Accurate Quantification and Quality Control of Illumina TruSeq Sequencing Libraries with the QX100™ Droplet Digital™ PCR (ddPCR™) System

Shawn P. Hodges, Nick Heredia, Austin P. So, Samantha Cooper, Dimitri Skvortsov, and Svilen Tzonev
Bio-Rad Laboratories, Digital Biology Center, Pleasanton, CA

Introduction

A ddPCR method has been developed to accurately quantify and quality control Illumina TruSeq libraries. A FAM/HEX duplex assay designed to Illumina's Truseq adapters allows for the detection of species composed of the P5 and P7 adapters. ddPCR quantification can exclude ill-formed libraries and adapter dimers, which would not yield valuable sequence information on the insert. The accurate quantification allows a researcher to load an optimal amount of library within the narrow optimal range, thereby maximizing the amount of useable sequencing reads without saturating the system.

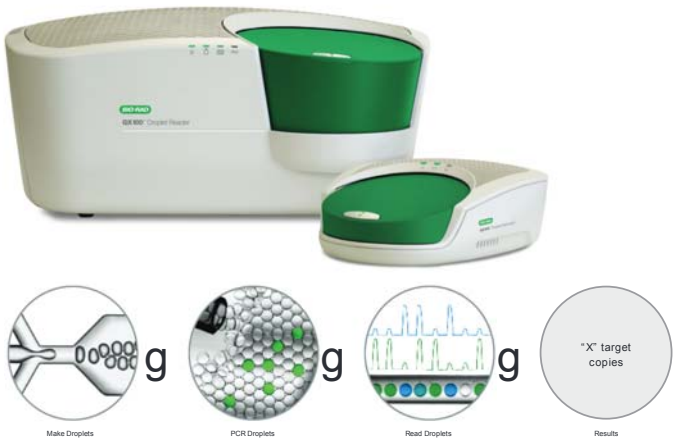


Fig. 1. QX100 Droplet Digital PCR system and ddPCR workflow.

Materials and Methods

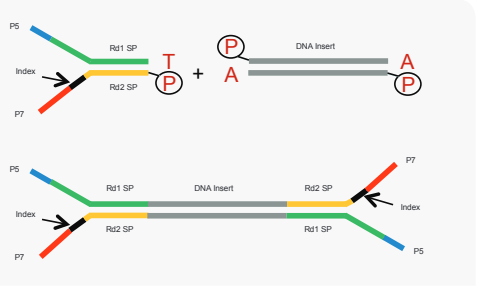
FirstChoice total human brain reference RNA (Life Technologies cat #AM6050), TruSeq RNA sample prep kit v2 (Illumina cat #RS-122-2001), ddPCR library quantification kit for Illumina TruSeq (Bio-Rad cat #186-3040), QX100 Droplet Digital PCR system (Bio-Rad cat #186-3001), MiSeq reagent kit v2 (50 cycles) (Illumina cat #MS-102-2001), MiSeq personal sequencer (Illumina), HiSeq personal sequencer (Illumina).

Libraries were constructed from total brain RNA with Illumina's TruSeq RNA sample preparation kit. 10 µg, 1.0 µg, or 1 ng of RNA was used as starting material for the "bad library," "good library," and "bad library," respectively, as shown in the ddPCR results. For the sequencing results, all libraries were constructed starting with 1.0 µg of RNA. Quantification and quality control (QC) of the libraries was done with Bio-Rad's ddPCR library quantification kit for Illumina TruSeq on the QX100 Droplet Digital PCR System. Titrations of total library concentrations per ddPCR were performed and loaded onto Illumina's MiSeq reagent kit, then sequenced on the MiSeq sequencer. Multiplexes of 12 and 8 libraries were then run at 5 pM total concentrations on MiSeq and HiSeq sequencers, respectively. Finally, MiSeq sequencing quality was compared when "bad" library species were included vs. excluded in the ddPCR quantifications for library loading.

Fig. 2. Library construction and evaluation workflow.

Step 1

The DNA inserts to be sequenced are hybridized and then ligated to Illumina's Y-adapters. After this step library fragments contain both P5 and P7 adapter sequences that bind to the Illumina flow cell and a barcoding index to uniquely identify samples.



Step 2

A library is constructed after ligation of the Illumina Y-adapters to the DNA inserts.

Step 3

Bio-Rad's FAM/HEX duplex assay for Illumina TruSeq library quantification and QC is designed using probes that bind to the adapter sequences.

ddPCR Results

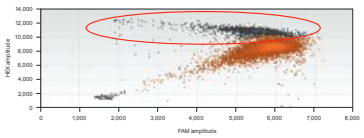


Fig. 3A. Bad library. Additional P7 adapter incorporated into library. These species can potentially bind sequencing flow cell and produce unusable mixed-reads.

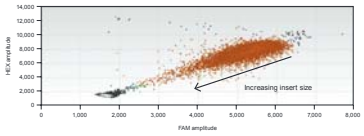


Fig. 3B. Good library. Libraries contain inserts flanked by one P5 and one P7 adapter. Relative size distribution established with smaller libraries having higher FAM/HEX amplitudes.

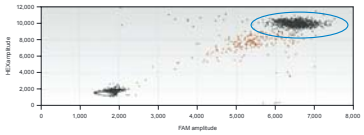


Fig. 3C. Bad library. Adapter-dimer species. These bind to the sequencing flow cell and do not yield insert sequences.

Sequencing Results

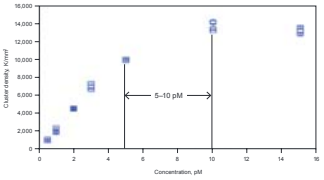


Fig. 4A. Cluster densities. The densities increase linearly, then saturate with increasing library loads.

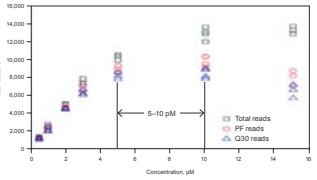


Fig. 4B. Total reads. Reads increase with increasing amounts of libraries loaded, but the quality decreases above 10 pM; the optimal loading amount was 5–10 pM.

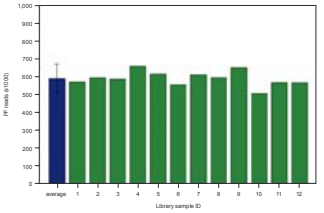


Fig. 4C. MiSeq run. Twelve libraries were diluted to 2 nM each based on ddPCR concentrations and loaded at 5 pM total. The result was 8.24% ± 1.15% PF vs. 8.3% PF expected.

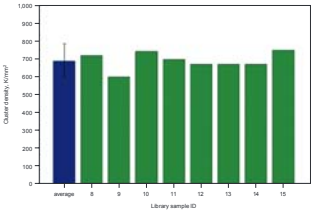


Fig. 4D. HiSeq run. Eight libraries yielded 823 ± 13.8 K/mm2.

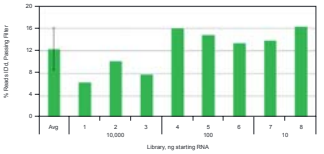


Fig. 4E. Library balancing with extraneous clusters. Libraries were balanced based on ddPCR data that include bad library clusters, resulting in poor balancing.

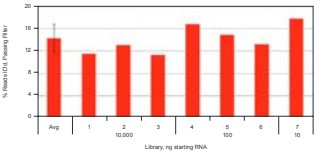


Fig. 4F. Library balancing. Libraries were balanced based on excluding libraries not yielding usable sequence data, which balanced the data.

Conclusions

- ddPCR accurately quantifies Illumina TruSeq libraries
- ddPCR serves as a quality control to determine whether there are substantial species that should not be sequenced, which saves time and money in sequencing
- Using ddPCR, optimal loading of the flow cell is achievable to maximize usable sequencing reads
- Multiplex balancing of libraries is shown such that passing filter reads are nearly equally represented
- Exclusion of bad libraries can be achieved by using ddPCR quantification, allowing more optimal loads onto the sequencer and improved multiplexing

Bulletin 6510 Rev A 1013