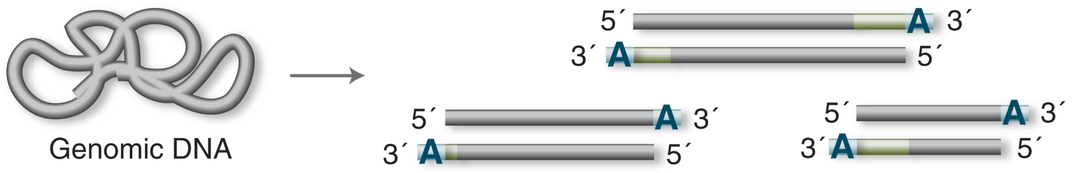
A Robust, Streamlined, Enzyme-based DNA Library Preparation Method Amenable to a Wide Range of DNA Inputs Style

Lynne Apone, Karen McKay, Vaishnavi Panchapakesa, Bradley Langhorst, Timur Shtatland, Chen Song, Pingfang Liu, Christine Sumner, Christine Rozzi, Melissa Arn, Daniela Munafo, Fiona Stewart, Eileen Dimalanta and Theodore Davis

New England Biolabs, Inc.
240 County Road, Ipswich, MA 01938, USA.

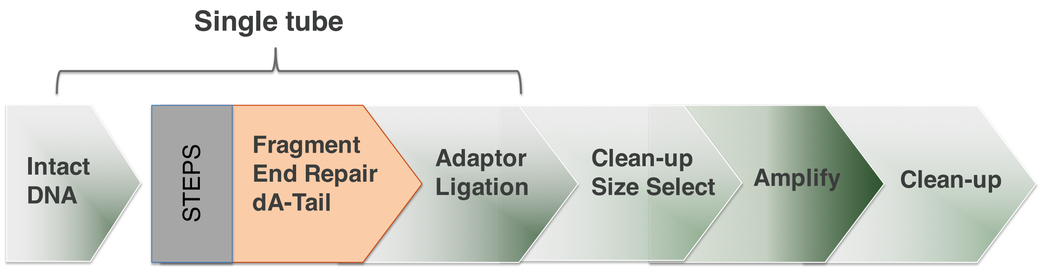
INTRODUCTION

In order for Precision Medicine to be widely implemented, robust, accurate, methods to process and interrogate large numbers of human samples are required. While advances in Next Generation sequencing (NGS) technologies have positioned it at the forefront of the Precision Medicine Initiative, it is ideally suited for deep molecular analysis of a single sample, rather than the processing and interrogation of thousands. In order to overcome this limitation, we have developed a robust, streamlined library construction method that integrates enzyme-based DNA fragmentation with end repair and dA-tailing. This method utilizes a single protocol for a wide range of DNA input amounts and types. In addition, it eliminates the need for expensive equipment to fragment DNA and numerous cleanup and liquid transfer steps, reducing time, cost and errors associated with library construction



METHODS

ENZYME BASED FRAGMENTATION COMBINED WITH NEBNEXT® ULTRA™ II DNA LIBRARY PREPARATION

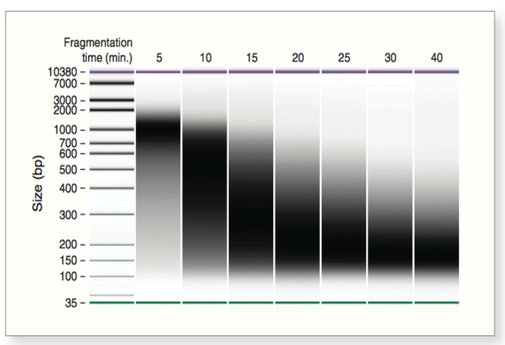


Genomic DNA isolated from a variety of sources was used to construct Illumina libraries. DNA (100pg - 500ng) was fragmented, end repaired and dA-tailed in a single step followed by adaptor ligation in the same tube. PCR amplified libraries were sequenced, reads aligned to the appropriate reference genome, and quality metrics generated using Picard tools.

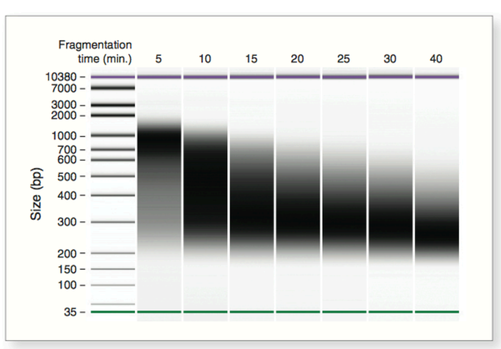
RESULTS

ULTRA II FS ENABLES TIME_DEPENDENT FRAGMENTATION

A



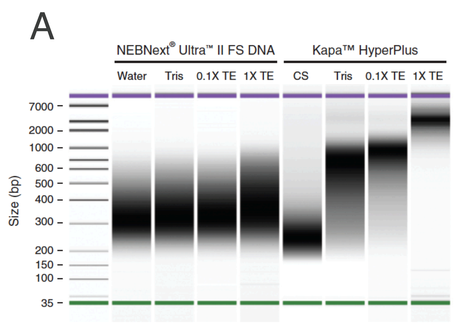
B



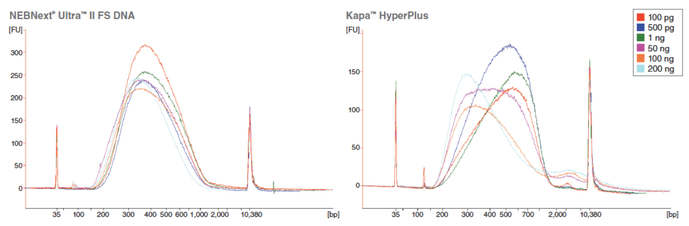
A) 100 ng of Human NA19240 genomic DNA was incubated with the NEBNext Ultra II FS Enzyme Mix and Reaction Buffer for 5, 10, 15, 20, 25, 30 and 40 minutes at 37°C, followed by 65°C for 30 minutes. After clean-up using NEBNext Sample Purification Beads, size was assessed using the Agilent Bioanalyzer. B) Libraries were constructed using the NEBNext Ultra II FS kit and 100 ng of Human NA19240 genomic DNA, with fragmentation times of 5, 10, 15, 20, 25, 30 and 40 minutes, and 4 PCR cycles. Size selection was not performed. After clean-up using NEBNext Sample Purification Beads, size was assessed using the Agilent Bioanalyzer. NEBNext Ultra II FS shows expected final library sizes consistent with the fragmentation sizes seen in A).

ROBUST FRAGMENTATION INDEPENDENT OF DNA AND BUFFER

A

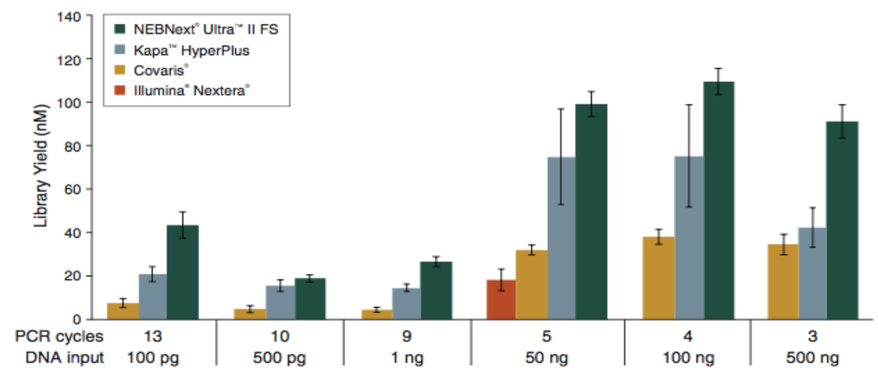


B



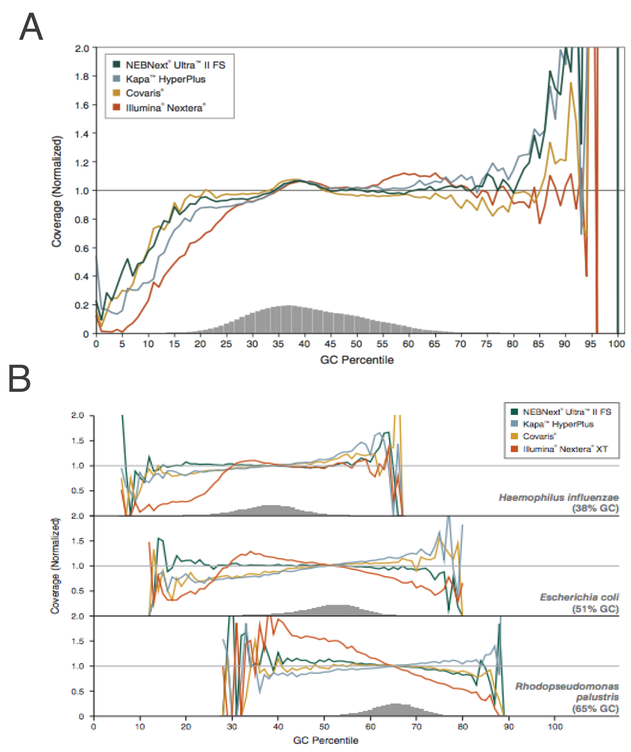
A) Libraries were made using 100 ng of human genomic DNA and the NEBNext® Ultra™ II FS kit or the Kapa™ HyperPlus Kit. Libraries were fragmented to generate 200 bp inserts (320 bp libraries) from DNA resuspended in H_2O, 10mM Tris, 0.1X TE or 1X TE . Kapa™ HyperPlus libraries were made using the recommended dilution of the supplied Conditioning Solution (CS), or using DNA in 10mM Tris, 0.1X TE or 1X TE, in the absence of either CS or 3X bead clean up. Library size distribution was assessed using the Agilent® Bioanalyzer®. B) Libraries were prepared from human genomic DNA using the input amounts shown. NEBNext® Ultra™ II FS libraries were prepared using a 20-min fragmentation time. For Kapa™ HyperPlus, the recommended bead clean-up was performed prior to the 20 min fragmentation. Library size was assessed using the Agilent® Bioanalyzer®. Low input (1 ng and below) libraries were loaded on the Bioanalyzer® without dilution. High input libraries were loaded with a 1:5 dilution in 0.1X TE.

ULTRA II FS PRODUCES THE HIGHEST YIELD LIBRARIES



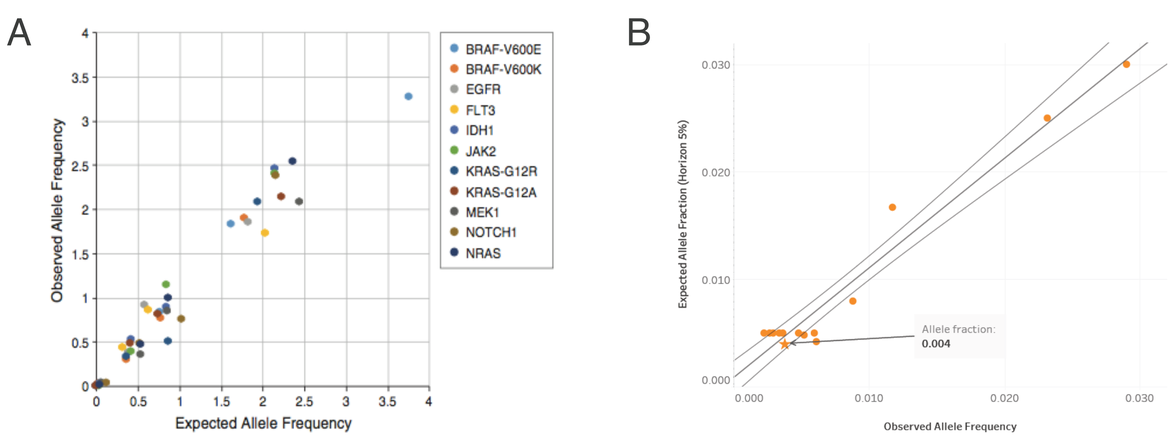
Libraries were prepared from Human NA19240 genomic DNA using the input amounts and numbers of PCR cycles shown. For NEBNext Ultra II FS, a 20-minute fragmentation time was used. For Kapa™ HyperPlus libraries, input DNA was cleaned up with 3X beads prior to library construction, as recommended, and a 20-minute fragmentation time was used. Illumina® recommends 50 ng input for Nextera, and not an input range; therefore, only 50 ng was used in this experiment. “Covaris®” libraries were prepared by shearing each input amount in 1X TE Buffer to an insert size of ~200 bp using a Covaris instrument, followed by library construction using the NEBNext Ultra II DNA Library Prep Kit (NEB #E7645). Error bars indicate standard deviation for an average of 3–6 replicates performed by two independent users.

SUPERIOR GC COVERAGE WITH HUMAN AND MICROBIAL GENOMES



A) Libraries generated using the indicated kits, 50ng of Human gDNA (NA19240), and 5 cycles of PCR were sequenced on an Illumina MiSeq® (2 x 76 bp). B) Libraries prepared from 1 ng of a genomic DNA mix containing Haemophilus influenzae, Escherichia coli (K-12 MG1655), and Rhodopseudomonas palustris were amplified for 9 cycles of PCR and sequenced on an Illumina MiSeq®. Reads generated in A and B were mapped using Bowtie 2.2.4 and GC coverage information was calculated using Picard’s CollectGCBiasMetrics (v1.117). Expected normalized coverage of 1.0 is indicated by the horizontal grey line, the number of 100bp regions at each GC% is indicated by the vertical grey bars, and the colored lines represent the normalized coverage for each library.

ULTRA II FS ENABLES ACCURATE VARIANT DETECTION



Tru-Q 1 (5% Tier) Reference Standard (Horizon® HD728) was spiked into wild type human DNA (NA19240) at varying expected allele frequencies. 100 ng DNA mix was used to generate NEBNext® Ultra™ II FS libraries, followed by target enrichment using the ClearSeq® Comprehensive Cancer XT panel (Agilent® #5190-8011). A) Droplet digital PCR (Bio-Rad® QX200) detected 11 of the known Tru-Q 1 mutations in the input DNA sample as well as the UII FS libraries at expected frequencies. B) Comparison of observed variant allele frequency (VAF) detected from sequencing data with the expected VAF from the 0.5% EAF. Libraries were sequenced on a MiSeq (2x75 bp) to an average depth of 2000x coverage. Reads were aligned with BWA-MEM (Li and Durbin, 2010) and duplicates were marked either by alignment coordinates alone or alignment coordinates and UMIs. Variants were detected with Scalpel (Narzisi et al., 2014), Mutect (Cibulskis et al., 2013), and Vardict (Lai et al., 2016) and confirmed by hand for low allele frequencies.

CONCLUSIONS

- Ultra™ II FS enables high-quality library construction from a broad range of DNA quantities and qualities
- Sequencing metrics of FS libraries are similar or superior to those produced with mechanically sheared DNA
- Ultra™ II FS maintains mutation abundance in samples without introducing bias
- These advances will allow greater use and adoption of NGS technologies in clinical and diagnostic settings

REFERENCES

Li, H., and Durbin, R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26:589–595.

Narzisi, G., et al. 2014. Accurate de novo and transmitted indel detection in exome-capture data using microassembly. *Nature Methods* 11:1033–1036.

Cibulskis, et al. 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology* 31:213–219.

Lai, Z., et al 2016. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Research*:gkw227.