ThruPLEX HV: a simplified system for preparation of molecular-tagged NGS libraries from FFPE and cell-free DNA

Joseph Rotsinger,* Julie Laliberte, Kensuke Abe, Sudakshina Chakrabarty, Patrick Martin & Andrew Farmer

Takara Bio USA, Inc., Mountain View, CA 94043, USA *Corresponding author: joseph_rotsinger@takarabio.com

Abstract

Creating next-generation sequencing (NGS) libraries from FFPE and cell-free DNA is critical to developing clinical assays. The systems used for generating the libraries must have simple, streamlined workflows that can accommodate clinical research samples without compromising accuracy. To satisfy these requirements, we have developed a complete, fast, and modular NGS library prep system that enables accurate, reproducible sequencing readouts from challenging sample types. For ease of use and automatability, the ThruPLEX® HV system features optional, tunable fragmentation of intact genomic samples from blood, tissue, or other sources. The fragmentation approach does not require additional enzymatic steps and results in highly reproducible fragment sizes optimized for Illumina® platforms.

Since accurate measurement of low-frequency mutations is critical when looking for rare alleles in heterogeneous samples (e.g., tissue biopsies or plasma), the ThruPLEX HV system includes optional molecular tags to ensure the most accurate data possible. These distinct molecular tags are well-balanced to ensure optimal representation and accuracy, which are critical considerations when looking for rare alleles. By combining the included molecular tags with deeper sequencing, this system can provide a greater degree of accuracy than is otherwise attainable. Additionally, ThruPLEX HV was designed to accommodate a large input volume, which improves mutation detection by increasing the complexity of the input and eliminates the need to concentrate precious DNA samples prior to library preparation. A final important consideration for low-frequency mutations is achieving even coverage throughout the genome in order to ensure optimal read depth at all relevant loci. To facilitate the necessary even coverage, our system has been optimized across a broad range of GC content.

Here, we illustrate the robust performance of the system with human-derived cell lines, severely degraded FFPE DNA, cell-free DNA samples, and microbial genomic DNA.

1 Simplified prep: 1 tube, 2 hours, 3 steps

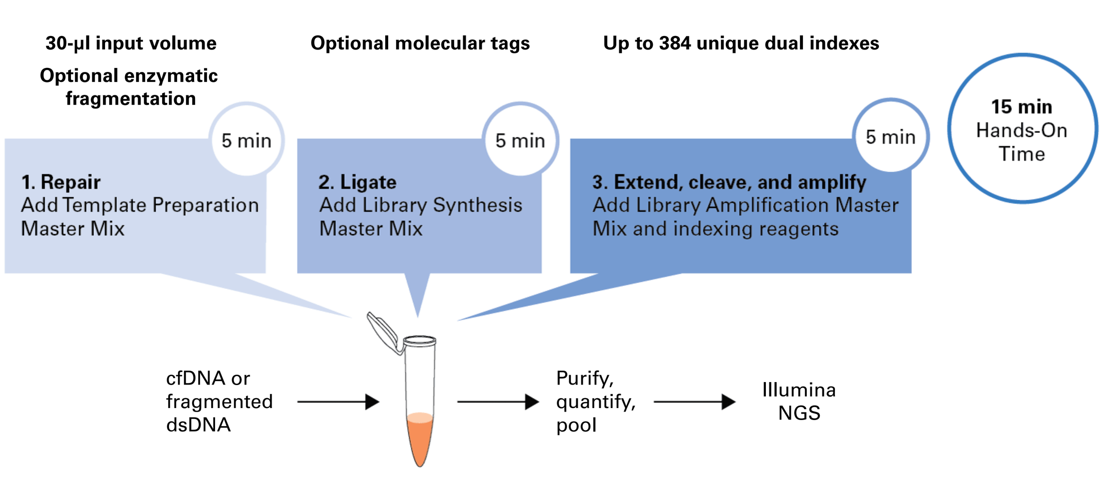


Figure 1. ThruPLEX HV single-tube library preparation workflow. The ThruPLEX HV workflow consists of three simple steps that take place in the same well or PCR tube, eliminating the need to purify or transfer the sample material. In this latest version of the technology, a higher input volume (30 µl) at the start of the protocol enables generation of higher complexity libraries and eliminates the need for sample concentration. An optional enzymatic fragmentation cocktail can be added to the initial repair step for simple, fast fragmentation of intact genomic samples. Adapters with molecular tags can easily be added at the ligation step. Unlike other ligation-based kits, the ligation step is not followed by a time-consuming bead purification step, helping save you time. Lastly, the kit will be available in configurations with up to 384 unique dual indexes for flexibility to work with any Illumina sequencer.

2 Tunable fragmentation without sonication saves time and increases throughput

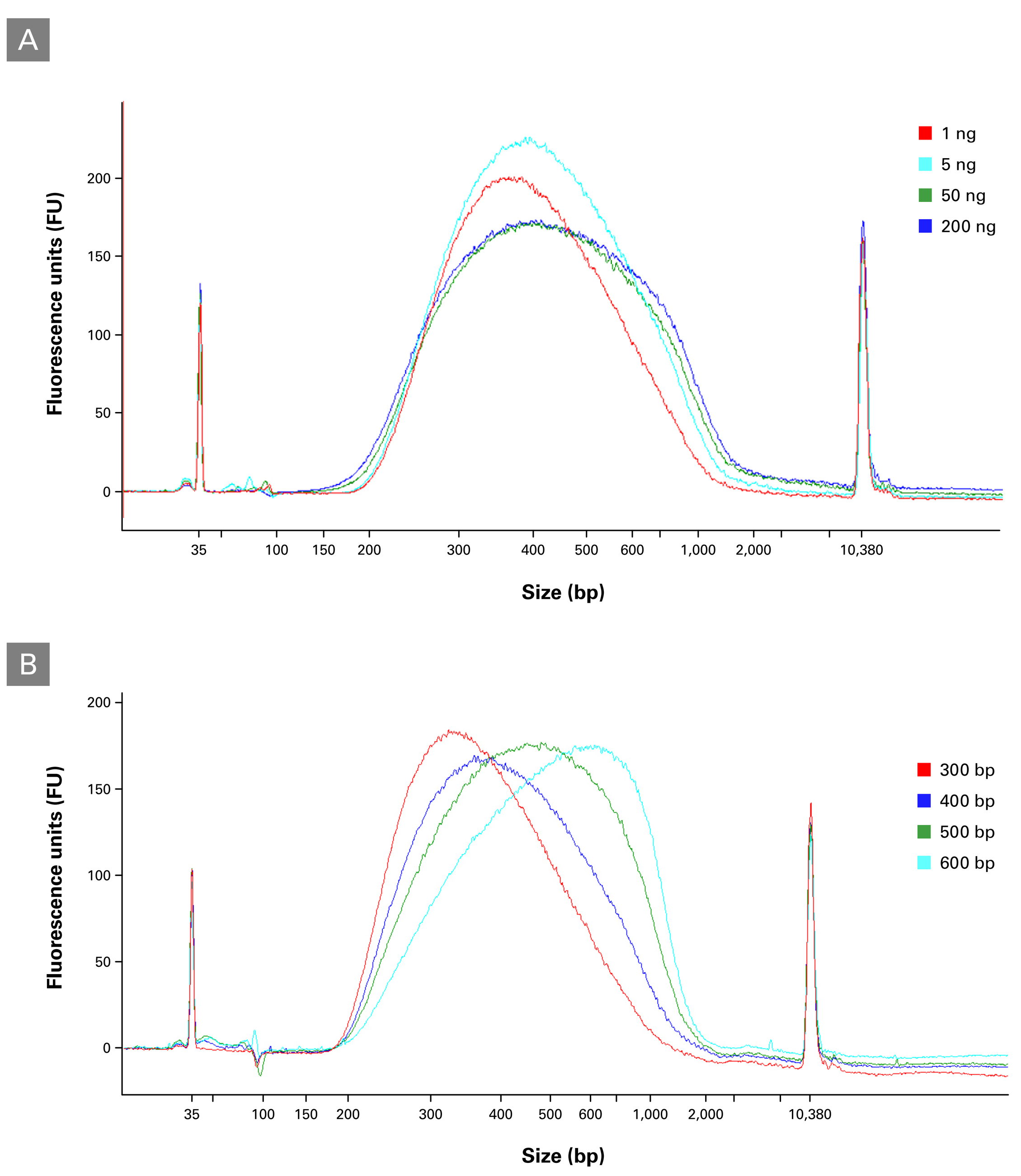


Figure 2: Size distribution of library fragments using ThruPLEX DNA-Seq HV with enzymatic fragmentation. Libraries were generated from Coriell Institute NA12878 human genomic DNA, using ThruPLEX DNA-Seq HV chemistry with enzymatic fragmentation. The size distribution of the library fragments were detected with an Agilent Bioanalyzer. **Panel A.** Size distribution of library fragments generated from 1-ng, 5-ng, 50-ng, and 200-ng inputs. **Panel B.** Size distribution of library fragments generated from a 10-ng input and different dilutions of the enzymatic fragmentation cocktail.

3 Consistent variant calls from highly degraded FFPE samples

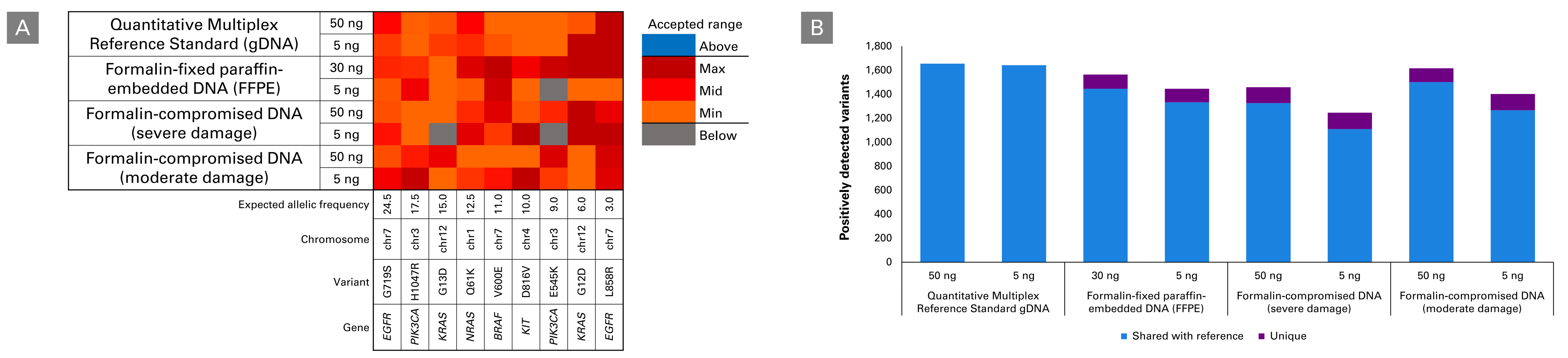
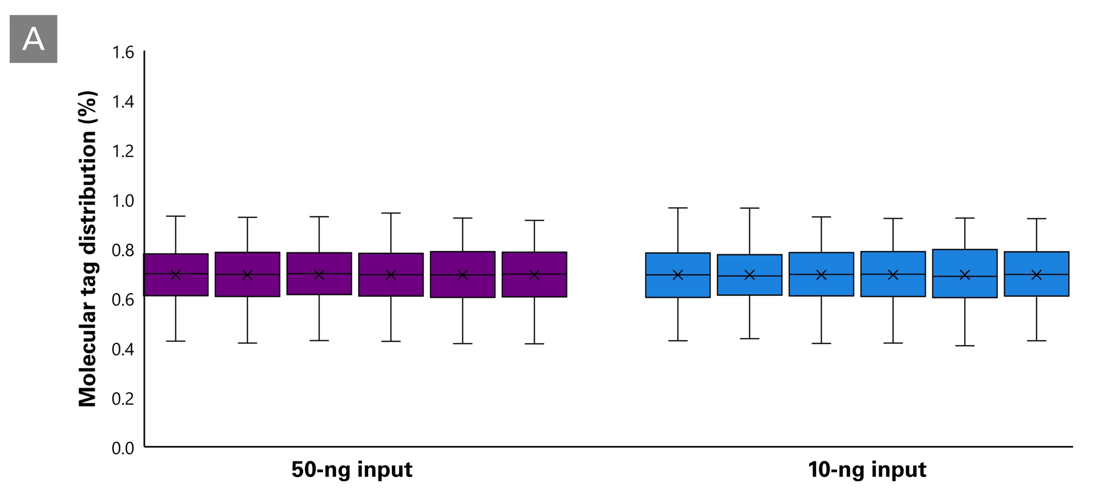


Figure 3: Positive detection of verified mutations using ThruPLEX DNA-Seq HV-generated libraries. Libraries were generated in triplicate with 50-ng and 5-ng inputs of Horizon Discovery DNA references including Quantitative Multiplex Reference Standard (gDNA; Cat. # HD701) and formalin-compromised material with moderate (Cat. # HD799) or severe (Cat. # HD803) damage, as well as with 30-ng and 5-ng inputs of formalin-fixed paraffin-embedded material (HD200). Libraries were amplified with ThruPLEX DNA-Seq HV, purified with AMPure beads, and pooled for target capture with the IDT xGEN Pan Cancer Panel (IDT, Cat. # 1056205) covering 800 kb of the human genome. Paired-end sequencing was performed with a NextSeq® 500/550 Mid Output Kit v2.5 (150 cycles). Each library was downsampled to 5 million total reads and aligned to hg19 using Bowtie 2. **Panel A.** Heat map of averaged replicate allele frequencies compared to Horizon-verified allele frequencies. The max and min acceptance range of allelic frequency denotes the accepted deviation range from the expected allele frequency (Mid) as provided by Horizon Discovery. The mean depth of coverage was 250X. **Panel B.** Positive variants detected with IDT's xGEN Pan Cancer Panel. Each bar represents variants called in 2 out of 3 or 3 out of 3 replicates. "Shared with reference" indicates these variants were also positively called in the reference control. "Unique" represents variants not called in the reference control.The mean depth of coverage was 200X.

4 Highly accurate detection of low-frequency mutations from cell-free DNA using optional molecular tags



B

				50-ng input						10-ng input					
Gene	Nucleotide	Amino acid	Cosmic ID	WT	WT	5%	5%	1%	1%	WT	WT	5%	5%	1%	1%
EGFR	2573T>G	L858R	COSM6224	0.0%	0.0%	6.5%	5.2%	1.3%	1.5%	0.0%	0.0%	5.7%	5.8%	1.3%	2.1%
EGFR	2369C>T	T790M	COSM6240	0.0%	0.0%	2.6%	5.6%	0.4%	1.3%	0.0%	0.0%	2.4%	1.9%	1.5%	1.9%
KRAS	35G>A	G12D	COSM521	0.0%	0.0%	4.0%	3.8%	1.3%	1.6%	0.0%	0.0%	3.5%	3.6%	1.5%	0.5%
NRAS	181C>A	Q61K	COSM580	0.0%	0.0%	8.6%	5.1%	2.4%	1.7%	0.0%	0.0%	6.4%	5.1%	2.5%	2.3%
PIK3CA	1633G>A	E545K	COSM763	0.0%	0.0%	4.2%	3.8%	0.9%	1.3%	0.0%	0.0%	4.2%	5.7%	miss	1.5%

Figure 4. Efficient variant detection from cell-free DNA. ThruPLEX Tag-Seq HV demonstrates excellent accuracy in detecting low-frequency variants down to 1% allele frequency. ThruPLEX Tag-Seq HV libraries were prepared from 10 and 50 ng of human cfDNA surrogate (Quan-Plex Patient-like ctDNA Panel, AccuRef, Cat. # ARF-1003CT). The libraries were pooled, and hybridization capture was performed using IDT xGen Pan-Cancer Panel v1.5 (IDT, Cat. # 1056205). The enriched libraries were sequenced with a NextSeq 500/550 High Output Kit v2.5 (150 cycles). The fastq files were downsampled to 8M reads, and the reads were then aligned to the human genome (hg19) using Bowtie 2. The alignment metrics and hybridization capture metrics were calculated using Picard tools. The average bait coverage was ~800X. Reads were grouped using fgbio and variants were called using VarDict. **Panel A**. Optimized, 144 discrete molecular tags show even distribution to ensure the broadest dynamic range and minimal bias in making variant calls. **Panel B.** Table of measured allele frequency and expected allele frequency using cell-free DNA surrogate input at 10 ng and 50 ng. Data shows an excellent correlation between the measured and expected allelic frequencies.

5 Simple, robust workflow ensures the right answer, every time

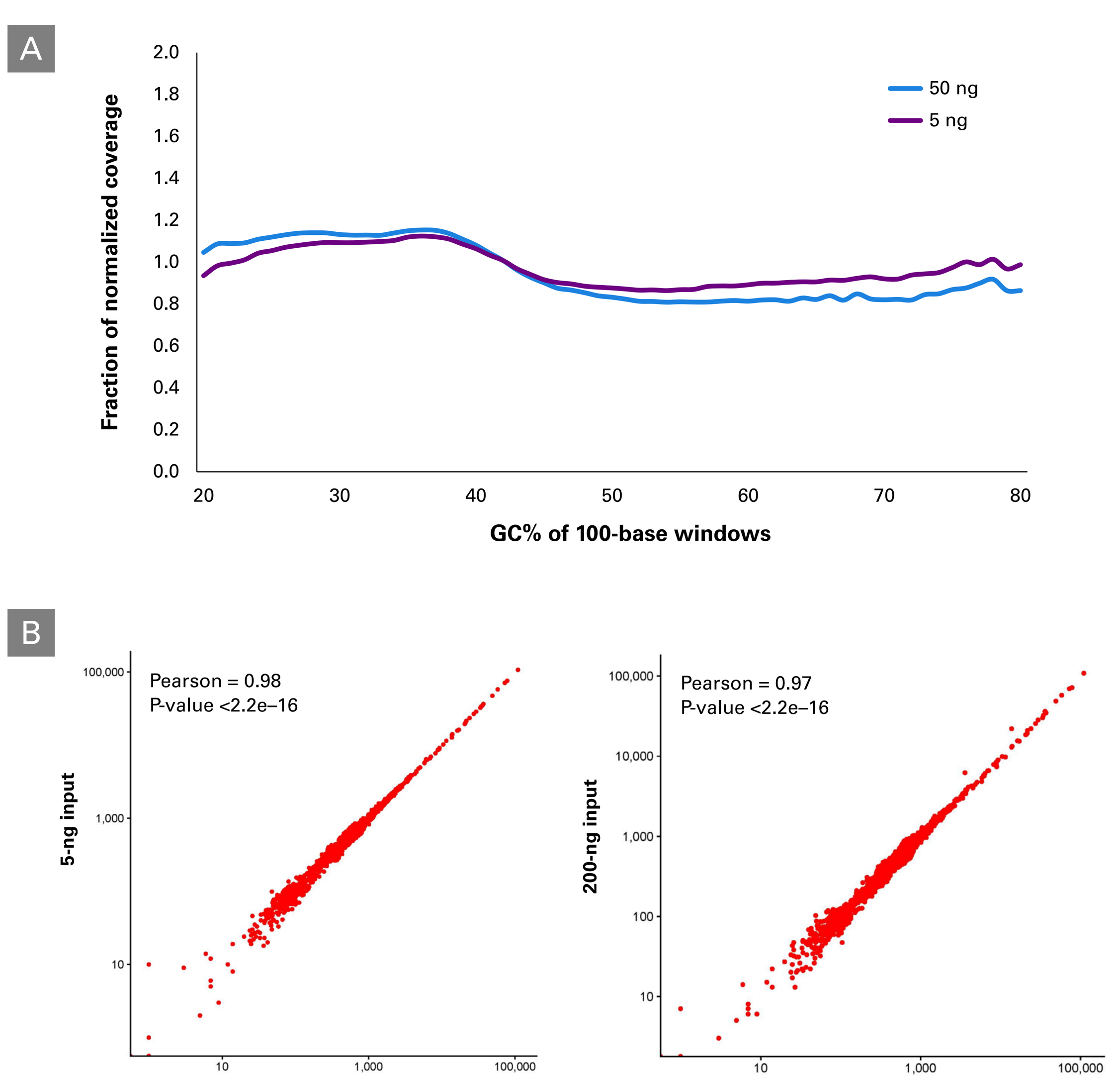
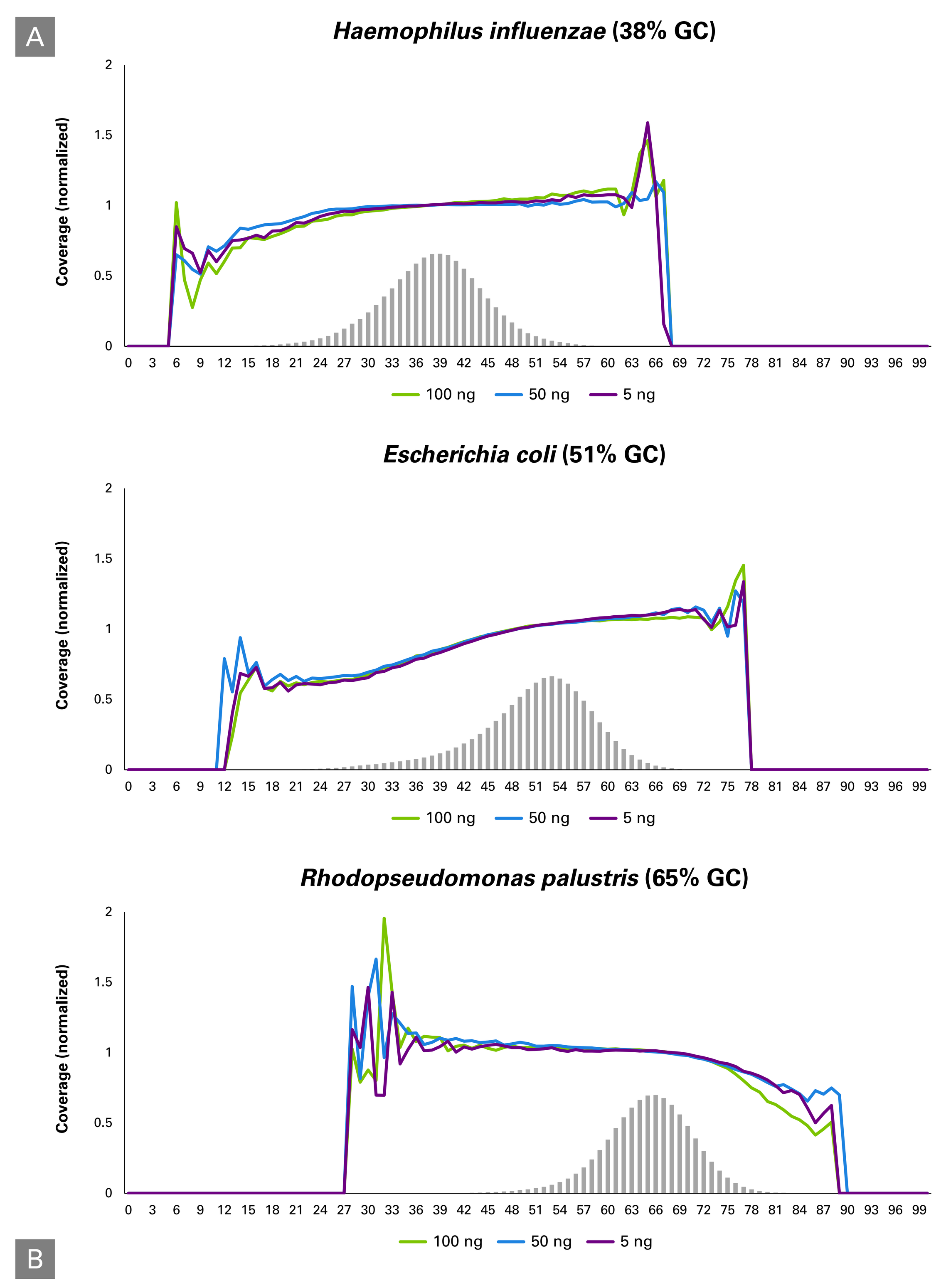


Figure 5. Ultra-reliable coverage and reproducibility. ThruPLEX DNA-Seq HV provides robust performance across a broad range of inputs and sample types. **Panel A.** Normalized genome coverage from Quantitative Multiplex Reference Standard (gDNA; Horizon Discovery) was calculated using CollectGcBiasMetrics (Picard). The distribution of reads across the genome is similar at inputs of 5 ng and 50 ng and shows good representation across a GC content range of 20–80%. **Panel B.** Correlation plots of replicate library preparations using ThruPLEX DNA-Seq HV. Comparison between two independent library preps (left) and between two different inputs (right) indicates high reproducibility of the system with FFPE DNA. Each library has 10M paired reads (20M total), and the bin size is 100 kb.

6 Reliable library prep regardless of input amount or microbial source



B

Input amount	Genome	Genome size	% GC	Total reads	% mapped	Mean coverage	% chimera	% duplication
100 ng	*H. influenzae*	1.83 Mb	38%	3.92×10^6	98.82%	132.1	3.18%	1.34%
	E. coli	4.7 Mb	51%	3.94×10^6	98.73%	52.2	3.91%	0.88%
	R. palustris	5.46 Mb	65%	3.94×10^6	98.79%	46.0	2.79%	0.91%
50 ng	*H. influenzae*	1.83 Mb	38%	3.51×10^6	99.00%	130.1	2.21%	1.34%
	E. coli	4.7 Mb	51%	3.94×10^6	98.96%	52.0	2.88%	0.90%
	R. palustris	5.46 Mb	65%	3.95×10^6	98.90%	45.8	2.17%	0.92%
5 ng	*H. influenzae*	1.83 Mb	38%	3.89×10^6	98.83%	131.3	2.60%	1.13%
	E. coli	4.7 Mb	51%	3.90×10^6	98.82%	51.6	2.41%	0.75%
	R. palustris	5.46 Mb	65%	3.91×10^6	98.69%	45.2	2.68%	0.74%

Figure 6. Robust performance from various input amounts and bacterial genomes. Libraries were amplified in triplicate with ThruPLEX DNA-Seq HV chemistry using 100-ng, 50-ng, and 5-ng inputs of *Haemophilus influenzae* 51907D-5, *Escherichia coli* 11303, or *Rhodopseudomonas palustris* BAA-98D-5 (ATCC). Libraries were purified with AMPure beads, and paired-end sequencing was performed with a MiSeq® Reagent Kit v3 (150 cycles). Each library was downsampled to 4M total reads and aligned to their respective genomes using Bowtie 2. **Panel A.** For each genome, GC coverage was calculated using Picard's CollectGCBiasMetrics. Normalized coverage is represented by the colored lines, while the expected GC content distribution using 100-bp windows is represented by the vertical gray bars. **Panel B.** Sequencing metrics representative of each library at their respective inputs.

Conclusions

- ThruPLEX HV is a simple, fast, accurate DNA-seq system with three addition-only steps that can be completed in two hours in one tube
- Optional, tunable enzymatic fragmentation does not add any steps or time to the library-prep workflow
- Well-balanced, discrete molecular tags are available to ensure the most accurate variant calling and copy number variation measurements
- A 30-µl input volume eliminates the need for sample concentration steps and enables higher complexity libraries
- The system excels in accurate variant calling from FFPE and cell-free DNA

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