Special Instructions for Preparing a Research Use Statement for Requesting Access to HeLa Cell Genome Sequence Data in dbGaP

When seeking access to the dbGaP HeLa Cell Genome Sequencing Studies, requestors should include the following items in their **Research Use Statements** in addition to describing the objectives, design, and analysis plan of the proposed research:

1. A statement explaining why the HeLa genome sequence data are of value to the proposed research (i.e., why these data are necessary to the research).
2. A statement about the commercial potential of the proposed research that addresses the following questions:
    A. Do you anticipate intellectual property (IP) or the development of commercial products or services?
    B. Is it foreseeable that IP or commercial products or services may arise from your research with HeLa cells?
    C. If your IP or commercial plans or expectations change, do you agree to notify the National Institutes of Health (NIH) under the terms of the HeLa Genome Data Use Agreement?
3. A statement about how the research findings from the proposed research will be disseminated (e.g., through publications, presentations).
    A. 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 investigators will include an acknowledgment of the source of the data in their publications and presentations.
    B. NIH recognizes that the expectation to disseminate research findings may not apply in some cases. For example,
        i. Data are to be used only to reproduce results: the HeLa genome sequence data are used solely for the purpose of verifying a publication or reproducing results, thus no new research findings are generated. ii. Data are used as a resource for teaching: the HeLa genome sequence data are used as a training resource, e.g., to teach a particular statistical method or analyses.
        iii. Data are used in preliminary studies: the HeLa genome sequence data are used in pilot studies or to generate preliminary findings for subsequent research that involves data solely from other sources. Even though the subsequent research findings may not be based on the HeLa genome sequence data, the publication should acknowledge Henrietta Lacks and her descendants if the HeLa sequence data informed the subsequent research.

    Requestors who do not plan to disseminate their findings should provide a justification in their Research Use Statement for consideration by the HeLa Genome Data Access Working Group.

In addition to the Research Use Statement, a requestor is required to also provide a **Non-Technical Research Summary**. The non-technical summary should describe the purpose and objectives of the proposed research in terms that are understandable to a lay reader. In addition to informing the HeLa Genome Data Access Working Group, which is composed of a multi-disciplinary scientific and clinical experts as well as members of the public, including members of the Lacks family, the non-technical summary is posted on the public portion of the dbGaP HeLa Cell Genome Sequencing Studies webpage.

Examples of Acceptable Non-Technical Summaries

Example 1

Short proteins (peptides) are displayed by cells to report to the immune system the internal workings of the cell. If something changes then these peptides change too. When all is well, these peptides are pieces of the cells own proteins. In an infection, they are parts of the microbe's proteins. We have found that after a virus infection the peptides from the cell do not return to normal. This can be harmful for transplants and other medical conditions if not everyone has the same peptides. We, therefore, want to know if these cell peptides vary among the population. Using the HeLa genome as a control (since the virus infected a derivative of HeLa cells), we will look for differences among the population at these cell peptide locations.

Example 2

Every day, cells in our body are affected by mutations that introduce small changes to the genetic material. The vast majority of these mutations is harmless, but some can cause serious problems such as cancer. It is well known that some external factors such as excessive sunlight or cigarette smoke accelerate the rate at which cells accumulate mutations. It is also understood that accumulation of mutations in the germ line are the driving force of evolution. However, questions remain how mutations are distributed across the genome, and whether this distribution is the same in body cells compared to those that are part of the reproductive system. We plan to investigate how a number of physical properties of the organization of the genome affect the distribution of mutations, and how these effects differ between cancer and germ-line cells.

Example 3

Cancer genomes are characterized by frequent rearrangements - that is, when chromosomes break and join up with other chromosomes. It is believed that in many cases, such rearrangements are the "drivers" of particular cancers. We have recently developed a novel method for identifying chromosomal rearrangements in cancer genomes. We propose to test this method using the HeLa genome as a gold standard, as the rearrangements in its genome are well-defined by prior work by us and others.