LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-675376

Lawrence Livermore National Laboratory-Completing the Human Genome Project and Triggering Nearly $1 Trillion in U.S. Economic Activity

J. S. Stewart

July 28, 2015

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Lawrence Livermore National Laboratory: Completing the Human Genome Project and Triggering Nearly $1 Trillion in U.S. Economic Activity

Human Genome Project

Prepared by LLNL under Contract DE-AC52-07NA27344.

Prepared by Jeffrey Stewart

August 10, 2015

LLNL-TR-675376

Lawrence Livermore National Laboratory: Completing the Human Genome Project and Triggering Nearly $1 Trillion in U.S. Economic Activity

Jeffrey Stewart

The success of the Human Genome project is already nearing $1 Trillion dollars of U.S. economic activity.[1] Lawrence Livermore National Laboratory (LLNL) was a co-leader in one of the biggest biological research effort in history, sequencing the Human Genome Project. This ambitious research effort set out to sequence the approximately 3 billion nucleotides in the human genome, an effort many thought was nearly impossible.

Deoxyribonucleic acid (DNA) was discovered in 1869, and by 1943 came the discovery that DNA was a molecule that encodes the genetic instructions used in the development and functioning of living organisms and many viruses. To make full use of the information, scientists needed to first sequence the billions of nucleotides to begin linking them to genetic traits and illnesses, and eventually more effective treatments. New medical discoveries and improved agriculture productivity were some of the expected benefits. While the potential benefits were vast, the timeline (over a decade) and cost ($3.8 Billion) exceeded what the private sector would normally attempt, especially when this would only be the first phase toward the path to new discoveries and market opportunities. The Department of Energy believed its best research laboratories could meet this Grand Challenge and soon convinced the National Institute of Health to formally propose the Human Genome project to the federal government. The U.S. government accepted the risk and challenge to potentially create new healthcare and food discoveries that could benefit the world and the U.S. Industry.

[1] (Tripp & Gruber, 2011)

Why LLNL Co-Led one of the Biggest Biological Research Efforts in History

How did LLNL get involved in DNA sequencing? LLNL began working on biomedical research as far back as the 1950s. Studying the effects of radiation on humans led to research on how genes were affected by radiation. LLNL formally began its biological research program in 1963. Interest in DNA mapping began at LLNL in 1985. Work was also being conducted at Los Alamos National Laboratory (LANL) and Lawrence Berkeley National Laboratory (LBNL). During this period LLNL developed world class expertise in researching cells and DNA leading to important contributions to cell research, some listed in the table below.

1963 LLNL begins a biological research program to investigate the effects of ionizing radiation on humans
1968 LLNL was the first to use a computer to create three-dimensional images of organelles, tiny working parts within the cell nucleus
1973 LLNL Cytophotometric Data Conversion System proved it could measure the DNA in individual chromosomes to great sensitivity
1974 LLNL made history when it successfully measured and sorted hamster chromosomes
1984 LLNL and LANL begin human chromosome-specific DNA libraries
1980s Fingerprinting Chemistry- automated florescence based strategy for fingerprinting each clone
1980s Algorithm Development- Software to process acquired fluorophore signals and convert the signal data to restriction fragment lengths for each cosmid
1980s Map construction- Creating the early stages of a map with the completed work of over 2500 cosmids.

These achievements gave LLNL recognition as one of the leading research institutions with both the experience and talent to play a major role in sequencing the human genome.

A blue-ribbon scientific advisory board was assembled by the Health and Environmental Research Advisory Committee (HERAC) to review the goal and research plans of the Human Genome project that

was put together by LLNL and its partner laboratories LANL and LBNL. The advisory board began their report with the following statement;

> *"It may seem audacious to ask DOE to spearhead such a biological revolution, but scientist of many persuasions on the subcommittee and on HERAC agree that DOE alone has the background, structure, and style necessary to coordinate this enormous, highly technical task. When done properly, the effort will be interagency and international in scope; but it must have strong central control, a base akin to the National Laboratories, and flexible ways to access a huge array of university and industrial partners. We believe this can and should be done, and that DOE is the one to do it."*

LLNL's recognition grew midway through the project when in 1996 LLNL scientist Dr. Elbert Branscomb was named director of the Joint Genome Institute.

Today there are many LLNL technologies that are still in use today including popular programs such as;

zPictures- Compares sequences of two species

Mulan- Compares sequences of multiple species

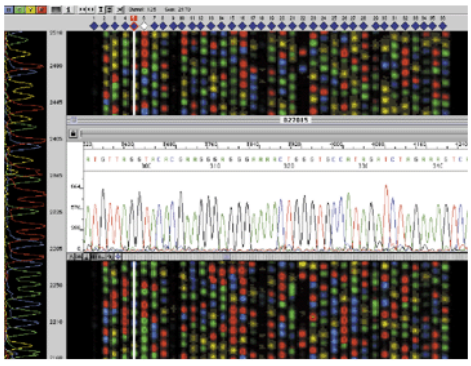
ECR Browser- Compares complete genomes of many species

The Sequencing Process

The basic unit scientists worked with is a single DNA nucleotide. Nucleotides consist of sugars and a phosphate group. Together these are called Bases. DNA bases can be one of four kinds and are attached to these sugars. These bases are Adenine (A), Thymine (T), Cytosine (C) and Guanine (G).

The Human Genome Project used a process called “Sanger Sequencing” that involved;

1) Taking DNA and shearing it into overlapping pieces and making clones comprised of 2,000 to 2 million base pairs.
2) Reactions are run; an enzyme copies the input DNA to produce fragments that differ from one another by one base pair. Four fluorescent dyes are added. Each dye is attracted to only one base.
3) An electric current is applied to the labeled fragments. The DNA itself has a negative charge; the many fragments migrate at different rates depending on their size with smaller ones moving faster than larger ones.
4) Once the fragments reach a certain point a scanning laser beam excites the dyes on the DNA fragments to indicate which base is on the end of each fragment with the smallest ones arriving first.
5) Then another group of labeled DNA fragments, one base group longer than the previous group, passes the laser scanner, indicating the identity of that terminal base.



The four colors in this chromatogram represent the four bases that make up our DNA: green is adenine (A), blue is cytosine (C), yellow is guanine (G), and red is thymine (T). Each fragment of DNA differs from the next fragment by one base, and the dye indicates the terminal base of each fragment. The order of the colors indicates the order of the bases and hence the sequence of the DNA.

6) The fluorescent signals are captured and digitized, resulting in a four color chromatogram showing peaks that represent each of the four DNA bases. The order of the colors is the order of the bases and the DNA is sequenced.

This technique proved to be the most efficient way to sequence the genome. At the same time it was described as the equivalent of taking multiple copies of a puzzle, cutting each puzzle into different sized individual pieces and assembling the pieces into one complete puzzle. The puzzle you are assembling is also invisible to the naked eye.

Industry, seeing a potentially large market on the horizon once the Human Genome project was launched began developing faster and cheaper technologies for DNA sequencing. These technologies began to enter the market soon after the successful completion of the Human Genome project in 2003. The pace of technological advancement has exceeded Moore's law in reducing the cost of sequencing (see figure 1). Sequencing technology improvements have lowered the cost an estimated $100 million in 2001 to below $10,000 by 2012.

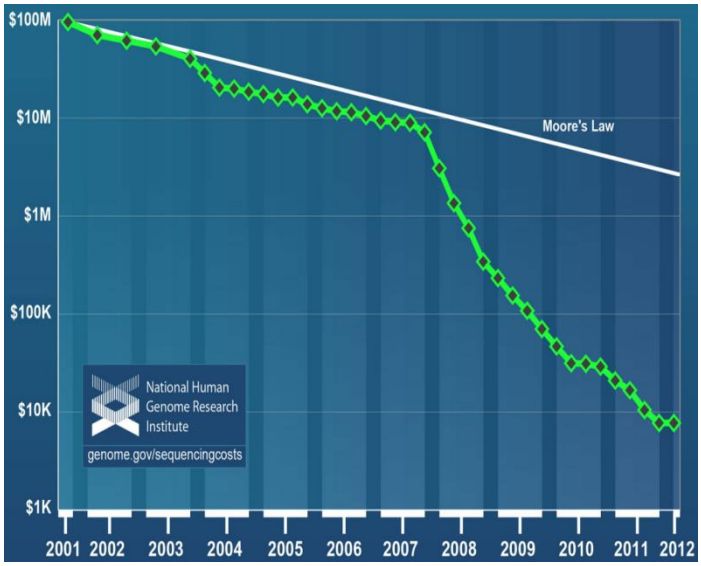


Figure 1 Cost of Sequencing Human Genome over time

Value to the U.S. Economy

The Human Genome program costs a total of $3.8 billion over 13 years program which ended in 2003. In 2010 dollars that translates to $5.6 billion. By 2010 the Human Genome project is estimated to have created over $798 Billion dollars in U.S. economic activity, a return on investment of $141 to $1 for every dollar the U.S. government invested in this program. [2]

The sectors of the U.S. economy that have seen the most expansion range from biomedical research, veterinary medicine, agriculture and even our legal systems. The most notable areas are listed in the table below.

Genetics and Genomics Tools, Technologies, Techniques and Services	Expanding Basic Scientific knowledge	Human Health and Medicine	Veterinary Medicine	Agriculture and Food	Industrial Biotechnology	Environmental Applications	Forensics, Justice and Security

Combined these areas are credited with using the success of the Human Genome project to create;

- *3.8 million job years between 1988-2010;*
- *$796 billion in U.S. economic activity from 1988-2010*
- *2010 alone- 310,000 jobs, $20 Billion in personal income, $67 billion in economic activity*

At the beginning of the Human Genome Project the cost to map each base pair was approximately $10. By the time the Human Genome Project was completed, the cost was down to under $0.10 to map a base pair. This improvement in technology has enabled many researchers and commercial entities to

[2] (Tripp & Gruber, 2011)

affordably find new scientific discoveries and create new commercial endeavors. In the article "Beyond the Genome: New Uses of DNA Sequencers", it states;

> *"Once upon a time sequencing the human genome took tens of millions of dollars and a warehouse full of DNA sequencing machines that analyzed samples throughout the day, and year after year. Now, less than a decade later, the same human genome sequence- the order of nucleotides or "letter" can be generated using a single machine that analyzes samples for a few days, and for about a 100-fold lower cost. The ability to sequence DNA faster and more cheaply comes from recent technological advancements representing the most significant technological metamorphosis in the history of modern genetics"*[3]

Mark Skousen, editor of Hedge Fund Tracker, follows the DNA markets and believes that the price and speed of sequencing will drop to a point in which they are common place in hospitals and doctor's offices. [4]

We have compiled a list of the top five companies with a hand in the DNA instrument and sequencing market and their market cap as of 2011.

1. Illumina (Nasdaq: ILMN): Biotechnology Industry. Market cap of $6.32B.

2. Life Technologies (Nasdaq: LIFE) : Biotechnology Industry. Market cap of $7.04B.

3. Pacific Biosciences of California (Nasdaq: PACB) : Biotechnology Industry. Market cap of $316.25M.

[3] (Eisenstadt, 2010)
[4] (Kapitall, 2011)

4. Fluidigm (Nasdaq: FLDM) : Scientific & Technical Instruments Industry. Market cap of $268.87M.

5. Complete Genomics (Nasdaq: GNOM) : Biotechnology Industry. Market cap of $276.25M.

LLNL's technology transfer and people have also directly contributed to the technology boom that was born out of the Human Genome project. These include two former LLNL employees, Dr. Bill Colston, co-founder of Quantalife, Inc and Dr Allan Northrup, co-founder of Cepheid. LLNL has a long history of helping U.S. economic security through Tech Transfer.

References

Bauer, S. (2014). The FBI Is Very Excited About This Machine That Can Scan Your DNA in 90 Minutes. *Mother Jones*.

Branscomb, E. (2000). The Joint Genome Insitute: Decoding the Human Genome. *Science and Technology Review* .

Cinkosky, M., & Johnson, R. P. (1989, october 12). A Natural Language Query System for Genebank. *Letter*. Livermore, ca, USA: LLNL.

Cook-Deegan, R. (n.d.). *The Gene Wars: Science, Politics, and the Human Genome.*

Eisenstadt, L. (2010). Beyond the Genome: New Uses of DNA Sequencers.

Energy, U. D. (1989, November 3-4). Human Genome Workshop- Contractor/Grantee. *Internal Document*. Sante Fe, New Mexico, USA: DOE.

Heller, A. (2005). Mining Genomes. *Science and Technology*.

Kapitall, A. C. (2011, August 19). DNA Sequencing: The Next High-Growth market in Biotech.

Nanoscale technologies to cut DNA sequencing costs. (2012, September 26). *Kurzweil Accelerating Intelligence*.

Pray, L. (2008). Discovery of DNA Structure and Function: Watson and Crick. *Scitable*.

Science, U. D. (2008). *Joint Genome Insitute 2008 Progress Report.* DOE.

Slezak, T. (2015, May-July). Overview of Human Genome History at LLNL. (J. Stewart, Interviewer)

Staden, R. (1979). A Strategy of DNA Sequencing Employing Computer Programs. *Nucleic Acids Research*, 2601-2610.

Tripp, S., & Gruber, M. (2011). *Economic Impact of the Human Genome Project.* Battelle Memorial Institute.

Walter, K. (2002). Biological Research Evolves at Livermore. *Science and Technology Review*.