HeLa Genome Data Access Working Group

Report to the Advisory Committee to the Director

December 14, 2017

Carrie D. Wolinetz, Ph.D.
Acting Chief of Staff and Associate Director for Science Policy
National Institutes of Health

Lisa A. Cooper, M.D., M.P.H., FACP
Bloomberg Distinguished Professor
James F. Fries Professor of Medicine and Director, Johns Hopkins Center to Eliminate Cardiovascular Health Disparities
Johns Hopkins University 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
  - Evaluation not based on scientific merit
- 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

Lisa A. Cooper, M.D., M.P.H., FACP (Co-Chair)
Bloomberg Distinguished Professor,
James F. Fries Professor of Medicine and Director, Johns Hopkins Center to Eliminate Cardiovascular Health Disparities
Johns Hopkins University School of Medicine

Carrie D. Wolinetz, Ph.D. (Co-Chair)
Acting Chief of Staff and Associate Director for Science Policy
National Institutes of Health

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
Bloomberg School of Public Health
Johns Hopkins University

David Lacks Jr.
Representative, Henrietta Lacks Family

Jeri Lacks-Whye
Representative, Henrietta Lacks Family

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

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

Veronica Spencer
Representative, Henrietta Lacks Family

HeLa Whole Genome Sequence Data in dbGaP

Principle Investigator	Institution	Project Title	Submission Year
Andrew Adey	Oregon Health and Science University	Construction of thousands of single cell genome sequencing libraries using combinatorial indexing	2017
Jay Shendure	University of Washington	Massively multiplex single-cell Hi-C	2016
Xun Xu	BGI-Shenzhen, China	Full-length single-cell RNA-seq applied to a viral human cancer: Applications to HPV expression and splicing analysis in HeLa S3 cells	2016
Erez Aiden	Baylor College of Medicine	A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping	2016
Jay Shendure	University of Washington	Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions	2014
Jay Shendure	University of Washington	The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line	2013
Lars Steinmetz	European Molecular Biology Laboratory	The Genomic and Transcriptomic Landscape of a HeLa Cell Line	2013

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
74	Evaluated by the HeLa Genome Data Access Working Group
59	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
Validation of Our In-house Software for Refined Protein Identification	**University of Geneva**	• **The investigator previously developed and implemented MzVar, a software tool that uses data obtained from mass spectrometry (MS), an analytical method that separates chemical species based on size and charge, to identify proteins. The investigator developed the software mainly using HeLa proteomic data available in public databases.** • **The investigator proposes to use HeLa cell genome sequence to validate the HeLa proteins identified using the MzVar software and to identify unstable proteins that are present in the HeLa genome sequence but absent in their HeLa proteome database assembled by MzVar.**	**CONSISTENT WITH DATA USE AGREEMENT**
Statistical Methods for Profiling Intra-tumor Heterogeneity from Sequencing Data	**University of Chicago**	• **Cancer development can result in tumors composed of cells similar to one another or tumors composed of diverse cell type. The cell type variations observed in tumors is known as tumor heterogeneity and has important consequences for cancer diagnosis and treatment.** • **The investigator requests access to HeLa cell genome sequence data to validate the investigator's computational model, which relies on sequencing to help characterize single cells within tumors to detect tumor heterogeneity.**	**CONSISTENT WITH DATA USE AGREEMENT**

Working Group Findings: Evaluation of Access Requests

Project Title	Requestor's Affiliation	Project Overview	Working Group Findings
Long Non-coding RNA Transcriptional Landscape in HeLa Cells	Center for Genomic Regulation	• Recent evidence suggests that a special class of RNAs, called long non-coding RNAs (lncRNAs), are implicated in various diseases. The investigators seek to understand the expression of lncRNAs and their role in cancer. • The investigators plan to use HeLa genome sequence to find out how genomic alterations affect the regulation of expression of lncRNAs in cancer cells. This should enable a better understanding of how cancer arises, and what role lncRNAs play in this process.	CONSISTENT WITH DATA USE AGREEMENT
Methods Development for the Comparisons of Three-dimensional Genome Structure	Carnegie-Mellon University	• The investigator seeks to develop a method to quantify the similarity and differences in 3D chromosome structure across cell types and different disease states. • The investigator proposes to use the HeLa cell genome sequence to develop and validate this method to reveal the role of chromosome structure in disease, as well as to improve the understanding of how chromosomes are organized in HeLa and other cell types.	CONSISTENT WITH DATA USE AGREEMENT

Working Group Findings: Evaluation of Access Requests

Project Title	Requestor's Affiliation	Project Overview	Working Group Findings
Identification of Structural Change of Chromosomes Related to DNA Copy Number Changes in Cancer	Gwangju Institute of Science /Technology	• Aberrant genomic alterations such as chromosome addition, deletion, or duplication occur frequently in cancer. Such aberrant genomic alterations are known as copy number aberrations (CNAs) and may contribute to cancer development. • To investigate the contribution of CNAs to cancer, the investigator requests to use the HeLa cell whole genome sequence to develop a computational method to search the cancer genome for CNAs. HeLa cell genomic structural images will be used to validate the CNAs identified by their computational approach.	CONSISTENT WITH DATA USE AGREEMENT
Determining Differences in the HeLa Genome at the CDR1-AS Locus	Stanford University	• Previous work by the investigator identified that a gene important for normal brain function is expressed at very low levels in HeLa cells, but not many other cell lines. The cause of the low expression level in HeLa cells is unclear. • The investigator proposes to compare the HeLa cell genome sequence to the human reference genome to see if genomic differences in HeLa cells can explain the low expression level of the gene.	CONSISTENT WITH DATA USE AGREEMENT

Working Group Findings: Evaluation of Access Requests

Project Title	Requestor's Affiliation	Project Overview	Working Group Findings
Mapping Controlled HeLa Transfection MS Data Over Customized Proteome Database Description	Norwegian University of Science and Technology	• The investigators previously identified and characterized the HeLa proteome (all of the proteins within HeLa cells) using a technique that tagged individual proteins in HeLa cells with a stable isotope of carbon and identified the tagged proteins using mass spectrometry (MS), an analytical technique that can identify chemicals in a mixture based on their physical characteristics. • The investigators propose to use HeLa cell genome sequence to validate their initial characterization of the HeLa proteome and compare the HeLa genome to the proteome to check for proteins not previously identified.	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
Validation of Our In-house Software for Refined Protein Identification	University of Geneva	CONSISTENT WITH DATA USE AGREEMENT
Statistical Methods for Profiling Intra-tumor Heterogeneity from Sequencing Data	University of Chicago	CONSISTENT WITH DATA USE AGREEMENT
Long Non-coding RNA Transcriptional Landscape in HeLa Cells	Center for Genomic Regulation	CONSISTENT WITH DATA USE AGREEMENT
Methods Development for the Comparisons of Three-dimensional Genome Structure	Carnegie-Mellon University	CONSISTENT WITH DATA USE AGREEMENT

Working Group Findings: Evaluation of Access Requests

Project Title	Requestor's Affiliation	Working Group Findings
Identification of Structural Change of Chromosomes Related to DNA Copy Number Changes in Cancer	Gwangju Institute of Science /Technology	CONSISTENT WITH DATA USE AGREEMENT
Determining Differences in the HeLa Genome at the CDR1-AS Locus	Stanford University	CONSISTENT WITH DATA USE AGREEMENT
Mapping Controlled HeLa Transfection MS Data Over Customized Proteome Database Description	Norwegian University of Science and Technology	CONSISTENT WITH DATA USE AGREEMENT