine whether the excess observed FDR was due to the inclusion of realistic fragment GC coverage in the main simulation, for this experiment fragments were instead drawn uniformly from positions on the transcripts. The dispersion-mean relationship was kept the same, drawing from the joint distribution of estimates on the GEUVADIS dataset ($n = 458$).



Supplementary Figure 3: As in Supplementary Figure 2, shown is the result of an additional experiment to assess the false discovery rate control for *sleuth* for the two largest sample sizes in the simulation. For this experiment, realistic fragment GC bias was used in the simulation, but the percent of genes with DGE, DTE and DTU was lowered from 10% to 5% each. This modification of the simulation helped to regain control of FDR for *sleuth*.