HeLa Genome Data Use Agreement

National Institutes of Health
August 6, 2013

Last updated 1/25/2025

Introduction

The National Institutes of Health (NIH) database of Genotypes and Phenotypes (dbGaP) contains human genomic data and related information from research participants. In order to protect the privacy of those participants, NIH can control access to the use of the data by other researchers. Data access will be provided to research investigators who, along with their institutions, have certified their agreement with the expectations and terms of access detailed below. The HeLa Genome Data Access Working Group of the Advisory Committee to the NIH Director (ACD) will be responsible for reviewing any applications for access to the HeLa genome data in dbGaP determine whether they satisfy all of the terms and restrictions described here. The HeLa Genome Data Access Working Group's findings will be reported to the ACD, and the ACD will make recommendations to the NIH Director about whether a request should be approved or disapproved.

Although NIH has taken steps to control access to whole genome sequence data from HeLa cell lines in order to respect the wishes of the Henrietta Lacks family, the family and the NIH recognize that other genomic data from HeLa cell lines are available in open access. These data could be combined to recreate large portions of the HeLa cell genome sequence. We also recognize that the HeLa cell line can be sequenced de novo at any time. We strongly discourage such actions and will work with other funding agencies, scientific societies, and publishers to help ensure that the research community abides by the data access process.

Terms of Access

1. Research Use

The Requester agrees that if access is approved, the Principal Investigator named in the Data Access Request (DAR) submitted to the NIH, those named in the "Senior/Key Person Profile" portion of the DAR, which should include the Information Technology Director or his/her designee, and any trainee or employee working on the proposed research project under the direct supervision of these individuals, shall become Approved Users of the requested dataset(s). Research use will occur solely in connection with the research project described in the DAR, which includes a 1-2 paragraph description of the research objectives and design.

If the DAR process expects a Cloud Use Statement for investigators interested in using Cloud Computing, investigators must provide a Cloud Use Statement about the Cloud Service Provider (CSP) and/or Third-party IT system and agree to secure the data according to the NIH Security Best Practices for Users of Controlled-Access Data. The Cloud Use Statement should at least state the name of the CSP and/or Third-party IT system, the security standard, and how the CSP and/or Third-party IT system will be used to carry out the work described in the Research Use Statement. If applicable, the investigator should describe the role of any Collaborators in using the CSP and/or Third-party IT system. If the Approved User(s) plans to collaborate with investigators outside the Requester, the investigators at each external site must submit an independent DAR using the same project title and Research Use Statement, and if the DAR process expects when using the cloud, a

Cloud Use Statement.

New uses of these data outside those described in the DAR will require submission of a new DAR; modifications to the research project will require submission of an amendment to this application (e.g., the addition of new aims related to the approved project, adding or deleting collaborators from the same institution, and the potential addition of new NIH genomic datasets to an approved project).

The DAR must also include a statement by the Requester about whether the research is intended or could be reasonably expected to result in a patent or commercial product or service. If so, the Requester should include a description of that patent or commercial product or service. In addition, the Requester must agree to notify the HeLa Genome Data Access Working Group of any changes to these commercial expectations. The Requester and all Approved Users may use the dataset(s) only in accordance with the parameters described here.

Research access to the requested dataset(s) is granted for a period of one (1) year as defined below. Specific Terms of Access

Access to the HeLa genome data in dbGaP may be permitted for the following:

- Health/Medical/Biomedical research use: Use for any research question related to health or biomedical research.

2. Institutional and Approved User Responsibilities

The Requester agrees through the submission of the DAR that the PI named in the DAR has reviewed and understands the principles for responsible research use and data handling of the genomic datasets as defined in the NIH Security Best Practices for Users of Controlled-Access Data and as detailed in this agreement. The Requester and Approved Users further acknowledge that they are responsible for ensuring that all uses of the data are consistent with federal, state, and local laws and regulations and any relevant institutional policies. The Requester certifies that the Approved User is in good standing with the institution and relevant funding agencies (i.e., no known sanctions) and is eligible to conduct independent research. Through submission of the DAR, the Principal Investigator also agrees to submit Annual Data Use Reports to the HeLa Genome Data Access Working Group describing the research use of the data as described under *"Research Use Reporting"* below.

It is anticipated that, at least in some cases, these datasets will be updated with additional information. Unless otherwise indicated, all statements herein are presumed to be true and applicable to the access and use of all versions of these datasets.

3. Public Posting of Approved User's Research Use Statement

Information about the PI and the approved research will be posted on an NIH web site. The information will include the Approved User's name and institution, project name, Research Use Statement, and a Non-technical Summary of the Research Use Statement. In addition, citations resulting from the use of NIH genomic datasets will be posted on NIH data repository websites. In addition, and if applicable, this information may include the Cloud Use Statement and name of the CSP and/or Third-party IT system. Citations of publications

resulting from the use of controlled-access data obtained through this DAR may also be posted on the dbGaP website.

4. Privacy

Approved Users agree not to attempt to contact family members of Henrietta Lacks.

5. Non-Transferability

The Requester and Approved Users agree to retain control over the data and further agree not to distribute data obtained through this DAR to any entity or individual not covered in the submitted DAR. NIH genomic datasets obtained through this DAR, in whole or in part, may not be sold to any individual at any point in time for any purpose.

Approved Users agree that if they change institutions during the access period, they will submit a new DAR in which the new institution agrees to the NIH Security Best Practices for Users of Controlled-Access Data before data access resumes. Any versions of data stored at the prior institution for the approved use will be destroyed and documented through a Final Data Use Report as described below. However, if advance written notice and approval by NIH are obtained to transfer responsibility for the approved research project to another Approved User within the same institution, the data may not need to be destroyed.

6. Data Security and Data Release Reporting

The Requester and Approved Users agree to keep the data secure and confidential at all times and to adhere to information technology practices in all aspects of data management to assure that only authorized individuals can gain access to the HeLa datasets. This agreement includes the maintenance of appropriate controls over any copies or derivatives of the data obtained through this DAR.

All data security practices and other terms of use defined in this agreement and the NIH Security Best Practices for Users of Controlled-Access Data for the raw data are expected to be followed for the derived data, including any transmission of the data. The Requester and Approved Users, including the institutional Information Technology Director or his/her designee, acknowledge that they have reviewed and agree to handle the requested dataset(s) according to the current NIH Security Best Practices for Users of Controlled-Access Data, including its detailed description of requirements for security and encryption.

Requesters and Approved Users agree to notify the NIH of any unauthorized data sharing, breaches of data security, or inadvertent data releases that may compromise data confidentiality within 24 hours of when the incident is identified. As permitted by law, notifications should include the known information regarding the incident and a general description of the activities or process in place to fully define and remediate the situation.

Within 3 business days of the notification, the Requester, through the Approved User and the Institutional Signing Official, agree to submit to the NIH a more detailed written report including the date and nature of the event, actions taken or to be taken to remediate the issue(s), and plans or processes developed to prevent further problems, including specific information on timelines anticipated for action.

All notifications and written reports of data security incidents should be sent to:

Email: helagenome@nih.gov

The NIH may, as permitted by law, also investigate any data security incident. Approved Users and their associates agree to support such investigations and provide information, within the limits of applicable local, state and federal laws and regulations. In addition, Requesters and Approved Users agree to work with the NIH to assure that plans and procedures developed to address identified problems are mutually acceptable consistent with applicable law.

7. Intellectual Property

In Assoc. for Molecular Pathology et al. v. Myriad Genetics, Inc., et al. 569 U.S. ___ (2013), the United States Supreme Court ruled that DNA sequences isolated from the genome are not patentable.

8. Research Dissemination and Acknowledgement of NIH Genomic Study Datasets

It is the intent of the NIH to promote the dissemination of research findings from NIH genomic dataset(s) as widely as possible through scientific publication or other appropriate public dissemination mechanisms. Approved Users are strongly encouraged to publish their results in peer-reviewed journals and to present research findings at scientific meetings, etc.

There will be no publication embargo period associated with data included in this compilation.

Approved Users agree to acknowledge the Contributing Investigator(s) who contributed the HeLa sequence data, and the primary funding organization that supported the contributing study in all oral and written presentations, disclosures, and publications resulting from any analyses of the data. Approved Users further agree that the acknowledgment shall include the dbGaP accession number to the specific version of the dataset(s) analyzed.
The following acknowledgment, or a variation of it reviewed by the HeLa Genome Data Access Working Group, will be made in any dissemination of research findings:

> "The genome sequence described/used in this research was derived from a HeLa cell line (*url to dbGaP*). Henrietta Lacks, and the HeLa cell line that was established from her tumor cells without her knowledge or consent in 1951, have made significant contributions to scientific progress and advances in human health. We are grateful to Henrietta Lacks, now deceased, and to her surviving family members for their contributions to biomedical research. This study was reviewed by the NIH HeLa Genome Data Access Working Group."

9. Research Use Reporting

To assure that NIH policies and procedures for genomic data use are adhered to, Approved Users agree to provide to the HeLa Genome Data Access Working Group annual feedback on how these data have been used and any results that have been generated as a result of access to the data, including patents and publications. This annual report must include a description by the Requester of any patents or commercial products or services that have been made as the result of research using the HeLa sequence.

Approved Users who are seeking renewal agree to provide specific information in a renewal DAR. Those not seeking renewal agree to provide specific information to the HeLa Genome Data Access Working Group via the contact information below. Annual Data Use Reports will provide information regarding potentially significant findings and publications or presentations that resulted from the use of the requested dataset(s), a summary of any plans for future research use, any violations of the terms of access described within this agreement and the implemented remediation, and information on any downstream intellectual property generated as a result of the data. Approved Users also may include general comments regarding topics such as the effectiveness of the NIH genomic data access process (e.g., ease of access and use), appropriateness of data format, challenges in following the policies, and suggestions for improving data access or the program in general if desired.

Approved Users agree to send the Annual Data Use Report prior to the anniversary of the Approved Access Date and specified within the manifest file provided to Approved Users by the NIH Data Repository at the time that data access is provided. It is agreed that the Annual Data Use Report will be shared with the NIH within the context of a renewal DAR, or via a letter signed by the Institutional Signing Official and the Approved User.

Annual Data Use Reports should be submitted to:

> The NIH by e-mail at helagenome@nih.gov, unless otherwise indicated in automated reminder messages from NCBI/dbGaP. Requests for continued data access should be made through dbGaP.

10. Non-Endorsement, Indemnification

The Requester and Approved Users acknowledge that although all reasonable efforts have been taken to ensure the accuracy and reliability of NIH genomic data, the NIH, and Contributing Investigators do not and cannot warrant the results that may be obtained by using any data included therein. The NIH and all contributors to these datasets disclaim all warranties as to performance or fitness of the data for any particular purpose.

No indemnification for any loss, claim, damage, or liability is intended or provided by any party under this agreement. Each party shall be liable for any loss, claim, damage, or liability that said party incurs as a result of its activities under this agreement, except that the NIH, as an agency of the United States, may be liable only to the extent provided under the Federal Tort Claims Act, 28 U.S.C. 2671 et seq.

11. Termination and Violations

This agreement will be in effect for a period of one (1) year from the date the dataset(s) are made accessible to the Approved User ("Approved Access Date"). At the end of the access period, Approved Users agree to destroy all copies, versions, and Data Derivatives of the requested dataset(s) on both local servers and hardware, and if Cloud Computing was used, delete the data and cloud images from Cloud Computing provider storage, virtual and physical machines, and databases in accord with the NIH Security Best Practices for Users of Controlled-Access Data. However, the Requester may retain only encrypted copies of the minimum data necessary at their institution to comply with institutional scientific data retention policy, law, and scientific transparency expectations for disseminated research results, and/or journal policies. A Requester who retains data for any of these purposes

continues to be a steward of the data and is responsible for the management of the retained data in accordance with the NIH Security Best Practices for Users of Controlled-Access Data, and any institutional policies. Any retained data may only be used by the PI and Requester to support the findings (e.g., validation) resulting from the research described in the DAR that was submitted by the Requester and approved by NIH. The data may not be used to answer any additional research questions, even if they are within the scope of the approved DAR, unless the Requester submits a new DAR and is approved by NIH to conduct the additional research. If a Requester retains data for any of these purposes, the relevant portions of Terms 4, 5, 6, and 11 remain in effect after termination of this Agreement. These terms remain in effect until the data is destroyed. In instances where NIH provides written notification that Data Derivatives should be transferred to a NIH controlled-access data repository; the transfer must be completed prior to Project Close-out.

Consideration will be given to a renewal of this agreement upon submission of a new DAR. Copies of HeLa genomic dataset(s) may not need to be destroyed if, with advance notice and review by the HeLa Genome Data Access Working Group, the project has been transferred to another Approved User. In this case, documentation must be provided that other Approved Users are using the dataset(s) under an active approved research project at the same institution.

The Requester and Approved User acknowledge that the NIH or the Institute may terminate this agreement and immediately revoke access to all NIH genomic datasets at any time if the Requester is found to be no longer in agreement with the policies, principles, and procedures of the NIH and the Institute.

The HeLa Genome Data Access form and this policy may be updated periodically by the NIH Director following the review of the HeLa Genome Data Access Working Group and the recommendation of the Advisory Committee to the NIH Director.

****

By submission of the attached DAR, the Requester attests to the Approved Users' qualifications for access to and use of HeLa sequence dataset(s) and certifies their agreement to the NIH principles, policies, and procedures for the use of the requested datasets as articulated in this document and as summarized in the dbGaP Approved User Code of Conduct, including the potential termination of access should a violation of any of these agreement terms be identified.

Requesters and the Principal Investigator further acknowledge that they have shared this document, the dbGaP Approved User Code of Conduct, and the NIH GWAS data sharing policies and procedures for access and use of genomic datasets with any Approved Users, appropriate research staff, and all other Key Personnel identified in the DAR.

Institutional Signing Officials acknowledge that they have considered the relevant NIH policies and procedures, that they have shared this document and the relevant policies and procedures with appropriate institutional organizations and have assured compliance with local institutional policies related to technology transfer, information technology, privacy, and human subjects research.

Appendix

Definitions of Terms

Annual Data Use Report: A report submitted to the NIH on the anniversary of access approval summarizing the analysis of NIH genomic datasets obtained through the Data Access Request and any significant findings derived from the work.

Approved User: In addition to the PI, approved users may include the PI, collaborators at the home institution who are named in the "Senior/Key Person Profile" portion of the DAR, the IT Director or designee named in the "Senior/Key Person Profile" portion of the DAR, and trainees or staff to these investigators.

Cloud Computing: The National Institute for Standards and Technology defines cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. For more information see NIST Special Publication 800-145.

Cloud Service Provider (CSP): A company or institution that offers some component of cloud computing to other businesses or individual, typically Infrastructure as a Service (IaaS), Software as a Service (SaaS) or Platform as a Service (PaaS), as defined by the National Institute of Standards and Technology. For more information see NIST Special Publication 800-145.

Contributing Investigator: The researcher who submitted the genomic dataset to dbGaP.

Data Access Request: SF 424 (R&R) cover pages and requested attachments, if any.

Data Derivative: any data including individual-level data or aggregate genomic data that stems from the original dataset obtained through dbGaP. Excepted from this term is summary information that is expected to be shared through community publication practices.

Final Data Use Report: A final report submitted to the NIH at the conclusion of the approved access period when no additional access is sought, or when leaving an institution. This report should summarize the analysis of genomic study datasets obtained through the Data Access Request and any significant findings derived from the work.

Information Technology Director: Someone with the authority to vouch for the IT capacities at an institution, or higher-level division of an institution (e.g., the School of Medicine).

Institutional Signing Official: The label, "Signing Official," is used in conjunction with the NIH eRA Commons and refers to the individual that has institutional authority to legally bind the institution in grants administration matters. The individual fulfilling this role may have any number of titles in the institution but is typically located in its Office of Sponsored Research or equivalent.

Requester: The home institution/organization for the Primary Investigator (PI) that will use the requested data.

Senior/Key Persons: Collaborators at the home institution, and the IT Director or designee.

Third-party IT system: A collection of computing and/or communications components and other resources that support one or more functional objectives of an organization.