

Sequencing Data Quality Assessment

Nate Olson

2017-07-18

Quality assessment of sequencing run summarizing number of reads per sample.

Two barcoded experimental samples have less than 50,000 reads 1. The rest of the samples with less than 50,000 reads are negative PCR controls (NTC). Sample E01JH0016 titration 5 position F9 of plate 1 initial 16S PCR failed.

Excluding the one failed reaction the total range in the observed number of sequences per sample is approximately 40,000 to 150,000 reads. The low variability in number of reads per sample suggests that normalization methods are less likely to impact the results compared other datasets with larger variability in the number of reads per sample.

TODO Figure out why E01JH0011 titration 3 position D2 plate 2 is also low, look at picogreen post normalization data.

Table 1: Summary statistics for number of reads per sample by biological replicate.

biosample_id	median	min_lib_size	max_lib_size
E01JH0004	106596	67844	152267
E01JH0011	79353	46328	140926
E01JH0016	87388	58014	145351
E01JH0017	92307	52302	143110
E01JH0038	86708	53874	123669
NTC	6629	1305	25349

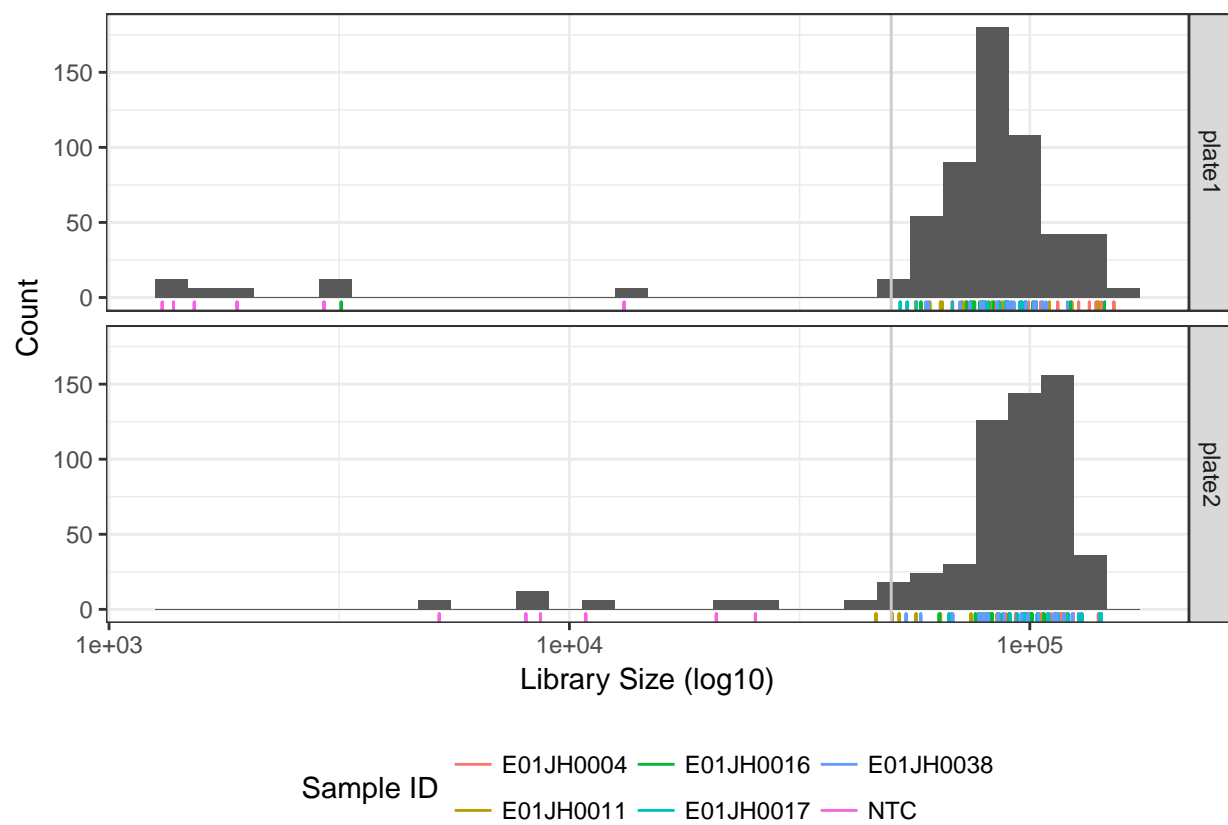


Figure 1: Number of reads per barcoded sample (Library Size), by read direction (X-facet) and replicate 16S PCR plate (Y-facet). Vertical line indicates 50,000 reads per barcoded sample.