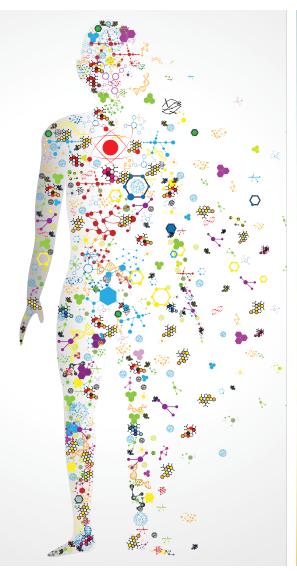


Methods Guide for Microbial Whole-Genome Sequencing

Sequence hundreds of organisms simultaneously, generate accurate reference genomes, and perform comparative genomic studies







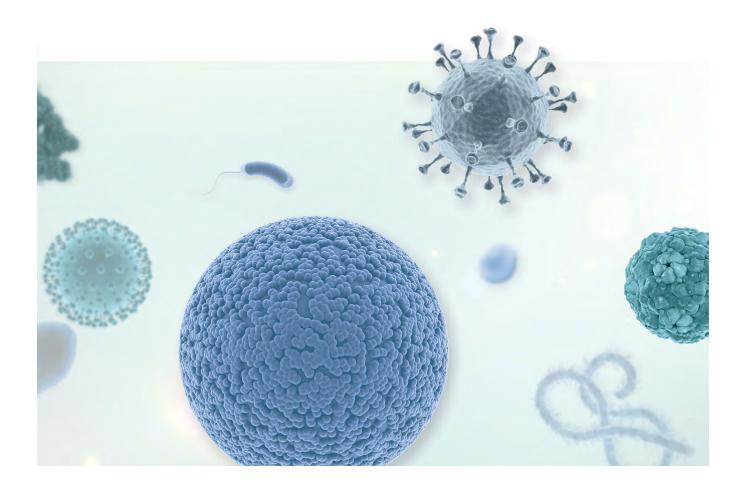
Introduction

Microbial whole-genome sequencing (WGS) is an important method for characterizing or assembling genomes of novel organisms, finishing genomes of known organisms, or comparing genomes across multiple samples. Sequencing entire bacterial, viral, and other microbial genomes is important for generating accurate reference genomes, for microbial identification, and enables comparative genomic studies.

Unlike capillary sequencing or PCR-based approaches, next-generation sequencing (NGS) allows microbiology researchers to sequence hundreds of organisms simultaneously. Additionally, NGS can identify low-frequency variants and genome rearrangements that may be missed or are too expensive to identify using other methods.

Use microbial WGS to:

- Analyze entire genomes and mobile genetic elements from microbial isolates and cultures comprehensively
- Sequence whole genomes directly from primary samples through targeted resequencing
- Discover new variants and detect low-frequency quasispecies in culture or clinical samples with an improved lower limit of detection compared to traditional methods
- Sequence multiple organisms in parallel using high-capacity NGS systems





De novo microbial genome sequencing

De novo WGS involves assembling a genome without the use of a genomic reference and is often used to sequence novel microbial genomes. Illumina sequencing systems provide exceptional raw read accuracy, read length, and read depth for high-quality draft and complete microbial genome assemblies.



Microbial whole-genome resequencing

Microbial whole-genome resequencing involves sequencing the entire genome of a bacteria, virus, or other microbe, and comparing the sequence to that of a known reference. Generating rapid and accurate microbial genome sequence information is critical for detecting low frequency mutations, finding key deletions and insertions, and discovering other genetic changes among microbial strains.



Sequencing isolates or primary samples

Sample type and previous knowledge of organisms of interest will determine the best sequencing method to use. For microbial isolates, WGS of DNA or whole-transcriptome sequencing (WTS) of RNA is recommended. These methods can detect known targets or offer hypothesis-free analysis of the sample. When analyzing complex primary samples, a hypothesis-driven, targeted resequencing approach like amplicon or enrichment sequencing is recommended when the target organism is either known or suspected. Hypothesis-free sequencing of complex primary samples requires approaches like shotgun metagenomics or metatranscriptomic sequencing.

Methods for sequencing isolates and primary samples

	Isolate		Primary sample		
Target Organism	DNA RNA (prokaryotes, eukaryotes, and DNA viruses) (viruses only)		DNA RNA (prokaryotes, eukaryotes, and DNA viruses) (viruses only)		
Known			Targeted DNA amplicon sequencing	Targeted RNA amplicon sequencing	
Suspected	WGS	WTSª	Custom DNA enrichment metagenomic sequencing	RNA enrichment metagenomic sequencing	
Unknown			Shotgun metagenomics	Metatranscriptomic sequencing ^a	

a. With optional ribosomal RNA (rRNA) depletion.

Applications

NGS has opened new doors in microbial genomics, revealing fresh insight into how microbes impact humans and the environment. The power to generate accurate reference genomes for microbial identification and/or other comparative genomic studies has widespread applications.

- Discover new microbes, including viruses
- Characterize difficult-to-culture organisms
- Investigate and understand outbreaks from infectious disease, antibiotic resistance, or foodborne pathogens
- Enable synthetic biology solutions to harness the power of nature to solve problems in medicine, manufacturing, and agriculture
- Analyze human, plant, animal, and environmental samples for beneficial microbes and pathogens
- Monitor quality in industrial bioprocess and vector quality control (QC)
- · Discover new enzymes and metabolic functions for industrial or medical use with systematic bioprospecting

Bacterial species change over time through various mechanisms, such as horizontal gene transfer and coevolution with other organisms. High-resolution gene sequencing can inform a wide range of studies, including mutagenesis, directed evolution, the spatial and temporal dynamics of epidemics, and mechanisms of disease transmission.

Uncover new taxa and biology

Extremophiles

Dr. Brian Hedlund uses microbial WGS to learn more about the unique microbes that populate hot springs. Through his multidisciplinary approach to the topic, Dr. Hedlund is discovering major new groups of life that have never been explored by science before. These organisms likely diverged from known lineages of microbes over a billion years ago. Thanks to genomics, he is discovering new classes of organisms, learning how to work with difficult-to-culture microbes, and learning new insights about microbial life in general.

Watch the video at www.illumina.com/company/video-hub/nH3dGo-DZTk.html

Infectious disease outbreaks and epidemiology

Ebola

Dr. Pardis Sabeti wants to empower public health communities and try to stop the next Ebola epidemic before it starts. To do this, she advocates using microbial WGS to identify infectious threats and track their spread in real time. She, along with her partners and sponsors, including Illumina, have successfully overcome many challenges and delivered and installed three sequencing systems in West Africa.

2 Watch the video at www.illumina.com/company/news-center/feature-articles/adventures-in-genomics--the-fight-against-ebola.html

Tuberculosis

In their paper, "Use of Whole Genome Sequencing to Determine the Microevolution of Mycobacterium tuberculosis during an Outbreak," Kato-Maeda et al demonstrated how WGS analysis of all *M. tuberculosis* isolates from a hospital outbreak of tuberculosis enabled researchers to trace the outbreak, uncovering the source and determining the directionality of transmission.

Read the article at journals.plos.org/plosone/article?id=10.1371/journal.pone.0058235

"WGS provided a link between cases that did not otherwise have an epidemiologic link."

Kato-Maeda et al

Antibiotic resistance detection and managing the spread

MRSA (methicillin-resistant Staphylococcus aureus)

Bartels et al used WGS to analyze a Methicillin-resistant Staphylococcus aureus (MRSA) outbreak in a neonatal ward and cases in the community, suggesting that WGS can potentially fill in gaps in epidemiological data.

Read the article at www.ncbi.nlm.nih.gov/pubmed/25955776

"WGS seems promising in outbreak investigations, and might add information on the spread of outbreak clones outside the hospital setting, where the connection between patients is frequently not clear." – Bartels et al

Foodborne pathogen surveillance and outbreak investigations

IZSLER Parma laboratory provides diagnostic and testing services for animal disease diagnostic testing, including zoonoses, official controls on food and feedstuffs, epidemiological and analytical support to health authorities for the design and management of official prevention and eradication plans, monitoring and surveillance of animal and zoonotic diseases for foodborne diseases, and applied and basic research in infectious diseases.

Read the article at www.illumina.com/content/dam/illumina-marketing/documents/icommunity/pongolini-izsler-interview-miseq-micro-770-2015-033.pdf

Harness the power of microbes to improve medicine, agriculture, and manufacturing

Synthetic biology is a growing field that involves redesigning organisms for useful purposes by engineering them to have new abilities. Researchers and companies around the world are harnessing the power of nature to solve problems in medicine, manufacturing, and agriculture. WGS is a foundational technology for this area of research and applied outcomes, enabling organism engineers to use many approaches, such as gene editing or synthetic gene synthesis, to build complex metabolic pathways in microorganisms designed to perform high-value functions. Applied applications include living medicines, cultured ingredients, enhanced agricultural microbiomes, and use in biosecurity initiatives.

Improve food and drug development and QC

Microbes are used in many food production processes, like cheese, wine, and beer. Microbes can be analyzed and characterized to understand how they influence important quality traits and must be closely monitored and managed to ensure that these traits of interest are maintained. WGS enables discovery of new strains that can provide new quality traits in food production and monitoring and quality control of strains in use.

Similarly, in biopharmaceutical research, risk monitoring is a key aspect in research and product development. WGS allows researchers to track cell line stability and conduct biosafety monitoring.

Yeast in beer

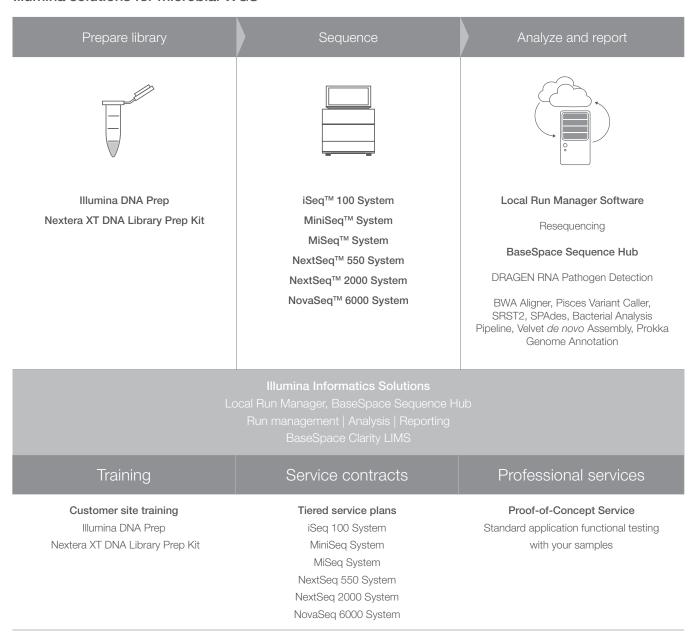
An international group of scientists, yeast producers, and brewers sequenced 96 yeast strains to learn how genetics can influence the flavor of beer. Their findings could help brewers make better beers.

Read the Cell article at www.cell.com/cell/fulltext/S0092-8674(16)31071-6

How it works

Microbial WGS has three basic steps: library preparation, sequencing, and analysis. During library preparation, DNA is fragmented and probes containing Illumina sequencing adapters hybridize to each fragment. Each individual fragment is then sequenced and the resulting nucleotide-level data is analyzed to generate an accurate reference genome, compare variations within and between species, and differentiate between organisms.

Illumina solutions for microbial WGS



Step 1 Choose a library prep method

Illumina recommends two library preparation kits for microbial WGS studies: Illumina DNA Prep and Nextera™ XT DNA Library Prep. Illumina DNA Prep provides a faster workflow with fewer steps. Nextera XT DNA Library Prep requires less input material and can generate longer insert sizes for sequencing.





Library prep kit	Illumina DNA Prep	Nextera XT DNA Library Prep Kit
Most important to me	Simple and versatile sample-to-data workflow	Less hands-on time and longer insert protocol
Least important to me	PCR-free workflow	Amplicons < 300 bp in length
Input quantity	1-500 ng	1 ng DNA
Multiplexing	384 samples	384 samples
Assay time	3-4 hrs	~5.5 hrs
Hands-on time	1.5 hrsª	15 min
Target insert size	~350 bp	300 bp-1.5 kb
Diversity	≥ 1B unique fragments	≥ 1B unique fragments
Genome coverage	Complete	Complete
Mechanism of action	Bead-linked transposome	Enzymatic DNA fragmentation

a. Denotes total workflow time, including DNA extraction, library preparation, and library normalization/pooling steps

Illumina DNA Prep*

Save time and reduce hands-on touch points with On-Bead Tagmentation, producing sequencing-ready libraries in less than three hours. Simplify daily operations with a kit that supports a broad DNA input range (1-500 ng), multiple DNA input types, and use with small (bacteria, Archaea, viruses, plasmids) to large genomes (human, plant, mouse). Libraries prepared with Illumina DNA Prep are compatible with all Illumina sequencing systems.

Learn more at www.illumina.com/illumina-dna-prep

Nextera XT DNA Library Prep Kits

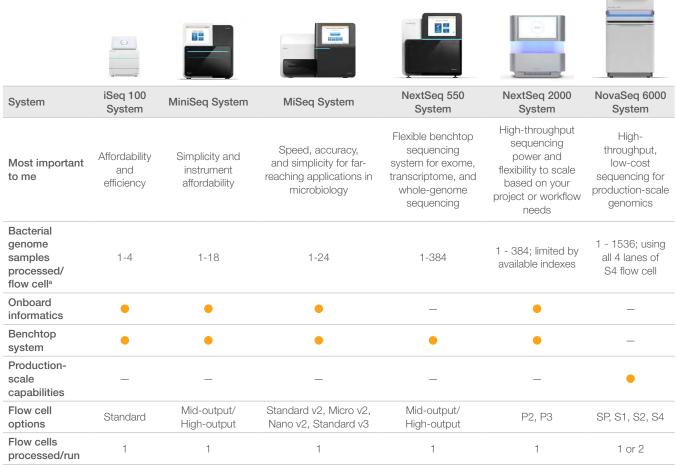
Generate sequencing-ready libraries from small genomes (bacteria, Archaea, viruses), amplicons, and plasmids in less than 90 minutes with only 15 minutes of hands-on time. Nextera XT DNA Library Prep Kits simultaneously fragment input DNA and tags the fragments with sequencing adapters in a single-tube enzymatic reaction. Nextera XT DNA Library Prep Kits require as little as 1 ng of input, supporting a wide array of sample types. Bead-based normalization eliminates the need for library quantification before pooling and sequencing. Libraries prepared with Nextera XT Kits are compatible with all Illumina sequencing systems.

2 Learn more at www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-xt-dna.html

^{*} Formerly available as the Nextera DNA Flex Library Preparation Kit

Step 2 Choose a sequencing system

Sequencing system selection begins with identifying the best flow cell type and read length to support given applications while matching data output, time to results, and price per sample to meet study needs.



a. Sample numbers may depend on genome size, read depth, and specific flow cell output. Calculations based on 5 Mb genome and 1M reads.



iSeq 100 System flow cell options and specifications					
Flow cell type i1 ^a					
Output/run	144 Mb-1.2 Gb				
Reads/run	4M				
Max read length 2 × 150 bp					
Cycles	300				

a. The i1 flow cell is included in the iSeq 100 Reagents; this is a ready-to-use cartridge prefilled with all the reagents needed for sequencing. Simply load the reagent cartridge directly onto the iSeq 100 System and start a run.

iSeq[™] 100 System

With the lowest price, smallest footprint, and fastest run time of any Illumina instrument, the iSeq 100 System offers an affordable option for researchers to expand the scope of their research. Sequence microbes, targeted genes, RNA transcripts, and more at the push of a button.

- Lowest Illumina sequencing system cost
- Suitable for small sample batches
- Fixed throughput and read length
- Suitable for microbial WGS and amplicon sequencing

Learn more at www.illumina.com/iSeq.

MiniSeq[™] System

Applying industry-leading sequencing technology in a simple, small, affordable system, the MiniSeq System supports a broad range of targeted DNA and RNA applications for examining single genes or entire pathways. An intuitive user interface, load-and-go operation, and onboard data analysis make it easy to learn and easy to use.

300

Affordable to acquire

Cycles

- Suitable for moderate sample batches
- Range of throughput and read length
- Suitable for microbial WGS and amplicon sequencing



∠ Learn more at www.illumina.com/miniseq									
MiniSeq System flow cell	options and specifications								
Flow cell type	Mid-output		High-output						
Output/run	2.1-2.4 Gb	1.7-1.9 Gb	3.3-3.8 Gb	6.6-7.5 Gb					
Reads/run	8M	25M	25M	25M					
Max read length	2 × 150 bp	1 × 75 bp	2 × 75 bp	2 × 150 bp					



MiSeq[™] System

75

Combining speed, high-quality data, and the longest read lengths from Illumina, the MiSeq System is ideal for sequencing targeted panels, amplicons, and small genomes.

300

• Suitable for moderate sample batches

150

- Range of throughput and offers the longest read lengths of any Illumina benchtop system
- Suitable for many microbial sequencing applications, including WGS

Learn more at www.illumina.com/miseq

MiSeq System flow cell options and specifications									
Flow cell type		Standard v2		Micro v2	Nan	o v2	Stand	lard v3	
Output/run	0.75-0.85 Gb	4.5-5.1 Gb	7.5-8.5 Gb	1.2 Gb	0.3 Gb	0.5 Gb	3.3-3.8 Gb	13.2-15 Gb	
Reads/run	15M	15M	15M	4M	1M	1M	25M	25M	
Max read length	2 × 25 bp	2 × 150 bp	2 × 250 bp	2 × 150 bp	2 × 150 bp	2 × 250 bp	2 × 75 bp	2 × 300 bp	
Cycles	50	300	500	300	300	500	150	600	





NextSeq[™] 550 System

The NextSeq 550 System delivers the power of high-throughput sequencing with the simplicity of a benchtop sequencer, transforming exome, transcriptome, and whole-genome sequencing into everyday research tools. High-quality data combined with versatile, streamlined DNA-to-data workflows enables low- and high-throughput studies supporting a range of project sizes and applications.

- Suitable for high sample batches
- Range of throughput and read lengths
- Suitable for many microbial sequencing applications, including WGS, shotgun metagenomic sequencing, and metatranscriptomic sequencing

Learn more at www.illumina.com/systems/sequencing-platforms/nextseq.html.

NextSeq 550 System flow cell options and specifications								
Flow cell type	Mid-out	put v2.5		High-output v2.5				
Output/run	16-19 Gb	32-39 Gb	25-30 Gb	50-60 Gb	100-120 Gb			
Reads/run	130M	130M	400M	400M	400M			
Max read length	2 × 75 bp	2 × 150 bp	1 × 75 bp	2 × 75 bp	2 × 150 bp			
Cycles	150	300	75	150	300			

NextSeq 2000 System

The NextSeq 2000 System enables researchers to explore current and emerging applications with higher efficiency and fewer restraints. The cost-efficient, high-throughput system delivers accurate results for emerging applications of increasing complexity. Combining a streamlined workflow with onboard informatics, this easy-to-use benchtop system makes NGS accessible to novice and expert users.

- Suitable for high sample batches
- Fixed throughput and read lengths
- Suitable for many microbial sequencing applications, including WGS, shotgun metagenomic sequencing, and metatranscriptomic sequencing
- On-board integrated informatics with DRAGEN™ Bio-IT Platform for rapid secondary analysis



∠ Learn more at www.illumina.com/nextseq2000

NextSeq 2000 System flow cell options and specifications								
Flow cell type		P2			P3ª			
Output/run	40 Gb	80 Gb	120 Gb	100 Gb	200 Gb	300 Gb		
Reads/run	400M	400M	400M	1B	1B	1B		
Max read length	2 × 50 bp	2 × 100 bp	2 × 150 bp	2 × 50 bp	2 × 100 bp	2 × 150 bp		
Cycles	100	200	300	100	200	300		

a. Available late 2020



NovaSeq[™] 6000 System

The NovaSeq 6000 System unleashes groundbreaking innovations that build upon proven Illumina sequencing by synthesis (SBS) chemistry. Get scalable throughput and flexibility for virtually any sequencing method, genome, and scale of project. Applications requiring large amounts of data can now be completed in a more cost-effective manner. For additional flexibility, the NovaSeq Xp workflow supports individual lane loading for sequencing different libraries in each flow cell lane. When combined with lower output flow cells, the same instrument can be used for less data intensive methods.

- High-throughput sequencing with unprecedented throughput at the lowest cost per sample for any Illumina sequencing system
- Range of throughput and read lengths
- Suitable for many microbial sequencing applications including WGS, shotgun metagenomic sequencing, and metatranscriptomic sequencing

Learn more at www.illumina.com/systems/sequencing-platforms/novaseq.html.

NovaSeq 6000 System flow cell options and specifications										
Flow cell type		SP			S1			S2		S4
Output/run	80 Gb	250 Gb	400 Gb	167 Gb	333 Gb	500 Gb	417 Gb	833 Gb	1250 Gb	3000 Gb
Reads/run	800M	800M	800M	1600M	1600M	1600M	4100M	4100M	4100M	10,000M
Cycles	100	300	500	100	200	300	100	200	300	300
Output per flow	cell									
2 × 50 bp		65-80 Gb			134-167 Gk)		333-417 GI	b	N/A
2 × 100 bp		134-167 Gb)	266-333 Gb		667-833 Gb		1600-2000 Gb		
2 × 150 bp		200-250 Gb)	400-500 Gb		1000-1250 Gb		Эb	2400-3000 Gb	
2 × 250 bp		325-400 Gb)		N/A			N/A		N/A

N/A = not applicable

Step 3

Add biological context

Analyze whole-genome data with Local Run Manager Software or BaseSpace™ Sequence Hub Apps.

Learn more, including example data sets, workflows, and supported file types at www.illumina.com/basespace

Pathogen detection



DRAGEN RNA Pathogen Detection App†

- Uses a combined human and virus reference to analyze pathogen data and create consensus FASTAs
- Creates pathogen coverage plots
- Offers integrated QC metrics



BWA^{† ‡} Aligner and Pisces Variant Caller

Performs read alignment and mutation detection for any genome with a known reference sequence.
 If none is available, the pipeline can still be used to analyze bacteria, fungi, or viruses after de novo assembly



SRST2[†]

 Reports the presence of sequence types (STs) from a multilocus sequence typing (MLST) database and/or reference genes from a database of sequences for virulence genes, resistance genes, and plasmid replicons



Bacterial Analysis Pipeline

Identifies species, MLST, plasmids, virulence, and antimicrobial resistance genes in bacteria

Isolate sequencing and assembly



Assembly workflow§

- Assembles small genomes (< 20 Mb); best suited for the assembly of bacterial genomes, such as E. coli
- Uses the EMBL Velvet algorithm and writes assembly results in FASTA format



SPAdes Genome Assembler

 Assembles genomes from standard bacterial isolates and single-cell multiple displacement amplification (MDA) preparations

[†] BaseSpace Labs Apps are developed using an accelerated development process to make them available to BaseSpace users faster than conventional Illumina Apps.

Illumina may provide support for BaseSpace Labs Apps at its sole discretion. BaseSpace Labs Apps are provided AS IS without any warranty of any kind. BaseSpace

Labs Apps are used at the user's sole risk. Illumina is not responsible for any loss of data, incorrect results, or any costs, liabilities, or damages that may result from the use
of a BaseSpace Labs App. For inquiries or feedback for this application, to provide a feature request, or to get support, email basespacelabs@illumina.com.

[‡] BWA: Burrows-Wheeler Aligner

[§] Available through Local Run Manager

Gene annotation



Prokka Genome Annotation[†]

- Annotates genes and identifies coding sequences in prokaryotic genomes
- Does not annotate eukaryotic genomes

Commercial third-party apps

bioMérieux EPISEQ

- Integrates comprehensive applications for microbiological research and routine applications
- Combines information from various genomic and phenotypic sources into one global database and conducts combined analyses
- Learn more at www.applied-maths.com/applications

ChunLab TrueBac ID

- Uses whole-genome sequences to identify > 12,000 species for microbial applications in research, applied, and industrial settings
- Learn more at www.truebacid.com

[†] BaseSpace Labs Apps are developed using an accelerated development process to make them available to BaseSpace users faster than conventional Illumina Apps. Illumina may provide support for BaseSpace Labs Apps at its sole discretion. BaseSpace Labs Apps are provided AS IS without any warranty of any kind. BaseSpace Labs Apps are used at the user's sole risk. Illumina is not responsible for any loss of data, incorrect results, or any costs, liabilities, or damages that may result from the use of a BaseSpace Labs App. For inquiries or feedback for this application, to provide a feature request, or to get support, email basespacelabs@illumina.com.

Services and support

Illumina consulting services provide laboratories with direct access to experienced Illumina teams to jump start successful sequencing workflow implementation.



Training

Onsite customer training courses provide detailed, hands-on training on Illumina library preparation solutions in the customer's laboratory. Participants will receive expert instruction and best practice guidance for performing and troubleshooting the basic protocol. By the end of the course participants will be able to prepare libraries efficiently and consistently for use on Illumina sequencing systems. Standard and personalized workflow training options are available.

2 Learn more at support.illumina.com/training/instructor-led/on-site.html.



Service contracts

A standard 1-year base warranty is included with every new Illumina instrument purchase, along with installation and basic applications training. Illumina also offers several tiered services plans to upgrade the base warranty to an enhanced service level or extend service coverage beyond the 1-year warranty. These plans are designed to optimize the customer experience and to protect instrument investment. Illumina Product Support Services Plans help researchers maximize performance and productivity with reliable, high-quality results at various cost-effective levels.

Learn more at www.illumina.com/services/instrument-services-training/ product-support-services/service-plans.html.



Professional services

The Proof-Of-Concept (POC) Service will run a subset of your samples with your desired system and application in a real-world demonstration of how Illumina NGS can empower your laboratory. Receive in-person consultation pre-POC to establish POC needs and goals, POC sequencing services of your samples at the Illumina Service Lab, and in-person consultation post-POC to review sequencing run QC, data, and reports based on your samples.

Contact your local account manager to learn more.

Ordering information

Library prep kits

Product	Catalog no.
Illumina DNA Prep	
Illumina DNA Prep (M), Tagmentation (24 samples)	20018704
Illumina DNA Prep (M), Tagmentation (96 samples)	20018705
Flex Lysis Reagent Kit	20018706
Nextera DNA CD Indexes (24 indexes, 24 samples)	20018707
Nextera DNA CD Indexes (96 indexes, 96 samples)	20018708
Nextera XT DNA Library Prep	
Nextera XT DNA Library Prep Kit (24 samples)	FC-131-1024
Nextera XT DNA Library Prep Kit (96 samples)	FC-131-1096
Nextera XT Index Kit (24 indexes, 96 samples)	FC-131-1001
Nextera XT Index Kit (96 indexes, 384 samples)	FC-131-1002
TruSeq™ Dual Index Sequencing Primer Kit, single read	FC-121-1003
TruSeq Dual Index Sequencing Primer Kit, paired-end read	PE-121-1003
Nextera XT DNA Library Prep Kit training at customer site	TR-204-0009

Sequencing Systems

System	Catalog no.
iSeq 100 System	20021532
MiniSeq System	SY-420-1001
MiSeq System	SY-410-1003
NextSeq 550 System	SY-415-1002
NextSeq 2000 System	20038897
NovaSeq 6000 System	20012850

Data analysis

Product	Catalog no.
BaseSpace Sequence Hub Professional Annual Subscription	20042109
BaseSpace Sequence Hub Enterprise Annual Subscription	15066411

