Use of the Agilent 4200 TapeStation System for Sample Quality Control in the Whole Exome Sequencing Workflow at the German Cancer Research Center (DKFZ)

Application Note

Nucleic Acid Analysis

Authors

Jana Petersen, Marie Beckhaus, and Stephan Wolf
DKFZ Genomics and Proteomics Core Facility,
High Throughput Sequencing Unit,
Heidelberg, Germany

Elisa Viering
Agilent Technologies, Inc.
Waldbronn, Germany

Abstract

This Application Note demonstrates the use of the Agilent 4200 TapeStation system as a quality control (QC) tool in a next generation sequencing (NGS) workflow. The whole exome sequencing workflow at the Genomics and Proteomics Core Facility of the German Cancer Research Center (DKFZ) includes several QC steps, of incoming samples, intermediate products during library preparation, final libraries, and sequencing data. These QC steps include quantification and qualification of the DNA, and can be done with the 4200 TapeStation System.

Agilent Technologies

Introduction

The German Cancer Research Center (DKFZ) is one of the largest biomedical research institutions in Germany. The High Throughput Sequencing Unit of the DKFZ Genomics and Proteomics Core Facility provides sequencing services to national cancer genome projects for multiple applications including whole exome sequencing.

Mandatory for the experimental success of whole exome sequencing is the quality of the incoming genomic DNA (gDNA) material and the DNA samples at various stages of the library preparation workflow. The High Throughput Sequencing Unit subjects the received gDNA samples to an incoming quality control (QC), prepares the libraries with additional QC steps, and performs the sequencing. The sequencing is followed by the final data QC analysis according to the workflow shown in Figure 1.

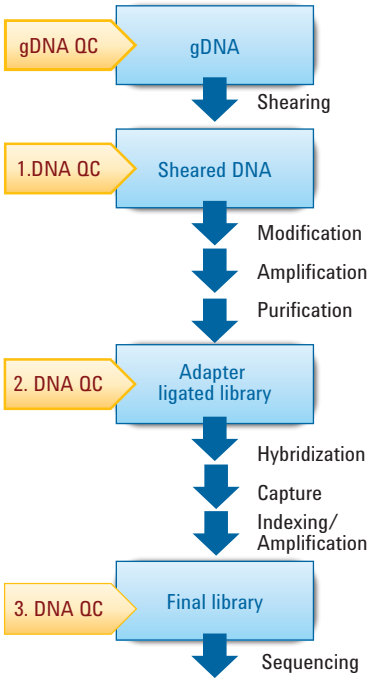

Figure 1. Workflow of the Agilent SureSelectXT automated target enrichment protocol for Illumina paired-end multiplexed sequencing established at the DKFZ High Throughput Sequencing Unit. The QC steps are performed by using the Agilent 4200 TapeStation system with the Agilent Genomic DNA and Agilent D1000 ScreenTape assays.

QC analyses consist of quantification and qualification of the DNA samples. The quantification of the DNA samples is performed using the Qubit assay based on fluorescence detection with a microplate reader. Integrity of the DNA samples is verified with an Agilent 4200 TapeStation system at the High Throughput Sequencing Unit. The use of both instruments enables processing of up to 96 samples at once. Data are then automatically transferred into a Laboratory information management system (LIMS) for further evaluation.

Experimental

Materials

The VolumeCheck was from Micronic (AR Lelystad, The Netherlands), the Qubit dsDNA HS Assay Kit was from Thermo Fisher Scientific Inc. (Waltham, MA, USA), and the FilterMax F3 Multi-Mode Microplate Reader was from Molecular Devices (Sunnyvale, CA, USA). The Covaris E220 instrument and the 96 micro TUBE Plate from Covaris (Woburn, MA, USA) were used. The HiSeq 3000/4000 PE Cluster Kit, cBot and the HiSeq 3000/4000 SBS Kits 50, 150, and 300 cycles, HiSeq 4000 System was obtained from Illumina, Inc. (San Diego, CA, USA). The Mastercycler Pro and the Concentrator plus (speed vac) from Eppendorf (Hamburg, Germany) and the centrifuge Universal 320/320R from Hettich (Beverly, MA, USA) was used.

Instruments

The following were obtained from Agilent Technologies (Waldbronn, Germany).

- Agilent 4200 TapeStation system (G2991AA)
- Agilent Bravo NGS workstation (G5522A)
- Agilent Genomic DNA ScreenTape (p/n 5067-5365)
- Agilent Genomic DNA Reagents (p/n 5067-5366)
- Agilent D1000 ScreenTape (p/n 5067-5582)
- Agilent D1000 Reagents (p/n 5067-5583)
- Agilent Sure-SelectXT Human All Exon v5 kit (p/n 5190-6210)

Samples

A batch of 88 gDNA samples from FFPE tumor tissue was provided by a customer. Eight samples were not further processed after failing the incoming QC check, and were replaced with positive and negative controls. The positive control used was commercially available Human Genomic DNA from Roche Diagnostics GmbH (Mannheim, Germany). For the negative controls, buffer instead of DNA was used for shearing.

Method

Library preparation for exome sequencing

The exome libraries were prepared according to the Agilent Low Input Sure-Select^{XT} Human All Exon v5 protocol following the workflow shown in Figure 1. The exome libraries were prepared using automated processing in a 96-well plate format using the Agilent Bravo NGS Workstation. The Covaris fragmentation was performed for 8 minutes, and the hybridization was done at 65 °C for 16.5 hours.

DNA analysis

The 4200 TapeStation system with the Agilent Genomic DNA and Agilent D1000 ScreenTape assays was used for DNA analysis according to the appropriate ScreenTape Quick Guides[2-3].

Sequencing

For sequencing, samples were normalized to 10 nM, and equimolar pooled on the Agilent Bravo NGS Workstation. For the normalization, the region molarity of the finished libraries determined by the 4200 TapeStation system can be used.

The 80 libraries were equimolar pooled to six 12-plexes and one 8-plex. Each pool was sequenced on two lanes using the Illumina HiSeq 4000 System with 100 bp paired end sequencing (PE100). The Illumina-specified 8 bp index tags, provided with Sure-Select^{XT} target enrichment kits, were used for demultiplexing.

Results and Discussion

After initial quality control, 80 customer samples were prepared according to the Agilent Low Input Exome Sequencing Human v5 protocol using automated processing in a 96-well plate format. The exome libraries were then sequenced on an Illumina HiSeq 4000 system.

Incoming gDNA QC

To determine if the samples were suitable for library preparation, a QC was performed at the beginning for the batch of 88 gDNA samples from FFPE tumor tissue by the High Throughput Sequencing Unit. This initial QC includes quantification and analysis with a 4200 TapeStation system and the Genomic DNA assay to determine DNA quality based on the DNA integrity number (DIN). The DIN algorithm is included in the 4200 TapeStation software, and provides a numerical assessment of the DNA sample by assigning a numerical score for DIN from 1 to 10. A high DIN indicates highly intact gDNA, and a low DIN a strongly degraded gDNA sample[4].

Figure 2 shows a representative subset of the data collected on the 4200 TapeStation system with the Genomic DNA ScreenTape assay. gDNA samples extracted from FPPE material are often of low DNA integrity, but can still be sufficiently intact for whole exome library preparation and successful sequencing. Depending on the quality of the DNA material, modified fragmentation protocols can be applied for optimized results.

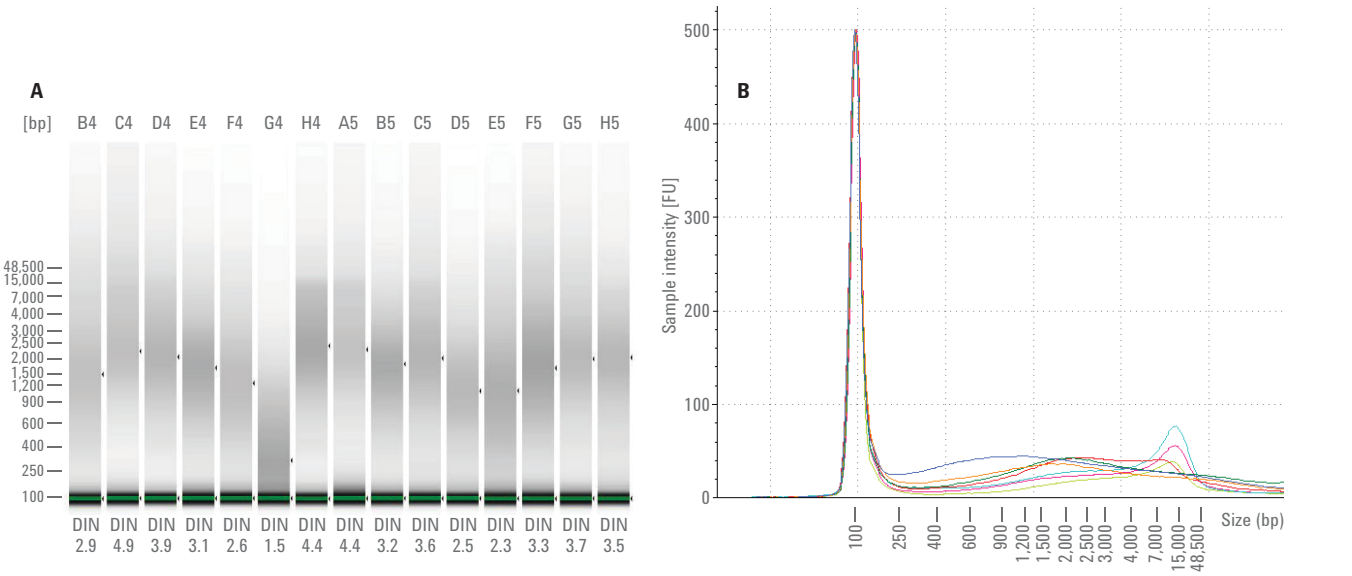

Figure 2. QC for gDNA extracted from FFPE tissue was performed by using the Agilent 4200 TapeStation system and the Agilent Genomic DNA ScreenTape assay. A) The gel image of 15 samples with the determined DIN on bottom and the well on top of each lane are shown. B) An overlay of the electropherograms of 15 samples. Both views enable to compare and visualize the integrity of the samples.

Based on empirical evidence, the High Throughput Sequencing Unit had established a DIN threshold of 7 for their incoming gDNA samples. Figure 3A shows the DNA integrity of all FFPE samples in this study. DIN values ranged from DIN 1.3 to 6.2, and are well below the QC threshold of DIN 7. Samples with lower DIN require a special review, and are only subjected to library preparation if there was approval to proceed from the customer. Depending on the quality, modified protocols can be used. However, in these cases, successful library preparation cannot be guaranteed.

The second criterion for QC of incoming samples is the calculation of total DNA abundance. A minimum of 450 ng gDNA is required for DNA library preparation. Samples were quantified in triplicate using the Qubit system. The Qubit assay is the established quantification method of the High Throughput Sequencing Unit. The concentration analysis of the 4200 TapeStation was used as control to double check the Qubit quantification.

To calculate the total amount of DNA present in the sample, the remaining sample volumes were determined. Based on the measured volume and concentration, the total available DNA amount was calculated (Figure 3B).

The DNA amount of most of the 88 tested gDNA samples was relatively low. Despite the low gDNA quality and the low DNA amount, it was decided to proceed with the library preparation for 80 of the samples. Eight samples were not further processed after the incoming QC, and were replaced with positive and negative controls.

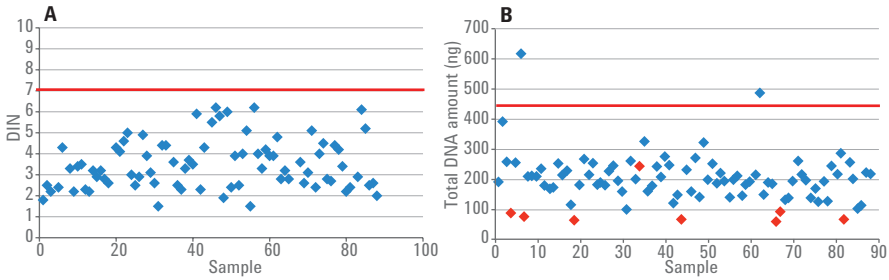

Figure 3. The incoming QC consisting of quality analysis was performed using the Agilent 4200 TapeStation system and the Agilent Genomic DNA ScreenTape assay. A) Qualification of the 88 gDNA samples with the DNA Integrity number DIN, the red line indicates the incoming QC criteria for DNA quality (DIN ≥ 7). B) Quantification of the 88 gDNA samples. The total DNA amount was calculated based on the measured concentration and the volume. The eight gDNA samples marked in red were not further processed. The red line indicates the incoming QC criteria for DNA quantity (≥ 450 ng).

Library preparation for exome sequencing

The exome libraries were prepared according to the Agilent Low Input Exome Sequencing Human v5 protocol following the workflow shown in Figure 1. Intermediate QC steps were taken throughout the protocol to monitor library preparation for sequencing, such as evaluation of DNA after fragmentation, analysis of adapter-ligated and amplified DNA, and lastly, qualification of the final library.

In the first step, gDNA was fragmented. Due to the relatively strong degradation of the samples (DIN value between 1.3 to 6.2), the Covaris fragmentation was performed for 8 minutes instead of 6 as described in the standard protocol[1]. Then, the size of the fragmented DNA was determined with the 4200 TapeStation system and D1000 ScreenTape assay.

Figure 4 shows 4200 TapeStation data of Covaris-treated DNA samples. In accordance with the NGS library preparation protocol[1], all sheared DNA samples have maximum peak size between 150 and 190 bp.

After fragmentation, DNA ends were modified for downstream target enrichment, including end-repair, A-tailing, and adaptor ligation. After the modification steps, the amplified DNA samples were purified using AMPure XP beads. The size and the concentration of the purified DNA were determined with the 4200 TapeStation system and the D1000 ScreenTape assay (Figure 5). The total DNA amount was calculated based on the available volume of 30 µL.

According to the Agilent Low Input Sure-Select^{XT} Human All Exon v5 protocol[1], the libraries should have a peak size between 225 and 275 bp. Only two samples were slightly below the recommended 225 bp. The hybridization protocol in the following section requires 750 ng of each amplified DNA library. Two DNA samples were slightly below the recommended total DNA amount (Figure 5). Three samples did not fulfill the QC criteria based on size or quantitation, but were processed through the workflow as automated library preparation did not allow the exclusion of individual samples.

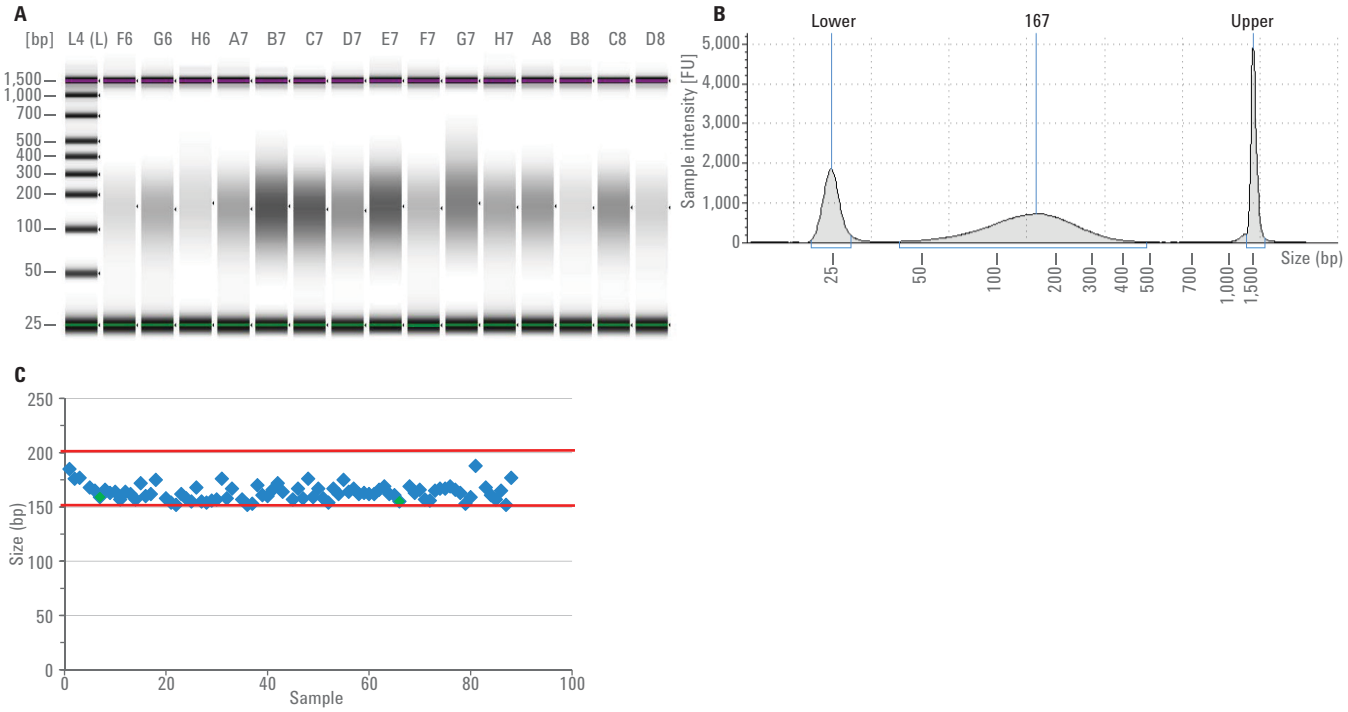

Figure 4. The sizing analysis of the fragmented DNA was performed using the Agilent 4200 TapeStation system and the Agilent D1000 ScreenTape assay. A) Gel view of 15 samples. B) Example of an electropherogram of one sample. The Agilent D1000 Assay includes a lower and upper marker. C) The size distribution of all 80 samples plus eight controls. The two positive controls are shown as green symbols. The red lines indicate the recommend size range (150 to 200 bp).

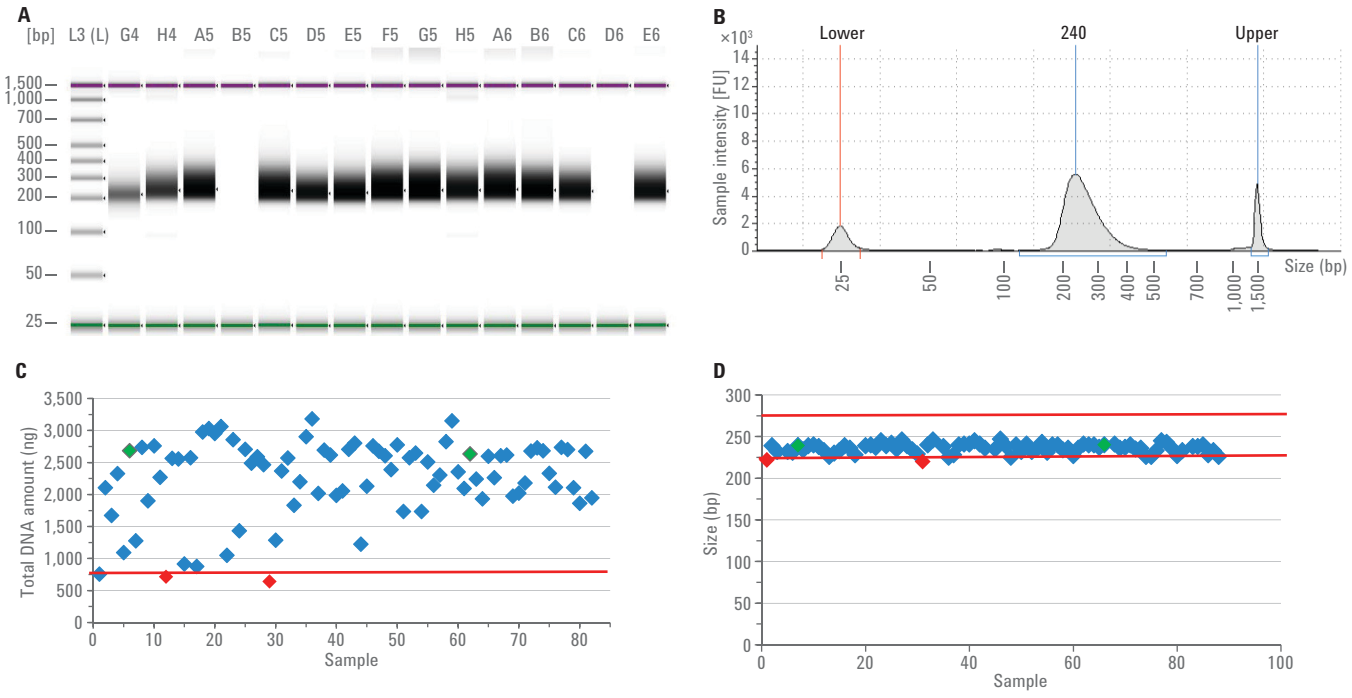

Figure 5. The sizing analysis of adapter-ligated library was performed using the Agilent 4200 TapeStation system and the Agilent D1000 ScreenTape assay. A) Gel view of 15 samples, lane B5 and D6 show negative controls. B) Example of an electropherogram of one sample. The Agilent D1000 Assay includes a lower and upper marker. C) The total DNA amount for all 80 samples plus eight controls. The two positive controls are shown as green symbols. The red lines indicate the threshold (750 ng). D) The size distribution of all 80 samples plus eight controls. The two positive controls are shown as green symbols. The red lines indicate the recommend size range (225 to 275 bp).

Due to the addition of index sequences, a size shift between adapter-ligated (Figure 5) and finished library (Figure 6) was expected and detectable on the 4200 TapeStation system.

The final libraries were expected to be sized between 250 and 350 bp with a minimum concentration of 2 ng/µL. The 4200 TapeStation and Qubit systems were used for quantification. This study exhibited successful DNA library preparation for all 80 samples, the six positive control samples, and even the three samples that were slightly outside the QC criteria after adapter ligation. All customer samples were used for sequencing.

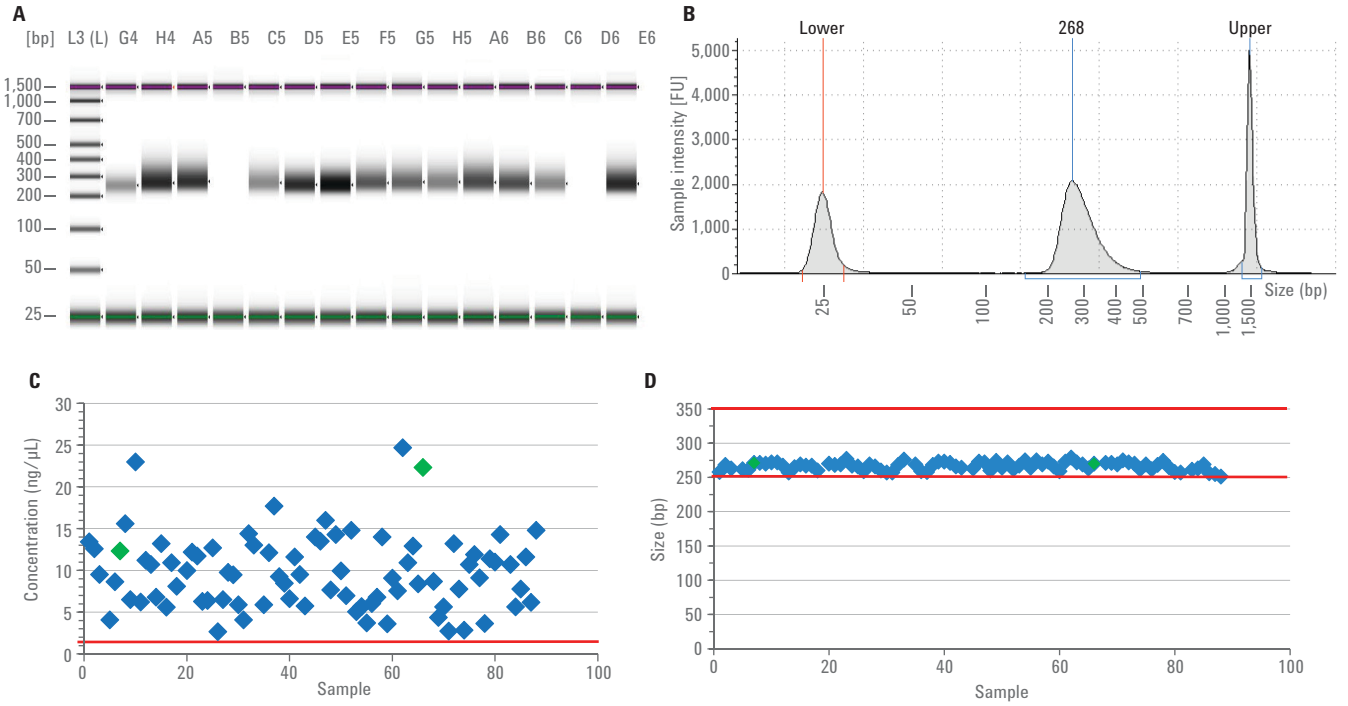

Figure 6. The sizing analysis of finished library was performed using the Agilent 4200 TapeStation system and the Agilent D1000 ScreenTape assay. A) Gel view of 15 samples, lane B5 and D6 show negative controls. B) Example of an electropherogram of one sample. The Agilent D1000 Assay includes a lower and upper marker. C) The distribution of the concentration for all 80 samples plus eight controls. The two positive controls are shown as green symbols. The red lines indicate the recommend concentration threshold (2 ng/µL). D) The size distribution of all 80 samples plus eight controls. The two positive controls are shown as green symbols. The red lines indicate the recommend size range (250 to 350 bp).

Sequencing

The 80 libraries were equimolar pooled; each pool was sequenced on two lanes using the Illumina HiSeq 4000 System (PE 100).

Table 1 summarizes the data QC criteria for whole exome sequencing.

Figure 7 summarizes the sequencing results. The mapping rate (dark blue) for all samples is approximately 100 %. Between 55 to 70 % of the mapped reads are on-target (green). The data of both sequenced lanes are merged to calculate the average target coverage (turquoise); the coverage for all samples is above 100 %. The duplicate rate (red) varies between 4 and 15 %. VerifyBAM FREEMIX (purple) is a tool to determine the contamination level of genomes. Only four of the 80 samples have a contamination level slightly above 2 %. Structurally, changes in DNA, which occur often in tumor samples, can cause this increase of the VerifyBAM FREEMIX level.

Despite the relatively low DNA integrity and total DNA amount of the gDNA starting material, it was possible to obtain meaningful sequencing results.

In summary, all samples were successfully prepped and sequenced. This was feasible due to the extensive monitoring of the process with multiple quality control steps, starting with the integrity evaluation of the incoming DNA material. Due to the high degradation of the DNA samples, an alternative fragmentation protocol was used. The DIN makes it possible to categorize the quality of DNA samples, and set thresholds for the downstream workflows to ensure a successful library preparation. Low DIN values allowed appropriate modifications of the fragmentation protocol, which assisted in successful library preparation. The QC of the intermediately processed samples during the library preparation enabled monitoring the success of fragmentation, ligation of adapters and indices, and quantification of the final product, to ensure that the final library is suitable for sequencing.

QC of finished libraries assisted in predicting sequencing results. In this sample set, all libraries fulfilled the quality criteria, and all libraries were successfully sequenced.

Additionally, the 4200 TapeStation system is a useful tool for troubleshooting the entire NGS library preparation workflow. Based on data obtained, steps to optimize established workflows may be taken. The low hands-on time and flexible 1-to-96-well plate format allows the 4200 TapeStation system to easily be included in any existing NGS workflow, especially for automated processes. Furthermore, a report with the results can be created for customers, and the data can be included in individual LIMS systems.

Table 1. Used data QC criteria for exome sequencing.

QC criterion	Range
Mapping rate	>90 %
Duplicate rate	≤10 %
On target mapped reads	>65 %
Average target coverage	>100x
Contamination level (VerifyBAM FREEMIX)	<2 %

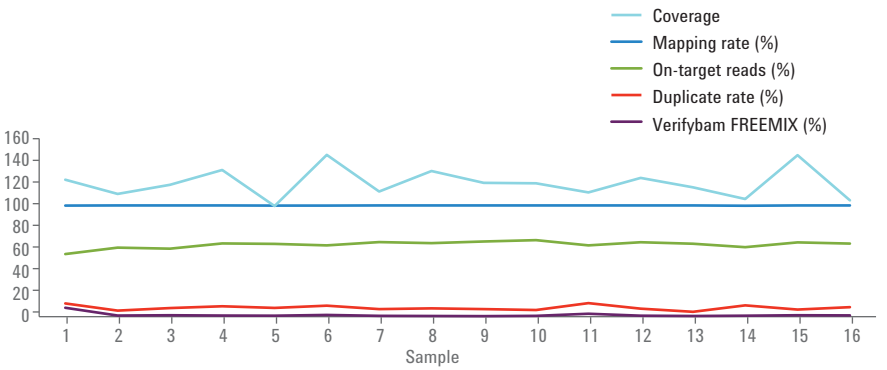

Figure 7. The QC values to determine the quality of sequencing data are mapping rate, on-target mapped reads, average coverage, duplicate rate, and the Verifybam FREEMIX contamination level. The graph shows an example with the sequencing quality criteria for 16 samples sequenced on one lane.

Conclusion

This Application Note demonstrates the successful integration of the Agilent Genomic DNA and Agilent D1000 ScreenTape assays using the Agilent 4200 TapeStation system as a QC tool into the whole exome sequencing workflow. The benefit of using the 4200 TapeStation system is that it enables optimization of the sequencing library preparation process from the beginning to the end with a final QC of finished libraries to ensure good sequencing results. The 4200 TapeStation system provides critical information at multiple key steps, from incoming gDNA material to predicting valuable sequencing results.

References

1. SureSelect^{XT} Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing Automated using Agilent NGS Workstation Option B, Version H0, *Agilent Technologies*, part number G7550-90000, **2015.**
2. Agilent Genomic DNA ScreenTape System Quick Guide for 4200 TapeStation System, *Agilent Technologies*, publication number G2991-90040, **2015.**
3. D1000 ScreenTape Assay Quick Guide for 4200 TapeStation System, *Agilent Technologies*, publication number G2991-90030, **2015.**
4. DNA Integrity Number (DIN) with the Agilent 2200 TapeStation System and the Agilent Genomic DNA ScreenTape Assay, *Agilent Technologies*, publication number 5991-5258EN, **2015.**

www.agilent.com/genomics/tapestation

Published in the USA, December 1, 2016
5991-7615EN

Agilent Technologies