NGS library QC by capillary electrophoresis on Applied Biosystems SeqStudio genetic analyzers

NGS library QC

Authors

Erich Klem, Michael Zianni,
Field Applications Scientists,
Thermo Fisher Scientific,
Carlsbad, California, USA

Keywords

Next-generation sequencing, library qualification, capillary electrophoresis, fragment analysis, genetic analyzer, multiplexing

Introduction

Next-generation sequencing (NGS) is a powerful tool that performs highly multiplexed, parallel sequencing of millions or billions of DNA fragments in a single run. This technology enables the rapid accumulation of data that previously took years to generate, greatly accelerating the pace of discovery. However, because NGS is often time consuming and expensive, rapid and cost-effective techniques that improve the odds of a successful sequencing run are highly valuable.

High-quality NGS data begins with high-quality, accurately quantified libraries. Numerous methods of library quantification are available, including real-time PCR (qPCR), digital PCR (dPCR), and fluorometric techniques such as Invitrogen™ Qubit™ assays. In contrast, library quality assessment (QC) involves evaluating the overall profile of the library to gauge its size distribution, purity, and integrity. There are comparatively few tools available for library QC, yet this step is just as critical as accurate quantification. Short fragments are often preferentially amplified during templating and cluster generation, so a large proportion of short fragments can produce low-quality sequencing data. Therefore, detecting excessive short fragments and incorrectly sized libraries as part of a robust library QC step before sequencing can save both time and money.

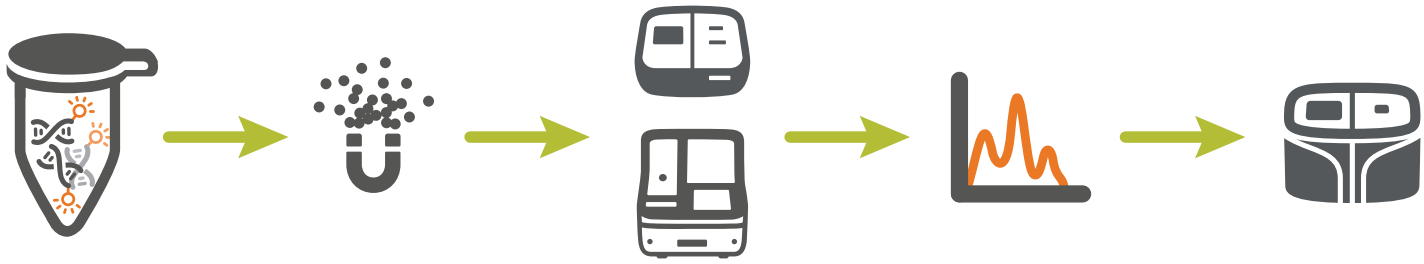

Figure 1. High-level workflow. NGS libraries are first amplified to incorporate fluorescent dyes for visualization by capillary electrophoresis, followed by an optional cleanup step. Library quantification using the Ion Torrent™ Ion Library TaqMan™ Quantitation Kit can be performed on labeled libraries so that quantification and QC are performed with the same material. The labeled libraries are run on an Applied Biosystems™ genetic analyzer with a sizing standard, then visualized using Applied Biosystems™ InnoviGene™ Suite, Applied Biosystems™ GeneMapper™ 6 Software, or Applied Biosystems™ Peak Scanner™ Software.

Capillary electrophoresis (CE) can size DNA fragments over a range of 20 to 1,200 base pairs with single-base resolution, providing accurate and precise information about library quality. Additionally, the multiplexing capability of CE allows the simultaneous evaluation of multiple libraries, increasing efficiency.

In this case study, we demonstrate that CE using Applied Biosystems™ genetic analyzer instruments and reagents offers a fast and cost-effective solution for NGS library QC. Furthermore, library QC by CE combined with quantification by qPCR or dPCR provides a powerful solution for library characterization.

Materials and methods

Instruments and polymers

This workflow was demonstrated on the Applied Biosystems™ SeqStudio™ Genetic Analyzer and the Applied Biosystems™ SeqStudio™ Flex Genetic Analyzer but may be performed on any Applied Biosystems genetic analyzer system with factory-installed G5 (DS-33) dye set filters. For the SeqStudio Flex Genetic Analyzer, Applied Biosystems™ POP-7™ polymer is recommended.

Library amplification and labeling

The oligonucleotide sequences used for Ion Torrent™ Ion AmpliSeq™ library amplification are listed in **Table 1**.

Primer sequences for other NGS platforms will be provided upon request.

After adapter ligation and purification, the libraries were amplified and labeled using the 1X Library Amp Mix from the Ion Torrent™ Ion AmpliSeq™ Library Preparation Kit Plus or the Invitrogen™ Platinum™ PCR SuperMix High Fidelity from the Ion Torrent™ Ion AmpliSeq™ Library Preparation Kit 2.0 and the primers listed in **Table 1**. Reaction mixes were formulated according to the instructions in **Table 2**.

Table 2. Reaction mix for aliquot or whole-library amplification

Component	Library Aliquot	Whole Library*
1X Library Amp Mix (black cap) or Platinum PCR SuperMix High Fidelity	72 µL	50 µL
10 µM A-Universal-LibQC	1.5 µL	1.0 µL
10 µM **dye-labeled** P1-LibQC	1.5 µL	1.0 µL
Library	25 µL	N/A
Total volume	**100 µL**	**52 µL**

*For whole-library amplification, the 52-µL amplification reaction mix can be added directly to the air-dried beads after purification. For further information about 1X Library Amp Mix, see *Ion AmpliSeq™ Library Kit Plus User Guide* (Publication no. MAN0017003). For Platinum PCR SuperMix High Fidelity, see *Ion AmpliSeq™ Library Kit 2.0 User Guide* (Publication no. MAN0006735).

Table 1. Ion AmpliSeq library amplification primers

Primer name	Sequence (5' to 3')	Purification	Stock conc. (in Low TE)
A-Universal-LibQC	CCATCTCATCCCTGCGTGTCTCCGACTCAG	HPLC	10 µM
<dye>*-P1-LibQC	**<dye>***-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT	HPLC	10 µM

* **<dye>**: add any of the following G5 dye set fluorophores to the 5' end: 6FAM™, VIC™, NED™, or PET™. Labeled primers can be ordered from thermofisher.com/5primelabeledprimers.

The libraries were amplified using the thermal cycling conditions described in **Table 3**.

Note: the number of cycles to perform depends on the starting library concentration. Use of 20 cycles is recommended for low-concentration libraries (<10 pM), but the number of cycles should be reduced for higher-concentration libraries[1].

Table 3. Thermal cycling conditions for library amplification and labeling

Temperature	Time	Cycles
98°C	120 seconds	1
98°C	15 seconds	20*
64°C	60 seconds	
10°C	Hold	1

* Reduce the number of cycles for libraries with a higher starting concentration

Library cleanup and dilution

Amplified and labeled libraries were either purified using Applied Biosystems™ MagMAX™ Pure Bind Beads prior to CE or loaded directly without cleanup.

Note: other paramagnetic SPRI-type beads may be suitable for cleanup but were not tested.

Bead-purified libraries were combined with Applied Biosystems™ Hi-Di™ Formamide and diluted Applied Biosystems™ GeneScan™ 600 LIZ™ dye Size Standard v2.0 or Applied Biosystems™ GeneScan™ 1200 LIZ™ dye Size Standard according to **Table 4**.

Table 4. Dilution of purified library and size standard

Component	Volume
LIZ dye size standard diluted 1:250 in Hi-Di Formamide	10 µL
Amplified library (post clean-up)	1 µL

Libraries loaded directly without cleanup were combined with H-Di Formamide and the indicated GeneScan LIZ dye size standard according to **Table 5**.

Table 5. Dilution of non-purified library and size standard

Component	Volume
Hi-Di Formamide	10 µL
LIZ dye size standard (undiluted)	1 µL
Amplified library (no cleanup)	1 µL

Multiplexing library QC on an Applied Biosystems genetic analyzer

With CE, up to four libraries can be analyzed in the same well (multiplexing) by using different dyes for different libraries. Additionally, libraries with non-overlapping size distributions can be multiplexed using the same dye, which further increases the number of libraries that can be analyzed in a single well. Dye-labeled libraries can be sequenced directly on Ion Torrent™ sequencing systems. If using a fluorophore-based NGS system, libraries labeled with NED™ may be suitable for sequencing, depending on the characteristics of the NGS system.

The diluted libraries were vortexed thoroughly, then centrifuged briefly. The mixture was incubated for 3 minutes at 95°C, then immediately placed on ice for at least 2 minutes followed by brief centrifugation.

Capillary electrophoresis and analysis

The diluted libraries were run on a SeqStudio genetic analyzer using the default fragment analysis or long fragment analysis protocols. The FSA files were analyzed using InnoviGene Suite software, GeneMapper software, and Peak Scanner software as specified.

Results

CE enables detection of poor-quality libraries

We first sought to determine whether CE can be used to differentiate between high- and low-quality libraries. Two whole-genome fragment libraries had been quantified before sequencing, but no quality checks were performed. Library A produced satisfactory results, whereas Library B yielded less than 2% final reads (98% of reads were filtered out) despite excellent templating and loading efficiencies. To assess the quality of the two libraries, Library A was labeled with PET and Library B was labeled with 6FAM prior to running on the genetic analyzer. Library A (**Figure 2A**, red) produced a broad smear above 200 bp, which represents the high-quality portion of the fragment library. In contrast, Library B (**Figure 2B**, blue) did not produce the expected profile, only several low-molecular-weight peaks representing overdigested products and concatemers. Therefore, CE can be used to distinguish a high-quality library from a low-quality one.

[1] The number of cycles was determined empirically by amplification of a known 6.8-pM *Escherichia coli* library. Briefly, tubes were removed from the thermal cycler after cycles 4 through 40 in 3-cycle increments, and the resultant libraries were quantified by qPCR and qualified by CE (data not shown).

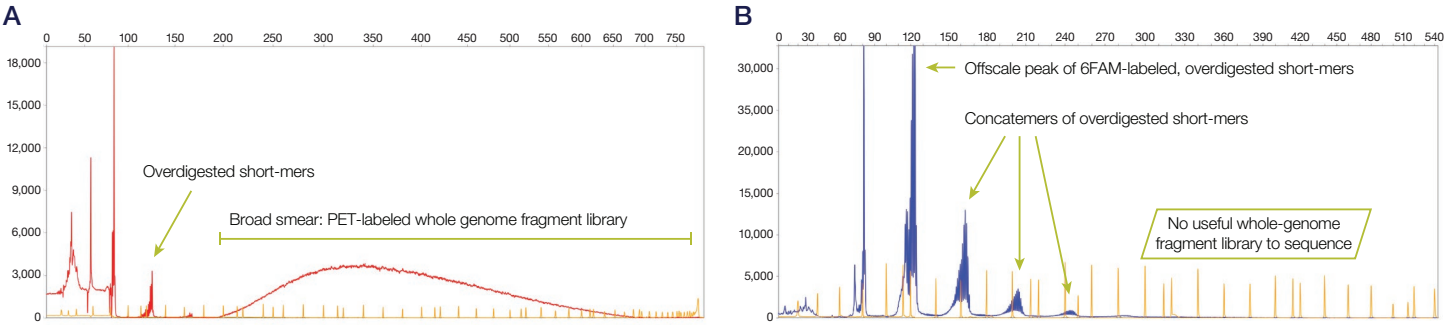

Figure 2. Discrimination of high- and low-quality libraries by CE. Libraries were run with the GeneScan 600 LIZ size standard, and data were analyzed using InnoviGene Suite software. **(A)** PET-labeled whole-genome fragment library that produced high-quality sequencing results. The broad peak above 200 bp represents the readable library. **(B)** FAM-labeled whole-genome fragment library that produced a failed sequencing run. The lack of a broad peak combined with shorter distinct peaks demonstrates overdigestion and no functional library for sequencing.

CE supports multiplexing of libraries in the same well

Next, to demonstrate different options for multiplexing, four libraries were prepared using the following Ion AmpliSeq panels: Ion Torrent™ Ion AmpliSeq™ Cancer Hotspot Panel v2, Ion Torrent™ Ion AmpliSeq™ SARS-CoV-2 Insight Research Assay, Ion Torrent™ Genexus™ Controls, and an Ion Torrent™ Ion AmpliSeq™ custom 375-bp assay. The libraries were labeled with the indicated dyes, run in multiplex with the Applied Biosystems™ GeneScan™ 600 LIZ™ size standard, and analyzed using both GeneMapper and Peak Scanner software. As shown in **Figure 3**, libraries with overlapping size distributions can be multiplexed using different dyes, and libraries with non-overlapping size distributions can be multiplexed using the same dye.

CE can QC libraries for different NGS platforms

To show that CE can be used to assess library quality for other NGS platforms, whole-genome and 16S fragment libraries were prepared using reagents from another leading NGS supplier and analyzed on a genetic analyzer. An example of a shorter-read, broadly size-selected library (whole genome) is provided in **Figure 4A**, and an example of a longer-read library with discrete fragment sizes (16S) is provided in **Figure 4B**.

Libraries can be analyzed with or without cleanup

Finally, to demonstrate that libraries do not need to be purified before CE, two libraries with similar size distributions were prepared using the Ion Torrent™ Ion AmpliSeq™ CarrierSeq™

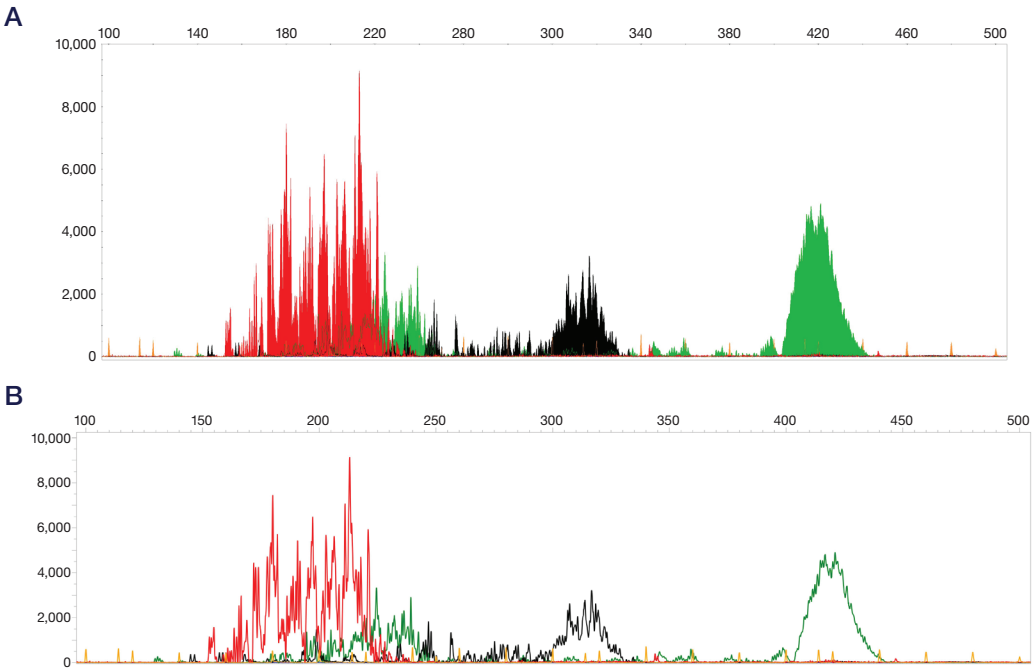

Figure 3. Multiplex library analysis using both GeneMapper and Peak Scanner software. Red indicates Ion AmpliSeq Cancer Hotspot v2 library labeled with PET; black indicates SARS-CoV-2 Insight Research library labeled with NED; green (≤300-bp peaks) indicates Genexus Controls library labeled with VIC; and green (350–450-bp peaks) indicates the Ion AmpliSeq custom 375-bp library. Libraries were run with the GeneScan 600 LIZ size standard. **(A)** Analysis performed using GeneMapper software. **(B)** Analysis performed using Peak Scanner software.

ECS Panel and the Ion Torrent™ Ion AmpliSeq™ CarrierSeq™ ECS ACMG Research Supplemental Panel - Tier 3. As shown in **Figure 5**, library cleanup is not required for QC by CE.

Discussion

In this case study, we describe a fast, economical, and effective method for qualifying various types of NGS libraries on Applied Biosystems genetic analyzers. An in-depth assessment of library size distribution, purity, and integrity can be obtained in less than 90 minutes, and a preview can be obtained in as little as 30 minutes via the live data capture on the instrument.

Compared with other library qualification techniques, the Applied Biosystems genetic analyzer systems offer high resolution across a wide size range, the ability to multiplex libraries, and multiple throughput options. Combined with a robust quantification method such as qPCR or dPCR, library qualification by CE can help you maximize your NGS budget by reducing the number of failed sequencing runs.

Selecting and optimizing settings for data analysis

Most NGS libraries contain a distribution of sizes rather than a single peak because they consist of a mixture of heterogeneous PCR products or broadly size-selected fragments. Therefore, the baseline window setting, which aids in normalizing baseline shifts, must be set based on the expected profile of the library being qualified. For broad smear-type libraries (such as those shown in **Figure 2A** and **Figure 4A**), start with a baseline window setting of 1,001 – 5,001. For libraries with discrete fragment sizes and spiky distributions (such as those shown in **Figure 3** and **Figure 4B**), a setting of 101 is often sufficient for robust peak detection. Compare the raw data with the analyzed data at various settings to further optimize analysis parameters for a specific library.

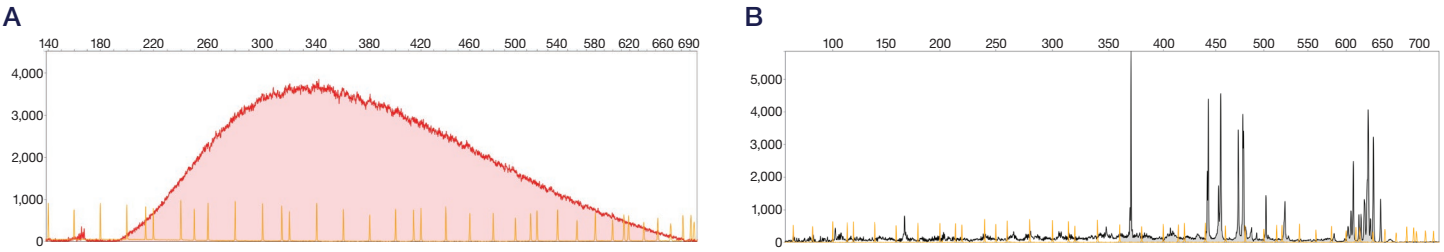

Figure 4. Use of CE for QC of non-Ion Torrent libraries. Libraries were prepared using reagents from another leading NGS supplier. **(A)** Whole-genome library labeled with PET (red), run with the Applied Biosystems™ GeneScan™ 1200 LIZ™ size standard, and analyzed using InnoviGene Suite software. **(B)** 16S fragment library labeled with NED (black), run with the GeneScan 1200 LIZ size standard, and analyzed using InnoviGene Suite software.

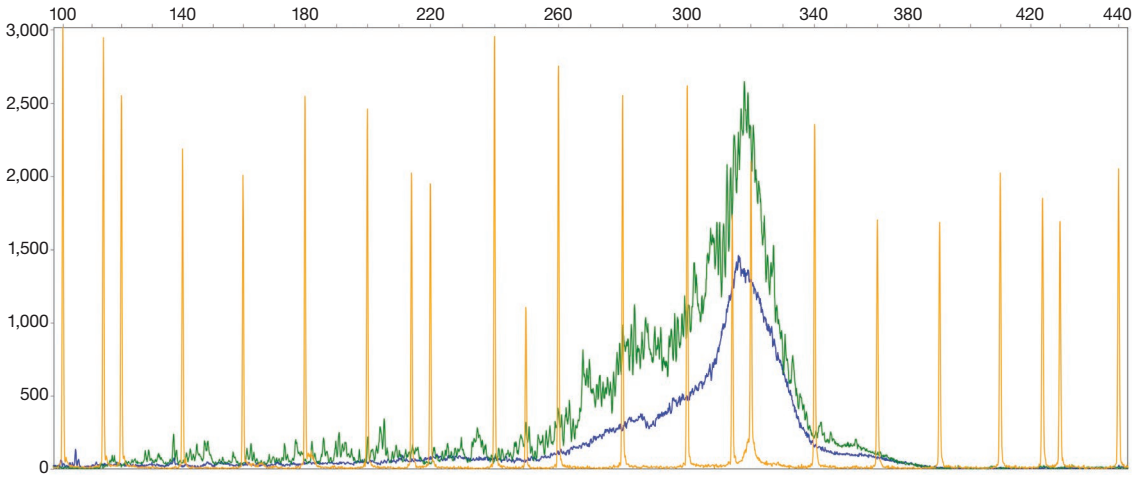

Figure 5. Resolution of multiplexed overlapping libraries without library cleanup. The CarrierSeq ECS library was labeled with 6FAM (blue), and the CarrierSeq ECS ACMG library was labeled with VIC (green) and run with the GeneScan 600 LIZ size standard. Data were analyzed using InnoviGene Suite software.

Ordering information

Description	Catalog No.
Equipment and software	
Applied Biosystems ProFlex PCR System, 96-well	4484075
Applied Biosystems VeritiPro Thermal Cycler, 96-well	A48141
Applied Biosystems SeqStudio Genetic Analyzer	A35644
Applied Biosystems SeqStudio 24 Flex Genetic Analyzer	A53630
Applied Biosystems SeqStudio 8 Flex Genetic Analyzer	A53627
Applied Biosystems 3730xl Genetic Analyzer	A41046
Ion Torrent Ion Chef Instrument	4484177
Ion Torrent Ion GeneStudio S5 Series	A38194, A38195 A38196
Applied Biosystems InnoviGene Suite	A40001762
Reagents and consumables	
5' labeled primers	450007*
Ion Torrent Ion AmpliSeq Library Kit Plus	4488990
Ion Torrent Ion AmpliSeq Library Kit 2.0	4475345
Applied Biosystems Hi-Di Formamide	4440753
Applied Biosystems GeneScan 600 LIZ dye Size Standard v2.0	4311320
Applied Biosystems GeneScan 1200 LIZ dye Size Standard	4408399
Invitrogen (Low) TE Buffer (10 mM Tris-HCl, pH 8.0, 100 µM EDTA)	12090015

*Order from thermofisher.com/5primelabeledprimers

Learn more at **thermofisher.com/ce**