

Supplementary file 2:

Comparison between PacBio run 1 (PCR products generated from samples with ~55, ~110 and ~255 CAGs) and PacBio run 2 (PCR products generated from samples with ~255 and ~470 CAGs) by comparing the sample with ~255 CAGs in each run

Two PacBio runs have been performed to establish the usefulness of PacBio sequencing to quantify somatic mosaicism in different organs of R6/2 mice with different numbers of CAG repeats:

- PacBio run 1 included several samples with ~55 CAGs, several samples with ~110 CAGs and a single with ~255 CAGs (cerebellum of a 6-week-old mouse)
- PacBio run 2 included several samples with ~255 CAGs (including the cerebellum of the 6-week-old mouse also in PacBio run 1) and several samples with ~470 CAGs.

We have used the sample included in both runs (cerebellum of the 6-week-old mouse with ~255 CAGs) as an internal control to compare the sequencing quality in each run (quality of the reads of insert - ROI). The sequencing quality was much higher in PacBio run 1 than in PacBio run 2 with the majority of ROIs being associated with average PHRED scores >30 for PacBio run 1 while the majority of the ROIs were associated with average PHRED scores <30 for PacBio run 2 (Figure 1 A & B). When looking at the quality of the ROIs as a function of their length it was also obvious that PacBio run 1 produced better quality sequences. Indeed, the median PHRED score for PacBio run 1 was >40 for ROI length < 1900 bp while the median PHRED score was mostly around 20 for PacBio run 2 without any particular relationship with ROI read length (Figure 1 C & D).

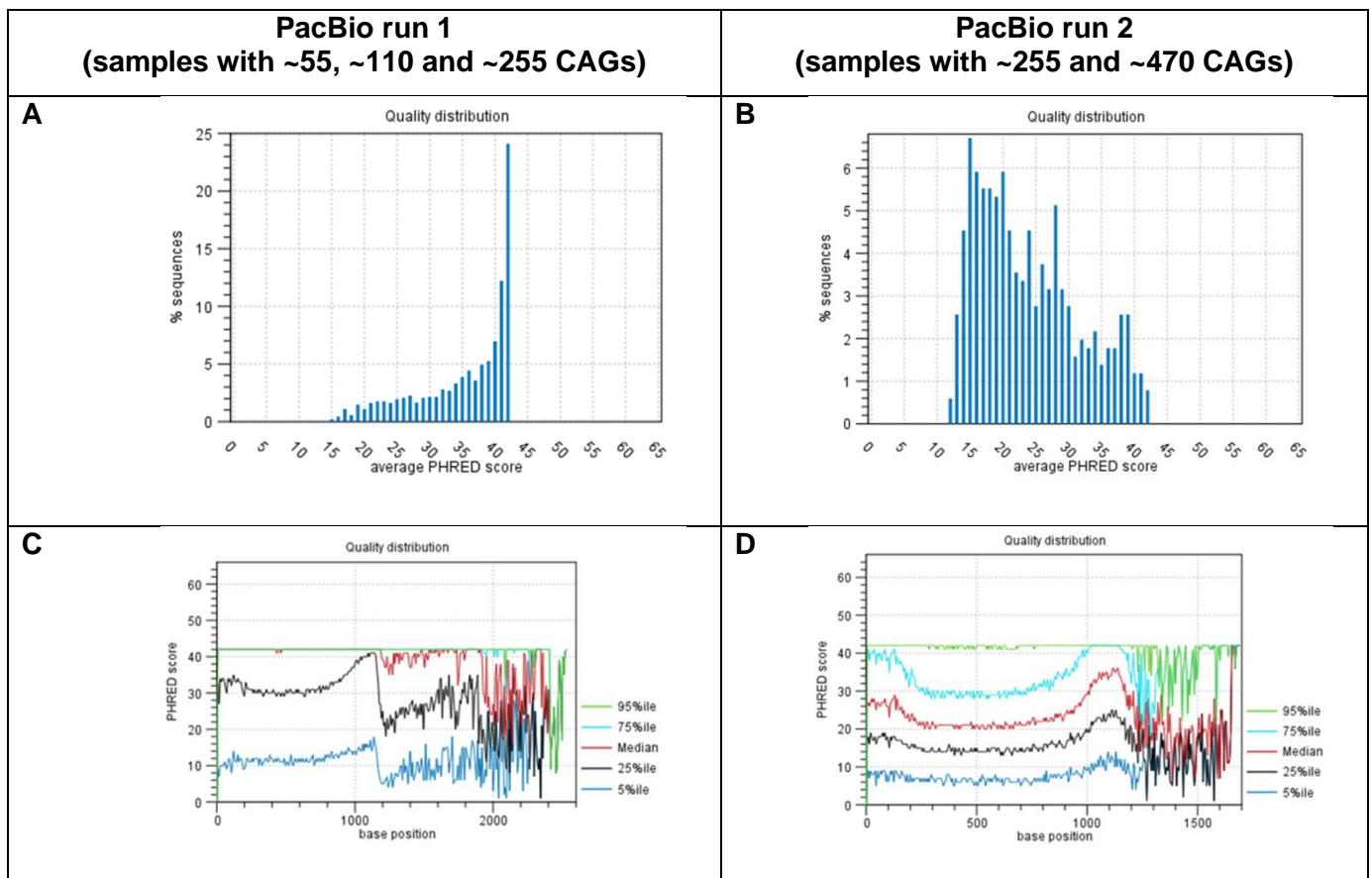


Figure 1: Sequencing quality distributions of reads of insert (ROI) for PacBio run 1 and PacBio run 2. A and C: PacBio run 1. B and D: PacBio run 2. A and B: frequency distribution of the average PHRED score of the ROIs. C and D: Median and quantile PHRED score as a function of the position of the base within the ROIs.