Validation of VERSA Mini NGLP for Automated Library Preparation with the Illumina Platform

Matthew J. Nesbitt[1], Matthew Kaplan[2], Sikander Gill[1], and Dong Liang[1]
[1]Aurora Biomed Inc., Vancouver, BC, Canada
[2]University of Arizona Genetics Core, Tucson, AZ, USA

Introduction

The plummeting cost of sequence acquisition that has resulted from advances in next generation sequencing (NGS) technologies has led to its use in new settings such as small research groups, contract research organizations, and the clinic. Widespread adoption of NGS is quickly approaching, in spite of an underdeveloped field of liquid handling workstations validated to complete the associated library preparation protocols. As this task requires a high number of sample manipulations, consistently reliable results depend on liquid handling automation.

The University of Arizona Genetics Core validated the use of the VERSA Mini NGLP (Aurora Biomed Inc.) for NGS library preparation. As an open-platform workstation, third party reagent kits can be used to drastically drop the cost per sample. In addition to this, the accompanying VERSAware user software enables control over aspiration/dispensing speeds and tip positioning that can be tailored to delicately handle the precious genomic sample.

Fig. 1. The VERSA Mini NGLP is ideal for library preparation because of its delicate sample handling. Reaction setups, incubations and purifications are executed in biosafety hood equipped with a UV lamp and HEPA.

Materials and Methods

Isolated gDNA from the Pocket Mouse (*Chaetodipus intermedius*) was used as a template for library preparation. Fragmentation was completed using a Covaris instrument (Woburn, MA, USA) and split into two equivalent volumes. One was processed manually using the Illumina (San Diego, CA) TruSeq library preparation kit (manual library). The other was processed in an automated fashion (automated library) with the VERSA Mini NGLP Workstation (Figure 1). Enzymatic modifications and purification steps were completed using the NEXTflex DNA Sequencing Kit from Bioo (Austin, TX).

Reaction setup for end repair, dA tailing, and barcoded adapter ligation were handled by the workstation, as were incubations and magnetic bead cleanups. Purified sequenceable libraries (one automated, one manual) were size selected via agarose gel electrophoresis and excision of a 400 - 500 bp target range. Amplification was then completed.

Both libraries were assessed to determine competency. A Bioanalyzer trace (Agilent Technologies, Santa Clara, CA) detailed the fragment size distribution. DNA concentrations were determined with the use of the PicoGreen reagent (Life Technologies, Carlsbad, CA). A qPCR using a KAPA Biosystems (Woburn, MA) Library Quant Kit determined the levels of adapter-bound fragments. After cluster generation and sequencing of the libraries using one lane on a HiSeq 2000 run, the data was characterized via bioinformatic analysis.

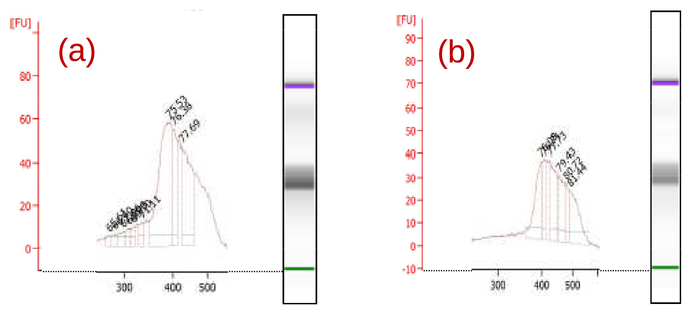

Fig. 2(a-b). Bioanalyzer traces for the automated (a) and manual (b) libraries. A dark electropherogram band in (a) proves the automated process efficiently recovered DNA within the range targeted by the size selection.

Results and Discussion

Library Preparation QC

Bioanalyzer traces for the manual and automated libraries indicated acceptable fragment size distribution profiles (Figure 2). Both libraries had similar traces; the manual library had an average fragment size of 423 bp compared to 417 bp for the automated library. A PicoGreen experiment determined total DNA concentrations of 157.16 nM and 180.66 nM for the automated and manual libraries, respectively (Table 1).

Table 1. Determination of total dsDNA concentration (PicoGreen) and adapter-bound fragments (qPCR).

Library	PicoGreen (nM)	qPCR (nM)
Automated	157.16	169.2
Manual	180.66	172.5

To ensure the library fragments were appropriately ligated to Illumina adapter sequences, a qPCR experiment using Library Quant Kits was completed. The C_t values (Figure 3) and concentration of adapter-bound fragments (Table 1) for the automated and manual libraries were highly similar and deemed appropriate for sequencing.

Both libraries passed all quality control filtration measures put in place by the University of Arizona. To confirm the libraries were a) representative of the original template and b) of necessary integrity to provide usable raw data, they were submitted to DNA cluster generation and sequencing on a single lane of a HiSeq 2000 instrument.

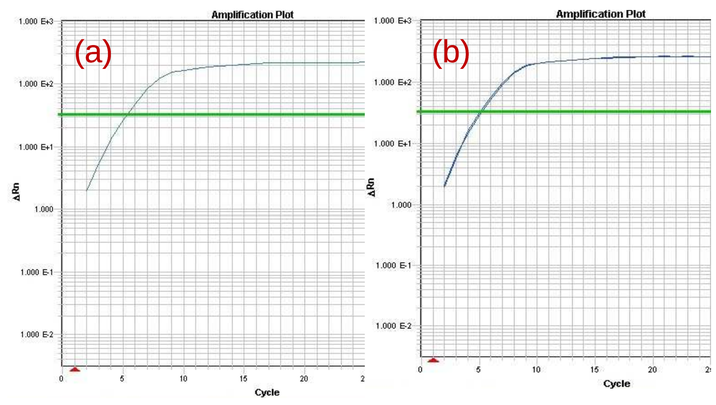

Fig. 3(a-b). Log (△Rn) vs. cycle count for qPCR analysis of adapter-bound fragments in the automated (a) and manual (b) libraries. A common threshold of 32.7 was used for both, with C_t values of 5.3 and 5.0 in (a) and (b), respectively.

Sequence Data QC

Resulting raw sequence data for 100 bp paired-end reads were sorted by barcode, converted to FastQ files and processed with Trimmomatic (Usadel lab, Max Planck Institute, Potsdam, GER). Adapter sequences were removed, leading and trailing bases were scanned, and a sliding window was used to trim reads at points over which average Q scores dropped below 15.

An understanding of the sequence characteristics was necessary to prove the automated method was viable. A FastQC analysis (Andrews lab, Barbraham Institute, Cambridge, UK) of the R1 read from the paired-end sequences revealed similarities between the two libraries for sequence quality and content. The per base sequence quality and overall read quality scores (Figure 4) illustrated that both sample preparations resulted in libraries of high integrity. In addition, both data sets indicated overall GC content to be 39% (data not shown). R2 sequence metrics mirrored those found in R1 (data not shown).

Differences between the sequence data sets generated with the two libraries did exist. The automated set contained approximately 75% more paired-end reads than the manual set (Table 2). This discrepancy could be attributed to a highly efficient DNA cluster generation with the automated library.

Both libraries appeared to have levels of sequence duplication that could have been introduced during the PCR amplification. FastQC analysis suggested that compared to the number of unduplicated reads, the manual and automated data sets had duplication levels of at least 92.7% and 77.3%, respectively, for the R1 reads (Figure 5). The R2 reads did not differ significantly.

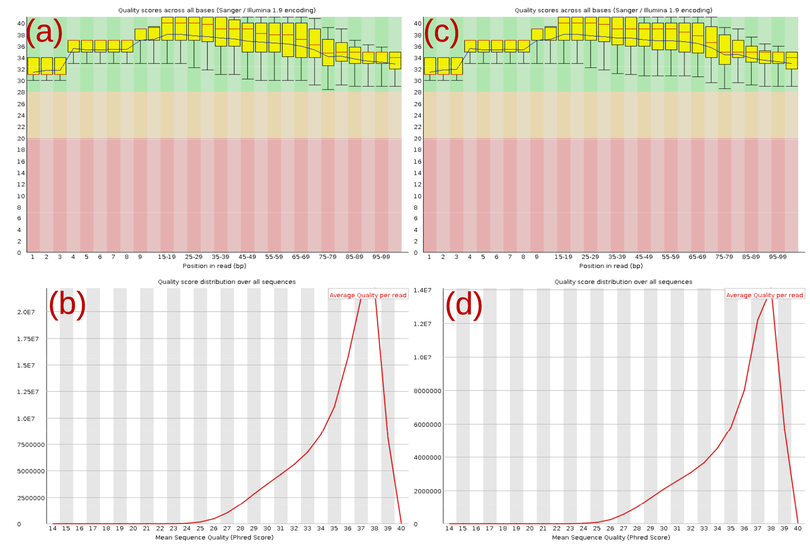

Fig. 4(a-d). Per base and per sequence Q scores for R1 ends of the automated (a, b respectively) and manual (c, d respectively) libraries. All median per base scores exceeded Q28, and the average read quality peaked at Q38 for both libraries.

Table 2. Paired-end reads for automated and manual data sets.

Library	Reads (M)
Automated	114.1
Manual	65.4

Validation work with the VERSA Mini NGLP Workstation suggests the instrument is a viable tool for executing NGS library preparations. The automated library and sequence data QC measures show that when compared to manual methods, VERSA offers comparable recovery of adapter-bound fragments and equivalent library integrity that produces high sequence quality scores. Furthermore the differences in read count and sequence duplication between the libraries suggests it is likely that VERSA produces libraries that are at least as complex as those produced manually.

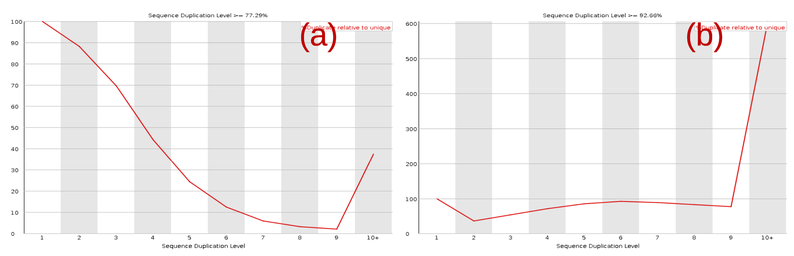

Fig. 5(a-b). Sequence duplication of R1 reads for the automated (a) and manual (b) data sets. The ratio of unique to duplicate reads is higher in the automated data set than the manual set.

Conclusions

The VERSA Mini NGLP workstation is capable of preparing DNA libraries for sequencing on the Illumina platform that are as competent as those produced with manual methods. Further more, automation provides the opportunity to scale levels of sample preparation while freeing up technician time for more complicated work.

Acknowledgments

The authors would like to thank Drs. Ryan Sprissler, Susan Miller and Joe Still at the University of Arizona for their wet bench and *in silico* work to generate the data used in this poster. Dr. Rajwant Gill provided technical support for the VERSA workstation, and Dr. Michael Nachman provided the genomic sample for experimentation.

For further information

Please address and email to Matthew Nesbitt or Sikander Gill (*info@aurorabiomed.com).*