Evaluation of DNA sample fragmentation methods for input into library construction protocol

B.D. Craig, B.A. Marosy, K.F. Doheny

Center for Inherited Disease Research (CIDR), IGM, JHU-SOM, Baltimore, MD

Introduction

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. CIDR currently uses a mechanical method to fragment DNA with the Covaris E210 for input into library construction workflows. Samples undergo independent customizable fragmentation events in specialized sample tubes. When processing large sample numbers, serial fragmentation adds considerable time to the overall workflow (up to 10 hours for a 96 sample plate). To reduce workflow time and cost in the lab for library construction, we evaluated the KAPA enzymatic fragmentation kit that is controlled by varying time and temperatures within the reaction. This product does not require specialized equipment or consumables and all samples in a plate will be processed simultaneously with the same reaction conditions. We evaluated reproducibility of the fragmentation events, the impact of DNA amount on the profile and compared these profiles to current Covaris programs. Additionally, CIDR receives samples in a variety of EDTA concentrations which is an inhibitor of the KAPA reaction. In order to incorporate this product as a drop-in replacement to current CIDR best practice workflows, a bead cleanup must be performed to remove all EDTA. This extra cleanup resulted in an average of 45% sample loss prior to fragmentation. Depending on DNA sample number and input amount required, the additional cleanup step could require more starting DNA and offset time saved over the Covaris protocol. Samples were run on the Agilent Bioanalyzer to evaluate the size profile and optimize the reaction time to fit the current Covaris protocols. Size distributions were not affected by DNA input amounts for experiments with the same reaction time and were somewhat consistent for the replicated time and input experiments. Our results show fragmentation profiles larger than the expected mode outlined in the default protocol suggesting more optimization to the reaction time will be required to better match the profiles of our current Covaris protocols and integrate with existing methods. Here we describe experiments involving additional evaluation of the reproducibility of fragment size profiles generated under the same reaction conditions and the evaluation of sequencing data (insert size, library complexity, and target coverage) from libraries created with the KAPA fragmented samples.

Full plate fragmentation replication comparison

Methods

- 96 Hap Map samples were normalized to 1000 ng in plate format
- Removal of EDTA in the samples was performed with a 2X bead cleanup with elution in 35uL Qiagen Elution Buffer
  - Recovery estimated by Nanodrop following cleanup is ~55%
- KAPA Fragmentation master mix was made per the KAPA Frag kit protocol (*KR1141-v2.15)* and dispensed to the samples using a 12 channel pipette
- The plate was gently vortexed, spun and put on a PE 9700 Thermal cycler with the default cycling parameters for 20 minutes
  - Following incubation, KAPA stop reagent was dispensed and the plate was vortex and spun
- After fragmentation, a 2X bead cleanup was performed per the KAPA protocol
- Product was run on a Caliper GX high sensitivity QC chip (75-1000 bp smear range) and compared with results using the CIDR production Covaris 80 second protocol (Figures 1a-d)

Results

- One sample failed fragmentation completely in both the KAPA and Covaris experiments
- Purity is a measure of the percentage of sample that has callable peak concentration within a user supplied size range with respect to the overall concentration bound by the lower and upper markers
  - Samples with lower purity calculated have more un-fragmented sample outside of the user defined size range
    - Samples with < 80% Purity (Figure 1a)
      - KAPA: ***19 / 95 (20%)***
      - Covaris: ***1 / 95 (1%)***
  - Overlaid electropherogram traces shown for KAPA samples with and without efficient fragmentation and Covaris fragmented samples (Figures 1b, c and d respectively)
  - Samples with inefficient fragmentation would need to be redone prior to starting an end repair task
- The range of the mean insert sizes (Figure 1e) shows more variability for the KAPA vs. Covaris samples (234 vs. 144 bp)

Figure 1a: Purity of the fragmentation reactions (x) vs. Mean size in bp according to the Caliper GX High Sensitivity Assay (y). Samples fragmented with KAPA shown in red (left trellis) and Covaris in blue (right trellis).

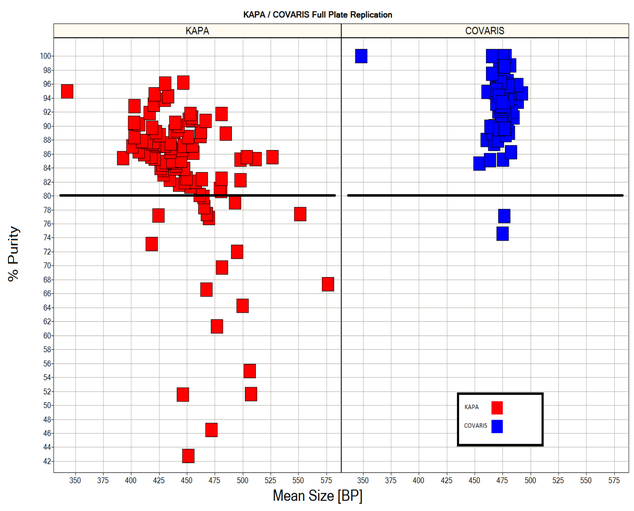
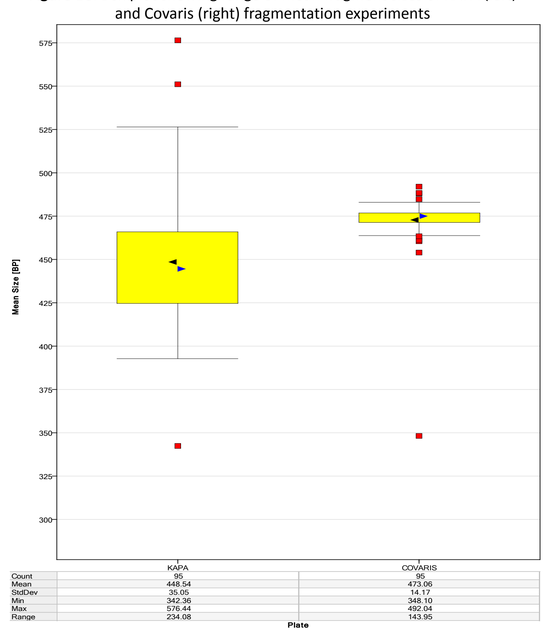


Figure 1e: Box plot showing range of mean fragment sizes of KAPA (left) and Covaris (right) fragmentation experiments



	KAPA	COVARIS
Count	95	95
Mean	448.54	473.06
StdDev	35.05	14.17
Min	342.36	348.10
Max	576.44	492.04
Range	234.08	143.95

Figures 1 b-d:
Caliper GX electropherogram trace overlay of KAPA samples with inefficient fragmentation (1b), KAPA samples with reproducible fragmentation (1c) and default Covaris samples profiles (1d). Red lines annotate the expected fragmentation range outlined by the default CIDR protocol

Figure 1b

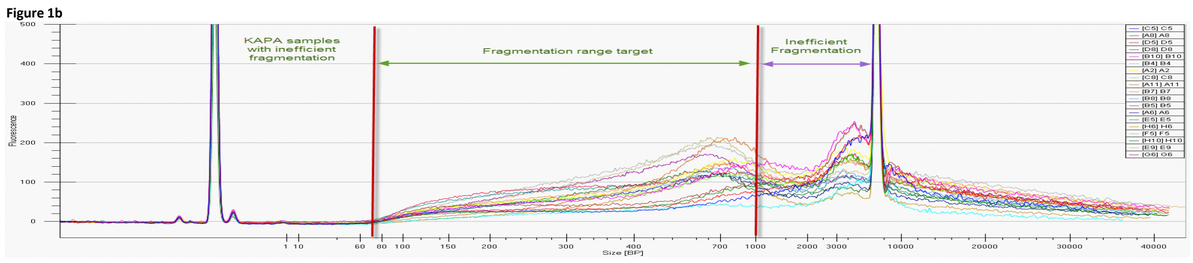


Figure 1c

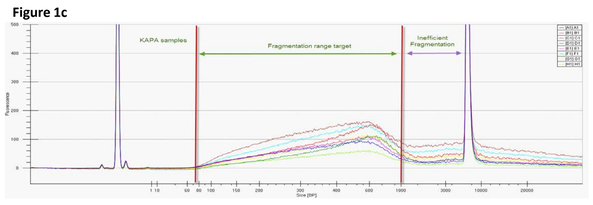
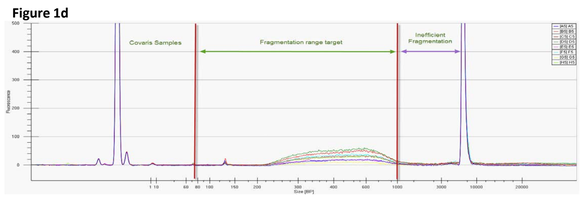


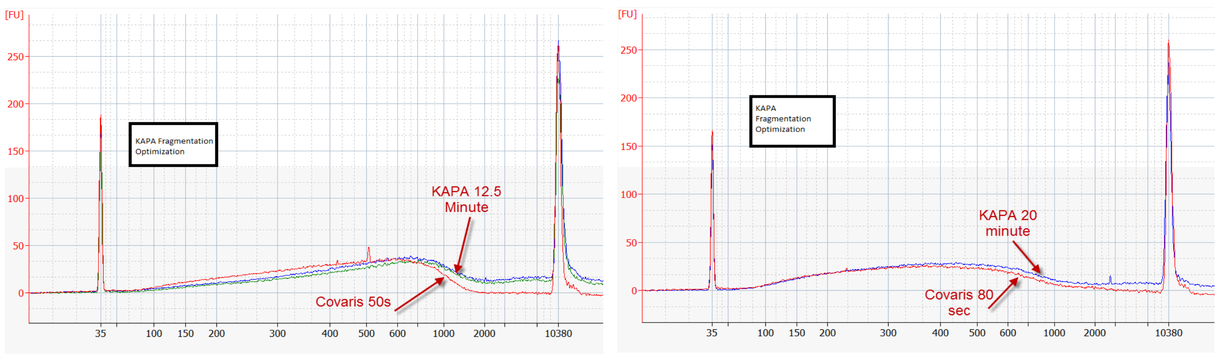
Figure 1d



Optimization of KAPA fragmentation parameters

- Initial experiments used the default fragmentation parameters outlined in the KAPA protocol
  - Fragmentation temperature 37 °C and incubation times of 5, 10, 20 and 30 minutes
  - Based on the fragmentation profiles from the default incubation parameters, time was adjusted to match profiles from the CIDR production Covaris programs
  - 12.5 and 20 minute incubation times most closely matched the 50 and 80 second Covaris profiles respectively (Figure 2)
  - These incubation times were selected for continuing with library construction, whole exome capture and sequencing comparisons to CIDR default Covaris programs

Figure 2: Fragmentation profiles optimized to fit current CIDR production Covaris programs. Covaris 50 sec vs. KAPA 12.5 min (left). Covaris 80 sec vs. KAPA 20 min (right)

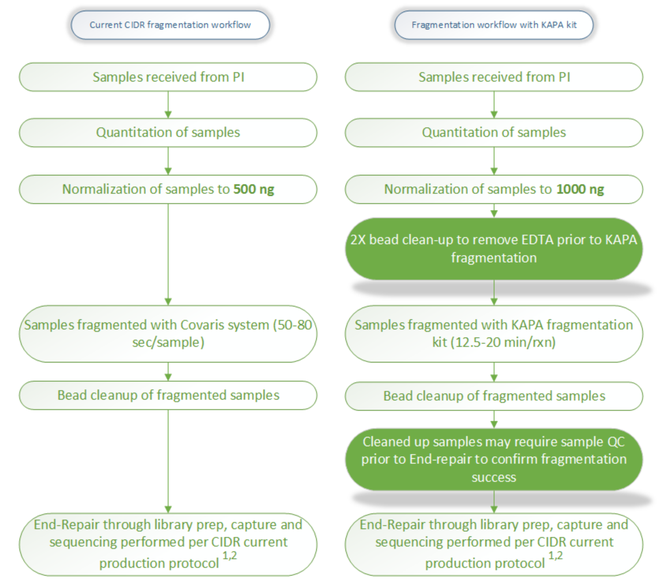


Fragmentation method sequencing QC comparison

Methods

- Control NA12891 was selected for library construction, whole exome capture and sequencing (Table 1)
  - Covaris fragmentation aliquots were normalized to CIDR default input requirements (500 ng)
    - One aliquot was sheared under both current Covaris programs (50 and 80 sec)[1,2]
  - KAPA fragmentation aliquots were normalized to 1000 ng to adjust for (~50%) sample loss due to the cleanup to remove EDTA (Figure 3)
    - One aliquot each received 12.5 or 20 min incubation time to target the shear profile of the Covaris 50 or 80 sec program respectively
- Fragmentation was performed per the KAPA Frag Kit *(KR1141-v2.15)* or CIDR Covaris protocol[1,2]
- Following fragmentation all samples were processed with current CIDR production protocols (Figure 3)
  - Agilent SureSelect™ XT library and whole exome capture using the V5 + UTR bait product
  - Clustering and sequencing performed on 2 lanes of an Illumina HiSeq™ 2500
- Analysis performed using the CIDRSeqSuite pipeline (http://www.cidr.jhmi.edu/next_gen_seq_serv/sequencing_qc_pipelines_workflow.pdf)
  - Down sampling of data was performed using Picard's DownsampleSam (http://broadinstitute.github.io/picard/) to the experiment with the lowest raw data yield to ensure equal comparison between methods

Figure 3: Current CIDR workflow for fragmentation of samples (left) and proposed modifications to replace Covaris with KAPA



Results

- Sequencing QC metrics comparing results for the standard Covaris programs and the potential KAPA counterparts are shown in Table 1
  - Increased yield and decreased % duplication in the KAPA samples are a result of the increased input DNA (more than 500 ng recovered after the pre-fragmentation cleanup
  - The mean insert size of the KAPA 12.5 minute sample is ~30 bp larger than the 50 sec Covaris equivalent suggesting additional modification to the incubation time may be needed

Table 1: Sequencing QC metrics comparing results for the standard Covaris programs (Left columns) and the potential KAPA counterparts (Right columns) . Color denotes fragmentation targets to be compared.

Results	Standard Covaris		KAPA Fragmentation	
Fragmentation Time	50 seconds	80 seconds	12.5 minutes	20 minutes
DNA Input [ng]	500	500	1000	1000
Yield at PCR Amp [ng]	1647	1316	2688	2659
% Selection	74	74	72	73
% Total Concordance	99.7	99.8	99.8	99.8
% Sensitivity (Het)	96.0	96.0	95.3	96.0
TiTv All Ratio	3.03	3.05	3.04	3.05
Count SNV On Target	49786	49750	49415	49558
Count SNV On Bait	90927	90935	89880	90444
% SNV On Bait in DBSNP138	99.1	99.1	99.1	99.0
Raw GB	6.3	6.3	6.3	6.3
Mean Target Coverage	38	39	36	39
% On Target @ 10X	94.3	94.2	93.4	93.9
% Zero Bases	0.66	0.72	0.73	0.77
Mean Insert Size	360	346	389	337
% Duplication	6.6	5.6	4.7	3.4

Discussion

- The KAPA Fragmentation Kit as a drop-in replacement for mechanical fragmentation methods does not appear to influence sequencing QC results (Table 1)
- Reproducibility of the KAPA fragmentation reaction is not as consistent within experiments as Covaris (Figure 1d)
  - Inefficient fragmentation and variable profiles were seen indicating additional troubleshooting and QC tasks (Figure 3) may be required
  - Shown here 20% of the full KAPA plate would need to be redone at the fragmentation task (Figure 1b) compared to 1% of the Covaris plate
- The presence of EDTA inhibiting the KAPA reaction requires an additional bead based cleanup task to be added before fragmentation
  - The automation and execution of the additional task negates most time and input DNA savings over the use of the Covaris
- The Covaris offers the flexibility to define different fragmentation profiles for samples of different qualities (i.e., partially degraded) on the same plate
  - Switching to an enzymatic fragmentation could require time consuming sample movements to avoid over fragmentation of partially degraded samples in an experiment
- At present, there is little cost benefit for high throughput sample processing with existing Covaris hardware
  - Sample processing estimates, previous equipment expenditures and consumable costs are projected to be lower than KAPA reagent and extra bead cleanup reagents required for this enzymatic fragmentation

References:
1-SureSelectXT Target Enrichment System for Illumina Paired-End Sequencing Library; SureSelectXT Target Enrichment for Illumina Multiplexed Sequencing Protocol v1.1.1, January 2011 p/n G7530-90000
2-Fisher *et al*; A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biology* 2011, 12:R1