APPLICATION NOTE

Illumina DNA Prep crude lysate protocol for NGS

Fast, accurate metagenomic profiling and high-quality *de novo* genome assembly data

illumina®

Introduction

Next-generation sequencing (NGS) technology has become an important tool in metagenomics research for organism identification, metagenomic profiling, and *de novo* assembly.[1,2] Illumina has introduced several workflow advances that support faster and more efficient library preparation, including tagmentation chemistry to consolidate DNA fragmentation and adapter tagging steps into a single reaction. With innovative on-bead tagmentation, Illumina DNA Prep shortens the library prep workflow by integrating DNA extraction, quantification, fragmentation, and library normalization steps.[3]

Though NGS–based whole-genome sequencing (WGS) has provided significant advantages in speed, accuracy, and depth of information to microbiology labs, DNA extraction and library preparation steps have remained a significant bottleneck in the NGS workflow. Preparation of NGS libraries for metagenomic studies typically begins with time-consuming and labor-intensive genomic DNA extraction steps. To address this challenge in metagenomics, Illumina offers Illumina DNA Prep and the Illumina Crude Lysate Protocol. This method supports quick and easy library preparation directly from crude lysate. Sequencing directly from crude lysate eliminates the time and cost associated with DNA extraction steps. In addition to increased speed and efficiency, Illumina DNA Prep offers exceptional flexibility for sample input type and cell lysis methods, including direct bacterial colonies, blood, and saliva.

This application note demonstrates the exceptional performance of Illumina DNA Prep* and the Illumina Crude Lysate Protocol using mock microbial communities and real-world human stool samples. Results show equivalent or better performance with crude lysates compared to extracted DNA and with Illumina DNA Prep compared to the Nextera™ DNA Library Preparation Kit in metagenomic profiling and *de novo* assembly.

* Data presented in this application note was generated using the Nextera DNA Flex Library Prep Kit, which has been replaced by Illumina DNA Prep. These two kits use the same tagmentation chemistry and have identical product performance specifications and kit configurations.

Methods

Sample sources

To represent mock microbial communities, 20 Strain Even Mix Genomic Material (ATCC, Catalog no. MSA-1002), 20 Strain Staggered Mix Genomic Material (ATCC, Catalog no. MSA-1003), and 20 Strain Even Mix Whole Cell Material (ATCC, Catalog no. MSA-2002) were used. The 20 Strain Even/Staggered Mix Genomic Material mixtures are composed of purified DNA extracted from 20 bacterial strains, while the 20 Strain Even Mix Whole Cell Material is a mixture of whole (unlysed) cells. The species in these mixtures were selected for their diverse genome size, range of GC content, and Gram stain profile. Additionally, the microorganisms in the 20 strain mixtures have fully sequenced and characterized genomes. To represent real-world metagenomic samples, stool samples were obtained from both healthy donors and patients undergoing drug treatment.

Crude lysate and DNA extraction from mock microbial communities

To prepare crude lysate, 200 µl from the 20 Strain Even Mix Whole Cell sample were processed as described in the Illumina Crude Lysate Protocol with the PureLink Microbiome DNA Purification Kit (Thermo Fisher Scientific, Catalog no. A29790). Supernatant from step H of the PureLink Microbiome DNA Purification Kit protocol, which includes the transfer of supernatant from homogenized cells to a clean tube, was used as crude lysate.[4] To prepare extracted DNA, an equivalent number of cells was processed with the PureLink Microbiome DNA Purification Kit using the full protocol. The Illumina Crude Lysate Protocol is compatible with various bacterial DNA extraction kits (Table 1).

Crude lysate and DNA extraction from stool samples

Stool samples (0.05 g) from donor patients were used to generate crude lysates and extracted DNA as described in the Illumina Crude Lysate Protocol from Stool Samples with the PureLink Microbiome DNA Purification Kit. Supernatant from step G of the PureLink Microbiome DNA Purification Kit protocol, which includes the transfer of supernatant from homogenized cells to a clean tube, was used as crude lysate.[5]

To prepare extracted DNA, an equivalent amount of starting material was processed with the PureLink Microbiome DNA Purification Kit following the full protocol. Stool samples can be challenging due to the presence of PCR inhibitors such as bile, salts, and polysaccharides. Inhibited library preps show an inability to pellet during clean-up steps or they may deliver low library yield. Decreasing the amount of lysate below the recommended amount may be needed for some highly inhibitory samples. The Illumina Crude Lysate Protocol is compatible with various stool sample purification kits (Table 1).

Table 1: DNA extraction and lysate prep kits

Kits for bacterial samples
PureLink Microbiome DNA Purification Kit (Thermo Fisher Scientific)
UltraClean Microbial DNA Isolation Kit (MO BIO Laboratories)
ChargeSwitch gDNA Mini Bacteria Kit (Thermo Fisher Scientific)
Kits for stool samples
PureLink Microbiome DNA Purification Kit (Thermo Fisher Scientific)
PowerSoil DNA Isolation Kit (MOBIO Laboratories)
PowerFecal DNA Isolation Kit (MOBIO Laboratories)
QIAamp DNA Soil mini kit (QIAGEN)

For detailed information about recommended collection and lysis methods to obtain crude lysate with these kits, and the recommended input amount of crude lysate, download the Illumina Crude Lysate Protocol Guide.

Library preparation and sequencing

To compare the accuracy and sensitivity of sequencing from crude lysate to extracted DNA, libraries were prepared with 5 µl lysate or 5 µl extracted DNA from the 20 Strain Even Mix Whole Cell sample or from stool samples. The lysate and extracted DNA samples were input to library prep using either the Illumina Crude Lysate Protocol or the standard protocol. Library preparation using the Illumina Crude Lysate Protocol significantly reduced the total amount of time and the number of touch points in the library preparation workflow (Figure 1). Using crude lysate as a direct input into Illumina DNA Prep also eliminated the need for costly, time-consuming DNA quantification kits.

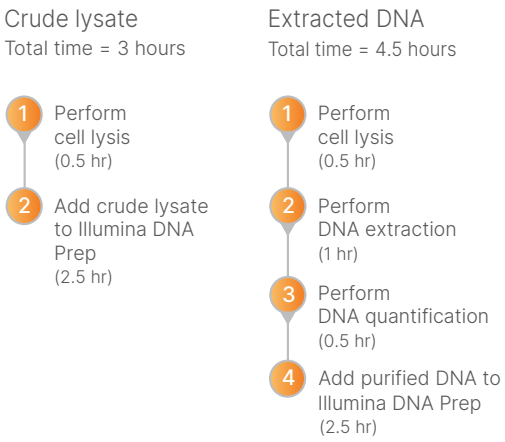

Figure 1: Workflow comparison—The Illumina Crude Lysate Protocol reduces the amount of time and total number of touch points required for library preparation. Workflow times were calculated assuming specific methods: DNA extraction (PureLink Microbiome DNA Purification Kit), DNA Quantification (Qubit). Times may vary depending on equipment used, kits used, sample batch number, automation procedures, and user experience.

To demonstrate data concordance between Illumina DNA Prep and the original Nextera DNA Library Preparation Kit (Illumina, Catalog no. FC-121-1030), 50 ng DNA from the 20 Strain Even Mix Genomic Material sample and 50 ng DNA from the 20 Strain Staggered Mix Genomic Material sample were added directly to the library prep kits. Libraries were prepared using the standard protocol.

All libraries were sequenced on the NextSeq™ 550 Sequencing System with a run configuration of 2 × 150 bp paired-end reads using the NextSeq 500/550 High Output v2 Kit (Illumina, Catalog no. FC-404-2004).

Data analysis

Metagenomic profiling metrics, including true positives, relative abundance, false positives, observed abundance, and expected abundance, were calculated on the One Codex platform.[6] Genome assembly bar graphs were generated with two software tools. First MEGAHIT v1.1.1.[7] performed *de novo* alignment with the FASTQ reads to generate contigs, then the contigs were assessed with Quality Assessment Tool for Genome Assemblies (QUAST)[8] with the known reference genomes. Metagenomic profiling stacked bar graphs were compiled with the DRAGEN™ Metagenomics pipeline.[9] The DRAGEN Metagenomics pipeline can be accessed in BaseSpace™ Sequence Hub, the Illumina genomics cloud-computing platform.

Results

Metagenomic profiling with mock community mixtures and stool samples

To establish the accuracy and sensitivity of Illumina DNA Prep in basic organism identification, a sequencing library was prepared with the 20 Strain Staggered Mix sample. Metagenomic profiling summary metrics—true positive, relative abundance, and false positives—were calculated on the One Codex platform. The results demonstrate excellent metagenomic profiling performance: all 20 organisms in the 20 Strain Staggered Mix were identified with no false positives (Table 2). The composition statistics with the 20 Strain Staggered Mix were also highly accurate with the observed abundance scores closely matching the expected abundance scores for all 20 species (Table 3). Furthermore, the observed abundance scores spanned four orders of magnitude and were able to detect species present in the sample down to 0.018%, which demonstrates high sensitivity with the Illumina DNA Prep sequencing workflow.

Workflow performance comparison with mock microbial communities

To compare the performance of the Illumina Crude Lysate Protocol to the standard protocol, sequencing libraries were prepared with crude lysate and extracted DNA from the 20 Strain Even Mix Whole Cell Material sample. Libraries made with crude lysate and extracted DNA generated equivalent metagenomic profiling summary statistics (Table 4).

Table 2: 20 Strain Staggered Mix metagenomic profiling summary statistics[a,b]

Library	True positives	Relative abundance	False positives
20 Strain Staggered Mix and Illumina DNA Prep	100%	100%	0

a. One Codex defines statistics as follows: True positives, percentage of organisms present (detected within two logs of the true abundance) in the control; relative abundance, the Pearson correlation coefficient between the known input organism abundances and the detected abundances (based on genome-size adjusted read counts); false positives, 100% less 10 percentage points for each "High" abundance false positive, 5 points for each "Moderate" one, and 1 point for each "Low" one; "trace" false positives do not count against the score, and the minimum possible score is 0%.
b. Data set down sampled from 20 million reads to 1 million reads.

Table 3: 20 Strain Staggered Mix metagenomic profiling composition statistics

Organism name	Observed abundance	Expected abundance
Rhodobacter sphaeroides	19.07%	18.00%
Escherichia coli	18.54%	18.00%
Staphylococcus epidermidis	18.26%	18.00%
Porphyromonas gingivalis	17.17%	18.00%
Streptococcus mutans	17.61%	18.00%
Pseudomonas aeruginosa	1.93%	1.80%
Clostridium beijerinckii	1.80%	1.80%
Bacillus cereus	1.15%	1.80%
Staphylococcus aureus	1.64%	1.80%
Streptococcus agalactiae	1.77%	1.80%
Acinetobacter baumannii	0.18%	0.18%
Propionibacterium acnes	0.21%	0.18%
Neisseria meningitidis	0.21%	0.18%
Lactobacillus gasseri	0.19%	0.18%
Helicobacter pylori	0.19%	0.18%
Bacteroides vulgatus	0.015%	0.018%
Enterococcus faecalis	0.021%	0.018%
Deinococcus radiodurans	0.017%	0.018%
Bifidobacterium adolescentis	0.013%	0.018%
Actinomyces odontolyticus	0.007%	0.018%

Table 4: Comparison of crude lysate and extracted DNA metagenomic profiling summary data[a]

Library	True positives	Relative abundance	False positives
Crude lysate	100%	33%	0
Extracted DNA	100%	32%	0

a. Data set down sampled from 20 million reads to 5 million reads.

Using the same crude lysate and extracted DNA data set, the *de novo* genome assembly quality was compared and the percentage of genome fraction assembled was calculated with QUAST. In general, a higher fraction of genome assembled is indicative of a higher quality genome assembly. For all 20 organisms analyzed, the crude lysate and extracted DNA libraries generated similar genome assembly results (Figure 2).

Workflow performance comparison with stool samples

The metagenomic profiles resulting from the Illumina Crude Lysate Protocol and the extracted DNA protocol are similar even with highly inhibitory, challenging stool samples (Figure 3). Unlike the mock microbial community samples, the stool samples contain a mixture of unknown species and an unknown species composition (percentage of each species). For the *de novo* assembly of metagenomics samples with unknown species, cumulative assembled length can be used to assess the quality of genome assembly. Assembly using QUAST revealed that the cumulative assembled lengths are highly similar between libraries prepared with the Illumina Crude Lysate Protocol from stool samples and the extracted DNA protocol (Figure 4).

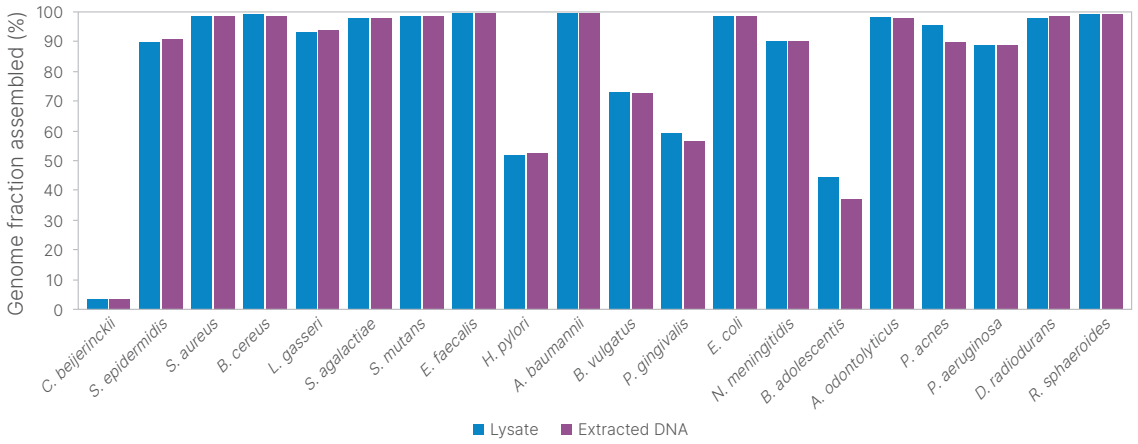

Figure 2: Comparison of crude lysate and extracted DNA genome assembly—Assembly was performed with MEGAHIT using 5 million, paired-end, 150-bp reads. Bar graph illustrates genome assembly performance for 20 organisms as reported by QUAST.

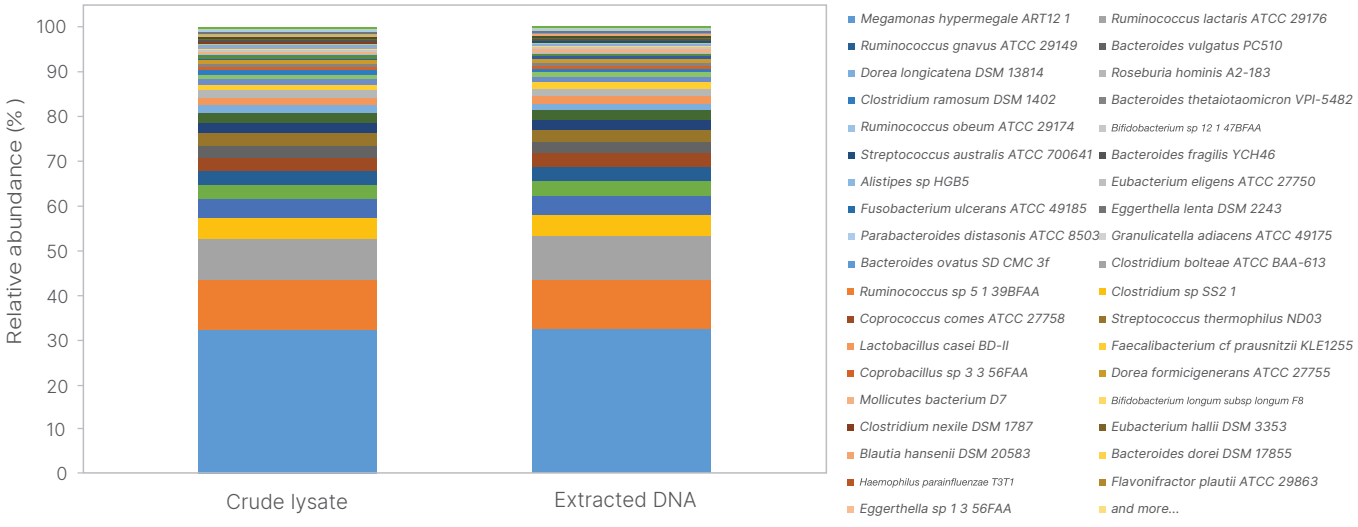

Figure 3: Comparison of crude lysate and extracted DNA metagenomic profiles—The abundance of each organism identified by the DRAGEN Metagenomics pipeline (using 3 million, paired-end, 150 bp reads) for the > 60 organisms identified in both stool libraries.

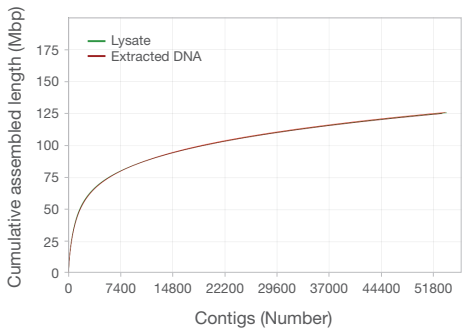

Figure 4: Protocol comparison of assembly with unknown species—Cumulative assembled length (Mbp) with the contigs ordered from largest to smallest. 20 million reads were used for assembly.

Illumina DNA Prep outperforms older kits for metagenomic profiling

The profiling performance of Illumina DNA Prep and the Nextera DNA Library Prep Kit were compared with libraries prepared from the 20 Strain Even Mix and the 20 Strain Staggered Mix Genomic Material samples. Summary metrics were calculated on the One Codex platform and the results demonstrate that Illumina DNA Prep delivers superior metagenomic profiling results compared to the Nextera DNA Library Prep Kit (Table 5). Unlike older kits, Illumina DNA Prep uses advanced on-bead tagmentation, which delivers uniform library yield, consistent insert sizes, and uniform genome coverage. These advantages work together to produce improved metagenomic profiling.

Table 5: Comparison of Illumina DNA Prep and Nextera library prep metagenomic profiling summary data

Library[a]	True positives	Relative abundance	False positives
20 Strain Even Mix Genomic Material			
Illumina DNA Prep	100%	97%	0
Nextera DNA Library Prep	100%	83%	0
20 Strain Staggered Mix Genomic Material			
Illumina DNA Prep	100%	100%	0
Nextera DNA Library Prep	100%	96%	0

a. 50 ng of DNA was input to library prep.

Illumina DNA Prep improves genome assembly compared to Nextera DNA Library Prep

Beyond assessments of organism detection and metagenomic profiling, the sequencing data were analyzed to compare genome assembly metrics. To evaluate genome assembly quality, the genome fraction assembly metrics were calculated by QUAST. The QUAST data show that Illumina DNA Prep outperformed the Nextera DNA Library Prep Kit for most of the 20 organisms in the 20 Strain Even Mix Whole Cell Material sample, especially for the AT rich organisms (Figure 5).

Crude lysate and extracted DNA libraries were prepared from stool samples using Illumina DNA Prep and the Nextera DNA Library Prep Kit. The metagenomics profiles resulting from the two library prep kits are similar, indicating that both provide excellent and comparable profile results (Figure 6).

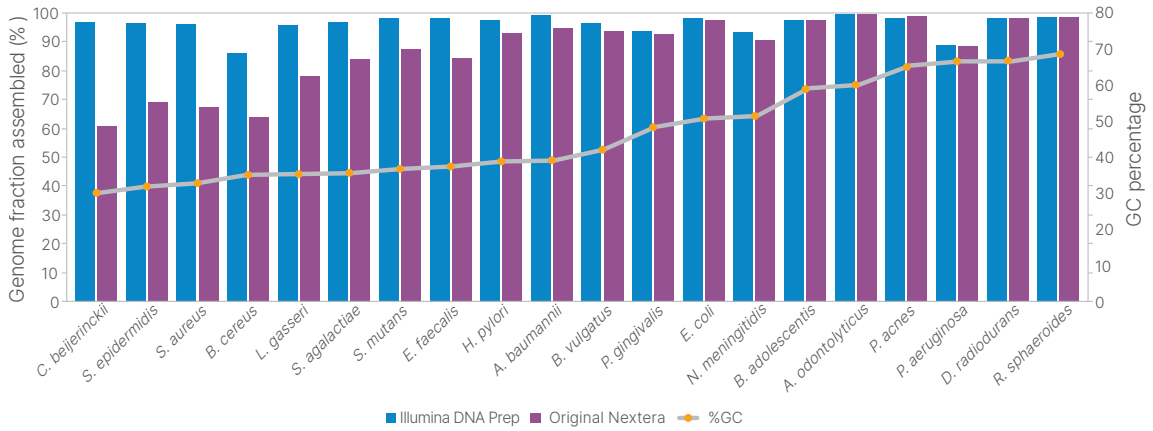

Figure 5: Comparison of Illumina DNA Prep and Nextera genome assembly—Genome fraction assembled with MEGAHIT (using 2 million, paired-end, 150 bp reads) for all 20 organisms. Organisms are listed in the order of increasing GC% from left to right.

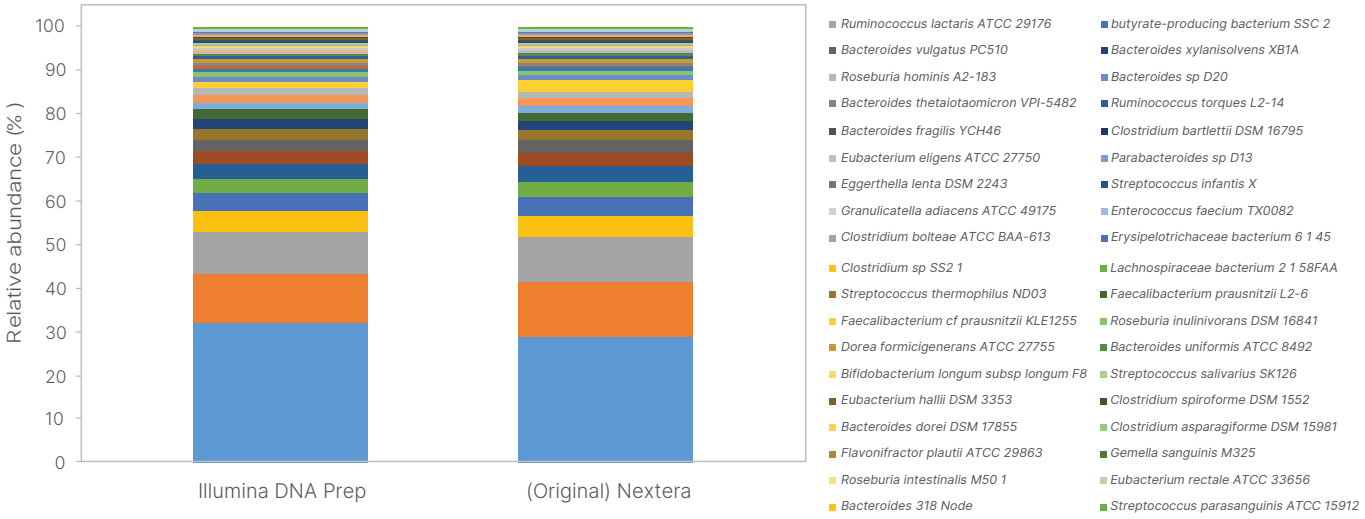

Figure 6: Comparison of Illumina DNA Prep and Nextera metagenomic profiles—The abundance of each organism identified by GENIUS metagenomics analysis (3 million, paired-end, 150 bp reads were used for analysis) for the > 60 organisms identified in both libraries.

Summary

Illumina DNA Prep combined with the Illumina Crude Lysate Protocol provide a fast and easy workflow that eliminates costly, time-consuming DNA extraction steps and delivers exceptional data quality. The Illumina Crude Lysate Protocol offers high sensitivity and accuracy for organism detection and metagenomic profiling, as well as high-quality *de novo* genome assembly comparable to extracted DNA. With the ability to support complex metagenomic mixtures such as stool or mock microbial community samples with the Illumina Crude Lysate Protocol, as well as saliva, blood, and direct bacterial colonies with additional demonstrated protocols, Illumina DNA Prep delivers a flexible, cost-effective approach for NGS-based metagenomic research.

Learn more

Illumina DNA Prep

Microbiome sequencing

Ordering information

Library prep	Catalog No.
Illumina DNA Prep (M) Tagmentation (24 Samples, IPB)	20060060
Illumina DNA Prep (M) Tagmentation (96 Samples, IPB)	20060059
Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 Indexes, 96 Samples)	20091654
Illumina DNA/RNA UD Indexes Set B, Tagmentation (96 Indexes, 96 Samples)	20091656
Illumina DNA/RNA UD Indexes Set C, Tagmentation (96 Indexes, 96 Samples)	20091658
Illumina DNA/RNA UD Indexes Set D, Tagmentation (96 Indexes, 96 Samples)	20091660
Nextera DNA CD Indexes (96 indexes, 96 samples)	20018708

References

1. Ranjan R, Asha RA, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun*. 2016;469:967–977.
2. Lloyd-Price J, Mahurkar A, Rahnavard G, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. 2017;550:61-66.
3. Illumina (2023). Illumina DNA Prep Data Sheet. Accessed August 1, 2023.
4. Invitrogen (2015). PureLink Microbiome DNA Purification Kit User Guide for Microbial Culture. Accessed August 1, 2023.
5. Invitrogen (2015). PureLink Microbiome DNA Purification Kit User Guide for Stool Samples. Accessed August 1, 2023.
6. One Codex platform. www.onecodex.com. Accessed August 1, 2023.
7. MEGAHIT v1.1.1. github.com/voutcn/MEGAHIT. Accessed August 1, 2023.
8. QUAST. quast.sourceforge.net/quast. Accessed August 1, 2023.
9. DRAGEN Metagenomics. illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/dragen-metagenomics-pipeline. Accessed August 1, 2023.

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-02087 v1.0