

Streamlined DNA-seq from challenging samples

- Streamlined, single-tube workflow without the need to concentrate samples »
- Reproducible and uniform library coverage, regardless of input level »
- Improved performance for regions with increasing GC content »

Introduction

A high-performance NGS library preparation system must have a simple, streamlined workflow that can accommodate a range of sample inputs without compromising accuracy. [ThruPLEX DNA-Seq HV](#) satisfies these requirements, with a complete, fast, and accurate system that enables reproducible sequencing readouts from challenging sample types. ThruPLEX DNA-Seq HV was designed to accommodate a large input volume and a higher amount of starting material than our original ThruPLEX DNA-seq kits, which improves coverage and mutation detection by increasing the complexity of the sample input and eliminating the need to concentrate precious DNA samples prior to library preparation. A final important consideration for low-frequency mutations is the ability to achieve 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 for improved coverage uniformity across a broad range of GC content.

Single-tube ThruPLEX DNA-Seq HV workflow

The ThruPLEX DNA-Seq HV workflow consists of three simple steps of reagent addition that take place in a single well or PCR tube with just 15 minutes of hands-on time, and yields indexed libraries from fragmented DNA within two hours (Figure 1). The libraries generated can be used directly for whole-genome sequencing applications or enriched using a custom panel for the leading target enrichment platforms. The implementation of a single-tube workflow increases throughput and prevents the loss of precious samples by eliminating the need for time-consuming bead purification or transfer of sample material. Additionally, eliminating input amount-based adapter dilution further minimizes hands-on time and allows ThruPLEX DNA-Seq HV to be one of the quickest and most consistent library preparation kits when compared to others that are commercially available (Table 1).

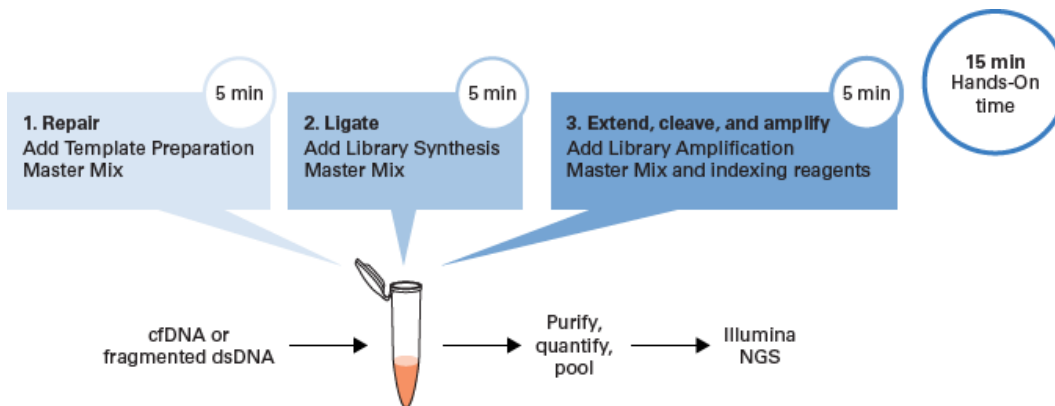


Figure 1. ThruPLEX DNA-Seq HV single-tube library preparation workflow. The ThruPLEX HV workflow consists of three simple steps that take place in the same well or PCR tube, eliminating the need to purify or transfer the sample material. In this latest version of the ThruPLEX technology, a higher input volume (30 μ l) and range of input amounts (up to 200 ng) at the start of the protocol enables the generation of higher-complexity libraries and eliminates the need for sample concentration.

| | ThruPLEX DNA-Seq HV | Kapa Hyper Prep | NEB Next Ultra II |
|------------------------------|---------------------|-----------------|-------------------|
| Hands-on time | 15 min | 20 min | 20 min |
| Total time | 2.4–2.6 hr | 2.5–2.7 hr | 3.1–3.2 hr |
| Single-tube workflow | Yes | No | No |
| Adapter dilution | No | Yes | Yes |
| Intermediate cleanup | No | Yes | Yes |
| Post-ligation size selection | No | No | Yes (>100 ng) |

Table 1. Comparison of three leading NGS library preparation chemistries. Total time is representative of the time required to amplify inputs of 5 ng and 200 ng with each chemistry to yield enough Illumina-compatible dual-indexed library for target enrichment. ThruPLEX DNA-Seq HV is the only single-tube workflow and the only chemistry which does not require adapter dilution, intermediate cleanup, or post-ligation size selection. The culmination of all these features is the quickest protocol with the least amount of hands-on time.

Results

Improved library preparation

Preparing NGS libraries from input material with a low starting concentration can lead to a low-complexity library pool. This complexity can diminish even further when hampered by a low input volume. The ThruPLEX DNA-Seq HV library preparation kit takes the coveted single-tube workflow of ThruPLEX DNA-seq and elevates the input volume from 10 to 30 μ l, while also increasing the input range to 200 ng.

Preparing NGS libraries from low-concentration samples makes PCR amplification essential in order to enrich a pool of sequenceable libraries and increase yield. Regions of high GC content contain strong secondary structures and can introduce bias by resisting denaturation and amplification. Introducing bias in the process of NGS library preparation can lead to low yields, uneven representation of coverage, and low coverage depth in regions of interest. Through reformulation and workflow optimization, ThruPLEX DNA-Seq HV ensures the accurate representation of the original material by removing bias and providing substantial improvements in coverage of regions with extreme base composition.

Uniform library coverage across input levels

Generating libraries which proportionally cover the complete sequence of a given sample is pivotal to creating an accurate representation of that input material. This becomes challenging when lower inputs are used, as the chances of uniformly covering the sample decreases. Additionally, at lower inputs, reproducibility is challenged. By consistently generating libraries of the recommended insert size for sequencing, the need for size selection is eliminated and the likelihood of achieving a higher base quality score is improved. The ThruPLEX DNA-Seq HV kit can generate sequenceable libraries with a consistent insert size across the input range (data not shown). ThruPLEX DNA-Seq HV also demonstrates coverage uniformity across the input range and coverage uniformity of replicates at the kit's lowest recommended input (Figure 2).

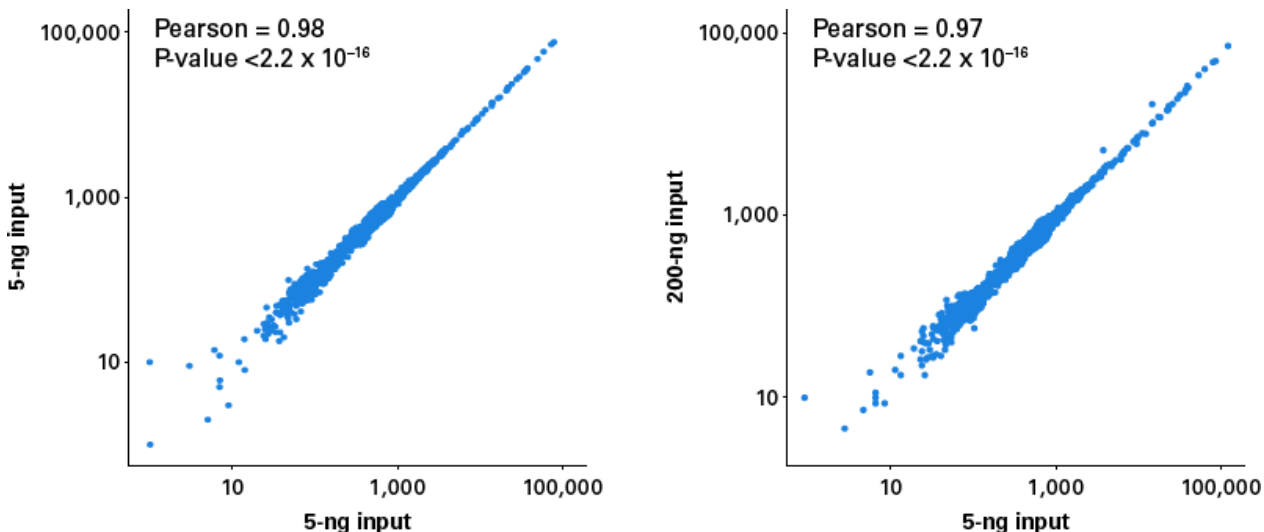
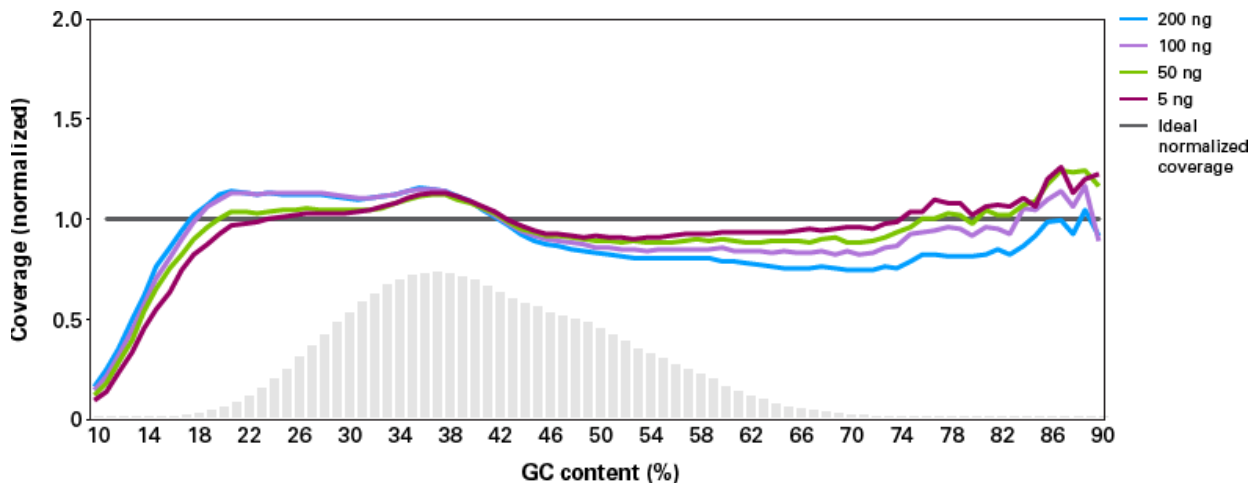


Figure 2. Reproducibility and uniform coverage across input levels. Correlation plots are shown for replicate library preparations generated by ThruPLEX DNA-Seq HV with 5, 50, and 200 ng of EMD Millipore gDNA and downsampled to 20 million total reads. Coverage of each 10-kb region of hg19 was compared across inputs. Comparison of two independent 5-ng library preps (left), and two different starting inputs of 5 and 200 ng (right), demonstrate the high reproducibility of the system.

Uniform GC coverage

While consistent, uniform coverage can be challenging, the robustness of a library preparation kit depends on its ability to accurately and uniformly cover a variety of challenging genomes with varying GC content. ThruPLEX DNA-Seq HV is able to produce consistent GC coverage spanning the input range for gDNA (Figure 3), as well as inputs with a variety of GC contents (Figure 4).

A

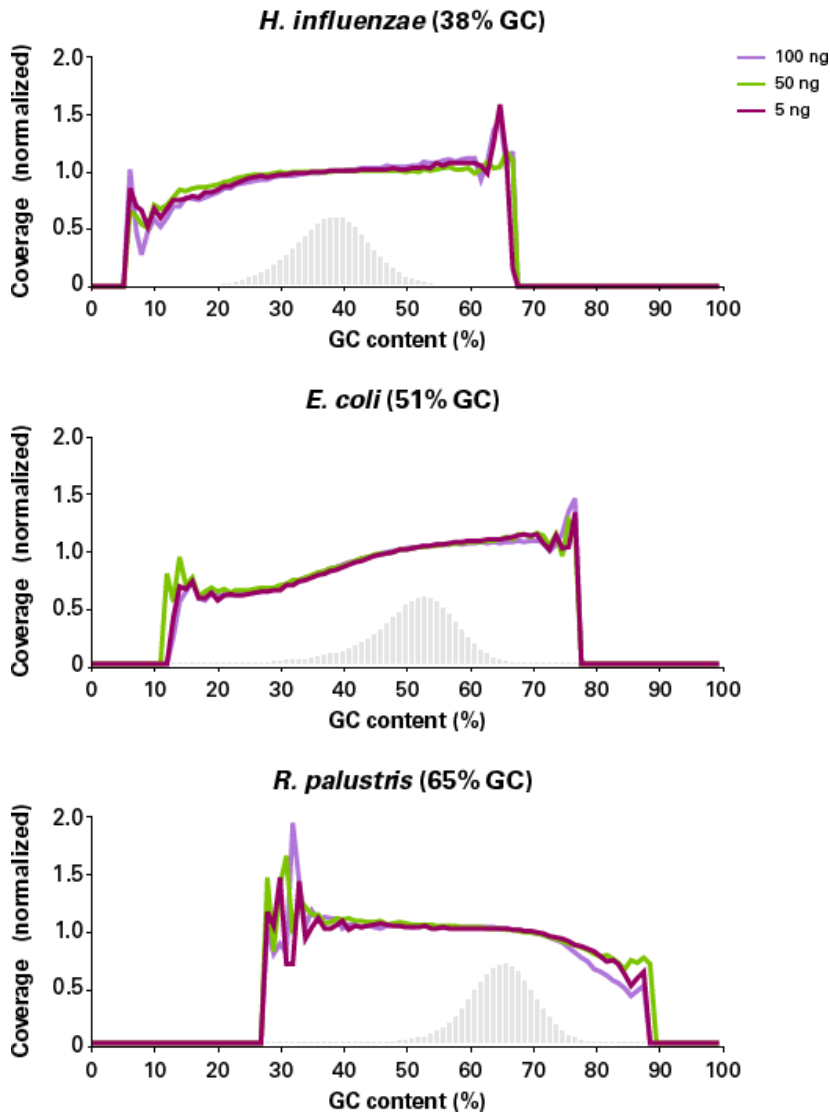


B

| Input | PCR cycles | Average yield (ng) | Total reads aligned | % reads aligned | % chimera | % duplicates |
|--------|------------|--------------------|---------------------|-----------------|-----------|--------------|
| 200 ng | 6 | 992.5 | 2.53×10^7 | 96.17% | 2.17% | 1.42% |
| 100 ng | 7 | 850.0 | 2.39×10^7 | 96.23% | 1.77% | 1.42% |
| 50 ng | 8 | 750.0 | 2.78×10^7 | 96.12% | 1.59% | 1.44% |
| 5 ng | 12 | 882.5 | 2.70×10^7 | 95.80% | 1.35% | 2.06% |

Figure 3. Consistent GC coverage across inputs. Panels A and B. Libraries were prepared in duplicate from 5, 50, 100, and 200 ng of gDNA from EMD Millipore. Libraries were generated following the ThruPLEX DNA-Seq HV protocol. Paired-end sequencing was performed on a NextSeq® 150 Cycle Mid Output (2 x 75 bp). The vertical gray bars represent the relative number of 100-bp regions at each GC%.

A



B

| Input | Genome | Total reads aligned | % mapped | Mean coverage | % chimera | % duplication |
|--------|----------------------|------------------------|----------|---------------|-----------|---------------|
| 100 ng | <i>H. influenzae</i> | 3.92 x 10 ⁶ | 98.82% | 132.1 | 3.18% | 1.34% |
| | <i>E. coli</i> | 3.94 x 10 ⁶ | 98.73% | 52.2 | 3.91% | 0.88% |
| | <i>R. palustris</i> | 3.94 x 10 ⁶ | 98.79% | 46.0 | 2.79% | 0.91% |
| 50 ng | <i>H. influenzae</i> | 3.92 x 10 ⁶ | 99.00% | 130.1 | 2.21% | 1.34% |
| | <i>E. coli</i> | 3.94 x 10 ⁶ | 98.96% | 52.0 | 2.88% | 0.90% |
| | <i>R. palustris</i> | 3.94 x 10 ⁶ | 98.90% | 45.8 | 2.17% | 0.92% |
| 5 ng | <i>H. influenzae</i> | 3.92 x 10 ⁶ | 98.83% | 131.3 | 2.60% | 1.13% |
| | <i>E. coli</i> | 3.94 x 10 ⁶ | 98.82% | 51.6 | 2.41% | 0.75% |
| | <i>R. palustris</i> | 3.94 x 10 ⁶ | 98.69% | 45.2 | 2.68% | 0.74% |

Figure 4. Uniform coverage in libraries with extreme base content. Panels A and B. Libraries were amplified in triplicate with ThruPLEX HV chemistry

using 100-, 50-, and 5-ng inputs of *Haemophilus influenzae* 51907D-5, *Escherichia coli* 11303, or *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 1.85 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 5.45 Mb in size, with 65% GC content.

Improved library coverage with the HV protocol

Libraries were generated from gDNA, with input amounts that span the overlapping input ranges of ThruPLEX DNA-Seq HV and ThruPLEX DNA-seq kits, using their respective protocols. The ThruPLEX DNA-Seq HV libraries show substantially better coverage at both inputs, as indicated by the improvements in uniform coverage (Table 2). These improvements in uniform coverage were made without sacrificing mapping, chimera, or duplication rates.

| | Input | Total reads aligned | %reads aligned | %chimera | %duplicate |
|---------------------|-------|------------------------|----------------|----------|------------|
| ThruPLEX DNA-Seq HV | 50 ng | 7.06 x 10 ⁶ | 96.6% | 3.05% | 0.73% |
| | 5 ng | 7.56 x 10 ⁶ | 96.9% | 2.47% | 0.84% |
| ThruPLEX DNA-Seq | 50 ng | 7.71 x 10 ⁶ | 97.1% | 3.46% | 0.78% |
| | 5 ng | 7.69 x 10 ⁶ | 97.0% | 2.76% | 0.85% |

Table 2. Improved library coverage uniformity with ThruPLEX HV technology. Libraries were prepared in triplicate from 5 and 50 ng of a quantitative multiplex reference standard consisting of gDNA pooled from HCT116, RKO, and SW48 cell lines from Horizon Discovery. Libraries were generated following ThruPLEX DNA-Seq HV or ThruPLEX DNA-seq protocols. Paired-end sequencing was performed on a NextSeq 150 Cycle Mid Output (2 x 75 bp) and total reads were downsampled to 8 million.

Conclusions

The [ThruPLEX DNA-Seq HV](#) library preparation kit for Illumina sequencing platforms elevates the ThruPLEX family by increasing starting input volume and expanding the amount of starting material when compared to previous ThruPLEX DNA-seq kits. Along with these improvements, this advanced system retains the coveted single-tube workflow with no intermediate cleanup, which is synonymous with ThruPLEX DNA-seq technology. Through workflow optimization and reformulation, this kit provides substantial improvements in coverage of regions with increasing GC content. ThruPLEX DNA-Seq HV is a simple, fast, and accurate system with three addition-only steps that can be completed in a single tube in just two hours.

Methods

DNA preparation

Human genomic DNA from EMD Millipore (69237) or Horizon Discovery (HD701) and bacterial genomic DNA from *Haemophilus influenzae* 51907D-5, *Escherichia coli* 11303, or *Rhodopseudomonas palustris* BAA-98D-5 (ATCC) were sheared on a Covaris M220 following the 250-bp shearing protocol. Sheared input material was evaluated for correct size on an Agilent 2100 BioAnalyzer using High Sensitivity DNA Reagents. The concentration of these samples was measured using a Qubit 2.0 Fluorometer with the dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific).

Library preparation

Libraries were prepared as per the manufacturer's instructions using the ThruPLEX DNA-Seq HV kit or the ThruPLEX DNA-Seq Kit. All libraries were generated using dual indexes. Amplified libraries were pooled and then purified using AMPure XP beads (Beckman Coulter) and eluted in low TE buffer for whole-genome sequencing (WGS). Purified libraries' size was assessed on the Agilent 2100 BioAnalyzer using High Sensitivity DNA Reagents. Libraries were quantified by qPCR using Takara Bio's Library Quantification Kit, or Qubit 2.0 Fluorometer with the dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific).

Illumina sequencing

Quantified post-PCR libraries were pooled and loaded onto an Illumina MiSeq® V3 or NextSeq 500/550 v2.5 flow cell for sequencing, following Illumina-recommended loading concentrations.

Data analysis

For WGS data, reads were merged using FastQC and downsampled to equal numbers across all samples using seqtk. Adapters were trimmed using trimmomatic and each library was aligned to hg19 using Bowtie 2 and converted to sam format. Sam files were converted to bam files using samtools. Library metrics were collected from bam files using the following: Picard Tools, MarkDuplicates, AlignmentSummaryMetrics, CollectInsertSizeMetrics, CollectGcBiasMetrics, and CollectWgsMetrics.

References

Chen Y. C., Liu T., Yu C. H., Chiang T. Y., Hwang C. C. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS One* **8**, e62856 (2013).

Tan, G., Opitz, L., Schlapbach, R. *et al.* Long fragments achieve lower base quality in Illumina paired-end sequencing. *Sci. Rep.* **9**, 2856 (2019).

Related Products

| Cat. # | Product | Size | License | Quantity | Details |
|--|----------------------|---------|---------|----------|---------|
| R400740 | ThruPLEX® DNA-Seq HV | 96 Rxns | | * | |
| <p>ThruPLEX DNA-Seq HV uses a simple, three-step workflow to generate high-complexity DNA libraries from standard samples and challenging sample sources such as FFPE and cell-free plasma DNA. This product contains reagents for 96 reactions and includes unique dual indexing (UDI) primers.</p> | | | | | |
| <div style="display: flex; justify-content: space-around;"> Documents Components </div> | | | | | |
| R400741 | ThruPLEX® DNA-Seq HV | 24 Rxns | | * | |

[Add to Cart](#)

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

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES. © 2020 Takara Bio Inc. All Rights Reserved. All trademarks are the property of Takara Bio Inc. or its affiliate(s) in the U.S. and/or other countries or their respective owners. Certain trademarks may not be registered in all jurisdictions. Additional product, intellectual property, and restricted use information is available at takarabio.com.