DNA fragmentation strategies for next-generation sequencing library preparation

Summary

- Two methods are employed for the fragmentation of DNA prior to preparation of libraries for next-generation sequencing on Illumina™ systems: physical shearing and enzymatic shearing
- Physical shearing is the most widely adopted method as it provides superior coverage of challenging genomic regions and consistent insert size
- Enzymatic shearing methods are highly scalable and require lower input amounts for PCR-free library generation

Introduction

Next-generation sequencing (NGS) has become ubiquitous in the investigation of disease at the molecular level. The technology involves millions of nucleic acid strands being read in parallel, one base at a time. Over the past 10 years, DNA sequencing systems have evolved from instruments with a throughput of several megabases per day to instruments with a throughput of terabases per day. This expansion has required development of technologies to improve the speed, ease of use, and scalability of methods for the production of NGS libraries.

Preparing nucleic acids for NGS systems involves a multistep library construction process (Figure 1). In the general workflow, the nucleic acid of interest is harvested, purified, fragmented, end-repaired, and A-tailed. Adapters are then ligated, and the libraries are cleaned, quantitated, normalized, and loaded onto a flow cell.

Two common methods are used to fragment DNA: physical shearing and enzymatic shearing. The choice of method depends on several factors, including the amount of nucleic acid available, the equipment available, and any budgetary restrictions.

Physical shearing uses shear force to break the covalent bonds that connect DNA together. When both strands are broken, the DNA is fragmented into smaller pieces. Methods that create shear force include sonication, nebulization, and focused acoustic shearing [1,2]. Sonication generates ultrasonic waves that disrupt the covalent bonds; however, this method requires large amounts of sample and can result in loss of nucleic acid, which makes it unpopular for fragmentation. Nebulization creates shear force by using compressed gas to force a nucleic acid solution through a small hole, where the nucleic acid is collected after being sheared to a homogeneous size. This method can result in significant loss of starting material and is no longer commonly used.

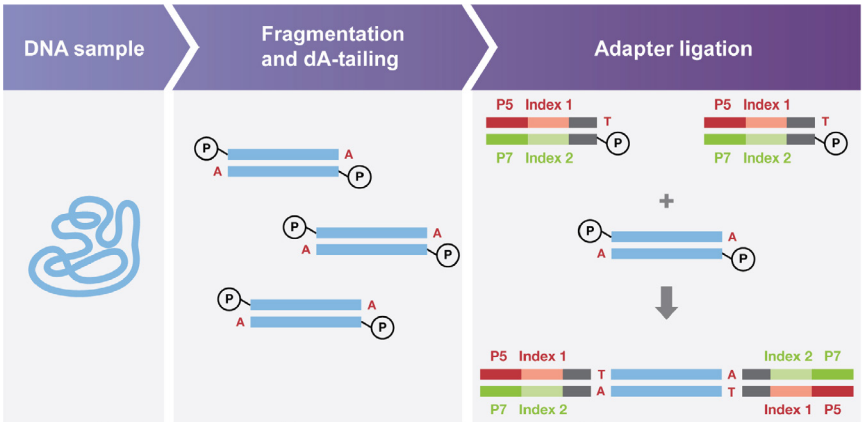

Figure 1. Overview of the library preparation process.

ThermoFisher SCIENTIFIC

The use of Adaptive Focused Acoustics™ (AFA) shearing, which was commercialized by Covaris, results in bubbles that collapse in upon themselves, thereby creating localized forces that break up the nucleic acid. AFA shearing has become the preferred physical shearing method because of its minimal sample loss, the limited chance of cross-contamination, and the homogeneous sizes of the fragments it produces. However, a possible barrier to adoption of AFA technology is the need to purchase a dedicated instrument.

Enzymatic shearing uses nucleic acid endonucleases or endonuclease cocktails to simultaneously break both strands of dsDNA or to generate nicks on each strand of dsDNA, leading to fragmentation. Endonuclease cocktails are used to overcome the inherent specificity of enzymes.

Results

Compared with enzymatic shearing, physical shearing is more tolerant of a broad range of input material; thus, quantitation errors with physical shearing cause less variation in fragment size. Physical shearing is also a less-biased fragmentation method than enzymatic shearing because of the randomness inherent to the mechanical process (Figure 2). Physical shearing results in less sample-to-sample variation because of the aforementioned factors, and is recommended for DNA from FFPE tissue. Since this DNA is already damaged, physical shearing minimizes additional complexity in genomic interpretation. A drawback to physical shearing is that it requires dedicated equipment that can be expensive, depending on the number of libraries being prepared or equipment being used.

Enzymatic shearing uses standard lab equipment; thus, no additional upfront equipment costs are associated with this method. However, enzymatic shearing can cause fragmentation bias because of the underlying endonuclease mixes. Furthermore, this method can also be sensitive to nucleic acid input amounts because of the relative abundances of enzymes and substrates, sometimes resulting in skewed representation in AT or GC coverage (Figure 3). The use of enzymatic fragmentation can cause sample-to-sample variation for numerous reasons, which can sometimes be difficult to discern.

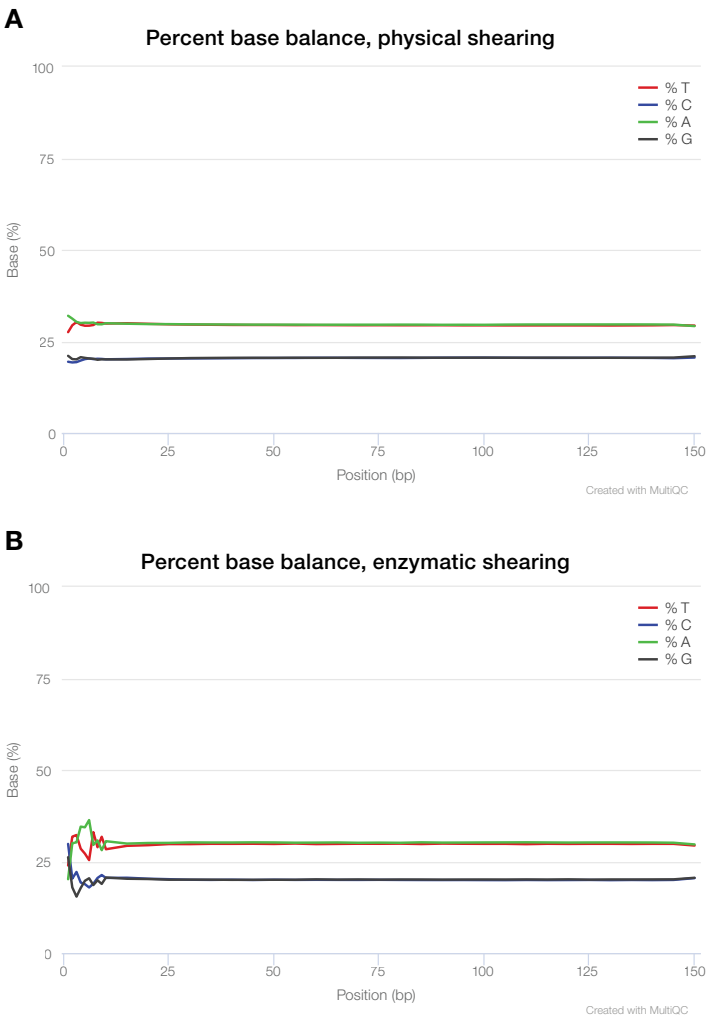

Figure 2. Percent of each base at each position during a sequencing run on a NovaSeq™ 6000 system. Libraries were prepared using Invitrogen™ Collibri™ DNA library preparation kits. **(A)** Physical shearing shows very little bias in base representation at the beginning of reads. **(B)** Enzymatic shearing shows minor base imbalance at the beginning of reads.

Conclusions

- AFA shearing results in more consistent, less biased libraries that are tolerant of broader input ranges compared to libraries created with other fragmentation methods
- Enzymatic shearing can be accomplished with standard laboratory equipment
- Library preparation with the Collibri PS DNA Library Prep Kit for Illumina Systems is recommended for use with FFPE samples

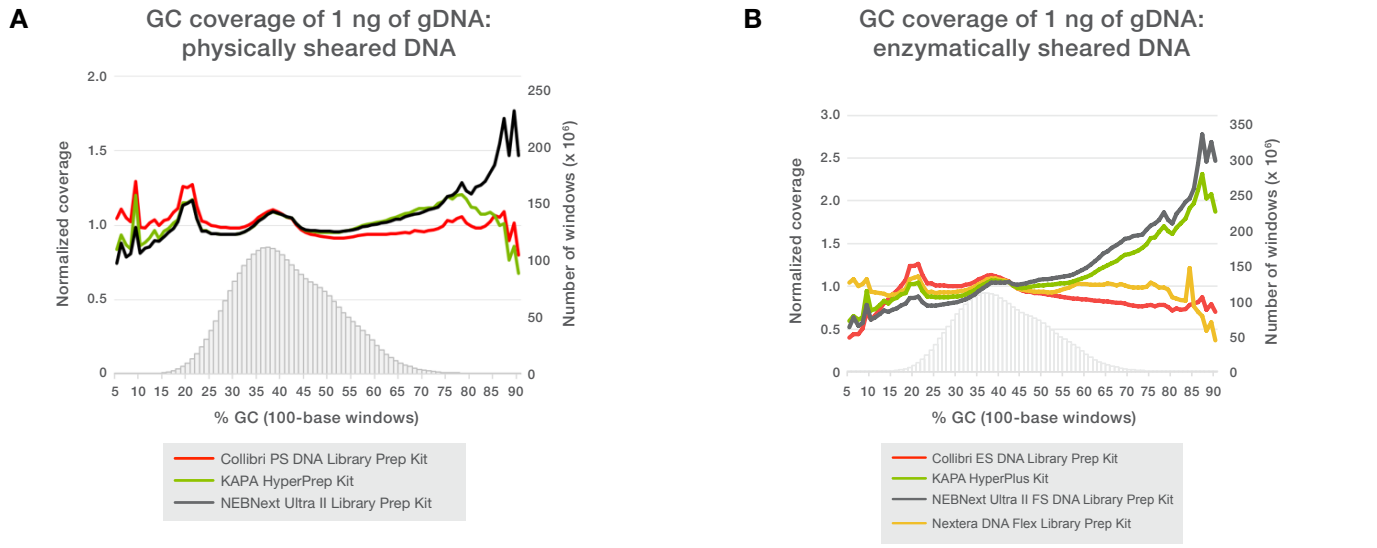

Figure 3. Coverage across GC content. Collibri DNA library preparation kits enable **(A)** uniform coverage across GC content with physical shearing (PS) and **(B)** representative coverage with enzymatic shearing (ES). Libraries were prepared following the manufacturer's instructions with the recommended PCR enzymes. Only Collibri DNA library preparation kits use Collibri Library Amplification Master Mix containing Invitrogen™ Platinum™ SuperFi™ DNA Polymerase. Libraries were sequenced on a NovaSeq 6000 system using an S4 flow cell.

Ordering information

Product		Quantity	Cat. No.
DNA-Seq kits for Illumina systems			
Collibri ES DNA Library Prep Kits	with CD Indexes	24 preps	A38605024
	with CD Indexes	96 preps	A38607096
	with UD Indexes, Set A (1-24)	24 preps	A38606024
	with UD Indexes, Set B (25-48)	24 preps	A43605024
	with UD Indexes, Set C (49-72)	24 preps	A43606024
	with UD Indexes, Set D (73-96)	24 preps	A43607024
Collibri PCR-Free ES DNA Library Prep Kits	with CD Indexes	24 preps	A38545024
	with CD Indexes	96 preps	A38603096
	with UD Indexes, Set A (1-24)	24 preps	A38602024
	with UD Indexes, Set B (25-48)	24 preps	A43602024
	with UD Indexes, Set C (49-72)	24 preps	A43603024
	with UD Indexes, Set D (73-96)	24 preps	A43604024
Collibri PS DNA Library Prep Kits	with CD Indexes	24 preps	A38612024
	with CD Indexes	96 preps	A38614096
	with UD Indexes, Set A (1-24)	24 preps	A38613024
	with UD Indexes, Set B (25-48)	24 preps	A43611024
	with UD Indexes, Set C (49-72)	24 preps	A43612024
	with UD Indexes, Set D (73-96)	24 preps	A43613024
	with UD Indexes, Set A-D (1-96)	96 preps	A38614196
Collibri PCR-Free PS DNA Library Prep Kits	with CD Indexes	24 preps	A38608024
	with CD Indexes	96 preps	A38610096
	with UD Indexes, Set A (1-24)	24 preps	A38609024
	with UD Indexes, Set B (25-48)	24 preps	A43608024
	with UD Indexes, Set C (49-72)	24 preps	A43609024
	with UD Indexes, Set D (73-96)	24 preps	A43610024
	with UD Indexes, Set A-D (1-96)	96 preps	A38615196

CD = combinatorial dual, UD = unique dual

invitrogen

Ordering information (continued)

Product	Quantity	Cat. No.
RNA-Seq kits for Illumina systems		
Collibri Stranded RNA Library Prep Kit for Illumina Systems	24 preps	A38994024
	96 preps	A38994096
Collibri Stranded RNA Library Prep Kit for Illumina Systems with H/M/R rRNA Depletion Kit	24 preps	A39003024
	96 preps	A39003096
ERCC RNA Spike-In Mix	1 kit	4456740
ERCC ExFold RNA Spike-In Mixes	1 kit	4456739
Library quantification		
Collibri Library Quantification Kit	100 rxns	A38524100
	500 rxns	A38524500
Qubit 4 Fluorometer, with WiFi	1 fluorometer	Q33238
Qubit 4 NGS Starter Kit, with WiFi	1 kit	Q33240
Library amplification		
Collibri Library Amplification Master Mix	50 rxns	A38539050
	250 rxns	A38539250
Collibri Library Amplification Master Mix with Primer Mix	50 rxns	A38540050
	250 rxns	A38540250

H = human, M = mouse, R = rat

References

1. Quail MA (2010) DNA: Mechanical breakage. In: *eLS*. Chichester: Wiley & Sons.
2. Knierim E, Lucke B, Schwarz JM et al. (2011) Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS One* 6(11):e28240.

Find out more at **thermofisher.com/collibri**

ThermoFisher SCIENTIFIC