Technical Overview

Agilent
Trusted Answers

Detection of Adapter Dimers in NGS Libraries with the Agilent Fragment Analyzer and TapeStation Systems

Introduction

An NGS library is the compilation of many DNA fragments pooled together for subsequent analysis. During the process of library preparation, DNA fragments are ligated to a known sequence, or adapter. These adapters allow the DNA fragments to bind to the flow cell during the sequencing run. The adapter is designed to include a primer binding site to initiate sequencing. Additionally, adapters can include unique sequences that can be used as indexes to help distinguish between multiplexed libraries during data analysis. However, sometimes two adapters may ligate to each other without an insert, forming a dimer that can be amplified, bind, and cluster on the flow cell, generating erroneous sequencing data (Figure 1). Even small amounts of adapter dimers in a library can contribute to detrimental sequencing results. The presence of adapter dimers can affect the accuracy of library quantification, leading to suboptimal flow cell loading and reducing clustering efficiency. Smaller fragments such as dimers are also preferentially sequenced over larger fragments, so even low levels of adapter dimers could contribute to issues such as reduced output, lower diversity, and decreased genome coverage[1,2]. Just 5% of adapter dimer contamination in a library can cause up to 50% of the reads to be adapter sequences[3]. Thus, it is highly recommended to minimize and remove adapter dimers from the library before sequencing.

While the optimal amount of adapter dimer in a library is zero, there are thresholds of adapter dimer in a library that may be present and still generate satisfactory sequencing results. As an example, random flow cell technologies from Illumina, including the MiniSeq and MiSeq, recommend that no more than 5% of the library be composed of adapter dimers. Patterned flow cells, such as the Illumina HiSeq X and NovaSeq 6000, are more sensitive and recommend a threshold of only 0.5% dimer[4]. Quality control (QC) steps are crucial to determining if a library contains any adapter dimers. The Agilent automated electrophoresis portfolio offers a suite of instruments that allow for accurate and reliable QC throughout NGS library preparation.

This technical overview highlights the sensitivity and resolution of the Agilent Fragment Analyzer and TapeStation systems to detect adapter dimers even lower than the recommended threshold of 0.5% and deliver an accurate assessment of the percentage of adapter dimer in the library. This information provides valuable data to evaluate the quality of the library and enables researchers to make important decisions regarding subsequent analysis.

NGS Library Preparation Schematic

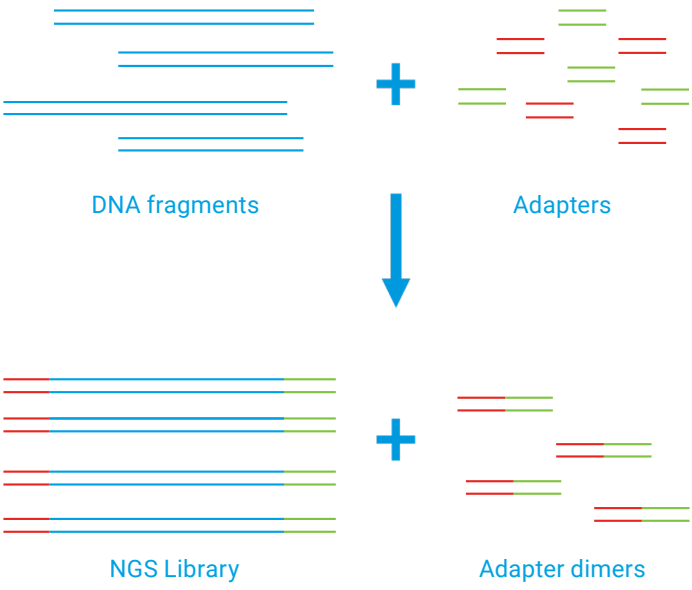


Figure 1. During the process of NGS library preparation, known DNA adapter sequences are ligated to the 5' and 3' ends of the DNA. Generally, one adapter will contain the primer sequence, while the other is used to bind the library to the flow cell for sequencing. Adapter dimers form when the two adapters ligate to each other instead of the target insert.

Experimental

A DNA library was prepared using the NEXTflex™ Rapid DNA-Seq Kit (Bioo Scientific, part number 5144-01). 0.8x SPRI bead cleanup was performed two times to eliminate any inherent dimers. An RNA library was prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, part number #E7530), following manufacturer's protocol.

A 150 bp NoLimits Fragment (Thermo Fisher Scientific) at 50 pg/µL was utilized to mimic an adapter dimer peak (hereafter referred to as adapter dimer). The fragment was added to aliquots of the NGS libraries such that the library concentration stayed consistently at 500 pg/µL and contained varying amounts of the 150 bp fragment: 0, 0.1, 0.2, 0.5, 1, and 5%. Controls were prepared for each fragment amount using 1x TE instead of library.

Samples were analyzed using Agilent automated electrophoresis instruments. Analysis on the Agilent 5200 Fragment Analyzer system was performed with the Agilent HS NGS Fragment kit (1-6000 bp) (part number DNF-474) and the Agilent HS Small Fragment kit (part number DNF-477). Samples were analyzed on the Agilent 4200 TapeStation system using the Agilent High Sensitivity D1000 ScreenTape (part number 5067-5584 and 5067-5585) and Agilent High Sensitivity D5000 ScreenTape (part number 5067-5592 and 5067-5593) assays.

The Agilent ProSize data analysis software was utilized to analyze the samples run on the Fragment Analyzer. The baseline was adjusted as necessary to ensure full integration of the smear sample. The minimum peak height was adjusted to 10 RFU, and the peak width to 40 seconds. Likewise, the TapeStation analysis software was used to analyze the samples run on the TapeStation. The peak width of the library was manually adjusted by moving the start and end points to include the entire smear and enable accurate percent integrated area calculations. Fragments that were visualized but not automatically detected were manually added as a peak. The sensitivity of both systems was analyzed based on their ability to visualize the fragment and to accurately report the percent concentration of the fragment peak compared to the total sample.

Results and discussion

Detection sensitivity of adapter dimer

The NGS library preparation workflow includes several instances at which QC checks are recommended, including evaluation of the integrity of both the starting sample and the final library prior to sequencing. One reason for this final QC step is to ensure that the library is devoid of small fragments such as adapter dimers, as they can significantly affect the sequencing data. Different sequencing platforms have a range of dimer that can be tolerated without interfering with the sequencing results. An acceptable threshold of dimer contamination can range from 5% to as little as 0.5%. The Fragment Analyzer and TapeStation systems can be used to detect and quantify very small amounts of dimer present within different types of NGS libraries. Each instrument uses various analysis kits and assays to assess samples covering a broad range of sizes and concentrations (Table 1).

Typically, short read NGS libraries are within a size range that is appropriate for analysis using the HS NGS kit for the Fragment Analyzer systems or the HS D5000 ScreenTape assay for the TapeStation systems. The HS NGS kit offers accurate sizing from 100 to 6,000 bp, while the HS D5000 ScreenTape assay ranges from 100 to 5,000 bp. In this example, an NGS library with a target insert size of 150 to 250 bp was analyzed with both systems and shown to have a peak size of about 370 bp following adapter ligation, with a distribution ranging from approximately 250 to 2,000 bp. After cleanup with SPRI beads, the library was absent of any inherent primer or adapter dimer. To assess the ability of the Fragment Analyzer and TapeStation to detect small percentages of adapter dimer, a 150 bp fragment was added to the library at 0.1% up to 5% of the concentration of the library. Each sample was analyzed on both systems in triplicate.

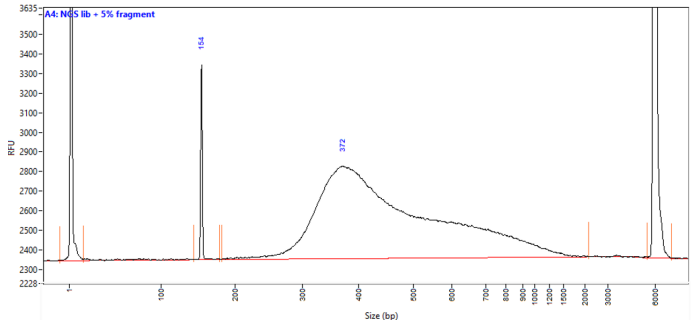
The ProSize data analysis software utilized with the Fragment Analyzer displays the results in both a digital gel format and an individual electropherogram of each sample. A Peak Table also records information about the sample, including the size, concentration, and percentage of each peak. Each sample is automatically analyzed, but the user can adjust settings such as the peak width, minimum peak height, and baseline setpoints to ensure that each peak is appropriately integrated. Figure 2 shows representative electropherogram images of the NGS library containing varying percentages of a 150 bp fragment from the Fragment Analyzer using the HS NGS kit. The adapter dimer fragment was detected at all variations tested, from 5% (Figure 2A) to 0.1% (Figure 2D). The percent total reported by the Fragment Analyzer was accurate for the 1%, 0.5%, and 0.1% samples, at 0.9%, 0.5%, and 0.1%. The adapter dimer fragment at 5% was reported at 4.2% of the total library concentration (Table 2). Additionally, the Fragment Analyzer displayed accurate sizing of the fragment, with a size of 153 or 154 bp in all sample wells and a percent error of less than 3%, within the kit specification of 5%.

The TapeStation analysis software also displays a digital gel image and electropherogram of each sample. The Peak Table provides the size, concentration, and percent integrated area for each peak. The width of a peak can be adjusted by moving the start and end points of the peak to ensure integration of the entire sample, and thus determine the accurate percent integrated area for each peak. With the HS D5000 assay, the TapeStation was able to detect the fragment at 5% to 0.5% of the total concentration, as shown in Figure 3. The adapter dimer fragment displayed an average size of 133 bp, slightly smaller than the expected 150 bp, with a percent error of less than 12%, within the kit specification of 15%. The percent integrated area reported by the TapeStation was within 0.3 percentage points from the expected 5%, 1%, and 0.5% samples (Table 2).

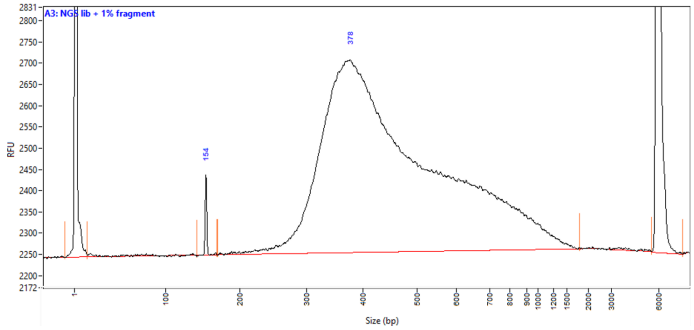
Table 1. Specifications of the kits used with the Agilent Fragment Analyzer systems and Agilent TapeStation systems for NGS library quality control.

Instrument and Kit	Sizing Range	Sizing Accuracy	Quantitative Range	Quantitative Accuracy	Resolution
Fragment Analyzer HS Small Fragment kit	50 – 1,500 bp	±5%	Smears: 100 – 5,000 pg/µL Fragments: 5 – 500 pg/µL	±25%	50 – 900 bp: 5% 900 – 1,500 bp: 10%
Fragment Analyzer HS NGS kit	100 – 6,000 bp	±5%	Smears: 50 – 5,000 pg/µL Fragments: 5 – 500 pg/µL	±25%	100 – 1,000 bp: 5% 1,000 – 6,000 bp: 10%
TapeStation HS D1000 ScreenTape assay	35 – 1,000 bp	±10%	10 – 1,000 pg/µL (sensitivity to 5 pg/µL)	±20%	35 – 300 bp: 15% 300 – 1,000 bp: 10%
TapeStation HS D5000 ScreenTape assay	100 – 5,000 bp	±15%	10 – 1,000 pg/µL (sensitivity to 5 pg/µL)	±25%	400 – 5,000 bp: 15%

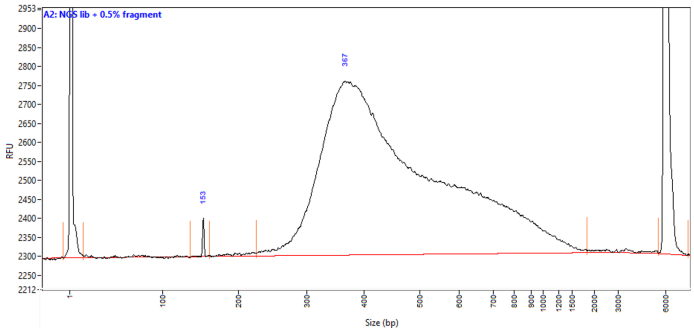
A. 5% Fragment



B. 1% Fragment



C. 0.5% Fragment



D. 0.1% Fragment

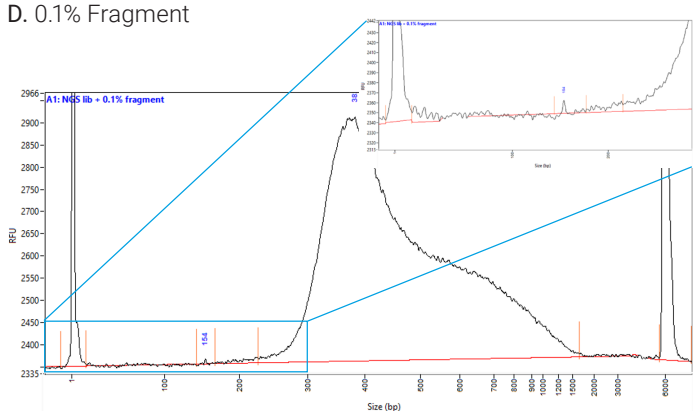
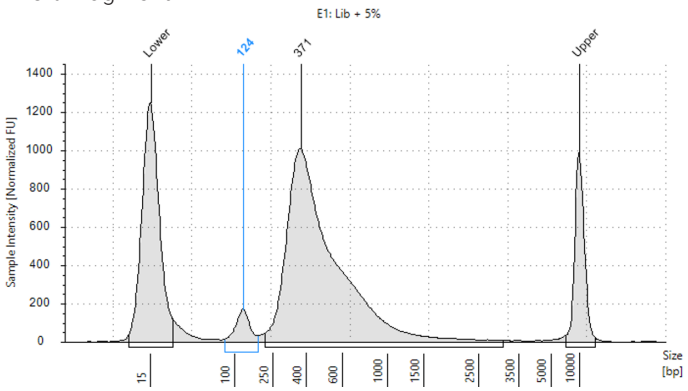


Figure 2. Representative electropherograms from the Agilent 5200 Fragment Analyzer system of an NGS library containing known amounts of adapter dimers using the Agilent HS NGS fragment kit. The 150 bp adapter dimer fragment can be visualized at A) 5%, B) 1%, C) 0.5%, and D) 0.1% of the total library concentration.

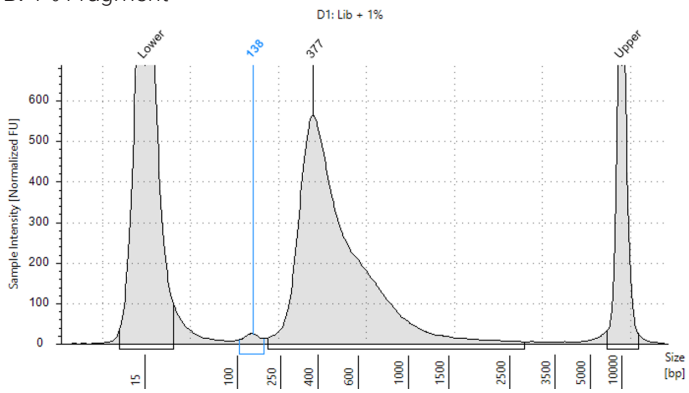
Table 2. Reported size and percentage of adapter dimer present within an NGS library analyzed on the Agilent 5200 Fragment Analyzer system and the Agilent 4200 TapeStation system.

	Fragment Analyzer HS NGS Kit		**TapeStation HS D5000 ScreenTape Assay**	
Expected Adapter Dimer	**Fragment Size (bp)**	**% (Conc.)**	**Fragment Size (bp)**	**% Integrated Area**
5%	153	4.2%	124	4.7%
1%	154	0.9%	138	1.3%
0.5%	153	0.5%	138	0.8%
0.1%	154	0.1%	Not detected	Not detected

A. 5% Fragment



B. 1% Fragment



C. 0.5% Fragment

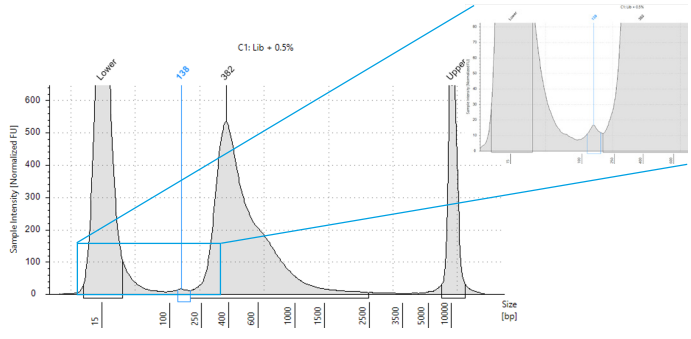


Figure 3. Representative electropherograms from the Agilent 4200 TapeStation system of an NGS library containing known amounts of adapter dimers using the Agilent HS D5000 ScreenTape assay. The adapter dimer fragment can be visualized at A) 5%, B) 1%, and C) 0.5% of the total library concentration.

Resolution of adapter dimers

Depending on the type of sample being analyzed, the automated electrophoresis systems offer a variety of kits and assays for different sizing ranges and separation resolution. Each assay utilizes different separation methods, gel chemistries, and sizing ladders. To demonstrate the differences in some of the kits that can be used for smaller molecular weight samples, RNA-Seq libraries containing varying amounts of adapter dimer were compared across both systems, with two kits of similar sizing ranges for each (Table 1). For the Fragment Analyzer, the samples were analyzed on the HS Small Fragment kit and the HS NGS kit. Samples were analyzed on the TapeStation using the HS D1000 assay and the HS D5000 assay.

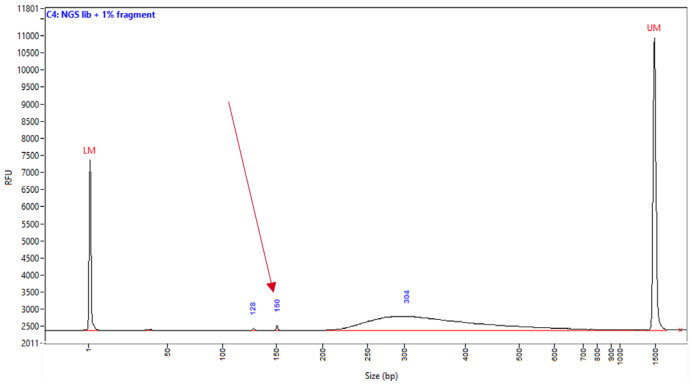
The RNA-Seq library preparation protocol used was optimized to generate approximately 200 bp inserts. As shown in the electropherogram images in Figure 4, the final library distribution ranged from approximately 200 to 600 bp, with the largest area of the curve at about 270 to 300 bp, depending on the analysis kit. The library contains a small amount of 130 bp dimer contamination, the 150 bp fragment, and the library, which can be seen at approximately 130 bp. This contamination was intentionally left in the sample, and a 150 bp fragment mimicking adapter dimer was spiked into the sample at varying amounts from 0.1% to 5% to examine the resolution capabilities of the analysis kits.

Shown in Figure 4 are representative examples of the library with a spike-in of 1% of the 150 bp fragment. The Fragment Analyzer provided high-resolution separation of the inherent 130 bp dimer contamination, the 150 bp fragment, and the library with both the HS Small Fragment (Figure 4A) and the HS NGS kits (Figure 4B). Both kits also showed accurate sizing and percent total calculations for the fragment from 0.2% to 5%, with the HS NGS kit also able to detect the fragment at as low as 0.1% of the total concentration (Table 3).

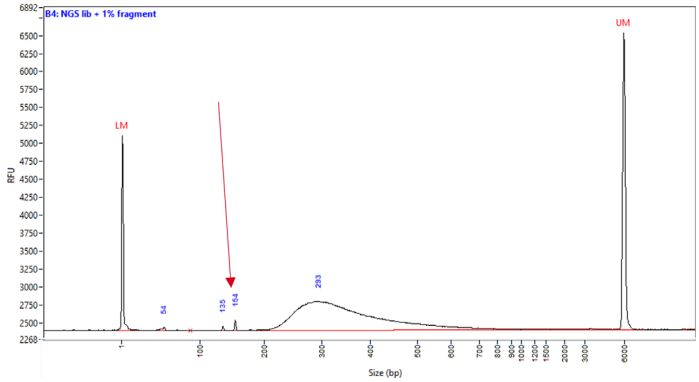
The TapeStation HS D1000 assay fully resolved the 130 bp dimer contamination, the 150 bp fragment, and the library (Figure 4C) and was able to detect the adapter fragment among the entire range from 0.1% to 5% spike-in (Table 3). Alternately, the HS D5000 assay was only able to visualize the fragment from the library at 5% of the concentration (Figure 4D). The concentration of the fragment in the 0.1% and 0.5% samples was below the assay's limit of detection of 5 pg/uL and could not be resolved from the library smear.

Evaluation of the same samples across multiple assays and instruments demonstrates the importance of choosing the most appropriate analysis method for different applications. The Fragment Analyzer and TapeStation data presented here can help users decide which method of QC is best for their samples depending on the type of sequencing they will be performing and the amount of dimer that is considered acceptable for their specific workflows.

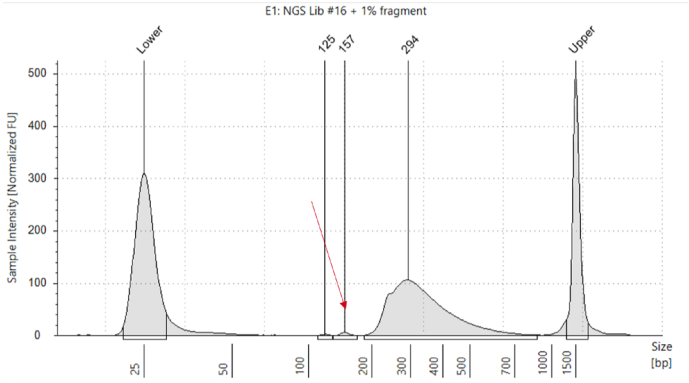
A. Fragment Analyzer HS Small Fragment kit. 1% adapter dimer.



B. Fragment Analyzer HS NGS kit. 1% adapter dimer.



C. TapeStation HS D1000 ScreenTape assay. 1% adapter dimer.



D. TapeStation HS D5000 ScreenTape assay. 5% adapter dimer.

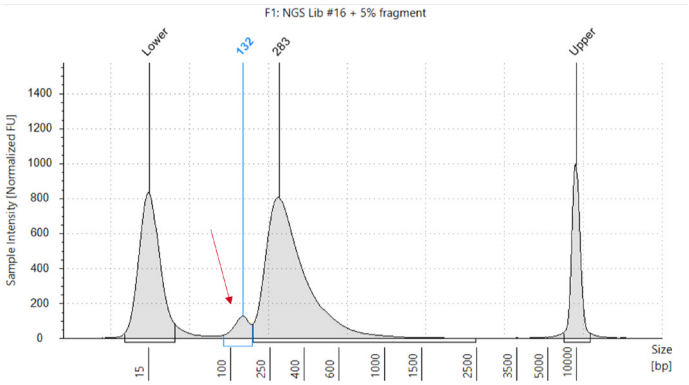


Figure 4. Representative electropherograms of an RNA-Seq NGS library with known amounts of adapter dimer contamination. The library containing 1% adapter dimer was visualized on the Agilent 5200 Fragment Analyzer with A) the HS Small Fragment kit and B) the HS NGS kit, and on the Agilent 4200 TapeStation system with the HS D1000 ScreenTape assay. D) 5% adapter dimer on the Agilent 4200 TapeStation system with the HS D5000 ScreenTape assay. The red arrow in each figure points to the 150 bp NoLimits DNA fragment spiked into the library to mimic the adapter dimer.

Table 3. Reported size and percentage of adapter dimer present within an RNA-Seq NGS library analyzed on the Agilent 5200 Fragment Analyzer system and the Agilent 4200 TapeStation system.

	Fragment Analyzer HS Small Fragment Kit		Fragment Analyzer HS NGS Kit		TapeStation HS D1000 ScreenTape Assay		TapeStation HS D5000 ScreenTape Assay	
Expected Adapter Dimer	**Fragment Size (bp)**	**% (Conc.)**	**Fragment Size (bp)**	**% (Conc.)**	**Fragment Size (bp)**	**% Integrated Area**	**Fragment Size (bp)**	**% Integrated Area**
5%	151	4.6%	154	4.9%	161	4.3%	132	5.8%
1%	150	0.9%	154	1.0%	157	0.7%	Not detected	Not detected
0.5%	151	0.4%	154	0.5%	159	0.4%	Not detected	Not detected
0.2%	151	0.2%	154	0.2%	170	0.1%	Not detected	Not detected
0.1%	Not detected	Not detected	154	0.1%	166	0.1%	Not detected	Not detected

Conclusion

As sequencing platforms are becoming more sensitive, steps must be taken to ensure that the sequencing libraries contain a minimum of undesired products that could impede the sequencing results. For example, the formation of adapter dimers can reduce the amount of adapters that are available to ligate to the target insert, resulting in fewer usable sequencing reads. Additionally, the small adapter dimer fragments present in the library will be preferentially sequenced over the target. As such, QC of the final library is vital to ensuring that the sample is of high integrity, the correct size, and contains minimal amounts of primer or adapter dimer.

This technical overview summarized the sensitivity and resolution that can be achieved with the Agilent Fragment Analyzer and the Agilent TapeStation by utilizing different analysis kits. The Agilent Fragment Analyzer HS NGS kits and the Agilent TapeStation HS D1000 kits both provide high sensitivity necessary to detect adapter dimer in traditional NGS libraries even lower than the 0.5% to 5% threshold recommended by sequencing cores. The Fragment Analyzer HS Small Fragment and HS NGS kits and the TapeStation HS D1000 kit provides high resolution to enable separation of low amounts of adapter dimer from the library smear. By providing valuable data regarding the quality of NGS libraries, the Agilent automated electrophoresis systems can help researchers make important decisions about their sequencing projects.

References

1. Adapter Dimers Causes, Effects, and How to Remove Them. *Illumina Knowledge article*, publication number 1911. **2023**.
2. How Short Inserts Affect Sequencing Performance. *Illumina Knowledge article,* publication number 3874. **2023**.
3. *Frequently Asked Questions*. Stanford Medicine, Genome Sequencing Center. https://med.stanford.edu/gssc/faq.html (accessed 2023-03-07).
4. *Primer Dimer Sequencing Requirements.* University of North Carolina at Chapel Hill, Integrated Genomics Core. https://Integrated /genomics/wp-content/uploads/sites/708/2022/02/Primer-Dimer-White-Paper_12-2021_FINAL.pdf (accessed 2023-03-07).

www.agilent.com/genomics/automated-electrophoresis

For Research Use Only. Not for use in diagnostic procedures.
PR7001-0902

This information is subject to change without notice.

Published in the USA, June 1 2023
5994-6079EN

Agilent
Trusted Answers