NIH Advisory Committee to the Director
December 12, 2019

Summary of HeLa Genome Data Access Requests

1. Project #10411, Characterizing the role of immune gene polymorphisms in cancer predisposition
   University of California, San Diego, CA

2. Project #11062, High-risk HPV infection and development of cervical cancer
   National Institutes of Health, Bethesda, MD

3. Project #22136, Spatial organization and disorganization in immortal cell genomes,
   Institut Pasteur, France

4. Project #23220, RNAseq analysis using HeLa genome sequence
   Missouri State University, Springfield, MO

This page intentionally left blank

National Institutes of Health
Advisory Committee to the Director
HeLa Genome Data Access Working Group

HeLa Genome Data Access Request: Project 10411

Working Group Finding	Consistent with the Data Use Agreement
Project Title	**Characterizing the role of immune gene polymorphisms in cancer predisposition**
Date Received	7/2/2019
Project Summary (Provided by NIH)	• Previous laboratory studies conducted by the Requestor demonstrated that genetic changes in key immune system genes play a role in how tumors develop and affect the ability of the immune system to attack tumors. • The investigator proposes to build upon this research by studying the genetic changes in more detail, how the changes are related the increased susceptibility that some individuals have to tumors, and how these individuals might respond to therapies that help the immune system attack tumors. • The Hela Genome Sequencing Studies data will be used to perform an analysis of HeLa cell surface protein "flags" that play an important role in the ability of immune system to identify and attack the cancer cell.
Institution	University of California, San Diego, CA
Collaborator(s)	None
Research Use Statement (Provided by Requestor)	The immune system plays an important role in tumor surveillance and individual heterogeneity in immune genes is known to contribute to cancer risk. The objective of our study is to characterize polymorphic variation impacting anti-tumor immune activity in order to study the role of immune variation in cancer predisposition and individual potential to respond to immunotherapy. We have previously shown that inherited and somatic variation affecting the MHC and KIR loci have direct impact on tumor genome evolution via tumor-immune interactions and can influence immune cell infiltration into tumors. We plan to extend this work by identifying associations between polymorphic variants and immune phenotypes, investigate mechanisms linking polymorphic variation to predisposition and response to checkpoint inhibitor therapy, comparing to non-cancer cohorts and developing of predictive models of risk and response to therapy using datasets available through dbGaP. In order to advance our research, we are applying for access to datasets that support investigation of polymorphic variation underlying gene regulation in key immune cell types including the Immune Variation Consortium (phs000815.v1.p1) and the Database of Immune Cell Expression (phs001703.v2.p1). To study the role of immune variation in the context of cancer predisposition and response to immune checkpoint inhibitor therapies in specific cancer types, we are requesting the following datasets (phs000178.v10.p8, phs000254.v2.p1, phs000290.v1.p1, phs000291.v2.p1, phs000179.v5.p2, phs000631.v1.p1, phs000362.v1.p1, phs000209.v13.p3, phs000422.v1.p1, phs000518.v1.p1,

HeLa Genome Data Access Request: Project 10411

	phs000632.v1.p1, phs000815.v1.p1, phs001493.v1.p1, phs001041.v1.p1, phs001451.v1.p1.c1, phs000452.v2.p1.c1, phs001618.v1.p1.c1, phs001519.v1.p1.c1, phs001565.v1.p1.c1, phs001512.v1.p1, phs001469.v1.p1, phs001038.v1.p1,phs001572.v1.p1). This includes sequencing datasets that will serve as non-cancer control populations. Cohorts that are listed only for specific tumor types (melanoma: phs001005.v1.p1.c1, phs000452.v2.p1.c2, phs001427.v1.p1.c1, phs001257.v2.p1.c1; lung cancer: phs001618.v1.p1.c2, phs001464.v1.p1.c2, phs001565.v1.p1.c3; bladder cancer: phs001565.v1.p1.c2) will only be analyzed in the context of the approved tumor type. We intend to add other relevant studies as they become available via dbGaP. Finally, we are requesting access to phs000640.v7.p1 to support experimental analysis of MHC genotype-specific neoantigen presentation using HeLa cells. This project aims to investigate interactions between the inherited immune system, the somatic evolution of tumors and the potential to respond to immune checkpoint inhibitors using genomic data from healthy individuals and from tumors in patients treated with immunotherapy in clinical trials of immune checkpoint inhibitors. Although we do not have plans to generate IP, it is possible that our investigations could lead to development of generic statistical methods that could be filed as IP by the University of California San Diego. All such IP will be freely available to the academic community. HeLa cell line data will not be used in the development of any statistical methods for this project. We do not anticipate that HeLa data will be used to develop any IP under this project. We agree to notify the NIH should there be any change to our plans or expectations for IP development as pertains to HeLa cell line data. All research findings generated from the requested dbGaP datasets will be broadly disseminated to the scientific community through publication and presentation at relevant conferences, or through other means where appropriate.
Non-Technical Summary (Provided by Requestor)	Our research suggests that genetic variation in an individual's genome in regions encoding immune system related genes affects their ability of their immune system to detect and eliminate cancer cells. We are working to optimize a computational pipeline to determine genetic variants impacting immune gene activities carried by an individual. Applying this methodology to a population of cancer patients and a non-cancer population, we will investigate how genetic variation affecting immune genes contributes to cancer risk. By additionally studying genetic variants in the context of immune phenotypes and checkpoint inhibitor response data, we hope to harness genetic variants to better identify patients that will benefit from immunotherapies. All findings generated using the requested data sets will be made available through publications, presentations or by other means where appropriate.

National Institutes of Health
Advisory Committee to the Director
HeLa Genome Data Access Working Group

HeLa Genome Data Access Request: Project 11062

Working Group Finding	Consistent with the Data Use Agreement
Project Title	**High-risk HPV infection and development of cervical cancer**
Date Received	10/9/2019
Project Summary (Provided by NIH)	• The investigator previously developed a method to detect the location of viruses that integrate, or cut-and-paste themselves, into other genomes. Viral integration into other genomes is random and, sometimes can promote cancer cell development. • The investigator proposes to use the HeLa Cell Genome Sequencing Studies, along with their new method, to verify viral integration and in some cases, better understand how viral integration at a particular location may have promoted cancer cell development.
Institution	National Institutes of Health, Bethesda, MD
Collaborator(s)	Internal
Research Use Statement (Provided by Requestor)	We are studying HPV infection-leading to development cervical cancer. Integration of HPV18 genome into HPV18-infected cervical cells cause host cell genome instability and leads to immortalization and transformation of the HPV18-infected cells, ultimately to cervical cancer. HeLa cells harbor the integrated HPV18 genome randomly in multiple chromosomes. The linearized HPV18 genome at E2 ORF region for HPV18 integration into cervical cells features increased transcription of viral E6 and E7 oncogenes and extensive alternative RNA splicing of viral transcripts. However, whether viral transcripts will be spliced to host gene transcripts to make chimeric virus-host transcripts to encode novel chimeric HPV18-host proteins remains largely unexplored. The proposed studies would like to understand the genetic information of HeLa cells that contributes to high-risk HPV18 infection and development of cervical cancer. As a model cell for other high-risk HPV infections, the HeLa genome sequence data are of value to the proposed research if we could access the sequence data for our study. Although the new chimeric virus-host splicing patterns or novel chimeric virus-host protein(s) could be discovered as a diagnostic or therapeutic biomarker from our study and we might file our findings as an NIH employee invention report through NCI Office of Technology Transfer, we are not sure if these chimeric virus-host products are common in all HPV18-induced cervical cancer patients or could be extended to other HPV infections such as HPV16, HPV31, HPV58, etc., due to the nature of random integration of the HPV genome into the host genome. Nevertheless, we will notify the potential, commercial findings to NIH under the terms of the HeLa Genome Data Use Agreement. However, we will disseminate our findings by publication if accumulated data are repeatable and reliable. We will then of course acknowledge the source of the HeLa Genome Data and the family in our publications and presentations accordingly.

National Institutes of Health
Advisory Committee to the Director
HeLa Genome Data Access Working Group

HeLa Genome Data Access Request: Project 11062

Non-Technical Summary (Provided by Requestor)	Cervical cancer genomes are characterized by frequent rearrangements - that is, when chromosomes break and join up with a linearized human papillomavirus genome. It is believed that such random integrations into host chromosomes of a tumorvirus genome are the "drivers" of cervical cancer at DNA level. However, the integrated genome transcription and RNA splicing may lead to production of chimeric host-virus transcripts and novel chimeric host-virus proteins at the posttranscriptional level. We have recently developed several novel methods for identifying viral integration and RNA splicing in the cervical cancer genome. We propose to test these methods using the HeLa genome as a gold standard, as the rearrangements and HPV18 integrations in its genome are well-defined by prior work by others. This study would assist us to understand how high-risk HPV infection to cause cervical cancer and might provide new biomarkers for cervical cancer diagnosis and future treatment.

National Institutes of Health
Advisory Committee to the Director
HeLa Genome Data Access Working Group

HeLa Genome Data Access Request: Project 22136

Working Group Finding	Consistent with the Data Use Agreement
Project Title	**Spatial organization and disorganization in immortal cell genomes**
Date Received	7/25/2019
Project Summary (Provided by NIH)	• The investigator's team has previously investigated the three-dimensional organization of genomes in various organisms such as bacteria and yeasts using special algorithms developed by the group. • The team would like to characterize loop structures in the HeLa genome with the goal of identifying the factors behind their formation and maintenance, and ultimately how the structures regulate gene expression and other vital functions.
Institution	Institut Pasteur, France
Collaborator(s)	Internal
Research Use Statement (Provided by Requestor)	We are a team working on the spatial organization of genomes in various organisms such as bacteria, yeasts and humans using genomic contact approaches such as Hi-C, 3Cseq and developing new algorithms to extract relevant biological information. The spatial organization of genomes is closely linked to biological functions such as replication, transcriptional regulation or segregation. Chromosome loops are an important feature of human architecture, they can connect together several different loci that are distant on the 1D representation of genomes (several hundred kb). We would like to investigate in detail the potential role of chromosome loops in the organization and disorganization of the Hela cell genome. We plan to systematically detect chromosome loop patterns in contact maps and correlate them with the presence of architectural proteins such as cohesin, CTCF or other genomic signals. These proteins have recently been shown to play a major role in chromosome 3D organisation. The generation of contact maps requires a high quality reference genome, so we will need the assembled genome from Landry's study, J.J.M., et al (2013). "The genomic and transcriptomic landscape of a HeLa cell line." G3-Genes Genomes Genetics 3(8): 1213-1224 (corresponding to dbGaP Study Accession phs000643.v1.p1.). We would like to re-analyze the contact data generated by Rao et al (Cell 159, 1665-1680, 2014) and deposited in dbGaP (phs001010.v4.p1) that have excellent spatial resolution and allow systematic loop detection as well massively multiplex Single-Cell Hi-C of HeLa Cells (phs001269.v3.p1). We hope to have a comprehensive view of all the chromosomal loops that can form in the Hela genome and identify the biological factors behind their formation and maintenance. Our computational study can provide new observations on this very particular type of cell and bring other findings to the long history of Hela cells. We do not anticipate any intellectual property (IP) or the development of

HeLa Genome Data Access Request: Project 22136

	commercial products or services related to our project and, in the event of a change, we undertake to notify the NIH. We plan to disseminate the results of the research we propose in scientific journals and, possibly, at scientific conferences. We will include the appropriate HeLa acknowledgement statement in all manuscripts and presentations.
Non-Technical Summary (Provided by Requestor)	The architecture of the human chromosome is a precise and sophisticated construction involving different proteins and mechanisms. Chromosome loops can connect different points and are an important feature of the spatial organization of the genome. Our project aims to detect all chromosomal loops present in the Hela genome that could explain its remarkable biological properties.

National Institutes of Health
Advisory Committee to the Director
HeLa Genome Data Access Working Group

HeLa Genome Data Access Request: Project 23220

Working Group Finding	**Consistent with the Data Use Agreement**
Project Title	**RNAseq analysis using HeLa genome sequence**
Date Received	10/21/2019
Project Summary (Provided by NIH)	• Previous research with cancer cells treated with quantum dots (man-made nanocrystals that can emit light) or metal nanoparticles such as platinum and palladium, reported slower cancer cell growth and perhaps cancer cell targeting. • The investigator proposes to use the HeLa Cell Genome Sequencing Studies to study the impact quantum dots and platinum and pallidum-based nanoparticles have on cellular physiology by measuring changes in HeLa cell gene expression in treated and untreated cells.
Institution	Missouri State University, Springfield, MO
Collaborator(s)	None
Research Use Statement (Provided by Requestor)	Our research objectives and plans: We would like to use the HeLa reference genome sequence for a high-throughput investigation we are conducting on HeLa cells, specifically for processing our short-read transcriptomic sequence data obtained from our RNA-seq experiment. The aim of our study is to investigate the toxicity and gain a better understanding of the effects of quantum dots, specifically green CdSe/ZnS and INP/ZnS quantum dots as well as platinum and palladium-based nanoparticles, on HeLa cells. Currently, quantum dots have a high potential use as a biosensor for detecting cancer, and recent studies have reported that quantum dots can inhibit the growth of cancer cells. Further, platinum and palladium-based nanoparticles are heavily used for cancer treatment, but their impact on cell physiology has not been well established. Together, we will treat these nanoparticles in growing budding yeast and mammalian cells to assess their effects on cell metabolism and transcriptome changes. Why HeLa genome sequence is required for our research? Previously, we used the human reference genome to map our HeLa transcripts, but the variation between the HeLa genome and human reference genome decreases the accuracy of gene expression data. With the HeLa reference genome, we will be able to accurately assess differentially expressed genes (DEGs) in HeLa cells when exposed to quantum dots. Through analysis of these DEGs and their functions, we aim to gain a better understanding on the mechanisms of toxicity used by these quantum dots in HeLa cells. Commercial potential from the project?

HeLa Genome Data Access Request: Project 23220

	We do not anticipate that intellectual property or commercial products/services will arise from this research. Just in case circumstances change, we will notify DbGap through the proper channels provided in the HeLa Genome Data Use Agreement Sheet. How the research will be disseminated? We expect a number of presentation at scientific avenues and the publication of our findings in scientific journals. All RNA-Seq data will be deposited to a database accessible to everyone, such as GEO. When publishing these RNA-Seq data, we will make sure to cite the source of HeLa sequence by using the acknowledgment statement available on the special instruction entitled "special instructions for preparing a research use statement for requesting access to HeLa Cell genome sequence data in dbGap". Additional compliance statement All HeLa sequences we obtain from DbGap will remain secure and confidential under our supervision. As an extra precautionary measure, the PI will be the only one accessing the data sets. We understand access and policies may be updated periodically at the discretion of the NIH Director and Advisory Committee following the completion of our DAR, and we are prepared to use the HeLa genome for the purposes mentioned above while respecting and abiding by the principles, policies and procedures laid out for us in the DbGap Approved User Code of Conduct.
Non-Technical Summary (Provided by Requestor)	Genes in HeLa cells are expressed at specific levels in normal and healthy conditions, but when in the presence of toxic materials, they change the amount they are expressed. We are trying to figure out if extremely small particles that could be used in cancer treatments, known as quantum dots, and other new anti-cancer drugs are toxic by exposing HeLa cells to these materials and observing which genes change their expression and by how much. We achieve this data by extracting the total RNA and comparing the sequences to a reference genome that represents the genome of the organism we tested on, in this case the organism was HeLa cells. With the gene expression data we could gather, we would be able to better understand how these materials are toxic to cancer cells which is important for their safe use and development.