

# QX100™ Droplet Digital™ PCR System for Accurate Quantification of Illumina TruSeq Libraries

Tech  
Note

Shawn Hodges, Samantha Cooper, Dimitri Skvortsov, Svilen Tzonev, Pengchin Chen, Austin So, and Nicholas Heredia  
Digital Biology Center, Bio-Rad Laboratories, Inc., 5731 W Las Positas Blvd, Pleasanton, CA 94588

Digital PCR

Bulletin 6343

## Introduction

The advent of next-generation sequencing (NGS) technologies has led to an explosion of accessible and affordable genetic information. The potential benefits of NGS are tempered by the time and cost associated with preparing, running, and analyzing a sequencing library (Buehler et al. 2010, White et al. 2009). The total number of possible reads for the Illumina MiSeq and HiSeq platforms is directly related to the concentration of the prepared library loaded on the flow cell. Loading an insufficient amount of library onto the flow cell results in underutilization of sequencer capacity while overloading the library can lead to poor data quality or complete failure of the sequencing run. Accurate quantification and quality measures of the library prior to sequencing are important for optimal loading of the flow cell surface.

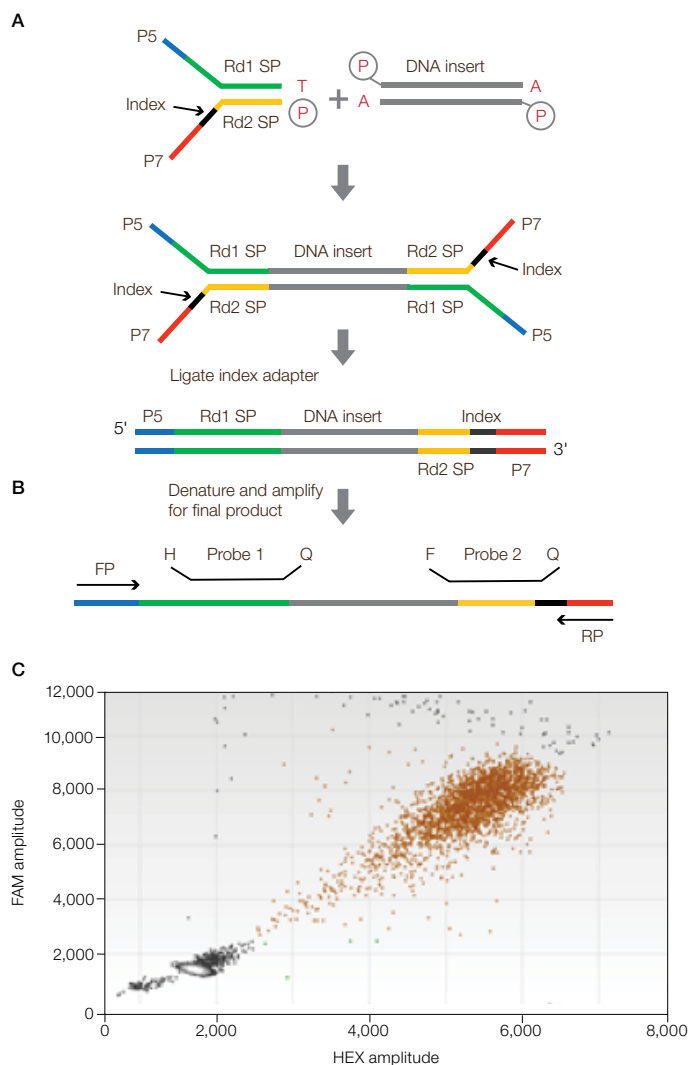
Accurate quantification is also required when combining different libraries into a single run. This combining, or “pooling,” approach is highly desirable because it increases sample throughput on NGS platforms. However, inadequate quantification will often cause poor balancing of these combined libraries, resulting in overrepresentation of some libraries and underrepresentation of others.

Here we demonstrate that Droplet Digital PCR (ddPCR™) with Bio-Rad’s QX100 ddPCR system can be easily incorporated into the library preparation workflow to accurately quantify and balance sequencing libraries on Illumina sequencers. ddPCR enables these highly precise measurements without the use of standards. In addition to accurate quantification, the data plots generated by the QX100 system are information rich and contain quality measures of the library construction that are not available when using other methods. Digital PCR enables consistent library loading and efficient utilization of the sequencing capacity of Illumina NGS platforms.

## Materials and Methods

### Illumina TruSeq Library Construction

The TruSeq DNA sample preparation kit v2 (Illumina, Inc.) contains Y-adapters that are ligated to the fragments to be sequenced. The library is constructed such that both the P5 and the P7 sequences of the adapters are added onto all fragments (Figure 1A). After enrichment by PCR amplification and subsequent purification steps, library concentrations



**Fig. 1. TruSeq V2 library protocol and corresponding design of the ddPCR library quantification kit for Illumina TruSeq assay.** **A**, TruSeq Y-adapters, made up of both P5 and P7 sequences and an index, are ligated to an adenylated DNA insert fragment. Libraries are enriched by PCR amplification, resulting in amplicons containing the P5 and P7 sequences, which are directionally oriented on either strand of the fragment. **B**, Bio-Rad’s ddPCR library quantification kit contains assays designed to span both the P5 and P7 sequences, allowing quantification of species possessing the adapter arms. **C**, an example of a ddPCR trace observed for an Illumina library quantified with the ddPCR library quantification kit assay. FP, forward primer; RP, reverse primer.

**BIO-RAD**

are determined by ddPCR and sequencing of the library is initiated. For this study, 12 uniquely indexed TruSeq DNA libraries and 12 RNA-Seq libraries were created from 1 µg of sonicated (200 bp average; Covaris, Inc.) human genomic DNA (female; Promega Corporation) and 10–4,000 ng total RNA from human brain (Life Technologies Corporation), respectively. The libraries were constructed according to the manufacturer's protocol.

#### TruSeq Library Quantification by ddPCR

The ddPCR library quantification kit for Illumina TruSeq (Bio-Rad Laboratories, Inc.) contains hydrolysis probes designed to detect and quantify both the P5 and the P7 adapter sequences (Figure 1B). The combined signals from each assay are used to confirm the formation of properly adapted library fragments. These signals can be seen as double-positive clusters on the 2-D fluorescence amplitude plot in QuantaSoft™ software (Bio-Rad) (Figure 1C).

ddPCR quantification of the samples was accomplished by preparing serial dilutions of purified libraries in TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). For quantification with the ddPCR library quantification kit, duplicate 20 µl digital PCR mixtures were prepared using 2x ddPCR supermix for probes (no dUTP) (Bio-Rad), 20x stock in the ddPCR library quantification kit assay, and 5 µl of the 10<sup>-6</sup>-, 10<sup>-7</sup>-, and 10<sup>-8</sup>-fold library dilutions. Each digital PCR mixture was converted into droplets, amplified by PCR, and analyzed with the QX100 ddPCR system (Bio-Rad). Thermal cycling conditions for the C1000 Touch™ thermal cycler (Bio-Rad) were 10 min at 95°C (hot-start), followed by 40 cycles of 30 sec at 94°C then 1 min at 60°C with ramp rates of 2.0°C/sec, 10 min at 98°C, and 4 or 12°C infinity hold. Following thermal cycling, droplets were analyzed with the QX100 droplet reader (Bio-Rad). The copies per microliter reported in QuantaSoft software were converted to nanomolar concentrations.

#### MiSeq Analysis of Libraries

Based on the ddPCR concentration results, individual TruSeq libraries were diluted to 2 nM in TE buffer and stored at –80°C. Prior to sequencing, aliquots were denatured with 0.1 N NaOH and neutralized with Illumina's HT1 buffer to 20 pM according to the manufacturer's recommended protocol (Illumina guide 2011). Neutralized libraries were mixed at various ratios, based on ddPCR results, with HT1 buffer ( $V_{\text{Total}} = 1$  ml) and 600 µl was loaded onto the MiSeq cartridge (Illumina) for sequencing.

#### Molarity Calculation from Copies per Droplet

Example of a molarity calculation from copies per droplet:

To backcalculate the stock concentration of the library from ddPCR concentrations within the range of 100–5,000 copies/µl, account for the dilution factors, including the fivefold dilution of the ddPCR reaction (4 µl of the dilution into 16 µl of supermix).

**Example:** A library at the 10<sup>-6</sup> dilution yielded 2,000 copies/µl in ddPCR. Multiply 2,000 by 10<sup>6</sup> and by 5 to account for the reaction dilution (2,000 x 10<sup>6</sup> x 5 = 10<sup>10</sup> copies/µl of original stock library).

- To obtain nM: (10<sup>10</sup> copies/µl) x (10<sup>6</sup> µl/L)/6.023 x 10<sup>23</sup> copies/mole = 1.66 x 10<sup>-8</sup> M or 16.6 nM

## Results and Discussion

Maximizing the flow cell cluster generation on the MiSeq and HiSeq NGS platforms requires accurate and precise library quantification. This maximization increases the number of usable reads and efficiently utilizes the high-throughput capabilities offered by next-generation sequencers.

#### Current NGS Library Quantification Methods

Both microfluidics-based electrophoresis (for example, Agilent 2100 bioanalyzer, Agilent Technologies, Inc.) and quantitative PCR (qPCR) using SYBR® Green chemistry are commonly employed to quantify libraries prior to sequencing. Neither method alone provides sufficient resolution to accurately determine library concentration.

Electrophoretic methods provide a measure of the fluorescence generated by intercalation of dye throughout the range of species present in the prepared library. This integrated fluorescence is calibrated to the fluorescence of size-based standards to derive a concentration value. However, the presence of poorly formed, single-stranded, and/or unamplifiable species can compromise this estimate because these species can comigrate with properly formed library fragments.

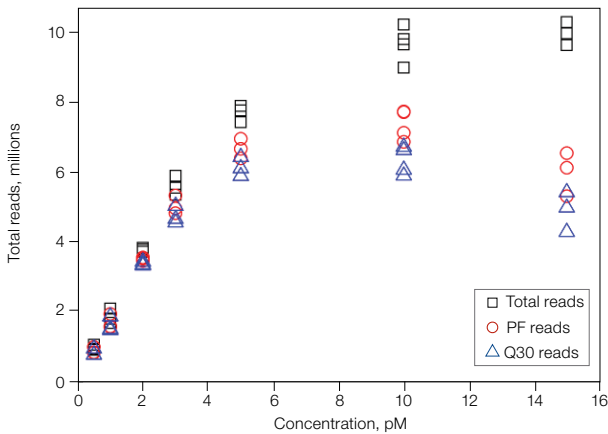
qPCR-based methods partially avoid these issues by detecting only amplifiable library species. To assess concentration with qPCR, external calibrants, which are typically synthetic and of fixed length or derived from previously characterized libraries, are required to create a standard curve. To correct for quantification cycle (Cq) differences that arise from the length-dependent fluorescence signal generated by intercalating dyes, a correction factor based on the average size of library fragments is employed. However, this correction factor does not fully capture the size distribution of library fragments, nor does it scale linearly with large size differences between library fragments and the standards. The presence of primer-dimers or background contaminants that affect amplification efficiencies can further skew concentration estimates.

By contrast, the QX100 ddPCR system provides the means to directly measure amplifiable library concentrations in terms of copies per microliter, which can be easily converted to molar concentrations. ddPCR achieves direct measurements by effectively counting amplified molecules in droplets. These

direct measurements remove the requirement for calibration against a standard curve and the need for further correction based on average library fragment length.

**NGS Library Quantification by ddPCR**

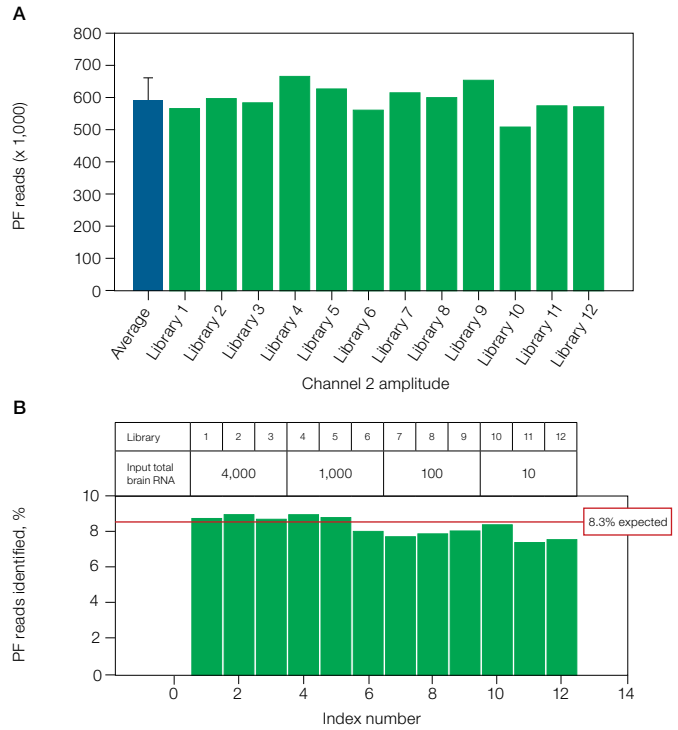
When performing NGS on the MiSeq platform, it is important to aim for a cluster density of approximately 800,000/mm<sup>2</sup> for optimal performance. ddPCR measurements were used to establish the functional relationship between input library concentration and the number of usable reads on the MiSeq platform. To calibrate performance, we used five libraries generated at different concentrations. After ddPCR quantification with the ddPCR library quantification kit, the five libraries were individually loaded at various concentrations between 0.5 pM and 15 pM onto the MiSeq sequencer. The resulting sequencing read metrics were compared with the concentration of library loaded (Figure 2). In these experiments, 8 pM concentration for loading the flow cell achieved approximately 800,000/mm<sup>2</sup> cluster density.



**Fig. 2. Impact of input library concentration on total usable reads.** Cluster density at 8 pM was approximately 800,000/mm<sup>2</sup>.

We examined the precision of ddPCR in balancing 12 TruSeq DNA libraries from human genomic DNA using concentration measurements obtained from the QX100 system with the ddPCR library quantification kit assay (Figure 3A). Based on the ddPCR concentration measurements, libraries with an average fragment length of 447 bp could be balanced to within less than 15% of each other with a confidence interval of 95%. Similar balancing results were observed when RNA-Seq libraries with an average fragment length of 280 bp were utilized (Figure 3B).

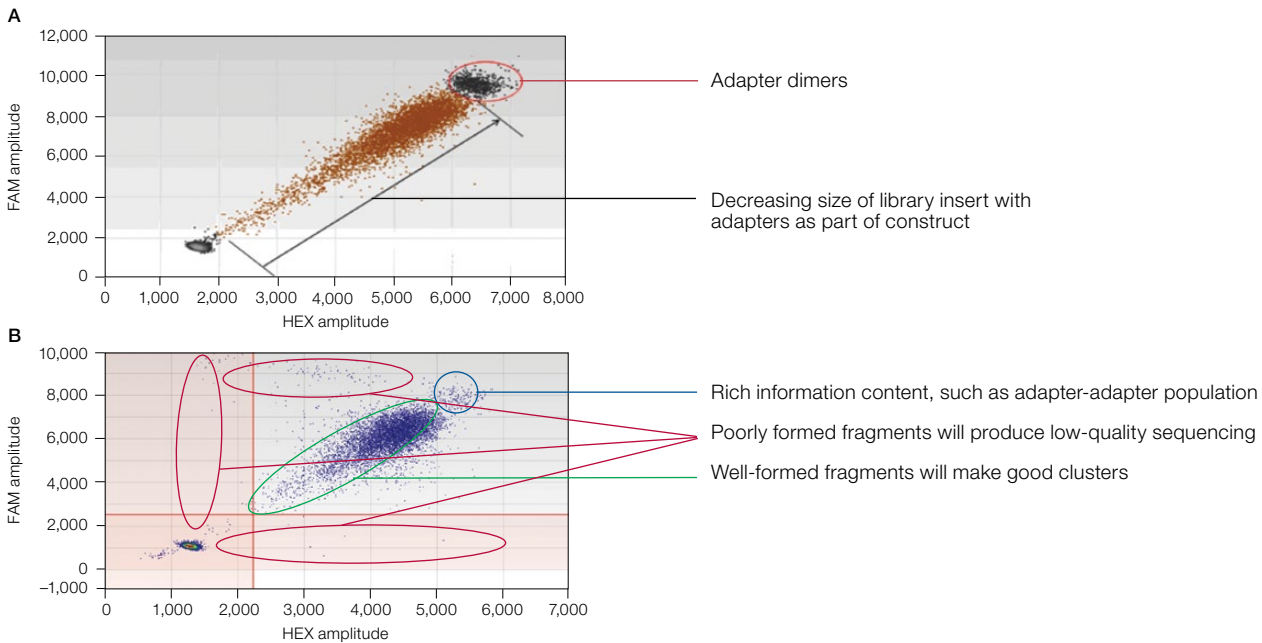
ddPCR is information rich and can provide quality metrics for the libraries generated by the user. In Figure 4, two-dimensional fluorescence amplitude plots are shown for the FAM and HEX probes in the ddPCR library quantification kit assay. The cluster of droplets, which is at a 45° angle from the negative droplet population, arises due to the different PCR efficiencies of different-sized amplicons in the adapted library.



**Fig. 3. Balancing TruSeq libraries using the ddPCR library quantification kit assay.** **A**, twelve TruSeq libraries were quantified by ddPCR and balanced to be equimolar for the sequencing run. Excellent balancing within less than 15% difference was achieved between all 12 libraries pooled in the same sequencing run (total number of reads passing filter for each library). **B**, twelve RNA-Seq libraries from varying input total RNA amounts (10–4,000 ng) were sequenced and balanced. Good balancing was obtained between the 12 indexed libraries and is plotted as percentage reads passing filter for the sequencing run vs. sample indices. PF, passing filter.

There is an inverse relationship between amplicon size/complexity and fluorescence amplitude. This relationship is what makes the droplet position within the data plots significant.

The position of the droplets within the 2-D plot enables the user to assess the overall quality of constructed libraries. As seen in Figure 4A, there is a secondary cluster of highest fluorescence amplitudes for both the FAM and HEX channels; this corresponds to the adapter-adapter ligated species. With this information, the user can quantify the adapter-adapter population using QuantaSoft software. Figure 4B further exemplifies different populations of adapted library molecules that could arise from aberrant processes during library construction and give poor or no sequencing information (shown in red). By quantifying potentially poorly formed library inserts, the user can exclude these abnormal populations from calculations to improve loading efficiency on the flow cell. If the library is of very poor quality, the user may choose to re-make the adapted library prior to sequencing, thus saving time and money by not sequencing poorly constructed libraries. The library quality information obtained from these 2-D plots will undoubtedly fuel further investigations into improvements of NGS sequencing, possibly by determining library fragment PCR efficiencies. Improvements



**Fig. 4. Two-dimensional plots of FAM and HEX fluorescence amplitudes of the ddPCR library quantification kit assay.** **A**, demonstration of how the ddPCR assay is rich in additional information. Smaller insert species have more efficient ddPCR reactions in the droplets, giving an inverse correlation between insert size and fluorescence amplitude in the ddPCR assay. **B**, illustration of how this amplitude information obtained from ddPCR can be detected and give a sense of overall quality of the library to be sequenced by observing different species in the fluorescence amplitude plots. QuantaSoft software can compute and exclude these populations from quantification for the purpose of accurately loading the flow cell.

in NGS workflows will likely result from the extremely accurate quantification possible with ddPCR. For example, it may be possible to eliminate secondary amplification steps if sufficient library material is generated for sequencing directly after library construction, thus avoiding unnecessary steps and any further skewing of fragment representation.

### Conclusions

Cluster densities, and therefore the number of reads, are directly tied to loading concentration of the prepared library. Moderate differences in loading can affect quality and read capacity of the MiSeq platform. The QX100 system complements NGS platforms by providing a means to measure absolute concentration of relevant molecular species. ddPCR, using the ddPCR library quantification kit for Illumina TruSeq, enables precise measurements to correlate loaded library concentration with cluster generation. ddPCR provides excellent balancing of pooled library samples, and it permits proper analysis to optimize the performance of NGS systems. Additional benefits of ddPCR include rich fluorescence amplitude data, which can correlate with library construction quality.

For more information, visit

[www.bio-rad.com/web/ddPCRTruSeqLibraries](http://www.bio-rad.com/web/ddPCRTruSeqLibraries).

### References

- Buehler B et al. (2010). Rapid quantification of DNA libraries for next-generation sequencing. *Methods* 50, S15–S18.
- Illumina, Inc. (2011). Sequencing Library qPCR Quantification Guide. Catalog #SY-930-1010, part #11322363 Rev C.
- White RA et al. (2009). Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics* 10, 116.

HiSeq, Illumina, MiSeq, and TruSeq are trademarks of Illumina, Inc. Illumina is not affiliated with Bio-Rad Laboratories, Inc.

Agilent is a trademark of Agilent Technologies, Inc. FAM is a trademark of Applied Biosystems, Inc. SYBR is a trademark of Life Technologies Corporation.

The QX100 Droplet Digital PCR system and/or its use is covered by claims of U.S. patents, and/or pending U.S. and non-U.S. patent applications owned by or under license to Bio-Rad Laboratories, Inc. Purchase of the product includes a limited, non-transferable right under such intellectual property for use of the product for internal research purposes only. No rights are granted for diagnostic uses. No rights are granted for use of the product for commercial applications of any kind, including but not limited to manufacturing, quality control, or commercial services, such as contract services or fee for services. Information concerning a license for such uses can be obtained from Bio-Rad Laboratories. It is the responsibility of the purchaser/end user to acquire any additional intellectual property rights that may be required.

Bio-Rad's thermal cyclers and real-time thermal cyclers are covered by one or more of the following U.S. patents or their foreign counterparts owned by Eppendorf AG: U.S. Patent Numbers 6,767,512 and 7,074,367.



**Bio-Rad  
Laboratories, Inc.**

Life Science  
Group

Web site [www.bio-rad.com](http://www.bio-rad.com) USA 800 424 6723 Australia 61 2 9914 2800 Austria 01 877 89 01 Belgium 09 385 55 11 Brazil 55 11 3065 7550  
Canada 905 364 3435 China 86 21 6169 8500 Czech Republic 420 241 430 532 Denmark 44 52 10 00 Finland 09 804 22 00  
France 01 47 95 69 65 Germany 089 31 884 0 Greece 30 210 9532 220 Hong Kong 852 2789 3300 Hungary 36 1 459 6100 India 91 124 4029300  
Israel 03 963 6050 Italy 39 02 216091 Japan 81 3 6361 7000 Korea 82 2 3473 4460 Mexico 52 555 488 7670 The Netherlands 0318 540666  
New Zealand 64 9 415 2280 Norway 23 38 41 30 Poland 48 22 331 99 99 Portugal 351 21 472 7700 Russia 7 495 721 14 04  
Singapore 65 6415 3188 South Africa 27 861 246 723 Spain 34 91 590 5200 Sweden 08 555 12700 Switzerland 026 674 55 05  
Taiwan 886 2 2578 7189 Thailand 1800 88 22 88 United Kingdom 020 8328 2000