DATA STORAGE POLICY

WGS and Direct RNA Sequencing Deliverables

Overview

Deliverables for Nanopore Whole Genome Sequencing (WGS) and Direct RNA Sequencing include large files in multiple data formats, such as FASTQ, BAM, and VCF. In general, file sizes per sample may reach up to 400 GB.

In addition, ONT's proprietary raw signal files (Pod5 format) are available upon request. These files are typically 1.0-2.5 TB in size and are only primary used for performing custom base calling using ONT's software tools (MinKNOW or Dorado) or compatible open-source alternatives.

Data Storage and Retention

1. **Data Acquisition and Initial Storage**
   Immediate Dual Storage: Upon receipt, raw data (FASTQ and/or BAM files) are immediately stored in two separate locations, a data acquisition server and a portable hard drive.

2. **Data Retention and Archiving**
   Data (FASTQ and/or BAM files) will be retained for 6 months for free and additional storage options are available. To utilize one of the options below, please complete the request form.

Data Type	5 Years	10 Years
WGS, without Pod5	$50.00	$100.00
WGS, with Pod5	$200.00	$400.00
Direct RNA, without Pod5	$25.00	$50.00
Direct RNA, with Pod5	$100.00	$200.00

3. **Data Security and Privacy**
   Access to the data is restricted to authorized personnel only.

4. **Data Transfer**
   Data will be transferred to the client using a secure pCloud drive. A download link will be provided at the time the final report is delivered. The data will remain available in pCloud for two weeks at no additional cost. Clients are advised to download their data within this period, as it will be automatically deleted after two weeks. If extended storage in pCloud is required, the cost will be $25/month and does not include Pod5 files. To utilize this option, please complete the request form.

Suggested Option

For long-term data storage, you may provide your own storage location and we can transfer the data to that location with your consent. A Data Transfer Consent Form is required. Please complete and email to support@poochonscientific.com.

Typical WGS Data File Sizes

Coverage for WGS may be 30-60X. For 50X coverage, for example, a single sample generates around 360 GB of data. The FASTQ files alone for a 50x WGS sample may be around 150 GB, the aligned BAM file may be around 200 GB, plus gVCF files and other files are around 2 GB. The Pod5 files, available upon request, may be around 1.8 TB of additional data for a single sample at 50X coverage.

Typical Direct RNA Data File Sizes

For Direct RNA sequencing, a single sample generates around 30 GB of data. The FASTQ files alone may be around 16 GB, the aligned BAM file may be around 14 GB, plus other files are around 200 MB. The Pod5 files, available upon request, may be around 900 GB of additional data for a single sample.