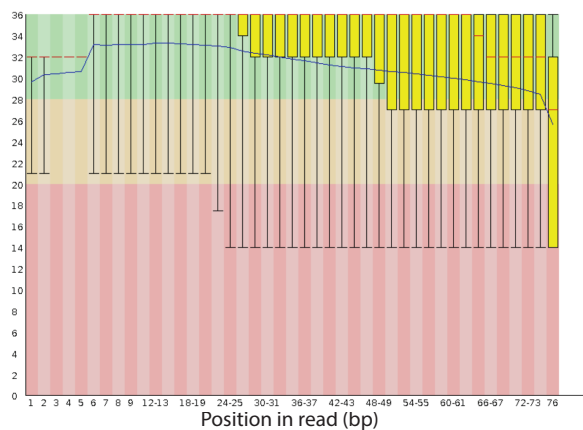
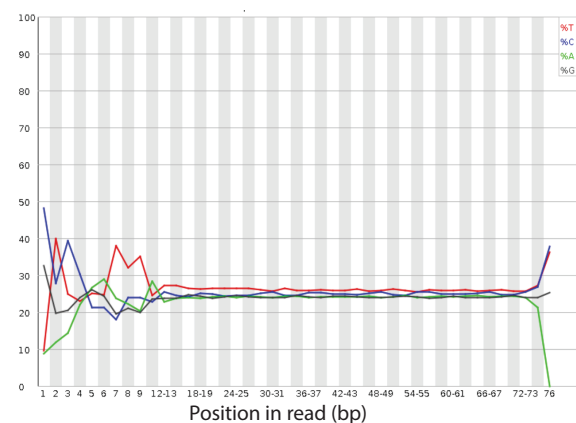


(a)

Quality scores across all bases (Sanger / Illumina 1.9 encoding)

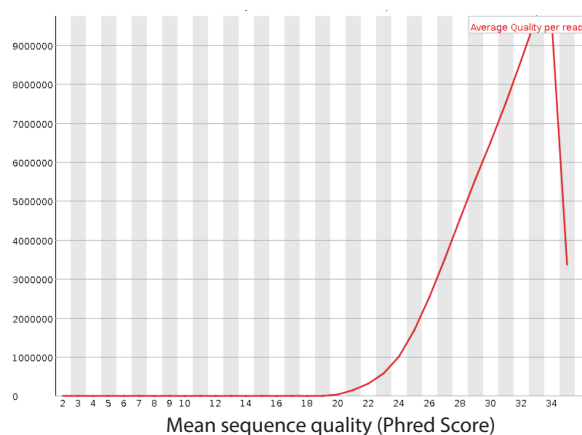


Sequence content across all bases



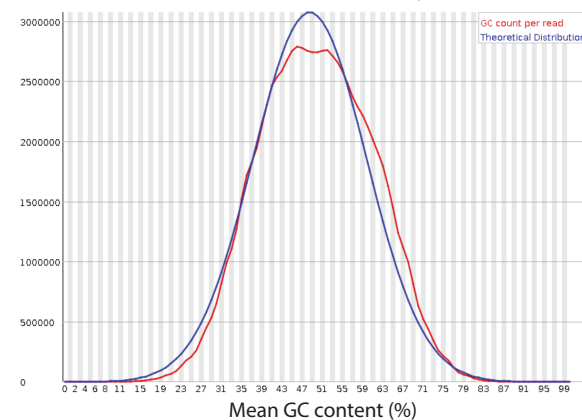
(b)

Quality score distribution over all sequences



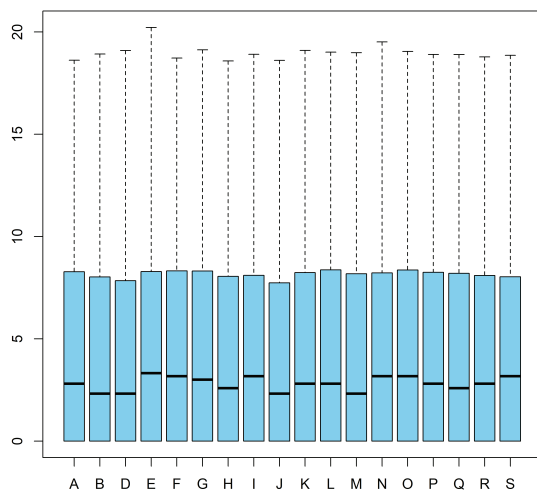
Mean sequence quality (Phred Score)

GC distribution over all sequences



Mean GC content (%)

Before Normalization



(c)

After Normalization

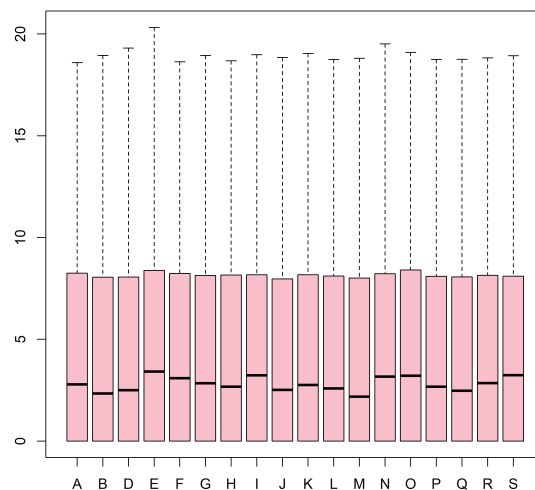


Figure S1. Quality report of raw data. (a) Average quality of sequencing for each position within the reads and the distribution according to their average quality. (b) Representation of the average proportion of each base for all reads and distribution of readings according to their composition in GC. (c) Normalization of raw data: before and after.