Improvements to the ThruPLEX HV kits

Introduction

Successful next-generation sequencing (NGS) experiments rely on a streamlined, straightforward workflow that accommodates a wide range of sample inputs and produces consistent results. ThruPLEX HV PLUS combines the complete, efficient, and accurate workflow of ThruPLEX HV with the ThruPLEX HV PLUS Enzymatic Fragmentation Module to produce Illumina®-ready libraries. With the same three-step, one-tube workflow as ThruPLEX HV, ThruPLEX HV PLUS fragments and repairs in parallel to decrease hands-on time and to remove extraneous steps for separate fragmentation (Figure 1). This industry-leading, single-tube workflow prevents sample loss and eliminates the need for time-consuming post-ligation bead purification.

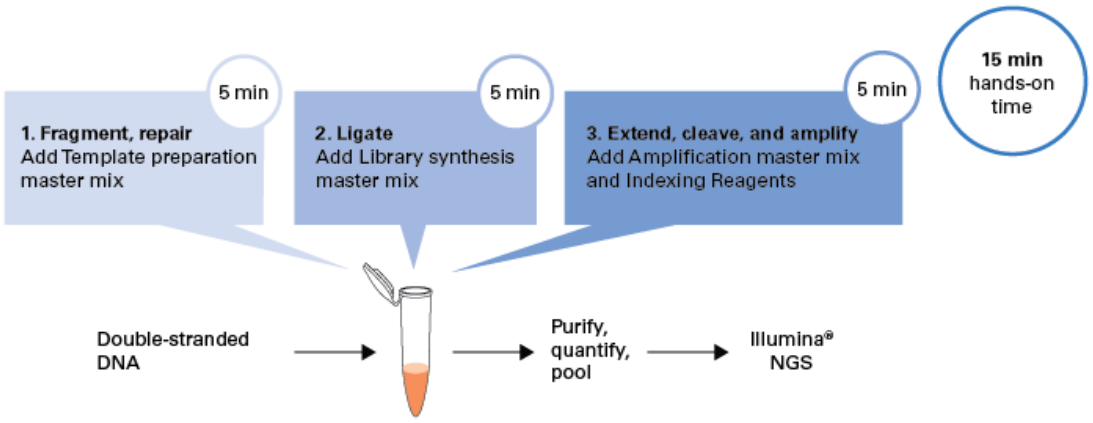

Figure 1. ThruPLEX HV PLUS single-tube library preparation workflow. The ThruPLEX HV PLUS workflow consists of three simple steps that take place in the same well or PCR tube, eliminating the need to purify or transfer the sample material. In this latest version of the ThruPLEX technology, an enzymatic fragmentation step at the start of the protocol streamlines the generation of higher-complexity libraries from double-stranded DNA, up to 200 ng in 30 µl.

The ThruPLEX HV PLUS kits accommodate up to 200 ng of starting input with a volume up to 30 µl, similar to its predecessor. The libraries produced by ThruPLEX HV PLUS kits are ready to be used directly for whole genome sequencing applications or enriched using a custom panel of the leading target-enrichment platforms.

Results

Familiar workflow, consistent performance—improved with fragmentation

The ThruPLEX HV PLUS library preparation system expands the leading single-tube workflow of the ThruPLEX HV kits with a modified template preparation step. In the same reaction conditions and time, it now includes an enzymatic fragmentation module. The incorporation of the enzymatic fragmentation module allows ThruPLEX HV PLUS kits to provide protocols optimal for generating DNA fragments of 300 and 450 bp. Fragment size can be modulated by simply varying the concentration of fragmentation enzyme. With built-in enzymatic fragmentation in the first step, separate enzymatic or mechanical fragmentation steps are not necessary.

The ThruPLEX DNA-Seq HV PLUS Kit performs comparably to the ThruPLEX DNA-Seq HV Kit. ThruPLEX DNA-Seq HV PLUS Kits demonstrate similar coverage uniformity as ThruPLEX DNA-Seq HV across the recommended input range (Figure 2). At the lowest recommended input value, 5 ng, the ThruPLEX HV PLUS Kit rivals its predecessor using mechanical shearing (Figure 2).

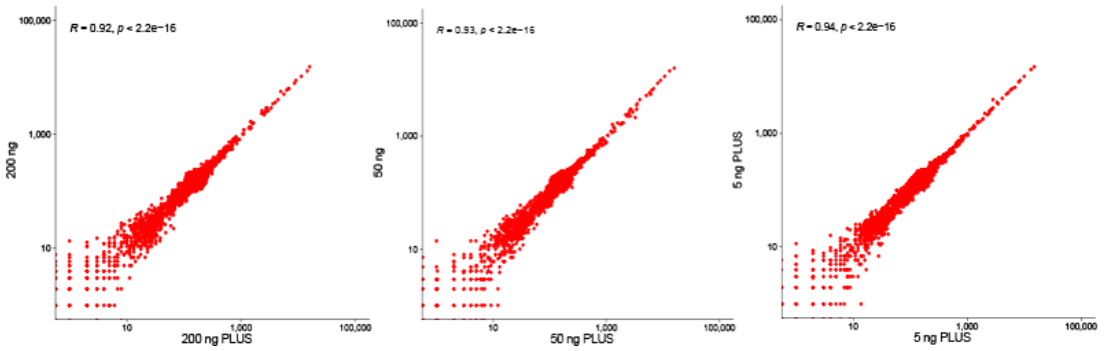

Figure 2. Reliable coverage and similar performance. ThruPLEX DNA-Seq HV (Y-axis) and ThruPLEX DNA-Seq HV PLUS (X-axis) provide robust performance across a broad range of inputs. Correlation plots are shown for library preparations generated by ThruPLEX DNA-Seq HV and ThruPLEX DNA-Seq HV PLUS with 5, 50, and 200 ng of NA12878 and downsampled to 5 million total reads. Coverage of each 100-kb region of hg19 was compared across inputs.

Uniform GC coverage

The robustness of a library preparation kit depends on its ability to accurately and uniformly cover an assortment of challenging genomes of varying GC content. Our ThruPLEX HV kits, which employ mechanical shearing of DNA samples, have been previously demonstrated to easily handle complex genomes. Like the ThruPLEX HV kits, ThruPLEX HV PLUS kits produce consistent GC coverage spanning the gDNA input range (Figure 3, Table 1). This impressive GC coverage can also be observed in microbial samples of varying GC content (Figure 4, Table 2).

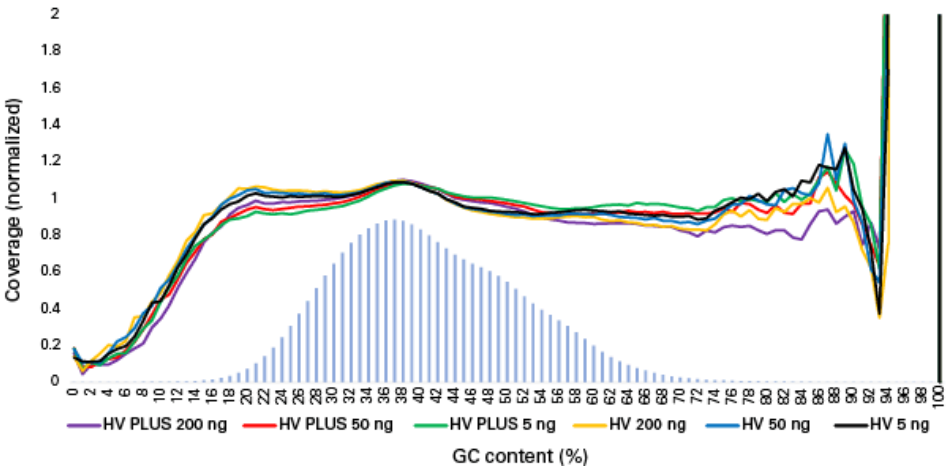

Figure 3. Consistent GC coverage across inputs. Libraries were prepared in triplicate from 5, 50, and 200 ng inputs of NA12878 gDNA. Libraries were generated following ThruPLEX HV PLUS (purple, red and green curves) and ThruPLEX HV (yellow, blue and black curves) protocols. Paired-end sequencing was performed on an Illumina NextSeq® 500/550 Mid Output Kit v2.5 (150 Cycles), and total reads were downsampled to 5 million total reads. The vertical blue bars represent the expected GC content distribution using 100-bp windows.

	Input	Total reads alligned	%reads aligned	% chimera	% duplicate
ThruPLEX HV	200 ng	4.14E+06	96.79%	0.60%	0.72%
	50 ng	4.83E+06	96.69%	0.49%	0.75%
	5 ng	4.84E+06	96.73%	0.50%	0.80%
ThruPLEX HV PLUS	200 ng	4.73E+06	95.10%	2.59%	0.82%
	50 ng	4.76E+06	96.03%	1.54%	0.77%
	5 ng	4.75E+06	96.11%	1.12%	0.91%

Table 1. Comparison of ThruPLEX HV and ThruPLEX HV PLUS. Processed data from ThruPLEX HV and ThruPLEX HV PLUS over a range of input DNA from

5 ng to 200 ng. % reads aligned refers to those successfully aligned to a reference genome. % chimera refers to the percentage of reads that align to two distinct portions of the genome. % duplicate refers to the percentage of reads originated from a single fragment of DNA, typically during library construction via PCR.

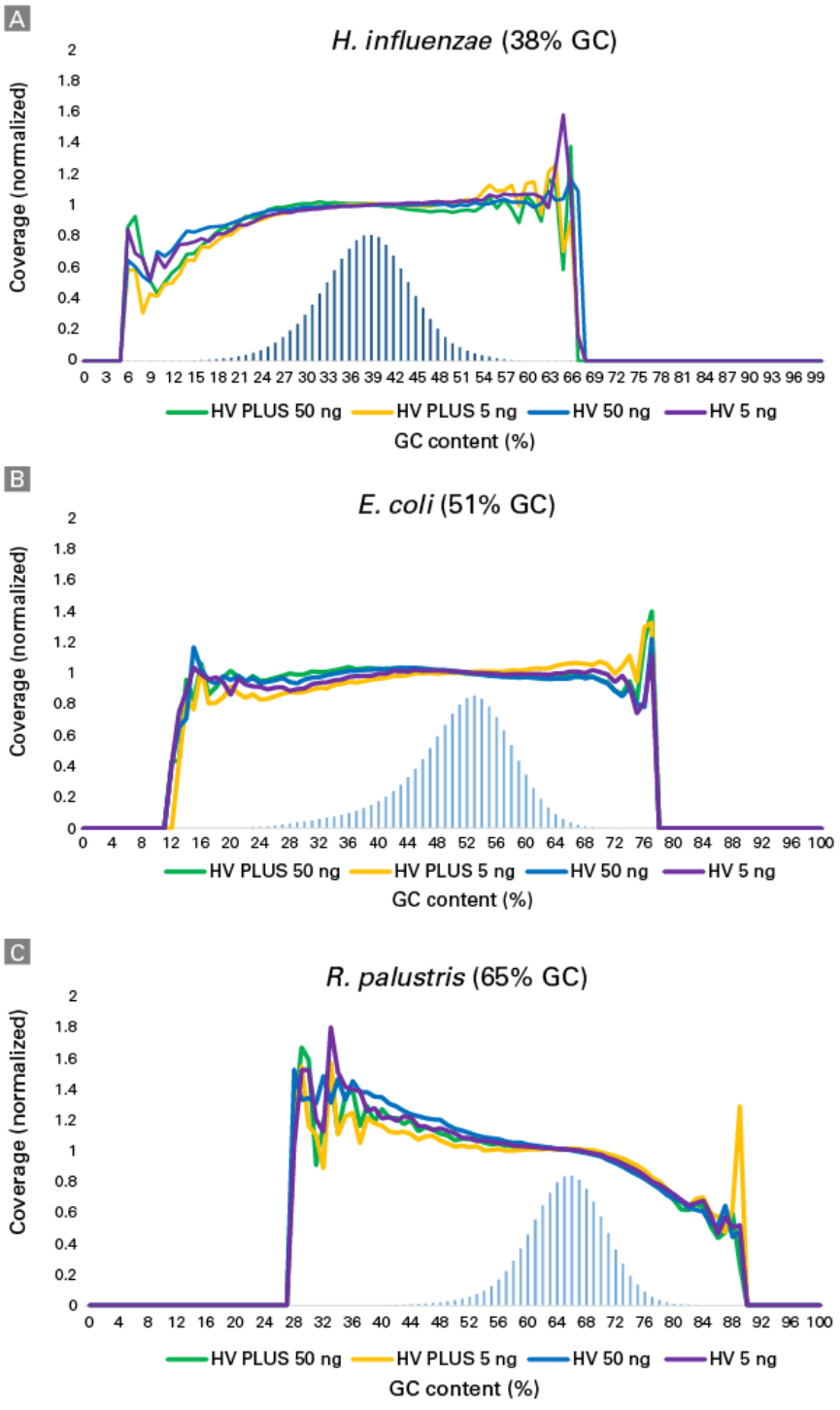

Figure 4. Uniform coverage in libraries with extreme base content. Panels A–C. Libraries were amplified in triplicate with ThruPLEX DNA-Seq HV (blue, purple curves) or ThruPLEX DNA-Seq HV PLUS (green, yellow curves) chemistry using 50-ng and 5-ng inputs of (**Panel A**) *Haemophilus influenzae* 51907D-5, (**Panel B**) *Escherichia coli* 11303, or (**Panel C**) *Rhodopseudomonas palustris* BAA-98D-5 (ATCC). After purification with AMPure beads, paired-end sequencing was performed on a NextSeq 150 Cycle Mid Output (2 x 75 bp). Normalized coverage is represented by the colored lines and the expected number of 100-bp regions at each %GC is represented by the vertical bars. The *H. influenzae* genome is 1.83 Mb in size, with 38% GC content. The *E. coli* genome is 4.7 Mb in size, with 51% GC content. The *R. palustris* genome is 5.46 Mb in size, with 65% GC content. ThruPLEX HV and ThruPLEX HV PLUS demonstrate similar performance when preparing libraries using microbial sample input with a variety of GC contents.

Chemistry	Starting input	Genome	Genome size	% GC	Total reads aligned	%reads aligned	Mean coverage	% chimera	%duplication
ThruPLEX DNA-Seq HV PLUS	50 ng	*H. influenzae*	1.83 Mb	38%	3.52E+06	94.5%	114	2.93%	3.72%
		E. coli	4.7 Mb	51%	3.56E+06	93.7%	44	4.58%	2.47%
		R. palustris	5.46 Mb	65%	3.46E+06	94.6%	33	4.83%	3.05%
ThruPLEX DNA-Seq HV	50 ng	*H. influenzae*	1.83 Mb	38%	3.77E+06	96.2%	115	2.47%	1.13%
		E. coli	4.7 Mb	51%	3.74E+06	95.2%	44	2.76%	0.88%
		R. palustris	5.46 Mb	65%	3.43E+06	96.8%	37	2.35%	1.14%
ThruPLEX DNA-Seq HV PLUS	5 ng	*H. influenzae*	1.83 Mb	38%	3.63E+06	94.4%	110	3.08%	3.19%
		E. coli	4.7 Mb	51%	3.62E+06	93.1%	42	4.30%	1.88%
		R. palustris	5.46 Mb	65%	3.54E+06	94.2%	35	4.35%	2.30%
ThruPLEX DNA-Seq HV	5 ng	*H. influenzae*	1.83 Mb	38%	3.68E+06	95.4%	111	2.88%	1.26%
		E. coli	4.7 Mb	51%	3.72E+06	95.0%	44	2.36%	0.90%
		R. palustris	5.46 Mb	65%	3.66E+06	96.5%	38	2.47%	1.05%

Table 2. Bacterial genomic data comparing ThruPLEX HV and ThruPLEX HV PLUS. Processed data over two sample DNA inputs of 5 ng and 50 ng for *H. influenzae*, *E. coli*, and *R. palustris* with the corresponding GC content of each genome. % mapped refers to those successfully aligned to a reference genome. Mean coverage is the average of the number of unique reads for a given nucleotide in a specific position in the reconstructed sequence. % chimera refers to the percentage of reads that align to two distinct portions of the genome. % duplicate refers to the percentage of reads originated from a single fragment of DNA, typically during library construction via PCR.

Conclusion

The ThruPLEX HV PLUS kits are the newest members of the ThruPLEX HV family. ThruPLEX HV chemistry is engineered and optimized to generate DNA libraries with high molecular complexity and balanced GC representation from input volumes of up to 30 µl. Through workflow optimization, reformulation, and addition of the ThruPLEX HV PLUS Enzymatic Fragmentation Module, ThruPLEX HV PLUS kits perform size-tunable enzymatic fragmentation in tandem with the ThruPLEX HV repair step. The NGS-ready libraries made in a single tube in about two and half hours require only 15 minutes of hands-on time.

The ThruPLEX HV PLUS kits demonstrate consistent performance comparable to the mechanical shearing-based ThruPLEX HV kits. Both chemistries produce consistent GC coverage spanning the input range for gDNA, as well as microbial sample input with a variety of GC contents. The addition of the ThruPLEX HV PLUS Enzymatic Fragmentation Module to the streamlined, reliable workflow of ThruPLEX HV expands the capabilities of whole genome sequencing and is ready for your challenging samples.

Methods

DNA preparation

Human genomic DNA (NA12878) and bacterial genomic DNA from *Haemophilus influenzae* 51907D-5, *Escherichia coli* 11303, or *Rhodopseudomonas palustris* BAA-98D-5 (ATCC) were left intact (ThruPLEX DNA-Seq HV PLUS kits) or mechanically sheared for correct size on a Covaris M220 following the 250- or 200-bp shearing protocol and evaluated on an Agilent 2100 BioAnalyzer using High Sensitivity DNA Reagents (ThruPLEX DNA-Seq HV kits). Concentrations of intact and sheared samples were measured using a Qubit 2.0 Fluorometer with the dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific).

Library preparation

Libraries were prepared according to the ThruPLEX DNA-Seq HV User Manual or ThruPLEX DNA-Seq HV PLUS User Manual. Amplified libraries were purified using AMPure XP (Beckman Coulter) and eluted in low-TE buffer for whole genome sequencing (WGS). Size of purified libraries was assessed by Agilent 2100 BioAnalyzer using High Sensitivity DNA Reagents. Libraries were quantified by Qubit 2.0 Fluorometer with Quant-IT dsDNA Assay Kit, high sensitivity (Thermo Fisher Scientific).

Illumina sequencing

Quantified post-PCR libraries were pooled and loaded onto an Illumina NextSeq® 500/550 Mid Output Kit v2.5 (150 Cycles) flow cell for sequencing. Libraries were loaded following Illumina's recommended loading concentrations.

Data analysis

Raw sequencing reads were downsampled to equal numbers across all samples using seqtk (v1.3-r106) and quality processed to remove adapters and low-quality bases using trimmomatic (v0.36). Quality processed reads were aligned to the UCSC hg19 reference genome with bowtie2 (v2.3.4.3) with default parameters. Resulting SAM files were sorted by coordinates using Picard SortSam (v2.18.3) and converted to BAM files with samtools view (v1.8). Duplicate reads were identified and marked from sorted BAM files with Picard MarkDuplicates (v2.18.3) and used as input to collect alignment, insert size, GC bias, and various WGS metrics with Picard AlignmentSummaryMetrics (v2.18.3), Picard CollectInsertSizeMetrics (v2.18.3), Picard CollectGcBiasMetrics (v2.18.3), and Picard CollectWgsMetrics (v2.18.3), respectively.

Related Products

Cat. #	Product	Size	License	Quantity	Details
R400783	ThruPLEX® DNA-Seq HV PLUS Kit	96 Rxns		*	
	ThruPLEX DNA-Seq HV PLUS integrates the size-tunable ThruPLEX HV PLUS Enzymatic Fragmentation Module into the simple, three-step ThruPLEX HV workflow to generate high-complexity DNA libraries from standard or challenging sample sources such as FFPE and cell-free DNA. This product contains reagents for 96 reactions and includes unique dual index (UDI) primers. Documents Components				
R400785	ThruPLEX® Tag-Seq HV PLUS Kit	96 Rxns		*	

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999