

Monitoring Library Preparation for Next-Generation Sequencing in Systems Biology Omics Analysis

Authors

Christine Voigt
Alacris Theranostics
Eva Graf
Agilent Technologies

Abstract

As part of the CanPathPro Project, Alacris Theranostics (www.alacris.de) investigates the genome and transcriptome of genetically engineered mouse models. The applied next-generation sequencing protocols are Agilent SureSelect^{XT} and Illumina TruSeq Stranded Total RNA. In the past, the Agilent 2100 Bioanalyzer system was used to perform sample quality control for extracted RNA as well as for DNA and final library quality control (QC) during next-generation sequencing (NGS) workflows. Quantification of extracted DNA was carried out using a fluorometric quantification method. The Agilent 4150 TapeStation system was evaluated as a new platform to monitor sample quality during library preparation. The QC steps of 30 samples were performed in parallel with the 2100 Bioanalyzer system for comparison.

Introduction

CanPathPro (www.canpathpro.eu), an EU Horizon 2020 project, is addressing the challenge of predictive modeling of biological data. During the project, bio-informatic and experimental tools are being developed and refined for the evaluation and control of modeling predictions for systems biology. Within CanPathPro, genetically engineered mouse models are used to follow the temporal molecular and signaling changes occurring during cancer development. This includes the histology of the tumor, proteomics, as well as the genome and transcriptome using next-generation sequencing (NGS). CanPathPro aims to develop a new systems biology platform for generating and testing cancer signaling hypotheses in biomedical research. Alacris Theranostics is working in the field of precision oncology, biomarker applications, systems biology, and drug response modeling and co-ordinates the CanPathPro project.

Experimental

Materials

The HiSeq 2500 system and kits for sequencing were obtained from Illumina (San Diego, CA, USA). The Agilent 4150 TapeStation system (p/n G2992AA) with Agilent Genomic DNA ScreenTape (p/n 5067-5365) and reagents (p/n 5067-5366), Agilent D1000 ScreenTape (p/n 5067-5582) with reagents (p/n 5067-5583), Agilent High Sensitivity D1000 ScreenTape (p/n 5067-5584) with reagents (p/n 5067-5585), and Agilent RNA ScreenTape (p/n 5067-5576) with reagents (p/n 5067-5577) from Agilent Technologies were used for sample quality control. The Agilent 2100 Bioanalyzer system was used with the Agilent RNA 6000 Nano kit (p/n 5067-1511), the Agilent DNA 1000

kit (p/n 5067-1504) and the Agilent High Sensitivity DNA kit (5067-4626). The Agilent SureSelect^{XT} reagent kit (p/n G9611A) and Agilent SureSelect^{XT} Mouse All Exon V2 (p/n G9496A) were used for DNA library preparation. The Illumina TruSeq Stranded Total RNA Library Prep Gold (order no. 20020599) and TruSeq RNA Single Indexes Set A (order no. 20020492) were used for RNA library preparation. The Qubit 2.0 Fluorometer (ThermoFisher Scientific) with the Qubit dsDNA HS assay kit (order no. Q32854, Fisher Scientific) and Qubit RNA HS assay kit (order no. Q32855, Fisher Scientific) was used for nucleic acid quantification. The Covaris Ultrasonicator S220 (KBiosciences) was used for shearing. The Thermal Cycler DNA EngineTetrad 2, BioRad was used for PCR amplification. Nucleic acids were extracted with the Qiagen AllPrep DNA/RNA/miRNA Universal Kit (catalog number 80224, Qiagen). Unless stated, all analyses were performed according to manufacturers' instructions.

Methods

Sample sources

Samples 1 to 15 were extracted from transgenic cancer mouse models for non-small cell lung cancer (NSCLC) and samples 16 to 30 from invasive lobular carcinoma (ILC) of the breast. Each cancer type is represented by two mouse models with different mutations in selected cancer-associated genes.

Quality control of starting material

The 4150 TapeStation system with Genomic DNA ScreenTape assay was used for sample integrity assessment and quantification of genomic DNA (gDNA) starting material. The quantification of gDNA samples was also performed using the Qubit system based on fluorescence detection. Samples were diluted 1:10 before quantification with the Qubit dsDNA HS assay. Based on

these concentrations, the DNA samples were normalized to the quantitative range of the Genomic DNA ScreenTape assay (10–100 ng/μL). The dilution factors for the TapeStation analysis ranged from 1 to 15.

RNA starting material was diluted 1:10 for quantification with the Qubit RNA HS assay. The samples were analyzed undiluted with the RNA ScreenTape assay for quantification and sample integrity assessment. For comparison, the samples were analyzed with the RNA 6000 Nano assay (2100 Bioanalyzer system).

DNA library preparation

The SureSelect^{XT} protocol¹ was used for whole exome sequencing (WES) as presented in Figure 1.

The samples were quantified using both the Genomic DNA ScreenTape assay and the Qubit dsDNA HS assay. Intact gDNA was fragmented using the Covaris instrument according to the SureSelect^{XT} protocol instructions. The treatment time was reduced to 240 seconds to increase the target size. Fragment length and size distribution of the sheared DNA was determined using the D1000 ScreenTape assay (4150 TapeStation system) and the DNA 1000 assay (2100 Bioanalyzer system). Successfully sheared samples were subjected to the next step; combining end repair, adenylation, and adapter ligation followed by amplification and purification (Figure 1). The adapter ligated samples were again qualified with the same assays. Successful adapter ligated samples were then hybridized and captured, followed by indexing and amplification. After purification, the libraries were quantified with the Qubit dsDNA HS assay, then normalized to approximately 2 ng/μL and quantified with the High Sensitivity (HS) D1000 ScreenTape assay. The sizes of the final libraries were automatically determined during the same step with the HS D1000 ScreenTape assay. Additionally, the

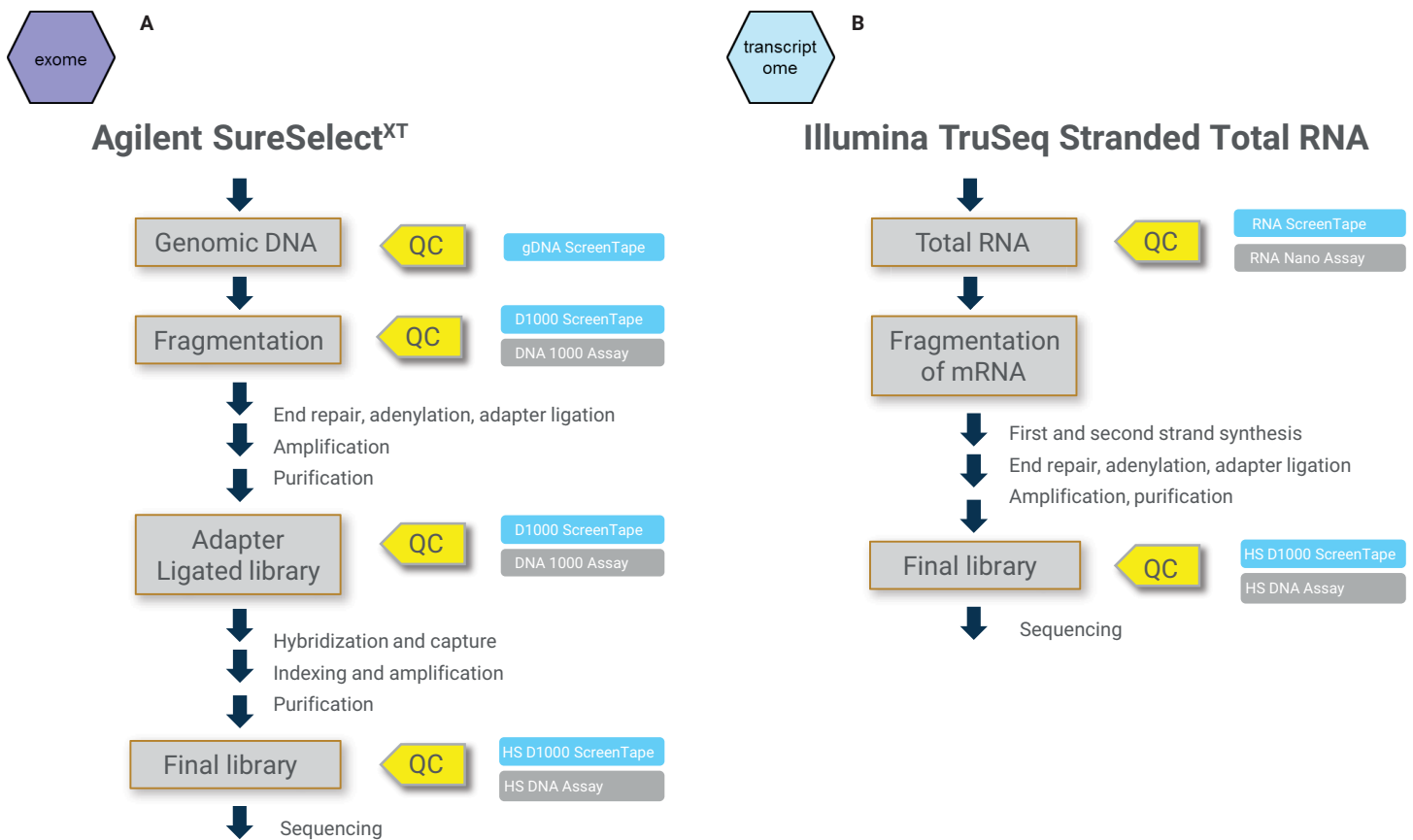


Figure 1. Overview of whole exome and transcriptome library preparation protocols. A) Workflow of the Agilent SureSelect^{XT} protocol B) Library preparation steps with the Illumina TruSeq Stranded Total RNA protocol. QC steps were performed with the Agilent 4150 TapeStation system and the ScreenTape assays as indicated in the blue boxes. Grey boxes refer to the assays used with the Agilent 2100 Bioanalyzer system.

library sizes were assessed with the High Sensitivity (HS) DNA assay using the 2100 Bioanalyzer system. The final libraries were normalized, pooled, and transferred to the Illumina HiSeq 2500 sequencer.

RNA library preparation

The TruSeq Stranded Total RNA protocol² was used to prepare libraries for transcriptome sequencing as presented in Figure 1 with total RNA as starting material. The integrity of the RNA samples was evaluated with the RNA ScreenTape assay (4150 TapeStation system) and compared to the results of the RNA 6000

Nano assay (2100 Bioanalyzer system). Both systems automatically provide quantitative data. The samples were additionally quantified after 1:10 dilution using the Qubit RNA HS assay. Depending on RNA integrity, determined with the RNA integrity number (RIN), the incubation time for total RNA fragmentation was modified (RIN > 3.5: 7 minutes; RIN 2.0–3.5: 2 minutes and RIN < 2.0: no fragmentation). Fragmented RNA samples were processed according to the Illumina TruSeq Stranded Total RNA protocol with the first and second strand synthesis, followed by end repair, adenylation, and adapter ligation. After

amplification, the yield of the purified final libraries was determined with the Qubit DNA HS assay. The samples were normalized to approximately 2 ng/μL (dilution factor 10 to 15) and analyzed on the HS D1000 ScreenTape assay to assess library size and concentration. The size was also determined with the HS DNA assay using the 2100 Bioanalyzer system.

All final libraries were pooled and then sequenced with the Illumina HiSeq 2500 sequencer.

Results and discussion

DNA library preparation

30 samples of genomic DNA extracted from genetically engineered mouse models were manually prepared according to the SureSelect^{XT} protocol.

Genomic DNA integrity

Quantity and integrity of gDNA were evaluated to ascertain that the starting material was suitable for library preparation. This first important QC step was performed with the 4150 TapeStation system and the Genomic DNA ScreenTape assay. Genomic DNA quality assessment based on the DNA integrity number (DIN) was previously demonstrated.³ The DIN algorithm is included in the TapeStation software providing an objective quality metric for standardized DNA integrity assessment. Signal distribution across the sizing range of the assay is converted into a numerical value ranging from 10 (intact DNA) to 1 (degraded DNA). Freshly extracted gDNA from mouse tissue is expected to depict high DIN values. The DIN values of the 30 gDNA samples ranged from 6.6 to 9.3 (data not shown). Since high quality could be verified for the starting material, all 30 samples were used for library preparation.

Genomic DNA quantification

The SureSelect^{XT} protocol is optimized for the preparation of libraries from 3 µg gDNA starting material. The protocol recommends the Qubit system for quantification, which was used to evaluate gDNA concentration. In parallel, the samples were quantified with the 4150 TapeStation system and the Genomic DNA ScreenTape assay. DNA samples were normalized before the analysis with the Genomic DNA ScreenTape assay. The DNA quantity determined by both methods is compared in Figure 2. The concentration data evaluated with the Genomic DNA ScreenTape assay was highly comparable with the Qubit data exhibiting an R² correlation value of 98.5%. This result shows that the Genomic DNA ScreenTape assay enables

reliable and accurate quantification for genomic DNA as starting material for NGS applications.

Characterization of sheared DNA

The first step of the SureSelect^{XT} protocol is fragmentation of gDNA by shearing with the Covaris ultrasonicator. The shearing protocol was adjusted to produce fragments with a larger target size to allow for sequencing in 2 × 100 bp mode. Optimal shearing in NGS workflows is typically verified with the evaluation of the size distribution and electropherogram pattern of fragmented DNA samples with the DNA 1000 assay and the 2100 Bioanalyzer system. In addition, the samples were analyzed with the 4150 TapeStation system and the D1000 ScreenTape assay.

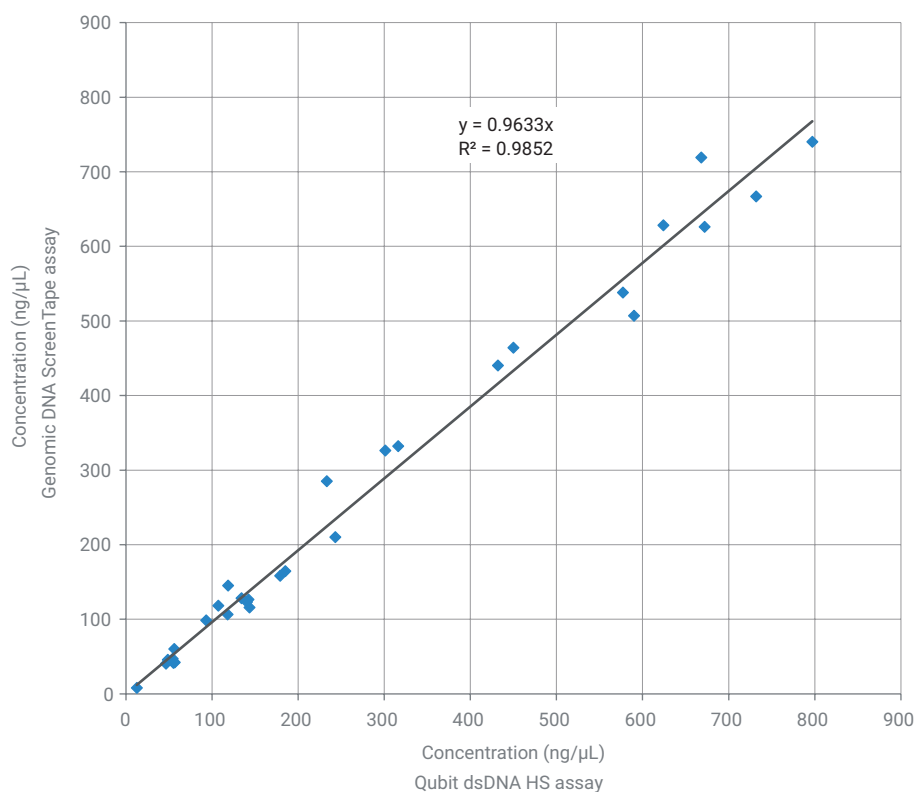


Figure 2. Comparison of gDNA concentrations measured with the Agilent Genomic DNA ScreenTape assay and with the Qubit dsDNA HS assay. The samples were diluted to match the assay ranges. The displayed concentrations represent the original sample concentrations.

Example electropherograms of sheared DNA analyzed with the 4150 TapeStation and the 2100 Bioanalyzer systems are shown in Figure 3. All samples displayed an even size distribution with no undesirable shouldering.

Due to the modified shearing protocol, the expected average size range of fragmented DNA is about 100 bp larger than recommended in the protocol (between 200 and 325 bp). The fragmented DNA samples showed a maximum peak size between 260 and 310 bp on both systems, which verified optimal shearing. The size of this QC step can be used as a core value for each sample to compare with the size after adapter ligation later in the SureSelect^{XT} workflow. The sizing results of the 2100 Bioanalyzer and the 4150 TapeStation systems correlated highly with an average deviation of 2.2% (Figure 4).

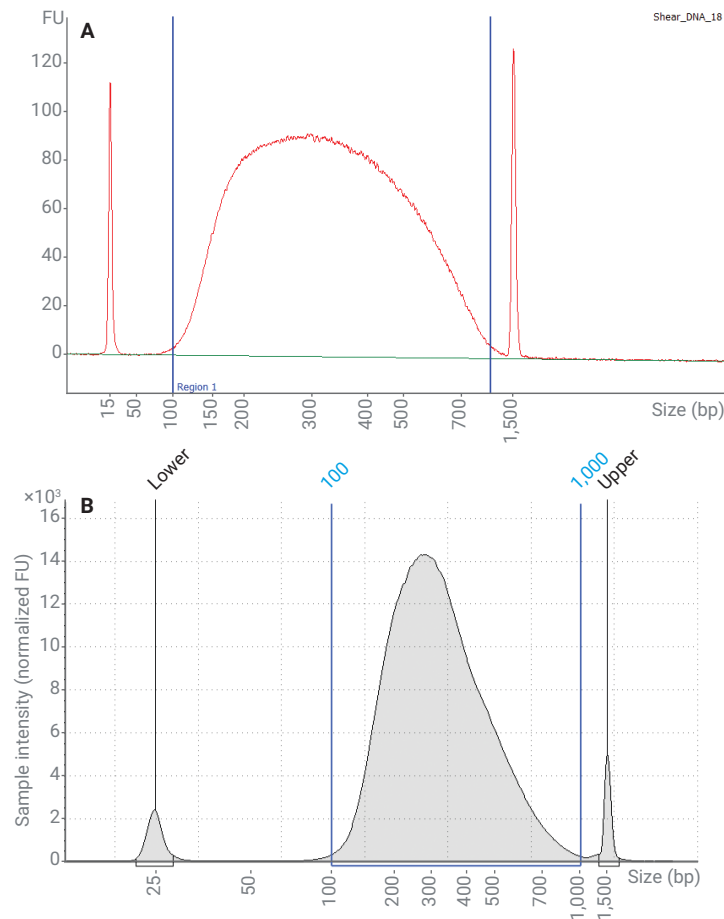


Figure 3. A fragmented DNA sample from the SureSelect^{XT} workflow showing even size distribution. A) Sheared DNA analyzed with the Agilent DNA 1000 assay, B) Electropherogram of the same sample analyzed with the Agilent D1000 ScreenTape assay.

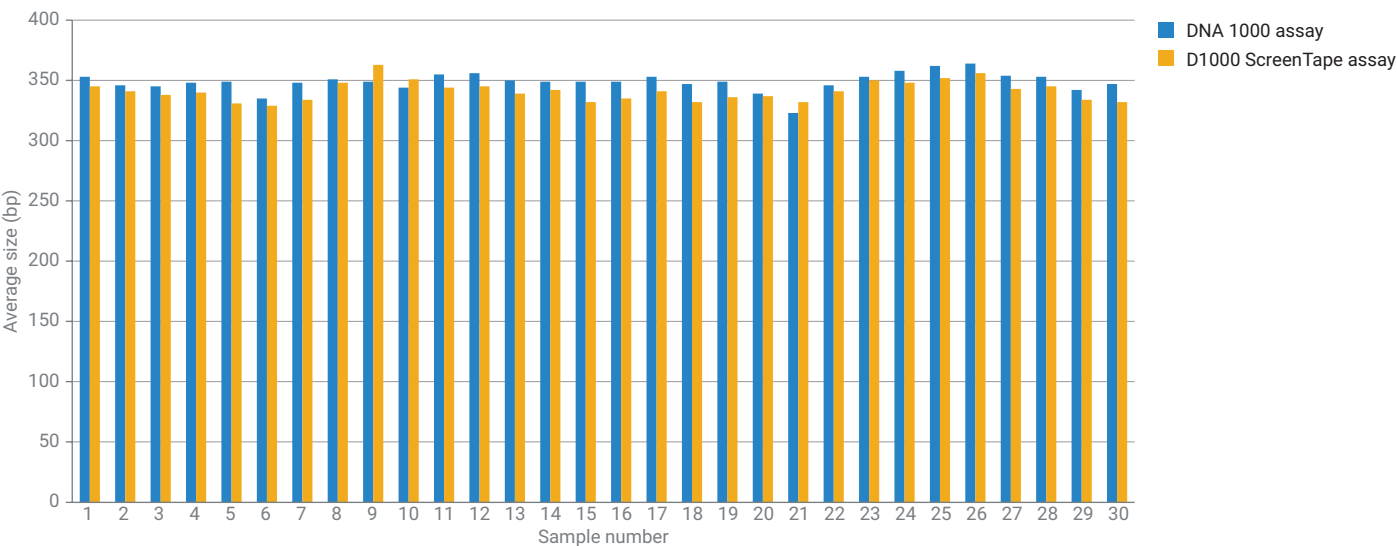


Figure 4. Average size of 30 sheared DNA samples with a region set from 100 to 1000 bp. The average size was determined with the Agilent DNA 1000 assay and the Agilent D1000 ScreenTape assay.

Qualification of precapture libraries

The next steps for sheared DNA in the SureSelect^{XT} protocol are comprised of end repair, adenylation, and adapter ligation followed by amplification. To verify efficient adapter ligation, purified adapter ligated DNA samples (precapture samples) were qualified with the DNA 1000 assay (2100 Bioanalyzer system) and the D1000 ScreenTape assay (4150 TapeStation system). The precapture libraries were analyzed by region analysis with boundaries set from 200 to 700 bp. The average region size ranged from 340 to 420 bp with a

maximum peak size at approximately 300 bp. All samples showed an expected size shift (approximately 30 bp) with regards to the sheared DNA samples, which is related to adapter ligation. The average size determined by both systems was highly comparable for all 30 samples with an average deviation of 2.7% (Figure 5). Hence, the average region size of libraries can be confidentially assessed using the region functionality with the DNA 1000 assay as well as with the D1000 ScreenTape assay.

Final library QC

Adapter-ligated samples were processed according to the SureSelect^{XT} protocol, including hybridization and capturing, followed by indexing and amplification. In general, final libraries are equimolar pooled for sequencing, which requires a QC step to ensure accurate and precise evaluation of library size and quantity. The libraries were analyzed with the HS DNA assay (2100 Bioanalyzer system) and with the HS D1000 ScreenTape assay (4150 TapeStation system). Electropherogram patterns with a broad single peak and the absence of artifacts represent high-quality libraries.⁴ All samples depicted high-quality libraries, example electropherograms are shown in Figure 6.

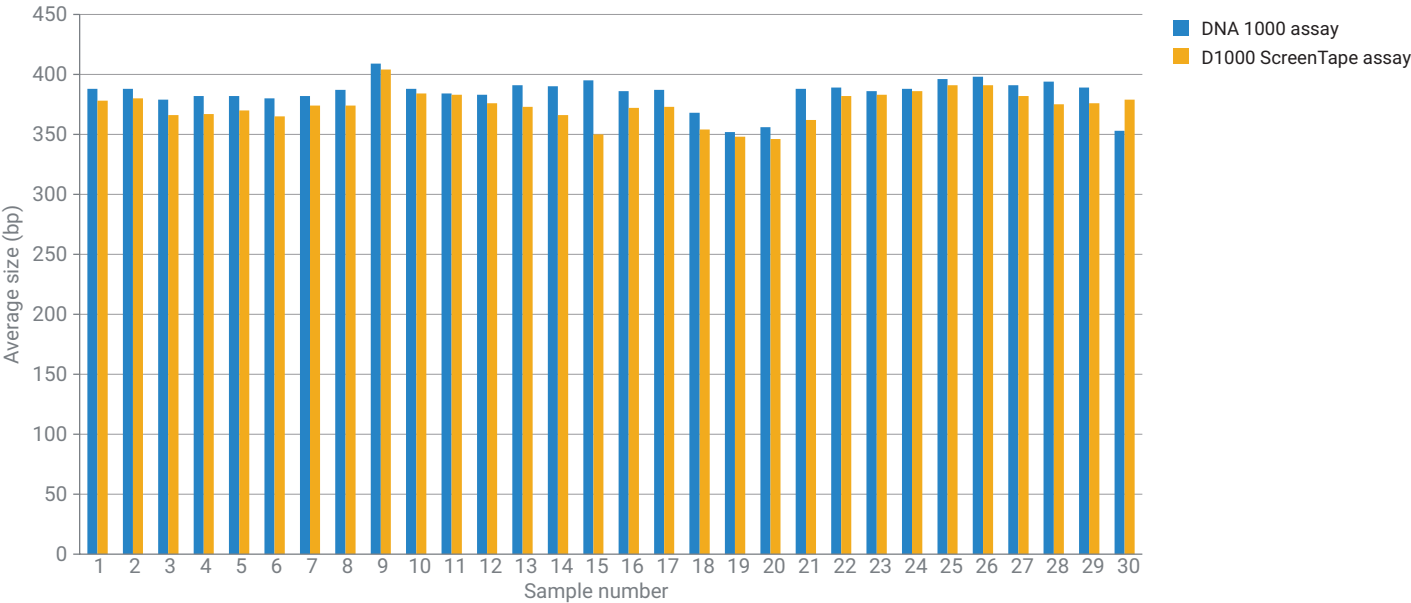


Figure 5. Comparison of the average size of 30 precapture libraries with a region set from 200 to 700 bp. The average size was determined with the Agilent DNA 1000 assay and the Agilent D1000 ScreenTape assay.

The maximum peak size ranged from 350 to 400 bp. Due to the increased fragment size that was aimed during shearing, the expected average size ranges from 325 to 400 bp. To determine the average size for molarity calculation, a region was set ranging from 200 to 600 bp. The sizing results of the HS D1000 ScreenTape assay were compared to the results determined by the HS DNA assay with the 2100 Bioanalyzer system. The results were highly comparable for all samples with an average deviation of 0.9% (Figure 7).

Both systems provide reliable results for final library size determination.

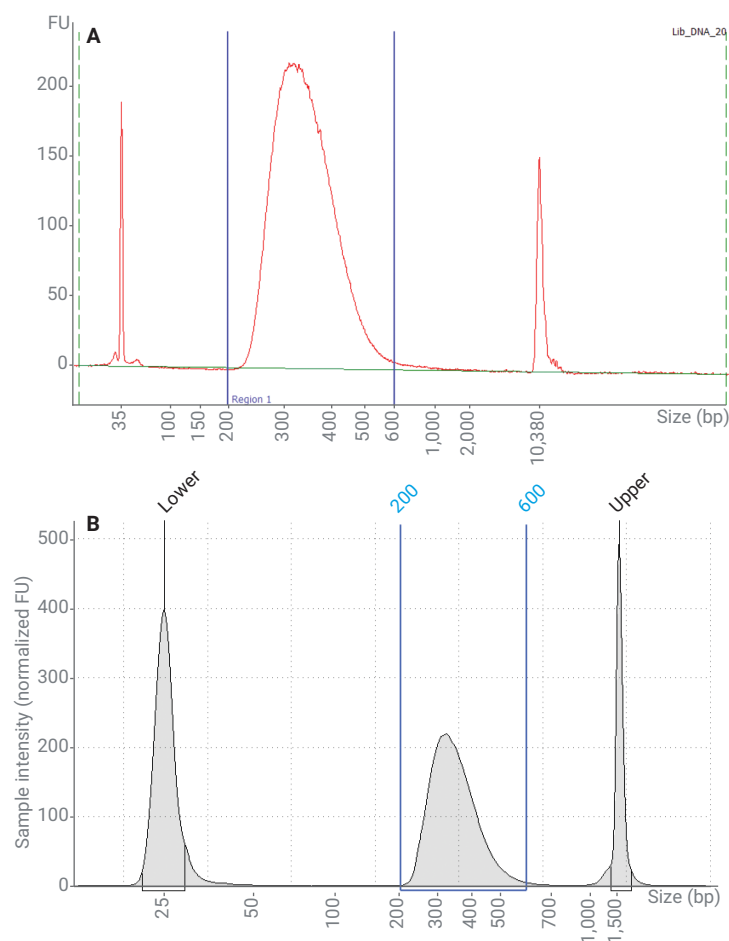


Figure 6. Example electropherograms of end products of the SureSelect^{XT} workflow. A) Final library analyzed with the Agilent HS DNA assay, B) Sample analyzed with the Agilent HS D1000 ScreenTape assay.

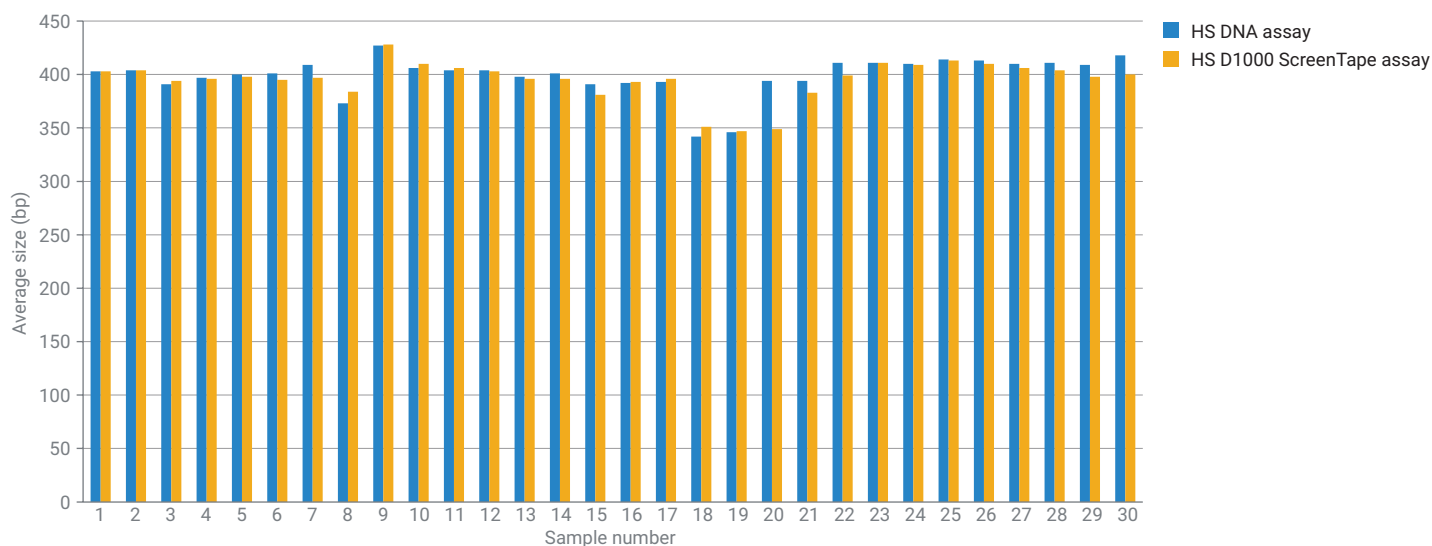


Figure 7. Sizing comparison of the final libraries generated with the Agilent SureSelect^{XT} workflow. The size was determined by region analysis ranging from 200 to 600 bp. Sizing data was determined with the Agilent HS DNA assay and compared to the Agilent HS D1000 ScreenTape assay.

Alacris Theranostics uses the average size provided by the TapeStation analysis software and the concentration determined with the Qubit assay to determine molarity data. The concentration data provided by the HS D1000 ScreenTape was compared to the quantification results attained from the Qubit dsDNA HS assay (Figure 8). The quantification with HS D1000 ScreenTape assay was highly comparable to Qubit results with an average deviation of 11.1%.

The electropherogram pattern of DNA smears and libraries can differ between the 4150 TapeStation system and the 2100 Bioanalyzer system due to differences in technology. Nevertheless, resulting sizing data was highly comparable between both systems. Furthermore, the electropherogram view of either system enabled visual evaluation of an even size distribution and the absence of artifacts. The quantification of final libraries with the 4150 TapeStation system provided

equivalent values compared to the Qubit system. Thus, the 4150 TapeStation system together with the ScreenTape assay portfolio can be confidentially used as a QC tool during library preparation with the SureSelect^{XT} protocol.

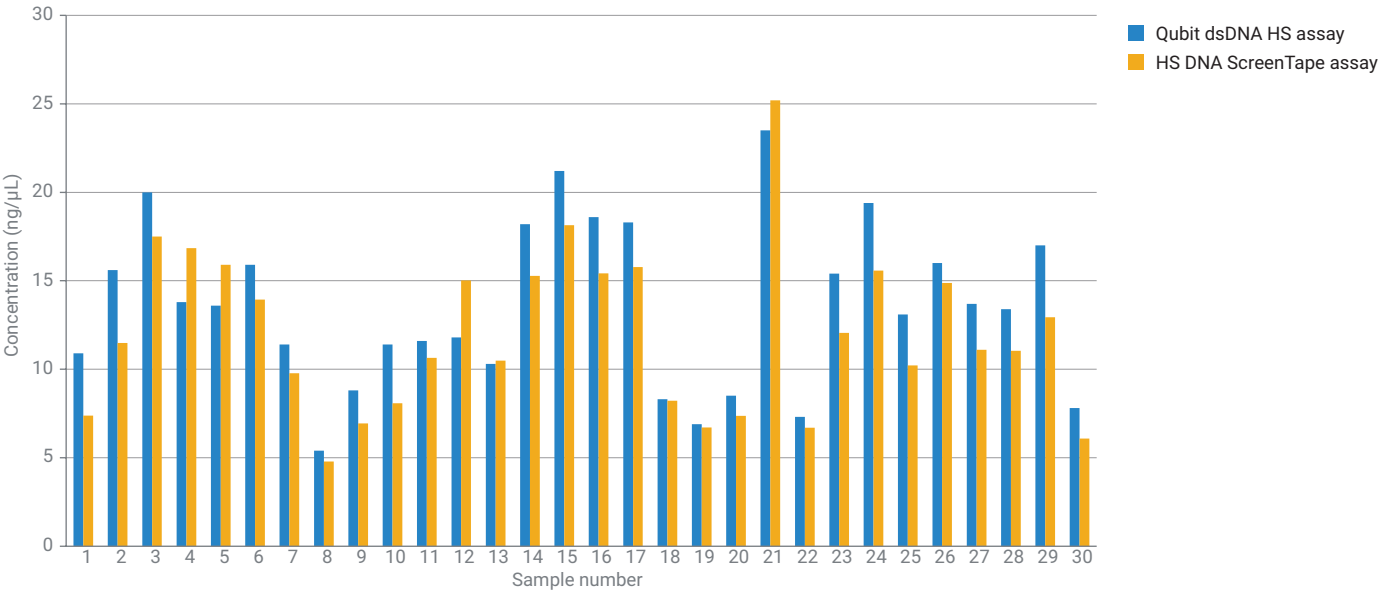


Figure 8. Quantification of SureSelect^{XT} final libraries with the Qubit dsDNA HS assay compared to the Agilent HS D1000 ScreenTape assay. The samples were diluted to match the Agilent HS D1000 ScreenTape assay range. The graph shows the original library concentration.

RNA library preparation

Total RNA samples of the same mouse models used for WES were subjected to library preparation for transcriptome analysis with the Illumina TruSeq Stranded Total RNA protocol (Figure 1). Final libraries were qualified, pooled, and then sequenced with an Illumina HiSeq 2500 system.

Total RNA integrity

RNA samples are often subject to degradation by enzymatical, chemical, or other environmental impacts. Therefore, RNA integrity analysis is a crucial step before any downstream application. In general, Alacris Theranostics successfully processes RNA starting material with a RIN above 6 for sequencing. RNA integrity of the starting material was determined with the RNA ScreenTape assay (4150 TapeStation system) and compared to the RNA 6000 Nano assay (2100 Bioanalyzer system) (Figure 9).

The RIN^e algorithm of the TapeStation software is equivalent to the RIN metric provided by the RNA assays of the 2100 Bioanalyzer system.^{5,6} The comparative analysis of the RNA samples verified that RIN and RIN^e are highly correlating with a slope of 0.997 and a linear regression (R^2) of 97.9%. Despite the partial degradation of some samples ($RIN < 6$), libraries were created from all RNA samples. The RNA could not be re-extracted since only very limited tumor tissue was provided. The Illumina TruSeq Stranded Total RNA protocol permits the use of partially degraded RNA as starting material by modifying the protocol according to the RNA integrity.

Total RNA quantification

Total RNA quantification is required to normalize the starting material before library preparation. The RNA samples were quantified with the RNA ScreenTape assay and compared to Qubit results (Figure 10). Both

systems resulted in similar total RNA concentrations of the samples with a slope of 1.017 and a linear regression of 97.9% showing high correlation. The two degraded RNA samples 6 and 7 were treated as outliers.

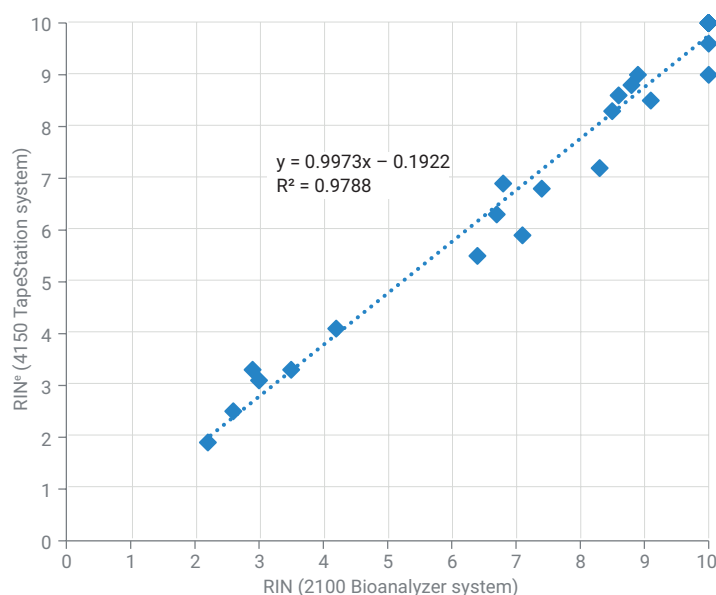


Figure 9. RIN^e and RIN correlation of 30 mouse total RNA samples analyzed with the Agilent RNA ScreenTape assay and the Agilent RNA 6000 Nano assay.

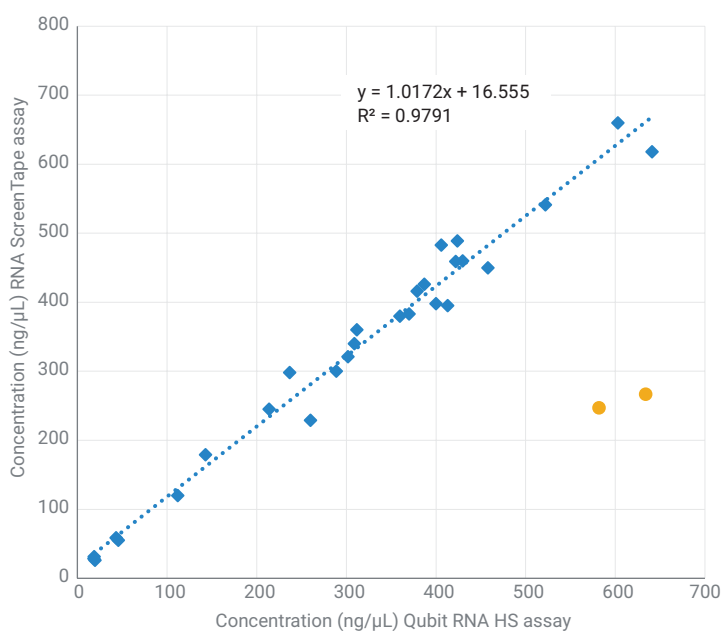


Figure 10. Quantification of 30 mouse total RNA samples with the Agilent RNA ScreenTape assay and the RNA HS assay (Qubit system). The two outliers in orange refer to samples 6 and 7 which exhibited a high level of degradation.

The determined concentration with the RNA ScreenTape assay of both samples was approximately 60% lower compared to the results of the Qubit RNA HS assay. During gel electrophoresis, RNA is separated by size and very small fragments migrate outside the detection window, thereby decreasing the measured concentration. Both samples were processed despite their lack of quality, however, sequencing results revealed insufficient performance.

Overall, the RNA ScreenTape assay can be considered as a very useful tool to determine both integrity and quantity of RNA starting material for sequencing in a single analysis step.

Characterization of final libraries

RNA samples were fragmented according to the Illumina TruSeq Stranded Total RNA protocol by applying modified fragmentation protocols for partially degraded RNA (RIN < 6). Fragmented RNA samples were processed through the first and second strand synthesis, followed by end repair, adenylation and adapter ligation, then amplification. The purified final libraries were characterized using the HS D1000 ScreenTape assay (4150 TapeStation system) in addition to the HS DNA assay (2100 Bioanalyzer system). Figure 11 shows example electropherograms for both systems, while Figure 12 compares the sizing results. The electropherograms of all samples reflected high-quality patterns showing a single broad library without any additional peaks. However, the two degraded RNA samples 6 and 7 depicted an anomalous library pattern, as shown in Figure 11C.

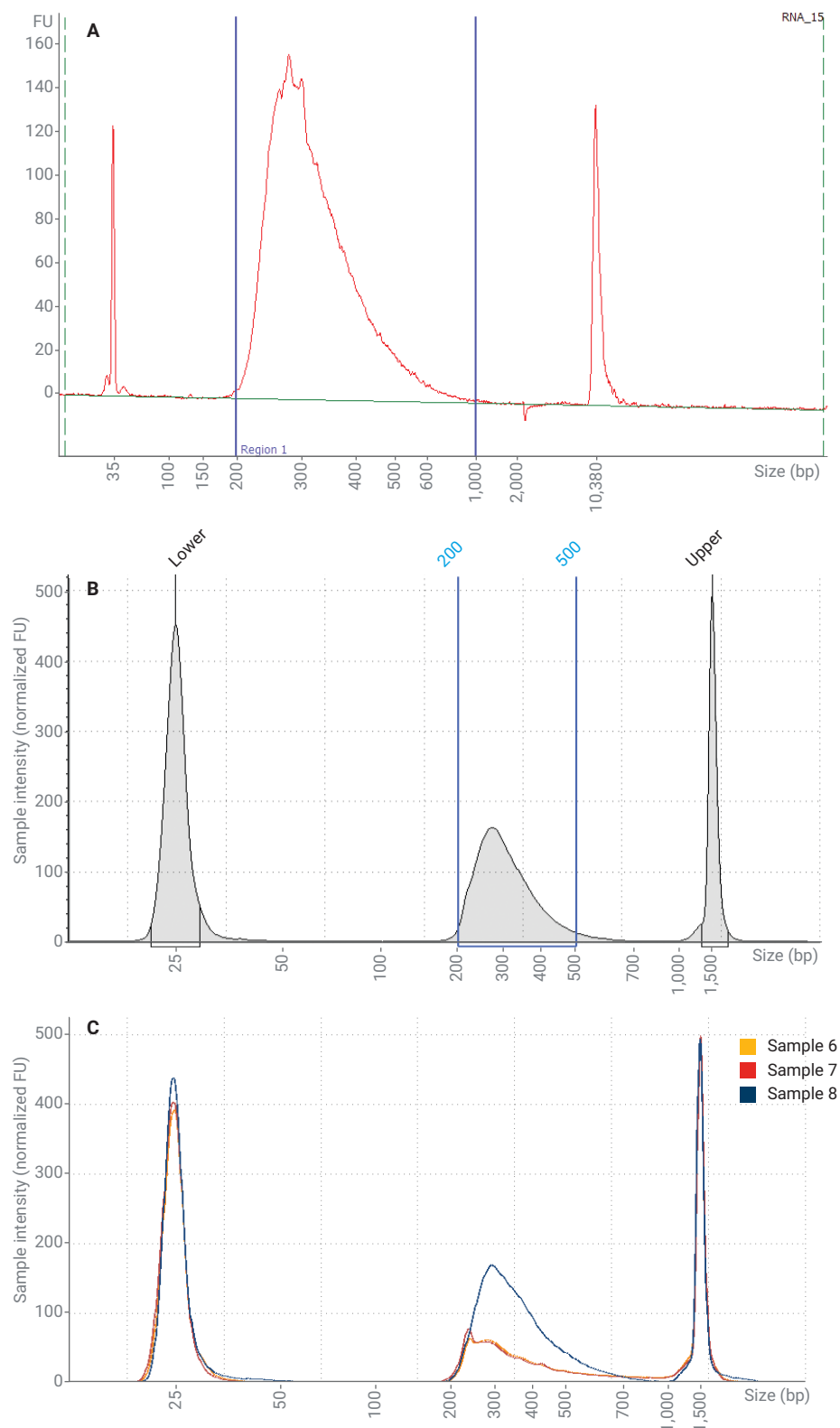


Figure 11. Comparison of electropherogram patterns of final libraries prepared from RNA. A) Final library analyzed with the Agilent DNA 1000 assay. B) Electropherogram of the same sample obtained with the Agilent D1000 ScreenTape assay. C) Overlay of a high-quality library (sample 8) and two libraries from degraded starting material (samples 6 and 7) analyzed with the D1000 ScreenTape assay.

The average size of the final libraries was determined by setting an analysis region from 200 to 500 bp and ranged from 300 to 390 bp. This matched the expected average size range between 250 and 400 bp. The sizing results of the HS D1000 ScreenTape assay were similar to the results determined by the HS DNA assay (2100 Bioanalyzer system), with an average deviation of 1.2%.

Library yield was automatically determined with the HS D1000 ScreenTape assay and was compared to results provided by the Qubit dsDNA HS assay (Figure 13). The libraries were diluted to the HS D1000 ScreenTape assay quantitative range. The data reflects the original library concentrations before dilution. The final library concentration determined by

both systems showed a high correlation, with an average deviation of 12.9%. This result reveals that the HS D1000 ScreenTape assay can be used for library yield determination equivalent to the Qubit system. During the same analysis step, library size can be determined with equal results compared to the HS DNA assay of the 2100 Bioanalyzer system.

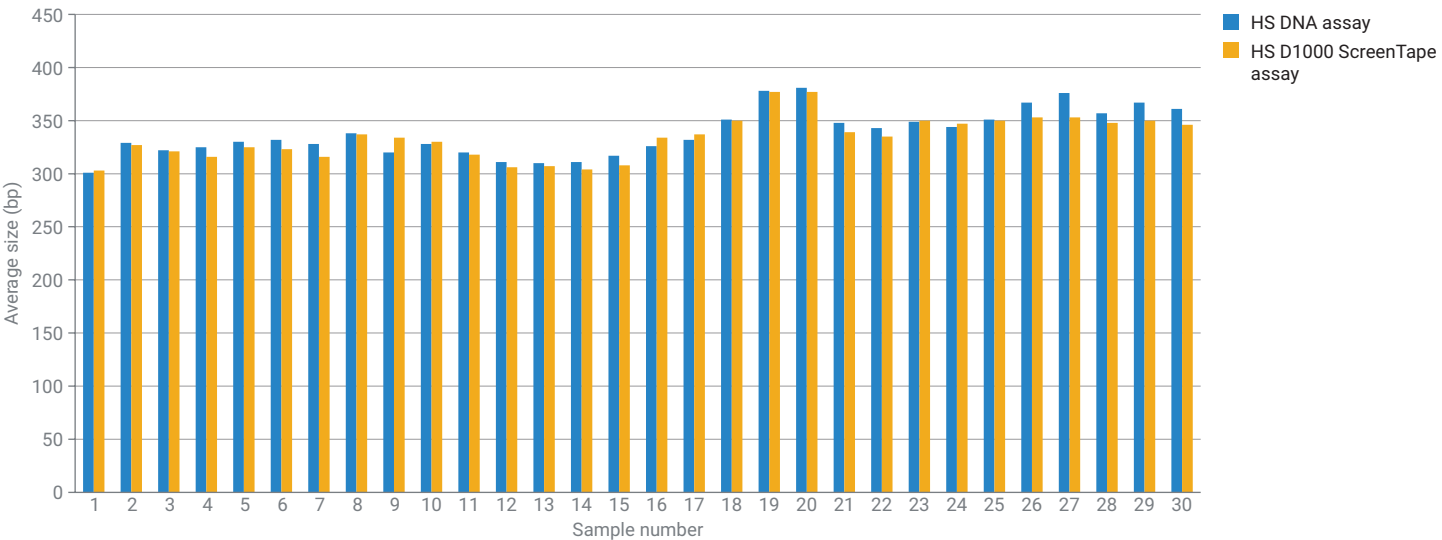


Figure 12. Sizing results for 30 final libraries of mouse RNA samples analyzed with the Agilent HS D1000 ScreenTape assay and the Agilent HS DNA assay.

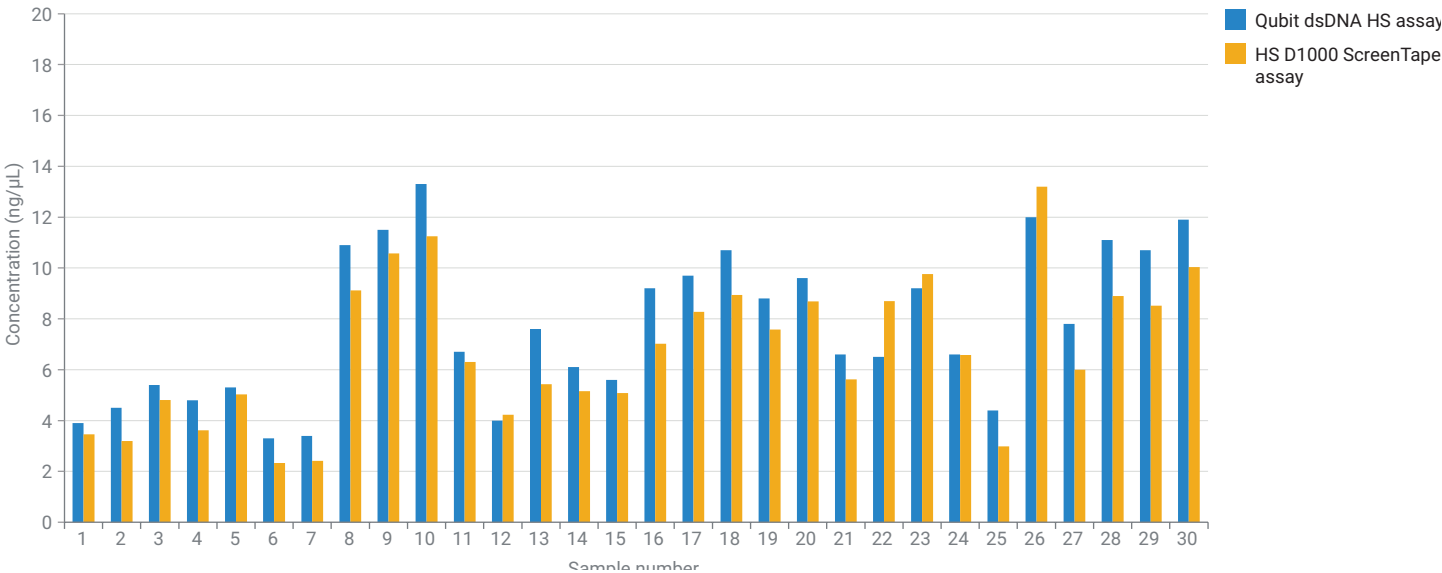


Figure 13. Quantification of 30 final libraries evaluated with the HS D1000 ScreenTape assay compared to data determined with the Qubit system. The data reflects the original library concentrations before dilution.

The final libraries depicted high yield and quality and were normalized, pooled, and sequenced. Two RNA samples with a high level of degradation showed low RIN and RIN^e numbers and deviating quantities when comparing Qubit and 4150 TapeStation system data. Libraries generated from these samples resulted in lower library yield and a different electrophoresis pattern and finally showed low library complexity.

Conclusion

Alacris Theranostics successfully evaluated the 4150 TapeStation system with the ScreenTape assay portfolio to monitor exome and transcriptome library preparation for systems biology omics analysis. 30 DNA and RNA samples were extracted from genetically engineered mouse models and sequenced as part of the CanPathPro project. The 4150 TapeStation system was compared with the 2100 Bioanalyzer and Qubit systems by performing quality control steps on the platforms in parallel. It was demonstrated that the 4150 TapeStation system is a reliable tool for sizing and quantification at any step of library preparation for exome and transcriptome sequencing. It can be used to qualify sheared DNA, precapture libraries, and final libraries with sizing results

equivalent to the 2100 Bioanalyzer system. Quantitative data of all employed ScreenTape assays highly correlated with Qubit results. In addition, the 4150 TapeStation system consistently delivered excellent RNA characterization for quality assessment. It provides the RNA integrity number equivalent (RIN^e), which is highly comparable to the well accepted RIN algorithm of the 2100 Bioanalyzer system. Poor sequencing of two degraded RNA samples could be anticipated by the interpretation of QC data assessed with the 4150 TapeStation system. The ready-to-use ScreenTape devices simplify the quality control steps that are part of the library preparation workflow, and significantly reduce hands-on-time compared to the 2100 Bioanalyzer system. Moreover, the 4150 TapeStation system enables reliable quality and quantity assessment of genomic DNA with the quality metric DIN. The 4150 TapeStation system can load any sample number from 1 to 16 samples per run without wasting consumables, thus providing a constant cost-per-sample. Together with the easy switch between different assays, this gives the user high flexibility to easily accommodate to project-driven throughput.

References

1. Agilent SureSelect^{XT} Target Enrichment System for Illumina Paired-End Multiplexed Sequencing, *Agilent Technologies User Manual*, publication number G7530-90000, **2017**.
2. TruSeq Stranded Total RNA Reference Guide, Illumina document no. 1000000040499 v00, **2017**.
3. Quality Control in Illumina Sequencing Workflows Using the TapeStation System, *Agilent Technologies Application Note*, publication number 5994-0327EN, **2018**.
4. Sample Quality Control in Agilent NGS Solutions, *Agilent Technologies Application Note*, publication number 5994-0127EN, **2018**.
5. Performance of the Agilent RNA ScreenTape and the High Sensitivity RNA ScreenTape Assay for the Agilent 2200 TapeStation System, *Agilent Technologies Technical Overview*, publication number 5991-3426EN, **2016**.
6. Performance Characteristics of the RNA and the High Sensitivity RNA ScreenTape Assays for the 4150 TapeStation System, *Agilent Technologies Technical Overview*, publication number 5991-1038EN, **2019**.

www.agilent.com/genomics/tapestation

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

This information is subject to change without notice.