A New Chapter for the Lacks Family

Kathy Hudson, PhD

Deputy Director for Science, Outreach, and Policy

National Institutes of Health

Henrietta Lacks

- Henrietta was born in 1920
- At 31, she was being treated for aggressive cervical cancer at Johns Hopkins
- Researchers took cells from a biopsy for research without her knowledge
- She died that year
- Her cells have been growing for 62yrs
- Family has been dealing with this since

HeLa's History

1951. George Gey first to get human cells to continuously divide in culture

1971. *Obstetrics and Gynecology* identifies Henrietta Lacks as the source of HeLa cells, publishes a photo of Henrietta Lacks

1976. McKusick paper, genetic characteristics of the HeLa cell; includes Lacks pedigree

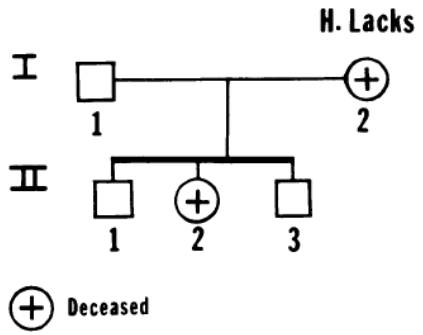
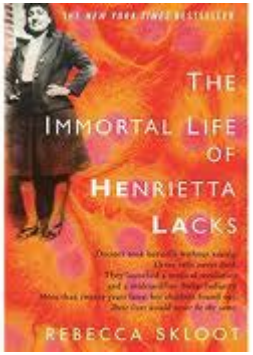


Fig. 1. The pedigree of the Lacks family.

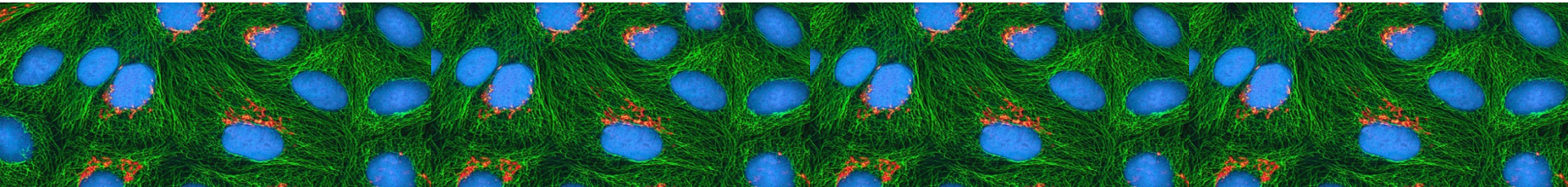
1997. BBC Documentary, *The Way of All Flesh*

2010. *The Immortal Life of Henrietta Lacks*, Skloot



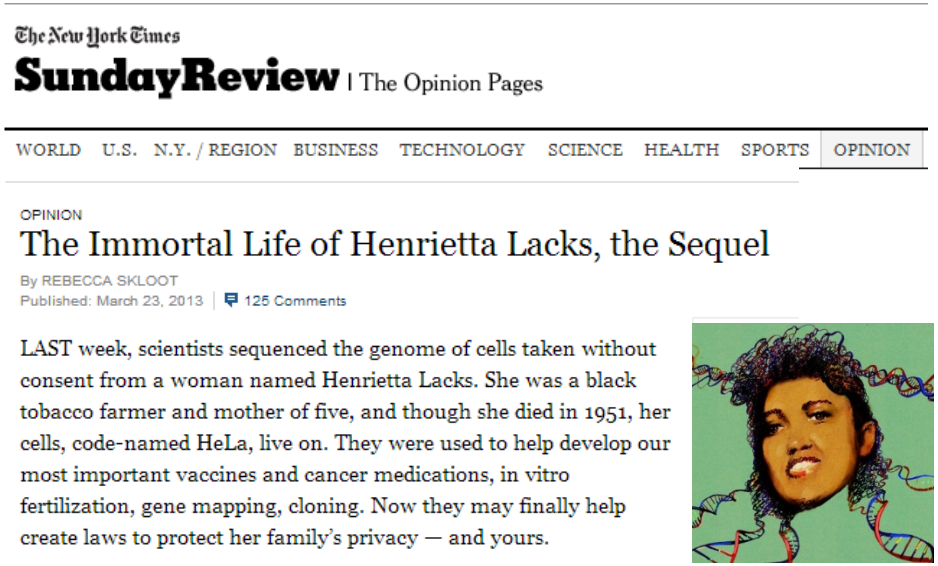
HeLa's Fame

- A Google search for "HeLa cell" generates **2.5 million results**
- **74,000 publications** citing HeLa cells
- In the last ten years, the majority of **Nobel Prizes** in medicine have used HeLa cells
- Accompanied the first man into **space** in 1961;
- And our **youngest scientists** are using HeLa too
  - HeLa cells and nanoscience to test new approaches to fight cancer *(2013 Semifinalist, Intel Science Talent Search)*
  - Screens for promising candidate compounds for a non-addictive painkiller *(2012 finalist, Siemens Competition in Math, Science, and Technology)*



Then…

- March 2013 – researchers in Germany posted the 1st HeLa whole genome sequence (EBI mirrored by NCBI)
- Lacks family asked that the sequence be removed – data taken down
- Another publication pending with Nature
- Growing public and media attention
- NIH reached out to the family



The New York Times
SundayReview | The Opinion Pages

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

OPINION

The Immortal Life of Henrietta Lacks, the Sequel

By REBECCA SKLOOT
Published: March 23, 2013 | 125 Comments

LAST week, scientists sequenced the genome of cells taken without consent from a woman named Henrietta Lacks. She was a black tobacco farmer and mother of five, and though she died in 1951, her cells, code-named HeLa, live on. They were used to help develop our most important vaccines and cancer medications, in vitro fertilization, gene mapping, cloning. Now they may finally help create laws to protect her family's privacy — and yours.

"I look at it as though these are my grandmother's medical records that are just out there for the world to see."

- Jeri Lacks-Whye, granddaughter

We Needed A Solution to Match the Problem

- HeLa cells and data are ubiquitous
- There are 1,300 gigabases of HeLa sequence in public data bases
- HeLa cells can be sequenced at any time
- The family has been through decades of unwanted intrusions and surprises
- No one had broken any laws
- Solution needed to advance science, respect family, and catalyze policy advances.

This is *not* precedent, HeLa is unique

NIH – Lacks Family Meetings

- NIH and the Lacks family - along with Ruth Faden, Dan Ford, and Rebecca Skloot - met 3 times: April 8th, May 6th, July 10th
- We talked about:
  - Lacks family's experiences, concerns, and hopes
  - What the HeLa genome can say about Henrietta and her family
  - Large amount of HeLa sequence data already public
  - Value of HeLa cells to science and medicine
  - Options for data access
  - Uniqueness of this situation

"The main goal was science and being part of the conversation"

- David Lacks Jr, grandson

August 7, 2013 – Agreement Reached

LETTER

doi:10.1038/nature12

The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line

Andrew Adey[1]*, Joshua N. Burton[1]*, Jacob O. Kitzman[1]*, Joseph B. Hiatt[1], Alexandra P. Lewis[1], Beth K. Martin[1], Ruolan Qiu[1], Choli Lee[1] & Jay Shendure[1]

The HeLa cell line was established in 1951 from cervical cancer cells taken from a patient, Henrietta Lacks. This was the first successful attempt to immortalize human-derived cells *in vitro*[1]. The robust growth and unrestricted distribution of HeLa cells resulted in its broad adoption—both intentionally and through widespread cross-contamination[2]—and for the past 60 years it has served a role analogous to that of a model organism[3]. The cumulative impact of the HeLa cell line on research is demonstrated by its occurrence in more than 74,000 PubMed abstracts (approximately 0.3%). The genomic architecture of HeLa remains largely unexplored beyond its karyotype[4], partly because like many cancers, its extensive aneuploidy renders such analyses challenging. We carried out haplotype-resolved whole-genome sequencing[5] of the HeLa CCL-2 strain, examined point- and indel-mutation variations, mapped copy-number variations and loss of heterozygosity regions, and phased variants across full chromosome arms. We also investigated variation and copy-number profiles for HeLa S3 and eight additional strains. We find that HeLa is relatively stable in terms of point variation, with few new mutations accumulating after early passaging. Haplotype resolution facilitated reconstruction of an amplified, highly rearranged region of chromosome 8q24.21 at which integration of the human papilloma virus type 18 (HPV-18) genome occurred and that is likely to be the event that initiated tumorigenesis. We combined these maps with RNA-seq[6] and ENCODE Project[7] data sets to phase the HeLa epigenome. This revealed strong, haplotype-specific activation of the proto-oncogene *MYC* by the integrated HPV-18 genome approximately 500 kilobases upstream, and enabled global analyses of the relationship between gene dosage and expression. These data provide an extensively phased, high-quality reference genome for past and future experiments relying on HeLa, and demonstrate the value of haplotype resolution for characterizing cancer genomes and epigenomes.

We generated a haplotype-resolved genome sequence of HeLa CCL-2 using a multifaceted approach that included shotgun, mate-pair and long-read sequencing, as well as sequencing of pools of fosmid clones[5] (Supplementary Table 1). To catalogue variants, we carried out conventional shotgun sequencing to 88× non-duplicate coverage and reanalysed 11 control germline genomes in parallel[8] (Supplementary Tables 2 and 3). Although normal tissue corresponding to HeLa is unavailable, the total number of single-nucleotide variants (SNVs) identified in HeLa CCL-2 ($n = 4.1 \times 10^6$) and the proportion overlapping with the 1000 Genomes Project[9] (90.2%) were similar to controls (mean $n = 4.2 \times 10^6$ and 87.7%, respectively), suggesting that HeLa has not accumulated appreciably large numbers of somatic SNVs relative to inherited variants. Indel variation was unremarkable after accounting for differences in coverage (Supplementary Fig. 1). Short tandem repeat profiles of HeLa also resembled controls, consistent with mismatch repair proficiency (Supplementary Fig. 2).

After removing protein-altering variants that overlapped with the 1000 Genomes Project or the Exome Sequencing Project[10], similar numbers of private protein-altering (PPA) SNVs were found in HeLa ($n = 269$) and controls (mean $n = 391$). Gene ontology anal found that all terms enriched for PPA variants in HeLa ($P \leq 0.01$) w also enriched in at least one control (except for 'startle response HeLa), suggesting that known cancer-related pathways are not p turbed extensively by point or indel mutations (Supplementary Fig. Although a previous study of the HeLa transcriptome[11] reported enrichment of putative mutations in cell-cycle- and E2F-related ge subsequently generated population-scale data sets contain all varia that we observed in these genes, suggesting that they are inherited benign rather than somatic and pathogenic.

The overlap between PPA variants and the Catalogue of Som Mutations in Cancer (COSMIC)[12] was similar for HeLa ($n = 1$) control genomes (mean $n = 2.6$). The gene-level overlap with Sanger Cancer Gene Census (SCGC)[12] was also similar for H ($n = 4$) and control genomes (mean $n = 8.7$). Canonical tumour s pressors and oncogenes were notably absent among the five SC genes with PPA variants in HeLa (*BCL11B* (B-cell CLL/lymphoma (zinc finger protein)), *EP300* (E1A binding protein p300), *FG* (fibroblast growth factor receptor 3), *NOTCH1* and *PRDM16* domain containing 16), Supplementary Tables 3–6). However, th are associated with HPV-mediated oncogenesis (*FGFR3*, *EP NOTCH1*) and may be ancillary to the dominant role of HPV on proteins in HeLa and other HPV$^+$ cervical carcinomas[13]. Mutation *FGFR3* have been noted previously in cervical carcinomas, altho infrequently and at different residues than observed here[14]. B *EP300* and *NOTCH1* are recurrently mutated in diverse cancers are involved in Notch signalling, a pathway that is dysregulated HeLa[15]. *EP300*, which encodes the transcriptional co-activator p interacts directly with viral oncoproteins such as HPV-16 E6 HPV-16 E7 (ref. 16). Although the in-frame deletion of a highly c served amino acid in *EP300* seems to be somatic (heterozygous with loss-of-heterozygosity (LOH) region), it is still possible that the oth are rare, inherited variants or passenger mutations. Further studies required to resolve their functional relevance and to assess whet these genes are recurrently altered in HPV$^+$ cervical carcinomas.

Aneuploidy and LOH, which are hallmarks of cancer genom were mapped in HeLa by constructing a digital copy-number profil kilobase resolution (Fig. 1, Supplementary Fig. 4 and Supplement Table 7). Read coverage profiles were segmented by a Hidden Mar Model (HMM) and recalibrated to account for widespread aneuplo (Supplementary Figs 5 and 6). Sixty-one per cent of the genome ha baseline copy number of three, and only a small minority (3%) has a c number of greater than four or less than two (Supplementary Table LOH encompassed 15.7% of the genome, including several entire ch mosome arms (5p, 6q, Xp, Xq) or large distal portions (2q, 3q, 6p, 1 13q, 19p, 22q) (Supplementary Fig. 7 and Supplementary Table 9), c sistent with previous descriptions of LOH in cervical carcinomas[17]. overall profile is consistent with published karyotypes of various H strains[4], suggesting that the hypertriploid state arose either during tu rigenesis or early in the establishment of the HeLa cell line.

[1]Department of Genome Sciences, University of Washington, Seattle, Washington 98115, USA.
*These authors contributed equally to this work.

COMMENT

COURTESY OF LACKS FAMILY

Henrietta Lacks' family gather around a historical marker dedicated to her in Virginia in 2011.

Family matters

Kathy L. Hudson and **Francis S. Collins** discuss how and why the US National Institutes of Health worked with the family of Henrietta Lacks, the unwitting source of the HeLa cell line, to craft an agreement for access to HeLa genome data.

In March, two of the most deeply held values in the medical-research community — public data-sharing and respect for research participants — collided when the genome of the ubiquitous cell line HeLa was published[1] and posted in a public database. Controversy ensued. The full sequence data could potentially uncover unwanted information about people whose identity is widely known: the family of the woman from whom this immortal line was derived 62 years ago, Henrietta Lacks.

So, since March, the US National Institutes of Health (NIH) in Bethesda, Maryland, has worked closely with Lacks' family. Together, we have crafted a path that addresses the family's concerns, including consent and privacy, while making the HeLa genomic sequence data available to scientists to further the family's commitment to biomedical research.

The agreement that we reached goes into effect this week. We hope that it, and its genesis, will spur broader discussions regarding consent for future use of biospecimens, with a goal of fostering true partnerships between researchers and research participants.

MEDICAL HISTORY

In 1951, physicians at Johns Hopkins Hospital in Baltimore, Maryland, took a biopsy from Henrietta Lacks, a 31-year-old African American woman who had an aggressive form of cervical cancer. This biospecimen was taken without her permission or knowledge; US regulations requiring consent ▶

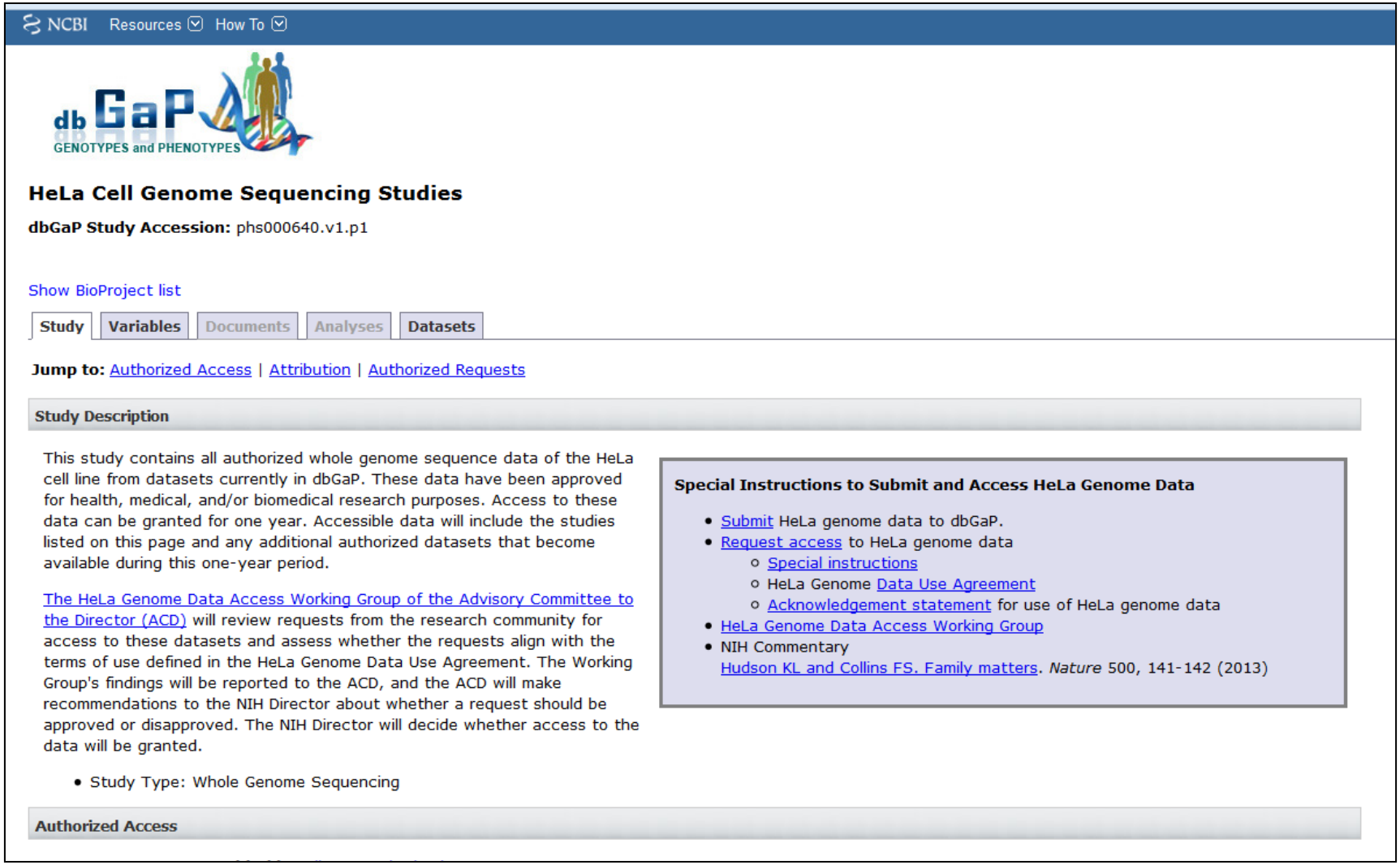
The Lacks Family and NIH: Working Together

Elements of the Agreement

NIH is requesting that **all researchers**:

- Apply for access to HeLa whole genome sequence
- Abide by terms defined by the Lacks family
  - Biomedical research only
  - No contact with family
  - Disclosure of commercial plans
  - Include acknowledgment in publications and presentations
- Deposit future whole genome sequence data into dbGaP
- This working group to review all requests

HeLa dbGaP page –



NCBI Resources How To

dbGaP GENOTYPES and PHENOTYPES

HeLa Cell Genome Sequencing Studies

dbGaP Study Accession: phs000640.v1.p1

Show BioProject list

Study | Variables | Documents | Analyses | Datasets

Jump to: Authorized Access | Attribution | Authorized Requests

Study Description

This study contains all authorized whole genome sequence data of the HeLa cell line from datasets currently in dbGaP. These data have been approved for health, medical, and/or biomedical research purposes. Access to these data can be granted for one year. Accessible data will include the studies listed on this page and any additional authorized datasets that become available during this one-year period.

The HeLa Genome Data Access Working Group of the Advisory Committee to the Director (ACD) will review requests from the research community for access to these datasets and assess whether the requests align with the terms of use defined in the HeLa Genome Data Use Agreement. The Working Group's findings will be reported to the ACD, and the ACD will make recommendations to the NIH Director about whether a request should be approved or disapproved. The NIH Director will decide whether access to the data will be granted.

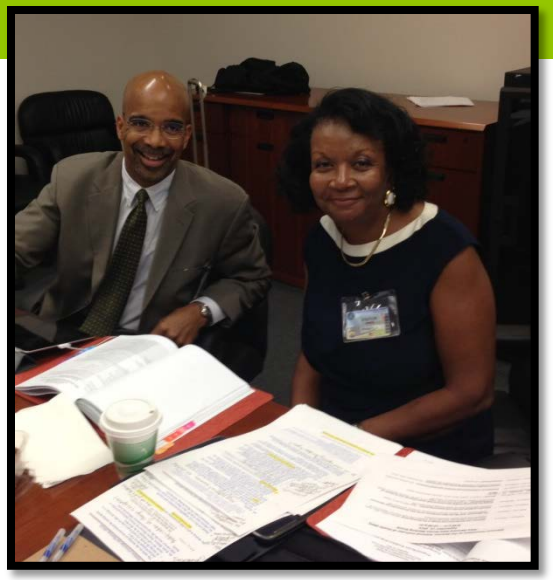
- Study Type: Whole Genome Sequencing

Special Instructions to Submit and Access HeLa Genome Data

- Submit HeLa genome data to dbGaP.
- Request access to HeLa genome data
  - Special instructions
  - HeLa Genome Data Use Agreement
  - Acknowledgement statement for use of HeLa genome data
- HeLa Genome Data Access Working Group
- NIH Commentary
  Hudson KL and Collins FS. Family matters. *Nature* 500, 141-142 (2013)

Authorized Access

HeLa Genome Data Access Working Group



Renee Jenkins, M.D. (CHAIR)
Professor and Chair Emeritus
Department of Pediatrics and Child Health
Howard University

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
Johns Hopkins University

David Lacks Jr.
Representative, Henrietta Lacks Family
Baltimore, MD

Veronica Spencer
Representative, Henrietta Lacks Family
Baltimore, MD

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