Improved library preparation with the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina®

Addressing lower input amounts and challenging sample types.

Introduction

Library preparation is a critical part of the next generation sequencing workflow; successful sequencing requires the generation of high quality libraries of sufficient yield and quality.

As sequencing technologies improve and capacities expand, boundaries are also being pushed on library construction. High performance is required from ever-decreasing input quantities and from samples of lower quality or those with varying GC content. At the same time, the need is increasing for faster, automatable protocols that perform reliably and do not compromise the quality of the libraries produced.

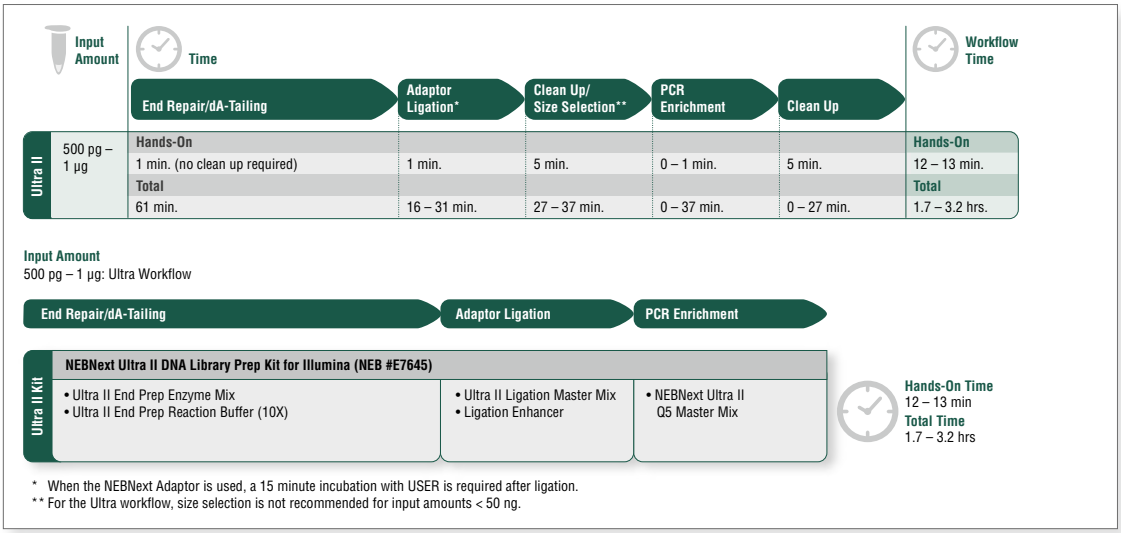
To meet these growing challenges, we have reformulated each of the reagents in our NEBNext Ultra DNA workflow to create the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB #E7645). This new kit utilizes a fast, streamlined, automatable workflow for high yield production of superior quality libraries, with picogram to microgram input amounts of DNA of varying quality. This new generation of NEBNext reagents requires fewer PCR cycles and also improves GC coverage.

Here we demonstrate the utility of the NEBNext Ultra II DNA Library Prep Kit for DNA library construction from a variety of sample types and for a number of applications.

The NEBNext Ultra II WorkFlow

The NEBNext Ultra II DNA Library Prep Kit contains 5 components. The workflow combines the End Repair and dA-Tailing steps and minimizes clean up steps, making the kit fast (~ 2.5 hours) and easy to use (Figure 1). The protocol can accommodate 500 picograms to 1 microgram of input DNA, which can be sheared by either mechanical or enzyme-based methods. The kit can also be used in PCR-free workflows. The protocol is compatible with adaptors and primers from the NEBNext product line ("NEBNext Oligos") or from other sources.

Figure 1. NEBNext Ultra II DNA Library Prep Kit workflow



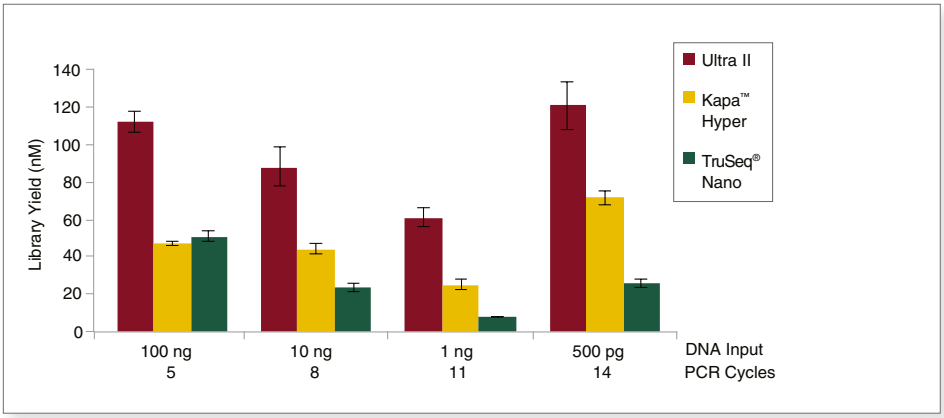
10/15

NEW ENGLAND BioLabs® GmbH

Improvements in Library Yield and Conversion Rate

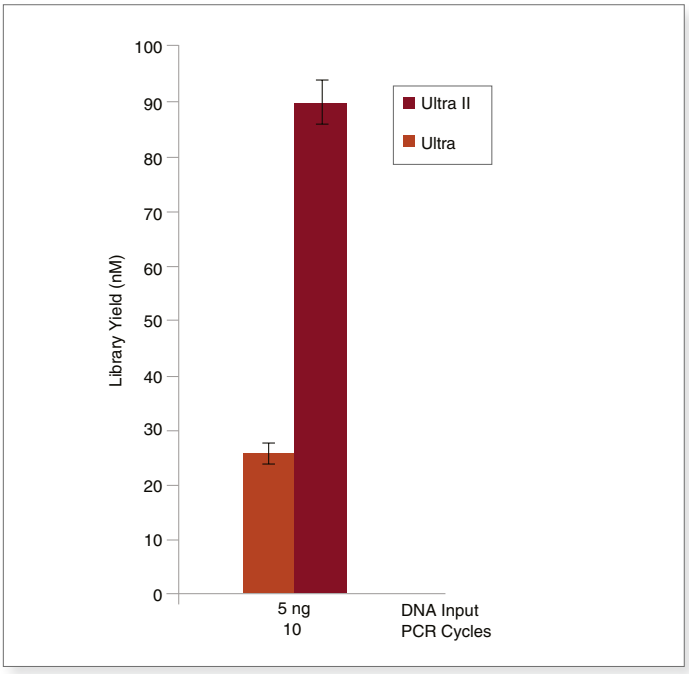
An important measure of the success of library preparation is the yield of the final library. The reformulation of each reagent in the library prep workflow enables substantially higher yields from the NEBNext Ultra II DNA Kit compared to other commercially available kits (Figure 2), as well as NEB's original Ultra DNA kit (Figure 3). Achieving sufficient yields for high quality cluster generation and sequencing from very low input amounts can be challenging, a fact compounded by the preference to amplify the library using as few PCR cycles as possible. NEBNext Ultra II overcomes this challenge, and users can now obtain higher library yields with lower input amounts, as shown in Figure 2.

Figure 2. NEBNext Ultra II produces the highest yield libraries from a broad range of input amounts.



Libraries were prepared from Human NA19240 genomic DNA using the input amounts and numbers of PCR cycles shown. Manufacturers' recommendations were followed, with the exception that size selection was omitted.

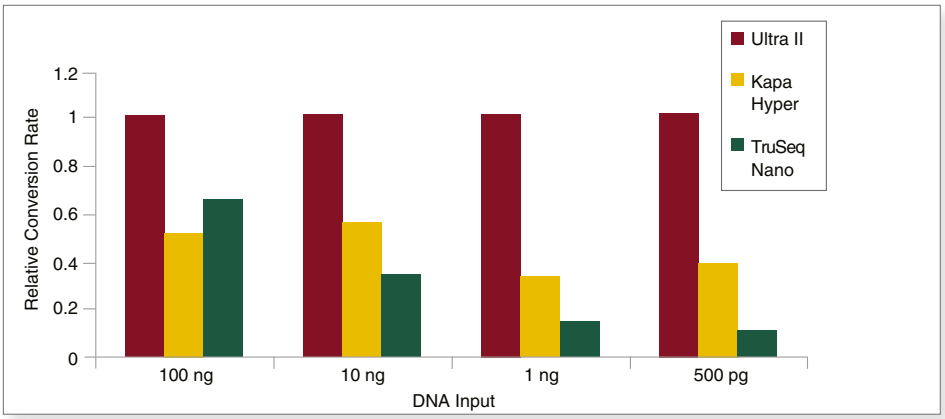
Figure 3. NEBNext Ultra II produces library yields several fold higher than the original Ultra kit.



Libraries were prepared from Human NA19240 genomic DNA using the lowest input amount recommended for the original NEBNext Ultra Kit (5 ng) and 10 PCR cycles. Significantly higher yields were achieved with the NEBNext Ultra II Kit than with the original Ultra Kit.

The efficiency of the End Repair, dA-Tailing and Adaptor Ligation steps during library construction can be measured separately from the PCR step by doing qPCR quantitation of adaptor-ligated fragments prior to library amplification. This enables determination of the rate of conversion of input DNA to adaptor-ligated fragments, i.e. sequenceable molecules. Therefore, measuring conversion rates is another way to assess the efficiency of library construction and also provide information on the diversity of the library. Again, NEBNext Ultra II enables substantially higher rates of conversion as compared to other commercially available kits (Figure 4).

Figure 4. NEBNext Ultra II produces the highest rates of conversion to adaptor-ligated molecules from a broad range of input amounts.



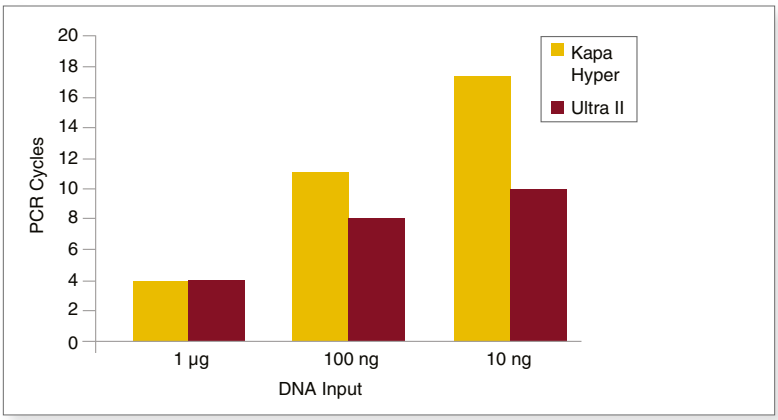
Libraries were prepared from Human NA19240 genomic DNA using the input amounts and library prep kits shown without an amplification step, and following manufacturers' recommendations. qPCR was used to quantitate adaptor-ligated molecules, and quantitation values were then normalized to the conversion rate for Ultra II. The Ultra II kit produces the highest rate of conversion to adaptor-ligated molecules, for a broad range of input amounts.

Minimization of PCR Cycles

In general, it is preferable to use as few PCR cycles as possible to amplify libraries. In addition to reducing workflow time, this also limits the risk of introducing bias during PCR. A consequence of increased efficiency of End Repair, dA-Tailing and Adaptor Ligation is that fewer PCR cycles are required to achieve the library yields necessary for sequencing or other intermediate downstream workflows.

For applications such as exome enrichment, very high library yields (1 µg or more) are generally required as input for the enrichment step. If library preparation is inefficient, a large number of PCR cycles may be required to achieve these required yields, especially when the input amount for the original library is low (Figure 5). This can result in production of a library that is not representative of the original sample. For more details on library preparation for target enrichment applications see pages 8 and 9.

Figure 5. Number of PCR cycles required to generate ≥ 1 µg amplified library for target enrichment.



Ultra II libraries were prepared from Human NA19240 genomic DNA using NEBNext Ultra II and the input amounts shown. Yields were measured after each PCR cycle and the number of cycles required to generate at least 1 µg of amplified library determined. Cycle numbers for Kapa Hyper were obtained from Kapa Biosystems webpage (https://www.kapabiosystems.com/product-applications/products/next-generation-sequencing-2/dna-library-preparation/kapa-hyper-prep-kits/) and plotted alongside the cycle numbers obtained experimentally for Ultra II.

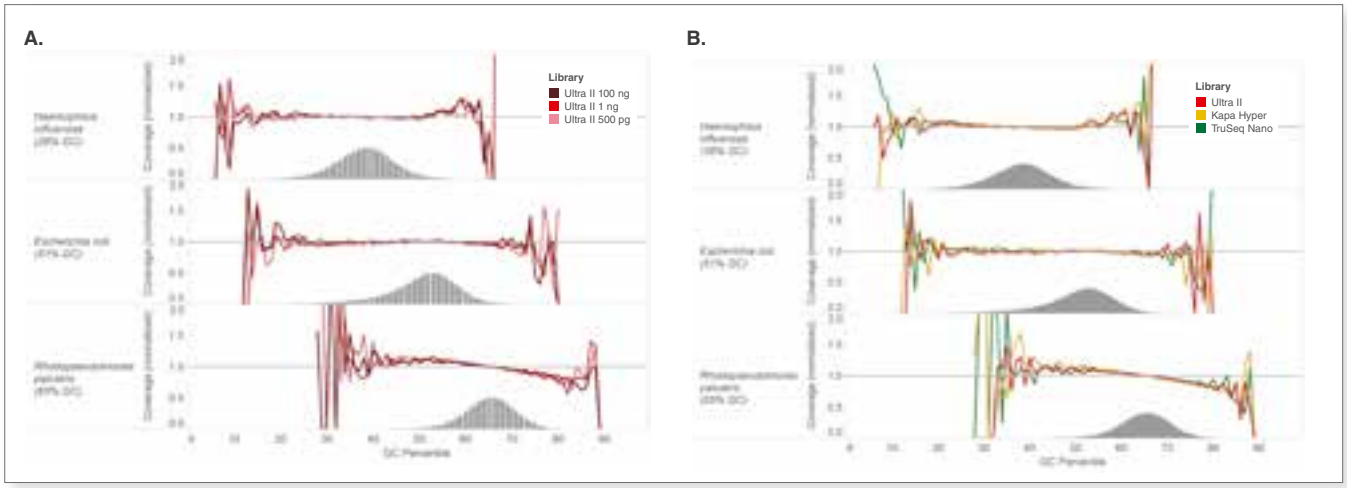
Improvements in Library Quality

While sufficient yield of a library is required for successful sequencing, quantity alone is not enough. The quality of a library is also critical, regardless of the input amount or GC content of the sample DNA. A high quality library will have uniform representation of the original sample, as well as even coverage across the GC spectrum.

Uniform GC Coverage

Libraries from varying input amounts of three microbial genomic DNAs with low, medium and high GC content (*H. influenza, E. coli* and *H. palustris*) were prepared using the NEBNext Ultra II Kit. In all cases, uniform coverage was obtained, regardless of GC content and input amount (Figure 6A). GC coverage of libraries prepared using other commercially available kits was also analyzed using the same trio of genomic DNAs. Again, NEBNext Ultra II provided good GC coverage (Figure 6B).

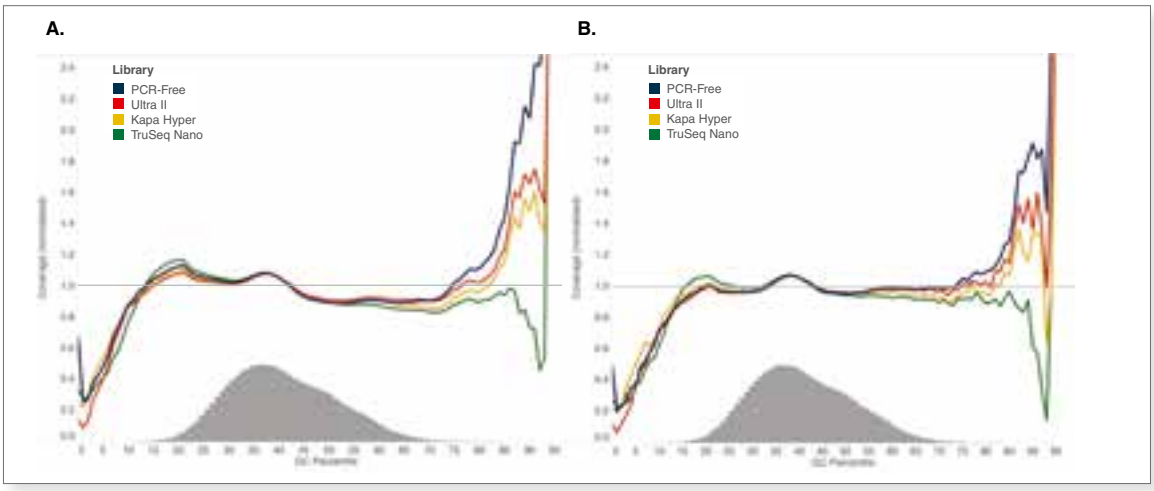
Figure 6. NEBNext Ultra II provides uniform GC coverage for microbial genomic DNA over a broad range of GC composition and input amounts.



*Libraries were made using 500 pg, 1 ng and 100 ng of the genomic DNAs shown and the Ultra II DNA Library Prep Kit **(A)** or using 100 ng of the genomic DNAs and the library prep kits shown **(B)**, and sequenced on an Illumina MiSeq®. 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 100 bp regions at each GC% is indicated by the vertical grey bars, and the colored lines represent the normalized coverage for each library.*

When amplification is required to obtain sufficient library yields, it is important to ensure that no bias is introduced, and that representation of GC-rich and AT-rich regions is not skewed in the final library. Comparison with libraries produced without amplification ("PCR-free") is a useful measure (1). In this example, a library was prepared from human genomic DNA using NEBNext Ultra II as well as other commercially available kits, and coverage was compared to a PCR-free library. Results demonstrate that the Ultra II library coverage is most similar to the PCR-free library, and also covers the range of GC content (Figure 7).

Figure 7. NEBNext Ultra II provides GC coverage of human DNA comparable to PCR-free libraries.



*Libraries were made using 100 ng of Human NA19240 genomic DNA and the kits shown, following manufacturers' recommendations. Libraries were sequenced on an Illumina NextSeq® 500 **(A)** or MiSeq **(B)**. 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 100 bp regions at each GC% is indicated by the vertical grey bars, and the colored lines represent the normalized coverage for each library. Ultra II provides GC coverage most similar to PCR-free libraries and enables coverage across the range of GC content.*

Figure 8. Ultra II libraries provide the highest quality sequencing data.

DNA INPUT	LIBRARY KIT	TOTAL READS	% MAPPED	% DUPLICATION	% CHIMERAS
100 ng	Ultra II	419,093,838	96	1.87	0.48
	Kapa Hyper	419,097,926	96	2.00	0.60
	TruSeq Nano	419,086,546	97	1.91	0.53
1 ng	Ultra II	226,860,968	96	3.96	0.44
	Kapa Hyper	226,857,578	96	11.40	0.53
	TruSeq Nano	226,857,754	97	34.80	0.41

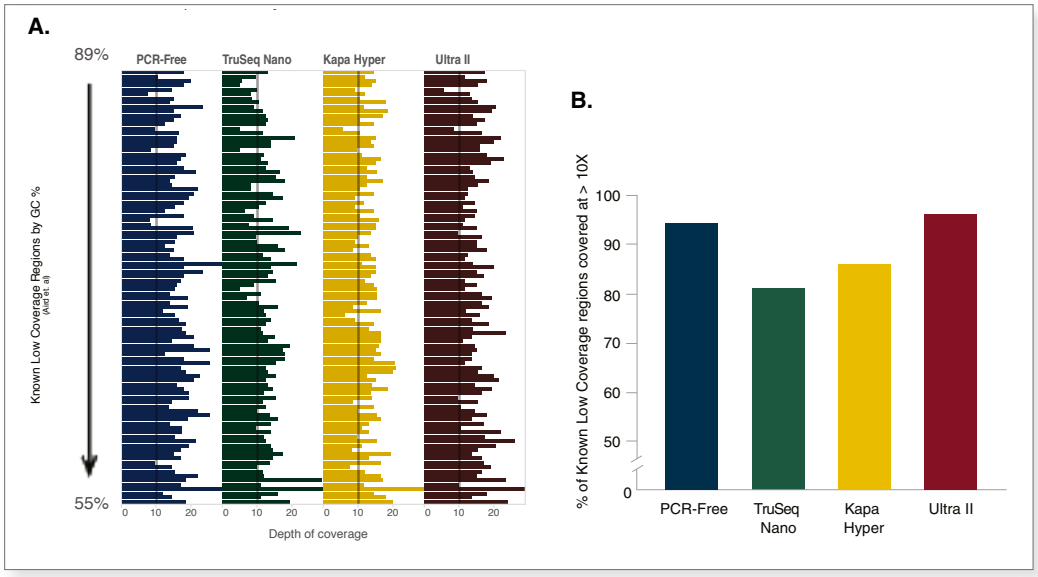
Libraries were prepared from Human NA19240 genomic DNA using the input amounts and library prep kits shown, following manufacturers' recommendations. Libraries were sequenced on the Illumina NextSeq 500. Reads were mapped to the hg19 reference using Bowtie 2.2.4. This data illustrates that the NEBNext Ultra II DNA Library Prep Kit enables high quality sequence data, even with very low input amounts.

% Mapped: *The percentage of reads mapped to Human hg19 reference.*
% Duplication: *The percentage of mapped sequence that is marked as duplicate.*
% Chimeras: *The percentage of reads that map outside of a maximum insert size or that have the two ends mapping to different chromosomes.*

Coverage of Known Low-Coverage Regions of the Human Genome

Regions of the human genome typically covered at a relatively low level have been identified (2), and the majority of these regions have high GC content. Library preparation can contribute to low and uneven sequence coverage, or even drop-outs, of these challenging regions. Depending on the polymerase used, PCR amplification of a library can result in under-representation of GC-rich regions, and libraries constructed by PCR-free workflows can provide more uniform coverage than amplified libraries (1). Improvements in efficiency and reduction in bias at each step in library preparation, including improved uniformity of library amplification over the full range of GC content improves the evenness of sequence coverage of these regions. Here we show a comparison of sequencing data from human genomic DNA libraries prepared with NEBNext Ultra II and other commercially available kits. Ultra II provided the highest and most uniform coverage of difficult sequence regions, as well as the coverage most similar to the PCR-free library (Figure 9).

Figure 9. NEBNext Ultra II provides the highest and most uniform coverage of difficult sequence regions.



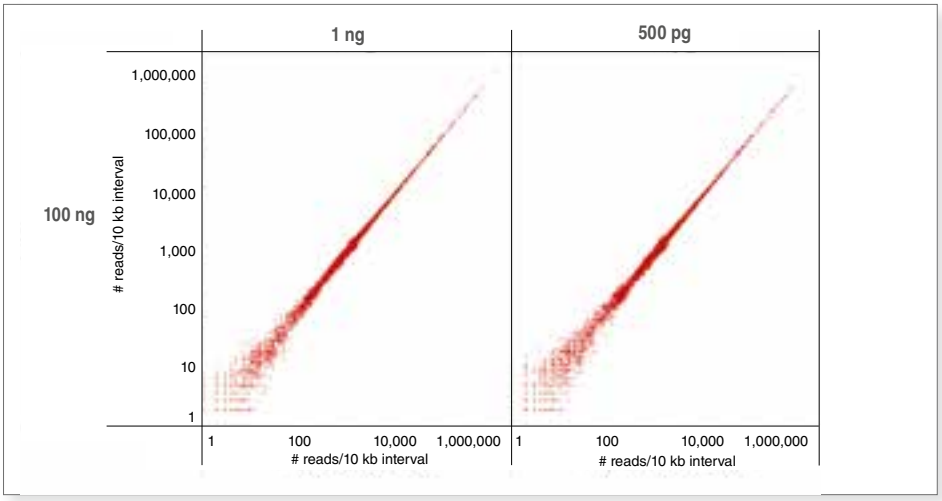
A: *Indexed libraries were prepared from 100 ng of Human NA19240 genomic DNA using a PCR-free workflow or the library prep kits shown, following manufacturers' recommendations. The PCR-free library was prepared using NEBNext Ultra II. Libraries were sequenced on the Illumina NextSeq 500. 420 million reads were randomly extracted from each dataset, to produce an average coverage of 10X. Reads were mapped to the hg19 reference genome using Bowtie 2.2.4. Reads on each region were counted using bedtools v2.19.1. The number of reads overlapping distinct difficult, low-coverage regions of the human genome (2) are shown for each library. Ultra II provides the highest and most uniform coverage of these difficult regions, and provides the coverage closest to that obtained with a PCR-free protocol.*

B: *From the 420 million 75 bp reads randomly extracted from each dataset, 10X coverage was expected. The percentage of difficult regions covered at ≥ 10X is shown for each library prep kit and for the PCR-free workflow. Ultra II provides the highest percentage of reads at ≥ 10X coverage and also provides the coverage closest to that obtained with a PCR-free protocol.*

Library Diversity

As described above, an ideal library will represent completely and proportionally the sequence of the input DNA. When library preparation is inefficient or when input amounts for a library are very low, there is a risk that the resulting library will lack this diversity, and that some sequences will be over- or under-represented. Comparison of the level of sequence coverage, in 10 kb intervals, achieved with libraries from different input amounts is a useful measure of the diversity of a library. The increased efficiency of each step in the NEBNext Ultra II library workflow improves the library diversity. Here we show comparisons of libraries prepared for 100 ng, 1 ng and 500 pg human genomic DNA prepared using NEBNext Ultra II. The results demonstrate consistently high coverage for the range of input amounts indicated (Figure 10).

Figure 10. Read depth correlation shows consistently high coverage for 500 pg–100 ng input amounts.

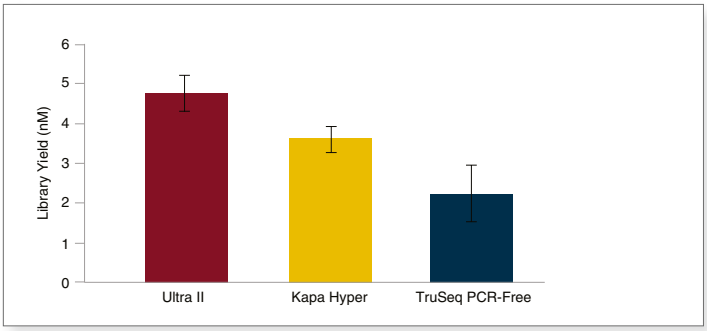


Libraries were prepared with 100 ng, 1 ng and 500 pg of human NA19240 genomic DNA. Each library was downsampled (sambamba view -s) to include 423 M reads and mapped to GRCh37 using Bowtie 2. Coverage of each 10 kb region of GRCh37 (as determined by bedtools coverage) was compared between low (500 pg and 1 µg) and 100 ng input. Most regions are covered by ~1,000 reads, as expected. Low and high coverage regions are well correlated.

Construction of libraries using PCR-Free Workflows

Construction of a library using a PCR-free workflow removes the risk of incorporation of bias during library amplification. However, the input amounts required to produce sufficient yield of a high diversity library without an amplification step can be substantial, and often limiting. Here we show comparison of yields and sequence quality for libraries generated using Ultra II and and other commercially available library preparation kits in workflows lacking a PCR step. (Figures 11 and 12). The improvements made in the efficiencies of the End Repair, dA-Tailing and Adaptor Ligation steps with NEBNext Ultra II enable the use of lower input amounts to generate high quality libraries constructed from a PCR-free workflow, and make the omission of the PCR step now feasible for nanogram level input amounts.

Figure 11. NEBNext Ultra II provides superior yields in PCR-free workflows



Libraries were generated from 100 ng of Human NA19240 genomic DNA using the library prep kits shown, following manufacturers' recommendations, and with no amplification step. Library yields were determined by qPCR using the NEBNext Library Quant Kit for Illumina. The NEBNext Ultra II kit produces the highest yields.

Figure 12. NEBNext Ultra II libraries provide the highest quality sequence data in PCR-free workflows

LIBRARY KIT	TOTAL READS	% DUPLICATION	% MAPPED	% CHIMERAS
Ultra II	3,685,029	0.02	97.00	1.10
Kapa Hyper	3,679,136	0.02	97.00	2.00
TruSeq PCR-Free	3,681,643	0.05	96.65	1.38

PCR-free libraries were generated from 100 ng of Human NA19240 genomic DNA using the library prep kits shown, following manufacturers' recommendations, and with no amplification step. This data illustrates that NEBNext Ultra II enables high quality sequence data in PCR-free workflows, even with low input amounts.

% Mapped: *The percentage of reads mapped to human hg19 reference.*
% Duplication: *The percentage of mapped sequence that is marked as duplicate.*
% Chimeras: *The percentage of reads that map outside of a maximum insert size or that have the two ends mapping to different chromosomes.*

Library preparation with FFPE-treated samples

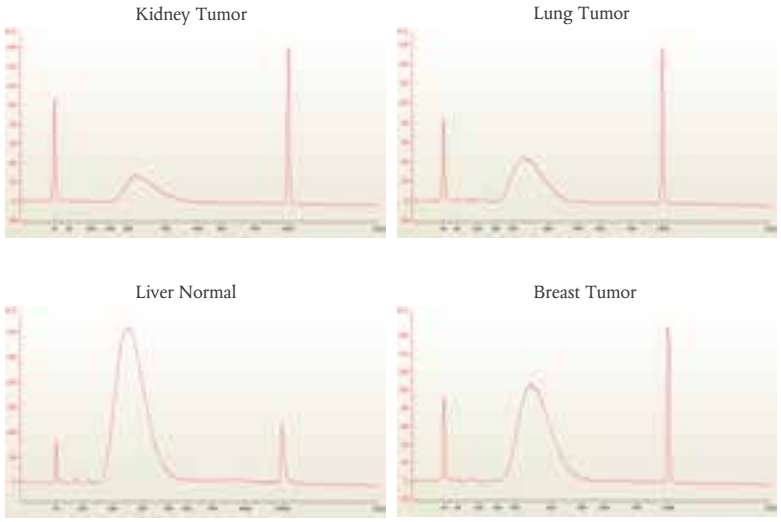
Archiving of clinical materials as Formalin-Fixed, Paraffin-Embedded (FFPE) samples is a common practice. However, the methods used for fixation and storage significantly damage the nucleic acids from these samples. Additionally, FFPE DNA samples are often available in only very small amounts. As a result, it can be challenging to construct high quality libraries in sufficient quantity to achieve good sequence data at the desired depth of coverage. NEBNext Ultra II libraries were made from low nanogram amounts of several FFPE DNA samples of varying age and quality and then sequenced. Bioanalyzer traces of the libraries and analysis of the sequencing data show that the quality of the libraries was high.

Figure 13. NEBNext Ultra II enables construction of high quality libraries from FFPE DNA samples.

A.

FFPE DNA	DNA INPUT (ng)	LIBRARY YIELDS IN ng	% MAPPED	% MAPPED IN PAIRS	% DUPLICATION	% CHIMERAS
Kidney Tumor	17	132	91.5	96.1	0.48	3.0
Lung Tumor	20	232	90.1	94.9	0.42	4.1
Liver Normal	20	691	92.6	94.7	0.33	8.6
Breast Tumor	30	514	91.9	95.1	0.37	4.5

B.



Libraries were prepared from 17–30 ng of human DNA extracted from the FFPE tissue samples listed, amplified using 10 cycles of PCR and sequenced on the Illumina MiSeq. This data illustrates that NEBNext Ultra II enables high quality sequence data, even with low input amounts of FFPE DNA.

A: *Reads were mapped to the hg19 reference genome using Bowtie 2.2.4.*
% Mapped: *The percentage of reads mapped to Human hg19 reference.*
% Mapped in Pairs: *The percentage of reads whose mate pair was also aligned to the reference.*
% Duplication: *The percentage of mapped sequence that is marked as duplicate.*
% Chimeras: *The percentage of reads that map outside of a maximum insert size or that have the two ends mapping to different chromosomes.*

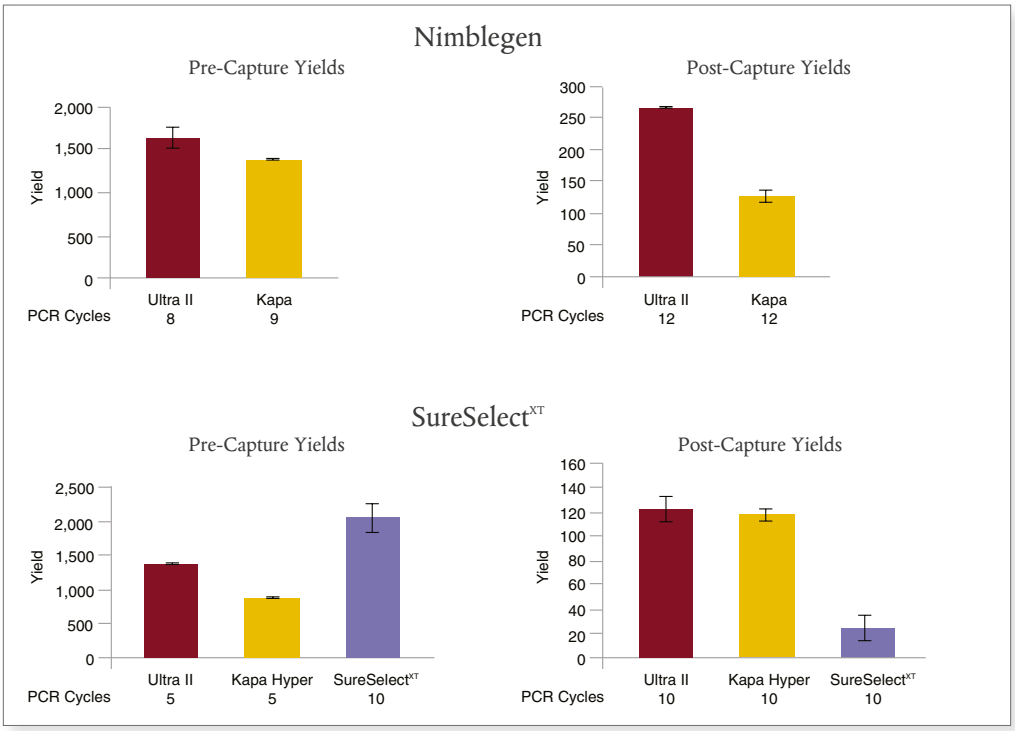
B: *Bioanalyzer® traces of each library show high quality libraries with minimal adaptor-dimer.*

Library Preparation for Target Enrichment Workflows

For hybridization-based target enrichment workflows the amount of library required can be large, generally in the microgram range. This can be challenging to achieve, especially when only small amounts of the original sample are available. As described in a previous section (see page 3, Figure 5), NEBNext Ultra II reduces the number of PCR cycles required to yield 1 µg or more of library, thereby enabling the use of low nanogram sample input amounts for library construction for target enrichment workflows.

Importantly, in addition to being of sufficient quantity, libraries must be of high quality to enable effective target enrichment and generation of high quality sequence data. Libraries were generated using NEBNext Ultra II and other commercially available library preparation reagents, followed by exome capture. Measurement of yields and sequencing metrics indicated superior performance from Ultra II libraries. (Figures 14 and 15).

Figure 14. NEBNext Ultra II provides superior yields for target enrichment applications.



For Nimblegen® enrichment, libraries were prepared with 100 ng of Human HG02922 genomic DNA, using NEBNext Ultra II with the NEBNext Adaptor and index primers, or the Kapa Library Preparation Kit Illumina Platforms with the SeqCap® Adapter Kit (Roche). Target enrichment was performed with NimbleGen SeqCap Human Exome v3.

For SureSelect^{XT} enrichment, libraries were prepared with NEBNext Ultra II or Kapa Hyper with the NEBNext Adaptor and index primers, or the SureSelect^{XT} Reagent Kit, Illumina (ILM) platforms with Herculase® II Fusion DNA Polymerase. Target enrichment was performed with SureSelect® Human All Exon V6.

Manufacturers' recommendations were followed, including use of the recommended number of PCR cycles for pre- and post-capture. NEBNext Ultra II libraries resulted in the highest post-capture yields for both enrichment workflows.

Figure 15. NEBNext Ultra II libraries provide superior results in exome enrichment.

	SAMPLE 1		SAMPLE 2	
Library Prep Kit	Ultra II	Kapa	Ultra II	Kapa
Total Reads	81 M	81 M	81 M	81 M
% Duplication	9.2%	14.7%	9.7%	12.4%
% Selected Base	90.3%	89.0%	89.7%	89.0%
Mean Target Coverage (x)	54	50	53	51
% Zero Coverage	0.37%	0.37%	0.10%	0.11%
Fold 80 Base Penalty	2.56	2.77	2.65	2.81
HS Library Size (molecules)	167.9 M	98.2 M	156.1 M	117.7 M

Pre-capture libraries were prepared with 100 ng of Human HG00096 (Sample 1) and HG02922 (Sample 2) genomic DNA, using NEBNext Ultra II with the NEBNext Adaptor and index primers, or the Kapa Library Preparation Kit Illumina Platforms with the SeqCap Adapter Kit (Roche). Hybrid selection of the human exome was performed with 1 µg of each library and SeqCap human exome v3 following the NimbleGen SeqCap protocol. Post-capture amplification was conducted with NEBNext Ultra II Q5 Hot Start Master Mix for Ultra II libraries, and Kapa HiFi Ready Mix for Kapa libraries. Libraries were pooled and sequenced on an Illumina NextSeq 500 instrument. Sequencing reads were mapped to human hg19 reference and analyzed using Picard HS Metrics tool.

% Duplication: *The percentage of mapped sequence that is marked as duplicate.*
% Selected Base: *On+Near Bait Bases/PF Bases Aligned.*
Mean Target Coverage: *The mean coverage of targets that received at least coverage depth = 2 at one base.*
% Zero Coverage: *The number of targets that did not reach coverage=2 over any base.*
Fold 80 Base Penalty: *The fold over-coverage necessary to raise 80% of bases in "non-zero-coverage" targets to the mean coverage level in those targets.*
HS Library Size: *The estimated number of unique molecules in the library.*

The NEBNext Ultra II Kit leads to lower duplication rate, higher percentage of selected base, lower Fold 80 Base Penalty, and significantly larger HS library size, suggesting a higher degree of uniformity of NEBNext Ultra II libraries.

Conclusion

The NEBNext Ultra II DNA Library Prep Kit for Illumina represents a substantial advance in library preparation for Illumina sequencing. Reformulated reagents increase the efficiency of each step in the workflow, and enable users to overcome many of the challenges previously associated with successful library preparation, such as:

- The use of input amounts in the picogram to microgram range
- Generation of higher yields
- The use of challenging sample types such as FFPE DNA
- The use of fewer PCR cycles
- Uniform GC coverage of the sample
- Fast, streamlined library preparation that is automation-friendly

References:
1. Kozarewa, I. et al. (2009). Amplification-free Illumina sequencing – library preparation facilitates improved mapping and assembly of (G+C) – biased genomes. *Nat. Methods* 6:291–295.
2. Aird, D. et al. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology* 12(2), R18.

www.neb-online.de
NEW ENGLAND BioLabs® GmbH enabling technologies in the life sciences
New England Biolabs GmbH, Brüningstr. 50, Geb. B852, 65926 Frankfurt/Main, Germany
Tel: +49/(0)69/305-23140 Toll Free: (Germany) 0800/246-5227 Toll Free: (Austria) 00800/246-52277 Fax: +49/(0)69/305-23149 e-mail: info.de@neb.com