NIH Advisory Committee to the Director
March 28, 2014

Summary of HeLa Genome Data Access Requests

1. Mining Exome Sequencing and Array Based Genome Wide Association Studies for Disease Target and Biomarker Discovery
   Bristol-Myers Squibb

2. TargetInfectX – Multi-Pronged Perturbation of Pathogen Infection in Human Cells
   University of Basel

3. Identifying Allele-specific Biases in HeLa Transcript Processing
   University of North Carolina, Chapel Hill

4. Genomic Engineering in Alzheimer's Disease
   Flanders Interuniversity Institute for Biotechnology

5. Allele-specific Analyses in the HeLa Genome
   Yale University

6. Allele-specific Gene Regulation
   Uppsala University

7. Use of HeLa Cells to Study Assembly of HIV-1
   Rockefeller University

8. Compare HeLa Genome with SH-SY5Y Neuroblastoma Cell Line
   University of Luxembourg

9. Genome editing of HeLa Cells to Study Endocytosis
   Medical Research Council

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

HeLa Genome Data Access Request

Working Group Finding	Inconsistent with Data Use Agreement • Requestor has no intention to disseminate findings • Requestor did not indicate how HeLa cell genome sequence data are valuable for the proposed research

Project Title	**Mining Exome Sequencing and Array Based Genome Wide Association Studies for Disease Target and Biomarker Discovery**
Date Received	February 3, 2014
Requestor's Organization	Bristol-Myers Squibb
Project Overview	• The aim of the project is to study how changes in the sequence of the HeLa genome affect gene activity. • HeLa cell genome sequencing data will be used as a control when looking for biomarkers in other genomic datasets.
Research Use Statement (Supplied by Requestor)	Our overall objective is to identify novel therapeutic targets and biomarkers for patient stratification through the analysis of multi-dimensional genomics data. We aim to identify target genes with genetic variations (especially loss of function and other functional variants identified via exome sequencing technologies) associated with diseases. Additional analyses of other genomic data will be integrated with the above gene set. The combined analysis results will be used to select and prioritize candidate therapeutic targets and biomarkers. Our study plan is: 1) combine genotype and phenotype data from NHLBI GO exome sequencing studies within dbGaP and internal GWAS and sequencing data sets to determine genetic association at SNP level and gene level; 2) apply additional institutional knowledge bases (including disease, pathway, transcriptome and proteome databases, clinical information) to the association results for validation and gene and biomarker identification. We plan to use public, commercial, and in-house developed bioinformatics tools to analyze the data. The proposed research is consistent with the data use agreement. The intended use of data is for internal target identification and patient selection biomarker identification. The proposed project does not involve patient identification, and all patient-identifiable data will be kept under strict internal control and will not be shared or transferred.
Addendum to Research Use Statement (Clarifications received from Requestor)	• Our interest in the HeLa cell genome sequencing is just for control purposes of looking for biomarkers mainly in the NHLBI data sets. • We are NOT interested in any of the biological aspects of HeLa cells. • We do not anticipate IP or the development of commercial products or services. • We do not foresee that IP or commercial products or services may arise from our research with HeLa cells. • Yes, we will notify NIH our IP or commercial plans or expectations change. • The findings will not be disseminated.

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

HeLa Genome Data Access Request

Non-Technical Summary (Supplied by Requestor)	Genome wide association study using array based technologies is a powerful tool for identifying genetic variants associated to diseases and other phenotypes. Exome sequencing technology further enhances our ability to study rare coding variants, as well as fine mapping disease causal loci in complex diseases. The goal of this project is to identify and prioritize disease targets and biomarkers through data integration. We plan to combine GWAS and exome sequencing data from dbGaP with additional proprietary and public data sets. This would increase the statistical power to identify true genetic associations. Further analysis with gene-pathway based methods and correlation with clinical information will help us prioritize target and biomarker candidates. Ultimately, the proposed project could lead to new therapeutics and biomarkers for patient stratification.

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

HeLa Genome Data Access Request

Working Group Finding	Consistent with Data Use Agreement

Project Title	**TargetInfectX – Multi-Pronged Perturbation of Pathogen Infection in Human Cells**
Date Received	November 12, 2013
Requestor's Organization	University of Basel
Project Overview	• The aim of the project is to design genome analysis tools that will enable identification of the parts of the cell that allow viruses to enter the cell. • HeLa cell genome sequencing data will provide the gene sequence information required to properly design the tools. • This work may lead to an improved understanding of infection and ways to fight it.
Research Use Statement (Supplied by Requestor)	TargetInfectX strives to advance research in the fields of infection biology, RNA interference, phenotypic screening and image analysis. Currently we are using the genomic transcript sequences from RefSeq and ENSEMBL for our gene mapping. However, our experimental studies are done with the HeLa CCL2 line. The HeLa CCL2 sequence differs from these sequences as it contains specific mutations. Having access to the full genome dataset in your database, would allow us to have the exact sequence for our high content screens. It would also allow us to determine with a better accuracy which cellular components are relevant for the entry of a pathogen into human cells. This dataset would then be the only one used in our research. TargetInfectX is a consortium involving three institutions.
Addendum to Research Use Statement (Clarifications received from Requestor)	• There is no plan in the foreseeable future to develop a commercial product or service or file (IP) based on our findings from the proposed research. The basic research done in our laboratory aims at identifying the human protein network underlying infection and to identify targets for novel host-directed anti-infectives. The results of our findings will be published in peer reviewed journals and presented at scientific meetings. • We do not expect our research to result in a commercialized product or service. • We are not expecting to change our plans regarding our intentions not to seek IP or commercialization. However, if such option arises we will promptly notify the NIH and the dbGaP. • We plan to publish or present our findings from the proposed research. We plan to publish in peer reviewed, world renowned scientific journals and present our findings at scientific meetings. We would of course acknowledge access to the dbGaP data.

HeLa Genome Data Access Request

	• IT director named in our application controls user access and maintains IT security at the consortium level (TargetInfectX). He would be directly responsible of protecting access to the genome sequence and the data collected from our experiments. The data and research findings will be stored on openBIS : web-enabled database with CROWD authentication. Access to the server is restricted on a per-user level with detailed read- and write- permissions. He is responsible for managing individual accesses. • I confirm the understanding that the listed collaborators must submit complementary requests.
Non-Technical Summary (Supplied by Requestor)	With the rise of antibacterial resistance and the emergence and re-emergence of infectious diseases of bacterial and viral origin, there is a tremendous need for new ways of tackling intracellular pathogens. Discovering and understanding the human infectome involved in pathogen entry into host cells and determining the signaling network needed for successful pathogen invasion will help identify new ways to fight infection. Our RTD uses the HeLa CCL2 line for experimental studies. Having an exact match between sequence in your dataset and the experimental data would enable us to draw conclusions with higher confidence.

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

HeLa Genome Data Access Request

Working Group Finding	Consistent with Data Use Agreement

Project Title	**Identifying Allele-specific Biases in HeLa Transcript Processing**
Date Received	November 14, 2013
Requestor's Organization	University of North Carolina, Chapel Hill
Project Overview	• HeLa cell gene activity will be examined by comparing HeLa cell genome sequencing data to existing HeLa gene activity data. • Results from this work will provide clues as to when genetic changes begin to cause disease, such as chronic obstructive pulmonary disease.
Research Use Statement (Supplied by Requestor)	Allele specific changes are the crux of phenotypic variation, and of genetic disease susceptibility. Many of the disease associated sequence variants identified by Genome Wide Association Studies do not occur in coding sequence, meaning that these variants exert their influence in the steps from transcription through translation. We aim to identify allelic bias in post-transcriptional processing events, and to estimate the step(s) at which bias occurs. Using an existing HeLa RNA-seq dataset from HeLa S3 whole cells, nuclei and cytoplasm we can identify transcripts from heterozygous alleles that are differentially distributed across the cell compartments [1]. Relative allelic imbalance in the nucleus indicates bias in a nuclear event like transcription or degradation, while relative imbalance in the cytoplasm points to bias in nuclear export or localization. Thorough genotype information is needed to confirm heterozygous transcripts and genomic copy number information is necessary to accurately gauge relative allelic imbalances in the highly polyploid HeLa genome. Phasing information will increase the accuracy of allelic bias calls. We have the resources to further validate any results by performing additional sub-cellular fractionation and RNA extraction on HeLa cells followed by RNA-seq. Analyzing transcript abundance in this way yields insights into where allelic differences begin and thus is invaluable in guiding research on the roots of genetic disease. One such disease of interest is COPD, which we can directly apply our results to by searching COPD associated genes for allele specific biases. We have no plans to patent or commercialize any results or data from this analysis.
Addendum to Research Use Statement (Clarifications received from Requestor)	• We are asking theoretical, basic questions about allele specific effects and their contribution to disease. Fundamentally, we aim to discover novel processes by which human genetic variation will affect gene regulation post-transcriptionally. Our findings may lead to new RNA therapeutic targets for rare genetic diseases and genetic predispositions. Thus our findings could reasonably be expected to identify new targets for therapeutic target development and new

HeLa Genome Data Access Request

	genetic markers for disease pre-disposition to help with the diagnosis of disease etiology. • At this stage we have no intentions to seek IP or commercialization and do not expect this decision to change. We will immediately inform NIH should that change. We aim to discover the allele-specific variation that affects RNA localization in cells. Should we find such a system in the HeLA cell line (which we hypothesize is likely) we intend to publish these findings in an open source academic journal such as PLoS Biology.
Non-Technical Summary (Supplied by Requestor)	The most basic question in genetics, how gene sequence begets function, is crucial to unraveling the causes of genetic disease. Genetic disease susceptibility means that some alleles increase the risk of certain diseases. Alleles are versions of a gene, where every individual has two alleles per gene, one from the father and one from the mother. If one allele promotes altered cellular function (i.e. a disease state), this must mean that this allele or its products perform differently at the cellular level. We aim to identify such allelic biases in transcripts, the RNA intermediates between genes and their corresponding proteins. Though there are many cellular processes at work that can introduce allelic bias, we will focus on identifying allelic bias among transcripts in the processing steps before protein translation. Provided with the fully characterized HeLa genome we will use an existing HeLa transcript data set to identify allelic bias in transcript processing. Results will provide clues to when disease associated alleles begin to differ from their benign counterparts.

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

HeLa Genome Data Access Request

Working Group Finding	Consistent with Data Use Agreement

Project Title	**Genomic Engineering in Alzheimer's Disease**
Date Received	December 13, 2013
Requestor's Organization	Flanders Interuniversity Institute for Biotechnology
Project Overview	• The aim of the project is to design tools for genomic engineering that will allow researchers to alter the sequence of genes in HeLa cell lines. • These tools will enable the project team to assess the function of genes that are involved in neurodegenerative diseases such as Parkinson's disease.
Research Use Statement (Supplied by Requestor)	Recently gene editing nucleases have emerged as valuable tools for genomic engineering applications. These techniques hold a huge amount of perspectives: increasing insight in disease pathology, studying cell interactions and pathways, identifying genes, and an improved system to use for gene therapy. The HeLa cell line will be of use in genomic engineering applications based on gene editing nucleases. Research will be conducted for neurodegenerative diseases such as Alzheimer's and Parkinson's disease.
Addendum to Research Use Statement (Clarifications received from Requestor)	• We are using the HeLa cell line in our lab because it is a well-documented and studied cell line, and specifically because we have a protocol for studying mitophagy optimised to work with HeLa cells. In the context of this study we want to design CRISPR and TALEN sites to target genes important in neurodegenerative diseases. • Our primary interest is basic research, understanding mechanism underlying neurodegenerative diseases. We do not expect to commercialise products based on this research. Of course, if we develop plans regarding IP around the HeLa genome sequence, we will confer with the NIH. We do expect to publish this research and when we do we will acknowledge the HeLa genome project. • Dr. X is our most senior IT specialist and as such responsible for the security of the computer systems that will be used to store and analyse the HeLa genome sequence. We can switch to his manager in the application, but she is not an IT expert.
Non-Technical Summary (Supplied by Requestor)	The goal of this project is to gain more insights in the pathology of neurodegenerative diseases such as Alzheimer's and Parkinson's disease. For example, what is the interaction between several genes? How are mechanisms regulated at molecular level? These, and other related questions regarding neurodegenerative diseases will be addressed via genomic engineering tools.

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

HeLa Genome Data Access Request

Working Group Finding	Consistent with Data Use Agreement

Project Title **Allele-specific Analyses in the HeLa Genome**

Date Received	December 9, 2013
Requestor's Organization	Yale University
Project Overview	• The aim of the project is to further develop a genome analysis software tool called AlleleSeq.1, which is used with non-cancer genomes to detect changes in a gene that can alter gene activity. • HeLa genome cell sequencing data will be used to modify the software to analyze cancer genomes, which are more complex than non-cancer genomes.
Research Use Statement (Supplied by Requestor)	For the detection of variants associated with allele-specific events, we have previously developed a tool: AlleleSeq.1 It first constructs a phased diploid genome of the individual based on his/her personal genomic variants. Subsequently, it maps reads from functional genomic assays performed on the same individual to his/her personal diploid genome. By statistically inferring differential reads counts on the alleles at heterozygous variant loci, we were able to detect variants that are associated with allele-specific binding and expression from ChIP-seq and RNA-seq experiments respectively. This pipeline enables us to perform allele-specific analyses on a healthy diploid genome from GM12878. On the other hand, a cancer genome is typically an aneuploid and laden with genomic duplications and rearrangements on the scale of single-nucleotide to chromosomal level. This presents a challenging scenario for both the phasing of a cancer genome and the detection of variation associated with allele-specific events. Henceforth, the HeLa cell genomic data is a uniquely valuable resource, as it is a phased cancer genome with high sequence coverage. Moreover, we are in possession of data from ChIP-seq experiments of various transcription factors and RNA-seq experiments from the ENCODE consortium, in which we are a member of and in which the HeLa cell line is a Tier 1 cell line.2,3,4 Using this phased HeLa genome and HeLa-specific functional genomic assays, we plan to continue adapt the tool to perform allele-specific analyses on cancer genomes. There are and will be no plans to commercialize or file an Intellectual Property (IP) based on the findings from the proposed research; all tools developed from the proposed research will be published and publicly available for the scientific community.
Addendum to Research Use Statement (Clarifications received from Requestor)	• Dr. X is the system administrator in the lab who manages all the computer infrastructure in the lab (these are not managed by other university IT). Since all the HeLa genome data will be stored on private internal lab computers he is the appropriate IT director as he manages these computers.

HeLa Genome Data Access Request

	• We agree to inform the NIH if our plans change regarding IP or commercialization.
Non-Technical Summary (Supplied by Requestor)	A typical healthy human genome is diploid, with each chromosome of paternal or maternal origin, or haplotypes. Haplotypic information is key in pursuing allele-specific analyses. However, in cancer genomes, abnormal karyotypes often make haplotypes extremely difficult to characterize. The HeLa cell line is the most utilized cancer cell line in the world. It is thus a particularly important model for cancer. The availability of a phased and aneuploidy-aware HeLa genome provides haplotypic information that is vital in allele-specific analyses of the cancer genome. Here, we plan to look into identifying candidate genomic variations in the HeLa genome that are associated with allele-specific binding and expression.

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

HeLa Genome Data Access Request

Working Group Finding	Consistent with Data Use Agreement

Project Title **Allele-specific Gene Regulation**

Date Received	December 20, 2013
Requestor's Organization	Uppsala University
Project Overview	• The aim of the project is to study how genetic variations found in the HeLa genome affect gene activity. • Genetic variations found by comparing HeLa gene activity data generated in previous experiments to HeLa cell genome sequencing data will provide clues into the cause of diseases involving cervical cell defects.
Research Use Statement (Supplied by Requestor)	Many or most genes vary in activity between alleles and people but how this is regulated is poorly known. We will use the sequence of the HeLa genome to look for polymorphic regions where gene-regulatory signals differ between alleles. The benefit of studying HeLa is that polymorphic sites are mapped to the two haplotypes of the genome so especially when two or more polymorphisms are close to each other i.e. within the distance of a next generation sequence read, it is possible to map such reads to the correct haplotype with high accuracy. Transcription factors are gene regulatory proteins that bind to specific motifs in the DNA sequence. Their binding sites can now be efficiently mapped by chromatin immunoprecipitation and large scale sequencing (ChIP-seq). In each region that a transcription factor binds the signal is generated from many individual sequence reads. Several transcription factors have been mapped by ChIP-seq in HeLa cells. We will take ChIP-seq data from HeLa S3 and align each read to the two different haplotypes and look for regions in which more reads map to one allele than the other. Since we are analyzing many polymorphic regions we will use appropriate correction for multiple testing. We and others have previously shown many polymorphic sites with different transcription factor binding exist in the genome and we will now make a map of those that exist in the HeLa genome. We will correlate allele-specific ChIP-seq data to RNA-seq and search for genes with allelic imbalance i.e. genes with higher transcription form one allele than the other which may be the consequence of the allele-specific transcription factor binding. We will also search for possible disease association e-g by investigating if polymorphisms with allele-specific transcription factor signal are present in the GWAS catalogue for diseases that can be explained by defects in cervix epithelial cells.
Addendum to Research Use Statement	• We do not have any plans to develop a commercial product or service or file Intellectual Property based on this research. • At this point I cannot foresee that the results of our research will result

HeLa Genome Data Access Request

(Clarifications received from Requestor)	in a commercialized product or service. • I will inform NIH if the situation would change so that we would seek IP or develop a commercial product or service. • As with any biomedical research project we plan to publish the results in established scientific journals and at scientific conferences. A synopsis of the results may also be published on the homepage of the university.
Non-Technical Summary (Supplied by Requestor)	The DNA sequence is mostly identical between two people and differs only in about 1 letter in 1 000. Most of the variable sites are neutral but some will lead to a difference either by changing the structure of a gene or by making a gene more active in one person than the other. We have two copies of the genome, one from our fathers and one from our mothers and also they often differ in sequence. Some proteins bind to DNA and regulate gene activity. In this project we will look for regions in the HeLa genome in which proteins bind stronger to one gene variant than the other i.e. where the signal differs between the genomes inherited from the two parents. Such DNA sequences regulate gene activity and may create differences between people.

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

HeLa Genome Data Access Request

Working Group Finding	Consistent with Data Use Agreement

Project Title	**Use of HeLa Cells to Study Assembly of HIV-1**
Date Received	January 10, 2014
Requestor's Organization	Rockefeller University
Project Overview	• The aim of the project is to understand how HIV activity could be disrupted by identifying and studying key cell proteins involved in the virus formation process. • HeLa cell genome sequencing data will be used to further studies on how HIV develops into a mature and active virus in HeLa cells.
Research Use Statement (Supplied by Requestor)	We have been studying the assembly of HIV-1 in HeLa cells for eight years. We need the genomic sequence information from the HeLa cells to better understand the dynamics of the interactions between the virus and host proteins.
Addendum to Research Use Statement (Clarifications received from Requestor)	• We were the first to demonstrate that we can image single viruses of HIV-1 as they assemble - an important step for understanding how to disrupt the assembly. We published this work in a number of prominent journals including Nature and Nature Cell Biology. Our work indicates that key host proteins of the cell are needed for assembly. • We need the genomic sequence information from the HeLa cells to better understand the dynamics of the interactions between the virus and host proteins. • We have no plans to develop a commercial product or service or file Intellectual Property (IP) based on your findings from the proposed research and we plan to put all the results out in the public. The results may be used at some point in the future by other people to develop a commercialized product or service, but that has not been a goal of our work. We do not expect our intentions to change: For seven years we have put all of our results out in the public. • We will definitely inform the NIH if our plans for IP or commercialization change.
Non-Technical Summary (Supplied by Requestor)	Many treatments that are effective for blocking the spread of viruses work by blocking how the virus assembles in cells. We are studying how they assemble so we can better design a strategy for interfering with HIV-1. To do so, we need to know the sequence of some of the proteins inside of the HeLa cells.

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

HeLa Genome Data Access Request

Working Group Finding	Consistent with Data Use Agreement

Project Title	**Compare HeLa Genome with SH-SY5Y Neuroblastoma Cell Line**
Date Received	February 11, 2014
Requestor's Organization	University of Luxembourg
Project Overview	• The aim of the project is to determine which human cell lines are suitable for studying neurodegenerative diseases, such as Parkinson's disease. • A comparative analysis will be conducted using genome data from a neuroblastoma cell line and the HeLa cell genome sequencing data and the results will help guide the selection of genetically optimal cell lines for use in disease research.
Research Use Statement (Supplied by Requestor)	Cell lines are widely used in translational biomedical research to study the genetic basis of diseases. One major approach for experimental disease modeling is genetic perturbation experiment that aims to trigger selected cellular disease states. In these types of experiments it is crucial to ensure that the targeted disease-related genes and pathways are genetically intact in the used cell line. In this work, we develop a framework, which integrates genetic mutations in the cell line and disease-specific network analysis to evaluate the suitability of the cell line for disease-specific models. We have already sequenced the SH-SY5Y neuroblastoma cell line and evaluated it in the context of various neurodegenerative diseases. We intend to compare the genetic suitability of SH-SY5Y with other human cell lines that are widely used as *in vitro* models. In this respect, the sequence data of the HeLa cell line, which is extensively used in genetic perturbation experiments, is extremely valuable. The methodology that we use to evaluate the integrity of genetic networks underlying the selected diseases and processes builds upon whole genome sequencing of the cell line and gene-gene interaction networks, which model specific diseases. Based on the distribution of mutations in these interaction networks, we score the suitability of the cell line to study the corresponding disease. This analysis already performed on SH-SY5Y will be reproduced with HeLa cell genome data. Such a comparative analysis will guide the selection of the genetically optimal human cell line to study a certain disease. We believe that this methodology can help researchers worldwide to make an educated choice of experimental cell culture models for translational research.
Addendum to Research Use Statement (Clarifications received from Requestor)	• We do not anticipate the development of commercial products or services. • We do not foresee that IP or commercial products or services may arise from our research with HeLa cells. • We agree to notify NIH under the terms of the HeLa Genome Data Use

HeLa Genome Data Access Request

	Agreement if our IP or commercial plans or expectations change. • We intend to disseminate research findings from the proposed research on HeLa. The interim findings of our research using another human cell line genome sequence – SH-SY5Y – are already accepted for publication: *Biryukov, M., Antony, P., Krishna, A., May, P., Trefois, C.* **Evaluation of cell line suitability for disease-specific perturbation experiments.** Springer 2014. Studies in Classification, Data Analysis and Knowledge Organization (2014).
Non-Technical Summary (Supplied by Requestor)	Compare different cell lines to study certain biological process or pathways.

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

HeLa Genome Data Access Request

Working Group Finding	Consistent with Data Use Agreement

Project Title **Genome Editing of HeLa Cells to Study Endocytosis**

Date Received	February 14, 2014
Requestor's Organization	Medical Research Council
Project Overview	• The aim of the project is to design molecular tools that will allow researchers to alter the sequence of genes in HeLa cells. • These methods will be used to alter the sequence of specific genes that play a role in a basic cellular process called endocytosis where cells engulf and absorb various molecules that are outside the cell.
Research Use Statement (Supplied by Requestor)	I study pathways for endocytosis. A major limitation in this field is reliance on overexpressed GFP fusion proteins to study dynamic processes in live cells. A second major limitation is reliance on siRNA depletion for loss-of-function experiments in live cells. The advent of genome editing technology based on use of TALE or cas9 site-specific nucleases to stimulate homologous recombination should address both of these issues, as we can edit endogenous genes to express fusion proteins and we can make specific gene knockouts. Much of the literature in our area is based on use of HeLa cells, for the usual reasons of experimental tractability. We have developed powerful genome editing tools on other cell lines, but application to HeLa cells has so far been hit and miss. We suspect that this is because the HeLa cell genomic sequence is often different from the standard curated human genome sequence. Access to the HeLa genomic sequence would resolve this problem, allowing use to design HeLa-specific gene targeting constructs for homologous recombination at specific loci.
Addendum to Research Use Statement (Clarifications received from Requestor)	• We definitely have no specific plans to develop a commercial product or service, or to file IP. • We are engaged in basic research and wish to access the HeLa genome sequence to further that research. We wish to access the HeLa genome sequence to facilitate genome engineering in this cell line, as this will help our basic research into endocytic mechanisms and caveolae. • There is no likely expectation that this will result in commercial application. • There is no expectation to change our plans, but if we should then we undertake to inform NIH. • We certainly plan to publish our research in peer reviewed journals in the usual manner. However, the aims of the research are to understand endocytic mechanisms and the title and content of any publication is likely to reflect that.

HeLa Genome Data Access Request

	• Use of the HeLa genome sequence will be a very useful asset in achieving this goal, but study of the specific properties of the HeLa genome is not one of our aims.
Non-Technical Summary (Supplied by Requestor)	Endocytosis, internalisation of molecules from the surface of the cell, is a basic property in the biology of all mammalian cells. Endocytosis controls many aspects of the cell, including inter-cellular signalling pathways that are important in cancer. Not all mechanisms for endocytosis are well understood, and this is our area of interest. Very recently it has become possible to alter the genome of mammalian cell lines in specific ways, for example removing a specific gene, and this is a very powerful way of looking at the function of this gene. In order to do this, we need to know the sequence of the gene that we wish to alter. It would be very useful for us to know the sequence of the HeLa genome, as it is likely to be different in some ways from the standard human genome sequence. Our basic research provides the first step along the pathway that will ultimately provide therapies relevant to human health.