Impact of NGS Library Sizing Accuracy on Sequencing

Whitney Pike, Chava Pocernich, Kit-Sum Wong, and Steve Siembieda
Advanced Analytical Technologies, Inc., 2450 SE Oak Tree Ct., Ankeny, IA 50021

ADVANCED ANALYTICAL®
Automating genomic discovery

Introduction

Quality control of next-generation sequencing (NGS) libraries is essential for obtaining good sequencing results. Capillary electrophoresis and microfluidics instruments have become increasingly important in determining the quality, quantity, and size of NGS samples due to the advantages of low sample input, high sensitivity, and short analysis time compared to agarose gel electrophoresis.

The Fragment Analyzer™ Automated CE System (Advanced Analytical Technologies, Inc.) is a capillary electrophoresis instrument with a flexible throughput platform capable of analyzing 12, 48, or 96 samples simultaneously. It has been used for the reliable qualification and quantification of nucleic acids over a wide range of concentrations and sizes. When employed in an NGS pipeline, the Fragment Analyzer can assess all sample types throughout library construction, from the initial assessment of sample integrity through final qualification.

Recently, we have analyzed the Bioanalyzer's High Sensitivity DNA ladder on both the Fragment Analyzer and the Agilent 2100 Bioanalyzer® microfluidics system and revealed vast differences between the separation profiles and data analysis method. Furthermore, the NGS library analysis indicated that the Bioanalyzer consistently reported higher smear sizing compared to the Fragment Analyzer. These disparate data indicate a fundamental difference in how the Fragment Analyzer and the Bioanalyzer calculate the concentration and size of DNA smears. Here we used an NGS library with a broad distribution to examine the sizing accuracy of the Fragment Analyzer and the Bioanalyzer, and compared the data to the sequencing results of the library.

Fragment Analyzer™ Automated CE System

Parallel Capillary Electrophoresis was performed on the Fragment Analyzer platform, in the 12-capillary array format. A high-output LED source provided an excitation wavelength of 470 nm with emission collected at 500-600 nm. The capillary array was conditioned with Capillary Conditioning Solution, and then filled with NGS Fragment Separation Gel (DNF-240). For each run, the pre-loaded method for the High Sensitivity NGS Fragment Analysis Kit (DNF-474) was selected and queued up as outlined in the user manual. Data analysis was performed using PROSize® Data Analysis software (Advanced Analytical).

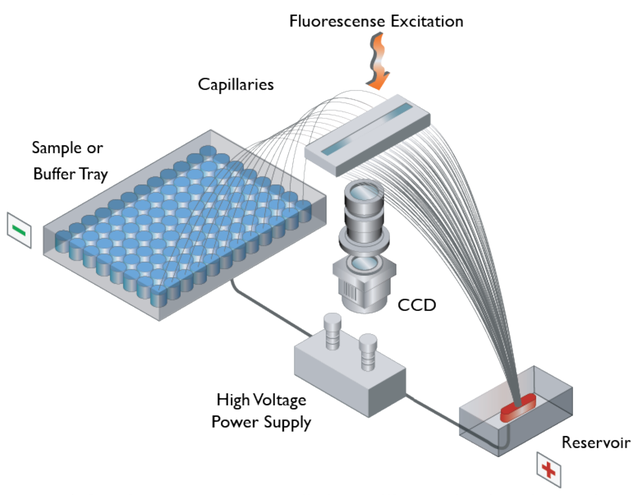
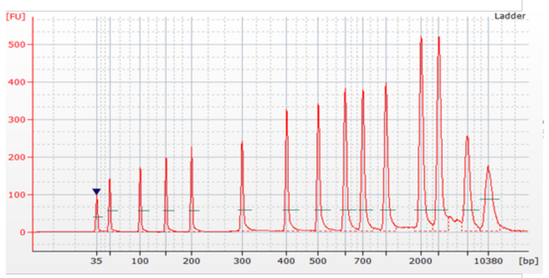


Figure 1: Advanced Analytical Technologies Fragment Analyzer uses parallel capillary electrophoresis and fluorescence detection of DNA/RNA to rapidly separate and quantify DNA or RNA.

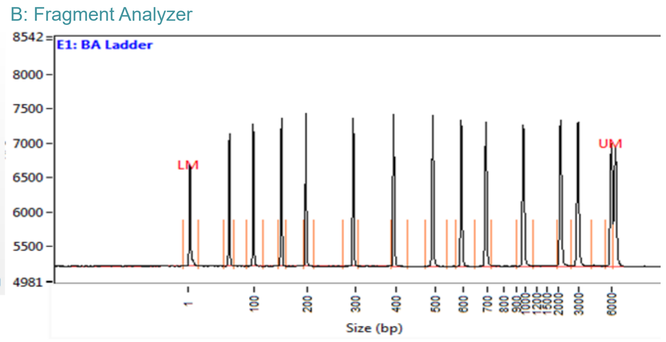
- 96-capillary, 48-capillary or 12-capillary capability
- Fully automated workflow enables continuous loading/exchange of up to three 96-well plates along with scheduling while the instrument is running
- Samples are injected by electrokinetic (voltage) mode
- Measures both quantity and quality of gDNA, NGS smears, DNA fragments and RNA

Ladder Quantification

A: Bioanalyzer



B: Fragment Analyzer



C.

Bioanalyzer Ladder		
Size (bp)	Conc. [pg/µL]	Molarity
35 (LM)	125	5,411.30
50	150	4,545.50
100	150	2,272.70
150	150	1,515.20
200	150	1,136.40
300	150	757.6
400	150	568.2
500	150	454.5
600	150	378.8
700	150	324.7
1,000	150	227.3
2,000	150	113.6
3,000	150	75.8
7,000	150	32.5
10,380 (UM)	75	10.9

D.

	Bioanalyzer				Fragment Analyzer		
Known Size [bp]	Corr. Peak Area	Conc. [pg/µL]	Ratio Conc./CPA		Corr. Peak Area	Conc. [pg/µL]	Ratio Conc./CPA
35 (LM)	1.50	125.00	83.33				
50	142.50	152.46	1.07		5.95	158.17	26.61
100	148.10	165.52	1.12		6.14	163.29	26.60
150	158.70	164.43	1.04		5.93	157.65	26.60
200	180.30	161.15	0.89		6.06	161.19	26.60
300	190.80	157.80	0.83		5.82	154.86	26.60
400	202.20	149.57	0.74		5.49	146.15	26.60
500	201.50	152.14	0.76		5.10	135.60	26.60
600	214.80	152.47	0.71		5.34	141.95	26.60
700	214.10	156.06	0.73		5.03	133.90	26.60
1,000	235.00	150.83	0.64		6.36	169.32	26.61
2,000	260.30	150.49	0.58		5.31	141.29	26.60
3,000	283.70	161.8	0.57		5.30	141.06	26.61
7,000	216.90	126.01	0.58				
10,380 (UM)	8.30	75.00	9.04				

Figure 2: A High Sensitivity DNA ladder (#5067-4627, Agilent Technologies) separated on the (A) Bioanalyzer and (B) Fragment Analyzer. *Calculation of the concentration is dependent on the integration of the peak area; therefore, peaks with equal corrected peak area (CPA) will have the same concentration. Although the Bioanalyzer reports all ladder peaks as having the same concentrations (C), the electropherogram (A) displays ladder fragments of varying peak heights and areas. The same ladder on the Fragment Analyzer (B) displays similar peak heights and areas for all the ladder fragments. Additionally, the ladder was run as a sample on the Bioanalyzer and the Fragment Analyzer, allowing for the calculation of the CPA (D). Despite similar concentrations reported on the Bioanalyzer, CPA varied greatly. The Fragment Analyzer reported similar corrected peak areas with corresponding similar concentrations and consistent concentration/ CPA ratios for each ladder fragment.*

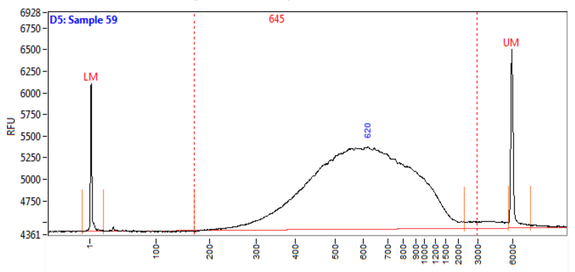
Library Size Variation Between Instruments

A:

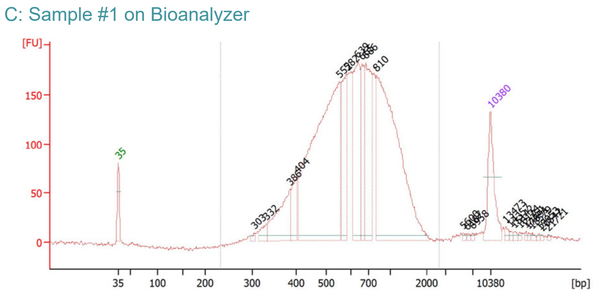
Sample ID	Fragment Analyzer Average Smear Size (bp)	Bioanalyzer Average Smear Size (bp)	Sizing % Increase on Bioanalyzer
1	645	698	8.2%
2	595	630	5.9%
3	411	438	6.6%
4	442	514	16.3%
5	467	545	16.7%
6	486	826	70.0%

The Fragment Analyzer consistently sizes the NGS Libraries smaller than the Bioanalyzer.

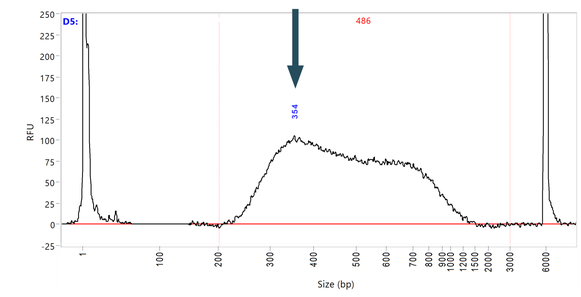
B: Sample # 1 on the Fragment Analyzer



C: Sample #1 on Bioanalyzer



D: Sample #6 on the Fragment Analyzer



E: Sample #6 on Bioanalyzer

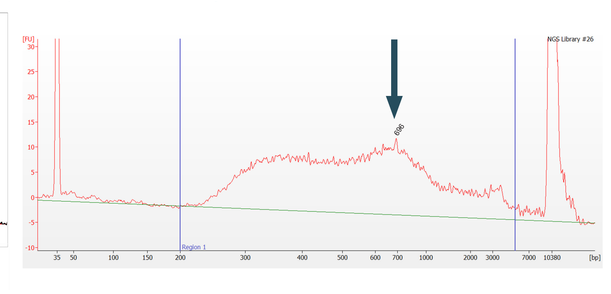
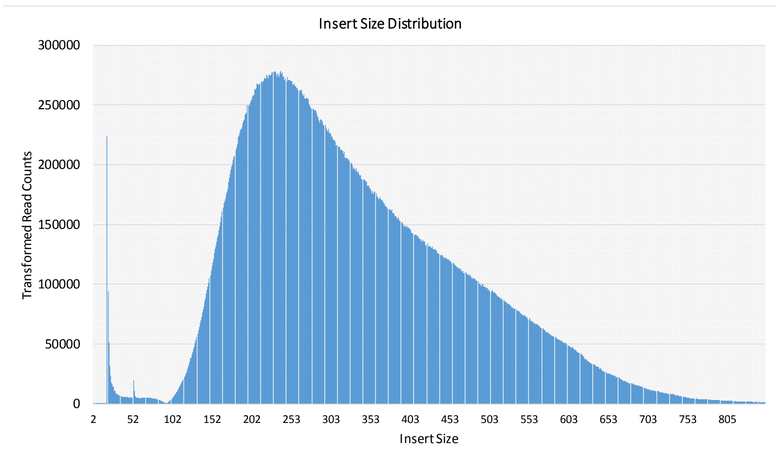


Figure 3: The average smear size determined by the Fragment Analyzer or the Bioanalyzer is dependent upon the distribution of the smear. A) NGS average smear size comparison between the Fragment Analyzer and the Bioanalyzer. All the smear sizes were consistently larger on the Bioanalyzer compared to the Fragment Analyzer. (B-E) Size distribution comparison of NGS samples #1 (B,C) and #6 (D,E) analyzed on the Fragment Analyzer (B,D) and Bioanalyzer (C,E). Sample #1 is a fairly symmetrical library, with a similar distribution across the smear, resulting in comparable average smear sizes between the two instruments. In contrast, Sample #6 is a more unevenly distributed smear, with the electropherograms from the Fragment Analyzer displaying a larger area of smaller molecular weight fragments compared to the Bioanalyzer (see black arrows, indicating the highest peak reported), resulting in the smaller sizing of the NGS smear.

The Fragment Analyzer Delivers More Accurate Library Sizing

To further investigate the discrepancy in the average smear size of libraries between the Fragment Analyzer and the Bioanalyzer, Library Sample #6 (Fig.3) was submitted for sequencing using the Illumina HiSeq 3000 platform, with a 150 paired end run. The library was prepared from Universal Reference Mouse RNA (Stratagene) using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England BioLabs). To overcome any possible sequencing bias towards smaller fragments, two internal control libraries were prepared using 200 bp and 600 bp NoLimits Fragments (Thermo-Fisher). These internal control libraries were spiked into the library of interest prior to sequencing, with 90% of the final pool composed of the library of interest, and the remaining 10% being equimolar concentrations of the two fragment libraries.

	Fragment Analyzer Average Smear Size (bp)	Bioanalyzer Average Smear Size (bp)	Sequencing Average Smear Size (bp)	% Difference Fragment Analyzer-Sequencing	% Difference Bioanalyzer-Sequencing
Prepared Library #6	486	826	--	--	--
Library minus Adapter	358	698	332	7.8%	52.4%



The sequencing data has an average smear size and a similar distribution to the Fragment Analyzer.

Figure 4: NGS Sequencing Analysis. *A) The average smear size of Library #6 reported by the Fragment Analyzer and the Bioanalyzer (see Fig. 3D,E) compared to the average insert size determined from sequencing results. Differences between the insert size obtained from sequencing compared to the Fragment Analyzer or Bioanalyzer are due to the adapters being trimmed off during Bioinformatics analysis, thus the expected size of the library minus the adapters is also listed for comparison to sequencing data. B) Bioinformatics analysis of the HiSeq run indicated that the insert size of the library was ~200bp. Internal controls were added to the library to eliminate sequencing bias towards smaller fragments and to determine a more accurate insert size. A linear equation was derived from the frequency count of the internal controls and was used to determine an average insert size free of any sequencing bias. This transformed insert size indicates that the average smear size of the library is 332 bp, which is consistent with what is seen on the Fragment Analyzer (358 bp). The Bioanalyzer reports an average smear size for the library of 698 bp (A), indicating that the Bioanalyzer does not appear to take into account the smaller portions of the library when reporting smear size. Because the Fragment Analyzer provides more accurate smear sizing, a more precise molarity calculation can be determined, leading to more efficient flow cell loading and optimum sequencing data generation.*

Summary

- Sequencing results indicated that the Fragment Analyzer provided accurate library distribution and size.
- Bioanalyzer smear size was higher than the Fragment Analyzer and the Sequencing results.

Acknowledgments

Iowa State University DNA Facility and ISUGIF