Journal of Life Sciences 11 (2017) 261-268
doi: 10.17265/1934-7391/2017.06.001

Implementation of Automated Sample Quality Control in Whole Exome Sequencing

Elisa Viering[1], Jana Molitor[1], Leslie Friedmann[1], Jana Petersen[2], Marie Beckhaus[2] and Stephan Wolf[2]

1. Agilent Technologies, Inc., Waldbronn 76337, Germany

2. DKFZ Genomics and Proteomics Core Facility, High Throughput Sequencing Unit, Heidelberg 69120, Germany

Abstract: Aims: High-quality DNA as input material for a NGS (next-generation sequencing) workflow is essential for the successful preparation of a DNA library. Additionally, DNA quality has a strong impact on sequencing results. Therefore, it is important to include QC (quality control) steps to assess size, concentration, molarity, and integrity of the DNA during the workflow. Material and Methods: The WES (Whole Exome Sequencing) workflow at the Genomics and Proteomics Core Facility of DKFZ (the German Cancer Research Center) was performed with several QC steps: QC of the input material, QC of intermediate products during library preparation, and QC of final libraries. The Agilent 4200 TapeStation system, which offers automated sample processing, was used to evaluate quantity, size, molarity, and integrity of the samples. Key Findings: The Agilent Genomic DNA ScreenTape assay offers an unbiased genomic DNA integrity assessment, which enables protocol adaption for optimized library preparation, for example, selection of a suitable shearing protocol. Additionally, QC steps during library preparation such as evaluation of library size, concentration, and molarity ensure maximal sequencing output. Significance: The automated and fast high-throughput analysis of genomic DNA with the 4200 TapeStation system helps to save labor time and costs. Additionally, the easy-to-use system can be integrated as a QC tool into the NGS workflow to ensure successful library preparation. QC steps enable the confirmation of suitable library size and concentration for the workflow.

Key words: NGS, FFPE (Formalin-fixed Paraffin-embedded), WES, DIN (DNA Integrity Number), QC, 4200 TapeStation system.

1. Introduction

The High Throughput Sequencing Unit of the DKFZ (the German Cancer Research Center) Genomics and Proteomics Core Facility provides sequencing services to international cancer genome projects for multiple NGS (next-generation sequencing) applications including WES (Whole Exome Sequencing).

Successful WES requires high-quality genomic DNA material as well as high-quality DNA samples during all stages of the library preparation workflow. The High Throughput Sequencing Unit subjects genomic DNA samples obtained by customers to an incoming QC (quality control), prepares the libraries including QC steps, and performs the sequencing. Sequencing is followed by final data QC analysis according to the workflow shown in Fig. 1.

This study describes the library preparation of 88 tumor samples extracted from FFPE (Formalin-fixed Paraffin-embedded) material for WES. Sample quality QC during the workflow includes the analysis of DNA concentration, size, and integrity. Quantification of DNA samples is performed using the Qubit assay based on fluorescence detection with a microplate reader. Integrity of DNA samples is verified with a 4200 TapeStation system. Both instruments enable simultaneous processing of up to 96 samples. The resulting data are automatically transferred into an LIMS (Laboratory Information Management System) for further evaluation [1].

2. Materials and Methods

2.1 Materials

The Qubit dsDNA HS Assay Kit was acquired from Thermo Fisher Scientific Inc. (Waltham, MA, USA).

Corresponding author: Leslie Friedmann, Ph.D., research field: microfluidics marketing & support.

A 96 micro TUBE Plate from Covaris (Woburn, MA, USA) was used. The HiSeq 3000/4000 PE Cluster Kit, cBot and the HiSeq 3000/4000 SBS Kits 50, 150, and 300 cycles were obtained from Illumina, Inc. (San Diego, CA, USA). The Genomic DNA ScreenTape (p/n 5067-5365); Genomic DNA Reagents (p/n 5067-5366); D1000 ScreenTape (p/n 5067-5582); D1000 Reagents (p/n 5067-5583); SureSelect^{XT} Human All Exon v5 kit (p/n 5190-6210) were obtained from Agilent Technologies (Waldbronn, Germany).

2.2 Instruments

The followings were obtained from Agilent Technologies (Waldbronn, Germany): 4200 TapeStation system (p/n G2991AA); Bravo NGS workstation (p/n G5522A). The VolumeCheck was acquired from Micronic (AR Lelystad, The Netherlands) and the Covaris E220 Focused-ultrasonicator was from Covaris (Woburn, MA, USA). The FilterMax F3 Multi-Mode Microplate Reader was obtained from Molecular Devices (Sunnyvale, CA, USA) and the HiSeq 4000 system from Illumina, Inc. (San Diego, CA, USA). The Mastercycler Pro and the Concentrator plus (speed vac) were received from Eppendorf (Hamburg, Germany) and the centrifuge Universal 320/320R from Hettich (Beverly, MA, USA).

2.3 Samples

A batch of 88 genomic DNA samples prepared from FFPE tumor tissue was provided by a customer. Human genomic DNA from Roche Diagnostics GmbH (Mannheim, Germany) was used as positive control during library preparation. For the negative controls, buffer without DNA was used.

2.4 Methods

For all applications, samples were prepared according to the manufacturer's protocol. Briefly, exome libraries were prepared according to the Agilent Low Input SureSelect^{XT} Human All Exon v5 protocol [2] following the workflow shown in Fig. 1. Exome libraries were automatically processed in a 96-well plate format using the Agilent Bravo NGS Workstation. Covaris fragmentation was performed for 8 minutes, and the samples were hybridized at 65 °C for 16.5 hours.

DNA samples were analyzed with the 4200 TapeStation system and the Agilent Genomic DNA ScreenTape and D1000 ScreenTape assays in accordance to the appropriate ScreenTape Quick Guides [3, 4].

For sequencing, samples were normalized to 10 nM and equimolarly pooled using the Bravo NGS Workstation. Molarity of the final library was determined by the 4200 TapeStation system using the

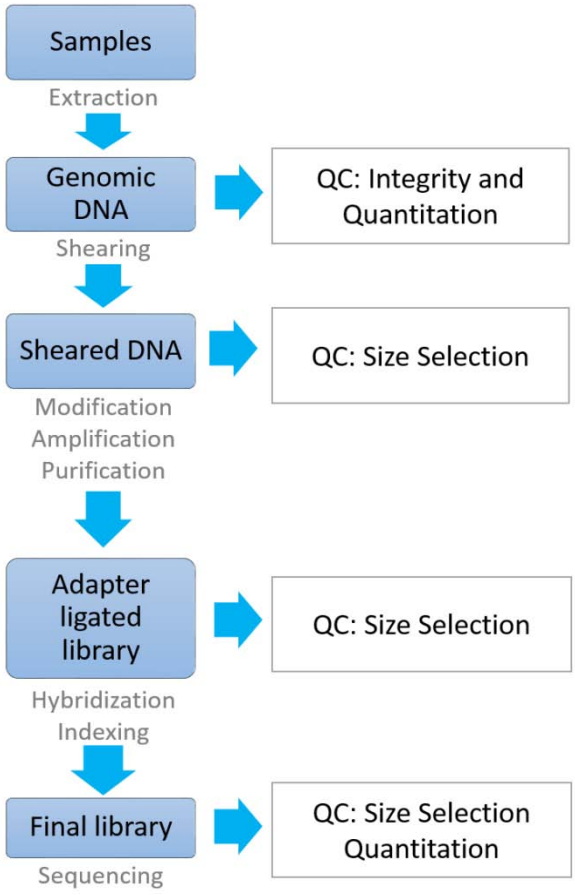

Fig. 1 Workflow of the Agilent SureSelect^{XT} automated target enrichment protocol for Illumina paired-end multiplexed sequencing that was established at the High Throughput Sequencing Unit at the DKFZ. QC steps are performed using the Agilent 4200 TapeStation system.

region analysis function. 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 PE100 (100 bp paired-end sequencing). The 8 bp index tags specified by Illumina, provided with SureSelect^{XT} target enrichment kits, were used for demultiplexing.

3. Results

3.1 QC of Genomic DNA

To determine if the obtained 88 genomic DNA samples prepared from FFPE tumor tissue were suitable for library preparation, QC was performed using the 4200 TapeStation system. Sample quantity and integrity based on the DIN (DNA integrity number) were assessed using the Genomic DNA ScreenTape assay. The DIN algorithm is a feature of the 4200 TapeStation software, and provides a numerical assessment of DNA sample integrity by assigning a score from 1 to 10. A high DIN value indicates highly intact genomic DNA, and a low DIN value degraded genomic DNA [5]. Fig. 2 shows a representative data subset of the genomic DNA analysis.

Based on empirical evidence, a DIN threshold of 7 for genomic DNA samples was established at the High Sequencing Unit. DIN values of all FFPE samples used in this study ranged from DIN 1.3 to 6.2, and are far below the QC threshold of DIN 7 (Fig. 3A).

The second QC criterion for incoming samples is the calculation of total DNA abundance. A minimum of 200 ng genomic DNA is required for DNA library preparation. Samples were quantified in triplicates using the Qubit system. The concentration analysis of the 4200 TapeStation system was used to confirm the Qubit quantification.

To calculate the total amount of DNA present in the sample, the remaining sample volume in the plate wells was determined using the Micronic VolumeCheck system. Based on the measured volume and the concentration obtained by Qubit measurement, the total available DNA amount was calculated (Fig. 3B).

The total DNA amount of most of the 88 tested genomic DNA samples was relatively low with less than 300 ng. Despite the low genomic DNA integrity and the low DNA amount, the customer decided to proceed with the library preparation for 80 of the samples. Eight samples that either had less than 100 ng of total DNA or were the paired sample of one of the excluded samples were not further processed and were replaced by positive and negative controls.

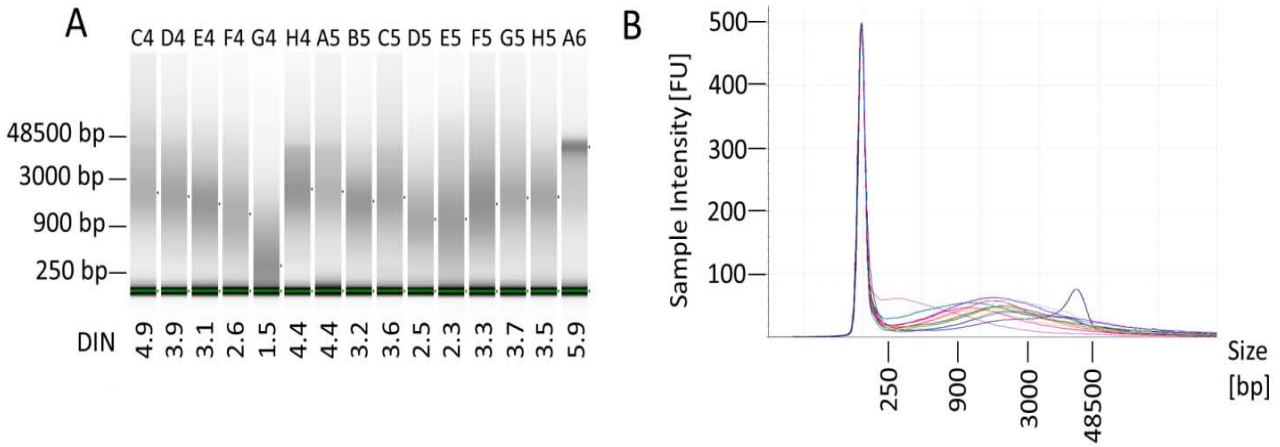

Fig. 2 QC for genomic DNA extracted from FFPE tissue was performed using the Agilent 4200 TapeStation system with the Genomic DNA ScreenTape assay. A: Gel image of 15 samples with the calculated DIN at the bottom and well number on the top of each lane. B: Overlay of the electropherograms of the same samples. Both views enable the visualization and comparison of the integrity of the samples.

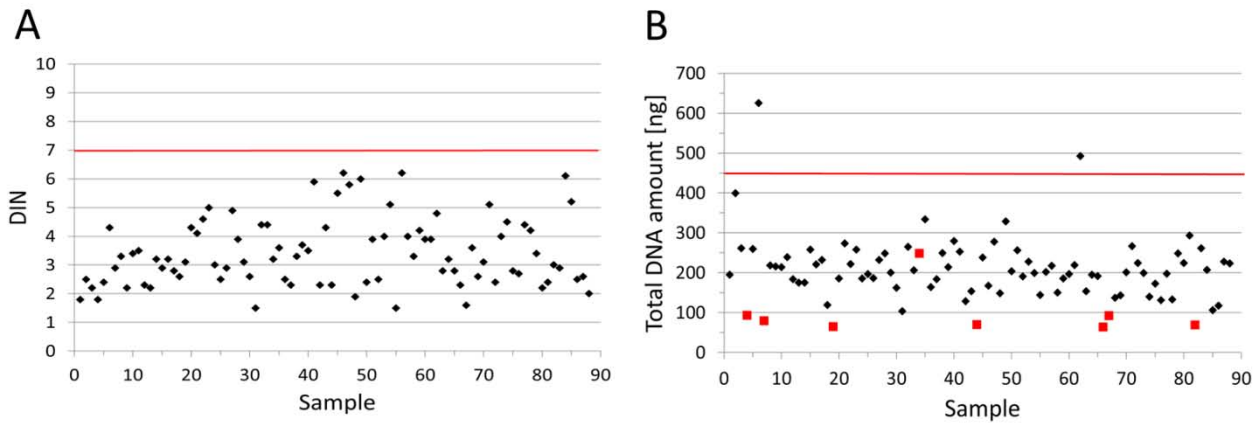

Fig. 3 QC of genomic DNA was performed using the Agilent 4200 TapeStation system with the Genomic DNA ScreenTape assay. A: DNA integrity assessment of 88 genomic DNA samples, the red line indicates the DKFZ threshold for genomic DNA quality (DIN≥ 7). B: Quantification of 88 genomic DNA samples. Red squares indicate genomic DNA samples that were excluded from further analysis. The red line indicates the threshold for DNA quantity (≥ 450 ng).

3.2 Library Preparation for Exome Sequencing

Exome libraries were prepared according to the Agilent Low Input SureSelect^{XT} Human All Exon v5 protocol [2] following the workflow shown in Fig. 1. Intermediate QC steps were used to monitor successful library preparation for sequencing. DNA was evaluated after fragmentation and after adapter-ligation. Before sequencing, the final library was analyzed with respect to concentration, size, and molarity.

In the first step, genomic DNA was fragmented by sonication. Due to the high level of sample degradation (DIN value between 1.3 and 6.2), the Covaris fragmentation was performed for eight minutes instead of six as recommended in the standard protocol [2, 7]. Then, the size of the fragmented DNA was determined with the D1000 ScreenTape assay (Fig. 4).

In accordance with the NGS library preparation protocol [2], all sheared DNA samples resulted in an optimal maximum peak size between 150 and 200 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 (Fig. 5). The total DNA amount was calculated based on the available volume of 30 µL.

According to the Agilent Low Input SureSelect^{XT} Human All Exon v5 protocol [2], the libraries are expected to have a peak size between 225 and 275 bp. The hybridization protocol requires 750 ng of each amplified DNA library. Two DNA samples were slightly below the recommended fragment size of 225 bp and two were slightly below the recommended total DNA amount (Fig. 5). These samples did not fulfill the QC criteria based on size or quantity but were processed through the workflow as automated library preparation did not allow the exclusion of individual samples.

Due to the addition of index sequences, a size shift between adapter-ligated (Fig. 5) and final library (Fig. 6) was expected and could be observed with the 4200 TapeStation system.

The final libraries were expected to have a maximum size between 250 and 350 bp with a minimum concentration of 2 ng/µL and both criteria were fulfilled by all samples. This study showed successful DNA library preparation for all 80 samples, the six positive control samples, and even the samples that

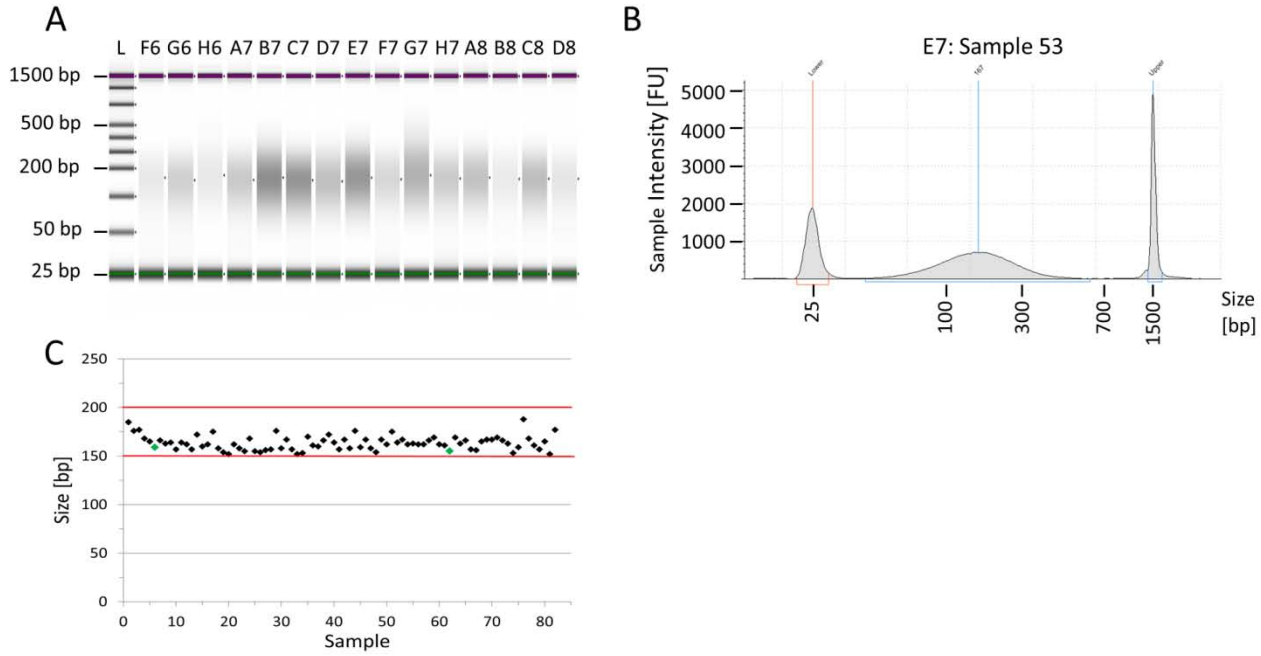

Fig. 4 Sizing of the fragmented DNA was performed using the Agilent 4200 TapeStation system and the D1000 ScreenTape assay. A: Gel view of 15 samples. B: Electropherogram of lane E7 as a representative sample with lower and upper marker as indicated. C: Average fragment size distribution of all 80 samples and eight controls. The two positive controls are shown in green. Red lines indicate the recommended size range (150 to 200 bp).

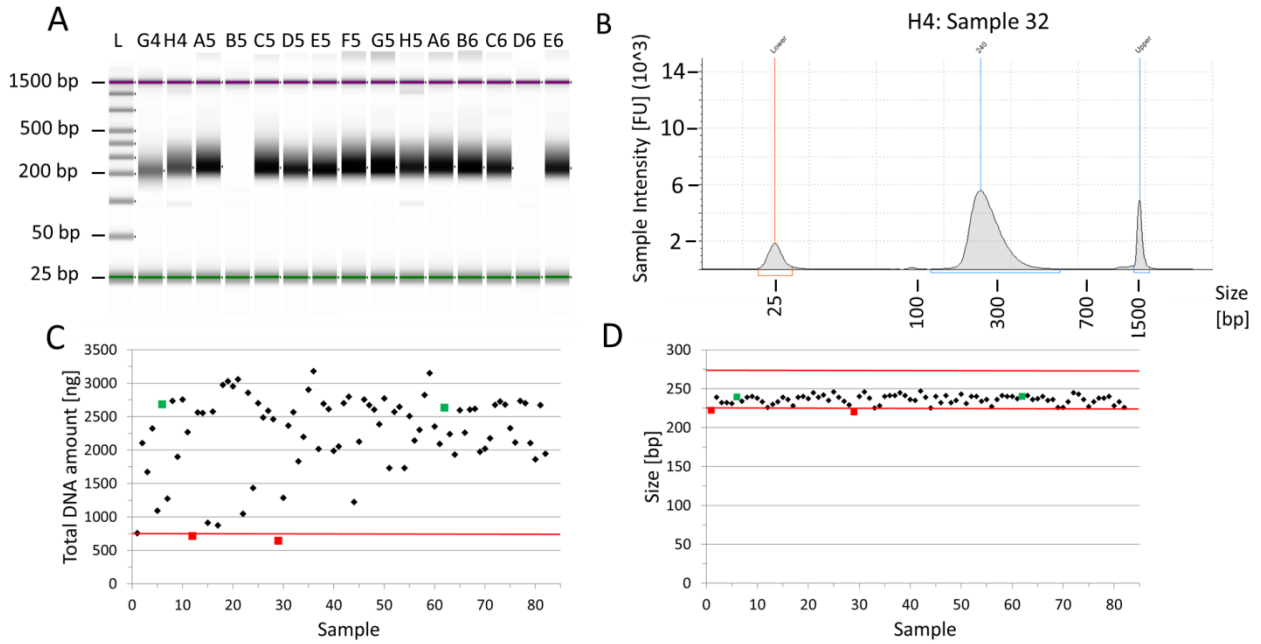

Fig. 5 Sizing of adapter-ligated library was performed using the Agilent 4200 TapeStation system and the D1000 ScreenTape assay. A: Gel view of 15 samples, lanes B5 and D6 show negative controls. B: Electropherogram of lane H4 as a representative sample with lower and upper marker as indicated. C: Total DNA amount for all 80 samples and eight controls. The two positive controls are shown in green, the two samples that were below the threshold (red line, 750 ng) are marked in red. D: Average fragment size distribution of all 80 samples and eight controls. The two positive controls are shown in green. The red lines indicate the recommended size range (225 to 275 bp) and samples marked in red are out of range.

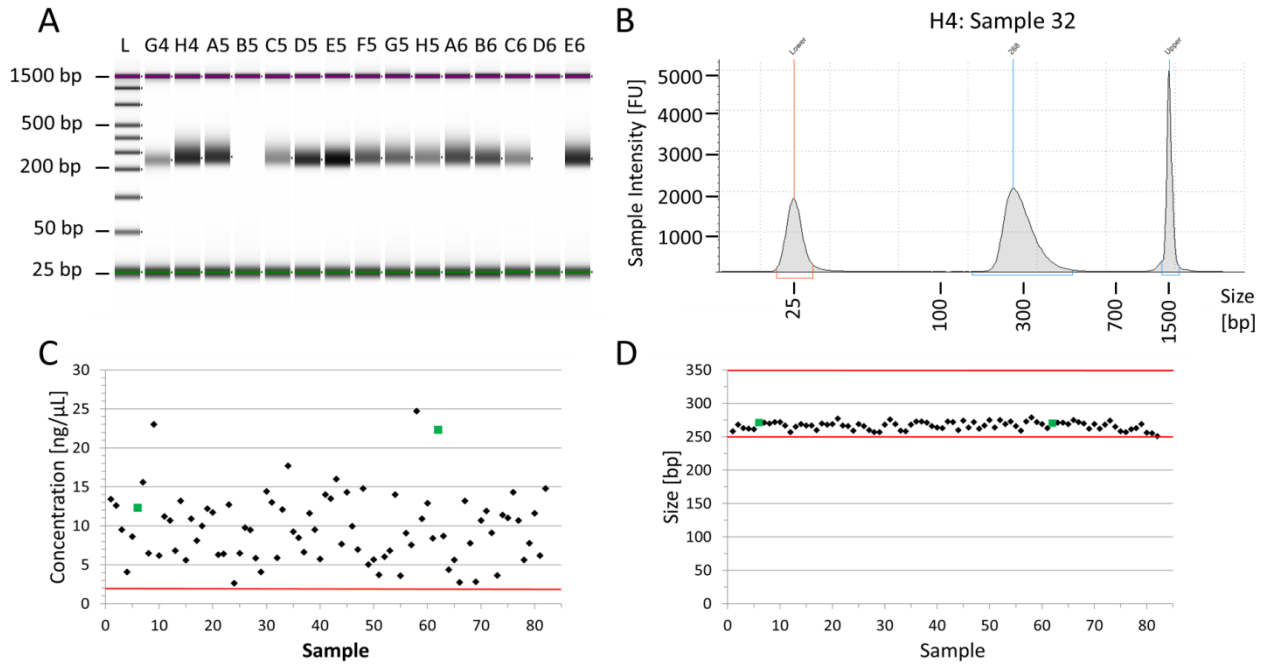

Fig. 6 Sizing analysis of the final library was performed using the Agilent 4200 TapeStation system and the D1000 ScreenTape assay. A: Gel view of 15 samples, lanes B5 and D6 show negative controls. B: Electropherogram of lane H4 as a representative sample with lower and upper marker as indicated. C: Distribution of the concentration for all 80 samples and eight controls. The two positive controls are shown in green. The red line indicates the recommended concentration threshold (2 ng/µL). D: Average fragment size distribution of all 80 samples and eight controls. The two positive controls are shown in green. The red lines indicate the recommended size range (250 to 350 bp).

were slightly outside the QC criteria after adapter ligation. All customer samples were used for sequencing.

3.3 Sequencing

All 80 libraries were equimolarly pooled and each pool was sequenced on two lanes using the Illumina HiSeq 4000 system (PE 100). The quality criteria for WES data are summarized in Table 1. The sequencing results are shown in Fig. 7. The mapping rate (dark blue) for all samples was approximately 100%. Between 55% and 70% of the mapped reads were on target (green). The data of both sequenced lanes were merged to calculate the average target coverage (turquoise) and the coverage for all samples was above 100%. The duplicate rate (red) ranged from 4% to 15%. VerifyBAM FREEMIX (yellow) is a tool to determine the contamination level of genomes. Only four of the 80 samples had a contamination level slightly above 2%. Structural changes in DNA, which often occur in tumor samples, can cause an increase of the VerifyBAM FREEMIX level.

4. Discussion

The success of WES mainly depends on the quality of the DNA samples. DNA sample QC is worthwhile at several steps of the library preparation workflow. In the study described above, the first QC was performed with the starting material, which was genomic DNA extracted from FPPE tissue. To determine sample quality, the 4200 TapeStation system and the Genomic DNA ScreenTape assay were used. The assay includes DNA integrity assessment by the DIN value, which enables the user to set a threshold to ensure successful library preparation.

Since genomic DNA from FFPE tissue is often degraded, all samples showed a DIN below seven but were still sufficiently intact for whole exome library preparation and successful sequencing. In a previous study, it was shown that FFPE DNA samples with a DIN above three can be successfully processed in an

Table 1 Data QC criteria for WES.

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

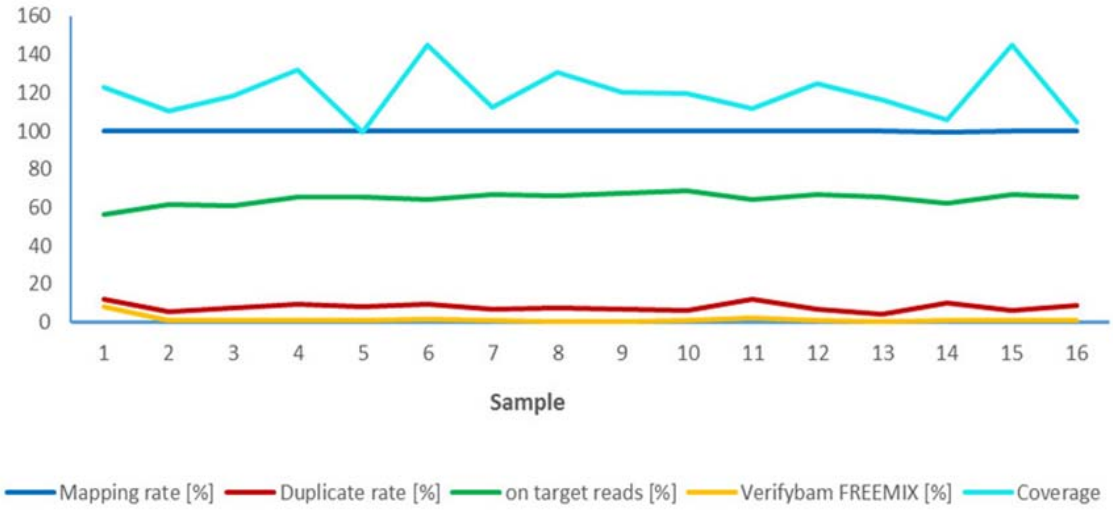

Fig. 7 Sequencing quality criteria are mapping rate, on-target mapped reads, average coverage, duplicate rate, and Verifybam FREEMIX contamination level. The graph shows an example for sequencing quality results of 16 samples. All samples were sequenced on one lane.

NGS workflow. Therefore, the DIN was integrated as selection criteria for downstream processes [6]. In general, the DIN threshold should be adapted depending on the sample origin and downstream applications.

Depending on the quality of the genomic DNA material, fragmentation protocols can be modified for optimized results, for example, using a time dose response experiment for obtaining the best treatment times [7]. Partially degraded genomic DNA has to be treated in a different way compared to intact genomic DNA to get optimal product sizes and finally a successful library preparation. In this case, the fragmentation time was extended from six to eight minutes based on DIN values below seven [7].

Calculation of the total DNA amount is another important step during the NGS process. To obtain optimal results during library preparation, the amount of DNA should be adjusted when starting the fragmentation. Too small amounts can cause a loss of DNA during the purification steps and therefore provide insufficient material for the library preparation. On the other hand, exceeding amounts of input material can lead to incomplete fragmentation.

The library preparation includes three additional QC steps using the 4200 TapeStation system and the D1000 ScreenTape assay. In these steps, the size of the fragmented DNA, the size and total amount of the adapter-ligated library, and the size and concentration of the final library were analyzed.

Depending on the NGS library preparation protocol, the fragmented DNA should have a specific size distribution because sequencing of libraries outside the recommended size range may lead to low read depth or even a lack of read coverage for specific sections of the sequence. Additionally, the size

determination of the DNA is an important step to control the adaptor ligation because a specific size shift between the post sheared and the adaptor ligated samples indicates the success of the ligation.

In general, sample QC of the intermediate steps of the library preparation enables to monitor the success of fragmentation, to monitor the ligation of adapters and indexes, and quantification of the final product—all of which ensure the final library is suitable for sequencing.

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

5. Conclusions

This study demonstrates successful integration of the Genomic DNA and D1000 ScreenTape assays with the 4200 TapeStation system as a QC tool into the WES workflow. The 4200 TapeStation system enables optimization of the sequencing library preparation from start to finish, including QC of final libraries to ensure optimal sequencing results.

In summary, the 4200 TapeStation system is a useful tool for QC and troubleshooting during the entire NGS library preparation workflow. The low hands-on time and flexible 1-to-96-well plate format enables smooth integration of the 4200 TapeStation system into any existing NGS workflow, especially into automated processes. Furthermore, a report with the results can be created for customers and the data can be included in individual LIMS systems [1].

6. Conflict of Interest Statement

The authors declare that there are no conflicts of interest.

References

[1] LIMS Integration of the Agilent 4200 TapeStation System. 2017. *Agilent Technologies*, publication number 5991-7984EN.

[2] SureSelect^{XT} Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing Automated Using Agilent NGS Workstation Option B, Version H0. 2015 *Agilent Technologies*, part number G7550-90000.

[3] Agilent Genomic DNA ScreenTape System Quick Guide for 4200 TapeStation System. 2015. *Agilent Technologies*, publication number G2991-90040.

[4] D1000 ScreenTape Assay Quick Guide for 4200 TapeStation System. 2015. *Agilent Technologies*, publication number G2991-90030.

[5] DNA Integrity Number (DIN) with the Agilent 2200 TapeStation System and the Agilent Genomic DNA ScreenTape Assay. 2015. *Agilent Technologies*, publication number 5991- 5258EN.

[6] The DNA Integrity Number (DIN) Provided by the Genomic DNA ScreenTape Assay Allows for Streamlining of NGS on FFPE Tissue Samples. 2014. *Agilent Technologies*, publication number 5991-5360EN.

[7] Covaris Quick Guide DNA Shearing with S220/E220 Focused-ultrasonicator. 2015. *Covaris*, part number 010308 Rev G.