RNA-seq benchmarking study on existing and new state-of-the-art sequencing technologies at the VIB Nucleomics Core

Kizi Coeck, Lim De Swert, Rekin's Janky, Kobe Lavaerts, Ruth Maes, Stéphane Plaisance, Emily Sheridan, Thomas Standaert, Jolien Vandewinkel, Stefaan Derveaux

VIB Nucleomics Core – Leuven, Belgium

Experimental design

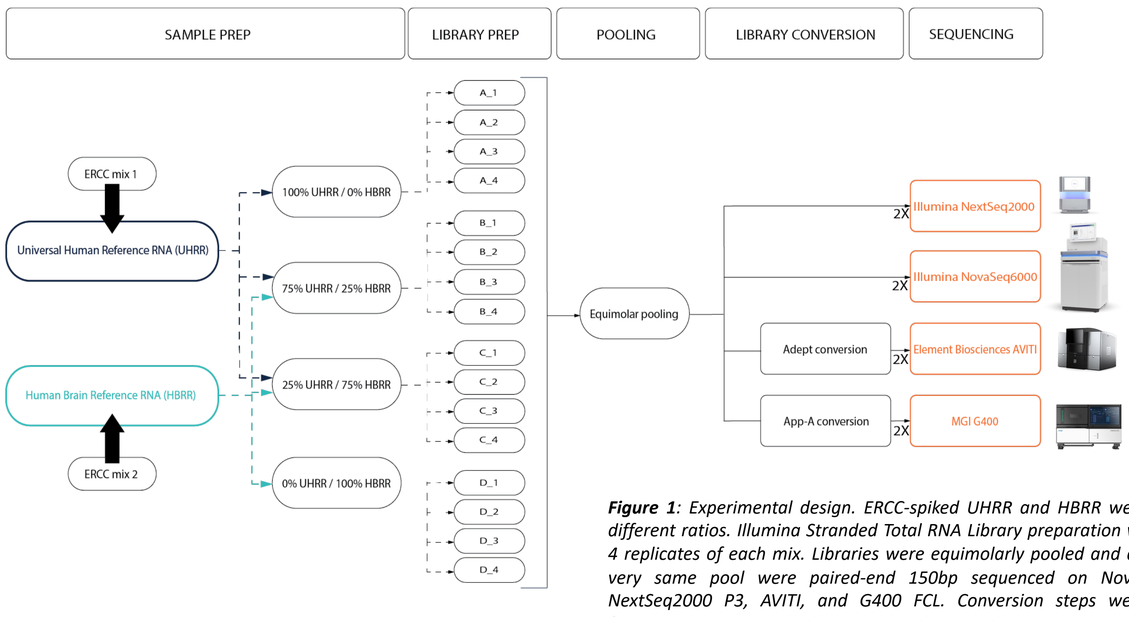

***Figure 1**: Experimental design. ERCC-spiked UHRR and HBRR were mixed at 4 different ratios. Illumina Stranded Total RNA Library preparation was applied to 4 replicates of each mix. Libraries were equimolarly pooled and aliquots of the very same pool were paired-end 150bp sequenced on NovaSeq6000 SP, NextSeq2000 P3, AVITI, and G400 FCL. Conversion steps were introduced following instructions by the providers to make the pool compatible with AVITI and G400 sequencing.*

Total RNA Library Prep QC

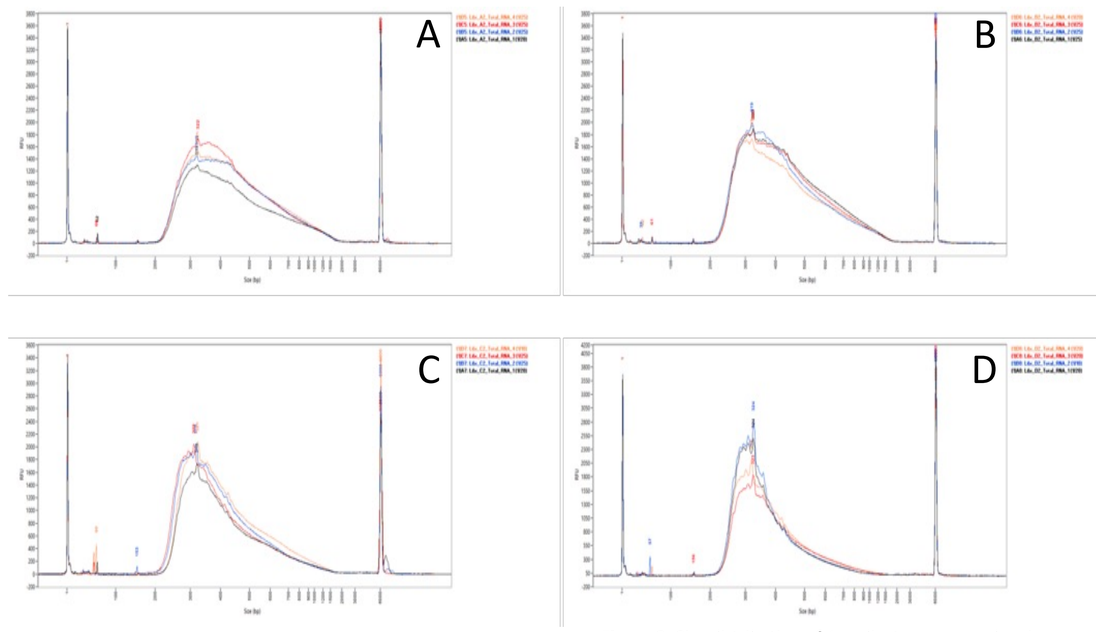

***Figure 2**: Size distribution of 4 replicates per sample mix.*

Pooling results

***Table 1**: Average read distribution with standard deviation across all replicates within each library group.*

Library	AVITI #1	AVITI #2	NovaSeq6000 #1	NovaSeq6000 #2	G400 #1	G400 #2	NextSeq2000 #1	NextSeq2000 #2
A	5,84% +/-1,18%	5,80% +/-1,19%	6,32% +/-1,14%	6,34% +/-1,10%	6,13% +/-0,31%	6,15% +/-0,27%	6,02% +/-0,57%	6,05% +/-0,56%
B	5,59% +/-1,74%	5,60% +/-1,77%	6,40% +/-0,45%	6,38% +/-0,45%	5,83% +/-0,39%	5,82% +/-0,41%	6,06% +/-0,40%	6,06% +/-0,40%
C	7,40% +/-0,34%	7,44% +/-0,33%	6,33% +/-0,26%	6,31% +/-0,24%	6,70% +/-0,12%	6,65% +/-0,13%	6,43% +/-0,36%	6,44% +/-0,34%
D	6,18% +/-1,63%	6,16% +/-1,61%	5,95% +/-1,22%	5,96% +/-1,13%	6,34% +/-0,10%	6,38% +/-0,11%	6,49% +/-0,57%	6,45% +/-0,59%

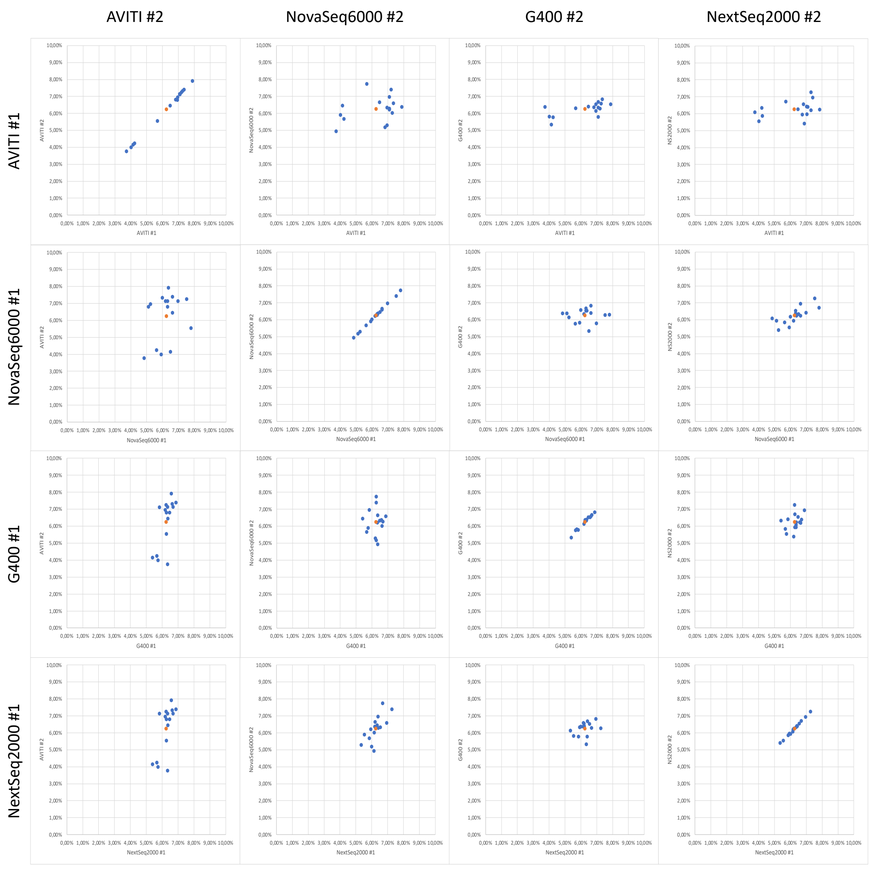

***Figure 3**: Run-to-run and platform-to-platform pairwise comparison of read distribution across all 16 libraries (% of total output, blue dots). Theoretically expected percentage of all libraries depicted as orange dot.*

Summary

- **Limited correlation in output between platforms** although the Bioanalyzer quality control plots show overlapping profiles for all 4 replicates within each sample.
- **G400 and NextSeq2000 show relatively the smallest variation** in read distribution, although AVITI and NovaSeq6000 reach similar standard deviations for some sample groups. The Adept conversion step has shown to be very reproducible (data not shown), so we expect that the underlying reason for the few outliers on the AVITI is related to small differences in library preparation reagents (adaptors and/or primers, currently under investigation).
- **High reproducibility between two flow cells run on each platform** in terms of read distribution and %duplicates.
- With the exception of few outlier samples that differ in output, **%duplicated sequences seems to cluster well for both AVITI and NovaSeq6000, while it's more spread out for the 2 other platforms**. We should note that %duplicate values in all plots are based on an exact match of (only) R1 150bp sequences. As sequencing errors will tend to create artificial diversity in a library (identical sequences will finally become different just because of such technical errors), higher percentages of sequencing errors will result in an underestimation of the real duplication levels. It's therefore important to take the corresponding read quality values also into account when comparing duplication levels between sequencing technologies.
- **NovaSeq6000 and NextSeq2000 shows different %duplicates.** Surprisingly, while NovaSeq6000 and NextSeq2000 use similar patterned FC technology and sequencing chemistry, the observed % duplicated sequences was very different, while Q30-scores are almost identical.
- **The highest Quality Scores were obtained on the AVITI, followed by the G400.** We observed for instance 98,5% bases with >Q30 for AVITI (corresponding to 95% >Q40), while this was 92%, 95% and 92,5% for NovaSeq6000, G400 and NextSeq2000, respectively.

Quality scores

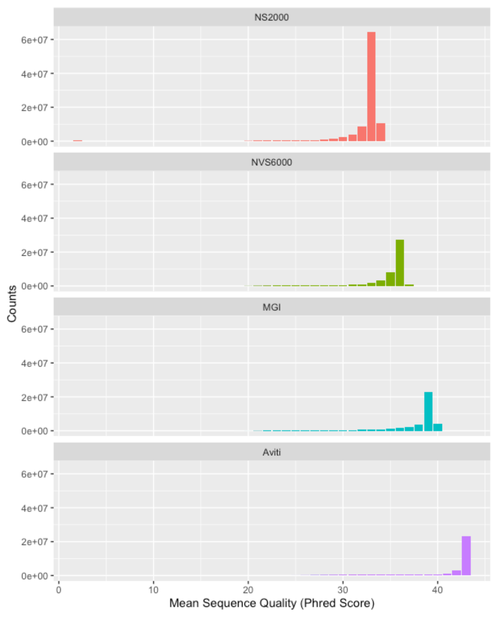

***Figure 4**: Read count versus mean Q-score of platforms*

%Duplicated sequences

***Figure 5**: %duplicate sequences in each run (2 per platform)*

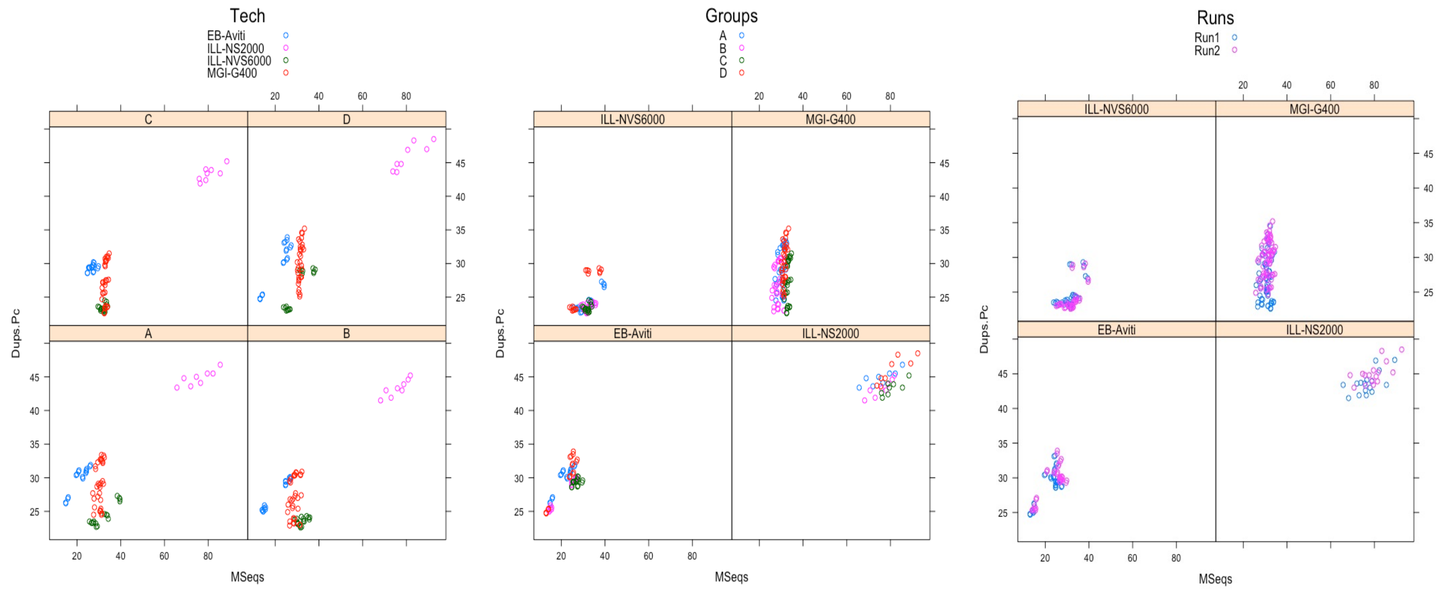

***Figure 6**: Evaluation of %duplicated sequences per library group (left), platform (middle), and run (right). Percentage duplicates has been calculated for each individual lane of each flow cell (AVITI: 2 lanes, NovaSeq6000: 2 lanes, G400: 4 lanes, and NextSeq2000: 1 lane – 2 flow cells per platform).*

As we just recently started with this RNA-seq benchmarking study, the evaluation is not yet finished and extra in-depth data-analysis and validation is still ongoing. If you are interested in reading more about it and/or if you would like to get an idea about other ongoing benchmarking studies at the VIB Nucleomics Core, please get into contact with us via connect.vib.be

VIB NUCLEOMICS CORE

SCIENCE MEETS LIFE