HeLa Genome Data Access Working Group

Report to the Advisory Committee to the Director

April 21, 2016

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

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

The HeLa Genome Data Use Agreement

Per the agreement between NIH and the Lacks family, NIH is requesting 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

Clyde Yancy, M.D., M.Sc. (co-chair)
Professor in Medicine-Cardiology and Medical Social Sciences
Chief, Division of Medicine-Cardiology
Northwestern University
Feinberg School of Medicine

Kathy Hudson, Ph.D. (co-chair)
Deputy Director for Science, Outreach, and Policy
National Institutes of Health

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

Lisa Cooper, M.D., M.P.H.
James F. Fries Professor of Medicine and Director, Johns Hopkins Center to Eliminate Cardiovascular Health Disparities
Johns Hopkins University School of Medicine

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

David Lacks Jr.
Representative, Henrietta Lacks Family

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

Robert Nussbaum, M.D.
Chief Medical Officer
Invitae Corporation

Veronica Spencer
Representative, Henrietta Lacks Family

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 or foresee IP or developing commercial products or services 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?

Types of Findings Reported by the Working Group

In evaluating a Data 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)

Status of Data Access Requests

Number of Requests	Status
48	Evaluated by the HeLa Genome Data Access Working Group
38	Approved by NIH Director
1	Disapproved by NIH Director
7	Disapproved by NIH staff (requestors did not respond to requests for clarifications regarding publication plans, IP, and/or the non-technical summary)
Number of New Requests	**Status**
7	Being reported to ACD today

Working Group Findings: Evaluation of Access Requests

Since the last ACD meeting, the Working Group has found 7 requests to be consistent with the HeLa Genome Data Use Agreement:

Project Title	Requestor's Affiliation	Project Overview	Working Group Findings
Methods Development for Phasing, Structural Variant Detection, and Epigenomics using HeLa as a Model System	Illumina, Inc.	• HeLa genome sequence data will be used to test a protocol designed to simultaneously measure maternal/paternal inheritance of certain genomic sequences and gene function in the HeLa cancer genome. • This research will provide a picture of the differences between normal and tumor cells to better understand why tumor cells behave differently.	CONSISTENT WITH DATA USE AGREEMENT
Improving Cancer Structural Variant Detection	Illumina, Inc.	• HeLa genome sequence will be used to assess weaknesses in genomic variant bioinformatics detection methods and to test the effectiveness of changes to the methods. • HeLa structural variants will be used to construct a simulated 'cancer genome' that can in turn be used to assess informatics methods and to estimate the impact of various changes. • Any improvements made to such detection methods will be beneficial to cancer research and ultimately help to identify accurate diagnosis and treatment of cancer.	CONSISTENT WITH DATA USE AGREEMENT

Project Title	Requestor's Affiliation	Project Overview	
Comparative Analysis and Sequence Profiling of eccDNA in the Nuclei of Malignant Transformed and Normal Human Cells to Measure CNV of Genes Encoding miRNA Important for Tumorgenesis	Adam Mickiewicz University, Poznań, Poland	• Extrachromosomal circular DNA (eccDNA) is found in all eukaryotic cells and consists of repetitive sequences of DNA. • The aim of this project is to obtain eccDNA sequence profiles of cancer cells and compare them to normal cells. • The results will be used to develop molecular diagnostic methods for ovary and cervical cancers.	CONSISTENT WITH DATA USE AGREEMENT
Methods Development to Study the Relationship Between Genetic Variation and DNA Double Strand Break in Cancer	MASSACHUSETTS INSTITUTE OF TECHNOLOGY	• The researchers are interested in using the HeLa data to determine the relationship between genomic mutations or breaks and cancer. • The results from the HeLa data will be compared to other cell lines to develop a computational method to investigate the potential impact of these mutations on the development of cancer.	CONSISTENT WITH DATA USE AGREEMENT

Project Title	Requestor's Affiliation	Project Overview	Working Group Findings
Quantitative Detection of Genome Structural Variants in Somatic Cells and Tissues by Next-Generation Sequencing	ALBERT EINSTEIN COLLEGE OF MEDICINE	• This research team has developed a novel analytical method to detect and quantify low-abundant, somatic DNA mutations, or mutation that occurs in any of the cells of the body except the sperm and egg cells. • The researchers plan to use dbGaP HeLa genomic data as a control to perform analysis of germline mutations and test their novel detection method. • The researchers also plan to sequence the HeLa genome from commercially available cells in order to detect erroneous mutations in the dbGaP HeLa sequence data that may have arisen from sample preparation.	CONSISTENT WITH DATA USE AGREEMENT
Chromosomal R-loop in the 3D Architecture of the Hela Genome	UNIVERSITY OF DEBRECEN, Hajdu-Bihar, Hungary	• An R-loop is a three-stranded nucleic acid structure, composed of a DNA- RNA mixture and the associated single-stranded DNA. R-loops have been implicated in various physiological and pathological processes regarding DNA metabolism. • Based on published findings, this researcher has determined the R-loop locations in HeLa cells and gaining access to the HeLa genome sequence would allow the investigator to validate these locations. • The results obtained from using the HeLa genomic sequence for validation purposes will be used as a control to test his 3D modelling algorithm in other cells types.	CONSISTENT WITH DATA USE AGREEMENT
Methods Development for Single Cell Epigenomics	Oregon Health and Science University	• The research team aims to develop new genomic technologies that will characterize any chemical changes found on DNA in a single cell. • HeLa cells will be used to develop and test the methods. Publicly available data on chemical changes found in HeLa will be used to validate the methods. • Once validated, the method will be applied to tumor samples that contain a mixture of cells.	CONSISTENT WITH DATA USE AGREEMENT

ACD Discussion, Vote, and Recommendations

Working Group Findings: Evaluation of Access Requests

Project Title	Requestor's Affiliation	Working Group Findings
Methods Development for Phasing, Structural Variant Detection, and Epigenomics using HeLa as a Model System	Illumina, Inc.	CONSISTENT WITH DATA USE AGREEMENT
Improving Cancer Structural Variant Detection	Illumina, Inc.	CONSISTENT WITH DATA USE AGREEMENT
Comparative Analysis and Sequence Profiling of eccDNA in the Nuclei of Malignant Transformed and Normal Human Cells to Measure CNV of Genes Encoding miRNA Important for Tumorgenesis	Adam Mickiewicz University, Poznań, Poland	CONSISTENT WITH DATA USE AGREEMENT
Methods Development to Study the Relationship Between Genetic Variation and DNA Double Strand Break in Cancer	MASSACHUSETTS INSTITUTE OF TECHNOLOGY	CONSISTENT WITH DATA USE AGREEMENT
Quantitative Detection of Genome Structural Variants in Somatic Cells and Tissues by Next-Generation Sequencing	ALBERT EINSTEIN COLLEGE OF MEDICINE	CONSISTENT WITH DATA USE AGREEMENT
Chromosomal R-loop in the 3D Architecture of the Hela Genome	UNIVERSITY OF DEBRECEN, Hajdu-Bihar, Hungary	CONSISTENT WITH DATA USE AGREEMENT
Methods Development for Single Cell Epigenomics	Oregon Health and Science University	CONSISTENT WITH DATA USE AGREEMENT