# Nextera™ Technology for NGS DNA Library Preparation: Simultaneous Fragmentation and Tagging by *In Vitro* Transposition

*Nick Caruccio, Haiying Grunenwald, and Fraz Syed, EPICENTRE Biotechnologies*

## Introduction

Over the past 5 years, the need to lower the cost of sequencing a human genome has led to development of many massively parallel sequencing platforms. The field of massively parallel sequencing, more commonly known as next-generation sequencing (NGS), has advanced rapidly and has made it possible to analyze the genome at previously unattainable depth. Some of the approaches for NGS have been commercialized and are now well-established. These NGS platforms employ slightly different technologies for sequencing, such as sequencing by ligation (Complete Genomics, Life Technologies, Polonator), sequencing by synthesis (Helicos, Illumina, Intelligent Bio-Sys), real-time sequencing by synthesis (Pacific BioSciences, VisiGen), and pyrosequencing (Roche/454™). Although different in their sequencing chemistries, most platforms adhere to a common library preparation procedure, with minor modifications, prior to a "run" on the instrument. This includes fragmenting the DNA (sonication, nebulization, or shearing), followed by DNA repair and end-polishing (blunt-end or A-overhang), and finally, platform-specific adaptor ligation. The library preparation procedure is one of the major bottlenecks for NGS, and typically results in significant sample loss with limited throughput due to the number of steps involved in making a library. To streamline the workflow, increase throughput, and reduce sample loss, we have developed Nextera™ technology*, a transposon-based method for preparing fragmented and tagged DNA libraries from as little as 50 ng of DNA. This flexible, scalable, and efficient technique can generate sequencer-ready libraries for multiple platforms in less than 2 hours.

**Covered by issued and/or pending patents.*

## Methods and Results

### *Overview*

Nextera technology employs *in vitro* transposition to prepare sequencer-ready libraries. In a modification of the classic transposition reaction, we used free transposon ends and a transposase to form a Transposome™ complex. When this complex is incubated with target double-stranded DNA (dsDNA), the target is fragmented and the transferred strand of the transposon end oligonucleotide is covalently attached to the 5′ end of the target fragment (Fig. 1). By varying the concentration of Transposome complexes, the size distribution of the fragmented and tagged DNA library can be controlled (Fig. 2).

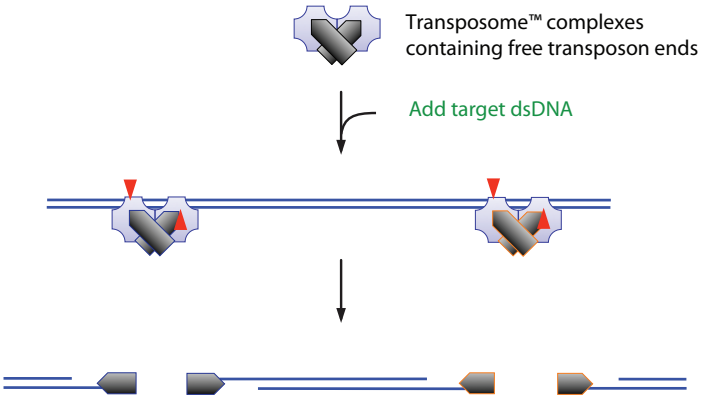



**Figure 1. Overview of Nextera™ fragmentation and tagging technology.** Transposon integration and strand transfer occur via a staggered, dsDNA break within the target DNA (red triangles). When free transposon ends are used in the insertion reaction, the target DNA is cleaved and tagged at the 5′ end with the transposon sequence. The resulting fragments have single-stranded gaps.

### *Creating tagged libraries*

Nextera technology can be used to generate di-tagged libraries, with optional barcoding, compatible with Roche/454 or Illumina/Solexa® platforms. To generate platform-specific libraries, either free transposon ends (Fig. 3A) or appended transposon ends (Fig. 3B) can be used. Platform-specific tags (and optional bar coding) are introduced by 10 cycles of PCR. The sequencing adaptors enable amplification by emulsion PCR (emPCR), bridge PCR (bPCR), and other methods. The amplified library can be subsequently sequenced using the appropriate primers.

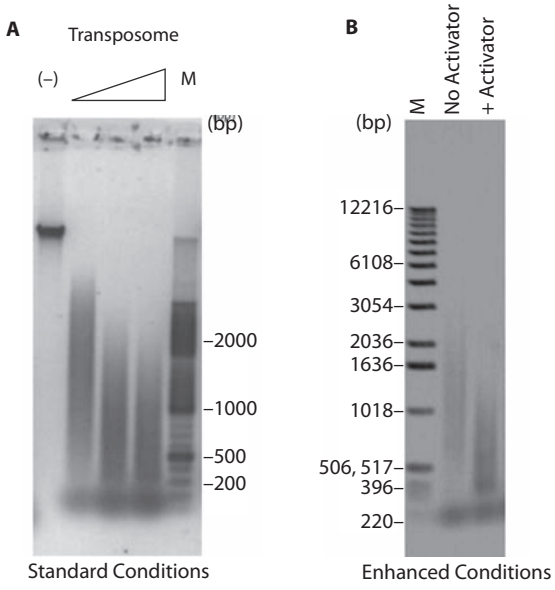



**Figure 2. Dose-dependent 5′-end-tagging and fragmentation of HeLa genomic DNA by Transposome™ complexes containing free transposon ends.** A) One microgram of genomic DNA was incubated with increasing amounts of Transposome complexes in 33 mM Tris-acetate (pH 7.6), 66 mM KOAc, 10 mM $Mg(OAc)_2$ for 2 hours. Reaction products were resolved by electrophoresis in a 1% agarose gel and staining with SYBR® Gold. B) By using enhanced reaction conditions, the reaction time in Fig. 2A was reduced to 5 minutes. Lanes M, DNA markers.

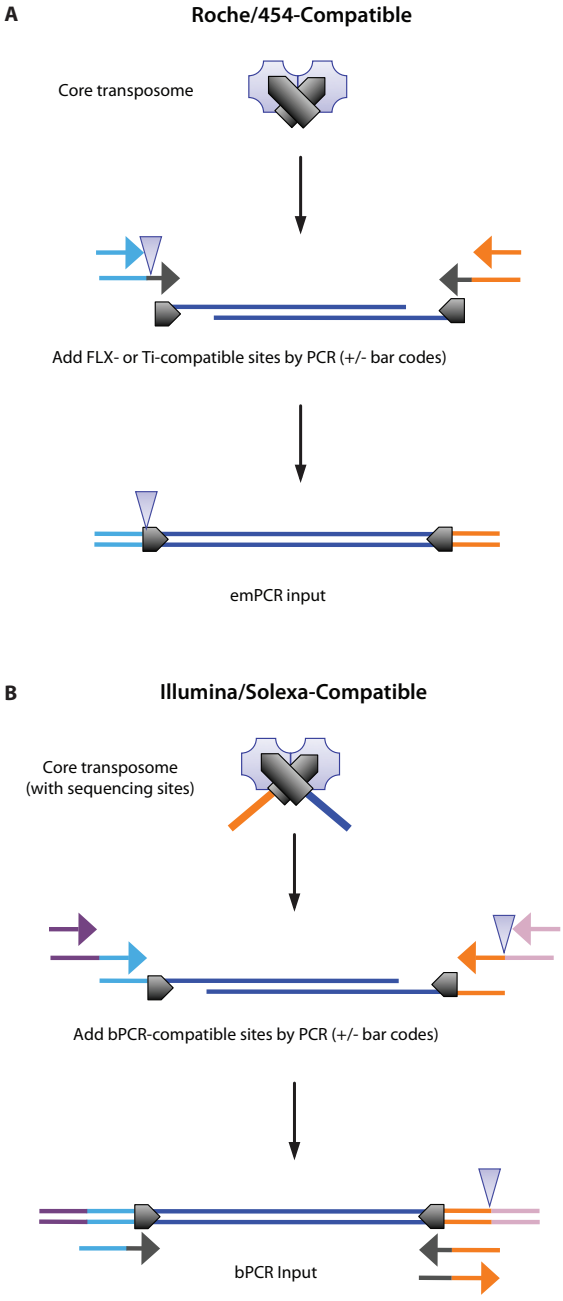



**Figure 3. Generation of platform-specific sequencing libraries.** A) Target DNA is fragmented and tagged with the core Transposomes™ containing free transposon ends. Suppression PCR with a four-primer reaction adds Roche/454-compatible adaptors sequences (blue and orange). Optional bar coding (triangle) is added between the upstream emPCR adaptor (blue) and the transposon end (gray). B) Target DNA is fragmented and tagged with the core Transposomes containing ends appended with sequencing primer sites (blue and orange). Suppression PCR with a four-primer reaction adds bPCR-compatible adaptors (purple and pink) to the core sequencing library. Optional bar coding (triangle) is added between the downstream bPCR adaptor (pink) and the core sequencing library adaptor (orange). Alternative sequencing primers are required for the Illumina/Solexa-compatible libraries: Read 1 (blue/gray arrow); Read 2 (orange/gray arrow); Index Read (gray/orange arrow).

### *Enrichment of di-tagged fragment libraries*

We used bacteriophage T7 genomic DNA to demonstrate di-tagging and enrichment of fragments containing both tags (Fig. 4). We modified transposon ends to contain Roche/454-compatible tags (A and B). Following the tagging and fragmenting reaction, the transposase was heat-inactivated, and 10 cycles of suppression PCR was performed. A 1:100 diluted sample was analyzed by quantitative PCR (qPCR). A mixed population of tagged DNA fragments was obtained, containing the desired A-B-tagged DNA, as well as DNA fragments with either A or B tags at both ends. qPCR showed enrichment of the A-B-tagged fragments in the amplified sample based on their lower $C_T$ values (Fig. 4).

### *Nebulization vs.* in vitro *transposition*

We prepared a di-tagged and enriched Nextera library from a 43-kb cosmid clone. We performed deep sequencing of the Nextera library and a control library (prepared by nebulization and the manufacturer's recommended protocol). Contig assembly, coverage, and accuracy of the Nextera library were comparable to the control library produced using nebulization (Fig. 5A and 5B). Neither nebulization nor Nextera libraries appeared to have significant GC bias with respect to depth of coverage (Fig. 5C).

## Conclusions

Current library preparation methods for NGS are time-consuming and prone to significant sample loss. Even prior to library preparation, the recovered DNA must be purified and end-polished. Nextera technology offers many advantages over current library preparation methods, such as a streamlined workflow that can result in significant time- and cost-savings. The procedure combines fragmentation, repair, and ligation steps. The method is scalable and requires as little as 50 ng of starting DNA, compared to 5-10 µg for current procedures. As described here, Nextera technology adapts *in vitro* transposition by using optimized transposases and incorporating engineered free transposon ends. By further manipulation, libraries containing complementary or independent adaptor sequences can also be efficiently constructed and amplified prior to sequencing on most NGS platforms.

## Acknowledgements

Sequencing data generated in collaboration with Hilary Morrison, Jessica Mark Welch, Ekaterina Andreishcheva, and William Reznikoff, Marine Biological Laboratory, Woods Hole, MA.

For more information or to learn more about products based on Nextera technology, visit: www.EpiBio.com/nextera

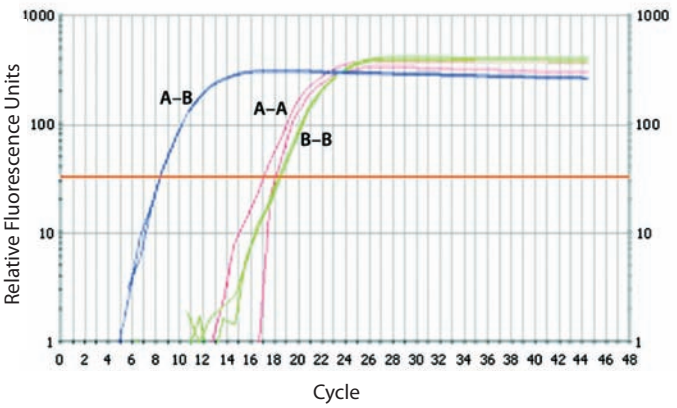
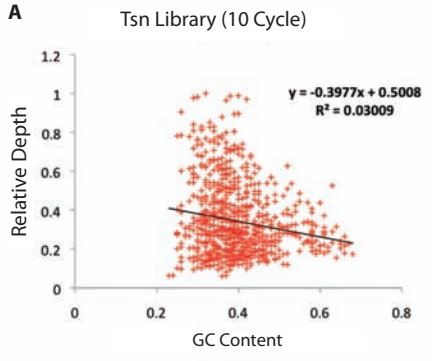
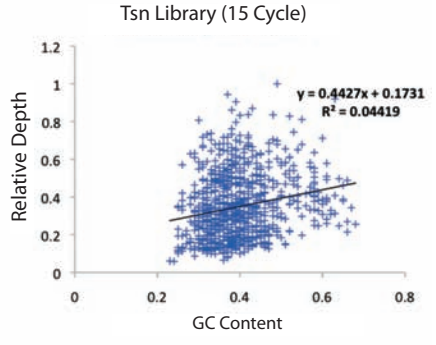



**Figure 4. Enrichment of A-B-tagged DNA fragments.** Transposon ends were modified to contain two unique Roche/454-compatible tags (A and B) and a library was prepared following the scheme in Fig. 3A. qPCR was performed using A and B primers for 45 cycles.

A




B

| | Tsn Library (10 cycle) | Tsn Library (15 cycle) |
|---|---|---|
| Number of Reads | 67823 | 63594 |
| Number of Bases | 9.51 x $10^6$ | 7.73 x $10^6$ |
| % Error | 0.51% | 0.49% |
| Average Read Length | 140 | 122 |
| Proportion Q40 Plus Bases | 0.999 | 0.999 |
| Average Coverage | 205 | 161 |

C

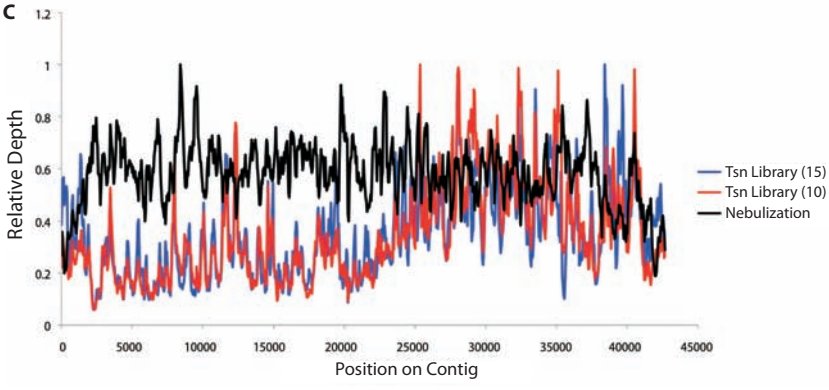



**Figure 5. Deep sequencing of a Transposome™-generated sequencing library.** A 43-kb cosmid DNA was fragmented and tagged with Transposomes containing transposon appended with Roche/454-compatible adaptor sequences. A+B di-tagged fragments were amplified and enriched by suppression PCR (10 or 15 cycles) using primers complementary to the sequencing adaptors. The recovered DNA fragments were used directly as input for Roche/454 FLX sequencing. A) Relative depth versus GC content. B) Deep sequencing assembly metrics for pyrosequencing dataset. Q40 is a quality score denoting the probability of a wrong base call at 1 in 10,000. C) Coverage plot indicating sequencing depth versus contig position.