

End to End Sample Quality Control for Next Generation Sequencing Library Preparation and SureSelect Target Enrichment on the Agilent 2200 TapeStation System

Application Note

Nucleic Acid Analysis

Author

Arunkumar Padmanaban
Agilent Technologies, Inc.
Bangalore, India

Abstract

The Agilent 2200 TapeStation system allows for the analysis of samples in the Next Generation Sequencing (NGS) workflow from the starting material, through to final quality control (QC) of the DNA library prior to sequencing. This Application Note follows the SureSelect library preparation protocol and demonstrates the performance and applicability of the Genomic DNA ScreenTape, D1000 ScreenTape, and High Sensitivity D1000 ScreenTape assays in analyzing samples from this workflow. The data demonstrates the 2200 TapeStation system as a reliable QC platform for sizing and quantification analysis of the libraries. The data also demonstrates that these DNA ScreenTape assays are comparable and within specification of the equivalent assays of the Agilent 2100 Bioanalyzer system.



Agilent Technologies

Introduction

Classical sequencing, based on the Sanger method, has been the predominant method for determining DNA sequences for more than 20 years¹. However, the rise of sequencing methodologies and technologies has led to massive increases in throughput over Sanger sequencing, and full genome sequences can be obtained in weeks, or even days². These technologies are collectively called next generation sequencing, or NGS.

As NGS technology and methods develop, the cost of sequencing has reduced substantially. However, the preparation of samples for sequencing is still complex and time consuming. Therefore, sample preparation for NGS needs to be carefully monitored to ensure a successful sequencing experiment. The Agilent SureSelect Target Enrichment Kit offers a complete, validated workflow of sample preparation for target specific sequencing with a range of capture sets including all exons, panels, kinome, custom, and other species. The library preparation involves isolating genetic material, shearing genomic DNA to generate shorter fragments, adding sequencing platform specific adaptors, hybridization to enrich targets, and tagging for multiplex sequencing. The whole preparation workflow needs to be quality controlled at several steps to ensure successful generation of the libraries before sequencing. The workflow overview is presented in Figure 1.

The Agilent 2200 TapeStation system is an ideal QC platform for quality and quantity assessment of NGS library preparation workflow solutions. The 2200 TapeStation system offers an automated electrophoresis system with rapid analysis time of 1 to 2 minutes per sample. This Application Note describes the quality control of NGS library preparation following the SureSelect protocol on 2200 TapeStation system with the Genomic DNA ScreenTape, D1000 ScreenTape, and High Sensitivity D1000 ScreenTape assays.

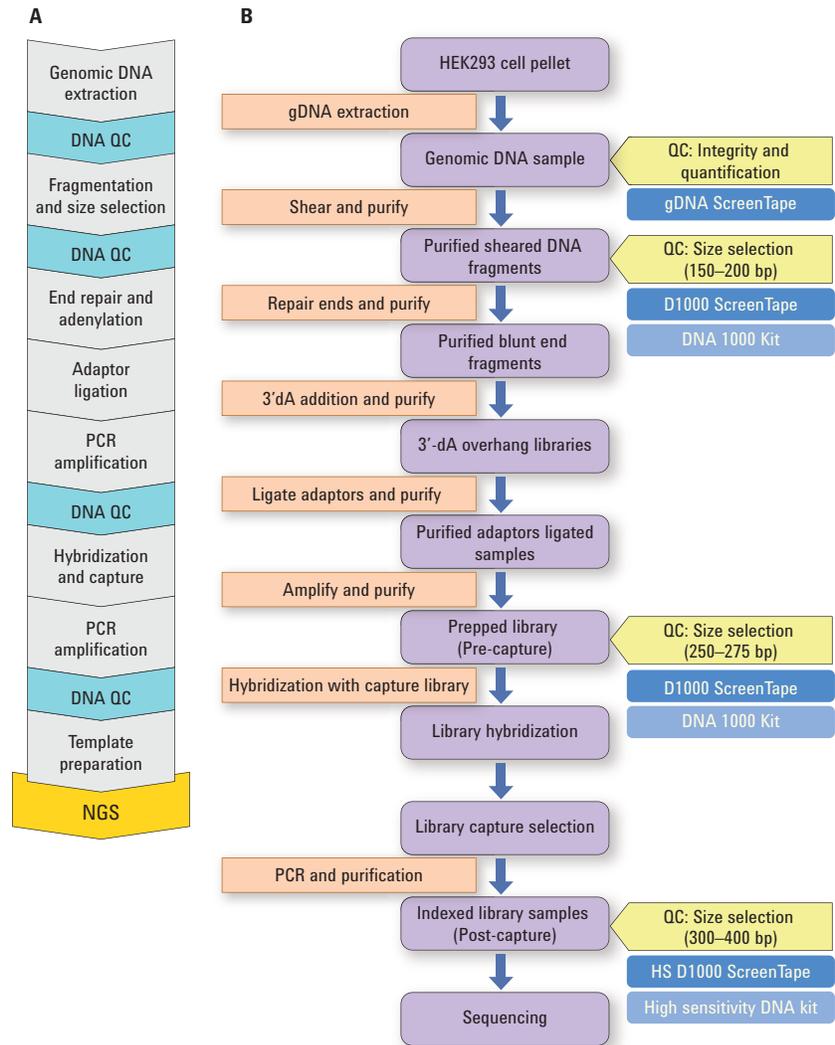


Figure 1. A) Workflow overview of the Agilent SureSelect^{XT} target enrichment system for Illumina paired-end sequencing library preparation highlighting the QC steps. B) Schematic diagram of the SureSelect library preparation workflow, detailing the recommended QC steps (yellow) with the associated Agilent 2200 TapeStation assays (dark blue) and Agilent 2100 Bioanalyzer assays (light blue).

Materials and methods

QC analysis

HEK293 cell lines from ATCC were cultured in MEM media containing 10 % FBS and 1 % Pen/Strep and incubated at 37 °C in 5 % CO₂ atmosphere. The cells were harvested, and genomic DNA was extracted using the Agilent DNA extraction kit (200600). Agilent 2200 TapeStation system (G2964AA), Genomic DNA ScreenTape and Reagents (p/n 5067-5365 and 5067-5366), D1000 ScreenTape and Reagents (p/n 5067-5582 and 5067-5583), High Sensitivity D1000 ScreenTape and Reagents (p/n 5067-5584 and 5067-5585), 2100 Bioanalyzer (G2939AA), DNA 1000 (p/n 5067-1504 and 5067-1505) and High Sensitivity DNA Kits (p/n 5067-4626 and 5067-4627) were obtained from Agilent Technologies. The Qubit 1.0 Fluorometer and NanoDrop 1000 (Thermo Scientific) were used for comparative QC analysis. The S220 Focused-ultra-sonicator (Covaris) was used for shearing the genomic DNA and was kindly provided by Genotypic Technology.

NGS Library preparation

NGS libraries for Illumina HiSeq and MiSeq Multiplexed platforms were prepared following the guidelines outlined in the Agilent SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library (version 1.6) protocol³ for a standard input of 3 µg DNA. SureSelect Reagent kit (G9611A) and SureSelect^{XT} Human All Exon V4+UTRs (5190-4636) capture library were obtained from Agilent Technologies and used according to the manufacturer's protocol. Agencourt AMPure XP kit (Beckman Coulter Genomics) and Dynabeads MyOne Streptavidin T1 (Life Technologies) were purchased and used as per manufacturer's guidelines. In general, the SureSelect protocol for NGS library preparation starts with extraction of genomic DNA. 3-µg of genomic DNA from a HEK293 cell pellet was sheared using a Covaris sonicator with the recommended shearing settings and purified. The purified fragments were end repaired and adenylated at the 3' end and further purified using magnetic beads. The adaptors specific for the Illumina platform were then ligated to the 3' end of the fragments and purified to remove excess unligated adaptors.

The adaptor-ligated fragments were hybridized with the biotinylated RNA library (capture baits) for 16 hours at 65 °C to enrich the targets. The captured libraries were then amplified and index tagged for multiplex sequencing.

All samples were run as triplicate on the 2200 TapeStation and the 2100 Bioanalyzer systems.

Results and Discussion

Current NGS technologies often require the generation of short fragment libraries of the DNA that is to be sequenced. The example of an Illumina sequencing protocol detailed here, shows that the initial fragment size range of DNA is recommended to be between 150–200 base pairs in length. Several steps were required to fully prepare the DNA for sequencing, as illustrated in Figure 1A and further detailed in Figure 1B. QC testing, at critical steps during this workflow, ensures the success of library generation. The 2200 TapeStation system checks the integrity of the starting genomic DNA, and assesses sample quality after integrated purification steps to the final QC step prior to sequencing on the NGS platform.

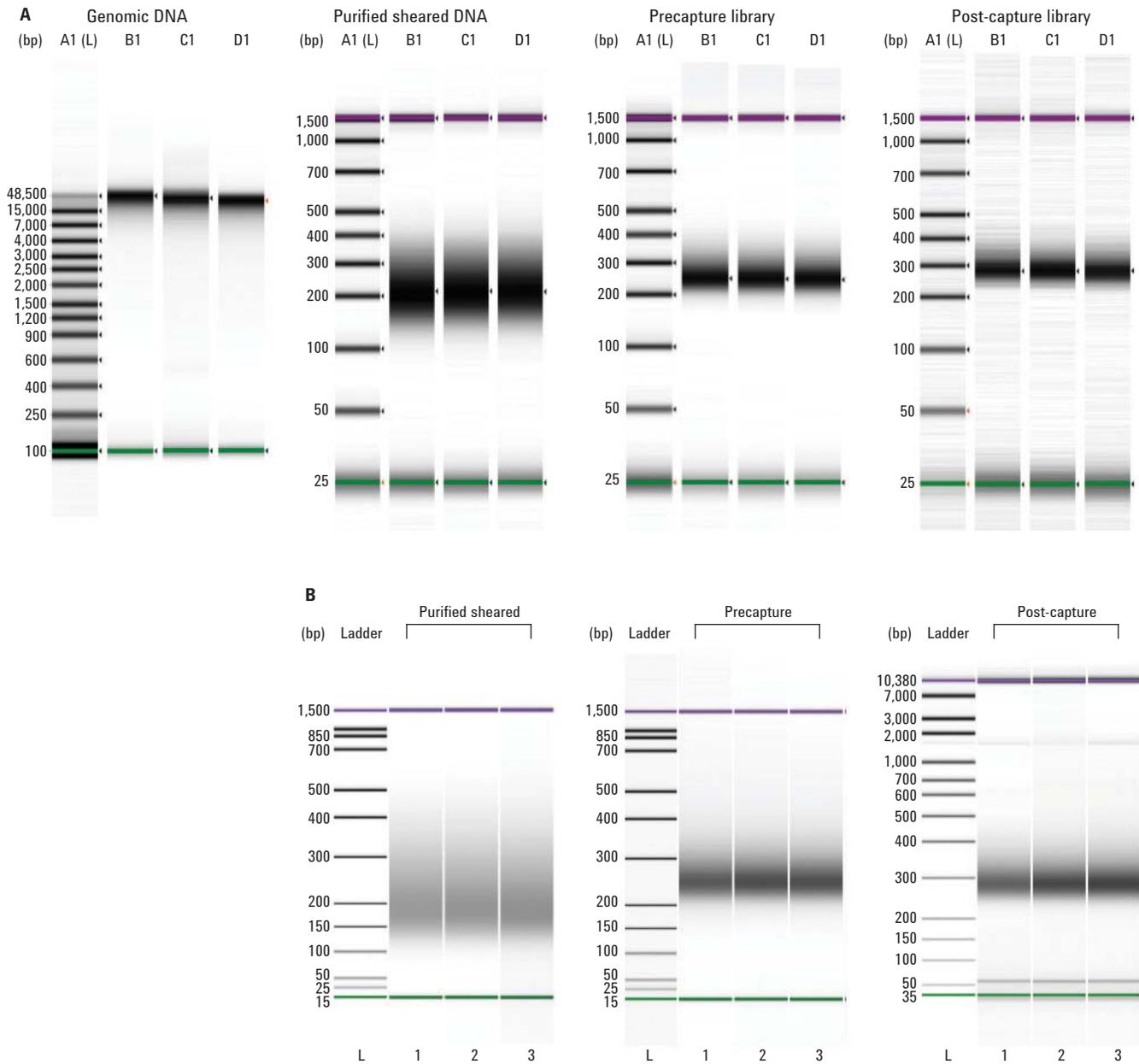


Figure 2. Gel images of the QC steps indicated in A obtained from the Agilent 2200 TapeStation system for genomic DNA, sheared DNA, precapture and post-capture library as analyzed on the Genomic DNA, D1000 and High Sensitivity D1000 ScreenTape respectively. Gel-like images of the QC steps indicated in B obtained from the Agilent 2100 Bioanalyzer system including sheared DNA, precapture and final library as analyzed on the DNA 1000 and High Sensitivity DNA assays. A genomic DNA assay is not available for the Agilent 2100 Bioanalyzer system.

Quantity and integrity check on starting genomic DNA material

The success of the NGS library preparation depends on the quality of the starting samples which, in this case, is genomic DNA. The genomic DNA was analyzed for quality and quantity using the Genomic DNA ScreenTape assay. Usually an OD_{260/280} ratio ranging from 1.8 to 2.0 obtained with the NanoDrop is used to qualify the purity of the samples, and the Qubit system is used to quantify the genomic DNA. In contrast, the 2200 TapeStation assay provides both sample quality and quantification within the same QC step. Sample purity can be determined by the gel image profile of the TapeStation Analysis Software with intact bands without any degradation products (Figure 3A). Data from the 2200 TapeStation system, NanoDrop, and Qubit are presented in Figure 3B and show that 2200 TapeStation system and Qubit are comparable in quantification. The measurement of genomic DNA with UV spectroscopy tends to overestimate the quantity due to other buffer components that may absorb in the UV spectrum⁴. Thus, in contrast to the other systems, the Genomic DNA ScreenTape assay with the 2200 TapeStation system offers both quality and quantity assessment in a single step using a sample volume of only 1 μ L.

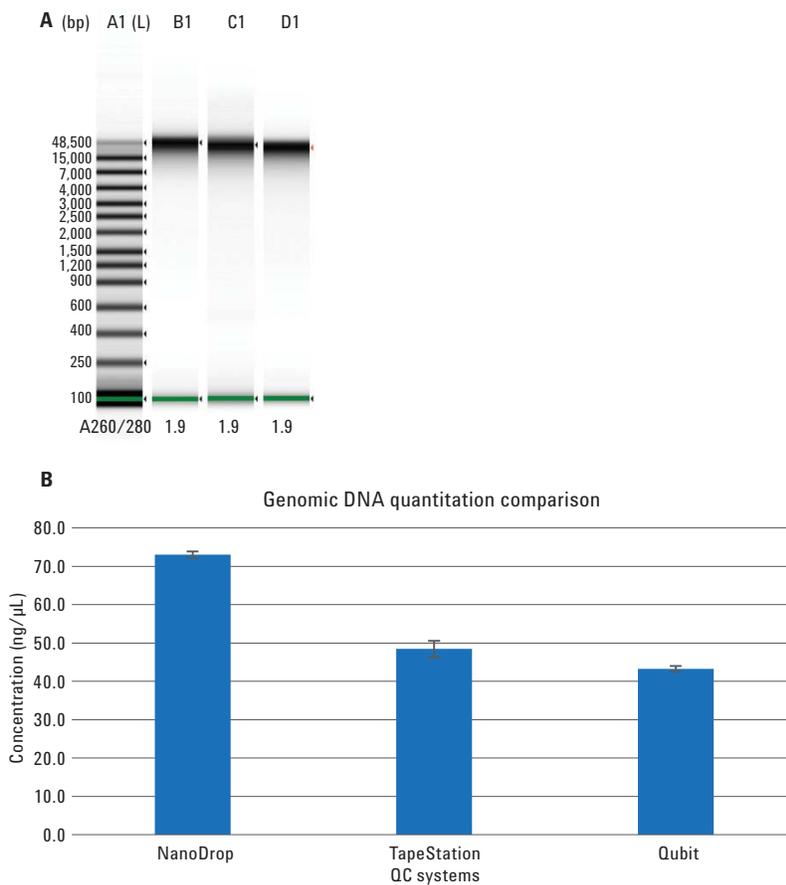


Figure 3. A) Agilent 2200 TapeStation system ScreenTape gel image of genomic DNA analysis. The Genomic DNA ScreenTape assay shows that the samples are intact, which is represented by the obtained A_{260/280} ratio from the NanoDrop (shown below each lane). B) Quantitation data from Genomic DNA ScreenTape assay compared against Qubit and NanoDrop.

SureSelect QC checkpoints

In addition to the initial genomic DNA QC of the starting material, the SureSelect protocol recommends two to three more checkpoints during the library preparation and target enrichment to ensure the correct quality of the samples (Figure 1):

- Optional purified and size-selected sheared genomic DNA is QC'd to check that the fragments give a single smear distribution with the median size ranging between 150 to 200 bp.
- Adaptor ligated, amplified, and purified adaptor ligated DNA libraries are also checked for even smear distribution with the median size ranging from between 225 to 275 bp. This step is often called Precapture QC.
- Final QC is recommended to be performed on the post-hybridized library (Post-capture QC) prior to sequencing. The median size range of the hybridized library should be between 250 to 350 bp.

The performance of the 2200 TapeStation system in quality control of these libraries was assessed by taking samples from each of the above steps and analyzing them in triplicate. The performance was compared to the 2100 Bioanalyzer system.

Sizing analysis

As indicated above, the sheared DNA, precapture, and post-capture DNA samples were analyzed on both the 2200 TapeStation and 2100 Bioanalyzer systems. The samples for the three QC steps were analyzed as triplicates and the collated data is presented in Figure 4. The data in the figure exhibits a very good correlation in sizing the libraries by the two systems.

The initial sheared DNA and precapture amplified libraries were analyzed using the D1000 ScreenTape assay for the 2200 TapeStation system and the DNA 1000 assay for the 2100 Bioanalyzer system.

The protocol recommends that the majority of the smear should be in the range of 150–200 bp for the initial sheared DNA samples. Using the 2200 TapeStation Analysis software for sizing evaluation the electropherograms (Figure 5A) shows that the majority of the smear falls within the recommended range, confirming successful shearing of the genomic DNA. The electropherogram is also comparable with the profile obtained with the 2100 Bioanalyzer system (Figure 5B).

The second QC step assesses the quality of the precapture amplified library before hybridization with capture baits. The SureSelect protocol indicates that the average smear size should be between 225–275 bp. The QC analysis of precapture amplification helps in troubleshooting any problems in the adaptor ligation prior to the hybridization step, saving time and reagent costs. The electropherograms show a single peak with an average size of approximately 250 bp indicating successful adaptor ligation (Figures 5C and 5D for 2200 TapeStation and 2100 Bioanalyzer system respectively).

The post-capture amplified libraries were analyzed using the High Sensitivity D1000 ScreenTape assay for the 2200 TapeStation system and the High Sensitivity DNA assay for the 2100 Bioanalyzer system. The SureSelect protocol recommends that the median size range of the hybridized library should be between 250 to 350 bp. The electropherogram profiles of the analyzed samples for both systems are presented in Figure 6. The electropherograms show a single peak with an average size of approximately 280 bp, indicating successful amplification of the libraries.

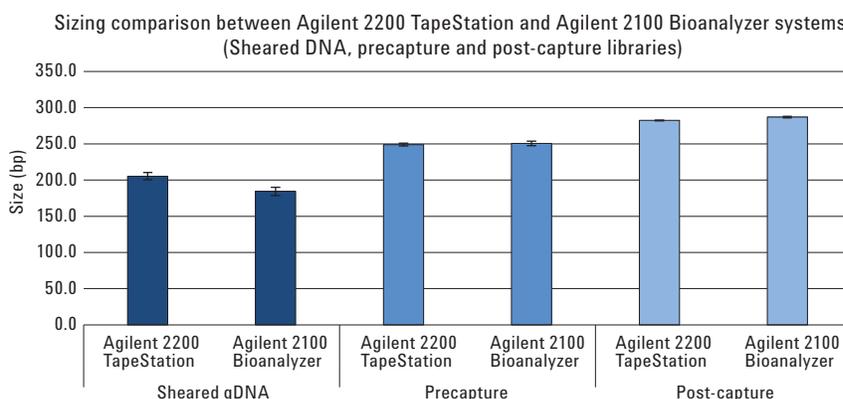


Figure 4. Sizing comparison between the Agilent 2200 TapeStation and Agilent 2100 Bioanalyzer systems. The average sizing from the triplicates were compared for sheared DNA, precapture library and post-capture libraries. Error bars shows the standard deviation.

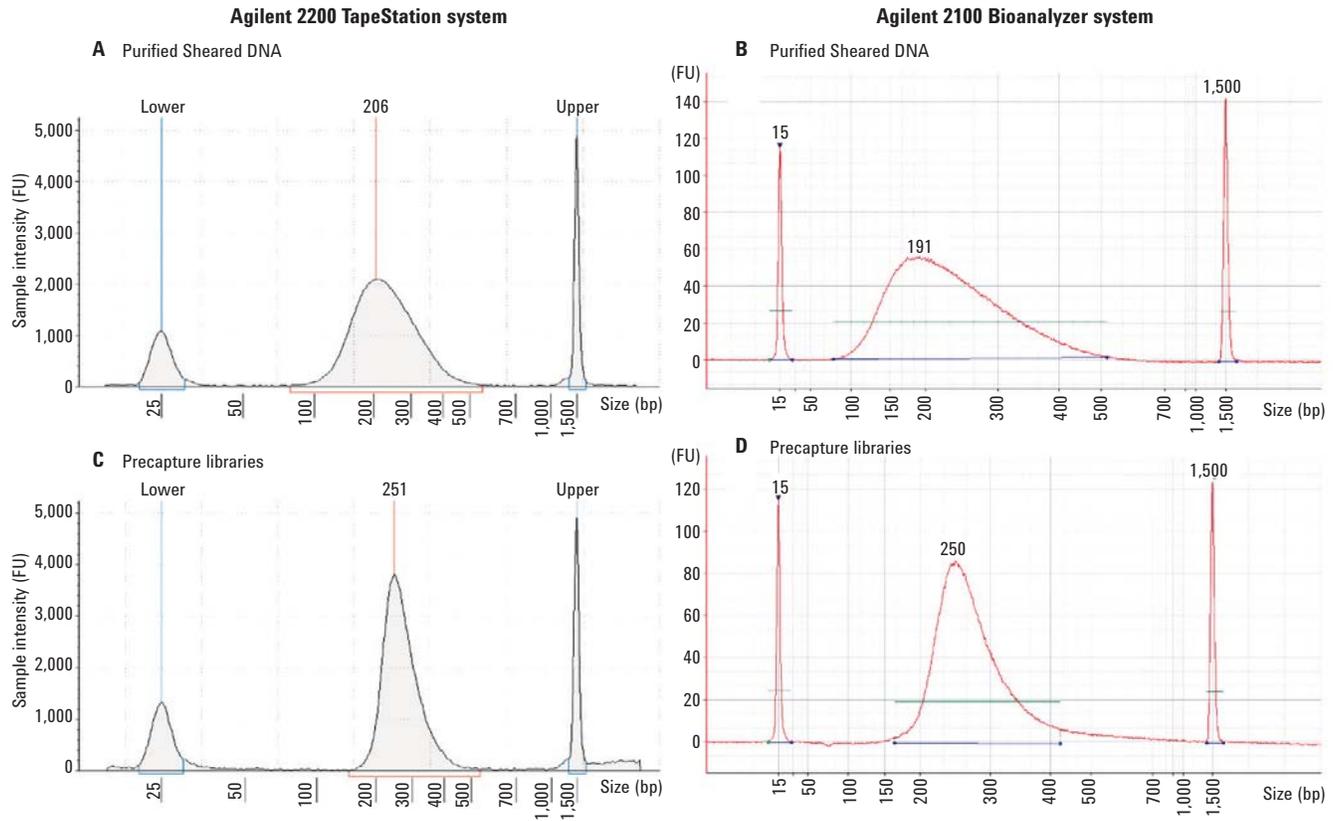


Figure 5. Electropherogram of purified sheared DNA (A and B) and precapture libraries (C and D) from the Agilent 2200 TapeStation (left column) and Agilent 2100 Bioanalyzer systems (right column).

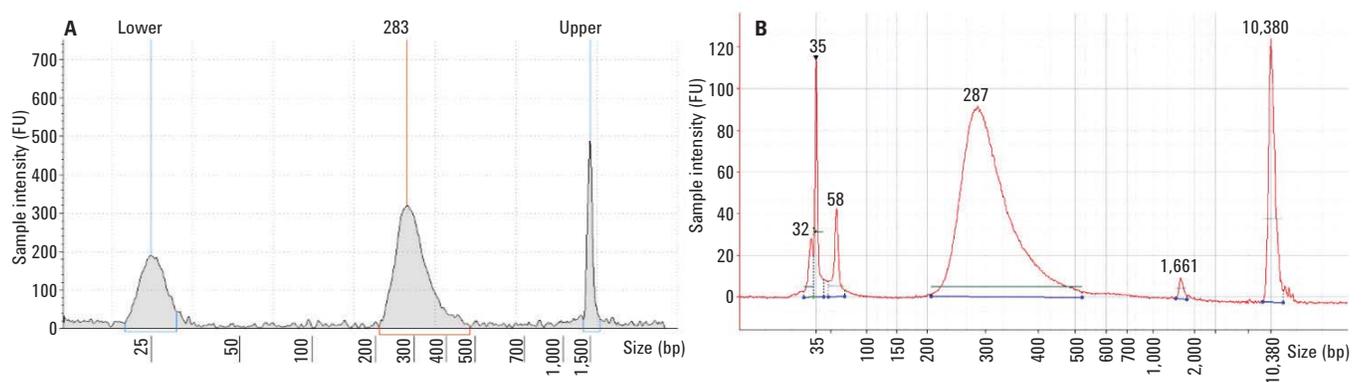


Figure 6. Electropherogram of post-capture amplification from the Agilent 2200 TapeStation system (A) shows a peak of 283 bp. For the same sample the Agilent 2100 Bioanalyzer system (B) shows a peak size of 287 bp.

Quantification analysis

The 2200 TapeStation and 2100 Bioanalyzer systems give quantitation data along with the sizing information. The quantification data from the QC analysis on sheared DNA and the precapture libraries using the standard assays is presented in Figure 7. The data show that quantification is comparable between the D1000 ScreenTape and the DNA 1000 assays of the 2200 TapeStation and 2100 Bioanalyzer systems, respectively.

The quantification of the post-capture library requires the use of high sensitivity assays since the amplification yields are at picogram levels. The quantification performance of the high sensitivity assays is presented in Figure 8. The data show comparable quantification comparing the results of the high sensitivity assays for the 2200 TapeStation and 2100 Bioanalyzer systems, respectively.

Optionally, the final library can be quantified using qPCR. However, some NGS users find that the molarity data obtained from the 2100 Bioanalyzer system at this stage is sufficient for their requirements. The quantification performance of the High Sensitivity D1000 ScreenTape assay was, therefore, assessed by analyzing a dilution series of a post-captured library, which was amplified for a higher number of PCR cycles. The molarity correlation for these diluted samples was compared with the High Sensitivity DNA assay and presented in Figure 9. The data show that the quantification of the post-capture samples can be reliably carried out using the High Sensitivity D1000 ScreenTape assay of the 2200 TapeStation system and is comparable to the High Sensitivity DNA assay of the 2100 Bioanalyzer system.

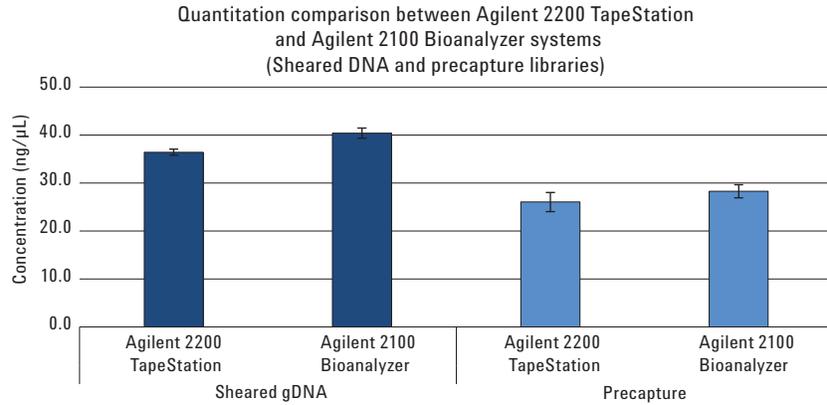


Figure 7. Quantitation comparison of sheared DNA and precapture DNA libraries from the SureSelect protocol on the 2200 TapeStation and 2100 Bioanalyzer systems.

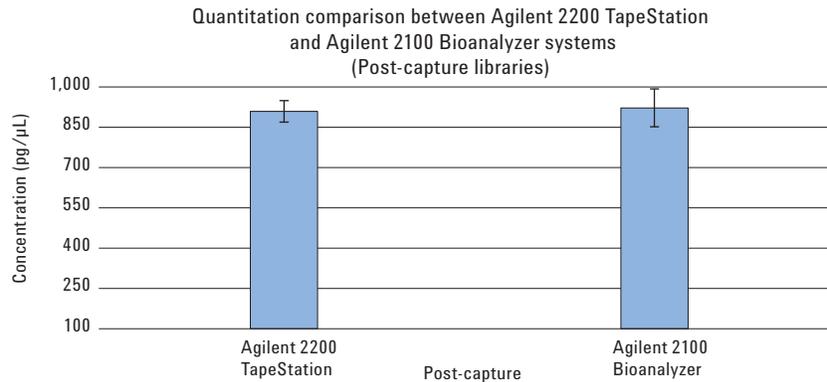


Figure 8. Quantitation comparison of postcapture DNA libraries using the high sensitivity assays on the Agilent 2200 TapeStation and Agilent 2100 Bioanalyzer systems.

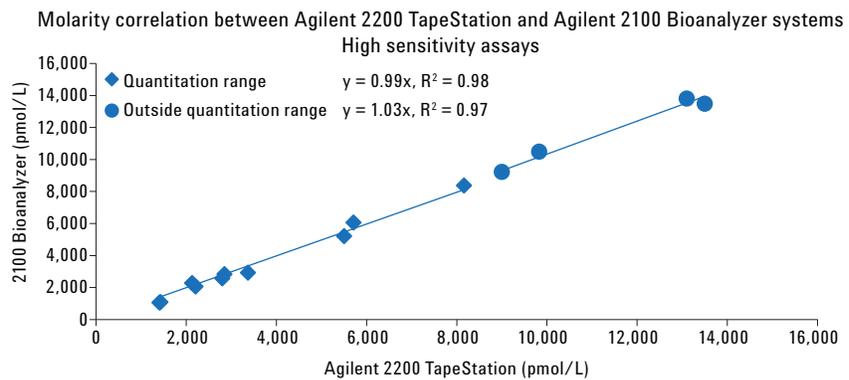


Figure 9. Graph showing correlation between the Agilent 2200 TapeStation and Agilent 2100 Bioanalyzer systems for molarity data obtained from a dilution series of the NGS post-capture amplified library.

Additional QC steps

In addition to the QC steps recommended in the SureSelect protocol, users often perform supplementary QC steps when undertaking NGS library preparation and target enrichment. These intermediate QC steps are useful to troubleshoot the progress of library synthesis. For example, if the libraries do not show a size shift after adaptor ligation, it indicates problem with adaptor ligation reaction. A subset of samples were taken throughout the NGS target enrichment process and analyzed using the D1000 ScreenTape assay on the 2200 TapeStation system and using the DNA 1000 assay on the 2100 Bioanalyzer system (Figure 10).

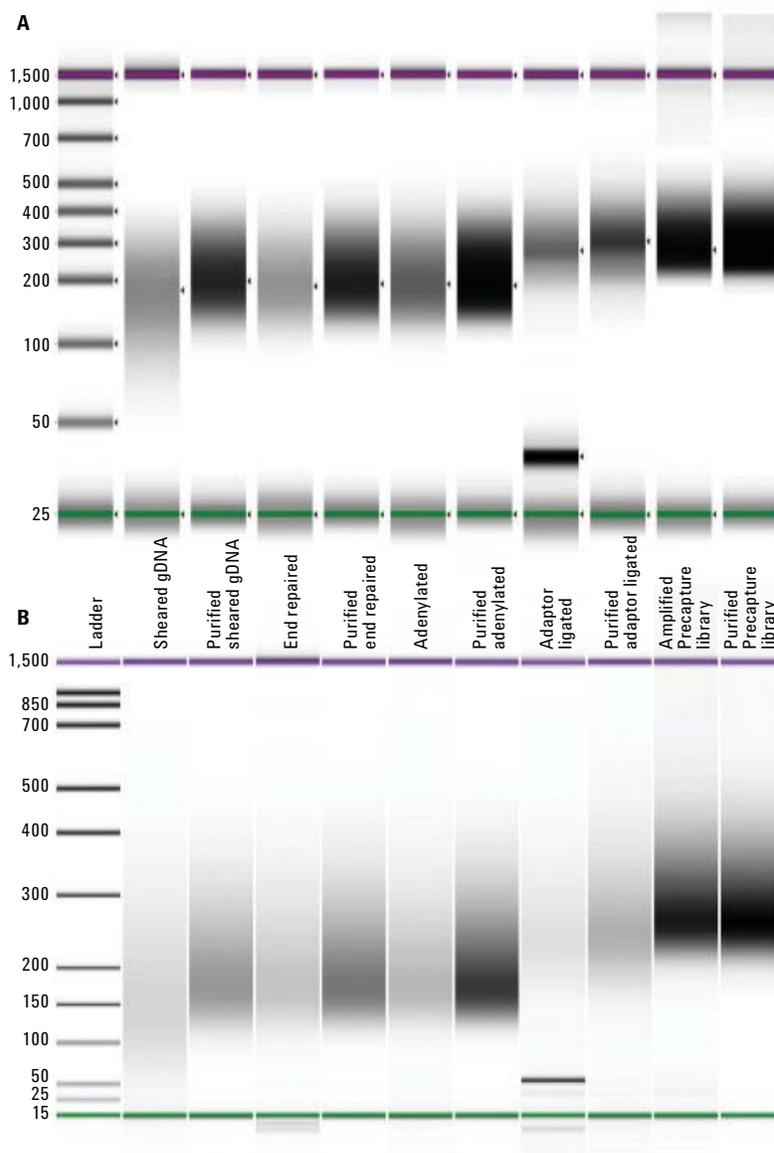


Figure 10. Gel images obtained from the D1000 ScreenTape assay of the Agilent 2200 TapeStation system (A) and DNA 1000 assay of the Agilent 2100 Bioanalyzer system (B) of samples taken from several stages within the NGS library workflow.

Adaptor ligation check

The NGS library preparation workflow involves addition of sequencing platform specific adaptors. The ligation of adaptors increases the sizes of the DNA fragments and unreacted adaptors can also be left over. The following purification step using magnetic beads should ensure the cleanup of the free adaptors, as carried through free adaptors can influence the quantification of the starting material processed for sequencing. Therefore, it is essential to check for the removal of excess adaptors. Electropherogram overlays from both systems (Figure 11) of purified adenylated and purified adaptor ligated samples show libraries with size shift indicating successful ligation of the adaptors. The overlay also shows no additional peaks, ensuring that the libraries are free of excess unreacted adaptors.

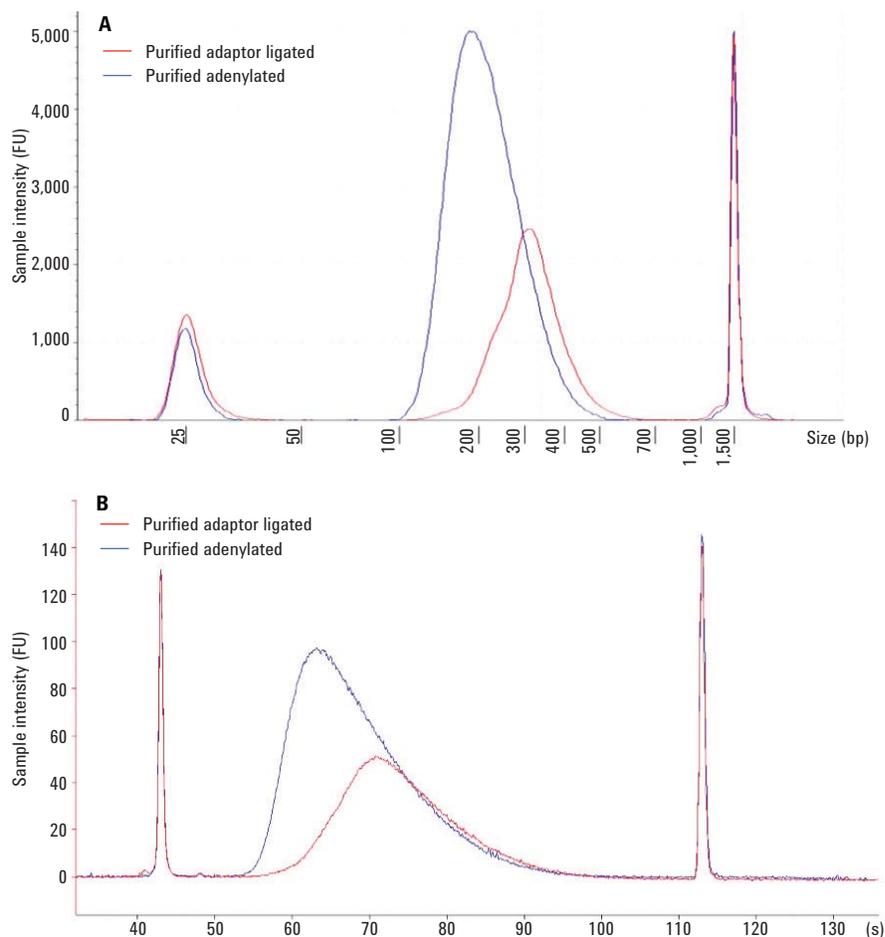


Figure 11. Electropherogram overlays of purified adenylated (blue) and purified adaptor-ligated (red) libraries showing higher size and no additional peaks from the Agilent 2200 TapeStation system (A) and the Agilent 2100 Bioanalyzer system (B).

Post hybridization PCR amplification optimization

The PCR amplification of post-captured library is often optimized for a particular number of amplification cycles. Under-amplification results in generating an insufficient amount of DNA for sequencing. Over-amplification can cause increased duplicates that will skew the sequencing coverage. An amplification of the post-captured libraries with 5, 8, and 10 PCR cycles was carried out and analyzed on the High Sensitivity D1000 ScreenTape assay. The electropherogram overlay (Figure 12) of the samples shows the impact of the PCR cycles on the amount of DNA generated. This study shows that the 2200 TapeStation system can be reliably used to optimize the PCR cycles due to fast analysis time and minimum sample requirement.

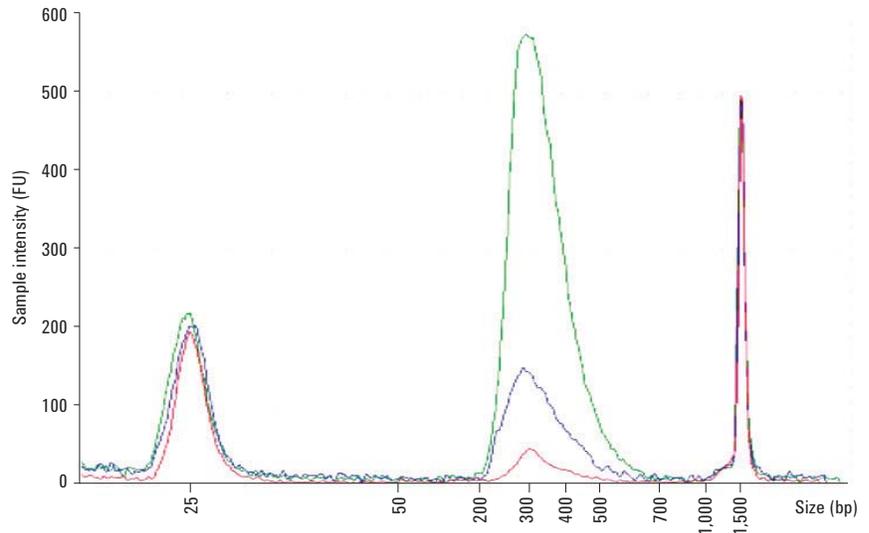


Figure 12. Analysis of increasing concentrations of amplified prepped post-captured library DNA using the Agilent 2200 TapeStation system after 5, 8, and 10 PCR cycles (blue, green, and red lines respectively).

Conclusion

By working through the SureSelect protocol, this Application Note shows that the Agilent 2200 TapeStation system, in combination with the Genomic DNA, D1000, and High Sensitivity D1000 ScreenTape assays, is ideal for the analysis of samples generated in the NGS library preparation workflow.

- Genomic DNA starting material can not only be assessed for quality, but also for quantity on the Genomic DNA ScreenTape assay, replacing two independent measurement methods with a single test.
- The sizing of NGS samples throughout the SureSelect target enrichment workflow can be reliably undertaken on the D1000 ScreenTape assays in conjunction with the 2200 TapeStation system. These size outputs match the equivalent assays on the Agilent 2100 Bioanalyzer system.
- Quantification of the final library can be performed using the High Sensitivity D1000 ScreenTape assay for the 2200 TapeStation system, which matches the respective outputs from the High Sensitivity DNA assay for the 2100 Bioanalyzer system.

References

1. O. Morozova, M.A. Marra
“Applications of next-generation sequencing technologies in functional genomics” *Genomics*, Volume 92, Issue 5, November **2008**, p. 255-264, ISSN 0888-7543.
<http://dx.doi.org/10.1016/j.ygeno.2008.07.001>.
2. “An Introduction to Next-Generation Sequencing Technology”
http://res.illumina.com/documents/products/illumina_sequencing_introduction.pdf
3. SureSelectXT Target Enrichment System for Illumina Paired-End Sequencing Library - Illumina HiSeq and MiSeq Multiplexed Sequencing Platforms, Protocol 1.6, http://www.chem.agilent.com/library/usermanuals/Public/G7530-90000_SureSelect_IlluminaXTMultiplexed_1.6.pdf
4. M. O’Neill, J. McPartlin, K. Arthure, S. Riedel, and Nd McMillan, “Comparison of the TLDA with the Nanodrop and the reference Qubit system” *J. Physics: Conference Series* Volume 307, Issue 1, **2011**, <http://iopscience.iop.org/1742-6596/307/1/012047>.

www.agilent.com/genomics/tapestation

This information is subject to change without notice.

© Agilent Technologies, Inc., 2014
Published in the USA, January 1, 2014
5991-3654EN

For Research Use Only. Not for use in diagnostic procedures.



Agilent Technologies