Use of the melting curve assay as a means for high-throughput quantification of Illumina sequencing libraries

Hiroshi Shinozuka, John W Forster

Background. Multiplexed sequencing is commonly performed on massively parallel short-read sequencing platforms such as Illumina, and the efficiency of library normalisation can affect the quality of the output dataset. Although several library normalisation approaches have been established, none is ideal for highly multiplexed sequencing due to issues of cost and/or processing time. **Methods.** An inexpensive and high-throughput library quantification method has been developed, based on an adaptation of the melting curve assay. Sequencing libraries were subjected to the assay using the Bio-Rad Laboratories CFX Connect™ Real-Time PCR Detection System. The library quantity was calculated through summation of reduction of relative fluorescence units between 86 and 95°C. **Results.** PCR-enriched sequencing libraries are suitable for this quantification without pre-purification of DNA. Short DNA molecules, which ideally should be eliminated from the library for subsequent processing, were differentiated from the target DNA in a mixture on the basis of differences in melting temperature. Quantification results for long sequences targeted using the melting curve assay were correlated with those from existing methods (R^2>0.77), and that observed from MiSeq sequencing (R^2 = 0.82). **Discussion.** The results of multiplexed sequencing suggested that the normalisation performance of the described method is equivalent to that of another recently reported high-throughput bead-based method, BeNUS. However, costs for the melting curve assay are considerably lower and processing times shorter than those of other existing methods, suggesting greater suitability for highly multiplexed sequencing applications.

Use of a melting curve assay as a means for high-throughput quantification of Illumina sequencing libraries

Hiroshi Shinozuka[1,2,*] and John W. Forster[1,2,3]

[1]Department of Economic Development, Jobs, Transport and Resources, Biosciences Research Division, AgriBio, Centre for AgriBioscience, 5 Ring Road, La Trobe University, Bundoora, Victoria 3083, Australia

[2]Dairy Futures Cooperative Research Centre, Australia

[3]La Trobe University, Bundoora, Victoria 3086, Australia

[*]Corresponding author: e-mail – hiroshi.shinozuka@ecodev.vic.gov.au

Keywords: second-generation sequencing technology; multiplexed sequencing; library preparation; library normalisation

Abstract

Background. Multiplexed sequencing is commonly performed on massively parallel short-read sequencing platforms such as Illumina, and the efficiency of library normalisation can affect the quality of the output dataset. Although several library normalisation approaches have been established, none is ideal for highly multiplexed sequencing due to issues of cost and/or processing time.

Methods. An inexpensive and high-throughput library quantification method has been developed, based on an adaptation of the melting curve assay. Sequencing libraries were subjected to the assay using the Bio-Rad Laboratories CFX Connect™ Real-Time PCR Detection System. The library quantity was calculated through summation of reduction of relative fluorescence units between 86 and 95°C.

Results. PCR-enriched sequencing libraries were suitable for this quantification without pre-purification of DNA. Short DNA molecules, which ideally should be eliminated from the library for subsequent processing, were differentiated from the target DNA in a mixture on the basis of differences in melting temperature. Quantification results for long sequences targeted using the melting curve assay were correlated with those from existing methods (R^2>0.77), and that observed from MiSeq sequencing (R^2 = 0.82).

Discussion. The results of multiplexed sequencing suggested that the normalisation performance of the described method is equivalent to that of another recently reported high-throughput bead-based method, BeNUS. However, costs for the melting curve assay are considerably lower and processing times shorter than those of other existing methods, suggesting greater suitability for highly multiplexed sequencing applications.

Introduction

Substantial reductions in both the cost and processing time of DNA sequencing have been achieved in the last decade, due to improvements of massively parallel short-read sequencing technologies, and this trend is expected to continue for the next several years or more (http://ark-invest.com/genomic-revolution/declining-costs-of-genome-sequencing). As a consequence, sequencing library preparation procedures, rather than the DNA sequencing process itself, provide the major bottleneck for large-scale experimentation (Rohland & Reich, 2012). As a cost-effective approach, single sequencing libraries are parallel-processed from individual samples, and assigned a unique oligonucleotide index. The libraries are then pooled for multiplexed processing in a single run (or lane) of a high-throughput sequencing platform. During this pooling step, equalisation of DNA quantity from each library is essential for an optimised sequencing output (Hosomichi et al., 2014).

Although several library normalisation methods have been established, all of them incur a relatively high cost and/or require long processing times for a large number of samples (Table 1) (Buehler et al., 2010; Campbell, Harmon & Narum, 2015; Harris et al., 2010; Hosomichi et al., 2014; Katsuoka et al., 2014; Kong et al. 2014; Rohland & Reich, 2012). Single libraries are commonly quantified with a well-established real-time PCR-based method (Buehler et al., 2010; Rohland & Reich, 2012) which, however, requires a relatively long processing time. The qMiSeq method was recently proposed as a highly accurate library titration method for multiplexed sequencing libraries (Katsuoka et al., 2014). In this method, the pooled library is first sequenced on the MiSeq instrument (Illumina, CA, USA) in order to estimate the quantity of single libraries, which are then re-pooled for sequencing on the HiSeq platform (Illumina).

Although this method is highly efficient when applied to an extensive sequencing experiment for an organism with a large genome, the per-sample cost approaches US$10 if the sample number is 93 or equivalent.

Melting curve (MC) assays identify the melting temperature (MT) point at which double-stranded DNA (dsDNA) dissociates into single-stranded DNA (ssDNA), through detection of reductions in dsDNA-specific fluorescence signals during the heating process (Ririe, Rasmussen & Wittwer, 1997). This assay can be performed on most real-time PCR instruments, and up to 96-1536 samples can be simultaneously processed, depending on the instrument format (Lennon et al. 2010). MT largely depends on the length of dsDNA, such that shorter dsDNA molecules, in general, dissociate at lower temperatures than longer ones (Ririe, Rasmussen & Wittwer, 1997). As a consequence, the MC assay is commonly used for quality control purposes in real-time PCR experiments that show presence/absence of undesirable short amplicons, such as PCR primer dimers.

An application of the MC assay for quantification of sequencing libraries is reported in this study. Libraries for multiplexed sequencing are generally enriched through PCR, and short DNA molecules, such as PCR primers, primer dimers, and amplicons derived from self-ligated sequencing adapters must be excluded from quantification. Although SPRI (solid phase reversible immobilisation)-based DNA size-selection is commonly used for this purpose (Harris et al., 2010), this approach requires a relatively large amount of consumables (micropipette tips and microtiter plates) and a long processing time. This procedure is not required for MC assay-based quantification, which permits exclusion of such short DNA molecules from library quantification, based on differences of MT. In addition, as real-time PCR instruments are commonly present

in molecular biology laboratories, this normalisation approach can be readily applied in labs and hence may not require the acquisition of a new instrument.

Materials & Methods

Preparation of short DNA and KAPA fragments

In order to generate short DNA amplicons (<300 bp), which are normally eliminated from sequencing libraries, an in-house library procedure was performed in the absence of DNA template. An in-house-designed double-stranded Y-shaped adaptor with a single thymine base extension (50 pmol) was treated with the Klenow fragment of *Escherichia coli* DNA polymerase I (conferring 3'→5' exonuclease activity) (New England Biolabs; NEB, MA, USA) in the presence of dATP, followed by self-ligation using T4 DNA ligase (NEB). The ligated products were purified with Lambda Exonuclease and Exonuclease I (NEB), and then subjected to PCR amplification for 32 cycles with the Phusion DNA polymerase kit (Thermo Fisher scientific, MA, USA) and in-house library preparation primers containing sequencing indexes. PCR amplification was also performed using the DNA standard 3, primer mix and qPCR master mix of the KAPA Library Quantification Kit Illumina Platforms (Kapa Biosystems, MA, USA), to generate amplicons of 452 bp in length. Sample mixtures (total volume: 20 µl), containing the short DNA or KAPA fragments, 1x PCR buffer, and SYBR Green I (Thermo Fisher Scientific), were prepared for the MC assay.

Preparation of sequencing libraries

Sequencing libraries for the MiSeq platform were prepared following a previously described high-throughput method based on use of the methylation-dependent restriction endonuclease MspJI (NEB) (10). Genomic DNA of strains of the most prevalent perennial ryegrass-associated fungal endophyte (*Epichloë festucae* var. *lolii*)

was directly amplified from mycelium using the REPLI-g Mini Kit (QIAGEN, Hilden, Germany). In the amplification solution of the kit, 16.7μM 5-methylcytosine (TriLink biotechnologies, CA, USA) was included. The amplicons were digested with MspJI in order to promote semi-random DNA fragmentation. The DNA fragments were treated with the Klenow fragment, followed by ligation of the in-house-designed adaptor using T4 DNA ligase. The DNA was then purified and size-selected using AMPure XP beads (Beckman Coulter, CA, USA), a portion of which was used for PCR enrichment with the in-house-designed primers using the Phusion polymerase kit. During this process, a unique sequencing string was attached to each sample. Using SYBR Green I as a fluorescence dye, amplification was monitored on the CFX Connect™ Real-Time PCR Detection System (Bio-Rad Laboratories, CA, USA).

MC assay-based dsDNA quantification

For the MC assay-based quantification procedure, 10 μl of DNA staining mixture, containing 0.1 μl 100x SYBR Green I, 2 μl 5x PCR buffer, 1.5 μl 500mM EDTA and 6.4 μl PCR-grade water, was prepared and added to 20 or 25 μl PCR products. The MC assay was performed on the real-time PCR instrument. Initial sample heating was performed at 75°C for 30 sec, and then samples were incubated at each degree point for 5 or 10 sec, followed by regular temperature increments of a single degree. Fluorescence measurement was performed after incubation at each degree point. The library concentration was calculated through summation of reduction of relative fluorescence units (dRFU) between 86 and 95°C.

Automated gel electrophoresis-based DNA quantification

Library quantification was performed with the 2200 TapeStation Instrument and the D1000 kits (Agilent Technologies, CA, USA). In order to reduce technical error, 2 µl of PCR product was added to 6 µl D1000 Reagent, instead of 1 µl DNA sample and 3 µl D1000 Reagent as described in the manufacturer's instruction. Quantity of DNA between 300-700 bp in length was manually determined and expressed as library quantity.

NanoDrop-based DNA quantification

PCR products were purified with 0.8 times volume of AMPure XP beads in order to remove DNA fragments shorter than 300 bp in length, and DNA was eluted in 10mM Tris-HCl buffer. The concentration of purified DNA was measured on the NanoDrop 1000 instrument (Thermo Fisher Scientific).

Massively-parallel short-read sequencing

Sequencing libraries were pooled based on the basis of quantification result. The pooled library was purified using 0.8 times volume of AMPure XP beads, and subsequently characterised with the TapeStation and Qubit instruments (Thermo Fisher Scientific). The pooled library was loaded and sequenced on the Illumina MiSeq platform, following the manufacturer's instruction. Output data were analysed using the PRINSEQ-Lite software (Schmieder & Edwards, 2011; http://prinseq.sourceforge.net/). The percentage of paired-end read numbers (PRN) among 66 to 75 indexes was calculated for evaluation of normalisation performance (Hosomichi et al., 2014). The normalisation degree at 50% (ND_{50}) was calculated as the ratio of single libraries of which the PRNs were 50% or more of the average PRN.

Results

Determination of a temperature range for MC assay

The short DNA or KAPA fragments were subjected to the MC assay. The results indicated a large portion of the short fragments, of which the majority were less than 300 bp in length, dissociated into ssDNA at temperatures less than 86°C, while the 452 bp fragment only began to dissociate at 86°C (Supplementary Figure S1a and S1b). The short DNA and KAPA fragments were also subjected to the assay in two different types of PCR buffers, demonstrating that components of the buffers slightly affected the MT of both DNA fragments (Supplementary Figure S1).

Comparison of the library quantification methods

The sequencing libraries were prepared from 67 *Epichloë festucae* var. *lolii* endophyte samples. Following adaptor ligation, the libraries were enriched through PCR (Supplementary Figure S2). Monitoring of the amplification process on the real-time PCR instrument revealed that most samples reached the PCR plateau phase within 14 reaction cycles. Using another portion of the adaptor-ligated products, PCR amplification was performed for 12 cycles in the absence of fluorescence dye, to generate sequencing libraries with a range of dsDNA concentration. The concentration of the 12-cycled products was measured using each of the three methods: MC assay-, TapeStation-, and NanoDrop-based quantification (Fig. 1). The DNA library concentrations varied from 0 (undetectable) to 7.51 ng/μl, obtained with the TapeStation-based method. The results from the three methods were correlated with one other (R^2>0.77). Comparison of the three methods indicated that a total dRFU value of 100 may be sufficient for relatively reliable quantification when the MC assay-based

183 method is used. The sum of dRFU values between 75 and 95°C exhibited a lower
184 correlation with the other quantification results (Supplementary Figure S3).
185
186 **MC-based library quantification and normalisation**
187 The 14-cycled libraries were subjected to the MC assay (Fig. 2), revealing that the total
188 dRFU values between 86 and 95°C varied less than those of the 12-cycled samples. A
189 value from a single sample was negative (-6), indicating failure of library preparation.
190 The sum of dRFU from the rest of samples was over 135. The 66 successful libraries
191 were pooled for multiplexed sequencing, according to the quantification result. From the
192 pooled library, 4.7 million paired-end reads were generated using a part of a sequencing
193 run on the MiSeq platform. The PRNs of the 66 libraries varied from 0.45 to 3.21, with
194 an average of 1.52, a standard deviation of 0.47 and coefficient of variation (CV) of 0.31
195 (Fig. 2). Except for the lowest library, the PRNs were over 0.77, and the ND_{50} was
196 consequently 0.98. Based on the obtained sequencing data, PRNs in the absence of a
197 normalisation procedure were also predicted. This simulation revealed that the PRNs
198 from 8 libraries (8/66 = 0.12) would be less than half of the average (1.52/2 = 0.76), if an
199 equal volume of libraries were pooled. The CV of PRNs for the un-normalised pool was
200 calculated to be 0.38.
201 For further validation, sequencing libraries were prepared from another set of 75
202 *Epichloë festucae* var. *lolii* strains. Following library enrichment based on PCR, single
203 libraries were subjected to the MC assay-based quantification procedure
204 (Supplementary Figure S2). The total dRFU values from libraries varied from 20 to 3537
205 (Fig. 3a), while that of the no-template PCR control was 15. From each library, the same
206 volume (5 µl) of the PCR products was pooled and then purified with the AMPure XP

beads. The pooled library was loaded on the MiSeq platform, and a total of 1.4 million reads were generated using a proportion of a MiSeq sequencing run. The PRNs of the sequencing output were correlated with the total dRFU values (R^2 = 0.82). A further correlation analysis indicated that the total dRFU values for the temperature range between 85 and 95°C were most strongly correlated to the PRNs (R^2 = 0.823), followed by the total dRFU values between 85 and 94°C (R^2 = 0.822), and between 86 and 95°C (R^2 = 0.821) (Supplementary Figure S4).

A simulation study was also performed using these results. Assuming that the total dRFU values between 86 and 95°C are in direct proportion to library concentrations, volumes of the PCR products for pooling were calculated. As the total dRFU values from two of the libraries were lower than 100 (20 and 26), these were excluded from the simulation. The simulated PRNs were obtained based on the calculated pooling volume and MiSeq output, and were between 0.5-2.27 (Fig. 3b). The average PRN was 1.37, and the ND_{50} and CV of PRNs were consequently 0.99 and 0.25, respectively. No strong correlations were observed between total dRFU values and PRNs after normalisation (R^2 = 0.25).

Discussion

PCR is commonly used as a target enrichment method for massively parallel sequencing (Hosomichi et al., 2014, Hosomichi et al., 2015). The MspJI-based DNA fragmentation method permits inexpensive sequencing library preparation from amplicon-based templates (Shinozuka et al. 2015). A combination of use of these techniques with the MiSeq platform allows low-cost sequencing analysis for the targeted loci. For instance, DNA sequences with a cumulative length of 15 kb could be enriched through PCR with locus-specific primers at a cost of US$3.4/sample, and a single library could be processed until the library enrichment stage for a cost of US$3.3 (Supplementary Table S1). Libraries from up to 384 samples can be sequenced on the MiSeq platform with the current Reagent Kit v3 (600 cycle; Illumina), generating an average of approximately 2,000 times read depth for each nucleotide position within the 15 kb-long regions. As the MiSeq kit costs less than US$1,500, the per-sample cost for this sequencing option would be under US$4. Compared with these examples, however, the per-sample costs of the existing high-throughput library normalisation methods, are relatively high (Table 1, Supplementary Table S2), even though the primary purpose is solely DNA quantification or normalisation. The MC assay-based library quantification procedure costs substantially less than the other methods; only 1/3 of the NanoDrop-based method. The processing time of the MC assay-based method is also considerably shorter than those of the other methods. For instance, 384 samples may be processed within 3 hours on a standard real-time PCR instrument (4×96-well microtiter plates), and the duration could be reduced to only 40 mins if a 384-well formatted instrument was made available. The MC assay-based library quantification would, therefore, be most suitable for sequencing experiments typified by large sample

number but relatively small-output, such as HLA typing on the MiSeq platform (Hosomichi et al., 2015).

As the quantification results from MC assay-, TapeStation-, and NanoDrop-based method were correlated with one other, each method can be used for the purpose of library normalisation. Similar to the previously reported 'over-estimation issue' (Simbolo et al. 2013), the quantification results obtained with the NanoDrop system were, however, much larger than those obtained with the other methods, especially when low-concentration libraries were subjected to analysis. Based on the NanoDrop-based method, concentrations of 6.51-9.06 ng/ul were calculated from 4 library samples, from which no DNA was detected with the TapeStation-based method. Similar results were observed when the result of NanoDrop-based quantification was compared with that of the MC assay. As the quantification result from the blank sample (10 mM Tris-HCl buffer without DNA) on the NanoDrop system was 0.17±0.25 ng/μl (3 replications), residual molecules from the SPRI-based DNA size-selection may have affected UV-visible spectrophotometry-based quantification.

During the MC assay, the target dsDNA and short DNA can be differentiated on the basis of differences in MT (Ririe, Rasmussen & Wittwer, 1997). The preliminary experiment with the short DNA and KAPA fragments suggested that the dRFU values at less than 86°C broadly represent the quantity of dsDNA shorter than 300 bp in length, while those at and over 86°C represent the quantity of target DNA. In the subsequent experiments, sequencing libraries with the average fragment size of approximately 500 bp in length were subjected to the MC assay-based quantification, and the correlation between the total dRFU values for the temperature range 86 and 95°C and other quantification method was demonstrated. As the values between 75 and

95°C exhibited a lower level of correlation, exclusion of the florescence signals from short dsDNA is essential for accurate library quantification (Supplementary Figures S3 and S4). The average fragment size of the sequencing library is determined in practice by multiple factors, such as types of the sequencing platforms, sequencing reagents, library preparation methods and experiment purposes. Optimisation of the temperature range of the MC assay, therefore, may be required, depending on the average fragment size of the sequencing library.

The length of DNA fragments is not the only factor determining the MT point, as the GC-content ratio and sequence complexity are also relevant (Ririe, Rasmussen & Wittwer, 1997). Differences between sequencing libraries for these parameters may hence influence accuracy of the MC assay-based quantification, and the method may not be suitable for a comparison of DNA quantities, if the DNA sequence contents are considerably different between the libraries. In general, libraries for multiplexed sequencing are, however, prepared in parallel from the identical targeted sequence(s) (Campbell, Harmon & Narum, 2015; Hosomichi et al., 2014, Shinozuka et al. 2015), and the DNA sequence contents of libraries should not be substantially different. For the scenario of different DNA sequence contents, NanoDrop- or TapeStation-based methods would be more suitable.

The CV of PRNs and ND_{50} from the first normalisation experiment were 0.31 and 0.98, respectively, and those from the second simulation experiment were 0.25 and 0.99, respectively. The result of the MC assay-based method may hence be almost equivalent to that of the BeNUS (Bead-based Normalisation for Uniform Sequencing depth) method (CV of PRNs = 0.31, ND_{50} = 0.93) (Hosomichi et al., 2014), although the normalisation result obtained with the qMiSeq method (CV of PRNs = 0.05) was

295 superior to the current result (Katsuoka et al., 2014).

296 The BeNUS method permits high-throughput automated library normalisation,
297 through reduction of the amount of SPRI beads necessary to capture a limited amount
298 of DNA (Hosomichi et al., 2014). This method may consequently be ineffective when
299 library concentrations are too low. The results of the current study suggest that the MC
300 assay-based method performs better for recovery of libraries with low concentration.
301 Due to the high-throughput nature of both the BeNUS and MC assays, a combination of
302 the two may provide a high-quality automated normalisation procedure suitable for an
303 even larger sample volume, numbering in the several thousands.

Acknowledgments

This work was supported by funding from the Victorian Department of Economic Development, Jobs, Transport and Resources. The authors would like to thank Dr. Emma Ludlow for provision of the perennial ryegrass-associated fungal endophyte materials, and Dr. Noel Cogan for technical advice.

Competing Interests Statement

The authors declare no competing interests

References

1. Buehler B, Hogrefe HH, Scott G, Ravi H, Pabón-Peña C, O'Brien S, Formosa R, Happe S. 2010. Rapid quantification of DNA libraries for next-generation sequencing. *Methods* 50:S15–S18. DOI: 10.1016/j.ymeth.2010.01.004.

2. Campbell NR, Harmon SA, Narum SR. 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources* 15:855–867. DOI: 10.1111/1755-0998.12357.

3. Harris JK, Sahl JW, Castoe TA, Wagner BD, Pollock DD, Spear JR. 2010. Comparison of Normalization Methods for Construction of Large, Multiplex Amplicon Pools for Next-Generation Sequencing. *Applied and Environmental Microbiology* 76:3863–3868. DOI: 10.1128/AEM.02585-09.

4. Hosomichi K, Mitsunaga S, Nagasaki H, Inoue I. 2014. A Bead-based Normalization for Uniform Sequencing depth (BeNUS) protocol for multi-samples sequencing exemplified by HLA-B. *BMC Genomics* 15:645. DOI: 10.1186/1471-2164-15-645.

5. Hosomichi K, Shiina T, Tajima A, Inoue I. 2015. The impact of next-generation sequencing technologies on HLA research. *Journal of Human Genetics*. DOI: 10.1038/jhg.2015.102.

6. Katsuoka F, Yokozawa J, Tsuda K, Ito S, Pan X, Nagasaki M, Yasuda J, Yamamoto M. 2014. An efficient quantitation method of next-generation sequencing libraries by using MiSeq sequencer. *Analytical Biochemistry* 466:27–29. DOI: 10.1016/j.ab.2014.08.015.

334 7. Kong N, Ng W, Foutouhi A, Huang C, Weimer BC. 2014. Quality control of high-throughput library
335 construction pipeline for KAPA HTP library using an Agilent 2200 TapeStation, *Agilent Technologies*,
336 publication number 5991-5141EN. *Available at* http://www.agilent.com/cs/library/applications/5991-
337 5141EN.pdf *(accessed 11 April 2016)*
338
339 8. Lennon NJ, Lintner RE, Anderson S, Alvarez P, Barry A, Brockman W, Daza R, Erlich RL,
340 Giannoukos G, Green L, Hollinger A, Hoover CA, Jaffe DB, Juhn F, McCarthy D, Perrin D, Ponchner
341 K, Powers TL, Rizzolo K, Robbins D, Ryan E, Russ C, Sparrow T, Stalker J, Steelman S, Weiand M,
342 Zimmer A, Henn MR, Nusbaum C, Nicol R. 2010. A scalable, fully automated process for
343 construction of sequence-ready barcoded libraries for 454. *Genome Biology* 11:R15. DOI:
344 10.1186/gb-2010-11-2-r15.
345
346 9. Ririe KM, Rasmussen RP, Wittwer CT. 1997. Product Differentiation by Analysis of DNA Melting
347 Curves during the Polymerase Chain Reaction. *Analytical Biochemistry* 245:154–160. DOI:
348 10.1006/abio.1996.9916.
349
350 10. Rohland N, Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for
351 multiplexed target capture. *Genome Research* 22:939–946. DOI: 10.1101/gr.128124.111.
352
353 11. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets.
354 *Bioinformatics* 27:863–864. DOI: 10.1093/bioinformatics/btr026.
355
356 12. Shinozuka H, Cogan NOI, Shinozuka M, Marshall A, Kay P, Lin Y-H, Spangenberg GC, Forster
357 JW. 2015. A simple method for semi-random DNA amplicon fragmentation using the methylation-

358 dependent restriction enzyme MspJI. *BMC Biotechnology* 15. DOI: 10.1186/s12896-015-0139-7.

359

360 13. Simbolo M, Gottardi M, Corbo V, Fassan M, Mafficini A, Malpeli G, Lawlor RT, Scarpa A. 2013.

361 DNA Qualification Workflow for Next Generation Sequencing of Histopathological Samples. *PLoS*

362 *ONE* 8:e62692. DOI: 10.1371/journal.pone.0062692.

Table legend

Table 1

Cost and processing duration assumptions for 92 to 96 samples using the library normalisation or quantification methods. SPRI denotes the SPRI bead-based DNA size-selection and purification before the quantification procedure. Size-specificity indicates that the method is able to exclude both unnecessary long (e.g. >1 kb) and short (e.g. <300 bp) DNA (++), or merely short DNA (+) from quantification. Library-specificity indicates that the method detects only dsDNA with the sequencing adaptors on both ends (+), or also dsDNA without the sequencing adaptor(s), from which clonal sequence clusters cannot be generated (-). Expenditure on instruments is not included in the per-sample cost. More details can be found in Supplementary Table S2.

Figure Legends

Fig. 1

Correlation analysis between the three library quantification methods: (a) MC assay- and TapeStation-based methods; (b) MC assay- and NanoDrop-based methods; and (c) TapeStation- and NanoDrop-based methods.

Fig. 2

Results of the MC assay (a) and high-throughput sequencing (b) from the 66 successful libraries. The horizontal axis shows the sample unique identifier. In the result from high-throughput sequencing, the blue circle indicates PRN from the MiSeq-derived output, and the red square shows the simulated PRN for un-normalised multiplexed sequencing.

Fig. 3

Correlation between the MC assay-based quantification and PRN of MiSeq sequencing from the 75 samples (a), and between the MC assay-based method and predicted PRN from the library normalisation simulation from the 73 successful samples (b).

Supplementary Data Legends

Supplementary Figure S1

Results of the TapeStation and MC assay. (a) The TapeStation results for the short DNA fragments, KAPA fragment and a single sequencing library, showing size distribution of dsDNA. The High Sensitivity D1000 kit (Agilent) was used for the short DNA and KAPA fragments. The y = 0 and x = 0 axes show the fluorescence intensity and DNA fragment size, respectively. 'Lower' and 'Upper' indicate signal peaks of upper and lower DNA size markers, respectively. (b) The melting peak plots from the three samples obtained for temperatures between 75 and 95°C. (c) Effect of PCR buffer on MT of the short DNA and KAPA fragments. The MC assay was performed using three PCR buffers; the Phusion DNA polymerase HF and GC buffers and MyFi^{TM} DNA polymerase buffer (Bioline, London, UK).

Supplementary Figure S2

Schematic explanation for the library quantification experiments and library pooling procedure.

Supplementary Figure S3

Correlation analysis between the sum of dRFU values between 75 and 95°C and quantification results from the TapeStation (a) and NanoDrop-based (b) methods.

Supplementary Figure S4

Coefficient of determination (R^2) between the PRN from MiSeq sequencing and sum of

417 dRFU values, depending on temperature ranges of the MC assay. The vertical and
418 horizontal axes denote lower and upper range of the temperatures (°C) of the MC
419 assay, respectively. The highest coefficient of determination (R^2 = 0.823), which is from
420 the PRN and dRFU values between 85 and 96°C, is marked with an asterisk.
421
422 **Supplementary Table S1**
423 Details for cost-reduced PCR amplicon sequencing library preparation with
424 fragmentation produced by the MspJI restriction endonuclease. A cumulative length of
425 15 kb of DNA sequence can be enriched through 5 PCRs (3 kb x 5 reactions) using the
426 MyFi™ DNA polymerase.
427
428 **Supplementary Table S2**
429 Details of the cost and processing duration assumption for the library normalisation or
430 quantification methods. Price information for the major reagents was obtained from the
431 websites of suppliers. Micropipette tips and 96-well microtiter plates are included in the
432 consumables, and the per-sample cost for each of those was estimated as US$0.04.
433 SPRI denotes the SPRI bead-based DNA size-selection and/or purification.

434 **Table 1**

435

Method	Method type	Size specificity	Library specificity	Per-sample cost (US$)	Processing time for 92-96 samples	Reference
SequalPrep	Normalisation	++*	-	0.91	1h30min	3, 7
BeNUS	Normalisation	++	-	1.04	1h30min	2
qMiSeq	Quantification	++	+	~10	c. 1 day	5
TapeStation	Quantification	++	-	2.75	2h15min	6
SPRI + real-time PCR	Quantification	+	+	1.41	4h50min	1, 4
SPRI + NanoDrop	Quantification	+	-	0.48	1h50min	3
MC assay	Quantification	+	-	0.15	40min	

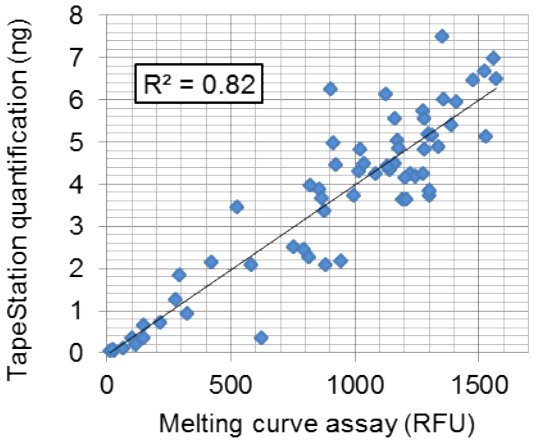
436 * DNA fragments between 100 bp and 20 kb are purified with the SequalPrep kit (Thermo Fisher Scientific)

437

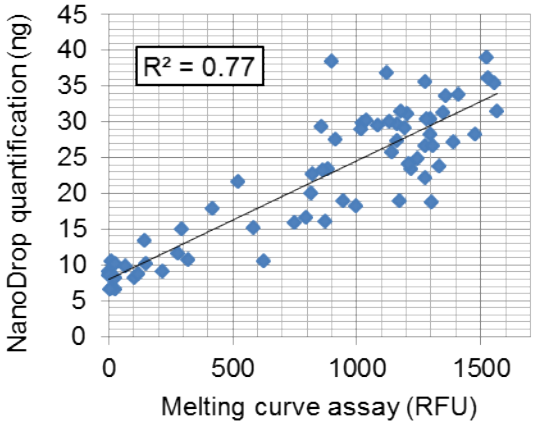
438

Figure 1

(a)


R² = 0.82
TapeStation quantification (ng)
Melting curve assay (RFU)
0
1
2
3
4
5
6
7
8
0
500
1000
1500

(b)


R² = 0.77
NanoDrop quantification (ng)
Melting curve assay (RFU)
0
5
10
15
20
25
30
35
40
45
0
500
1000
1500

(c)

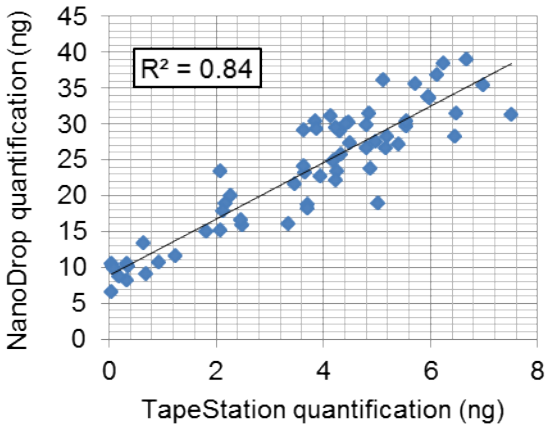
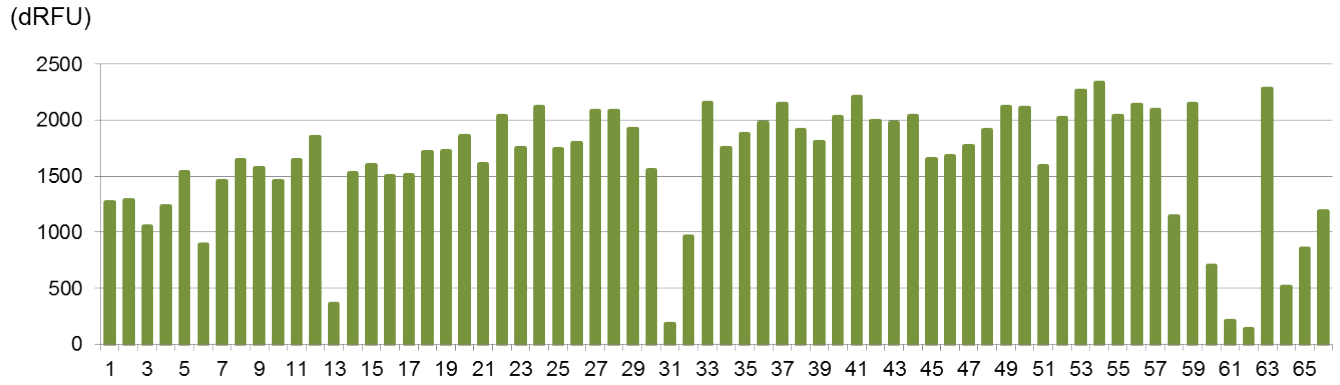

R² = 0.84
NanoDrop quantification (ng)
TapeStation quantification (ng)
0
5
10
15
20
25
30
35
40
45
0
2
4
6
8

Figure 2

(a)


(dRFU)
2500
2000
1500
1000
500
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65

(b)

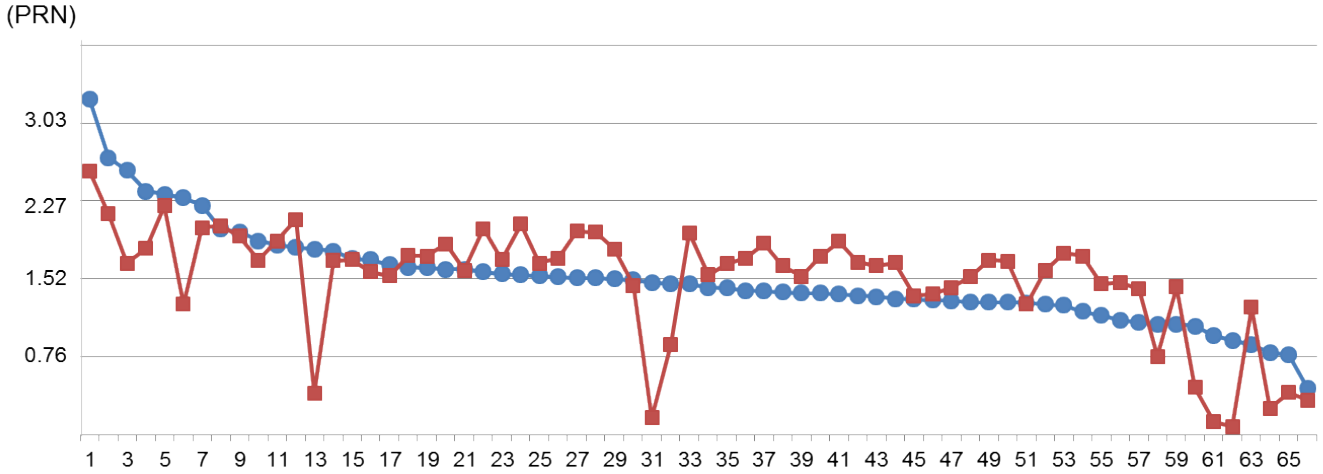
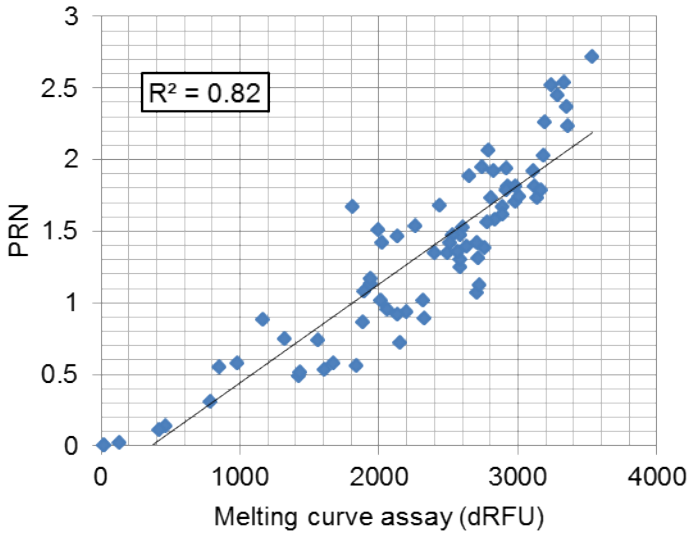
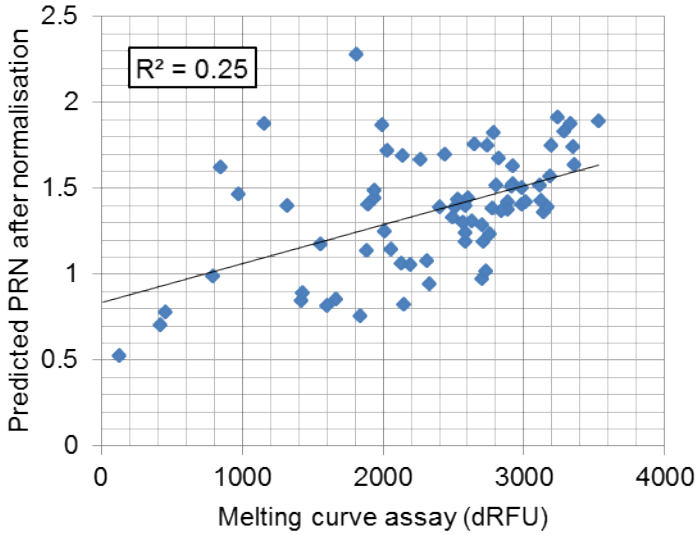

(PRN)
3.03
2.27
1.52
0.76
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65

Figure 3

(a)


R² = 0.82
PRN
Melting curve assay (dRFU)
0
0.5
1
1.5
2
2.5
3
0
1000
2000
3000
4000

(b)


R² = 0.25
Predicted PRN after normalisation
Melting curve assay (dRFU)
0
0.5
1
1.5
2
2.5
0
1000
2000
3000
4000