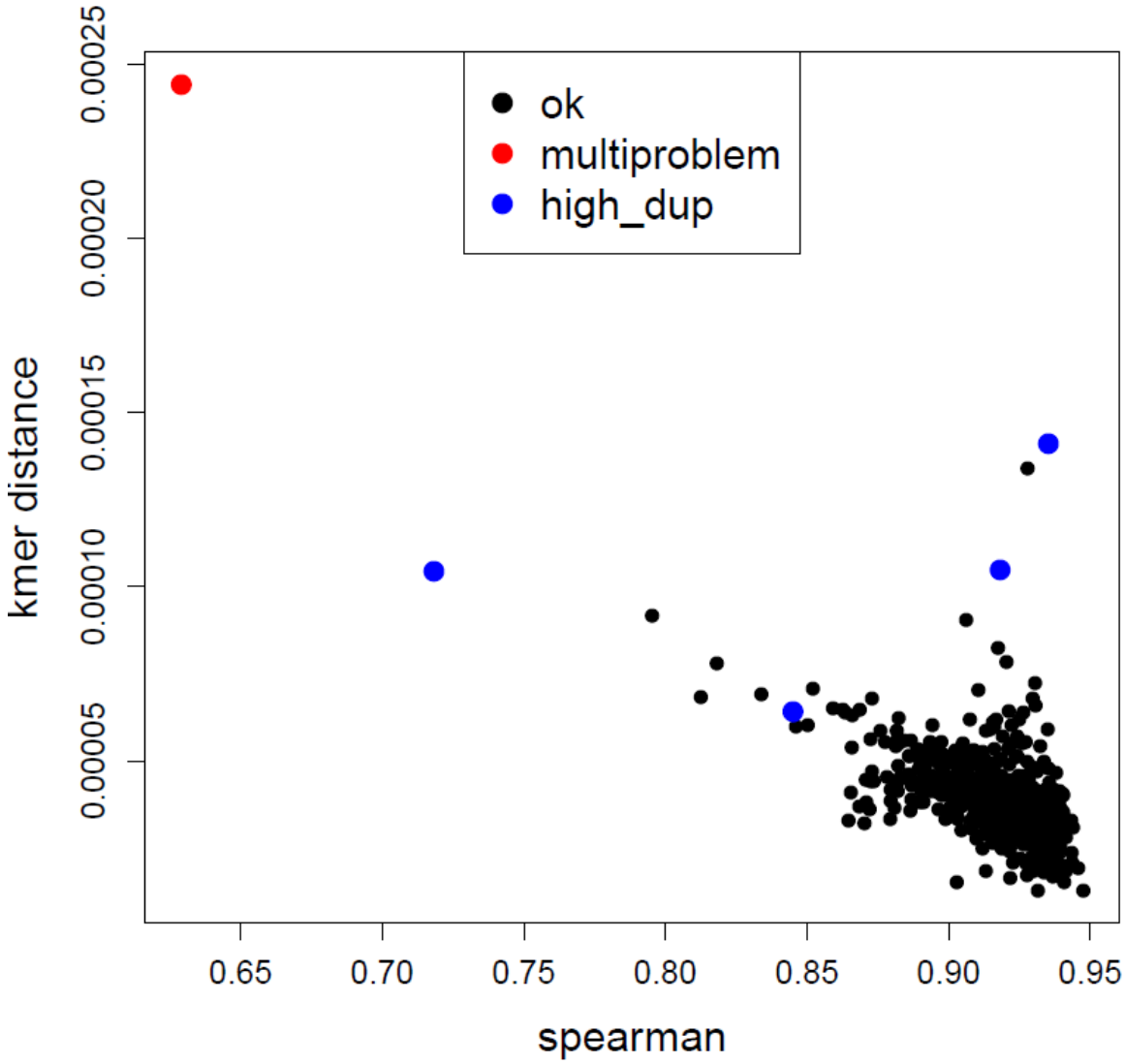
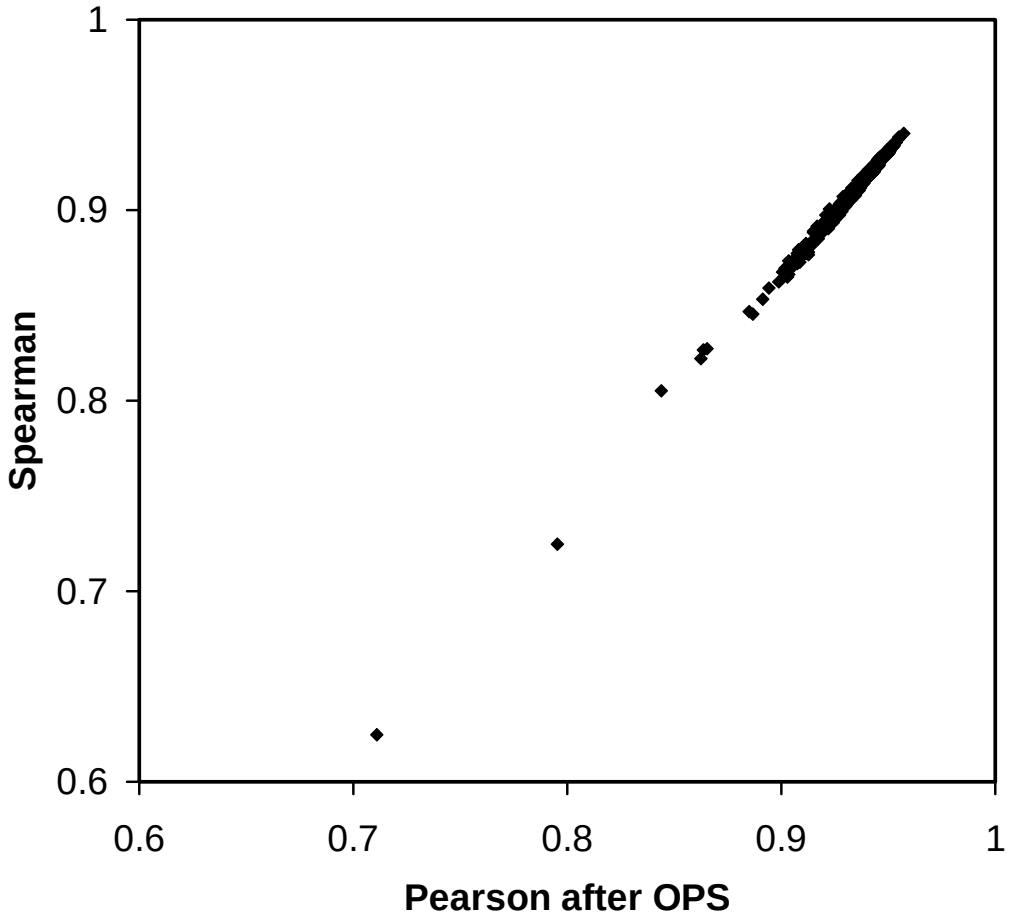
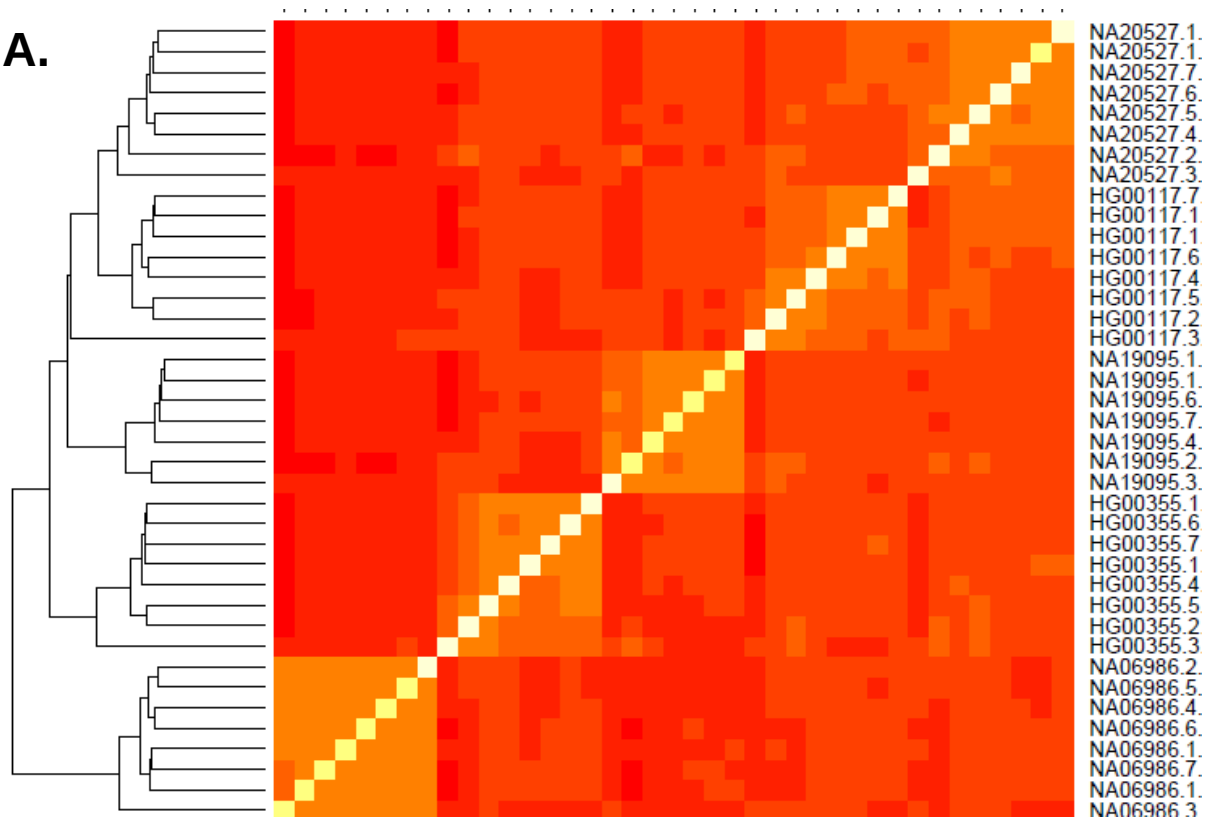
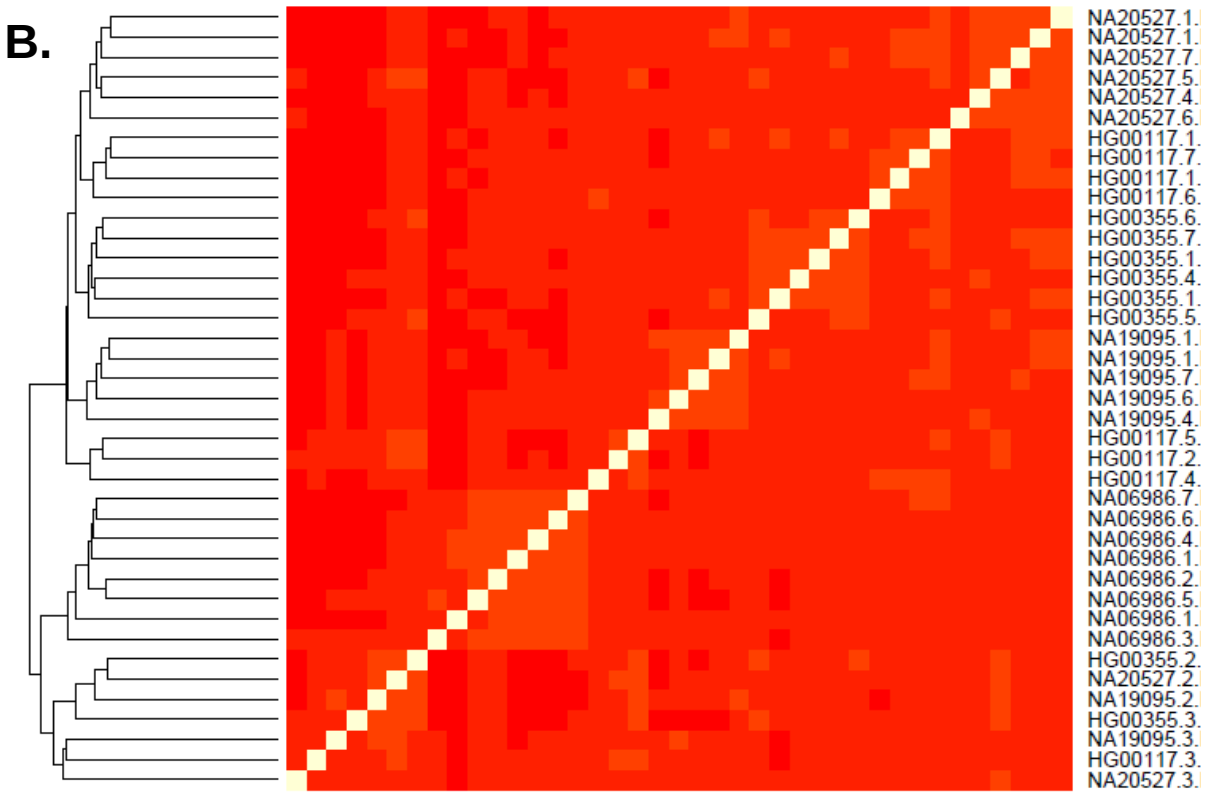


Suppl. Figure 1: Median of all pairwise Spearman correlations per sample based on exon quantifications (x-axis) plotted against median of k-mer distance (k=9) based on raw sequence reads. In red the sample NA18861.4 which failed all QC tests; in blue samples with high read duplication rates.



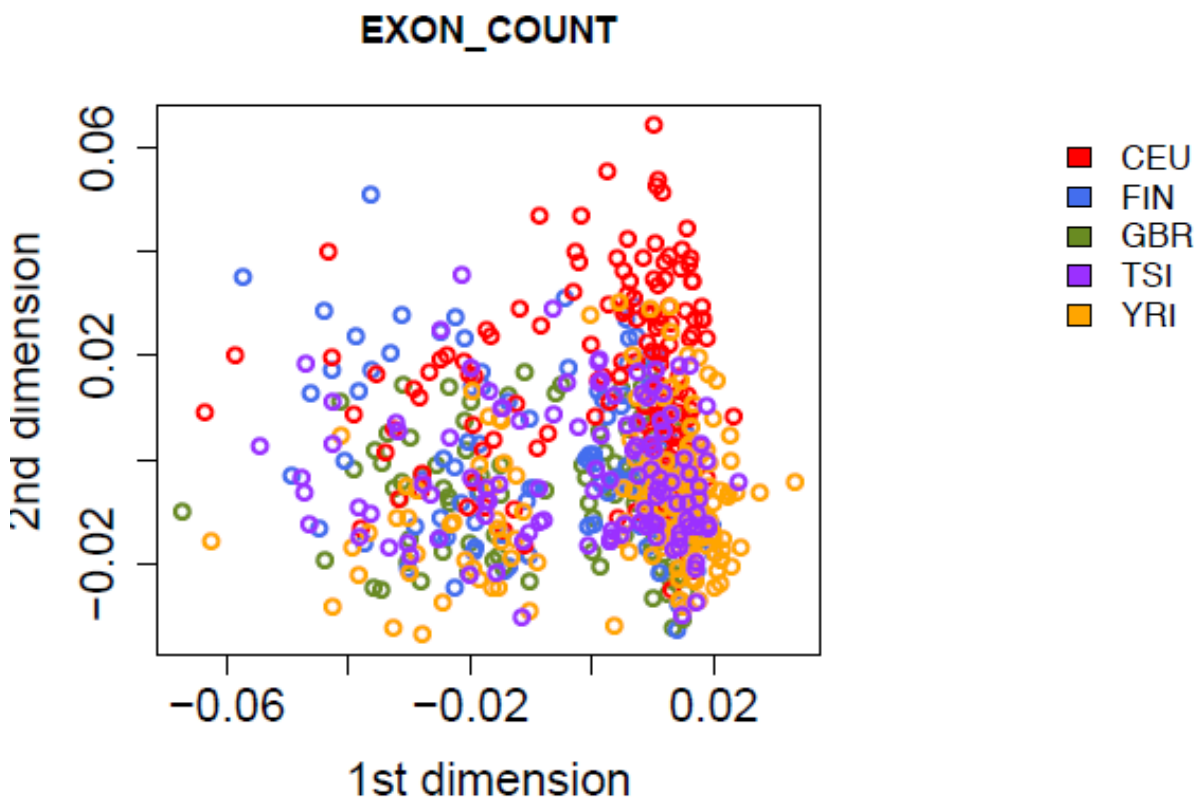
Suppl. Figure 2: D-statistics (median of all pairwise correlations per sample) for exon quantifications calculated from Pearson correlations after OPS transformation (i.e. raising all transcript quantifications to the power of 0.11) (x-axis) plotted against D-statistics calculated from Spearman correlations.



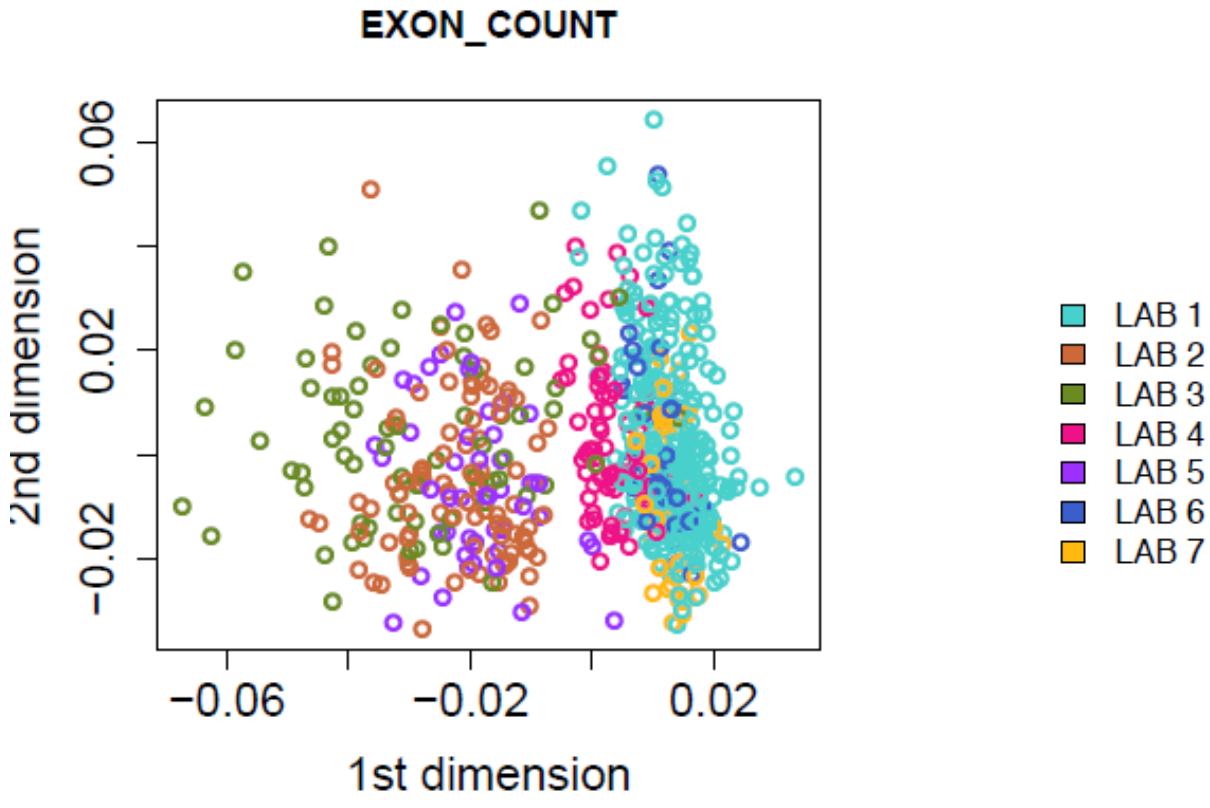


Suppl. Figure 3: Heatmap and clustering of Pearson correlations (after OPS transformation) for exon (A) and transcript (B) quantifications. White is highest correlation, yellow is intermediate, red is lowest. Samples are indicated with their HapMap identifier followed by their sequence laboratory identifier. Sequence runs in different laboratories generally cluster by sample rather than by laboratory. The intra-sample correlations are higher for exon than for transcript quantifications.

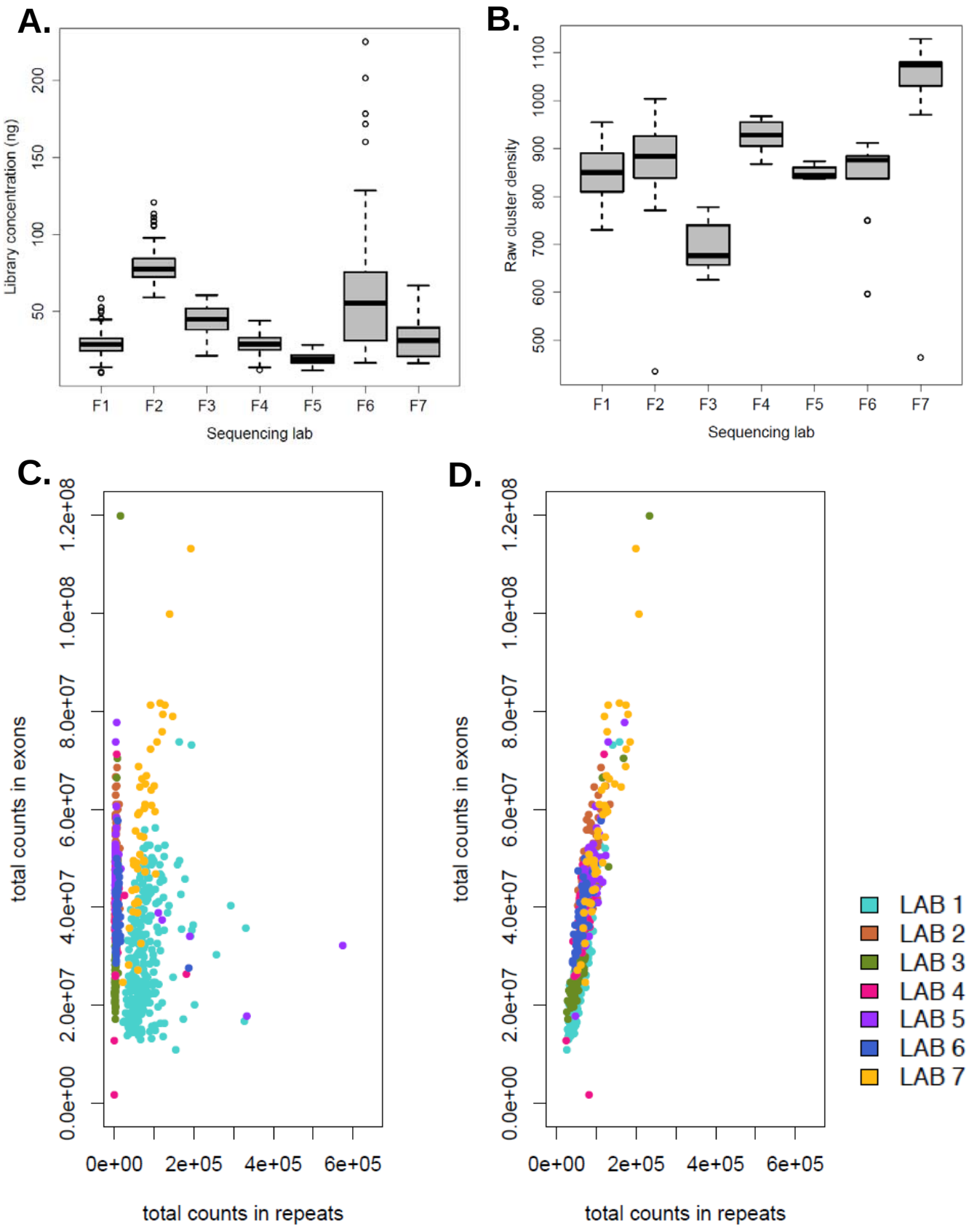
A.



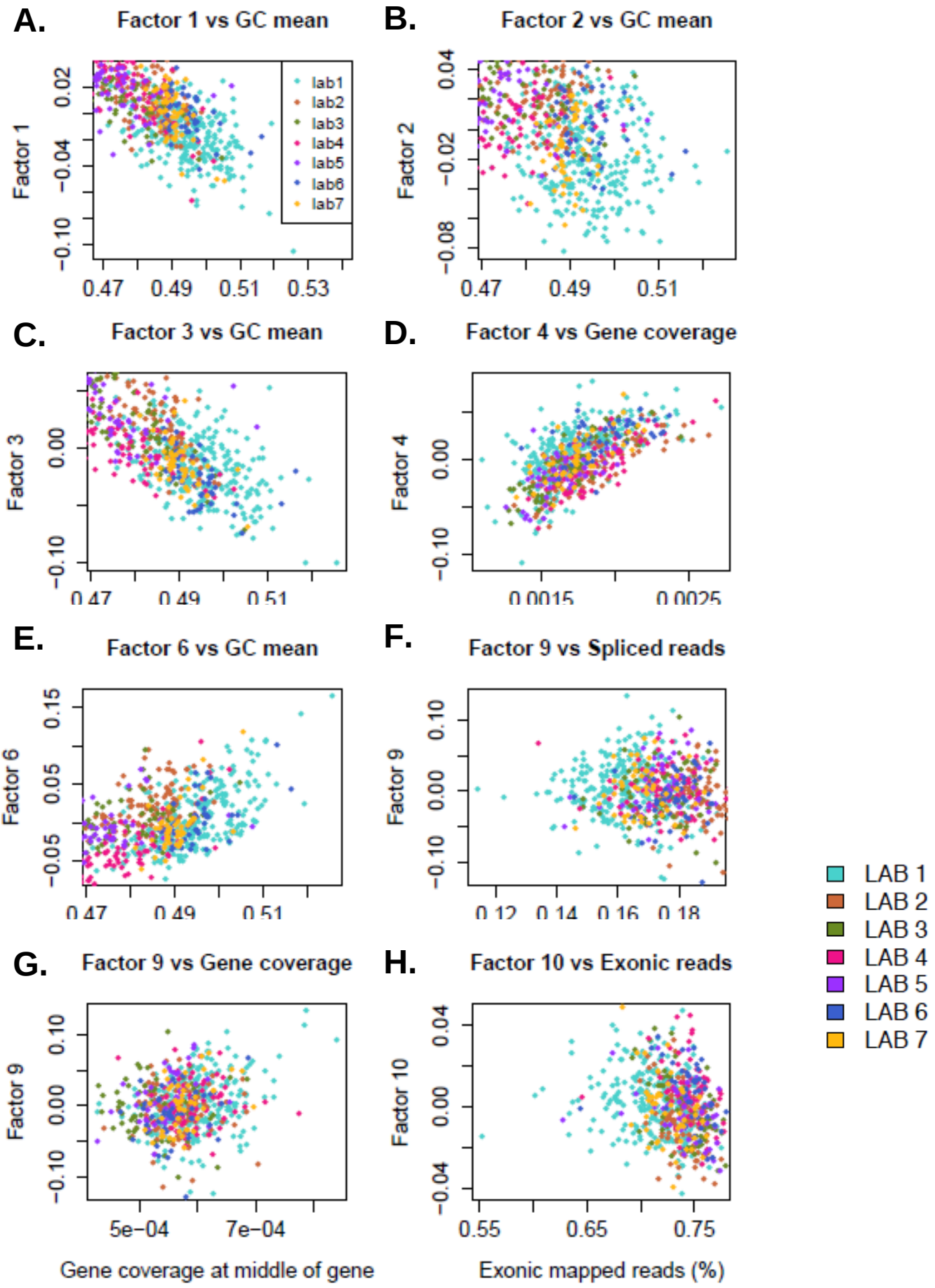
B.



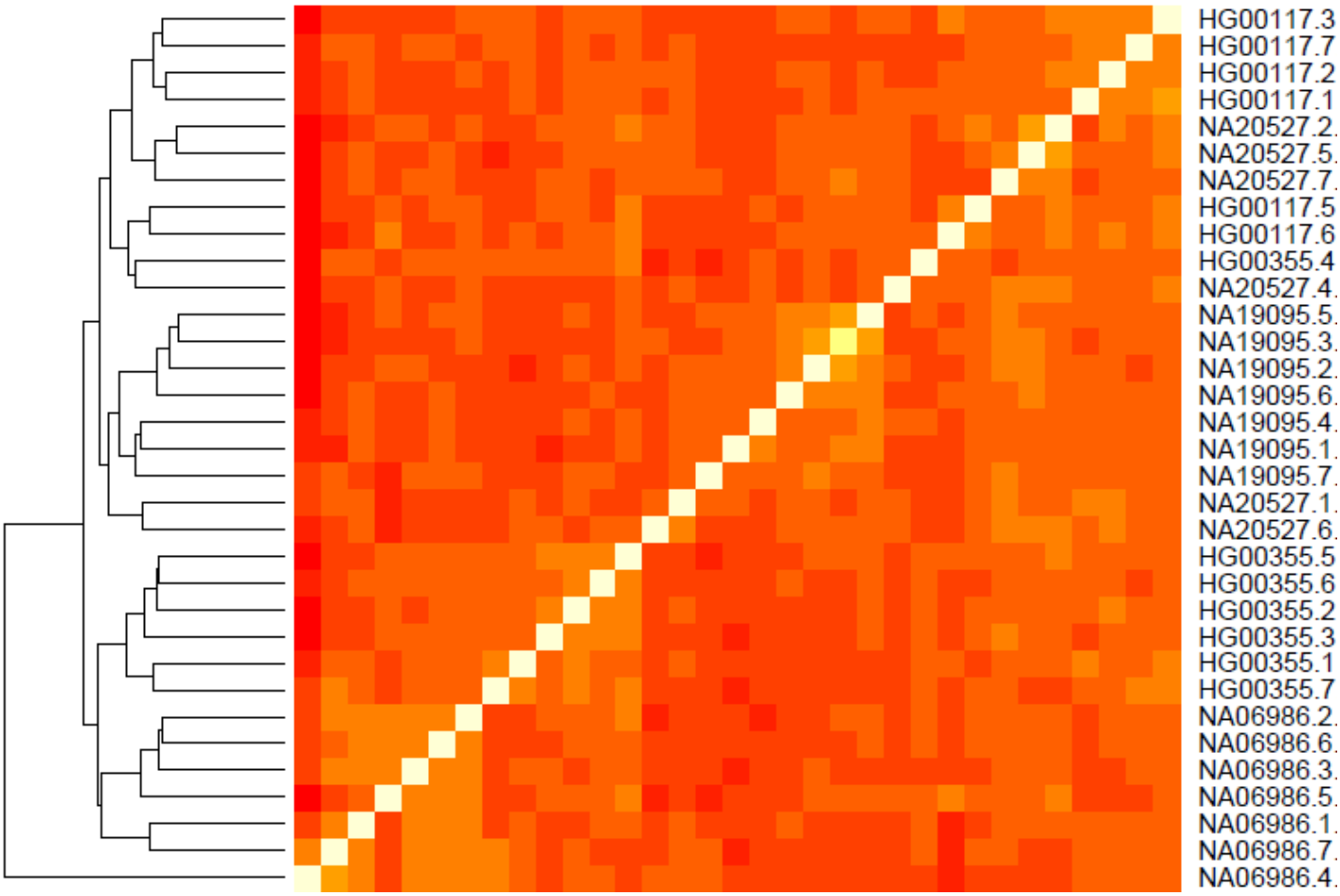
Suppl. Figure 4: Multidimensional scaling of pairwise sample correlations (Pearson correlations after OPS) based on exon quantifications colored by population (A) or laboratory (B).



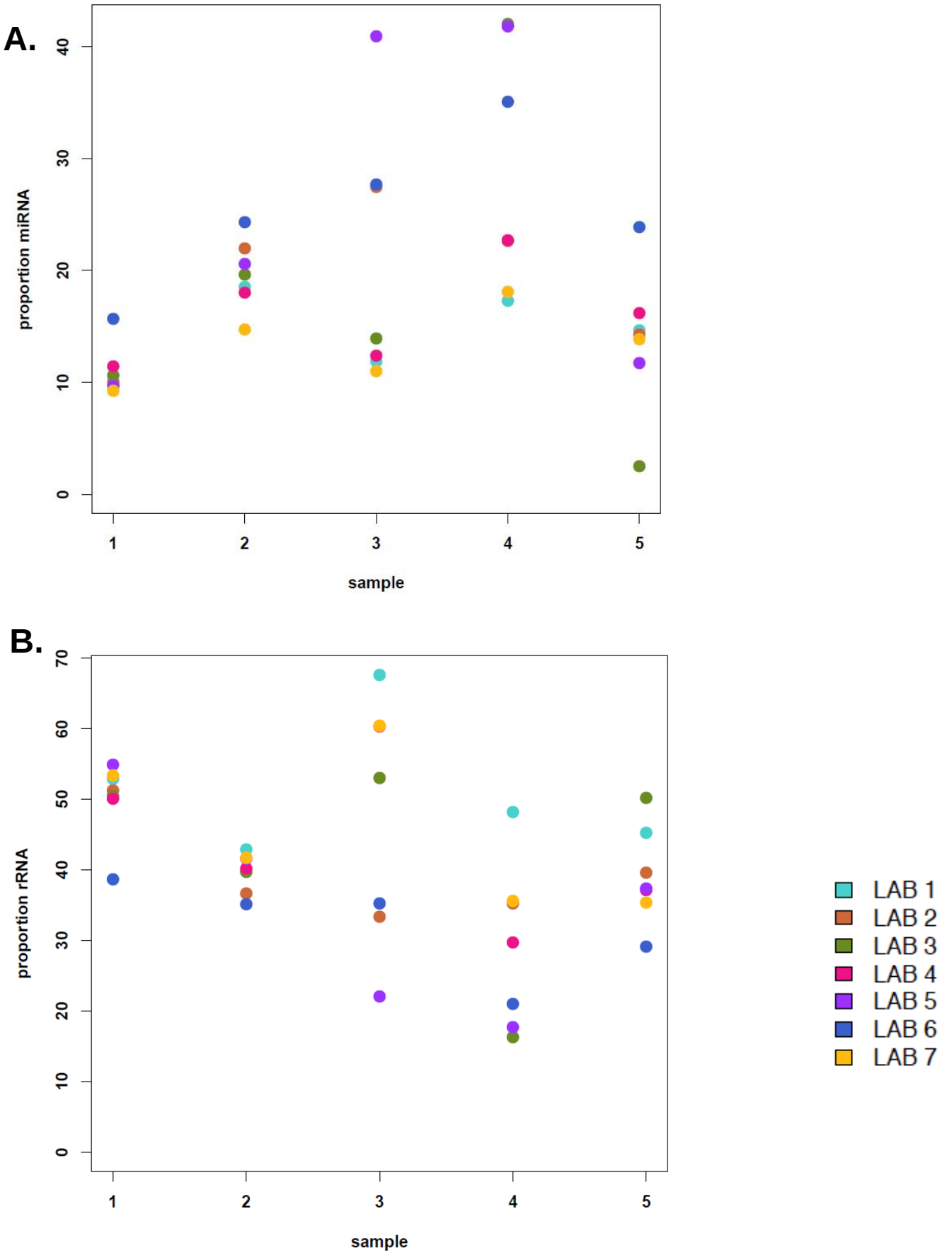
Suppl. Figure 5: Additional sample preparation differences between laboratories. A. Boxplot of library concentrations across different laboratories; B. Boxplot of raw cluster densities across different laboratories; C. Sum of the number of reads in repetitive regions outside genes (based on RepeatMasker, see Lappalainen et al., submitted) (x-axis) plotted against the sum of the counts in exons (y-axis) for all samples, colored according to the laboratory. D. same as C, but now counts in rRNAs are not included in the repeat counts.



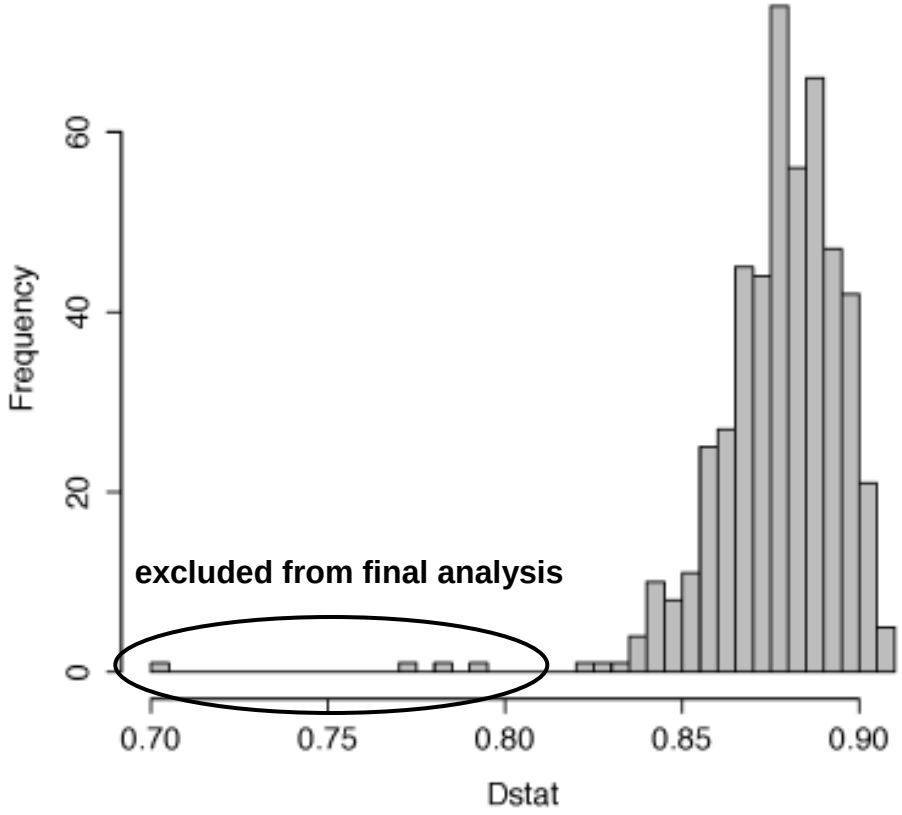
Suppl. Figure 6: Correlation of sample characteristics most strongly associated with PEER factors 1 (A), 2 (B), 3 (C), 4 (D), 6 (E), 9 (F,G), and 10 (H) colored by laboratory.



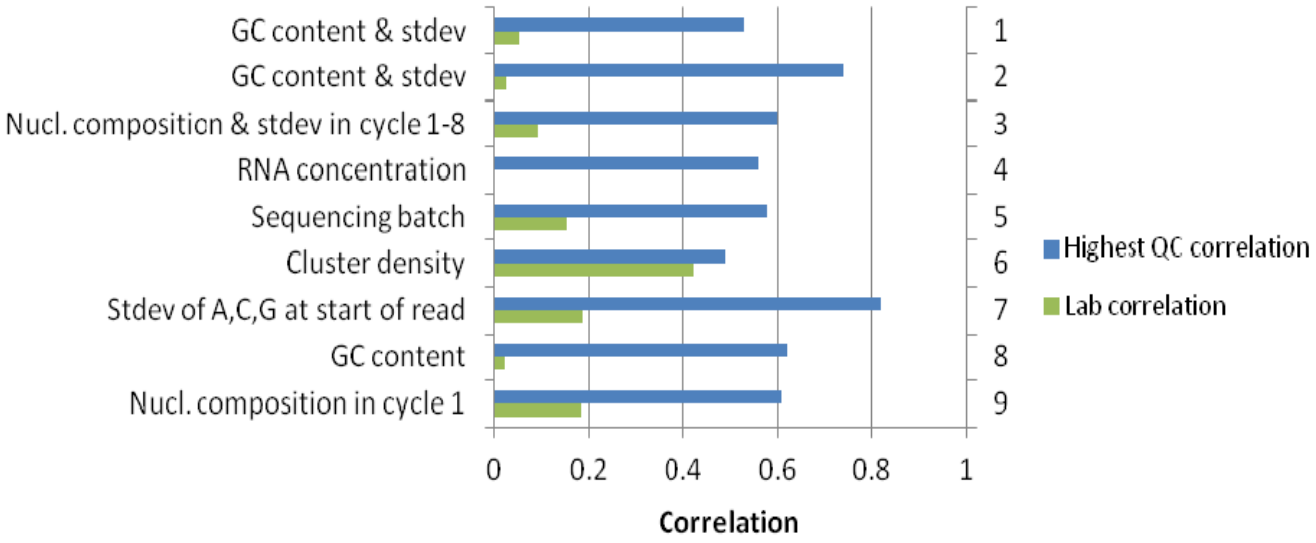
Suppl. Figure 7: Heatmap and clustering of Pearson correlations (after OPS transformation) for miRNA quantifications. White is highest correlation, yellow is intermediate, red is lowest. Samples are indicated with their HapMap identifier followed by their sequence laboratory identifier.



Suppl. Figure 8: Proportion of miRNA (A) and rRNA (B) reads in the five samples replicated in all seven sequencing laboratories (indicated with different colors). There is more variation between samples than within samples, indicating that differences in proportion of miRNAs as percentage of the total small RNAs have been introduced before the preparation of the samples for sequencing.



Suppl. Figure 9: Histogram of D-statistics (median of pairwise Pearson correlation after OPS transformation) for miRNA quantifications



Suppl. Figure 10: Most important sources of miRNA sample variation for each PEER factor, strength of these correlations (blue bars) and the correlation of the laboratory effect to each PEER factor (green bars). For numerical factors Spearman correlations are shown. For categorical variables the categories are first transformed into factors that are used together with each PEER factor in a linear regression. From the linear regression the R2 value is extracted and used to measure the correlation.