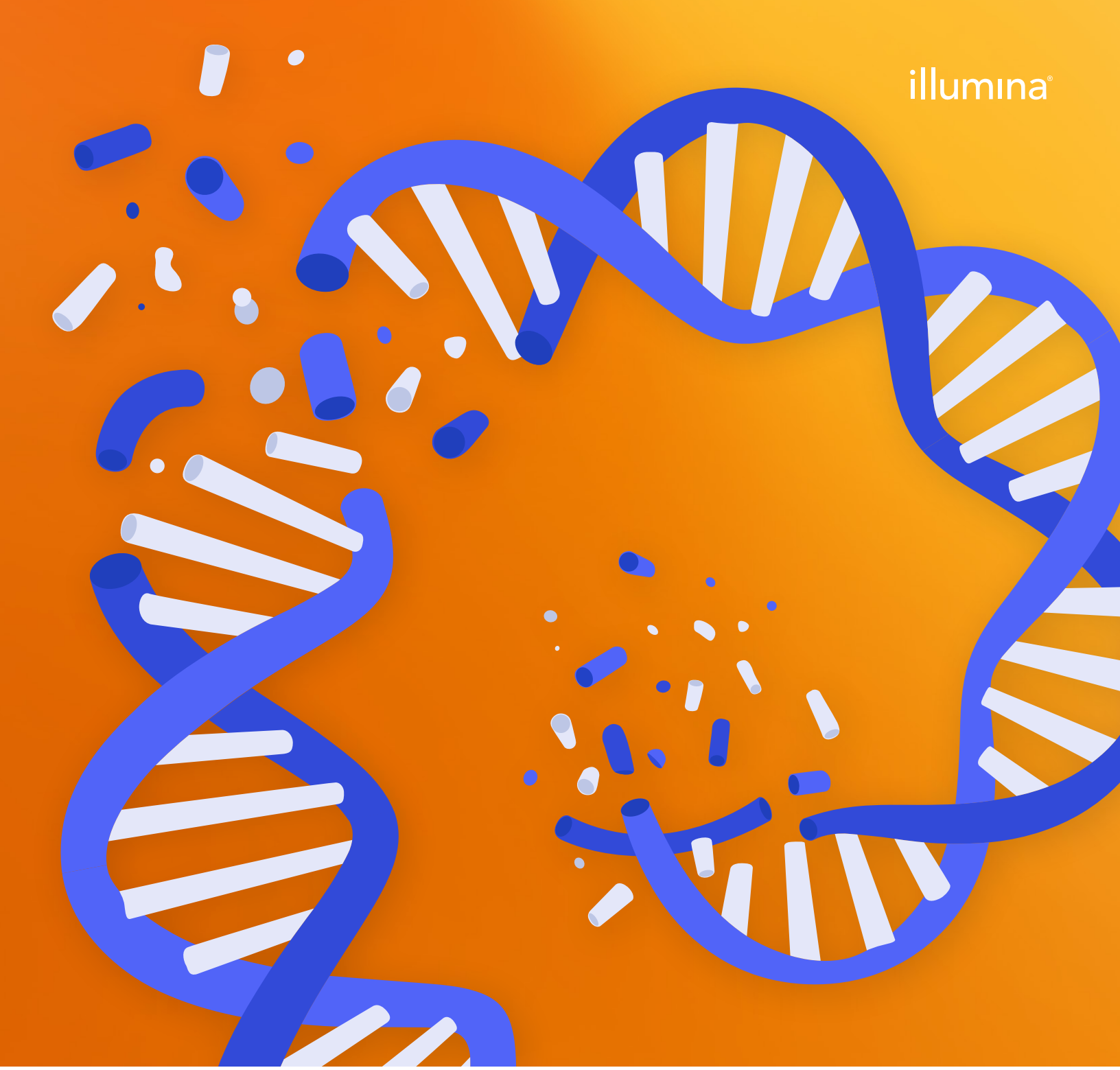


Next-generation sequencing and microarray methods to comprehensively understand cancer

Use next-generation sequencing and microarray tools to investigate the involvement of different “omes” in cancer.

Table of contents

An introduction to NGS: A methods guide for cancer research

Cancer is a complex disease that cannot be explained by a single factor. Dynamic interactions between cancer cells and their microenvironments influence all levels of cellular regulation, including the genome, epigenome, transcriptome, and proteome. Tools like NGS and microarrays have revolutionized the understanding of cancer by providing researchers with unparalleled, comprehensive characterization of this multidimensional complexity. Ultimately, these technologies and methodologies provide avenues for future improvements in patient care and outcomes.

Table 1: Each ome and its implications for cancer research

Biological layer	Implications for cancer research
Genome	Genomic analysis provides information on a person's susceptibility to cancer and cancer-related changes to DNA. In research studies, it can provide scientists, oncologists, and pathologists with the genetic profile of the tumor.
Epigenome	Studying the epigenome allows researchers to understand mechanisms that control gene activity through nongenetic changes such as DNA methylawtion.
Transcriptome	RNA sequencing explores the expressed changes to DNA that underlie disease mechanisms for specific cancers.
Proteome	Interrogating the proteome provides the link between cancer-specific genes or gene expression patterns and important biomarkers.
Immunome	Molecular expression profiling of immune cell populations to identify immune cell–associated genes and proteins associated with the cancer microenvironment.
Microbiome	Characterization of microbial communities can provide insights into members that may regulate inflammation, immune activity, and other pathways that may influence disease onset, disease progression, and response to anticancer therapy.

Why NGS in cancer research?

While there have been several effective tools to interrogate parts of the genome, transcriptome, proteome, and epigenome, NGS is the only tool that allows scientists to characterize the entirety (or a substantial portion) of any "ome" in one experiment. Since NGS does not depend on predesigned probes to determine nucleic acid sequences, it is an unbiased tool for uncovering novel insights into difficult-to-treat and little-understood cancers. With NGS, researchers are not limited by previous understanding or a limited number of targets.

NGS also allows researchers to detect low-frequency molecular events associated with carcinogenesis, cancer growth, and metastasis. These important, yet low-frequency molecular events could be completely missed using traditional molecular methods.

General NGS approaches

Researchers can leverage four general approaches when it comes to studying biological omes: **bulk-cell analysis**, **single-cell analysis**, **spatial analysis**, and **metagenomic analysis**. Each views cancer at a different level of resolution and has a distinct use case depending on what information must be obtained from a sample.

Bulk-cell analysis: Bulk-cell analysis allows scientists to study pooled cell populations, tissue sections, or biopsies.

Single-cell analysis: Single-cell analysis studies a given ome at the resolution of a single cell.

Spatial analysis (also called spatial genomics): Spatial analysis captures omic information of single cells within an unperturbed tissue environment.

Metagenomic analysis: Metagenomic analysis sequences every gene in every organism of a complex microbial community present within a tissue, organ, or tumor.

Which approach should you use?

The approach you use will depend on your research objectives.

Bulk-cell analysis measures the average gene or protein expression across a population of cells. This NGS approach is the simplest and most cost-effective way to gain functional comprehensive data on factors that play important roles in disease. For instance, using bulk-cell data sets, researchers were able to show that the reduced expression of two specific methyltransferases correlates with improved pancreatic cancer outcomes.[1]

Furthermore, researchers can combine bulk-cell analyses with methods like flow cytometry to thoroughly characterize the differences between cell populations. In this example, cells can be sorted into distinct populations based upon known biomarker expression. Nucleic acids can then be extracted from each population, followed by bulk-cell sequencing and yielding omic data for each sorted pool of cells. Combining techniques holds excellent opportunities for cancer researchers to uncover insights into tumor cell populations.

Bulk-cell analysis, however, does not fully account for the inherent heterogeneity of the tumor microenvironment. Tumors do not consist of one cell type, but instead contain many distinct cancerous and noncancerous cell populations, all of which influence tumor biology and may not express any known biomarkers. Therefore, while bulk-cell analysis is useful, researchers can use **single-cell sequencing** in cancer research to further reveal pertinent genomic, transcriptomic, proteomic, and epigenomic distinctions between individual cells.[2] Pathways that are undetectable at the bulk-cell level may be expressed in small, but important cell populations. Single-cell sequencing adds the resolution necessary to understand these populations.

While both bulk-cell and single-cell analyses have deepened our understanding of cancer biology, these methods do not provide information about the cell's position within the native tissue microenvironment. Experimental results from bulk- and single-cell sequencing lack information about positional factors that influence the behavior of cells. Cell–cell interactions and the location of rare populations such as stem cell niches are just a few examples of positional effects that influence cancer.

> It is also well known that the cell isolation process can impact cell behavior, including eliciting stress responses that introduce unique biases into an experiment.[3]

To overcome these challenges, researchers can depend on **spatial analysis**. Spatial analysis studies the gene and protein expression of individual cells in the unperturbed microenvironment and architecture of a tumor section. Whether you're studying RNA or protein, the great advantage of using sequencing-based spatial analysis is that an exponential number of unique barcodes can be applied to a variety of targets, including cell identifiers, antibody identifiers, unique molecular identifiers, and, especially, spatial locations within a tissue sample. As such, spatial analysis is an excellent tool for studying tumor cell populations, the spatial niches where those cells interact and how those interactions drive cancer.[4]

When taking a genomics approach to characterizing the microbiome, **metagenomic analysis** can be used to assess the full genomic profile of all organisms comprising microbial communities. This high-throughput, powerful sequencing approach can provide insights into members of the microbiome that have been linked to the regulation of hormonal, metabolic, and immunologic pathways that may interact with and influence cells around and within the tumor microenvironment.[5] In addition to metagenomics analysis, researchers can select targeted sequencing, like 16S ribosomal RNA sequencing, to determine the genus-level identity of microorganisms within a sample. Further, metatranscriptomics analysis, which employs NGS-powered RNA-Seq, allows researchers to identify microbiome-associated metabolic processes and pathways that are active within the localized microbiome and may influence cancer biology.

What does an NGS workflow look like?

NGS workflows generally start with nucleic acid isolation, followed by library preparation. A library is a collection of similarly sized nucleic acid fragments with known adapter DNA sequences attached to the 5' and 3' ends. Library preparation enables sample compatibility with Illumina sequencing systems. Finally, the library is sequenced, data is collected from the sequencing systems and analyzed.

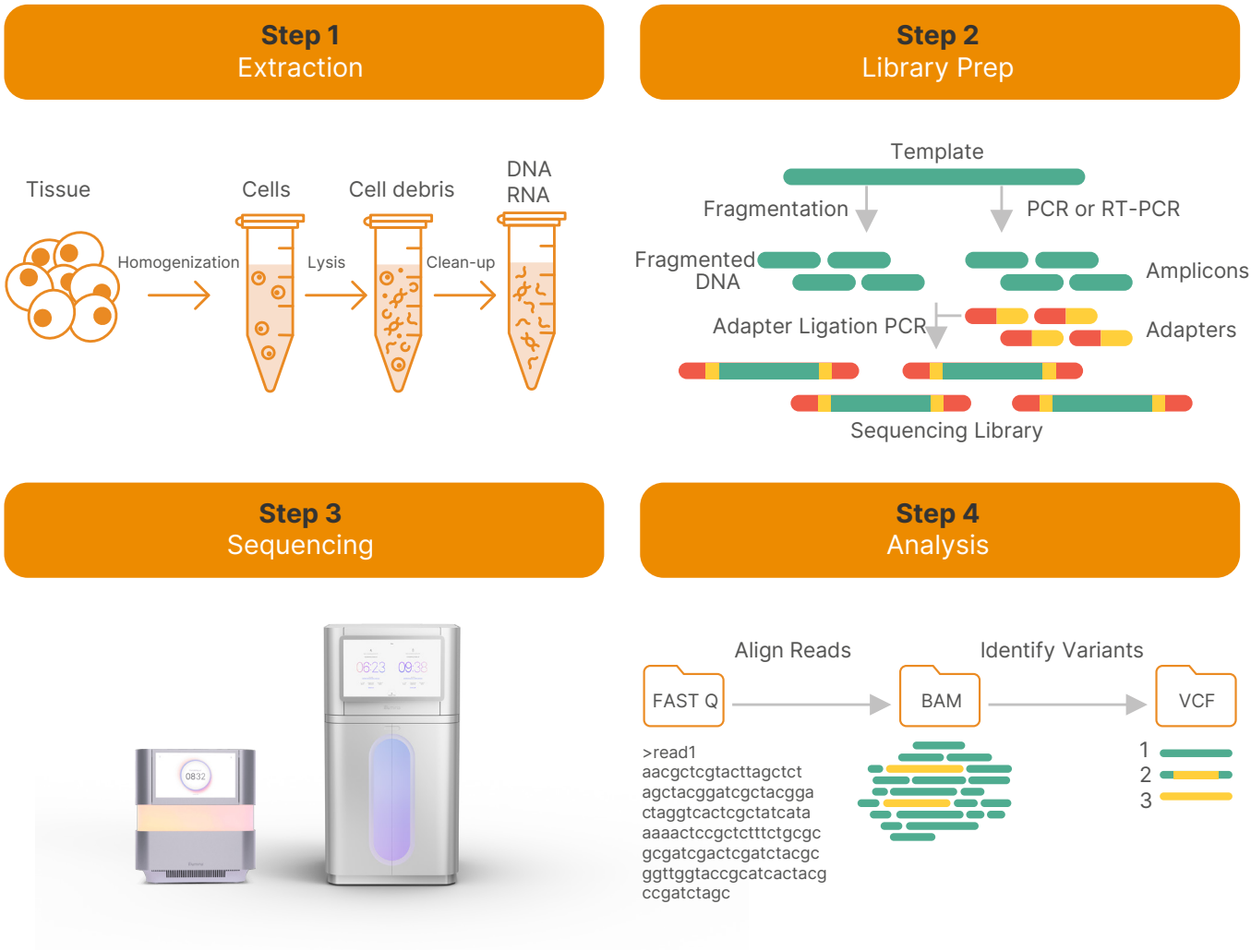


Figure 1: A general NGS workflow

Illumina supports end-to-end solutions, starting with sample preparation and ending with data analysis. Our NGS technology also provides you with:

- The capacity to analyze many omics, at multiple resolutions, on devices likely already present at your research institution, core lab, or local service provider
- Accurate sequence reads
- High-throughput capabilities at a low cost per sample as compared to other service providers
- The ability to scale your NGS studies to include more than one ome
- The opportunity to combine a large number of analytes into a single workflow

Overview of NGS-based workflow for cancer research

STEP 1 Library preparation

Following nucleic acid isolation, scientists can prepare nucleic acid libraries. Library preparation using our kits ensures that your samples will hybridize to the flow cells used in our sequencing systems. Library preparation kits are available for a range of applications, including large-scale, untargeted sequencing applications (eg, whole-genome sequencing), comprehensive targeted sequencing (eg, whole-exome sequencing), or sequencing panels (eg, cancer gene panels like TruSight™ Oncology 500).

STEP 2 Sequencing

Once your libraries are prepared, they are hybridized to special slides known as **flow cells**. The flow cells that contain hybridized nucleic acid fragments are subsequently sequenced. Regardless of your research question, our sequencing systems are flexible and can help you reach those answers using simple pushbutton workflows.

STEP 3 Data analysis

During sequencing, millions or sometimes billions of clusters or "reads" are generated via clonal amplification based on the original nucleic acid fragments that were hybridized to the flow cell. Analyzing sequencing data involves primary, secondary, and tertiary analysis steps. **Primary analysis** is completed automatically on many Illumina systems and results in files containing base calls. Base calls are signals that are used to infer the order of the nucleotides in the sample that was sequenced. **Secondary analysis** steps involve demultiplexing, alignment, and genetic characterization. Finally, data visualization, variant annotation, as well as biological and clinical interpretation is completed via **tertiary analysis**.

Researchers can analyze Illumina cancer research data using two Illumina software solutions that do not require bioinformatics expertise: BaseSpace™ Sequence Hub and DRAGEN™ bioinformatics tools. BaseSpace Sequence Hub contains data analysis tools that don't require coding or command-line expertise. The software can help you through both secondary and tertiary analyses via a series of apps. For instance, researchers can depend on the DRAGEN secondary analysis apps via BaseSpace Sequence Hub. DRAGEN analysis offers tools for somatic variant calling and biomarker readouts in genome, exome, and panel experiments (DRAGEN Somatic app); transcriptome analysis (DRAGEN RNA and DRAGEN Single Cell RNA); and methylome analysis (DRAGEN Methylation).

DRAGEN analysis is also available outside of the BaseSpace ecosystem and is preinstalled onboard Illumina NextSeq™ 1000 and NextSeq 2000 Systems and NovaSeq™ X Series for pushbutton analysis of some workflows. Furthermore, Illumina BaseSpace Correlation Engine software is a user-friendly tool for tertiary analyses. Illumina also offers the Illumina Connected Analytics analysis software solutions for experienced bioinformaticians and researchers who process data at a large scale.

Demultiplexing: This is a step in the analysis process where the researcher uses barcode information to separate and identify which sequences come from which cells or samples when multiple single cells or bulk samples are sequenced together.

Alignment: In this step, researchers align sequences with reference sequences for comparison.

Genetic characterization: In genetic characterization, data sets are analyzed for genomic variants, gene expression, protein expression, chromatin accessibility, etc.

Examples of NGS- and array-based methods used in cancer research

Table 2: NGS- and array-based methods used in cancer research

Biological level	NGS-based method	Implication
Multiome	Single-cell sequencing	Single-cell sequencing allows researchers to understand drivers of cancer at the DNA, RNA, epigenetic, and protein levels (individually or as multiomic experiments) at the resolution of a single cell.
	Cellular indexing of transcriptomes and epitopes (CITE-Seq)	CITE-Seq studies both protein and RNA in single tumor cells, allowing researchers to tie gene expression to protein biomarkers associated with the cancer type.
	Bulk epitope and nucleic acid sequencing (BEN-Seq)	BEN-Seq assays both protein and RNA. It interrogates tumor heterogeneity to characterize tumor cells, infiltrating immune cells, and the microenvironment.
	Spatial sequencing	Spatial transcriptomics combines RNA profiling with immunofluorescence imaging for studying cells in their native environment.
Genome	Whole-genome sequencing (WGS)	WGS helps identify a comprehensive list of cancer-driving genetic events.
	Whole-exome sequencing (WES)	WES is a cost-effective sequencing method that looks at cancer-driving genetic events in the coding portion of the genome.
Epigenome	Methylation arrays	Methylation arrays quantitatively interrogate methylation sites across the genome at single-nucleotide resolution.[6]
	Assay for transposase-accessible chromatin sequencing (ATAC-Seq)	ATAC-Seq studies the chromatin regulatory landscape in cancers to elucidate the role of epigenetics in cancer progression.[7]
Transcriptome	Bulk RNA sequencing (Bulk RNA-Seq)	Bulk RNA-Seq studies the average transcriptomic profile of a population of cancer cells. This can be targeted (eg, mRNA-Seq) or untargeted (whole transcriptome).
	T-cell receptor (TCR) Sequencing	TCR sequencing is a targeted RNA panel that investigates T-cell diversity and clonal expansion by sequencing T-cell receptor beta chain arrangements. TCR sequencing characterizes tumor-reactive T-cells to understand the central players in immunosurveillance.
	Cell-free RNA sequencing (cfRNA-Seq)	cfRNA-Seq studies and characterizes RNA in biofluids such as blood or plasma.
	Mutational profiling for neoantigen prediction	Mutational profiling identifies tumor-specific peptides that may be capable of inducing an immune response.
Proteome	Olink proximity extension assay (PEA)	Olink PEA enables high-throughput screening of a large protein panel for biomarker detection and drug discovery studies.
Microbiome	16S and ITS Sequencing	16S sequencing provides targeted sequencing of ribosomal rRNA to identify microbial species down to the genus-level. Incorporating the amplicon spanning the ITS1 region of the rRNA cistron into this targeted sequencing approach enables species-level identification of fungal species within a microbial sample.
	Shotgun metagenomics	Shotgun metagenomics enables the comprehensive sequencing of all microbial genomes, including cancer-associated viruses, phages, fungi, and bacteria, present in a complex sample, such as biopsies of human tumors.
	Metatranscriptomics	Bulk RNA-Seq can be used to reveal the presence of microorganisms and their gene expression profiles.

Multiomic methods

Method 1 Single-cell sequencing

Single-cell sequencing reveals pertinent genomic, epigenomic, transcriptomic, and proteomic differences amongst individual cells within the tumor microenvironment. While studying the average omic profiles of cell populations remains a useful and informative tool for cancer research, cells comprising the tumor microenvironment are heterogeneous—containing diverse cancer and noncancer cell populations that contribute to cancer biology. Profiling individual cells may therefore be important to identify and characterize rare cell populations that may not be detectible via some bulk sequencing approaches. While studying the average gene or protein expression of cancer cells remains a useful and informative tool, tumor microenvironments are heterogeneous and it is important for scientists to understand that differential expression in unique cell populations may have been missed with a bulk sequencing method. Profiling single cells helps researchers gain insights into how unique populations impact cancer and how they might be exploited therapeutically.

Key benefits of single-cell sequencing over bulk-cell sequencing include:

- Detecting functional cell populations in the tumor microenvironment.
- Uncovering the impact of non-cancerous cell populations such as immune cells (eg, B-cells and T-cells) and fibroblasts on tumor biology.
- Understanding the effects of epigenetic heterogeneity in cancer progression.
- Constructing the evolution of somatic variants from tumor samples.
- Identifying and characterizing cancer stem cell populations.

Potential applications of single-cell sequencing in cancer research

Identification of biomarkers associated with targeted therapeutics like immunotherapy	Immune checkpoint blockage has been shown to induce durable responses in patients across multiple types of cancer.[8,9] Unfortunately, only a fraction of patients responds to this type of therapy. The antitumor immune response is complex. Identifying new biomarkers may help to differentiate “responders” to immunotherapy from non-responders.
Understanding tumor cell heterogeneity	A tumor is a complex ecosystem of heterogeneous cells that includes cancer cells, endothelial cells, fibroblasts, and infiltrating immune cells, among other cell populations. Single-cell sequencing allows scientists to understand the differences between and among these cell populations and the mechanistic factors that drive expression.[10]
Uncovering the genetic drivers of relapse	A barrier to effective cancer treatment is the failure to prevent a relapse after initial treatment, even after complete remission. Using targeted single-cell DNA sequencing to study cells present in pretreatment and relapsed samples, cancer researchers can elucidate the molecular factors that underlie a relapse.

Understanding rare cell populations

Researchers are now linking rare cell populations in tumors with cancer progression and poor patient prognosis.[11] In certain cases, these rare cell populations within a tumor promote antitumor immunity.[12] In either case, identifying and understanding the role these rare cell populations play in cancer can be helpful to advancing treatments.

Identifying specific tumor-reactive T-cell profiles

CD8+ T-cells recognize neoantigens that are presented on the surface of tumor cells. This is a major mechanism that explains how immunosurveillance in the tumor microenvironment is modulated. The recognition of neoantigens by diverse T-cell receptors, which originate from V(D)J recombination, is a hallmark of the adaptive immune response. Using single-cell RNA-Seq (scRNA-Seq), researchers can elucidate specific TCR sequences that drive neoantigen recognition, as well as clonal expansion and TCR diversity. Consequently, this may help to drive an understanding of the key players in cancer immunosurveillance and immunotherapy.

scRNA-Seq: Step-by-step overview

In this guide, we provide an overview of single-cell RNA sequencing (scRNA-Seq) in cancer research. For a complete overview of single-cell sequencing techniques, download the single-cell sequencing eBook here.

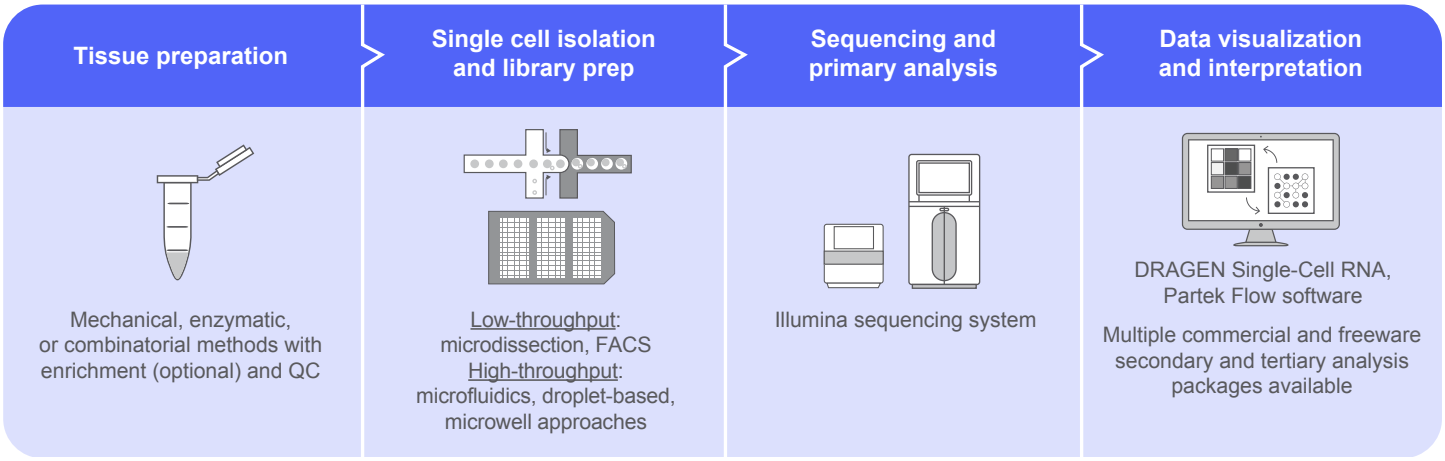


Figure 2: Single-cell analysis workflow

STEP 1 Single-cell tissue preparation

Successful single-cell sequencing requires viable individual cells. However, recovering intact, viable, and whole cells from tumors is difficult with the cells regularly contaminated with extracellular DNA/RNA from lysed cells. This can result in artifacts and lead to the generation of unreliable data. Some cancer researchers have found it more useful to isolate cell nuclei instead of individual cells. Isolation of nuclei enables preparation of fixed tissues and reduces the dependence on cell viability and integrity. After cells are dissociated from solid tumors, there are several techniques that may be used to isolate single cells.

STEP 2 Single-cell isolation

Advances in microfluidics technologies have enabled high-throughput single-cell profiling where researchers can cost-effectively examine hundreds to tens of thousands of cells per experiment. However, many use cases and methods do not require specialized equipment or high throughput, and new approaches have made single-cell sequencing accessible for nearly any research lab. This guide covers some of the most common single-cell methods, but our eBook on single-cell sequencing provides more detailed insights into a wide array of single-cell approaches.

Table 3: Single-cell isolation methods compared

Method	Commercial offerings	Advantages
Droplet fluidics platforms	• **10x Genomics Chromium Controller and Chromium X Series** • **Mission Bio Tapestri Platform**	Unique molecular identifiers (UMIs) and cell barcodes enable cell- and gene-specific identification, low cost per cell, and extensive support from commercial providers.
Microwells	• **BD Rhapsody Single-Cell Analysis System**	Supports imaging and short-term cell culture. Ideal for adherent cells.
Flow cytometry	• **Any commercial cell sorter (eg, BD FACSAria III Cell Sorter)**	Ideal for adherent primary tumor cells and cell culture lines.
Split-pool combinatorial barcoding	• **Parse Bioscience Evercode Whole-Transcriptome** • **Scale Biosciences ScaleBio Single Cell RNA Sequencing Kit**	Fixation of samples reduces batch effects. Only standard laboratory equipment is required. Compatible with a wide range of cells and samples. Relatively low cost per sample.
Templated emulsification	• **Fluent Biosciences PIPseq Single Cell RNA kit**	Does not require specialized microfluidic devices, expertise, or hardware for single-cell encapsulation and barcoding of cDNA.

STEP 3 Library preparation

As mentioned earlier, DNA libraries are prepared from isolated single cells and used in the actual sequencing experiment. Depending on the ome you plan to study, there are various commercial offerings you can use for to prepare libraries (Table 4).

Table 4: Commercial library preparation offerings or published references for each NGS method

Biological level	Method	Description	Commercial offering/reference
Genome	Targeted DNA sequencing (DNA-Seq)	Detect known variants using targeted panels.	• Mission Bio Tapestri Platform
	Whole-genome sequencing (WGS)	Simultaneously detect single nucleotide variants and copy number variants from the same cells.	• BioSkryb ResolveDNA Whole Genome Amplification Kit
Epigenome	Assay for transposase-accessible chromatin sequencing (ATAC-Seq)	Understand genome-wide chromatin accessibility within a single cell.	• Illumina Tagment DNA TDE1 Enzyme and Buffer Kits • 10x Genomics Chromium Nuclei Isolation Kit • Scale Biosciences Single Cell Methylation Kit
Transcriptome	RNA sequencing (RNA-seq)	Capture mRNA by 3’ and 5’ ends, enabling sequencing of the coding transcriptome with strand-specific information.	• 10x Genomics Chromium Single Cell Gene Expression Solution (3’ WTA) • BD Rhapsody Whole-Transcriptome Analysis (WTA) Amplification Kit • Parse Evercode Whole-Transcriptome Kit • Honeycomb HIVE scRNAseq Kit • Scale Biosciences Single Cell RNA Kit • Fluent Biosciences PIPseq Single Cell RNA Kit
	Fixed RNA-Seq	Oligonucleotide probes capture protein-coding gene targets across the transcriptome.	• 10x Genomics Chromium Single Cell Gene Expression Flex Fixed RNA assay
	Immune-repertoire sequencing (IR-Seq)	IR-Seq is a targeted sequencing method used to quantify the composition of B- or T-cell antigen receptor repertoires.	• 10x Genomics Chromium Single Cell Immune Profiling Solution
Proteome	Antibody sequencing (Ab-Seq)	DNA-tagged antibodies enable protein profiling by NGS.	• BD AbSeq Assay • BioLegend TotalSeq-A, B, or C reagents
Multiome	Cellular indexing of transcriptomes and epitopes (CITE-Seq)	CITE-Seq uses oligonucleotide-labeled antibodies to convert protein detection into a quantitative assay by NGS.	• BioLegend TotalSeq-A, B, or C reagents • 10x Genomics Single Cell Immune Profiling or Single Cell Gene Expression
	Combined whole genome amplification (WGA) and primary template-directed amplification (PTA)	WGA and PTA (with full-transcript reverse transcription) allows for comprehensive single-cell multiomic analysis.	• BioSkryb ResolveOme

STEP 4 Sequencing

Following single-cell isolation and library preparation, DNA libraries are sequenced using flow cells and the appropriate NGS method.

The number of cells being evaluated will vary depending on the experimental design. Researchers must consider their ability to detect desired cell types within a quantity of cells when designing single-cell sequencing experiments. Table 5 below shows the recommended number of read pairs per cell or nucleus using various single-cell sequencing methods.

A "read" refers to the nucleotide sequence from a fragment of DNA detected by the sequencer.

Table 5: The recommended number of read pairs per cell using different single-cell sequencing methods

Single-cell Method	Recommended number of reads[a] per cell
3' gene expression	15,000–50,000
5' gene expression	50,000
RNA-Seq	10,000–50,000
10x Flex	10,000
IR-Seq	5000
scATAC-Seq	50,000 per nucleus
Methyl-Seq	1M reads per cell
Antibody sequencing (Ab-Seq)	100 per antibody per cell
Targeted DNA-Seq	Depends on the panel size

a. Reads refers to either single-end reads or read pairs.

Using fixed single-cell RNA-Seq as an example, Table 6 below shows how many read pairs a sequencing system could detect per flow cell and the number of cells that can be sequenced per flow cell.

Table 6: Number of reads, cells per flow cell, and read length assuming 10,000 reads per cell and a 10x Genomics Flex protocol

System	Illumina flow cell	Reads[a] per flow cell	No. of cells per flow cell
NextSeq 1000/2000 Systems	P2	400M	40K
	P3[b]	1.2B	120K
	P4[b]	1.8B	180K
NovaSeq 6000 System	SP	800M	80K
	S1	1.6B	160K
	S2	4B	400K
	S4	10B	1M
NovaSeq X Series	1.5B	1.6B	160K
	10B	10B	1M
	25B	26B	2.6M

a. Reads refers to either single-end reads or read pairs.
b. P3 and P4 flow cells only available on the NextSeq 2000 System.

STEP 5 Single-cell analysis and visualization

After the single-cell sequencing run is complete, data analysis can be performed. Generally, the analysis pipeline for single-cell sequencing experiments involves three phases:

1. Primary analysis: Raw sequencing data files are converted for downstream analysis.
2. Secondary analysis: This includes demultiplexing, alignment, and genetic characterization.
3. Tertiary analysis: This is where data is visualized and interpreted.

There is no one correct way to execute the analysis pipeline for single-cell sequencing experiments. Many approaches and software programs are available for each step in the pipeline. The analysis pipeline you choose will be based on your research objective. The single-cell eBook and this video demonstrate a single cell RNA analysis workflow and interpretation of results using Illumina Connected Analytics.

			DRAGEN Single Cell RNA	Partek Flow
1	**Primary analysis:** File conversion .bcl file to .fastq file	**File conversion** Raw data files (BCL) are converted to FASTQ format for downstream analysis.	✓	✓
2	**Secondary analysis:** Demultiplexing (if applicable)	**Demultiplexing** If the samples were multiplexed for sequencing, resulting read files are demultiplexed prior to downstream analyses.	✓	✓
3	**Sequence alignment**	**Sequence alignment** The reads are mapped and aligned to a reference genome.	✓	✓
4	**Data set QC and filtering**	**Data set QC and filtering** Noncellular barcodes and low-quality cells are excluded from downstream analysis by various metrics	✓	✓
5	**Initial genetic characterization**	**Genetic characterization** Quality-controlled data sets are analyzed for genomic variants, gene expression, chromatin accessibility, protein expression, etc.	✓	✓
6	**Tertiary analysis:** Data visualization and interpretation	**Data visualization** Multidimensional data plots enable the clustering of cells and identification of subpopulations.	Provides basic visualization and cell type clustering, adding immediate biological context to data	✓

Figure 3: Sequencing data analysis pipeline and capabilities of the DRAGEN Single Cell RNA app and Partek Flow software

Downstream tertiary analysis and visualization solutions

There are many options for single-cell tertiary analysis tools, including free open-source analysis tools developed by academic labs. Most free programs are based on popular programming languages like R and Python. There are also “plug-and-play” packages that allow researchers to use preconfigured analysis workflows and commercial offerings. The tools you choose will depend on your research goals and experimental objectives.

Method 2 Cellular indexing of transcriptomes and epitopes (CITE-Seq)

CITE-Seq uses oligonucleotide-labeled antibodies to measure proteins and RNA in single cells in the same experiment. CITE-Seq is a high-throughput multiomics tool that allows cancer researchers to study protein expression and the intricacies of the cellular transcriptome both at the single-cell level and for spatial analysis.[13]

This combined proteomics/transcriptomics approach allows researchers to tie RNA expression directly to a cancer phenotype.

Key benefits of using CITE-Seq over independent transcriptomic (ie, RNA-Seq) or proteomic (ie, mass spectrometry) approaches include:

- Provides the entire picture of cell function by accurately measuring both transcript and protein expression.
- Adding a proteomics view to RNA analysis allows for measurement of post-transcriptional and translational modifications, such as protein degradation, isoform detection, and glycosyation.
- Solves the problem of detecting a limited number of proteins while using single-cell sequencing to study transcripts in an unbiased way by using barcoded antibodies to analyze a near-limitless number of cell markers and transcripts in a single run.

Potential applications of CITE-Seq in cancer research

Disease monitoring during remission

Relapse rates for certain cancers like acute myeloid leukemia are high. CITE-Seq could be used to study and monitor patients in remission to identify biomarkers that may be related to relapse.[13]

Study therapy resistance mechanisms through the course of treatment

Some cancer types resist even the most potent cancer medications. CITE-Seq might provide useful single-cell and spatial insights into the mechanisms underlying the resistance.[14]

A step-by-step overview of the CITE-Seq workflow

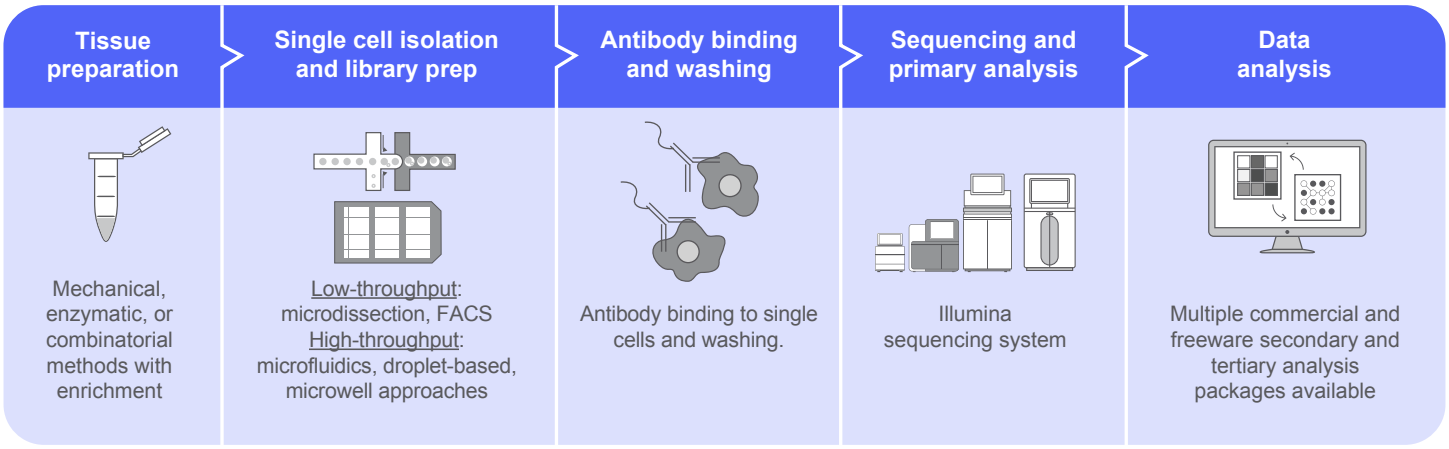


Figure 4: The general CITE-Seq workflow

STEP 1 Sample preparation

Whether you plan on doing single-cell analysis or spatial analysis, there are kits to support your experimental path. Several third-party products can be used to prepare libraries that are compatible with Illumina systems (see table below).

Table 7: Recommended CITE-Seq library preparation kits

Single-cell analysis library prep kits
BioLegend TotalSeq-A, -B and -C Reagents
BD AbSeq Assay
10x Single Cell Profiling
10x Genomics Single Cell Gene Expression

STEP 2 Sequencing

Table 8: Recommended sequencing systems for CITE-Seq

Sequencing system[a]	Advantages
NextSeq 1000/2000 Systems	Cost efficient, mid-throughput
NovaSeq X Series	High throughput that scales with your needs

a. Both systems can be used for single-cell and spatial analysis.

STEP 3 Data analysis

The secondary or tertiary data analysis method you will use for CITE-Seq data will depend on the library prep method.

Table 9: Data analysis software for CITE-Seq based on the library preparation method

Data analysis method	CITE-Seq library prep kit	Secondary/tertiary analysis
Single-cell analysis	BioLegend TotalSeq-A, -B, and -C Reagents	BioLegend Multiomics Analysis Software
	BD AbSeq Assay	BD SeqGeq Software
	10x Genomics Chromium Single Cell Immune Profiling	10x Genomics Cell Ranger Software
	10x Genomics Single Cell Gene Expression	10x Genomics Cell Ranger Software

Method 3 Bulk epitope and nucleic acid sequencing (BEN-Seq)

BEN-Seq enables researchers to study both the RNA and protein profiles of cell populations. BEN-Seq uses oligonucleotide-linked antibodies to detect proteins within a sample. The oligos can then be used to prepare libraries that can consequently be sequenced.

The key benefits of BEN-Seq over single parameter methods (eg, single-cell genomics) include:

- The ability to detect significantly more proteins (hundreds) more efficiently than traditional methods like Western Blotting or ELISA.
- The opportunity to achieve similar levels of accuracy for protein measurement as flow cytometry with the added benefit of measuring RNA expression on a high-throughput scale.

Potential applications of BEN-Seq in cancer research

Researchers will be able to perform multiomics experiments for a more comprehensive study of cancer cells. Various tumor processes, including initial formation, progression, immune evasion, angiogenesis, and metastasis, involve dynamic changes in the genome, transcriptome, and proteome of cancer cells. BEN-Seq enables simultaneous RNA and protein analysis for detailed cancer cell or tumor analysis.

BEN-Seq provides researchers with another avenue to discover new biomarkers and targets for therapy. By integrating protein and RNA analyses into a single assay, BEN-Seq has the potential to accelerate identification of prognostic, predictive, and diagnostic cancer biomarkers, as well as uncover novel targets for therapeutics.

Researchers can link tissue-specific gene expression with specific protein biomarkers.

The ability to measure the functional consequences of cancer mutations is another excellent reason to leverage BEN-Seq.

BEN-Seq: Step-by-step overview

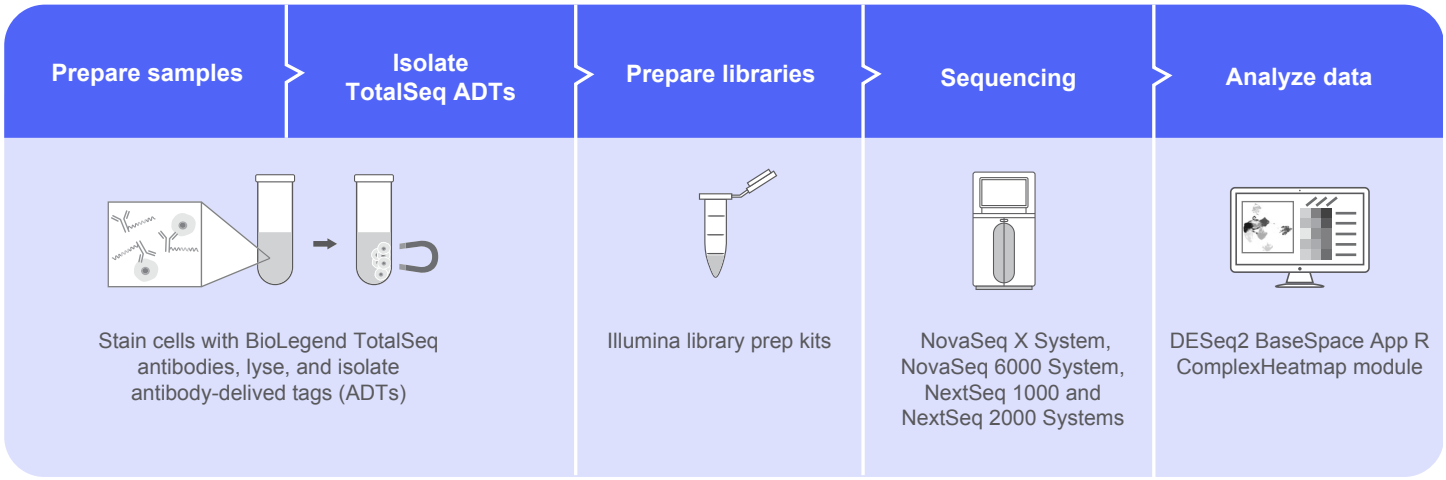


Figure 5: An overview of the BEN-Seq workflow

STEP 1 Sample preparation and antibody staining

Illumina recommends following the protocol for BioLegend's TotalSeq Kit designed specifically for protein analysis. Illumina recommends an input of 8000 cells. RNA and protein-associated libraries (antibody-derived tags or ADT) are separated using streptavidin-magnetic beads loaded with complementary oligonucleotides according to the protocol.

STEP 2 Recommended library preparation materials

Table 10: Recommended library preparation for BEN-Seq

Single-cell analysis library prep kits
BioLegend TotalSeq-A Reagents
Illumina Stranded mRNA Prep
Illumina RNA Prep with Enrichment for formalin-fixed, paraffin-embedded (FFPE) samples

STEP 3 Recommended sequencing systems

Table 11: Recommended sequencing systems for BEN-Seq

Sequencing system	Illumina flow cell	Reads[a] per flow cell	Reads[a] per sample (RNA panel)	Reads[a] per sample (protein panel)	No. samples per flow cell	Recommended read length
NextSeq 1000/2000 Systems	P2	400M	50M	1M	~8	2 × 101 bp
	P3[b]	1.2B			~23	
	P4[b]	1.8B			~35	
NovaSeq 6000 System	S1	1.6B			~31	2 × 101 bp
	S2	4.1B			~80	
	S4	10B			~196	
NovaSeq X Series	1.5B	1.6B			~31	2 × 101 bp
	10B	10B			~196	
	25B	26B			~510	

a. Reads refers to single-end reads or read pairs.
b. P3 and P4 flow cells only available on the NextSeq 2000 System.

STEP 4 Data analysis and visualization

Illumina recommends using the DESeq2 on BaseSpace Sequence Hub for data analysis and the third-party R ComplexHeatmap module software for data visualization.

Table 12: Data analysis and visualization software for BEN-Seq

Software	Application	Input
DESeq2	Differential expression analysis on aligned RNA samples	Output from RNA-Seq Alignment app
R ComplexHeatmap	Flexible generation, arrangement, and annotation of heat maps	

Method 4 Spatial sequencing

Spatial sequencing allows cancer researchers to study gene and/or protein expression within the natural tumor microenvironment and architecture. A common method for spatial sequencing is to use a microscope slide that is coated with spatially barcoded bead arrays to capture RNA from frozen histological tissue sections. The captured RNA then undergoes reverse transcription into cDNA that can be sequenced. Spatial transcriptomics and proteomics allow researchers to get as close to the realistic processes that happen in whole organismal systems as possible.

Key benefits of using spatial transcriptomics over a method like single-cell RNA-Seq include:

- The conservation of the spatial context for gene expression, precluding the need for tissue dissociation.
- RNA profiling can be combined with immunofluorescence or histochemical staining and imaging on the same sample.
- Obtaining results rapidly with streamlined, automation-compatible workflows.

Potential applications of spatial analysis in cancer research

Identify biomarkers related to tissue location

Cellular position within a tissue or tumor is lost in bulk-cell and single-cell analysis. Spatial analysis preserves morphological context to enable researchers to identify and profile therapeutically important biomarkers.

Map unique tumor and tumor microenvironment molecular profiles

Tumor dissociation before bulk-cell or single-cell analysis destroys the morphology of the tissue. Spatial analysis enables delineation and mapping of the tumor in its surrounding microenvironment.

Understanding tumor cell microenvironment

In addition to the various cell populations found within a tumor, it's important to understand the structure and function of the extracellular matrix that defines a tumor's microenvironment. Insights into the tumor microenvironment may help cancer researchers understand why particular cell populations exist in a location and why genes are regulated differently in those environments.

Track and profile tumor-infiltrating lymphocytes

Tumor-infiltrating lymphocytes (TILs) have been shown to be correlated with improved clinical outcomes for various cancers.[15] However, the ability to monitor TILs is challenging with other NGS methods. Spatial analysis provides cancer researchers with the opportunity to identify and profile TILs within the morphological context of the tumor for a highly accurate assay of immune function in response to cancer.

Spatial analysis: Step-by-step overview

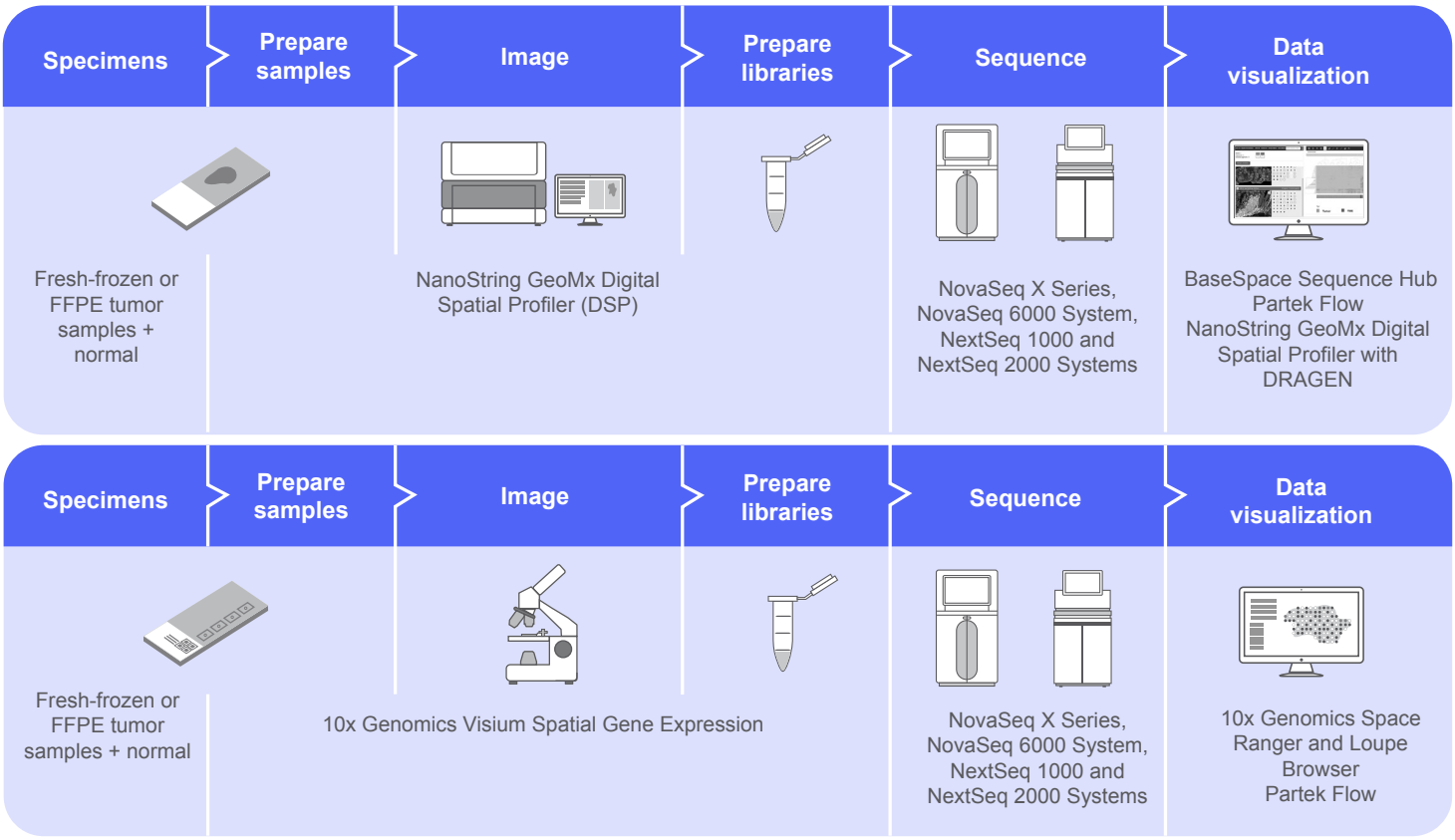


Figure 6: An overview of the spatial analysis workflow

STEP 1 Sample preparation

Researchers use tissue sections mounted on slides for spatial analysis. Illumina recommends following the respective protocols for either the 10x Genomics Visium Spatial Gene Expression or NanoString GeoMx Digital Spatial Profiler (DSP) for sample preparation. Options are available for targeted gene expression analysis. Additional devices such as the 10x CytAssist may be useful tools to prepare spatial samples from common clinical research sample types, such as FFPE.

Table 13: Sample preparation kits for spatial analysis

Provider	Panel	Description
10x Genomics Visium or Visium HD	Spatial Gene Expression	Captures whole-transcriptome gene expression
	Human Pan-Cancer Panel	Targets approximately 1200 genes associated with tumors, the tumor microenvironment, and immune response
NanoString GeoMx DSP	GeoMx Whole Transcriptome Atlas	Enables profiling of over 18,000 protein coding genes
	GeoMx Cancer Transcriptome Atlas	Enables profiling of over 1800 gene related to the immune response, tumor biology, and the microenvironment

STEP 2 Imaging

If you prepare your samples using NanoString GeoMx DSP, you can image both protein and RNA using RNAscope. For samples prepared using 10x Genomics Visium, imaging can be performed on a standard microscope.

Table 14: Imaging in spatial analysis

Product	NanoString GeoMx DSP	10x Genomics Visium
Most important to me	Imaging for both protein and RNA using RNAscope	Imaging can be performed with a standard microscope

The following application notes provide additional information on using 10x Genomics Visium and NanoString GeoMx DSP:

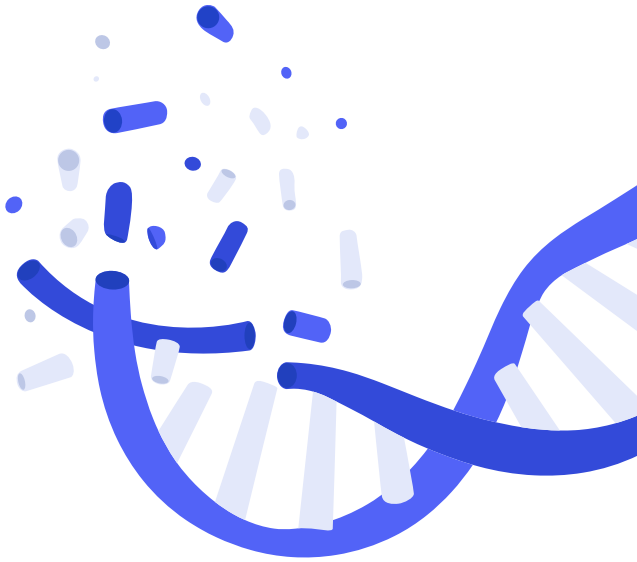
- Resolve The Whole-Transcriptome Within Tissue Architecture (10x Visium)
- High Resolution Spatial Transcriptomics of Complex Tissues (GeoMx DSP)

STEP 3 Library preparation

Illumina recommends following the respective manufacturer's protocol for library preparation.

Table 15: Library preparation methods for spatial analysis

Product	NanoString GeoMx DSP	10x Genomics Visium
Most important to me	Allows researchers to sequence a region of interest on the slide	Allows researchers to sequence the whole slide
Sample type	Fresh-frozen or FFPE	Fresh-frozen or FFPE



STEP 4 Sequencing

We recommend the following sequencing systems based on your research needs.

Table 16: Sequencing systems for spatial analysis

Product	NextSeq 1000 and NextSeq 2000 Systems	NovaSeq 6000 System and NovaSeq X Series
Most important to me	Instrument affordability and desktop footprint	Low cost/sample
10x Genomics reads[a] per flow cell	150M reads per section, four sections per slide 275M reads per capture area fully covered with tissue for Visium HD	
NanoString reads[a] per flow cell	200M reads per 24 ROI[b] for whole-transcriptome atlas 50M reads per 24 ROI[b] for cancer transcriptome atlas	
Recommended read length for 10x Genomics	2 × 151 bp	2 × 151 bp
Recommended read length for NanoString	2 × 35 bp	2 × 35 bp

a. Reads can refer to single-end reads or read pairs.
b. ROI, regions of interest.

STEP 5 Visualization

Illumina recommends using the respective analysis software for each manufacturer.

Table 17: Data visualization recommendations for spatial analysis

Software	Application	Input
10x Genomics Space Ranger	Automated overlay of spatial gene expression data on tissue images and analysis	10x Genomics Visium Spatial Gene Expression data files
10x Genomics Loupe Browser	Visual exploration of spatial expression data	Data output from Space Ranger software
NanoString GeoMx DSP Data Analysis	Automated analysis and visualization of spatial data	NanoString GeoMx data files
Partek Flow	Easy-to-use interface and rich interactive visualizations for start-to-finish spatial transcriptomics analysis	10x Genomics Visium and Xenium, NanoString CosMx data files

Genomics methods

Method 5 Whole-genome sequencing (WGS)

WGS provides an overview of the genetic make-up of an organism or tissue. In cancer research, it allows researchers to study cancer-driving genetic events that go beyond protein-coding variants.

Key benefits of WGS over targeted molecular panels include:

- Assessing the full genomic backbone of an organism or tissue for unbiased analysis and potential discovery of novel cancer-associated genes
- Detecting genomic signatures that may not yet be linked to cancer phenotypes
- Discovering noncoding regions of the genome that may influence cancer progression
- Discovering noncoding regions of the genome that may influence cancer progression
- Uncovering genome-wide integration sites of oncogenic viruses

Potential application of WGS in cancer research

Elucidate the role of noncoding mutations in cancer	The link between somatic mutations and cancer progression has largely focused on the role of coding mutations on cellular processes. Yet only 2% of the human genome codes for proteins. Recent research shows that noncoding DNA variants can also impact carcinogenesis. While most noncoding mutations may simply be passenger mutations, some may be regulatory elements that impact cancer progression. WGS can identify coding and noncoding somatic mutations. Using computational techniques, researchers can classify the importance of these mutations to cancer progression.
Link somatic mutational signatures with cancer progression	Somatic mutations that contribute to cancer can arise for various reasons and through various processes. Each process may lead to a distinct mutational signature. WGS is a sensitive method to identify these mutational signatures.[16,17] Understanding these signatures can lead to novel insights into cancer progression and new therapeutic targets.
Identify the role of viral integration in cancer	Researchers have known about the role of genomic viral integration in cancer progression for many years. Unfortunately, systematic studies that help scientists understand the role of viral integration is lacking. WGS can be used to identify viral integrations across the genome. This hypothesis-free approach can elucidate associations between certain viruses and cancer progression.

WGS: Step-by-step overview

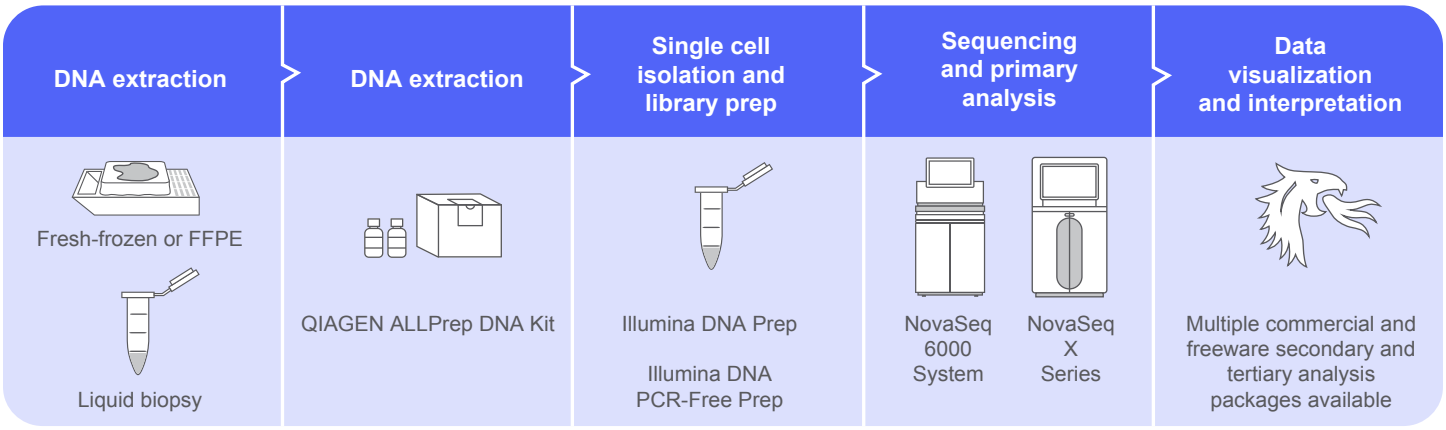


Figure 7: An overview of the whole-genome sequencing workflow

STEP 1 Recommended extraction methods

There are several DNA extraction methods that can be used for tissue and liquid biopsy samples. For tissues samples, Illumina recommends the QIAGEN AllPrep DNA Kit. This kit provides high-quality and high-quantity DNA.For analysis of circulating tumor cells (CTCs), cell-free DNA (cfDNA), and circulating tumor (ctDNA) using liquid biopsy, the Illumina DNA Prep kit or Illumina DNA PCR-Free Prep kit are recommended.

STEP 2 Library preparation

Table 18: Recommended library prep kit for WGS

Product	Illumina DNA Prep
Input requirements	100–500 ng
Total library prep time	3–4 hours
Sample type	gDNA from blood, FFPE and fresh-frozen samples
Sample index sets	384 unique dual indexes

STEP 3 Recommended sequencing systems

Table 19: Recommended sequencing system for WGS

System	Illumina flow cell	Reads[a] per flow cell	Reads[a] per sample	No. samples per flow cell	Recommended read length	Recommended read depth
NovaSeq 6000 System	S1	1.6B	400M	4	2 × 101 bp	At least 30× coverage
	S2	4.1B		9		
	S4	10B		24		
NovaSeq X Series	1.5B	1.6B		~4	2 × 101 bp, up to 150 bp	
	10B	10B		~24		
	25B	26B		~64		

a. Reads refers to either single-end reads or read pairs.

STEP 4 Secondary analysis

Illumina recommends using the DRAGEN pipelines on the NovaSeq X Series, on the BaseSpace Sequence Hub, or on a DRAGEN server, to obtain somatic variant calls and gene expression data. When using BaseSpace Sequence Hub, you can monitor runs in real time while securely streaming data from the instruments to the ecosystem for pushbutton analysis.

Secondary analysis with DRAGEN on Illumina Connected Analytics

The use of multiomics methods has allowed researchers to generate data at unprecedented levels. This large amount of data being generated far outpaces any organization's ability to extract relevant biological insights.

We built Illumina Connected Analytics for researchers who are interested in analyzing data at scale. Connected Analytics is a comprehensive cloud-based data management and analysis platform that allows you to share, aggregate, and explore large volumes of sequencing data in a secure, scalable, and flexible environment.

Connected Analytics analysis pipelines are highly customizable and integrate DRAGEN analysis tools. Connected Analytics is also incredibly secure, making it a great option for clinical research institutions.

- Direct integration with the data generation workflow using Illumina sequencing systems
- Powerful secondary analysis with the DRAGEN analysis tools
- Scalable data aggregation
- Secure data storage
- A dynamic and interactive data science environment for advanced machine learning and artificial intelligence

For more information on using Connected Analytics, read the Illumina Connected Analytics guide.

Table 20: Secondary analysis software for WGS

Pipeline	Application	Input
DRAGEN Somatic pipeline	Somatic variant detection in tumor samples; includes tumor-only and tumor–normal modes. This pipeline supports UMIs, HRD scoring, MSI and TMB calling, and HLA typing.	Tumor DNA FASTQ Normal DNA FASTQ

Enhanced WGS with Illumina Complete Long Reads

For germline variant calling in cancer research, long reads can help with phasing of inherited alleles and can enhance coverage of transposable elements, repeats, or other difficult-to-map regions. Illumina Complete Long Reads make long-read sequencing accessible and streamlined by enabling both short reads and long reads on the NovaSeq platforms. Illumina Complete Long Read Prep, Human is our most accurate and most comprehensive human whole-genome assay. Illumina Complete Long Read Prep with Enrichment, Human is a flexible, cost-effective solution to enhance coverage with targeted long reads where they matter most.

Method 6 Whole-exome sequencing (WES)

WES allows researchers to analyze the portion of the genome responsible for coding proteins (the exome). Sequencing the cancer exome provides useful information about the coding mutations that contribute to cancer progression.

Key benefits of WES include:

- Identifying variants across a wide range of applications.
- Because the exome represents less than 2% of the genome, sequencing it is a cost-effective alternative to whole-genome sequencing (4–5 Gb of sequencing data versus up to 90 Gb of data for the whole human genome).
- It provides complete coverage of the coding regions of the genomes, enabling deeper sequencing that allows for detection of rare SNPs and variants.
- Because of the above, the data generated from WES is more manageable and is easier to analyze.

Potential applications of WES in cancer research

Biomarker discovery	Cancer researchers can use WES to uncover tumor-associated gene expression profiles.[18]
Capturing variant data from blood	Historically, most cancers are diagnosed after a tissue biopsy. This often requires an invasive process (sometimes surgical) to acquire the biopsy. Discovering genetic variants in blood samples of patients using WES is minimally invasive and may help researchers capture predictive or diagnostic biomarkers of disease.[19]
Informing research into cancer susceptibility	Late-stage diagnosis of cancer is associated with poorer prognosis, but biological features associated with early disease aren't always clear. Research into factors associated with cancer susceptibility and early incidence is an important step toward future translational insights and targetable genes or pathways.[20,21]

WES: Step-by-step overview

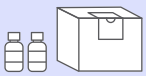
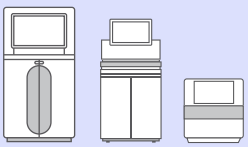
DNA extraction	Single cell isolation and library prep	Sequencing and primary analysis	Data visualization and interpretation
			
QIAGEN ALLPrep DNA Kit	Illumina DNA Prep with Exome 2.5 Enrichment	NovaSeq X Series, NovaSeq 6000 System, NextSeq 1000 and NextSeq 2000 Systems	Multiple commercial and freeware secondary and tertiary analysis packages available

Figure 8: An overview of the WES workflow

STEP 1 DNA extraction

DNA can first be extracted from the sample using the QIAGEN ALL Prep DNA Kit or the Illumina Flex Lysis Reagent Kit

STEP 2 Library preparation

After DNA extraction, libraries for WES can be prepared using the Illumina DNA Prep with Exome 2.5 Enrichment.

Table 21: Library preparation for WES

Product	Illumina DNA Prep with Exome 2.5 Enrichment
Input requirements	50–1000 ng
Total library prep time	6.5 hours
Sample type	High-quality gDNA from blood or saliva
Sample index sets	384 dual indexes

STEP 3 Sequencing

Illumina recommends the NovaSeq 6000 System or NovaSeq X Series for WES.

Table 22: Sequencing systems for WES

System	Illumina flow cell	Reads[a] per flow cell	Reads[a] per sample	Recommended read length	Recommended read depth	Exomes per single flow cell run
NovaSeq 6000 System	S1	1.6B	~40M	2 × 101 bp	100× mean target coverage	~40
	S2	4.1B				~102
	S4	10B				~250
NovaSeq X Series	1.5B	1.6B				~40
	10B	10B				~250
	25B	26B				~650

a. Reads refers to either single-end reads or read pairs.

STEP 4 Data analysis

We recommend using the DRAGEN platform either on the BaseSpace Sequence Hub or on a DRAGEN server to obtain data from WES. In BaseSpace Sequence Hub, you can monitor runs in real time while securely streaming data directly from the instruments into the ecosystem.

Epigenomics methods

Method 7 Assay for transposase-accessible chromatin sequencing (ATAC-Seq)

Cancer researchers can use ATAC-Seq to study epigenetic features across the genome, without prior knowledge of regulatory elements. ATAC-Seq exposes genomic DNA to Tn5, a highly active transposase that preferentially inserts into open chromatin sites and adds sequencing primers. Subsequent NGS analysis, which includes genomic or transcriptomic profiling, provides insights into chromatin accessibility across the genome. While several traditional methods like chromatin immunoprecipitation sequencing (ChIP-Seq), formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-Seq), or DNase I hypersensitive sites sequencing (DNase-Seq) can be used to study regions of chromatin–DNA interaction sites, ATAC-Seq illuminates regions of open chromatin.

Key advantages of ATAC-Seq:

- Researchers can avoid sensitive enzymatic digestion of chromatin or rigorous validation of antibodies by using the Tn5 transposase.
- Researchers can obtain results rapidly, with a streamlined workflow that can be completed in < 3 hours.
- It is possible to interrogate precious samples with input requirements as low as 500–50,000 cells.
- ATAC-Seq can be paired with RNA-Seq to directly determine if regions of open chromatin are being expressed.

Potential applications of ATAC-Seq in cancer research

Interrogate the chromatin regulatory landscape in cancers	Mutations in regulatory elements that affect transcription factor binding can result in increased or decreased chromatin accessibility to help drive tumor formation or cancer progression. ATAC-Seq enables surveillance of chromatin accessibility to identify patterns of gene regulation specific to different cancer types.
Map nucleosome redistribution during cancer progression	Nucleosomes are the structural subunit of chromatin, enabling packaging of long DNA molecules into more compact forms. They play a central role in epigenetic regulation. Changes in nucleosome positioning throughout the genome have been associated with dysregulation of gene expression that can contribute to tumor formation and cancer progression. ATAC-Seq provides genome-wide mapping of nucleosome distribution to help understand the etiology of cancer.
Elucidate the role of epigenetics in drug resistance	Epigenetics plays a key role in tumor progression and drug resistance. Drug-resistant tumor clones have been known to co-opt epigenetic pathways to suppress gene expression that drives drug sensitivity. ATAC-Seq can be used to profile these epigenetic changes to drive a deeper understanding of drug resistance mechanisms.
Find epigenetic therapies	Developing therapies for the treatment of cancers influenced by epigenetic mechanisms are an essential aspect of ongoing cancer treatment efforts. ATAC-Seq can be a useful tool in identifying new drug targets that influence the epigenome, such as enzymes responsible for the addition and removal of cancer-associated epigenetic modifications.

ATAC-Seq: Step-by-step overview

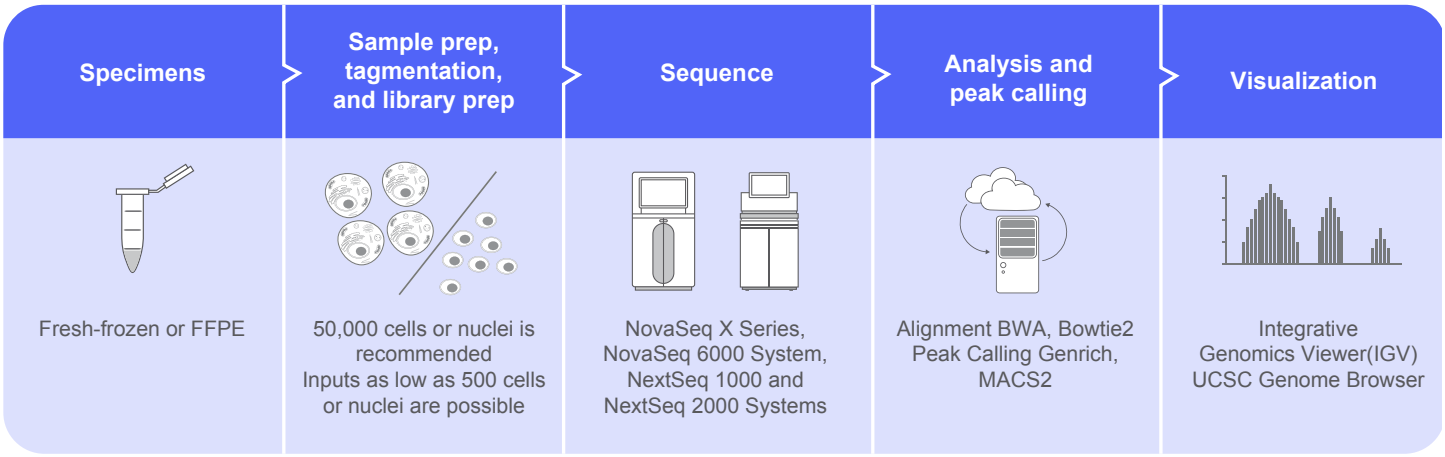


Figure 9: An overview of the ATAC-Seq workflow

STEP 1 Sample preparation

ATAC-Seq is performed on intact cells or nuclei, not on isolated DNA. The recommended minimum input requirements for ATAC-Seq of at least 50,000 cells or nuclei. While researchers can get results from a lower cell/nuclei input (as few as 500 cells), this may introduce artifacts due to low library complexity.

STEP 2 Tagmentation and amplification

Illumina recommends following the protocol included with the Illumina Tagment DNA TDE1 Enzyme and Buffer Kit (product numbers 20034197; 20034198) to generate a sequencing-ready library, with read depths greater than 50M reads per sample.

For more details on ATAC-Seq, refer to ATAC-Seq: A Method for Assaying Chromatic Accessibility Genome-Wide.

STEP 3 Recommended sequencing systems

Table 23: Sequencing systems for ATAC-Seq

Sequencing system	Illumina flow cell	Reads[a] per flow cell	Reads[a] per sample	No. samples per flow cell	Recommended read length
NextSeq 1000/2000 Systems	P2	400M	50M	8	2 × 151 bp
	P3[b]	1.2B		24	
	P4[b]	1.8B		36	
NovaSeq 6000 System	S1	1.6B		~30	
	S2	4.1B		~74	
	S4	10B		~180	
NovaSeq X Series	1.5B	1.6B		32	
	10B	10B		200	
	25B	26B		520	

a. Reads refers to either single-end reads or read pairs.
b. P3 and P4 flow cells only available on the NextSeq 2000 System.

STEP 4 Analysis

Illumina recommends using the Burrows-Wheeler Aligner (BWA) app on BaseSpace Sequence Hub for primary sequencing data analysis and the third-party Genrich or MACS2 software applications for secondary analysis.

Table 24: Recommended data analysis software for ATAC-Seq

Primary analysis: Alignment		
Software	Application	Input
Burrows-Wheeler Aligner (BWA)	Alignment to reference genome	FASTQ files
Bowtie2		
Secondary analysis: Peak calling		
Software	Application	Input
Genrich	Analysis of alignment files to call peaks of significant enrichment	Output files from alignment application
Model-based Analysis of ChIP-Seq (MACS2)		

STEP 5 Visualization

Illumina recommends the IGV app on BaseSpace Sequence Hub or UCSC Genome Browser for visualizing called peaks in a genomic context.

Table 25: Visualization software for ATAC-Seq

Software	Application	Input
Integrative Genomics Viewer (IGV)	Visual exploration of genomic data	Any project can be used as input, including BAM, VCF, BED, BW, and BEDGRAPH
UCSC Genome Browser		Multiple file types, including BAM, BED, and more
Partek Flow software		FASTQ

Studying the epigenome in cancer research

DNA methylation plays an essential role in regulating gene expression and facilitates responses to environmental stimuli. The accumulated mutations in cancerous cells normally precipitate changes in the DNA methylome and contribute to the abnormal genetic regulation observed in these cells. Alterations in DNA methylation are seen in all cancers, enabling the tumors to be subtyped according to their distinct methylation signatures. In general, hypermethylation represses transcription of the promoter regions of tumor suppressor genes while hypomethylation is linked with the expression of oncogenes.[6]

Method 8 Methylation arrays

Methylation arrays allow researchers to quantitatively interrogate methylation sites across the epigenome of cancer cells at single-nucleotide resolution.

Array-based solutions provide comprehensive genome-wide coverage that includes, but is not limited to, CpG islands, CHH sites, enhancers, open chromatin, and transcription factor binding sites. As a high-throughput research method, the cost is minimal per sample compared with methylation sequencing alternatives. Methylation array protocols have a user-friendly, streamlined workflow with > 98% assay reproducibility and support for FFPE samples, increasing applicability of methylation arrays to biobanked tissues.

Illumina Infinium™ MethylationEPIC v2.0 Kit is a genome-wide methylation screening tool that interrogates 935,000 CpG sites across the most biologically significant regions of the human methylome, including the human tumor epigenome. Moreover, the Infinium Mouse Methylation BeadChip Kit provides comprehensive coverage of the mouse methylome by analyzing 285,000 CpG sites and can be used for studying patient-derived xenografts, the tumor microenvironment, and pre-clinical testing of therapeutic agents. FFPE or fresh-frozen tissue are processed with bisulfite conversion using the Infinium workflow. Samples are subsequently scanned on Illumina iScan™ System. The NextSeq 550 System allows researchers to transition between array scanning and high-throughput sequencing.

Potential application of methylation arrays in cancer research

Characterize epigenetic patterns in cancer	Classify tumors and their subtypes	Screen cancer drug candidates before clinical trials
Studies of cancer epigenetics, such as aberrant methylation and altered transcription factor binding, can provide insight into important tumorigenic pathways.[22] Methylation arrays can assist translational researchers in tumor classification applications.	Methylation arrays enable the precise identification of tumor subtypes based on their distinct methylation profiles. This allows new tumor subtypes to be discovered. Tumor subclassification can also be used to effectively stratify patients into clinical trial groups.[23] Moreover, methylation profiles can be analyzed using the DKFZ classifier, a free online tool for DNA-methylation-based classification or tumors.[24]	The efficacy of new therapeutic agents can be measured by the reversion of DNA methylation abnormalities to a healthy state.[25] For preclinical testing, the Mouse Methylation BeadChip can be used for mouse models of cancer.

Methylation arrays: Step-by-step overview

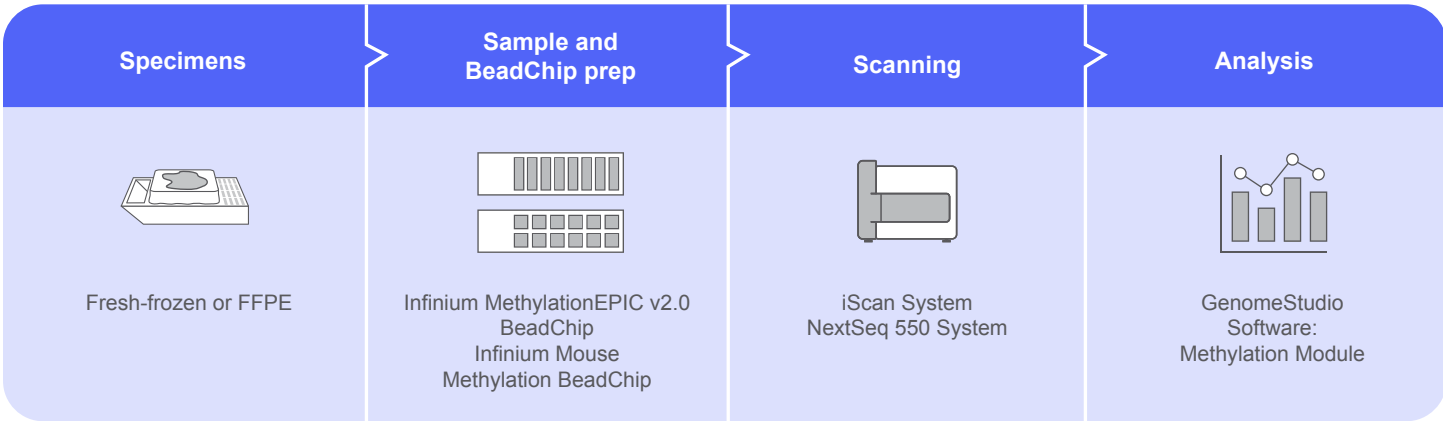


Figure 10: An overview of the workflow for methylation arrays *NextSeq 550: Only supported for MethylationEPIC v2.0.

STEP 1 Sample preparation

Illumina recommends an input of 250 ng genomic DNA. FFPE and fresh-frozen sections are supported.

STEP 2 BeadChip hybridization

After DNA samples are extracted, bisulfite conversion is performed with rapid manual or automated protocols from Zymo Research. Samples then undergo the Infinium workflow and are hybridized onto BeadChips. The Infinium MethylationEPIC v2.0 is used is used for analysis of human samples and the Infinium Mouse Methylation BeadChip supports studies of mouse models of cancer.

Table 26: Infinium MethylationEPIC BeadChips

Product	Infinium MethylationEPIC v2.0 BeadChip	Infinium Mouse Methylation BeadChip
Minimum input requirement	250 ng DNA	250 ng DNA
No. of markers	> 935K methylation sites per sample	> 285K methylation sites per sample
Sample type	Fresh-frozen or FFPE	Fresh-frozen or FFPE
No. of samples per array	8	12

STEP 3 Recommended array scanning systems

Table 27: Array scanning systems for methylation arrays

Product	iScan System	NextSeq 550 System[a]
Most important to me	High-throughput processing of BeadChips	Dual sequencing and array scanning capability
Scan time per BeadChip	20 minutes	40 minutes
Scan time per sample	~2.5 minutes	5 minutes

a. NextSeq 550 System only supports scanning for Infinium MethylationEPIC v2.0 and Infinium CytoSNP-850K BeadChip.

STEP 4 Analysis and visualization

Illumina provides a Methylation Module in GenomeStudio™ software for basic quality control. Freely available software packages such as SeSAMe and minfi provide the most advanced analysis capabilities, including normalization methods, differential methylation analysis, and visualization. For CNS and Sarcoma classifiers, raw MethylationEPIC data from tumor samples can be directly uploaded to MolecularNeuropathology.org or analyzed by the DKFZ classifier for quick and accurate classification.

Transcriptomics methods

Method 9 Bulk RNA sequencing(RNA-Seq)

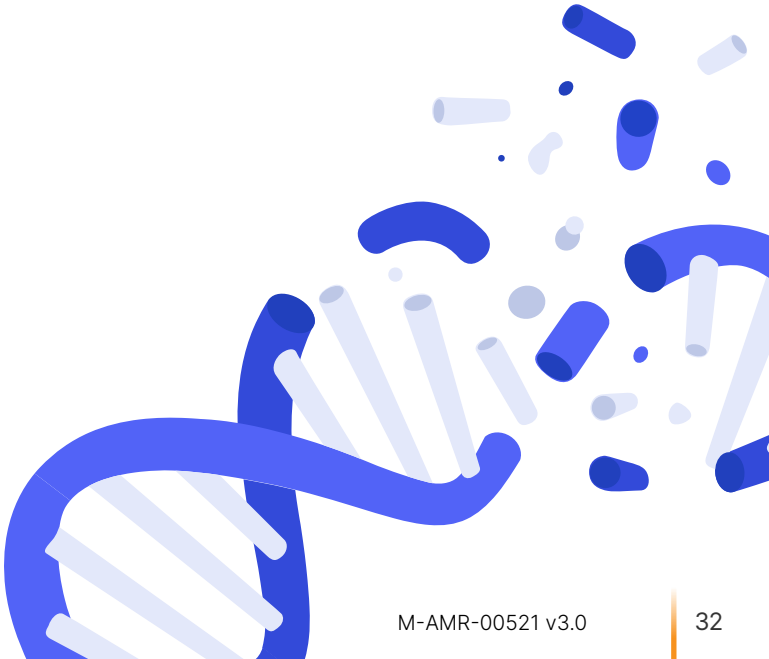
Bulk RNA sequencing (RNA-Seq) allows researchers to measure the average RNA expression within a certain cell population. In comparison to other NGS methods, bulk RNA-Seq has exceptionally well-defined workflows and has been implemented in thousands of publications. The ease of bulk RNA-Seq implementation combined with its comparatively low cost make it an ideal method for transcriptomics researchers who are new to NGS. Bulk RNA-Seq adds significant quality to translational and clinical cancer research and its use in such applications has been increasing.

Researchers have long depended on tools like qPCR to profile gene expression and regulation. However, methods like qPCR are inherently limited as researchers can only characterize a small number of targets that are preselected before an experiment. Bulk RNA-Seq removes substantial bias from transcriptomic experiments by addressing not only targeted hypotheses, but also transcripts and pathways that weren't even considered prior to experimental design. Field-defining projects such as The Cancer Genome Atlas were possible due to the comprehensive, large-scale transcriptomic characterization potential of RNA-Seq.

While bulk RNA-Seq lacks the resolution of single cell and spatial sequencing approaches, the method is frequently used to compare gene expression pattens between distinct cell populations (eg, different sections of the same tumor sample). Researchers use flow cytometry to sort cells into distinct populations and then use bulk RNA-Seq to identify transcriptomic differences between those populations. Another way cell populations can be isolated for RNA-Seq experiments is to use microdissection or laser dissection techniques to isolate different regions of tumors.

Key advantages of using bulk transcriptomics include:

1. Easily accessible and well-defined end-to-end workflows.
2. Relatively lower cost per experiment compared to other NGS methods.
3. Reduced experimental bias and amplified discovery power as compared to targeted methods such as qPCR.
4. Sensitive and accurate measurement of gene expression at the transcript level.
5. Both qualitative and quantitative data.
6. Scalable for large studies with high sample throughput.
7. High specificity and accuracy.
8. Maintains and tracks strand-specific information for both mRNA and Total RNA workflows.
9. Provides excellent performance with degraded RNA from samples such as FFPE tissue.



Potential applications of bulk RNA-Seq in cancer research

Differentiate driver mutations from passenger mutations

Although cancer cells contain many mutations, only a few contribute to cancer initiation and progression.[26] Researchers can use bulk RNA-Seq to examine the complete set of RNA transcripts produced by a genome to detect mutations. Data from the RNA-Seq experiment can then reveal whether these mutations result in transcriptomic changes that either drive cancer or simply act as passenger mutations.

Identify potentially druggable pathways

RNA-Seq can reveal pathways that are up- or down-regulated in cancer. This functional information is crucial for identification of molecular targets for precision therapeutics. Targeting upregulated pathways, for example, is a common method for suppressing tumor growth.[27,28]

Discovery of pathways associated with disease

Investigations into the transcriptomic differences between cancer samples and noncancerous tissue have been shown to be useful in differentiating cancer subtypes, assessing the impact of mutations, identifying biomarkers, and other variables.[29]

Assessment of therapy response

Bulk RNA-Seq can identify genes and pathways associated with biological response or lack of response to novel cancer therapies such as immunotherapeutics in model systems or retrospective study of tissue samples.[30]

STEP 1 RNA isolation

There are several options for RNA isolation kits depending on your sample type. We recommend the QIAGEN RNeasy Blood & Tissue Kit, QIAGEN RNeasy FFPE RNA, and QIAGEN RNeasy PowerMicrobiome kit.

STEP 2 RNA library prep

Table 28: Illumina recommended library prep kits for RNA-Seq

Product	Illumina Stranded mRNA Prep	Illumina Stranded Total RNA Prep w/ Ribo-Zero™ Plus	Illumina RNA Prep with Enrichment
Input requirements	25–1000 ng	1–1000 ng (10 ng for FFPE)	10–20 ng (20–100 ng for FFPE)
Total library prep time	6.5 hours	7 hours	9 hours
Sample type	High-quality RNA from any species with polyA tails	High-quality or degraded sample types, including FFPE; Ribo-Zero Plus includes rRNA depletion for human, mouse, rat, bacteria Gram +/-, and human beta globin transcripts	High-quality or degraded sample types, including FFPE; species compatibility dependent on panel used
Sample index sets	384 UDIs	384 UDIs	384 UDIs

STEP 3 Sequencing

The Illumina recommendation is calculated in terms of millions of reads to be sampled. Recommended read lengths depend on the protocol used for library preparation and the nature of the transcripts of interest. For example, 2 × 75 bp reads are compatible with stranded RNA prep protocols and 2 × 100 bp reads are best for RNA enrichment protocols. For optimal detections of events like fusions, 2 × 150 bp reads may be needed.

Detecting rarely expressed genes often requires an increase in the depth of coverage (the number of times a nucleotide is read during a sequencing experiment). Sequencing depth may be increased or decreased depending on the sample and rarity of transcripts of interest. Bioinformatic downsampling following control runs may be useful in determining whether events are detectable with fewer reads. There are several options to explore based on your RNA-Seq application. Highlighted below are recommendations based on common RNA-Seq projects.

Table 29: Benchtop sequencing recommendations for RNA sequencing

Sequencing system	Illumina library preparation kit	Illumina flow cell	Reads[a] per flow cell	Reads[a] per sample	No. samples per flow cell
NextSeq 1000/ 2000 Systems	Stranded Total RNA Prep w/ Ribo-Zero Plus	P2	400M	50M	8
		P3[b]	1.2B		24
		P4[b]	1.8B		36
	Stranded mRNA Prep	P2	400M	25M	16
		P3[b]	1.2B		48
		P4[b]	1.8B		72
	RNA Prep with Enrichment	P2	400M	25M	16
		P3[b]	1.2B		48
		P4[b]	1.8B		72

a. Reads refers to either single-end reads or read pairs.
b. P3 and P4 flow cells only available on the NextSeq 2000 System.

Table 30: Production-scale sequencing recommendations for RNA sequencing

Sequencing system	Illumina library preparation kit	Illumina flow cell	Reads[a] per flow cell	Reads[a] per sample	No. samples per flow cell	Recommended read length
NovaSeq 6000 System	Stranded Total RNA Prep w/ Ribo-Zero Plus	S1	1.6B	50M	32	2 × 75 bp
		S2	4.1B		82	
		S4	10B		200	
	Stranded mRNA Prep	S1	1.6B	25M	64	2 × 75 bp
		S2	4.1B		164	
		S4	10B		400	
	RNA Prep with Enrichment	S1	1.6B	25M	64	2 × 100 bp
		S2	4.1B		164	
		S4	10B		400	
NovaSeq X Series	Stranded Total RNA Prep w/ RiboZero Plus	1.5B	1.6B	50M	32	2 × 75 bp
		10B	10B		200	
		25B	26B		520	
	Stranded mRNA Prep	1.5B	1.6B	25M	64	2 × 75 bp
		10B	10B		400	
		25B	26B		1040	
	RNA Prep with Enrichment	1.5B	1.6B	25M	64	2 × 100 bp
		10B	10B		400	
		25B	26B		1040	

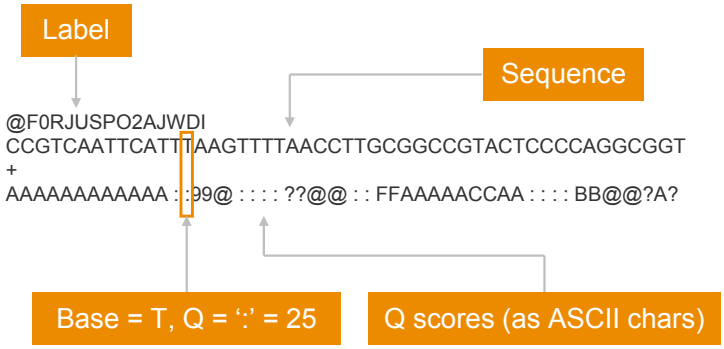
a. Reads refers to either single-end reads or read pairs.

STEP 4 Analysis

DRAGEN RNA and DRAGEN RNA Differential Expression Apps, Partek Flow software, and BaseSpace Correlation Engine for mRNA.

Primary analysis

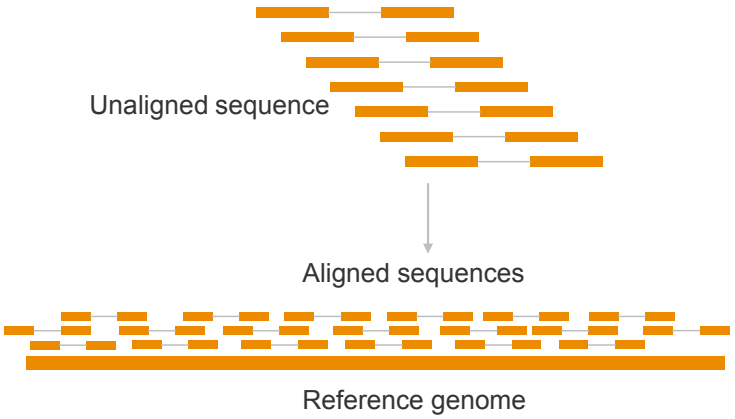
Base calling: Onboard software converts fluorescent signals from sequencing into As, Ts, Gs and Cs.



Secondary analysis

Demultiplexing: Separating samples based upon their unique barcodes.

Alignment: Determining where reads match within a reference genome.



Tertiary analysis

Making sense out of data by **visualizing** it, characterizing **expression patterns**, or subjecting it to more complex algorithms.

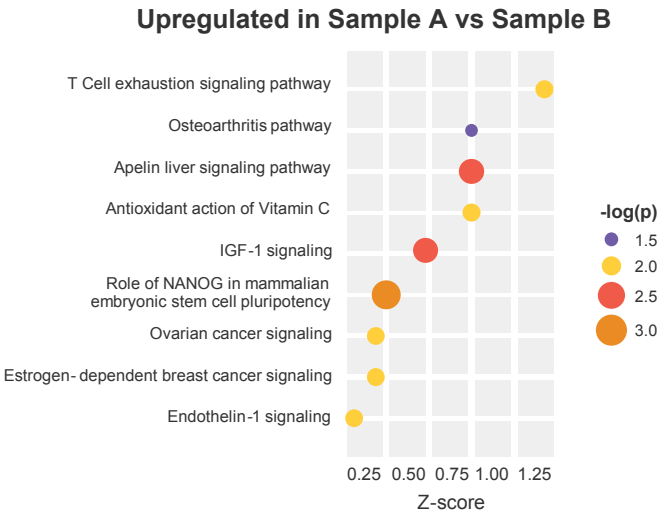


Figure 11: Data analysis involves primary, secondary, and tertiary analysis to ultimately visualize results

Method 10 Cell-free RNA sequencing (cfRNA-Seq)

NGS has been historically used to characterize DNA, RNA, and protein from cell or tissue samples. However, it is becoming clear that cancers also release analytes into the bloodstream and into other biofluids.[31] One such analyte is cell-free RNA (cfRNA), which has been shown to be released in exosomes and vesicles from both late- and early-stage tumors.[32]

Cell-free RNA is often abundant in blood plasma following a liquid biopsy. Advancements in cfRNA isolation and library preparation have allowed scientists to study and characterize cfRNA via NGS. Recent studies have shown that when it is found within exosomes, cfRNA is protected from degradation by RNases. Researchers now use cfRNA as a biomarker to identify specific cancers, detect cancer initiation, reveal the cell or tissue of origin, elucidate molecular characteristics, and monitor tumor therapeutic response due to its stability and low sensitivity to preanalytical variables.[31,32] Recent studies have indicated that cfRNA profiling can differentiate between cancerous, noncancerous, and precancerous samples.[33]

Like bulk RNA-Seq, cfRNA sequencing and isolation can be accomplished without specialized equipment. Illumina currently supports end-to-end cfRNA workflows using common, commercial RNA isolation kits and standard Illumina RNA library prep kits. The workflows are described in greater detail in the following Illumina resources:

- Improved Detection of Circulating Transcripts
- Advancement in Translational Biomarker Identification from Cell-free RNA
- Illumina Liquid Biopsy and NGS eBook

Potential applications of cfRNA-Seq in cancer research

Detect cancer

Recent studies have shown the applicability to cfRNA-Seq in distinguishing cancer and precancer conditions from noncancerous states.[33] There is the significant potential for high-impact studies to further investigate the extent to which cfRNA sequencing can be used for cancer detection across diverse cancer types.

Detect tumor tissue of origin

All cells throughout an organism have the same DNA (except for mutations) but differ significantly in expressed RNAs. Unlike cfDNAs which generally lack SNPs or other features that differ between tissues, cfRNAs exhibit high tissue specificity. There is therefore an exciting opportunity to study the capacity of cfRNAs to localize cancer.

Determine cancer subtype

Cancer is a collection of over 100 different diseases that differ substantially in their molecular characteristics, progression, treatment options, and prognosis. Even cancers that are associated with the same organ differ greatly. For instance, germ cell ovarian tumors have a 5-year survival rate of 93% compared to a rate of only 48% for epithelial ovarian cancer.[31] Research into noninvasive subtype differentiation methodologies like cfRNA sequencing has important implications for earlier, precise diagnosis and treatment of different cancer subtypes.

Monitor tumor response to treatment

Precision therapeutics are ineffective in up to 75% of patients, in the long term. Yet mechanisms related to poor long-term response are often elusive.[34] Liquid biopsy–based analytes such as cfRNA signatures have been proposed as potential biomarkers of treatment response due to documented release from cancer cells and relative ease of collection when compared to traditional biopsy methods.[31,35] This is an emerging research area with the potential to uncover mechanisms of tumor response or lack of response to therapeutic intervention.

cfRNA sequencing: Step-by-step overview

STEP 1 Blood collection

We recommend the cell-free DNA or RNA blood collection tubes (BCT) from Streck.

The RNA tubes are optimized for cfRNA, but researchers can use the DNA tubes if they already routinely use those tubes.

STEP 2 cfRNA isolation

We recommend the QIAGEN RNeasy Kit, Qiagen exoEasy Kit or the QIAamp Circulating Nucleic Acid Kit.

The minimum volume of blood/plasma we recommend for adequate cfRNA isolation is 2 ml. Below this volume, we observe a significant increase in sample failure.

STEP 3 Library preparation

For library prep, we recommend Illumina RNA Prep with Enrichment.

Table 31: Library preparation for cfRNA-Seq

Product	Illumina RNA Prep with Enrichment
Input requirements	~5 ng cfRNA (~5 ml blood plasma)
Total library prep time	~9.5 hr
Sample type	ctRNA derived from blood, other biofluids
Sample index sets	384 indexes available

STEP 4 Sequencing

Illumina recommends sequencing cfRNA on NextSeq 1000/2000, NovaSeq 6000, or NovaSeq X instruments.

Table 32: Recommended sequencing specifications for cfRNA-Seq

Sequencing system	Illumina flow cell	Reads[a] per flow cell	Reads[a] per sample	Sample number per flow cell	Recommended read length	Recommended read depth
NextSeq 1000/2000 Systems	P2	400M	150M	2	2 × 101 bp, or up to 2 × 150 bp for events like fusions 50–75 bp reads may be sufficient for larger studies or counting applications	100M–150M reads
	P3[b]	1.2B	150M	8		
	P4[b]	1.8B	150M	12		
NovaSeq 6000 System	S1	1.6B	150M	8–10		
	S2	4.1B		22–27		
	S4	10B		53–66		
NovaSeq X Series	1.5B	1.6B	150M	10		
	10B	10B		66		
	25B	26B		173		

a. Reads refers to either single end reads or read pairs.
b. P3 and P4 flow cells only available on the NextSeq 2000 System.

STEP 5 Analysis

Table 33: Data analysis software for cfRNA-Seq[36]

Analysis software	Usage
Seqtk	Downsampling
TopHat2	Mapping to genome
g.Profiler	Enrichment analysis
EdgeR or DESeq2	Differential expression analysis
All inclusive secondary analysis solution	
DRAGEN RNA pipeline	Rapid alignment, splice junction mapping and quantification, and fusion detection
Tertiary analysis	
Illumina Correlation Engine	Comparison to other tissue expression datasets

Special notes on the cfRNA-Seq workflow

1. Illumina has seen the most success across sample type and use cases using an enrichment protocol, but ribosome-depleted total RNA can also be used.
2. Sequencing depth depends on whether enrichment was used. For sequencing post enrichment, we generally recommend 100–150 million reads per sample. The flow cell you choose depends on how many samples you are running.
3. Broadly speaking, cfRNA-Seq exhibits more background noise and has more subtle changes than standard RNA-Seq. A general benchmark is to use twice the read depth for cfRNA-Seq for comparable tissue-level analyses when using an enrichment protocol.
4. Researchers should anticipate needing three times the sequencing depth for cfRNA-Seq when using ribosome-depleted samples that have not been enriched.
5. While 100M–150M reads per sample work in most cases, additional reads may be required for specific and uncommon occurrences. It is recommended to use bioinformatics to determine what number of reads are necessary for your sample type after an initial 100M–150M read pilot run. Recent studies have used up to 750M reads per sample to identify specific cancer characteristics.[30]
6. Researchers should note that read length may depend on the experiment goal. If you are just looking at RNA levels across a population, shorter reads, such as a 50 bp, may be sufficient. If you are looking for splicing events, or mutations, we recommend read lengths of 150 bp.

Method 11 T-cell receptor (TCR) sequencing

T cells express T-cell receptors (TCR) that mediate the recognition of major histocompatibility complex (MHC)-antigen complexes. This generates antigen-specific adaptive immune responses to cancer. Everyone possesses a unique, yet highly diverse TCR repertoire that allows the immune system to recognize a wide array of self and foreign antigens. TCR sequencing elucidates TCR antigen specificities. By characterizing TCR diversity using NGS methods, researchers can understand T-cell–specific immune function and responses to cancer.

Key advantages of TCR sequencing over traditional methods such as PCR and flow cytometry include:

- The ability to target all three complementary-determining regions (CDRs) for comprehensive coverage of the TCR region.
- Detecting T-cell clones at a far greater sensitivity with fewer false positives and false negatives.
- Interrogating precious samples with RNA inputs as low as 25 ng.

Potential applications of TCR sequencing in cancer research

Identify and investigate minimal residual disease (MRD)

MRD describes the small number of cancer cells that persist in a body post-treatment. The number of cells may be so small that they do not cause symptoms and may be challenging to detect by traditional methods, such as histology. Accurate measurement of MRD is important for understanding the biology of cancer progression, and relapse following treatment. Targeted TCR sequencing provides rapid, highly sensitive detection of MRD in acute lymphoblastic leukemia.[37,38,39] Researching MRD with NGS can show how cancer cells respond to treatments, evaluate remissions, and identify alternative therapies that may be counteract novel tumor characteristics.

Characterize tumor-reactive T-cell profiles

Analysis of TCR sequences can reveal the clonal content of tumor-infiltrating T-cells, which has been shown to correlate with improved clinical outcomes in various cancers.[40,41] TCR repertoire sequencing elucidates TCR diversity to help drive an understanding of the central players in immunosurveillance and immunotherapy.

TCR sequencing: Step-by-step overview

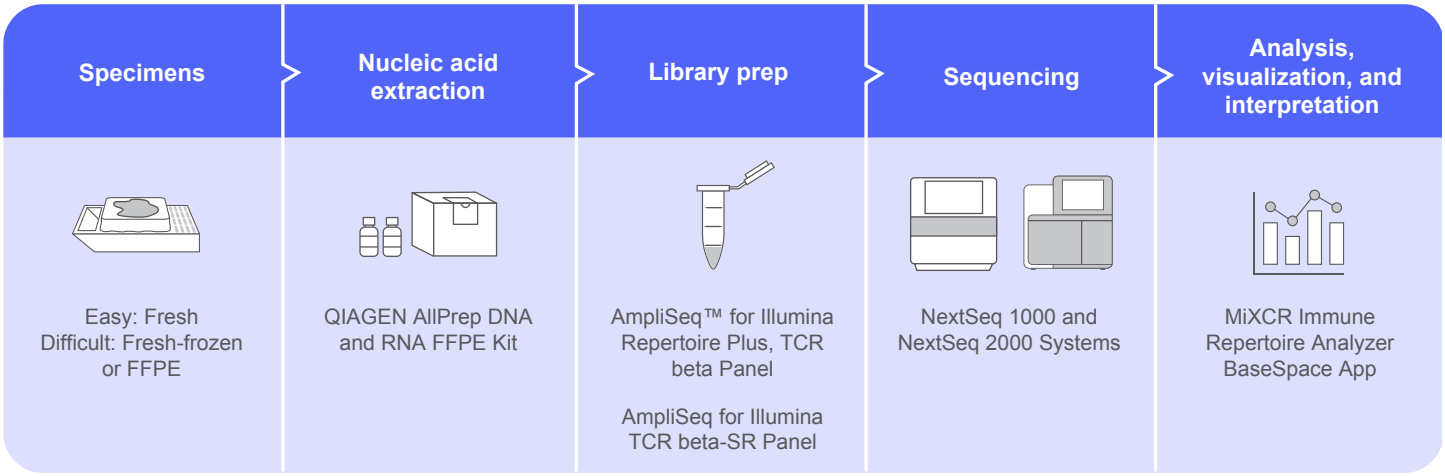


Figure 12: An overview of the TCR sequencing workflow

STEP 1 Recommended extraction methods

There are several DNA and RNA extraction methods that can be used with FFPE samples. Illumina recommends the QIAGEN AllPrep DNA/RNA FFPE Kit. In our internal studies, we found that this kit extracts high-quantity and high-quality DNA and RNA in the same workflow.

STEP 2 Recommended library prep methods

Table 34: Recommended library preparation methods for TCR sequencing

Product	AmpliSeq for Illumina Immune Repertoire Plus, TCR beta Panel	AmpliSeq for Illumina TCR beta-SR Panel
Minimum input requirements	25 ng	10 ng
Total library prep time	5–6 hours	5–6 hours
Sample type	Fresh-frozen	Fresh-frozen and FFPE compatible
Sample index sets	384 unique dual indexes	384 unique dual indexes

STEP 3 Recommended sequencing systems

Table 35: Recommended sequencing systems for TCR sequencing

Product	NextSeq 1000 and NextSeq 2000 Systems
Most important to me	Instrument affordability and desktop footprint
Recommended number of reads[a]	400M total PE reads and 30M PE reads per sample
Recommended read length	2 × 151 bp or 2 × 300 bp

a. Reads refers to single-end reads or read pairs.

STEP 4 Secondary analysis

We recommend using the MiXCR Immune Repertoire Analyzer App on BaseSpace Sequence Hub for TCR sequencing analysis.

Table 36: Recommended secondary analysis software for TCR sequencing

Software	Application	Input
MiXCR Immune Repertoire Analyzer	V(D)J segment mapping, alignment, mutation analysis	Any sequencing data type with any level of TCR coverage

Method 12 Mutational profiling for neoantigen prediction

Mutational profiling provides researchers with a comprehensive view of the mutational profile of a tumor sample. It can help to identify tumor-specific sequences and, ultimately, peptides or neoantigens that may be capable of inducing an immune response.

Key benefits of using NGS for mutational profiling of a tumor sample include:

- Unbiased screening for overall mutational load may predict patient response to immuno-oncology therapies.
- Detection of low-expressing targets with high confidence.

Potential applications of mutational profiling in cancer research

Identify responders and nonresponders of immune checkpoint blockers	Research shows that immune checkpoint blockades can induce durable responses in patients across different cancer types. Unfortunately, only a fraction of patients responds to this therapy.[1] Multiple biomarkers have been proposed to stratify responders from nonresponders, but there have been limitations with each. Since neoantigens play a direct role in inducing antitumor immunity, the accurate profiling of these molecules may be an effective way of identifying biomarkers that predict responses to immunotherapies.
Drive personalized vaccine development	Recent studies have illustrated the promise of neoantigen-based vaccines.[42] Unfortunately, while clinical trials are ongoing in several different tumor types, multiple challenges remain. The first step to developing an effective vaccine is to accurately identify putative neoantigens. Optimizing this technique is the key to developing personalized vaccines and fully harnessing the promise of this new therapeutic approach.
Identify tumor-reactive T-cells	The adoptive cell transfer of neoantigen-targeting tumor-infiltrating lymphocytes (TILs) into patients is an important emerging therapeutic for many cancer types. The key to this approach is to identify the appropriate neoantigen-reactive T-cells that can mediate a durable antitumor immune response. Identifying neoantigen targets is central to identifying the relevant T-cell population.

Mutational profiling: Step-by-step overview

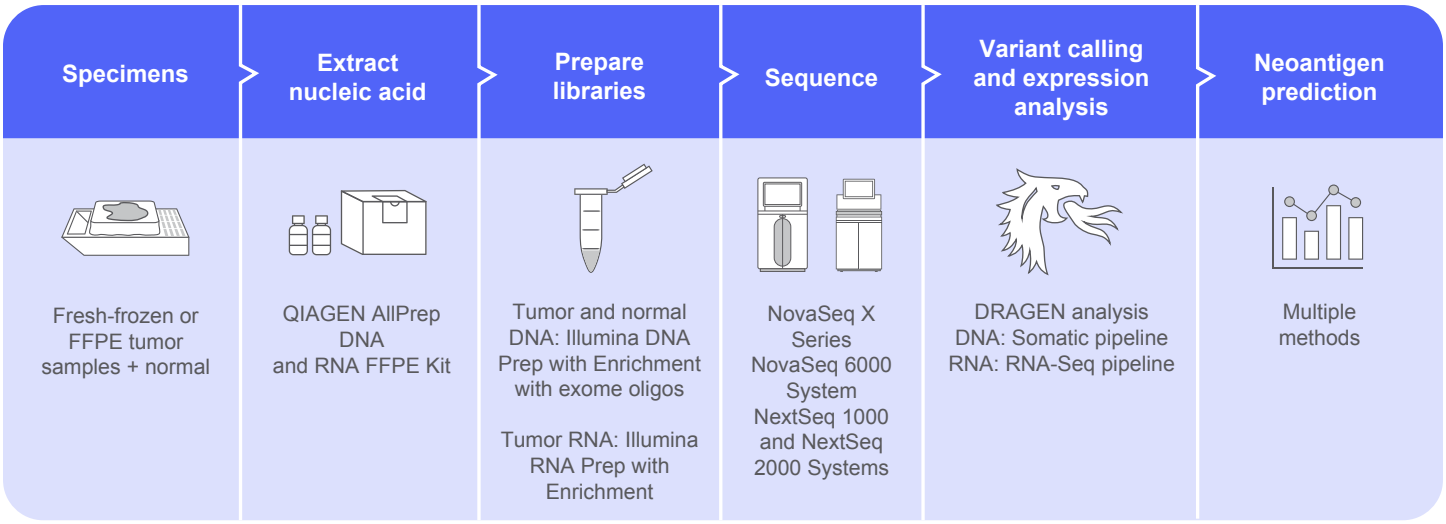


Figure 13: An overview of the mutational profiling workflow

STEP 1 Recommended extraction methods

There are several DNA and RNA extraction methods that can be used for FFPE samples. Illumina recommends the QIAGEN AllPrep DNA/RNA FFPE Kit. In our internal studies, this kit extracts high-quantity and high-quality DNA and RNA in the same workflow.

STEP 2 Recommended library prep methods

Table 37: Recommended library prep kits for mutational profiling. Turnaround time for 24 samples with 3-plex enrichment.

Product	Illumina Stranded mRNA Prep	Illumina Stranded Total RNA Prep w/ Ribo-Zero Plus	Illumina RNA Prep with Enrichment
Input requirements	25–1000 ng	1–1000 ng (10 ng for FFPE)	10–20 ng (20–100 ng for FFPE)
Total library prep time	6.5 hours	7 hours	9 hours
Sample type	High-quality RNA from any species with polyA tails	High-quality or degraded sample types, including FFPE; Ribo-Zero Plus includes rRNA depletion for human, mouse, rat, bacteria Gram +/-, and human beta globin transcripts	High-quality or degraded sample types, including FFPE; species compatibility dependent on panel used
Sample index sets	384 UDIs	384 UDIs	384 UDIs

STEP 3 Recommended sequencing systems

Recommended read lengths for mutational profiling depend on the protocol used for library preparation and the nature of the transcripts of interest. For example, 2 × 75 bp reads are compatible with stranded RNA prep protocols and 2 × 100 bp reads are best for RNA enrichment protocols. For optimal detections of events like fusions, 2 × 150 bp reads may be needed.

Table 38: Recommended sequencing systems for mutational profiling

Sequencing system	Illumina flow cell	Reads[a] per flow cell	Reads[a] per sample	No. samples per flow cell	Recommended read length	Recommended read depth
NextSeq 1000/2000 Systems	P2	400M	160M	2–3	2 × 101 bp	160M reads
	P3[b]	1.2B		~8		
	P4[b]	1.8B		~12		
NovaSeq 6000 System	S1	1.6B		10		
	S2	4.1B		~26		
	S4	10B		62		
NovaSeq X Series	1.5B	1.6B		10		
	10B	10B		62		
	25B	26B		162		

a. Reads refers to single-end reads or read pairs.
b. P3 and P4 flow cells only available on the NextSeq 2000 System.

STEP 4 Secondary analysis

Illumina recommends using the DRAGEN pipelines on BaseSpace Sequence Hub, Illumina Connected Analytics, a DRAGEN server, or onboard a sequencing system (as applicable) to obtain somatic variant calls and gene expression data. When using BaseSpace Sequence Hub, you can monitor runs in real time while securely streaming data directly from the instruments into the ecosystem for pushbutton analysis.

Table 39: Recommended analysis software for mutational profiling

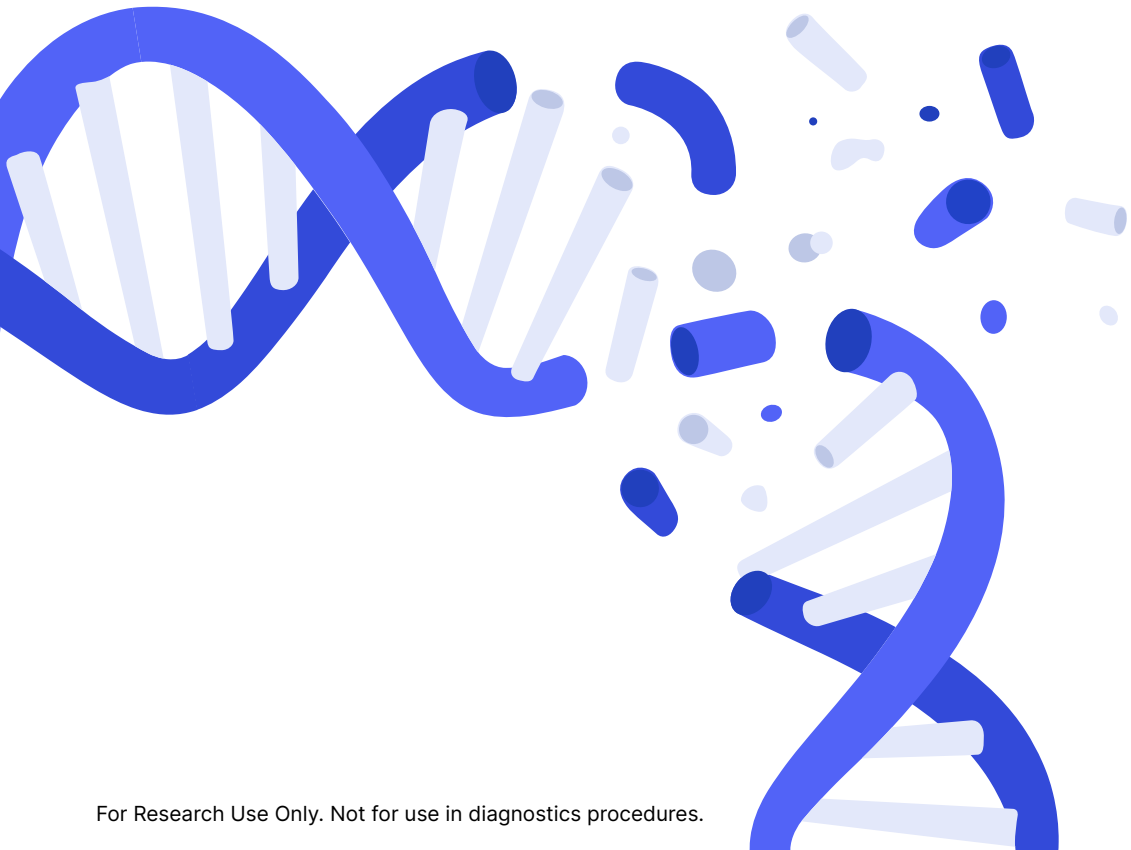
Pipeline	Application	Input
DRAGEN Enrichment pipeline	Rapid alignment of reads from targeted enrichment experiments	Tumor DNA FASTQ
DRAGEN Somatic pipeline	Somatic variant detection in tumor samples; includes tumor-only and tumor–normal modes	Tumor DNA FASTQ Normal DNA FASTQ
DRAGEN RNA pipeline	Rapid alignment, splice junction mapping and quantification, and fusion detection	Tumor RNA FASTQ
DRAGEN RNA Differential Expression pipeline	Pairwise differential gene expression analysis. Available on BaseSpace Sequence Hub only	Tumor RNA FASTQ

STEP 5 Tertiary analysis

Before neoantigen prediction, you must use the somatic variant calls and gene expression data for human leukocyte antigen (HLA) typing, peptide processing, and major histocompatibility complex (MHC)-binding prediction. There are many options for the user in each of these steps. For a review of these options and instructions for use, read Best practices for bioinformatic characterization of neoantigens for clinical utility.

Table 40: Tertiary analysis tools for mutational profiling

Analysis step	Description
HLA typing	The use of exome data to determine a patient's HLA alleles and corresponding MHC complexes
Peptide processing	A generation of small peptides using a sliding window that is applied to the mutant protein sequence
MHC-binding prediction	An analysis of peptide affinity toward the MHC complexes identified
Neoantigen prioritization	The prioritization of selected peptides based on variant frequency, binding affinity, and other factors



Proteomics methods

Method 13 Olink proximity extension assay (PEA)

Olink PEA technology powers the coupling of NGS with proteomics, enabling researchers to gain a fuller picture of biological phenomena from samples. With Olink PEA technology, each protein in a ~5400 protein panel is targeted by two antibodies, each with a complementary, unique DNA oligonucleotide. When the correct pair of antibodies binds to their target protein, the complementary DNA probes hybridize to form a double-stranded DNA "barcode" that is further exteneded and amplified using sample-specific primers.

Key benefits of using Olink PEA for proteomics analysis over traditional large-scale proteomics methods (ie, mass spectrometry) and conventional immunoassays include:

- Rapid, simultaneous detection of ~5400 proteins, including both circulating and intracellular proteins.
- Amplification of signal using DNA oligonucleotides bound to protein-specific antibodies allows for the use of low sample volumes, such as plasma and serum samples.
- High sensitivity enables detection of low-abundant proteins over a wide range of concentrations.
- Reduces nonspecific antibody binding.
- Offers multiplexing scalability without the loss of analytical specificity or sensitivity.

Potential applications of Olink PEA in cancer research

Transition cancer biomarkers from discovery into the clinic

The ability of Olink PEA to detect protein biomarkers from small sample volumes enables rapid proteomic discovery and profiling of cancer biomarkers from clinical samples. This technology can address the current bottleneck faced when transitioning novel cancer biomarkers from discovery to the clinic.

Retrospective and temporal analysis of cancer progression

Olink PEA sensitivity enables retrospective analysis of clinical samples collected at cancer diagnosis to provide insight into temporal changes in cancer biomarker abundances over the course of cancer treatment.[43]

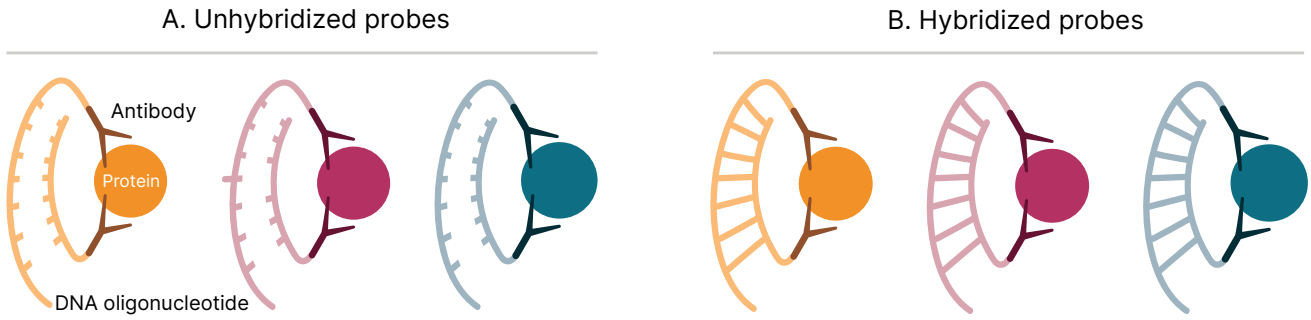


Figure 14: Olink PEA technology

Olink PEA: Step-by-step overview

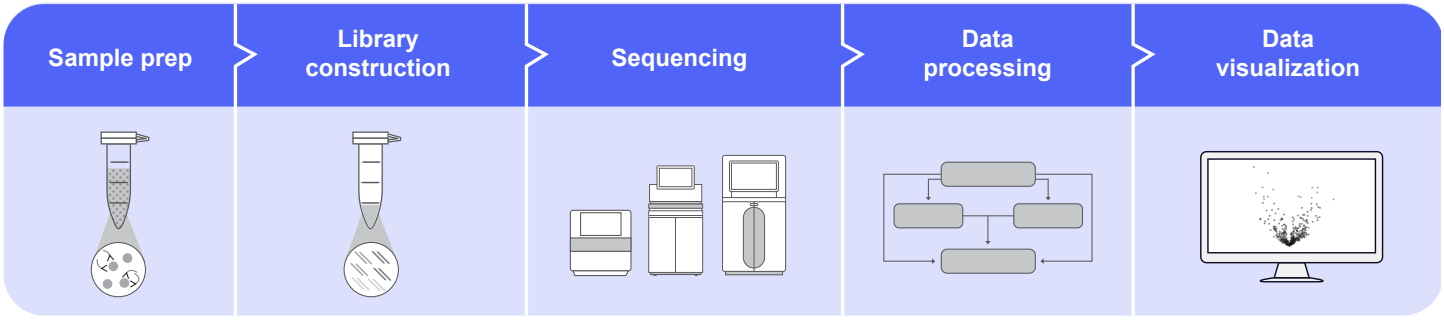


Figure 15: General overview of proteomic analysis with Olink PEA technology

STEP 1 Sample preparation and antibody staining

Samples can be isolated from a wide range of clinical sources, including serum, plasma, and cerebrospinal fluid. Sample preparation begins with the immunoreaction, during which the sample is diluted based on protein abundance and incubated with antibody probes overnight. During this incubation, the paired antibodies bind their target protein and the complementary DNA oligonucleotides hybridize. A variety of panels that detect different targets are available, depending on your area of interest within oncology.

Table 41: Antibody panels for Olink PEA analysis

Provider	Kit	Therapeutic area of study	Relevant available target 384-plex panels
Olink	Explore 3072	Oncology	• Oncology • Oncology II • Inflammation • Inflammation II

STEP 2 Library construction

After the immunoreaction, the sample is amplified via PCR and different sample dilutions are pooled. Compatible sequencing adapters and sample-specific indexes are then added to the amplicons via a secondary PCR step. Amplicons are then purified using purification beads, analyzed for quality, normalized, and pooled for sequencing analysis.

STEP 3 Sequencing

Table 42: Sequencing systems for proteomic analysis using Olink PEA technology

Sequencing system	Illumina flow cell	Reads[a] per flow cell	Reads[a] per sample	No. samples per flow cell	Cycles
NextSeq 1000/2000 Systems	P2	400M	26M	~15	100
	P3[b]	1.2B		~46	
	P4[b]	1.8B		~69	
NovaSeq 6000 System	S1	1.6B		~61	35–100
	S2	4.1B		~157	
	S4	10B		384	
NovaSeq X Series	1.5B	1.6B		61–62	35
	10B	10B		384	

a. Reads refers to single-end reads or read pairs.
b. P3 and P4 flow cells only available on the NextSeq 2000 System.

STEP 4 Data processing

Table 43: Olink Data Processing Software packages for analysis of sequencing data

Olink Data Processing Software	Analysis
bcl2count	BCL files are translated into protein counts
NPX Manager	Relative quantification, inter- and intra-assay normalizations, QC analyses
MyData (cloud-based)	Relative quantification, inter- and intra-assay normalizations, QC analyses

STEP 5 Data visualization

Table 44: Olink Data Visualization Software packages for visualization of sequencing data

Olink Data Visualization Software	Type	Capabilities
Analyze (local)	Local	Tables, figures, heat maps, volcano plots, principal component analysis (PCA), plots, box plots, and more
Insights Stat Analysis (web-based)	Web-based	

Microbiome methods

Method 14 Shotgun metagenomics

Emerging evidence supports that intratumoral bacteria have a pervasive presence within and across solid tumors, modulating immune cell recruitment into the tumor microenvironment. A seminal study investigating 1500 human tumors representing seven different tumor types found bacteria were localized within both cancer cells and immune cells, and bacterial species were varied depending on tumor type.[44]

Shotgun metagenomics is an NGS-powered approach to study microbial communities within the tumor microenvironment. Unlike traditional 16S and ITS sequencing, which target specific genomic regions, shotgun metagenomics sequences all microbial DNA within a sample. After sequencing, bioinformatics tools are used to align sequences to known genomic locations to discover both the identity and function of microbes within a community simultaneously.

Key benefits of shotgun metagenomics over 16S/ITS sequencing and WGS include:

- The depth of sequencing achieved allows for species-level designation of microbes within a community, as opposed to genus-level designation achieved by targeted methods.
- Unlike WGS, shotgun metagenomics does not require cultured organisms prior to sequencing.
- Provides comprehensive analysis of the microbial sequences within a diverse community to allow for the *de novo* discovery and characterization of new mutations, antimicrobial resistance markers, and even novel microbial species.

Potential applications of shotgun metagenomics in cancer research

Understanding antitumor immunity within the tumor microenvironment (TME)	The microbiota of the gastrointestinal tract produces a variety of small molecules and metabolites that affect many physiological processes, including inflammation and immunity. Further, the gut microbiota has been shown to reprogram the immunity of the TME by engaging innate and adaptive immune cells.[45]
Predicting responses to immunotherapy	Recent studies have established the gut microbiome as a key regulator of the efficacy of cancer immunotherapies, especially immune checkpoint inhibitors (ICIs), through the assessment of therapeutic responses in patients who have undergone antibiotic treatment (reducing microbiome diversity) or fecal-microbiota transplantation (altering the microbiome).[45]
Identification of microbial drivers of cancer	Infection with oncogenic viruses, which insert their viral genome into host DNA near cellular oncogenes, has long been known to drive cancer formation by disrupting the regulation of the cell cycle. Moreover, bacterial infections have been linked to cancer by inducing chronic information. Shotgun metagenomics has enabled insights into novel links between viruses, bacteria, and cancer.[46]

Shotgun metagenomics: Step-by-step overview

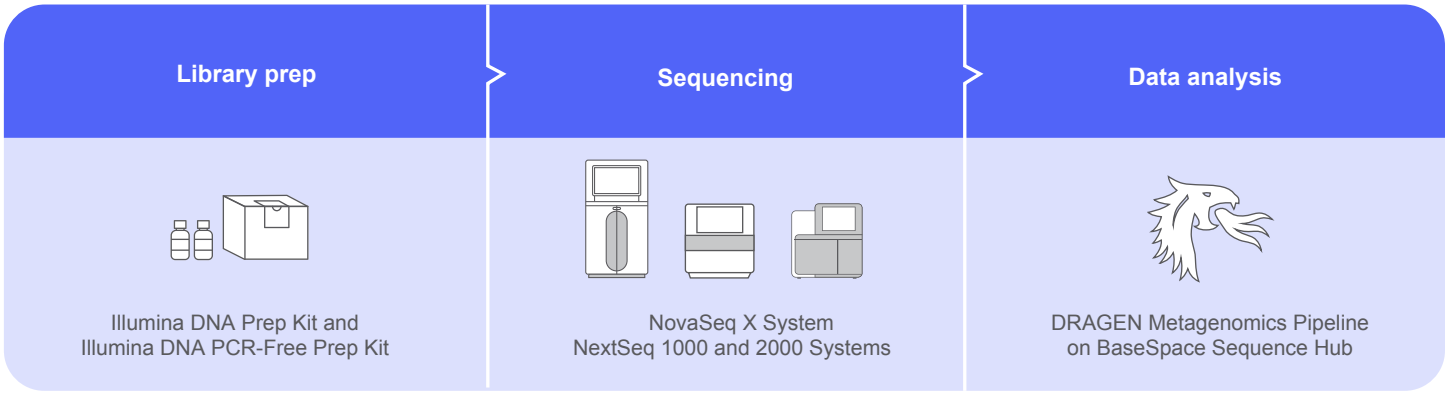


Figure 16: General overview of shotgun metagenomics workflow

STEP 1 Library preparation

Microbial samples can be prepared using the Illumina DNA Prep kit or the Illumina DNA PCR-Free Prep kit.

Table 45: Library preparation for shotgun metagenomics

Product	Illumina DNA Prep	Illumina DNA PCR-Free Prep
Input requirements	1–500 ng	25–300 ng
Total library prep time	1–1.5 hours	1.5 hours
Sample type	Microbial DNA	Microbial DNA from blood, low-input samples, saliva
Sample index sets	384 unique dual indexing and 96 combinatorial dual indexing	384 unique dual indexing

STEP 2 Sequencing

Table 46: Sequencing systems for shotgun metagenomics

Sequencing system	Illumina flow cell	Reads[a] per flow cell	Read[a] size	No. samples per flow cell		
				Taxonomic profiling	Strain-level detection	AMR gene family profiling
NextSeq 1000/2000	P1	100M	150–300 bp	30	2	1
	P2	400M		120	8	5
	P3[b]	1.2B		360	24	15
	P4[b]	1.8B		540	36	22

a. Reads refers to single-end reads or read pairs.
b. P3 and P4 flow cells only available on the NextSeq 2000 System.

STEP 3 Data analysis

The DRAGEN Metagenomics pipeline performs taxonomic classification of reads and provides both single sample and aggregate reporting.

Method 15 Metatranscriptomics

Metatranscriptomics leverages NGS to profile the complete set of microbial RNA transcripts isolated from a sample to provide insights into active, functional pathways that may modulate the tumor microenvironment.

Key benefits of metatranscriptomic analysis compared to shotgun metagenomics:

- Delves deeper than identifying the microbial composition alone, revealing active pathways of a microbial sample to uncover new functions within the intratumoral microbiome.
- Detects cancer-associated RNA viruses that may be missed by shotgun metagenomics.

Potential applications of metatranscriptomics in cancer research

Revealing functional associations between the intratumoral microbiome and tumorigenesis

Metatranscriptomics analysis of microbial samples isolated from tumors can identify novel correlations between the expression of certain microbial genes with enhanced cancer progression and metastasis.[47]

Identification of functional intratumoral microbial pathways as novel cancer biomarkers

Analysis of metatranscriptomic microbial signatures, such as activated proinflammatory and oxidative pathways, can be used to distinguish high-risk patients from healthy controls for certain cancers.[48]

Metatranscriptomics: Step-by-step overview

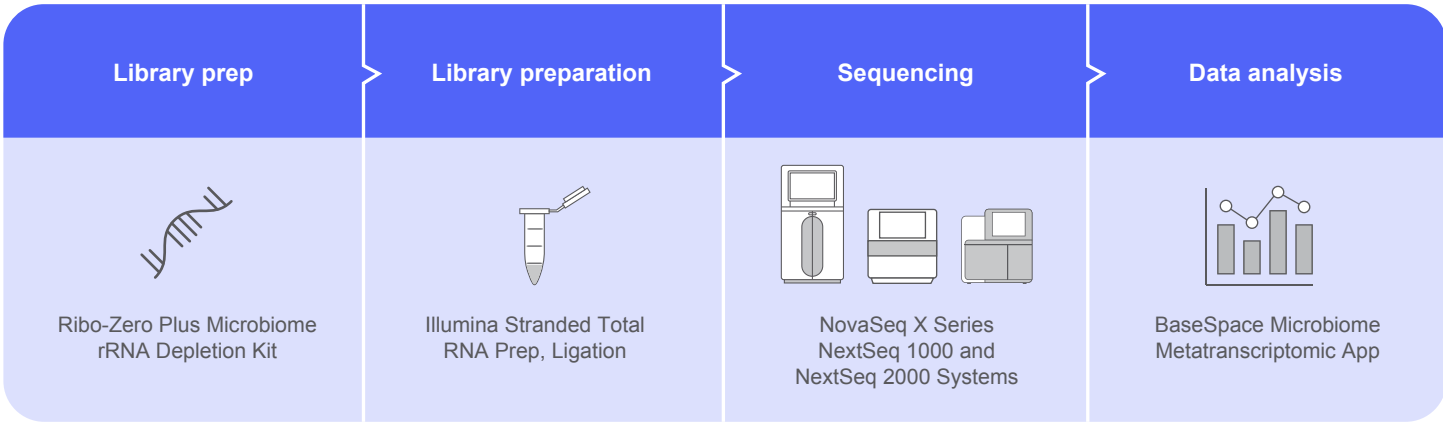


Figure 17: General overview of a metatranscriptomics workflow

STEP 1 rRNA depletion

When total RNA is extracted from a microbiome sample, the vast majority is ribosomal RNA (rRNA). Depletion of these unwanted rRNAs is critical to reveal metatranscriptomic changes in complex microbial samples. The Ribo-Zero Plus Microbiome rRNA Depletion Kit enables fast and efficient depletion of undesirable host and pan-bacterial rRNA from complex microbial samples for metatranscriptomics research.

STEP 2 Library preparation

Microbial samples can be prepared using the Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus Microbiome kit

Table 47: Library preparation for metatranscriptomic analysis

Product	Illumina Stranded Total RNA Prep, Ligation with Ribo-Zero Plus Microbiome
Input requirements	1–1000 ng total RNA
Total library prep time	7 hours
Sample type	Coding RNA plus multiple forms of noncoding RNA from standard-quality RNA samples
Sample index sets	UP to 384 unique dual indexes

STEP 3 Sequencing

Table 48: Sequencing systems for metatranscriptomic analysis

Sequencing system	Illumina flow cell	Reads[a] per flow cell	Reads[a] per sample	No. samples per flow cell
NextSeq 1000/2000 System	P1	100M	3M	33
	P2	400M	3M	133
	P3[b]	1.2B	3M	400
	P4[b]	1.8B	3M	600

a. Reads refers to single-end reads or read pairs.
b. P3 and P4 flow cells only available on the NextSeq 2000 System.

STEP 4 Data analysis

The easy-to-use BaseSpace Microbiome Metatranscriptomics App simplifies data analysis and accurate visualization of results.

References

1. Dawkins JB, Wang J, Maniati E, et al. Reduced Expression of Histone Methyltransferases KMT2C and KMT2D Correlates with Improved Outcome in Pancreatic Ductal Adenocarcinoma. *Cancer Res*. 2016;76(16):4861-4871. doi:10.1158/0008-5472.CAN-16-0481

2. Method of the Year 2019: Single-cell multimodal omics. *Nat. Methods*. 2020;17(1):1. doi:10.1038/s41592-019-0703-5

3. O'Flanagan CH, Campbell KR, Zhang AW, et al. Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *Genome Biol*. 2019;20(1):210. doi:10.1186/s13059-019-1830-0

4. Ji AL, Rubin AJ, Thrane K, et al. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell*. 2020;182(6):1661-1662. doi:10.1016/j.cell.2020.08.043

5. National Cancer Institute. Human Microbiome and Cancer. https://dceg.cancer.gov/research/how-we-study/microbiomics. Accessed May 17, 2024.

6. Das PM, Singal R. DNA methylation and cancer. *J Clin Oncol*. 2004;22(22):4632-4642. doi:10.1200/JCO.2004.07.151

7. Osmanbeyoglu HU, Shimizu F, Rynne-Vidal A, et al. Chromatin-informed inference of transcriptional programs in gynecologic and basal breast cancers. *Nat Commun*. 2019;10(1):4369. doi:10.1038/s41467-019-12291-6

8. Wei SC, Duffy CR, Allison JP. Fundamental Mechanisms of Immune Checkpoint Blockade Therapy. *Cancer Discov*. 2018;8(9):1069-1086. doi:10.1158/2159-8290.CD-18-0367

9. Zhao J, Chen AX, Gartrell RD, et al. Immune and genomic correlates of response to anti-PD-1 immunotherapy in glioblastoma. *Nat Med*. 2019;25(3):462-469. doi:10.1038/s41591-019-0349-y

10. GEN, Genetic Engineering & Biotechnology News. Revealing the Tumor Microenevironment's Vulnerabilities. genengnews.com/sponsored/revealing-the-tumor-microenvironments-vulnerabilities/. Published April 4, 2023. Accessed May 1, 2023.

11. Rozenblatt-Rosen O, Regev A, Oberdoeffer, P, et al. The Human Tumor Atlas Network: Charing Tumor Transitions across Space and Time at Single-Cell Resolution. *Cell*. 2020;181(2):236-249. doi:10.1016/j.cell.2020.03.053

12. Poropatich K, Dominguez D, Chan WC, et al. OX40+ plasmacytoid dendritic cells in the tumor microenvironment promote antitumor immunity. *J Clin Invest*. 2020;130(7):3528-3542. doi:10.1172/JCI131992

13. Ediriwickrema A, Aleshin A, Reiter JG, et al. Single-cell mutational profiling enhances the clinical evaluation of AML MRD. *Blood Adv*. 2020;4(5):943-952. doi:10.1182/bloodadvances.2019001181

14. McMahon CM, Ferng T, Canaani J, et al. Clonal Selection with RAS Pathway Activation Mediates Secondary Clinical Resistance to Selective FLT3 Inhibition in Acute Myeloid Leukemia. *Cancer Discov*. 2019;9(8):1050-1063. doi:10.1158/2159-8290.CD-18-1453

15. Noble F, Mellows T, McCormick Matthews LH, et al. Tumor infiltrating lymphocytes correlate with improved survival in patients with oesophageal adenocarcinoma. *Cancer Immunol Immunother*. 2016;65:651-662. doi:10.1007/s00262-016-1826-5

16. Bailey MH, Meyerson WU, Dursi LJ, et al. Retrospective evaluation of whole exome and genome mutation calls in 746 cancer samples. *Nat Commun*. 2020;11(1):4748. doi:10.1038/s41467-020-18151-y

17. Rusch M, Nakitandwe J, Shurtleff S, et al. Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome. *Nat Commun*. 2018;9(1):3962. doi:10.1038/s41467-018-06485-7

18. Végran F, Berger H, Boidot R, et al. The transcription factor IRF1 dictates the IL-21-dependent anticancer functions of TH9 cells. *Nat Immunol*. 2014;15(8):758-766. doi:10.1038/ni.2925

19. San Lucas FA, Allenson K, Bernard V, et al. Minimally invasive genomic and transcriptomic profiling of visceral cancers by next-generation sequencing of circulating exosomes. *Ann Oncol*. 2016;27(4):635-641. doi:10.1093/annonc/mdv604

20. Réda M, Richard C, Bertaut A, et al. Implementation and use of whole exome sequencing for metastatic solid cancer. *EBioMedicine*. 2020;51:102624. doi:10.1016/j.ebiom.2019.102624

21. Crosby, D. Delivering on the promise of early detection with liquid biopsies. *Br J Cancer*. 2022;126(3):313-315. doi:10.1038/s41416-021-01646-w

22. Van der Auwera I, Yu W, Suo L, et al. Array-based DNA methylation profiling for breast cancer subtype discrimination. *PLoS One*. 2010;5(9):e12616. doi:10.1371/journal.pone.0012616

23. Lehmann BD, Colaprico A, Silva TC, et al. Multi-omics analysis identifies therapeutic vulnerabilities in triple-negative breast cancer subtypes. *Nat Commun*. 2021;12(1):6276. doi:10.1038/s41467-021-26502-6

24. Lyskjaer I, De Noon S, Tirabosco R, et al. DNA methylation-based profiling of bone and soft tissue tumours: a validation study of the 'DKFZ Sarcoma Classifier'. *J Pathol Clin Res*. 2021;7(4):350-360. doi:10.1002/cjp2.215

25. Pedersen CA, Cao MD, Fleischer T, et al. DNA methylation changes in response to neoadjuvant chemotherapy are associated with breast cancer survival. *Breast Cancer Res*. 2022; 24, 43. doi:10.1186/s13058-022-01537-9

26. Pon JR, Marra MA. Driver and passenger mutations in cancer. *Annu Rev Pathol*. 2015;10:25-50. doi:10.1146/annurev-pathol-012414-040312

27. Chen K, Zhang Y, Qian L, Wang P. Emerging strategies to target RAS signaling in human cancer therapy. *J Hematol Oncol*. 2021;14(1):116. doi:10.1186/s13045-021-01127-w

28. Bode AM, Dong Z. Recent advances in precision oncology research. *NPJ Precis Oncol*. 2018;2:11. doi:10.1038/s41698-018-0055-0

29. Nassiri F, Liu J, Patil V, et al. A clinically applicable integrative molecular classification of meningiomas. *Nature*. 2021;597(7874):119-125. doi:10.1038/s41586-021-03850-3

30. Newell F, Pires da Silva I, Johansson PA, et al. Multiomic profiling of checkpoint inhibitor-treated melanoma: Identifying predictors of response and resistance, and markers of biological discordance. *Cancer Cell*. 2022;40(1):88-102.e7. doi:10.1016/j.ccell.2021.11.012

31. Larson MH, Pan W, Kim HJ, et al. A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. *Nat Commun*. 2021;12(1):2357. doi:10.1038/s41467-021-22444-1

32. Skog J, Würdinger T, van Rijn S, et al. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol*. 2008;10(12):1470-1476. doi:10.1038/ncb1800

33. Roskams-Hieter B, Kim HJ, Anur P, et al. Plasma cell-free RNA profiling distinguishes cancers from pre-malignant conditions in solid and hematologic malignancies. *NPJ Precis Oncol*. 2022;6(1):28. doi:10.1038/s41698-022-00270-y

34. Personalized Medicine Coalition. The personalized medicine report. Opportunity, challenges, and the future. personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/The_PM_Report.pdf. Published 2017. Accessed May 17, 2024.

35. Kilgour E, Rothwell DG, Brady G, Dive C. Liquid Biopsy-Based Biomarkers of Treatment Response and Resistance. *Cancer Cell*. 2020;37(4):485-495. doi:10.1016/j.ccell.2020.03.012

36. Munchel S, Rohrback S, Randise-Hinchliff C, et al. Circulating transcripts in maternal blood reflect a molecular signature of early-onset preeclampsia. *Sci Transl Med*. 2020;12(550):eaaz0131. doi:10.1126/scitranslmed.aaz0131

37. Kotrova M, Trka J, Kneba M, Bruggemann M. Is next-generation sequencing the way to go for residual disease monitoring in acute lymphoblastic leukemia? *Mol Diagn Ther*. 2017;21(5):481-492. doi:10.1007/s40291-017-0277-9

38. Wu J, Jia S, Wang C, et al. Minimal Residual Disease Detection and Evolved IGH Clones Analysis in Acute B Lymphoblastic Leukemia Using IGH Deep Sequencing. *Front Immunol*. 2016;7:403. doi:10.3389/fimmu.2016.00403

39. Wright G, Watt E, Inglott S, et al. Clinical benefit of a high-throughput sequencing approach for minimal residual disease in acute lymphoblastic leukemia. *Pediatr Blood Cancer*. 2019;66(8): e27787. doi:10.1002/pbc.27787

40. Valpione S, Mundra PA, Galvani E, et al. The T cell receptor repertoire of tumor infiltrating T cells is predictive and prognostic for cancer survival. *Nat Commun*. 2021;12(1):4098. doi:10.1038/s41467-021-24343-x

41. Zhigalova EA, Izosimova AI, Yuzhakova DV, et al. RNA-Seq-Based TCR Profiling Reveals Persistently Increased Intratumoral Clonality in Responders to Anti-PD-1Therapy. *Front Oncol*. 2020;10:385. doi:10.3389/fonc.2020.00385

42. Blass E, Ott PA. Advances in the development of personalized neoantigen-based therapeutic cancer vaccines. *Nat Rev Clin Oncol*. 2021;18(4):215-229. doi:10.1038/s41571-020-00460-2

43. Ghorbani A, Avery LM, Sohaei D, et al. Discovery of novel glioma serum biomarkers by proximity extension assay. *Clin Proteomics*. 2023;20(1):12. doi:10.1186/s12014-023-09400-5

44. Nejman D, Livyatan I, Fuks G, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science*. 2020;368(6494):973-980. doi:10.1126/science.aay9189

45. Lu Y, Yuan X, Wang M, et al. Gut microbiota influence immunotherapy responses: mechanisms and therapeutic strategies. *J Hematol Oncol*. 2022;15(1):47. doi:10.1186/s13045-022-01273-9

46. Borchmann S. An atlas of the tissue and blood metagenome in cancer reveals novel links between bacteria, viruses and cancer. *Microbiome*. 2021;9(1):94. doi:10.1186/s40168-021-01039-4

47. Feng Y, Ramnarine VR, Bell R, et al. Metagenomic and metatranscriptomic analysis of human prostate microbiota from patients with prostate cancer. *BMC Genomics*. 2019;20(1):146. doi:10.1186/s12864-019-5457-z

48. Banavar G, Ogundijo O, Toma R, et al. The salivary metatranscriptome as an accurate diagnostic indicator of oral cancer. *NPJ Genom Med*. 2021;6(1):105. doi:10.1038/s41525-021-00257-x

The time to get started is now

Here's why you should partner with Illumina to study your cancer research question.

Our NGS reagents, equipment, and protocols have been validated in hundreds of high-impact peer-reviewed journals for cancer research.
https://pubmed.ncbi.nlm.nih.gov/?term=illumina+NGS+AND+cancer&filter=simsearch2.ffrft

Illumina has the broadest range of applications to enable NGS analyses on one device and is built to support emerging changes in the field.

Illumina supports over 10,000 labs across 115 countries. Our 24/7 technical support professionals are ready to help you with any project.

Learn more about how to combine these methods

Multiomics allows you to explore your cancer research question even further. Learn more on our website.
www.illumina.com/morewithmultiomics2

illumina

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

M-AMR-00521 v3.0