**Why high-quality data is important**

Alt-text Figure 3 - Flowchart showing the steps involved in the pre-processing Quality Control step

Flowchart of pre-processing quality control. Raw reads are the input. Single-ended and pair-ended reads follow the procedures: Remove low quality bases - QScore > Intensifying read lengths/GC content check > Removing reads with “N” > Cross-species contamination check > Adapter/primer removal. The output is high-quality reads for downstream analysis.

Alt-text Figure 4 - Quality score plots

A box-plot of sequencing cycle vs position in read (bp). Each sequencing read is represented by a box-plot. The higher the box-plot is in the cycle axys, the better is the quality of the reads. The image shows one example of good quality and one bad quality score plots.

Alt-text Figure 5 - Sequence content plot

Line-plots representing % vs position in sequence. Lines represent the % content of the bases T (red), C (blue), A (green) and G (black). Three examples are shown. Good data shows 4 colour-coded lines in parelell, representing a constant % of each base across the sequence. Bad data is represented by an “bisulfide contamination” plot, were all bases show irregular peaks across the sequence. The RNA-seq bad data shows irregular peaks a the beginning of the sequence, then the lines cross each other in multiple points throughout the sequence.