

Application Note

CRISPR amplicon screening

Authors

Nancy Nabils
Senior Applications Scientist

Jennifer Pavlica
Applications Team Manager

Rachel Kasinskas
Director of Scientific Support
and Applications

Roche Sequencing & Life Science
Wilmington, MA, USA

Shipra V. Gupta
Automation Specialist

Leslie Grimmer
Scientist

Maura Berkeley
Lab Supervisor

Zachary Herbert
Associate Director

Molecular Biology Core Facilities,
Dana-Farber Cancer Institute
Boston, MA, USA

Date of first publication:
August 2018

Publication date of this version:
August 2022

Targeted CRISPR indel screening using an automated KAPA HyperPrep workflow

Discovery of the CRISPR system has enabled simple and cost-effective gene editing with significant application potential in biomedical research. An obligate step in many gene editing experiments is verification of the DNA sequence at the targeted genomic locus. Rapid adoption of CRISPR technology coupled with application across multiple model systems necessitates quick, accurate, and cost-effective sequencing of edited cell lines and/or organisms. Described here is a service provided by the Molecular Biology Core Facility (MBCF) at the Dana-Farber Cancer Institute using the KAPA HyperPrep Kit for automated screening of CRISPR target amplicons.

Introduction

The clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system was adapted from a bacterial immune system into a powerful gene editing technology, allowing for direct modification of DNA in most organisms.^{1,2} Application of this system has shown great promise in genomics, allowing researchers to comprehensively interrogate the function of various genes across diverse phenotypes.

Introduction of the Cas9 endonuclease and a target-specific guide RNA (gRNA) into mammalian cells results in a Cas9-mediated double strand break (DSB), repaired either by non-homologous end joining (NHEJ) or homology-directed repair (HDR). The error-prone NHEJ pathway results in variable short insertions and deletions (indels) near the DSB, while the HDR pathway uses a repair template to edit the DSB in a more specific and precise manner. In gene editing experiments, in which both NHEJ and/or HDR may play a role, the resulting heterogeneity of indels introduced at Cas9 target sites—coupled with variable allelic editing efficiencies—results in cellular populations with a highly mixed genetic profile at the target site. Thus, an obligate step following gene editing is to characterize the new genomic sequence in clonal isolates, by first expanding the cell pool into single-cell clones and then sequencing the region spanning the target Cas9 cut site. (Figure 1).

As many researchers desire to edit multiple targets—in some cases using multiple gRNAs per target across multiple cell lines—a high-throughput sequence verification methodology is essential. For service providers such as core facilities, the methodology must be robust, such that a single process may be applied to samples obtained from multiple groups and therefore varying widely in their characteristics. Described here is a solution for the automated, high-throughput screening of CRISPR amplicons utilizing the KAPA HyperPrep Kit for robust library preparation prior to NGS. This methodology is available at the Molecular Biology Core Facility (MBCF) at the Dana-Farber Cancer Institute.

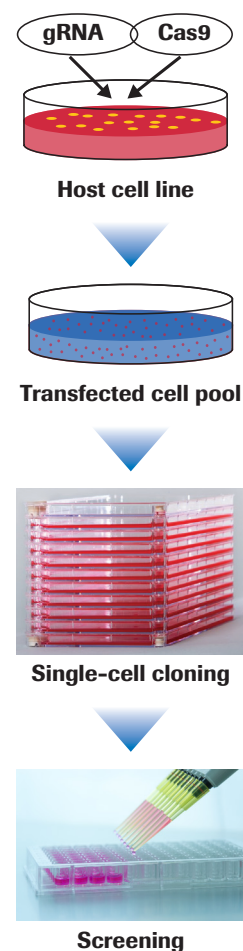


Figure 1: Schematic of gene editing and single-cell expansion



Workflow design and methods

For high-throughput NGS library preparation methodology at MBCF, the following properties were desired:

- Library preparation must be amenable to automation on liquid handling instruments.
- Library adapter strategy must enable pooling of up to 96 samples for sequencing on an Illumina® MiSeq® instrument.
- Library preparation chemistry must be robust in order to:
 - minimize the need for substantial sample QC,
 - minimize the need to normalize input sample quantities, and
 - eliminate the need to modify library construction parameters based on input sample quantity, quality, and length.
- Library preparation workflow must exhibit a high 'pass rate' of samples successfully constructed into libraries that proceed to sequencing.
- Libraries must yield high-quality sequencing metrics.

To address these properties, a KAPA HyperPrep method was automated on the Biomek® FX (Figure 2). The instrument run-time for processing 96 samples using this method is ~3.5 hr.

Sample submission guidelines and QC. Samples submitted to the MBCF should constitute user-generated dsDNA amplicons spanning region(s) of interest with lengths (225 – 275 bp) and input quantities (5 µL at 50 ng/µL) that support NGS library preparation (Table 1). Care should be taken during amplicon primer design to ensure that the relevant sequence information will be obtained from 2 X 150 bp paired-end reads.

The Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific) was used for the quantification of DNA prior to and after library preparation. Library fragment size distributions were analyzed with a TapeStation® D1000 ScreenTape Assay (Agilent Technologies). The KAPA Library Quantification Kit was used to quantify normalized libraries and library pools prior to sequencing.

Sequencing. Libraries were sequenced (2 x 150 bp) on a MiSeq, using V2 chemistry (Illumina). PhiX was used as a spike-in according to Illumina's recommendations.

Table 1: Sample submission guidelines for CRISPR indel screening service at MBCF

Sample submission guidelines	
Sample	dsDNA amplicon
Concentration	50 ng/µL
Length	225 – 275 bp

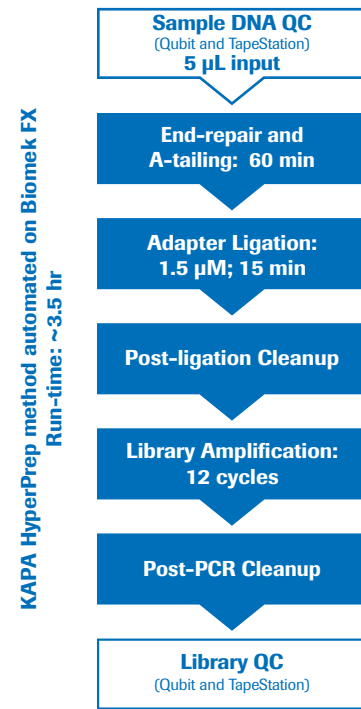


Figure 2: Schematic of automated library preparation workflow. Following input QC, samples were processed according to manufacturer recommendations on a Biomek FX instrument.

Results and discussion

Summarized in Table 2 are the input amounts and fragment lengths of 51 samples submitted to MBCF for CRISPR screening, as well as final library yields. Assessment of these samples indicated that while 28 of the 51 samples (~55%) fell within the submission guidelines for length, the remaining samples were shorter than was specified, though the range was tightly distributed ($\pm 11\%$ of the mean; Table 2, Figure 3A). When examining input concentration, only one sample was submitted as requested. The majority of samples were submitted at a much lower concentration than specified and varied widely, with a 50-fold difference between the lowest and highest concentration samples (Figure 3A). Regardless of this variability, an equivalent volume (5 µL) of each sample was input into library preparation. Ligation reactions were performed with the same stock concentration of adapter (1.5 µM), and the same number of PCR cycles (12) was applied across all samples.

Table 2: Summary of input characteristics and final library yields for 51 samples.

	Sample conc. (ng/µL)	Total input (ng)	Input size (bp)	Final library yield (nM)
Mean (\pm SD)	15.2 (\pm 7.2)	76.2 (\pm 35.9)	229.6 (\pm 26.4)	369.4 (\pm 96.8)
Min	1.1	5.5	176.0	247.7
Max	53.0	264.8	265.0	713.9

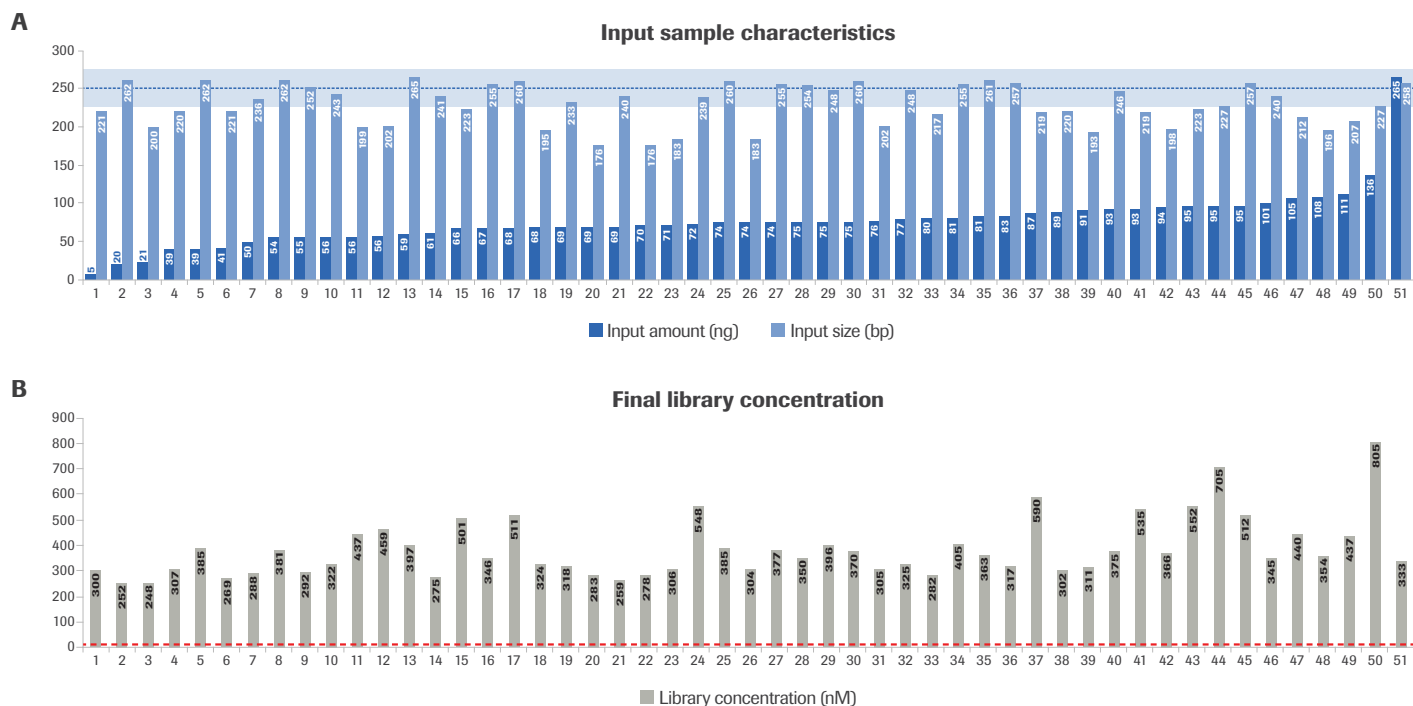


Figure 3: Successful library preparation across diverse input quantities and sizes. (A) QC assessment of input quantity, calculated from input concentrations (measured by Qubit) and input length (measured by TapeStation). Data for 51 samples is shown, sorted by input quantity. The blue rectangle across the top indicates the submission guideline for fragment length and the dashed blue line indicates the submission guideline for input quantity. **(B)** QC assessment of final amplified libraries measured by Qubit is plotted in the same order as in A. The dashed red line indicates the concentration threshold required to proceed to sequencing.

Evaluation of the final library concentrations indicated that all samples produced library yields that greatly exceeded requirements for sequencing, regardless of input amount and amplicon length. (Figure 3B). Surprisingly, a trend towards increased yields from higher input samples was not observed. This may be due to (1) the use of an adapter concentration optimized for the low-input samples but sub-optimal for the higher-input samples, and (2) the use of 12 PCR cycles, which was likely high enough that the amplification reactions from higher inputs reached saturation.

The fragment size distribution of final libraries was evaluated to confirm that libraries were appropriately sized and free of adapter-dimer contamination. None of the libraries examined exhibited any adapter-dimers and all libraries (100%) exhibited the appropriate fragment size (input size + ~130 bp added length from the dual index adapter; data not shown).

Representative TapeStation® traces from input samples of various qualities paired with their final library traces are shown in Figure 3. High-quality input samples are characterized as displaying a single amplicon without any contaminating non-specific products and/or primer dimers (Figure 4A, left panel). High and medium quality inputs exhibited high-quality final libraries with a single dominant peak ~130 bp larger than the input (Figure 4A – 4B, right panels). The low-quality input exhibited a number of potentially non-specific peaks containing molecules that were concomitantly ligated to adapters and carried through library preparation (Figure 4C). Despite the presence of these additional peaks, the dominant peak in the final library corresponded to the dominant peak in the input sample. Thus, it is expected that the high read coverage obtained for this sample will yield sufficient data for subsequent analysis.

Sequencing the 51 prepared libraries resulted in high-quality sequencing metrics (Table 3). Greater than 90% of clusters passed quality filters, and 93% of bases exhibited quality scores \geq Q30.

De-multiplexing of the data indicated that each sample obtained an average of >600,000 reads. This far exceeds the coverage required for sequence verification of individual cell clones (generally <100X). To increase cost efficiency, the MBCF currently offers multiplexing of up to 96 samples to obtain a minimum of 50,000 reads per sample. Note that higher sequencing coverage depths may be useful to quantify the editing efficiency in the mixed-cell population obtained prior to single-cell expansion. When sequencing libraries from lower-quality samples, high per-sample coverage such as that obtained here can compensate for reads mapped to non-specific products, ameliorating the need for extensive primer optimization and/or additional sample clean-ups prior to library preparation.

Table 3: Overview of sequencing parameters and summary of data quality

Sequencing overview and metrics	
Sample	Equimolar mix of libraries
Quantification	KAPA Library Quantification Kit
Spike-in	20% PhiX
Sequencing method	MiSeq (PE 2 X 150)
Cluster PF	92.8%
Reads \geq Q30	93.4%
Mean reads per sample	675,760

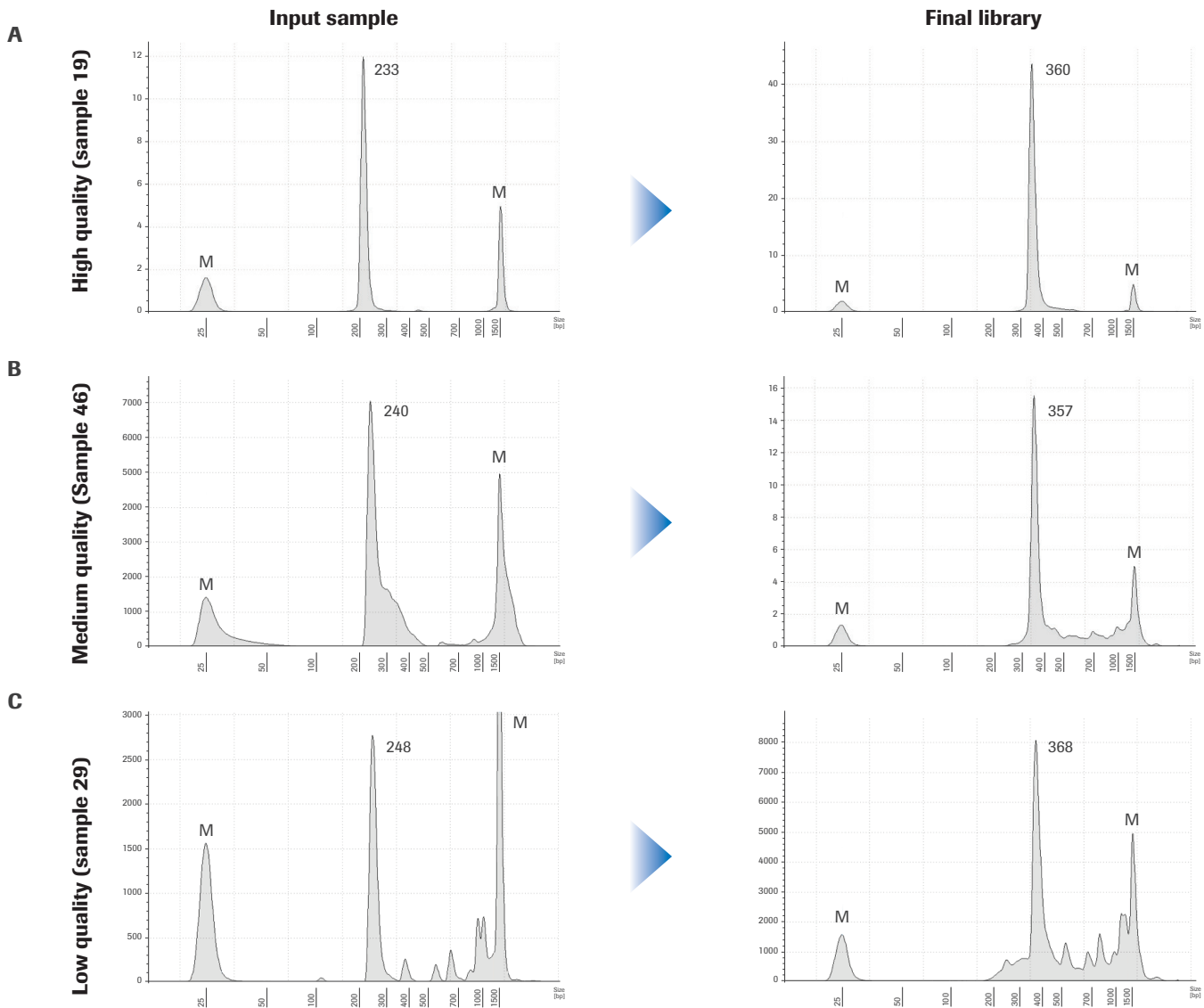


Figure 4: Successful library preparation across diverse input qualities. Tapestation traces are shown for representative input samples (left) and the resulting libraries (right). Samples range from high quality (top) to low quality (bottom). Sample numbers correspond to those in Figure 3. Upper and lower size markers are indicated by "M."

Summary

KAPA HyperPrep is a robust workflow that can accommodate a wide range of amplicon input amounts and sizes, thus eliminating the need for significant input QC and sample-specific workflow modifications. Availability of the HyperPrep method on a number of liquid handlers, including the Biomek FX, enables high-throughput CRISPR amplicon indel screening, available at the Dana-Farber Cancer Institute MBCF. Automated library preparation, provided with minimal QC and/or minimal sample-specific interventions, enables a rapid service turnaround time of 1 week from sample submission to data delivery.

Published by:

Roche Sequencing Solutions, Inc.
4300 Hacienda Drive
Pleasanton, CA 94588

sequencing.roche.com

References

1. Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J., Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. (August 2012).
2. Ran FA., Hsu PD., Wright J., Agarwala V., Scott DA., Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* (November 2013).