

OD Example Plan: Generation of HeLa cell Whole Genome Sequence Data**DATA MANAGEMENT AND SHARING PLAN**

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on [sharing.nih.gov](https://www.nih.gov/data-management/data-sharing). The Plan is recommended not to exceed two pages. Text in italics should be deleted. There is no "form page" for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format* shown below.

Element 1: Data Type**A. Types and amount of scientific data expected to be generated in the project:**

Type	Species	Platform/ Source	Amount
HiFi long-read whole-genome sequencing data	Human	PacBio/HeLa Cell Line, obtained from ATCC	One cell line, 700 MB

B. Scientific data that will be preserved and shared, and the rationale for doing so:

All the scientific data (and accompanying metadata described in 1.C.) will be preserved and shared. These data will complement existing whole-genome data that are available from the HeLa cell line maintained in dbGaP.

Metadata, other relevant data, and associated documentation:

Metadata: QC Metrics, relevant metadata pertaining to the sample (obtained from ATCC), and other metadata required for dbGaP deposition.

Associated Documentation: Study methods and protocols.

Element 2: Related Tools, Software and/or Code:

We are not generating any new tools, software and/or code for processing or analyzing these HiFi WGS data (i.e., FASTQ files); we will make use of open-source tools that are freely available to the scientific community.

Element 3: Standards:

Data Type	Standard
HiFi long-read whole-genome sequencing data	FASTQ

Element 4: Data Preservation, Access, and Associated Timelines**A. Repository where scientific data and metadata will be archived:**

All scientific data and metadata will be deposited to dbGaP.

B. How scientific data will be findable and identifiable:

Our dataset will be registered and deposited in dbGaP and assigned a phsID. Data will be findable and identifiable via the standard data indexing tools in dbGaP/NCBI. We will reference the accession number(s) for our dataset(s) in all relevant future publications.

C. When and how long the scientific data will be made available:

We will meet the data submission and release timeframes specified by the NIH Genomic Data Sharing and Data Management and Sharing Policies, as described on NIH's data sharing website and NHGRI's data sharing policies and expectations webpage. In accordance with the NIH Genomic Data Sharing Policy and NHGRI's Expectations for Data Submissions and Release timelines for level 2 data, we will begin submitting genomic data no later than 3 months after data is generated and quality measures has been assessed. Genomic data will be released 6 months after they are submitted to dbGaP in accordance with the level 2 data release timeline. Metadata and associated documentation will be submitted along with the genomic data files, and the dataset will be released in full by the time any results, supported in whole or in part by this award, at time of associated

OD Example Plan: Generation of HeLa cell Whole Genome Sequence Data

publication. If we do not publish on these data or a portion of the data, they will be released by the end of this award. Level 3-4 data will not be generated for this project.

Currently, dbGaP has no process for deleting or retiring data sets; data will be available for as long as dbGaP preserves the dataset.

Element 5: Access, Distribution, or Reuse Considerations

A. Factors affecting subsequent access, distribution, or reuse of scientific data:

The NIH-Lacks Family Agreement (Agreement) expects NIH-funded investigators who generate HeLa cell whole genome sequence data (DNA or RNA) to submit those data to NIH through dbGaP to the HeLa cell Genome Sequencing Studies. The data are distributed according to the HeLa Genome Data Use Agreement under Health, Medical, and Biomedical (HMB) purposes. Through the NIH-Lacks Family Agreement, an Institutional Certification with the designated HMB data use limitation has been completed for the HeLa cell Genome Sequencing Studies and is held by dbGaP. No additional Institutional Certification is needed.

B. Whether access to scientific data will be controlled:

Yes, data will be controlled. Access requests for HeLa Cell Genome Sequences Studies are evaluated by the Advisory Committee to the Director (ACD) HeLa Genome Data Access Working Group to assess whether terms of use align with the HeLa Genome Data Use Agreement. The requests are then reviewed by the ACD to recommend to the NIH Director to approve or disapprove. The NIH Director makes the final access decision.

C. Protections for privacy, rights, and confidentiality of human research participants:

Upon receipt of an NIH Award, the data for this study will be protected by a Certificate of Confidentiality.

Element 6: Oversight of Data Management and Sharing:

The study PI will be overseeing execution of this Data Management and Sharing Plan. The Study PI will be assessing quality metrics and will deposit all scientific data and metadata according to the timelines provided above. Progress on data sharing will be reported in the Research Performance Progress Report.