HeLa Genome Data Access Working Group

Report to the Advisory Committee to the Director

June 2014

Renee Jenkins, MD

Professor and Chair Emeritus, Department of Pediatrics and Child Health

Howard University

The HeLa Genome Data Use Agreement

Per the agreement between NIH and the Lacks family, NIH requests that ***all researchers***:

- Apply for access to HeLa whole genome sequence in the database of Genotype and Phenotype (dbGaP)
- Abide by terms outlined in the HeLa Genome Data Use Agreement, such as:
  - Data can only be used for biomedical research only; this does not include the study of population origins or ancestry
  - Requestors are not to make contact with the Lacks family
  - Requestors are to disclose any commercial plans
  - Requestors are to include an acknowledgment in publications and presentations
- Deposit future whole genome sequence data into dbGaP

Role of HeLa Genome Data Access Working Group

- Evaluate requests to access HeLa cell genome data in dbGaP for consistency with the terms of the HeLa Genome Data Use Agreement
- Report findings to the Advisory Committee to the Director
- Make recommendations to the ACD on changes to the terms specified in the HeLa Genome Data Use Agreement

HeLa Genome Data Access Working Group Roster

Renee Jenkins, M.D. (Chair)
Professor and Chair Emeritus
Department of Pediatrics and Child Health
Howard University

Russ B. Altman, M.D., Ph.D.
Professor, Bioengineering, Genetics, & Medicine
Director, Biomedical Informatics Training Program
Stanford University

Ruth Faden, Ph.D., M.P.H.
Philip Franklin Wagley Professor in Biomedical Ethics
Director, Johns Hopkins Berman Institute of Bioethics
Johns Hopkins University

Kathy Hudson, Ph.D.
Deputy Director for Science, Outreach, and Policy
National Institutes of Health

David Lacks Jr.
Representative, Henrietta Lacks Family
Baltimore, MD

Richard M. Myers, Ph.D.
President, Director and Faculty Investigator
HudsonAlpha Institute

Robert Nussbaum, M.D.
Professor of Medicine
Chief of Division of Genomic Medicine
University of California, San Francisco

Veronica Spencer
Representative, Henrietta Lacks Family
Baltimore, MD

Clyde W. Yancy, M.D.
Professor in Medicine-Cardiology and Medical Social Sciences
Chief, Division of Medicine-Cardiology
Northwestern University
Feinberg School of Medicine

Working Group Evaluation Criteria

- Is the proposed research focused on health, medical, or biomedical research objectives?
  - Is the proposed research related to determining the ancestry or population origins of HeLa cells?
- Are there any plans to develop intellectual property? Specifically:
  - Does the requestor anticipate IP or developing commercial products or services?
  - Does the requestor foresee that IP or commercial products may arise from the proposed research?
  - Has the requestor agreed to notify NIH if their plans for IP or commercial products change?
- Are there any plans to publish or present findings?

Through special instructions, the requestor is advised to address these items in their Research Use Statement, in addition to describing the objectives, design, and analysis plan of the proposed research and providing a statement explaining why the HeLa cell genome sequence data is valuable for the proposed research. Plans for IP or commercialization of a product or service is not used by the Working Group to make a final evaluation; this information is obtained for disclosure purposes only.

Types of Findings Reported by the Working Group

In evaluating an Access Request, the Working Group will report a finding as:

- **Consistent** with the Data Use Agreement
- **Inconsistent** with the Data Use Agreement
- **Conditional** (will be consistent with the Data Use Agreement if NIH staff find that additional information obtained from the Requestor is satisfactory)
- **Pending** (will require a re-evaluation from the Working group once additional information is obtained from the Requestor)

Updates on HeLa Data Access Requests

- 23 data access requests evaluated by the Working Group
- 18 data access requests have been approved by the NIH Director
- 1 data access request was disapproved by the NIH Director
- 1 data access request is pending
- 3 data access requests are being reported to you today

Virtual ACD Meeting: March 28, 2014

- Enabled ACD review of HeLa Working Group finding in a flexible and asynchronous way
- Consistent with Working Group finding, the ACD recommended that Dr. Collins accept 8 requests and reject 1 request (though requestor can revise and resubmit)
- ACD recommended that changes be made to clarify that dissemination of findings is a requirement

Revisions to the Special Instructions Document

- The following statement was added:
  - "In keeping with NIH's commitment to transparency and enabling access to data from NIH-funded research and in order to enable the family of Henrietta Lacks to be aware of research findings generated with HeLa genome sequence data, NIH expects that research findings based on the HeLa genome sequence data will be disseminated and that the source of the data will be appropriately acknowledged."
- The document states that NIH recognizes that the expectation to disseminate research findings may not apply in some cases
  - Examples: data are being used to reproduce results, data are being used as a teaching resource, data are used in preliminary studies
- Requestors who do not plan to disseminate their findings are now asked to provide a justification in their request, for consideration by the Working Group

Working Group Findings: Evaluation of Access Requests

Since the last ACD meeting, the Working Group has found three requests consistent with the HeLa Genome Data Use Agreement

Project Title	Requestor's Affiliation	Working Group Findings
Somatic Mutations	Center for Genomic Regulation, Barcelona, Spain	***CONSISTENT WITH DATA USE AGREEMENT***
Identifying Impact of Genetic Variants on Transcription Factor Binding Sites in the Human Genome	University of British Columbia	***CONSISTENT WITH DATA USE AGREEMENT***
lncRNA and Chromatin Interactions in Human Cancer Cells	Jackson Laboratory	***CONSISTENT WITH DATA USE AGREEMENT***

ACD Vote and Recommendations

Three Data Access Requests:

Project Title	Requestor's Affiliation	Project Overview	Working Group Findings
Somatic Mutations	Center for Genomic Regulation, Barcelona, Spain	• The aim of the project is examine how the structure of one's genome may affect their susceptibility to DNA mutations. • They will also investigate whether mutation rates in cancers cells differ from mutation rates observed in non-cancer cells. • HeLa genome data, taken together with other dbGaP genome data from cancer and non-cancer samples, will allow the researchers to carry out a comprehensive assessment of how overall DNA structure affect mutation rates.	***CONSISTENT WITH DATA USE AGREEMENT***
Identifying Impact of Genetic Variants on Transcription Factor Binding Sites in the Human Genome	University of British Columbia	• The aim of the project is to examine how changes at particular sites in the genome affect the control of gene activity. • These particular sites are where special proteins called transcription factors interact with the DNA and control the activity of genes. • dbGaP HeLa data will be combined with public HeLa data available through the ENCODE project to perform more sensitive analyses than could be done previously with just the public data.	***CONSISTENT WITH DATA USE AGREEMENT***
lncRNA and Chromatin Interactions in Human Cancer Cells	Jackson Laboratory	• The aim of the project is to map the interactions of a type of RNA, called long non-coding RNA (lncRNA), with the DNA and DNA-associated proteins of HeLa cells. • lncRNAs are known to play a specific role in gene regulation, and the role of this type of RNA is not well characterized in cancer cells. • HeLa cell genome data will be used in analyses to identify the sites in the genome that these molecules associate and interact with.	***CONSISTENT WITH DATA USE AGREEMENT***

Workshop on Scientific and Ethical Issues Related to Open-Access HeLa Genomic Data

Report to the
Advisory Committee to the Director

June 2014

Kathy Hudson, PhD

Deputy Director for Science, Outreach, and Policy

National Institutes of Health

Background

- The current HeLa policy applies to whole genome data only
- Since release of the HeLa Policy, questions surfaced about applicability of the Policy to other HeLa genomic data types
- Other HeLa genomic data types are currently in open-access and include, for example, epigenetic or RNAseq data
- Should the policy be expanded to include other HeLa genomic data types?

Workshop Agenda Overview

- HeLa Genomic Data Currently in Open Access and Future Considerations
- Scientific value of HeLa genomic data
- The information revealed by and the privacy risks of different types of open-access HeLa genomic data
- Ethical implications of open versus controlled data access
- Applying the NIH HeLa Genomic Data Policy to Other HeLa Genomic Data Types
- Relevance of the NIH HeLa Genomic Data Policy to the Sharing of Other Human Genomic Data

Availability of HeLa Genomic Data

- A Google search for “HeLa cell” generates 13 million results
- ~80,000 publications citing HeLa cells
- There are 1,700 gigabases of HeLa sequence in NIH public data bases
  - Epigenetic
    - 633,000,000,000 bases
    - 814 microarrays
  - RNAseq (gene expression)
    - 892,000,000,000 bases
    - 3737 microarrays
  - Short Genomic Regions
    - Over 192,000,000,000 bases
- Should we put these type of HeLa genomic data in controlled-access, prospectively?

Pros and Cons of Access to HeLa Genomic Data in dbGaP

Pros	Cons
Minimizing harms to the Lacks family by limiting data uses and users	Takes time and certain credentials to be granted access - investigators view this as a burden
Provide transparency regarding who is using the data and for what purpose	Might slow research or decrease the number of people who would use the data to advance knowledge

Open access repositories have millions of users a day while the HeLa genomic data in controlled access has been downloaded 9 times in the last 4 months

Preferences of Participating Lacks Family Members

- Desire that scientists have efficient and effective access to data
- Concerned about delaying or halting the progress of science with HeLa cells
- Want to know about scientific developments with the use of HeLa cells

Preliminary Workshop Outcomes

- HeLa Genome Data Use Agreement
  - Not much enthusiasm for expanding policy
- Open vs Controlled Access to HeLa Genomic Data
  - HeLa genomic data, beyond whole genome sequences, do not need to be kept in controlled access; the family does not want to halt science and still wants to see HeLa sequences being used to advance science
- HeLa Cell Research Collection
- HeLa Cell Research Symposia
- Information sharing with the Lacks family
  - Periodic summary of how HeLa cells and sequences are being used in research

Next Steps

- Working Group discussion at their next meeting
- NIH to consult with the Lacks family
- Present the Working Group findings to the ACD
- ACD to make a recommendation to the NIH Director
- NIH Director will make a final decision on how to move forward

Workshop Participants

ACD Discussion