GENOMICS SERVICES
NEW YORK GENOME CENTER

NYGC is an independent non-profit implementing advanced genomic research to improve diagnosis and treatment of serious diseases.

Through our collaborative approach, findings are integrated with world-class physician-scientists.

We utilize top-quality equipment and highly experienced personnel to offer superior integrated genomics capabilities.

The following services are available to support your genomics research.

NEXT-GENERATION RESEARCH AND CLINICAL SEQUENCING FACILITY

• NYGC has 16 Illumina HiSeq X instruments to enable sequencing of 28,000 30x human whole genomes or 7,000 tumor/normal pairs at 80x/40x each year and 12 Illumina HiSeq 2500s to allow annual sequencing of tens of thousands of whole exome, RNA, methylation, or custom samples.

• Our high-throughput processes and rapid turnaround times are achieved by extensive automation through wet-lab robotics and highly parallelized analysis pipelines. This, coupled with the highly experienced scientific staff, yields top quality data.

• NYGC also has a CLIA certified clinical lab that satisfies New York State CLEP regulations. For information on approved tests please visit www.nygenome.org

TECHNOLOGY INNOVATION LABORATORY

• Composed of engineers, molecular biologists and chemists working together to develop cutting-edge technologies to optimize genomics research and clinical applications.

• Expertise in: single-cell sequencing, microfluidics, molecular biology, current and future next-generation sequencing technologies, chemical synthesis, and surface chemistry.
NEW YORK GENOME CENTER
SERVICES
HIGH-PERFORMANCE COMPUTING INFRASTRUCTURE

- Efficient and highly parallelized analysis of large-scale data sets achieved using a high-performance infrastructure containing 7,000 cores of computing power and 16Pb of online storage and 5Pb of redundant archival storage.

- Secure 2N storage facility provides a high level of protection against failure of any component.

BIOINFORMATICS EXPERTISE

- Expert bioinformaticians answer a wide range of biological questions by combining streamlined pipelines and varied technical approaches. Typical service includes standard pipeline analysis of sequencing data (eg, through variant annotation or gene quantification); in addition bioinformatics support is available for advanced analysis through to biological answers either on a fee-for-service or a collaborative basis.

- Automated pipelines identify diverse types of variants from a multitude of sample types. This includes, somatic analysis of tumor and matched-normal samples, familial analysis to identify de novo variation in an individual, metagenomic analysis, epigenetic analysis, and differential gene/transcript analysis.

- Additional areas of expertise include variant curation through literature review, pathway analysis, cell-type specific biology, extended pedigree analysis, Mendelian and complex disease studies, pathogen identification and interpretation, integration of patients’ DNA and RNA data, and fusion transcript detection.

DEDICATED PROJECT MANAGERS

- Each project will be assigned to a dedicated project manager with expertise in genomics and genomic technologies to provide a single point of contact from your initial inquiry through data delivery.

- Project Managers provide updates as samples pass through NYGC to give you up-to-date information on the progress of your project.

- Project Managers will communicate and address questions that are generated once data is provided.

To learn more about working with NYGC contact us at service@nygenome.org or 646-977-7222.
WHOLE GENOME SEQUENCING

Human WGS - Germline

- NYGC's whole genome service provides investigators with high quality whole genome data at the most competitive rate.
- NYGC has completed thousands of whole genomes on Illumina’s HiSeq X sequencing platform.

<table>
<thead>
<tr>
<th>COVERAGE OPTIONS:</th>
<th>Minimum of 30x</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIBRARY PREPARATION:</td>
<td>Illumina TruSeq Nano DNA or Illumina TruSeq PCR-Free DNA</td>
</tr>
<tr>
<td>DELIVERABLES:</td>
<td>We will provide mapped data (bam) and annotated variant calls (VCF)</td>
</tr>
<tr>
<td>TURN AROUND TIME:</td>
<td>8-10 weeks</td>
</tr>
</tbody>
</table>

Human WGS - Cancer

- NYGC's whole genome cancer service is designed to deliver the most high quality, complete profile to understand the molecular drivers of the cancer.
- Cancer WGS services include tumor only, tumor-normal and integration with RNA data options at a variety of different coverages.

<table>
<thead>
<tr>
<th>COVERAGE OPTIONS:</th>
<th>Suggested coverage of 40x/80x</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIBRARY PREPARATION:</td>
<td>Illumina TruSeq Nano DNA or Illumina TruSeq PCR-Free DNA</td>
</tr>
<tr>
<td>DELIVERABLES:</td>
<td>We will provide mapped data (bam) and annotated variant calls (VCF)</td>
</tr>
<tr>
<td>TURN AROUND TIME:</td>
<td>8-10 weeks</td>
</tr>
</tbody>
</table>

Human WGS Quality Control

- Samples sequenced at NYGC are subject to an array of rigorous in line quality assurance protocols.
- **Contamination checks** (software and chips); necessary for accuracy, confirmation that the right sample is present throughout, up-front analysis for sample contamination. Net results geared toward delivering a high level of accuracy.
- **Joint Genotyping**; data from each individual being sequence is compared with the total population and the related family members. Net results geared toward delivering a high level of accuracy, and ongoing analysis for sample identification.
- **Concordance checking** against the Illumina genotyping chip. Net results geared toward delivering a high level of accuracy and assurance of sample identification.
**Non-human WGS**

- NYGC’s non-human whole genome service can be customized to your organism of interest. Past projects have included mouse, rat, bacteria, virus many and other non-model organisms.
- De novo assembly is available as an additional service for genomes without a reference.

<table>
<thead>
<tr>
<th>LIBRARY PREPARATION:</th>
<th>Illumina TruSeq Nano DNA or Illumina TruSeq PCR-Free DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>DELIVERABLES:</td>
<td>We will provide mapped data (bam) and annotated variant calls (VCF)</td>
</tr>
<tr>
<td>TURN AROUND TIME:</td>
<td>8-10 weeks</td>
</tr>
</tbody>
</table>

**Whole Exome Sequencing**

**Human WES**

- NYGC uses Agilent SureSelect capture kits to provides optimal coverage by targeting ~21,000 genes.
- Whole exome sequencing plus 3’ and 5’ untranslated regions is also available and allows interrogation of regulatory genome regions.

<table>
<thead>
<tr>
<th>LIBRARY PREPARATION:</th>
<th>Agilent SureSelect XT, 51mb (exon) or 71mb (exon + UTR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DELIVERABLES:</td>
<td>We will provide mapped data (bam) and annotated variant calls (VCF)</td>
</tr>
<tr>
<td>TURN AROUND TIME:</td>
<td>8 weeks</td>
</tr>
</tbody>
</table>

**Mouse WES**

- NYGC has workflows to perform mouse specific alignment, variant calling, and annotation.

<table>
<thead>
<tr>
<th>LIBRARY PREPARATION:</th>
<th>Agilent SureSelect Mouse Exome</th>
</tr>
</thead>
<tbody>
<tr>
<td>DELIVERABLES:</td>
<td>We will provide mapped data (bam) and annotated variant calls (VCF)</td>
</tr>
<tr>
<td>TURN AROUND TIME:</td>
<td>8 weeks</td>
</tr>
</tbody>
</table>
**Human PDX WES & RNA**

- NYGC has designed analysis pipelines to deliver the highest quality human data from patient derived xenographs.

- Uses oligo-dT beads to enrich the polyadenylated RNA molecules in the sample.

- Read lengths available are paired-end 50bp and 125bp.

- Current library preparation protocols preserve the strand information.

- Agilent SureSelect capture kits provide optimal coverage by targeting ~21,000 genes.

- Whole exome sequencing plus 3' and 5' untranslated regions is also available and allows interrogation of regulatory genome regions.

- Read lengths available are paired-end 50bp and 125bp.

- Ribosomal depletion in library preparation minimizes the presence of rRNA.

**Deliverables:**

We will provide mapped data (bam) and annotated variant calls (VCF)

**Turn Around Time:**

8 weeks

---

**RNA Sequencing**

**mRNA**

- Uses oligo-dT beads to enrich the polyadenylated RNA molecules in the sample.

- Read lengths available are paired-end 50bp and 125bp.

- Current library preparation protocols preserve the strand information.

**Library Preparation:**

Illumina TruSeq Stranded mRNA

**Deliverables:**

We will provide mapped data (bam) and differential expression analysis

**Turn Around Time:**

8 weeks
Total RNA

- Total RNA allows the study of protein coding RNA as well as long ncRNAs.
- Read lengths available are paired-end 50bp and 125bp.
- Ribosomal depletion in library preparation minimizes the presence rRNA.

LIBRARY PREPARATION: KAPA Stranded RNA-Seq with RiboErase

DELIVERABLES: We will provide mapped data (bam) and differential expression analysis

TURN AROUND TIME: 8 weeks

SERVICE OPTIONS

NYGC recognizes that one-size doesn't fit all projects. This catalog is designed to be an overview of standard NYGC offerings. For other services not listed, please inquire for more details at service@nygenome.org or 646-977-7222.
LANE SEQUENCING

NYGC Lane Sequencing Services refers to sequencing of investigator-prepared specialized libraries. Lane Sequencing can be used for any library preparation protocol not supported by NYGC high-throughput services. NYGC has designed its Lane Sequencing offering to be flexible and support the diverse application needs of the community while maintaining high quality results.

Example Uses of Lane Sequencing

- Custom targeted enrichment panels
- Single cell sequencing
- Metagenomics
- ChIP-seq
- HITS-CLIP
- ddRAD-seq
- shRNA screen libraries
- 16S libraries
- and many more!

<table>
<thead>
<tr>
<th>Type of Service</th>
<th>Turnaround Time</th>
<th>Quality Control</th>
<th>NYGC Warranty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Service</td>
<td>2-4 weeks</td>
<td>NYGC performs QC (^1)</td>
<td>NYGC will provide re-runs only if sequencer failure occurs</td>
</tr>
</tbody>
</table>

\(^1\) Library QC with Bioanalyzer (Fragment analyzer, Tape station, etc.), Qubit or Picogreen, and qPCR (using the appropriate Kapa Library Quant kit for your qPCR platform).
# LANE SEQUENCING OPTIONS

## HiSeq 2500 High Output

<table>
<thead>
<tr>
<th>Run Type</th>
<th>Maximum yield per lane (gigabases)</th>
<th>Minimum Submission</th>
</tr>
</thead>
<tbody>
<tr>
<td>1x50bp</td>
<td>11</td>
<td>8 lanes</td>
</tr>
<tr>
<td>1x100bp</td>
<td>23</td>
<td>8 lanes</td>
</tr>
<tr>
<td>1x125bp</td>
<td>28</td>
<td>8 lanes</td>
</tr>
<tr>
<td>2x50bp</td>
<td>23</td>
<td>4 lanes</td>
</tr>
<tr>
<td>2x100bp</td>
<td>45</td>
<td>8 lanes</td>
</tr>
<tr>
<td>2x125bp</td>
<td>56</td>
<td>4 lanes</td>
</tr>
</tbody>
</table>

1. *Illumina HiSeq® SBS Kit v4 reagents are used to provide maximal sequence output per lane, approximately 250 million paired-end reads.*

## HiSeq 2500 Rapid Run

<table>
<thead>
<tr>
<th>Run Type</th>
<th>Maximum yield per lane (gigabases)</th>
<th>Minimum Submission</th>
</tr>
</thead>
<tbody>
<tr>
<td>1x50bp</td>
<td>6</td>
<td>2 lanes</td>
</tr>
<tr>
<td>1x100bp</td>
<td>13</td>
<td>2 lanes</td>
</tr>
<tr>
<td>1x125bp</td>
<td>16</td>
<td>2 lanes</td>
</tr>
<tr>
<td>1x150bp</td>
<td>38</td>
<td>2 lanes</td>
</tr>
<tr>
<td>1x250bp</td>
<td>63</td>
<td>2 lanes</td>
</tr>
<tr>
<td>2x50bp</td>
<td>13</td>
<td>2 lanes</td>
</tr>
<tr>
<td>2x100bp</td>
<td>25</td>
<td>2 lanes</td>
</tr>
<tr>
<td>2x125bp</td>
<td>31</td>
<td>2 lanes</td>
</tr>
<tr>
<td>2x150bp</td>
<td>75</td>
<td>2 lanes</td>
</tr>
<tr>
<td>2x250bp</td>
<td>125</td>
<td>2 lanes</td>
</tr>
</tbody>
</table>

1. *Illumina HiSeq® SBS Kit v2 reagents are used to provide maximal sequence output per lane, approximately 150 million paired-end reads.*
NYGC currently runs 11 application-specific fully automated analysis pipelines that process data from instrument through automated annotation.

Importantly, these pipeline components have been geared toward reducing compute time for generated data.

**Constitutional Whole Genome and Whole Exome Analysis**

- Alignment
- SNV and Indel Calling
- Structural Variant Calling (Whole Genome only)
  - Includes: copy number variants, translocations, inversions, and other complex variants
- Extended Familial Analysis

**Cancer Whole Genome and Whole Exome Analysis**

- SNV and Indel Calling
- Structural Variant Calling

**RNA-sequencing Analysis**

- Differential Expression Analysis
- Functional Annotation
- Fusion Gene Discovery

**Mitochondrial and Functional Analysis**

**Microbiome Analysis**

**Epigenetics Analysis Pipelines**
DATA DELIVERY, STORAGE, AND COMPUTE

Data Delivery Options

- NYGC has designed a full service solution to work with and store your data following sequencing and analysis.
- SFTP/SCP: Access data via your own FTP client software, or SCP/SFTP from a command line.
- Robust, high-speed data delivery via Globus’ managed data transfer service.
- Other options also available.

Data Storage

Data is delivered via NYGC’s data delivery portal and stored for three months. Additional storage options are also available.

Compute

After NYGC has produced raw data or performed initial analysis, investigators may wish to complete their own additional analysis. Investigators have the opportunity to use NYGC’s high-performance computing cluster to analyze large data sets that require capacity beyond what is locally available.

For more information, please inquire at service@nygenome.org or visit us online at www.nygenome.org