NEW YORK GENOME CENTER® | Collaboration Research Discovery

Whole Genome Sequencing

NYGC offers comprehensive short-read whole genome sequencing (WGS) services for both **germline** and **somatic** studies, including sample quality control, library preparation, sequencing, bioinformatics analysis, and secure cloud-based data delivery. Our highly automated, GCLP-compliant laboratory workflows and validated analysis pipelines generate high-quality, reproducible results that are compatible with major genomic research initiatives.

Our standard germline WGS service delivers 30X coverage (or any desired coverage) using PCR-free library preparation and the Illumina NovaSeq X+ platform. PCR-based workflows are also available for low-input or challenging samples, including FFPE-derived DNA. Our somatic WGS service compares matched tumor-normal pairs at any desired coverage, with comprehensive analysis of SNVs, indels, copy number variants, structural variants, microsatellite instability, HLA type, mutational signatures, and ancestry. We also offer optimized workflows for **cell-free DNA (cfDNA)** and can provide custom downstream analyses to support a wide range of cancer research applications. Paired Plus-Minus Sequencing (**ppmSeq**) enhances WGS accuracy by leveraging information from both strands of each DNA molecule to distinguish true low-frequency variants from sequencing and sample-preparation artifacts, making it particularly valuable for applications such as circulating tumor DNA (ctDNA), minimal residual disease (MRD), and other rare variant detection studies.

DNA Submission Requirements - Short Read WGS

The following are the standard minimum DNA input requirements intended to maximize sequencing performance and data quality. If your samples fall outside these specifications, please contact us before submission. In many cases, we can recommend alternative workflows or evaluate samples on a project-specific basis.

Germline (PCR-free library preparation)

- Minimum of **1 µg** of unamplified, RNase-treated, high molecular weight DNA
- Submitted in **50–100 µL** of EB (Elution Buffer)
- DNA concentration determined by **PicoGreen** (or equivalent fluorescence-based assay)

Somatic (tumor-normal or tumor-only, including FFPE)

- Minimum of **500 ng** of unamplified, RNase-treated, high molecular weight DNA
- Submitted in **60–75 µL** of EB (Elution Buffer)
- DNA concentration determined by **PicoGreen** (or equivalent fluorescence-based assay)

Cell-free DNA (cfDNA)

- Minimum of **50 ng** of unamplified cfDNA
- Submitted in **30–60 µL** of EB (Elution Buffer)
- DNA concentration determined by **PicoGreen** (or equivalent fluorescence-based assay)

ppmSeq for Ultima Genomics

- Minimum of **500 ng** of unamplified gDNA or **50 ng** cfDNA
- Submitted in **30–60 µL** of EB (Elution Buffer)
- DNA concentration determined by **PicoGreen** (or equivalent fluorescence-based assay)

Standard Germline Analysis & Deliverables - Illumina Short Read WGS

All whole genome sequencing (WGS) data are processed using NYGC's validated, production-grade analysis pipeline, which follows the Functional Equivalence framework developed to enable harmonization with major genomic initiatives, including CCDG and TOPMed. Our standardized workflow includes sequence alignment, quality assessment, duplicate marking, variant calling, filtering, and comprehensive variant annotation using industry-standard tools and best practices. This robust, reproducible pipeline ensures high-quality, consistent results that are compatible with major public genomic repositories. Full details can be found here.

Analysis includes:

- Alignment of sequencing reads to the human reference genome (GRCh38)
- Duplicate read marking and quality recalibration
- Germline SNV and indel calling
- Structural variant detection
- Comprehensive variant annotation, including:
    - Variant effect predictions
    - Population allele frequencies
    - dbSNP identifiers
    - Conservation scores
    - Pathogenicity predictions
    - ClinVar annotations

Deliverables:

- Sequencing data (CRAM)
- Germline variant calls (VCF)
- Annotated variant report (tab-delimited)
- Structural variant calls (VCF)
- Mean coverage meeting the requested service level

Standard Somatic Analysis & Deliverables - Illumina Short Read WGS

Our somatic analysis pipeline is designed to accurately identify genomic alterations by comparing tumor and matched normal samples. Using industry-standard best practices and multiple complementary variant callers, the pipeline detects single nucleotide variants (SNVs), insertions and deletions (indels), copy number variants (CNVs), and structural variants (SVs). Results are

integrated to generate high-confidence variant calls, with additional analyses including microsatellite instability (MSI), HLA typing, mutational signatures, and ancestry inference. Developed and maintained by NYGC's experienced Computational Biology team, the pipeline has been applied across a wide range of cancer types and produces high-quality, reproducible results suitable for both discovery research and translational studies. Full details can be found here.

Analysis includes:

- Alignment to the human reference genome (GRCh38)
- Duplicate marking and quality recalibration
- Somatic SNV and indel calling using multiple callers
- Copy number variant (CNV) detection
- Structural variant (SV) detection

Deliverables:

- Raw somatic variant calls (VCF)
- Combined ("AllSomatic") SNV/indel files (VCF, MAF, and tab-delimited)
- High-confidence variant calls (VCF)
- CNV and SV results
- Structural variant files (BEDPE)
- HTML summary report with sequencing and variant statistics
- Mean coverage meeting the requested service level

Standard Patient-Derived Xenograph (PDX) Analysis & Deliverables - Illumina Short Read WGS

Our PDX workflows are designed to deliver high-quality genomic data from complex tumor models, enabling robust characterization of both the human tumor genome and model-specific genomic features. Advanced bioinformatic approaches help distinguish human-derived tumor signals from host background, providing accurate and actionable genomic insights.

Analysis includes:

- Separation of mouse and human sequencing reads
- Alignment of human reads
- Duplicate marking and quality recalibration
- Somatic SNV and indel calling
- Germline variant filtering
- CNV and structural variant detection
- Comprehensive annotation of somatic variants and structural variants

Deliverables:

- FASTQ files
- BAM files
- Mouse read percentage for each sample

- Raw somatic variant calls (VCF)
- Combined variant files (VCF, MAF, and tab-delimited)
- CNV and SV results
- Processed structural variant files (BEDPE)
- PDF summary report
- Expected sequencing depth

Standard cfDNA Analysis & Deliverables - Illumina Short Read WGS

Our cfDNA WGS workflow enables comprehensive profiling of genomic alterations, including single nucleotide variants, small insertions and deletions, copy number changes, and structural variants. By providing a genome-wide view of tumor-associated changes, cfDNA WGS supports research development efforts focused on biomarker discovery, disease monitoring, and precision oncology applications.

Analysis includes:

- Adapter trimming
- Alignment to the human reference genome (GRCh38)
- Duplicate marking and quality recalibration
- Tumor fraction estimation
- Ploidy assessment
- Copy number analysis using ichorCNA

Deliverables:

- FASTQ files
- BAM file
- ichorCNA analysis results
- Mean coverage meeting the requested service level

Standard ppmSeq Analysis & Deliverables - Ultima Genomics Short Read WGS

NYGC processes Ultima ppmSeq data using Ultima's validated secondary analysis workflow to generate variant-calling-ready CRAM files and ppmSeq-specific quality control metrics. Optional downstream analysis can be performed using our validated germline, somatic, and rare variant bioinformatics pipelines, enabling highly accurate detection of low-frequency variants for applications including cfDNA, minimal residual disease (MRD), and cancer genomics research.

Analysis includes:

- Adapter trimming
- Alignment to the human reference genome (GRCh38)
- Tagging of ppmSeq-specific reads
- QC of ppmSeq-specific metrics

Deliverables:

- CRAM files

Frequently Asked Questions – Short Read Whole-Genome Sequencing (WGS)

What information is needed to start a project?

To begin planning a WGS project, we typically request:

- Number and type of samples
- Sample source and preservation method
- Desired sequencing coverage
- Required timeline
- Data delivery requirements
- Any planned downstream analyses

How are projects initiated?

Once project scope and service requirements have been finalized with your NYGC Project Manager, NYGC will provide a formal service quote outlining the agreed-upon deliverables, pricing, and project details. To begin project execution and submit samples, NYGC requires a signed copy of the quote and an approved purchase order (PO). For large-scale projects, additional discussions regarding customized terms and conditions may be accommodated.

What is the expected turnaround time?

All turnaround times can be found on our '[main research services page]'.

What is a standard submission size?

Our standard Illumina short-read WGS projects typically begin with a minimum of 100 samples. For larger studies, including projects with 500+ samples, we work with PIs to develop a sequencing and delivery plan using rolling batches to optimize turnaround time and data availability. Volume discounts can apply.

For smaller projects, please contact us at ProjectManagement@nygenome.org to discuss project-specific options.

How should DNA be submitted?

For U.S.-based submissions, once a project has been quoted and a purchase order has been received, we provide NYGC-barcoded tubes in a matrix rack, an electronic sample manifest, and an overnight return shipping label for sample submission. Note that the provided tube orientation is critical, and tubes should not be rearranged.

Your Project Manager will coordinate international submissions, including shipping requirements and documentation.

What DNA input requirements do you recommend?

We recommend submitting high-quality genomic DNA that meets our standard quantity and quality requirements. Specific requirements may vary depending on sample type and requested sequencing depth. Our team will review your sample information during project planning to confirm suitability before submission.

Will you check DNA quality and quantity before library preparation?

Yes. All projects include an initial quality control (QC) assessment to confirm DNA quantity and evaluate DNA quality. QC results, including DNA integrity metrics, will be shared with you for review before proceeding with library preparation.

Do you have experience working with DNA from saliva or buccal swabs?

Yes. We have experience working with DNA from saliva and buccal samples. These sample types may contain microbial contamination, which can reduce the proportion of sequencing reads that map to the human genome and may impact final human genome coverage.

Samples will be sequenced to achieve the requested target coverage based on the expected performance of high-quality genomic DNA; however, final human genome coverage cannot be guaranteed for samples with substantial contamination.

If microbial contamination is a concern, we offer optional 16S analysis, for an additional charge, during initial QC to assess bacterial content. We may also investigate unexpected alignment results if QC metrics indicate possible contamination from non-human sources.

Do you offer whole exome sequencing (WES)?

NYGC does not currently offer whole exome sequencing (WES) or targeted sequencing assays. As sequencing technologies have advanced, whole genome sequencing (WGS) has become increasingly cost-effective while providing a more comprehensive view of genomic variation, including information beyond protein-coding regions that may be missed by exome-based approaches. Our team can work with you to determine whether whole genome sequencing or another available sequencing strategy is the best fit for their research objectives.

How do you handle low-quality DNA or FFPE samples?

For FFPE-derived samples, DNA undergoes a repair step prior to library preparation to improve sequencing performance. If a sample produces a library that does not meet sequencing quality standards, it may be removed from sequencing, and library preparation costs will still apply.

High-coverage sequencing projects using FFPE samples should be discussed during project planning, as degraded DNA can result in increased duplication rates and may limit the ability to achieve very high sequencing depths.

What sequencing coverage options are available?

We offer a range of sequencing depths depending on study goals. Our team can help determine the appropriate coverage based on your research objectives, sample type, and downstream analysis needs.

What quality metrics will be provided?

We provide sequencing and QC metrics to help evaluate data quality, including metrics related to sequencing performance, alignment, coverage, and sample quality.

What species are supported?

NYGC's standard sequencing and bioinformatics workflows support human, mouse, and patient-derived xenograft (PDX) samples. Our standard library preparation protocols have been validated primarily for mammalian samples. Projects involving other species may also be supported. Our scientific team will review your project requirements and, where appropriate, develop customized sequencing and bioinformatics workflows to meet your research objectives.

How is data returned?

We support data delivery to a variety of client-managed endpoints and storage environments. Before delivery, we can perform a test transfer to confirm connectivity and ensure efficient transfer of large sequencing datasets.

What data formats are provided?

Data delivery formats can be customized based on project requirements. Common deliverables include standard alignment files (CRAM) and associated quality metrics. Additional analysis outputs can be discussed during project planning.

Do you provide secondary analysis?

Yes. Our standard sequencing services include secondary analysis options, as described in the Analysis and Deliverables sections above. These services provide processed data and analytical outputs designed to support downstream research needs. For projects requiring additional customization, our bioinformatics team offers consulting services to address specific research questions, develop tailored analytical approaches, and generate publication-ready visualizations and reports.

For more collaborative engagements involving deeper scientific partnership, custom analyses, or method development, collaboration agreements may be established. Please contact our Project Management team at ProjectManagement@nygenome.org to discuss your project needs.

Can samples be submitted anonymously/de-identified?

Yes. **Samples submitted for research sequencing services must be de-identified prior to submission.** Sample identifiers should not contain personal health information (PHI) or other identifying information. As part of our sample intake process, submitted sample information is reviewed for compliance with these requirements. If identifying information is detected, the

sample record will be flagged and the submitting party will be notified so that the information can be corrected. Any records containing unintended identifying information will be removed and handled according to established data management procedures.

For clinical sequencing services, please contact us at https://www.nygenome.org/clinical-genetics/.

How are samples tracked throughout the workflow?

NYGC uses a comprehensive laboratory information management system (LIMS) and integrated informatics infrastructure to provide secure, end-to-end tracking of every sample throughout the sequencing process. From initial sample receipt through library preparation, sequencing, quality control, and data delivery, each sample is tracked with complete process traceability. Our production laboratory has used the LIMS platform since 2015, with integrations across laboratory workflows, instrumentation, robotics, sequencing platforms, and reporting systems. This enables automated tracking of sample status, workflow progression, and quality metrics while supporting thousands of samples across hundreds of active projects. Sequencing operations are supported by additional monitoring and reporting tools that provide real-time visibility into sequencing performance, data generation, quality control metrics, and workflow status. Automated pipelines track key processing steps, including demultiplexing, alignment, quality assessment, and issue resolution, ensuring that potential challenges are identified and addressed promptly. Before data delivery, NYGC's Project Management and Computational Biology teams review sequencing performance, coverage metrics, and project-specific quality requirements to confirm that deliverables meet established standards. This integrated approach provides clients with confidence that their samples are handled securely, tracked throughout the workflow, and delivered with reliable, high-quality results.