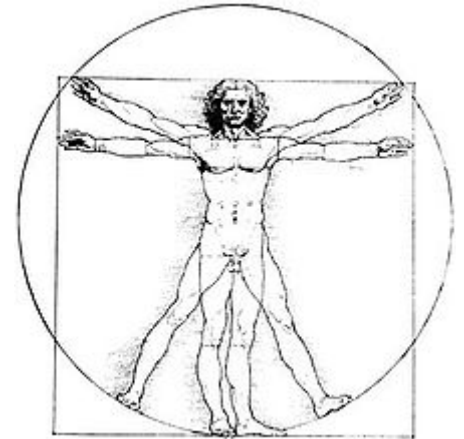


Human Genome Project

The **Human Genome Project (HGP)** was an international scientific research project with the goal of determining the sequence of nucleotide base pairs that make up human DNA, and of identifying and mapping all of the genes of the human genome from both a physical and a functional standpoint.^[1] It remains the world's largest collaborative biological project.^[2] After the idea was picked up in 1984 by the US government when the planning started, the project formally launched in 1990 and was declared complete in 2003^[3]. Funding came from the US government through the National Institutes of Health (NIH) as well as numerous other groups from around the world. A parallel project was conducted outside government by the Celera Corporation, or Celera Genomics, which was formally launched in 1998. Most of the government-sponsored sequencing was performed in twenty universities and research centers in the United States, the United Kingdom, Japan, France, Germany, Spain and China^[4]



Logo HGP; Vitruvian Man, Leonardo da Vinci

The Human Genome Project originally aimed to map the nucleotides contained in a human haploid reference genome (more than three billion). The "genome" of any given individual is unique; mapping the "human genome" involved sequencing a small number of individuals and then assembling these together to get a complete sequence for each chromosome. Therefore, the finished human genome is a mosaic, not representing any one individual.

Contents

Human Genome Project

- History

- State of completion

Applications and proposed benefits

Techniques and analysis

- Findings

- Accomplishment

Public versus private approaches

Genome donors

Developments

Ethical, legal and social issues

See also

References

Further reading

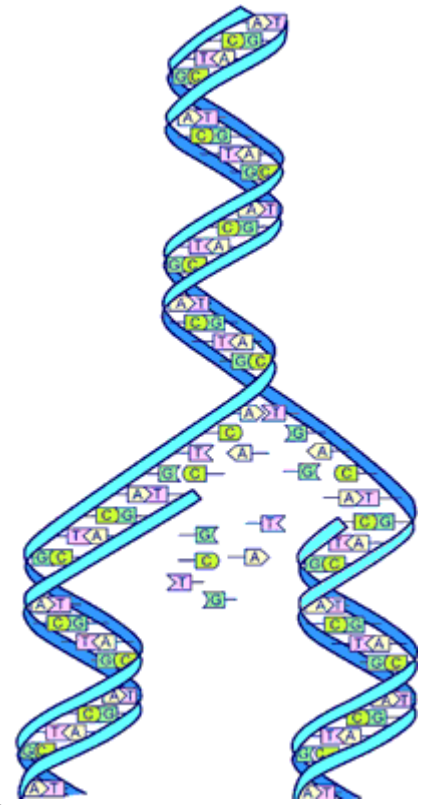
External links

Human Genome Project

History

The Human Genome Project was a 15-year-long, publicly funded project initiated in 1990 with the objective of determining the DNA sequence of the entire euchromatic human genome within 15 years.^[5] In May 1985, Robert Sinsheimer organized a workshop to discuss sequencing the human genome,^[6] but for a number of reasons the NIH was uninterested in pursuing the proposal. The following March, the Santa Fe Workshop was organized by Charles DeLisi and David Smith of the Department of Energy's Office of Health and Environmental Research (OHER).^[7] At the same time Renato Dulbecco proposed whole genome sequencing in an essay in *Science*.^[8] James Watson followed two months later with a workshop held at the Cold Spring Harbor Laboratory

The fact that the Santa Fe workshop was motivated and supported by a Federal Agency opened a path, albeit a difficult and tortuous one,^[9] for converting the idea into a public policy in the United States. In a memo to the Assistant Secretary for Energy Research (Alvin Trivelpiece), Charles DeLisi, who was then Director of the OHER, outlined a broad plan for the project.^[10] This started a long and complex chain of events which led to approved reprogramming of funds that enabled the OHER to launch the Project in 1986, and to recommend the first line item for the HGP, which was in President Reagan's 1988 budget submission,^[9] and ultimately approved by the Congress. Of particular importance in Congressional approval was the advocacy of Senator Peter Domenici, whom DeLisi had befriended.^[11] Domenici chaired the Senate Committee on Energy and Natural Resources, as well as the Budget Committee, both of which were key in the DOE budget process. Congress added a comparable amount to the NIH budget, thereby beginning official funding by both agencies.



Alvin Trivelpiece sought and obtained the approval of DeLisi's proposal by Deputy Secretary William Flynn Martin. This chart^[12] was used in the spring of 1986 by Trivelpiece, then Director of the Office of Energy Research in the Department of Energy, to brief Martin and Under Secretary Joseph Salgado regarding his intention to reprogram \$4 million to initiate the project with the approval of Secretary Herrington. This reprogramming was followed by a line item budget of \$16 million in the Reagan Administration's 1987 budget submission to Congress.^[13] It subsequently passed both Houses. The Project was planned for 15 years.^[14]

Candidate technologies were already being considered for the proposed undertaking at least as early as 1985.^[15]

In 1990, the two major funding agencies, DOE and NIH, developed a memorandum of understanding in order to coordinate plans and set the clock for the initiation of the Project to 1990.^[16] At that time, David Galas was Director of the renamed "Office of Biological and Environmental Research" in the U.S. Department of Energy's Office of Science and James Watson headed the NIH Genome Program. In 1993, Aristides Patrinos succeeded Galas and Francis Collins succeeded James Watson, assuming the role of overall Project Head as Director of the U.S. National Institutes of Health (NIH) National Center for Human Genome Research (which would later become the National Human Genome Research Institute). A working draft of the genome was announced in 2000 and the papers describing it were published in February 2001. A more complete draft was published in 2003, and genome "finishing" work continued for more than a decade.

The \$3-billion project was formally founded in 1990 by the US Department of Energy and the National Institutes of Health, and was expected to take 15 years.^[17] In addition to the United States, the international consortium comprised geneticists in the United Kingdom, France, Australia, China and myriad other spontaneous relationships.^[18]

Due to widespread international cooperation and advances in the field of genomics (especially in sequence analysis), as well as major advances in computing technology, a 'rough draft' of the genome was finished in 2000 (announced jointly by U.S. President Bill Clinton and the British Prime Minister Tony Blair on June 26, 2000).^[19] This first available rough draft assembly of the genome was completed by the Genome Bioinformatics Group at the University of California, Santa Cruz, primarily led by then graduate student Jim Kent. Ongoing sequencing led to the announcement of the essentially complete genome on April 14, 2003, two years earlier than planned.^{[20][21]} In May 2006, another milestone was passed on the way to completion of the project, when the sequence of the last chromosome was published in Nature.^[22]

State of completion

The project was not able to sequence all the DNA found in human cells. It sequenced only "euchromatic" regions of the genome, which make up 92% of the human genome. The other regions, called "heterochromatic" are found in centromeres and telomeres, and were not sequenced under the project.^[23]

The Human Genome Project was declared complete in April 2003. An initial rough draft of the human genome was available in June 2000 and by February 2001 a working draft had been completed and published followed by the final sequencing mapping of the human genome on April 14, 2003. Although this was reported to cover 99% of the euchromatic human genome with 99.99% accuracy, a major quality assessment of the human genome sequence was published on May 27, 2004 indicating over 92% of sampling exceeded 99.99% accuracy which was within the intended goal.^[24] Further analyses and papers on the HGP continue to occur.^[25]

Applications and proposed benefits

The sequencing of the human genome holds benefits for many fields, from molecular medicine to human evolution. The Human Genome Project, through its sequencing of the DNA, can help us understand diseases including: genotyping of specific viruses to direct appropriate treatment; identification of mutations linked to different forms of cancer; the design of medication and more accurate prediction of their effects; advancement in forensic applied sciences; biofuels and other energy applications; agriculture, animal husbandry, bioprocessing; risk assessment; bioarcheology, anthropology and evolution. Another proposed benefit is the commercial development of genomics research related to DNA based products, a multibillion-dollar industry

The sequence of the DNA is stored in databases available to anyone on the Internet. The U.S. National Center for Biotechnology Information (and sister organizations in Europe and Japan) house the gene sequence in a database known as GenBank, along with sequences of known and hypothetical genes and proteins. Other organizations, such as the UCSC Genome Browser at the University of California, Santa Cruz,^[26] and Ensembl^[27] present additional data and annotation and powerful tools for visualizing and searching it. Computer programs have been developed to analyze the data, because the data itself is difficult to interpret without such programs. Generally speaking, advances in genome sequencing technology have followed Moore's Law, a concept from computer science which states that integrated circuits can increase in complexity at an exponential rate.^[28] This means that the speeds at which whole genomes can be sequenced can increase at a similar rate, as was seen during the development of the above-mentioned Human Genome Project.

Techniques and analysis

The process of identifying the boundaries between genes and other features in a raw DNA sequence is called genome annotation and is in the domain of bioinformatics. While expert biologists make the best annotators, their work proceeds slowly, and computer programs are increasingly used to meet the high-throughput demands of genome sequencing projects. Beginning in 2008, a new technology known as RNA-seq was introduced that allowed scientists to directly sequence the messenger RNA in cells. This replaced previous methods of annotation, which relied on inherent properties of the DNA sequence, with direct measurement, which was much more accurate. Today, annotation of the human genome and other genomes relies primarily on deep sequencing of the transcripts in every human tissue using RNA-seq. These experiments have revealed that over 90% of genes contain at least one and usually several alternative splice variants, in which the exons are combined in different ways to produce 2 or more gene products from the same locus.

The genome published by the HGP does not represent the sequence of every individual's genome. It is the combined mosaic of a small number of anonymous donors, all of European origin. The HGP genome is a scaffold for future work in identifying differences among individuals. Subsequent projects sequenced the genomes of multiple distinct ethnic groups, though as of today there is still only one "reference genome."

Findings

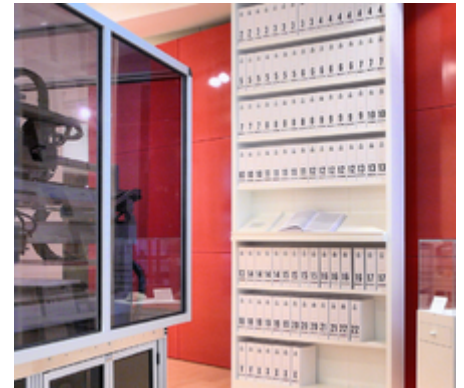
Key findings of the draft (2001) and complete (2004) genome sequences include:

1. There are approximately 22,300^[29] protein-coding genes in human beings, the same range as in other mammals.
2. The human genome has significantly more segmental duplications (nearly identical, repeated sections of DNA) than had been previously suspected.^{[30][31][32]}
3. At the time when the draft sequence was published fewer than 7% of protein families appeared to be vertebrate specific.^[33]

Accomplishment

The Human Genome Project was started in 1990 with the goal of sequencing and identifying all three billion chemical units in the human genetic instruction set, finding the genetic roots of disease and then developing treatments. It is considered a megaproject because the human genome has approximately 3.3 billion base-pairs. With the sequence in hand, the next step was to identify the genetic variants that increase the risk for common diseases like cancer and diabetes.^{[16][34]}

It was far too expensive at that time to think of sequencing patients' whole genomes. So the National Institutes of Health embraced the idea for a "shortcut", which was to look just at sites on the genome where many people have a variant DNA unit. The theory behind the shortcut was that, since the major diseases are common, so too would be the genetic variants that caused them. Natural selection keeps the human genome free of variants that damage health before children are grown, the theory held, but fails against variants that strike later in life, allowing them to become quite common. (In 2002 the National Institutes of Health started a \$138 million project called the HapMap to catalog the common variants in European, East Asian and African genomes.)^[35]



The first printout of the human genome to be presented as a series of books, displayed at the Wellcome Collection, London

The genome was broken into smaller pieces; approximately 150,000 base pairs in length.^[34] These pieces were then ligated into a type of vector known as "bacterial artificial chromosomes", or BACs, which are derived from bacterial chromosomes which have been genetically engineered. The vectors containing the genes can be inserted into bacteria where they are copied by the bacterial DNA replication machinery. Each of these pieces was then sequenced separately as a small "shotgun" project and then assembled. The larger, 150,000 base pairs go together to create chromosomes. This is known as the "hierarchical shotgun" approach, because the genome is first broken into relatively large chunks, which are then mapped to chromosomes before being selected for sequencing.^{[36][37]}

Funding came from the US government through the National Institutes of Health in the United States, and a UK charity organization, the Wellcome Trust, as well as numerous other groups from around the world. The funding supported a number of large sequencing centers including those at Whitehead Institute, the Wellcome Sanger Institute (then called The Sanger Centre) based at the Wellcome Genome Campus, Washington University in St. Louis and Baylor College of Medicine.^{[17][38]}

The United Nations Educational, Scientific and Cultural Organization (UNESCO) served as an important channel for the involvement of developing countries in the Human Genome Project.^[39]

Public versus private approaches

In 1998, a similar, privately funded quest was launched by the American researcher Craig Venter, and his firm Celera Genomics. Venter was a scientist at the NIH during the early 1990s when the project was initiated. The \$300,000,000 Celera effort was intended to proceed at a faster pace and at a fraction of the cost of the roughly \$3 billion publicly funded project. The Celera approach was able to proceed at a much more rapid rate, and at a lower cost than the public project because it relied upon data made available by the publicly funded project.^[40]

Celera used a technique called whole genome shotgun sequencing, employing pairwise end sequencing^[41] which had been used to sequence bacterial genomes of up to six million base pairs in length, but not for anything nearly as large as the three billion base pair human genome.

Celera initially announced that it would seek patent protection on "only 200–300" genes, but later amended this to seeking "intellectual property protection" on "fully-characterized important structures" amounting to 100–300 targets. The firm eventually filed preliminary ("place-holder") patent applications on 6,500 whole or partial genes. Celera also promised to publish their findings in accordance with the terms of the 1996 Bermuda Statement, by releasing new data annually (the HGP released its new data daily), although, unlike the publicly funded project, they would not permit free redistribution or scientific use of the data. The publicly funded competitors were compelled to release the first draft of the human genome before Celera for this reason. On July 7, 2000, the UCSC Genome Bioinformatics Group released a first working draft on the web. The scientific community downloaded about 500 GB of information from the UCSC genome server in the first 24 hours of free and unrestricted access.^[42]

In March 2000, President Clinton announced that the genome sequence could not be patented, and should be made freely available to all researchers. The statement sent Celera's stock plummeting and dragged down the biotechnology-heavy Nasdaq. The biotechnology sector lost about \$50 billion in market capitalization in two days.

Although the working draft was announced in June 2000, it was not until February 2001 that Celera and the HGP scientists published details of their drafts. Special issues of Nature (which published the publicly funded project's scientific paper)^[43] and Science (which published Celera's paper^[44]) described the methods used to produce the draft sequence and offered analysis of the sequence. These drafts covered about 83% of the genome (90% of the euchromatic regions with 150,000 gaps and the order and orientation of many segments not yet established). In February 2001, at the time of the joint publications, press releases announced that the project had been completed by both groups. Improved drafts were announced in 2003 and 2005, filling in to approximately 92% of the sequence currently.

Genome donors

In the IHGSC international public-sector HGP, researchers collected blood (female) or sperm (male) samples from a large number of donors. Only a few of many collected samples were processed as DNA resources. Thus the donor identities were protected so neither donors nor scientists could know whose DNA was sequenced. DNA clones from many different libraries were used in the overall project, with most of those libraries being created by Pieter J. de Jong's. Much of the sequence (>70%) of the reference genome produced by the public HGP came from a single anonymous male donor from Buffalo, New York (code name RP11).^{[45][46]}

HGP scientists used white blood cells from the blood of two male and two female donors (randomly selected from 20 of each) – each donor yielding a separate DNA library. One of these libraries (RP11) was used considerably more than others, due to quality considerations. One minor technical issue is that male samples contain just over half as much DNA from the sex chromosomes (one X chromosome and one Y chromosome) compared to female samples (which contain two X chromosomes). The other 22 chromosomes (the autosomes) are the same for both sexes.

Although the main sequencing phase of the HGP has been completed, studies of DNA variation continue in the International HapMap Project, whose goal is to identify patterns of single-nucleotide polymorphism (SNP) groups (called haplotypes, or "haps"). The DNA samples for the HapMap came from a total of 270 individuals: Yoruba people in Ibadan, Nigeria; Japanese people in Tokyo; Han Chinese in Beijing; and the French Centre d'Etude du Polymorphisme Humain (CEPH) resource, which consisted of residents of the United States having ancestry from Western and Northern Europe.

In the Celera Genomics private-sector project, DNA from five different individuals were used for sequencing. The lead scientist of Celera Genomics at that time, Craig Venter, later acknowledged (in a public letter to the journal Science) that his DNA was one of 21 samples in the pool, five of which were selected for use.^{[47][48]}

In 2007, a team led by Jonathan Rothberg published James Watson's entire genome, unveiling the six-billion-nucleotide genome of a single individual for the first time.^[49]

Developments

The work on interpretation and analysis of genome data is still in its initial stages. It is anticipated that detailed knowledge of the human genome will provide new avenues for advances in medicine and biotechnology. Clear practical results of the project emerged even before the work was finished. For example, a number of companies, such as Myriad Genetics, started offering easy ways to administer genetic tests that can show predisposition to a variety of illnesses, including breast cancer, hemostasis disorders, cystic fibrosis, liver diseases and many others. Also, the etiologies for cancers, Alzheimer's disease and other areas of clinical interest are considered likely to benefit from genome information and possibly may lead in the long term to significant advances in their management.^{[35][50]}

There are also many tangible benefits for biologists. For example, a researcher investigating a certain form of cancer may have narrowed down their search to a particular gene. By visiting the human genome database on the World Wide Web, this researcher can examine what other scientists have written about this gene, including (potentially) the three-dimensional structure of its product, its function(s), its evolutionary relationships to other human genes, or to genes in mice or yeast or fruit flies, possible detrimental mutations, interactions with other genes, body tissues in which this gene is activated, and diseases associated with this gene or other datatypes. Further, deeper understanding of the disease processes at the level of molecular biology may determine new therapeutic procedures. Given the established importance of DNA in molecular biology and its central role in determining the fundamental operation of cellular processes, it is likely that expanded knowledge in this area will facilitate medical advances in numerous areas of clinical interest that may not have been possible without them.^[51]

The analysis of similarities between DNA sequences from different organisms is also opening new avenues in the study of evolution. In many cases, evolutionary questions can now be framed in terms of molecular biology; indeed, many major evolutionary milestones (the emergence of the ribosome and organelles, the development of embryos with body plans, the vertebrate immune system) can be related to the molecular level. Many questions about the similarities and differences between humans and our closest relatives (the primates, and indeed the other mammals) are expected to be illuminated by the data in this project.^{[35][52]}

The project inspired and paved the way for genomic work in other fields, such as agriculture. For example, by studying the genetic composition of *Triticum aestivum*, the world's most commonly used bread wheat, great insight has been gained into the ways that domestication has impacted the evolution of the plant.^[53] Which loci are most susceptible to manipulation, and how does this play out in evolutionary terms? Genetic sequencing has allowed these questions to be addressed for the first time, as specific loci can be compared in wild and domesticated strains of the plant. This will allow for advances in genetic modification in the future which could yield healthier, more disease-resistant wheat crops.

Ethical, legal and social issues

At the onset of the Human Genome Project several ethical, legal, and social concerns were raised in regards to how increased knowledge of the human genome could be used to discriminate against people. One of the main concerns of most individuals was the fear that both employers and health insurance companies would refuse to hire individuals or refuse to provide insurance to people because of a health concern indicated by someone's genes.^[54] In 1996 the United States passed the Health Insurance Portability and Accountability Act (HIPAA) which protects against the unauthorized and non-consensual release of individually identifiable health information to any entity not actively engaged in the provision of healthcare services to a patient.^[55]

Along with identifying all of the approximately 20,000–25,000 genes in the human genome, the Human Genome Project also sought to address the ethical, legal, and social issues that were created by the onset of the project. For that the Ethical, Legal, and Social Implications (ELSI) program was founded in 1990. Five percent of the annual budget was allocated to address the ELSI arising from the project.^{[17][56]} This budget started at approximately \$1.57 million in the year 1990, but increased to approximately \$18 million in the year 2014.^[57]

Whilst the project may offer significant benefits to medicine and scientific research, some authors have emphasized the need to address the potential social consequences of mapping the human genome. "Molecularising disease and their possible cure will have a profound impact on what patients expect from medical help and the new generation of doctors' perception of illness."^[58]

See also

- [1000 Genomes Project](#)
- [100,000 Genomes Project](#)
- [Chimpanzee Genome Project](#)
- [ENCODE](#)
- [EuroPhysiome](#)
- [Genome Compiler](#)
- [HUGO Gene Nomenclature Committee](#)
- [Human Brain Project](#)
- [Human Connectome Project](#)
- [Human Cytome Project](#)
- [Human Microbiome Project](#)
- [Human proteome project](#)
- [Human Variome Project](#)
- [List of biological databases](#)
- [Neanderthal Genome Project](#)
- [Sanger Institute](#)
- [The Genographic Project](#)

References

1. Robert Krulwich (2001-04-17). *Cracking the Code of Life*(<https://www.pbs.org/wgbh/nova/genome/>)(Television Show). [PBS](#).
2. "Economic Impact of the Human Genome Project – Battelle"http://battelle.org/docs/default-document-library/economic_impact_of_the_human_genome_project.pdf(PDF). Retrieved 1 August 2013.
3. <https://www.genome.gov/11006943/human-genome-project-completion-frequently-asked-questions>(<https://www.genome.gov/11006943/human-genome-project-completion-frequently-asked-questions/>)Missing or empty |title= (help)
4. "Human Genome Project Completion: Frequently Asked Questions"<http://www.genome.gov/11006943>
genome.gov.
5. "Human Genome Project: Sequencing the Human Genome | Learn Science at Scitable"<http://www.nature.com/scitable/topicpage/dna-sequencing-technologies-key-to-the-human-828>*www.nature.com*. Retrieved 2016-01-25.
6. Sinsheimer, Robert (1989). "The Santa Cruz Workshop, May 1985". *Genomics*. **5**: 954–956. doi:10.1016/0888-7543(89)90142-0(<https://doi.org/10.1016%2F0888-7543%2889%2990142-0>)
7. DeLisi, Charles (October 2008). "Conferences That Changed the World". *Nature*. **455**: 876–877. Bibcode:2008Natur.455..876D (<http://adsabs.harvard.edu/abs/2008Natur.455..876D>). doi:10.1038/455876a(<https://doi.org/10.1038%2F455876a>) PMID 18923499 (<https://www.ncbi.nlm.nih.gov/pubmed/18923499>).
8. Dulbecco, Renato (1986). "Turning Point in Cancer Research, Sequencing the Human Genome"*Science*. **231** (4742): 1055–1056. Bibcode:1986Sci...231.1055D(<http://adsabs.harvard.edu/abs/1986Sci...231.1055D>) doi:10.1126/science.3945817(<https://doi.org/10.1126%2Fscience.3945817>)PMID 3945817 (<https://www.ncbi.nlm.nih.gov/pubmed/3945817>)
9. Gene Wars, Op.Cit. p. 102.
10. "Search" (https://repository.library.georgetown.edu/handle/10822/556935/discover?filtertype=author&filter_relational_operator>equals&filter=DeLisi%2C+Charles)*georgetown.edu*
11. "President Clinton Awards the Presidential Citizens Medals" (http://clinton5.nara.gov/WH/new/html/Mon_Jan_8_141714_2001.html) *nara.gov*.
12. <http://www.wpainc.com/Archive/Trivelpiece/HGP%20Presenation.jpg>
13. DeLisi, Charles (2008). "Meetings that changed the world: Santa Fe 1986: Human genome baby-steps"*Nature*. **455** (7215): 876–877. Bibcode:2008Natur.455..876D (<http://adsabs.harvard.edu/abs/2008Natur.455..876D>). doi:10.1038/455876a(<https://doi.org/10.1038%2F455876a>) PMID 18923499 (<https://www.ncbi.nlm.nih.gov/pubmed/18923499>).

14. DeLisi, Charles (1988). "The Human Genome Project"*American Scientist* **76**: 488. [Bibcode:1988AmSci..76..488D](http://adsabs.harvard.edu/abs/1988AmSci..76..488D) (<http://adsabs.harvard.edu/abs/1988AmSci..76..488D>)
15. DeLisi, Charles (2001). "Genomes: 15 Years Later A Perspective by Charles Deli, HEP Pioneer" (<https://web.archive.org/web/20040904183906/http://genome.gsc.riken.go.jp/hgmis/publicat/hgn/v11n3/05delisi.html>) *Human Genome News*. **11**: 3–4. Archived from the original (<http://genome.gsc.riken.go.jp/hgmis/publicat/hgn/v11n3/05delisi.html>) on September 4, 2004 Retrieved 2005-02-03.
16. "About the Human Genome Project: What is the Human Genome Project" (http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml) The Human Genome Management Information System (HGMIS). 2011-07-18 Retrieved 2011-09-02
17. Human Genome Information Archive: "About the Human Genome Project" (http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml) U.S. Department of Energy & Human Genome Project program Retrieved 1 August 2013.
18. Collins F; Galas D (1993-10-01). "A New Five-Year Plan for the United States: Human Genome Program" (<http://www.genome.gov/10001476>) National Human Genome Research Institute Retrieved 1 August 2013.
19. "White House Press Release" (http://www.ornl.gov/sci/techresources/Human_Genome/project/clinton1.shtml) Retrieved 2006-07-22.
20. Noble, Ivan (2003-04-14). "Human genome finally complete" (<http://news.bbc.co.uk/1/hi/sci/tech/2940601.stm>) *BBC News*. Retrieved 2006-07-22.
21. Kolata, Gina (15 April 2013). "Human Genome, Then and Now" (<https://www.nytimes.com/2013/04/16/science/the-human-genome-project-then-and-now.html>). *The New York Times*. Retrieved 24 April 2014.
22. "Guardian Unlimited | UK Latest | Human Genome Project finalised" (<https://www.theguardian.com/uklatest/story/0,,-5829253,00.html>) *The Guardian*. London. Retrieved 2006-07-22.
23. "The Human Genome Project FAQ" (<http://www.genoscope.cns.fr/spip/The-Human-Genome-Project.html?artsuite=1#FAQ2>). *Genoscope*. Centre National de Séquençage Retrieved 12 February 2015.
24. Schmutz, Jeremy; Wheeler, Jeremy; Grimwood, Jane; Dickson, Mark; Yang, Joan; Caoile, Chenier; Bajorek, Eva; Black, Stacey; Chan, Yee Man; Denys, Mirian Escobar, Julio; Flowers, Dave; Fotopulos, Dea; Garcia, Carmen; Gomez, Maria; Gonzales, Eideilyn; Haydu, Lauren; Lopez, Frederick; Ramirez, Lucia; Retter, James; Rodriguez, Alex; Rogers, Stephanie; Salazar Angelica; Tsai, Ming; Myers, Richard M. (2004). "Quality assessment of the human genome sequence" (<http://www.nature.com/nature/journal/v429/n6990/abs/nature02390.html>) *Nature*. **429** (6990): 365–368. [Bibcode:2004Natur.429..365S](http://adsabs.harvard.edu/abs/2004Natur.429..365S) (<http://adsabs.harvard.edu/abs/2004Natur.429..365S>). [doi:10.1038/nature02390](https://doi.org/10.1038/nature02390) (<https://doi.org/10.1038/nature02390>) PMID [15164052](https://pubmed.ncbi.nlm.nih.gov/15164052/) (<https://pubmed.ncbi.nlm.nih.gov/15164052/>)
25. "Landmark Human Genome Project Papers" (http://web.ornl.gov/sci/techresources/Human_Genome/project/journals.shtml). *ornl.gov*.
26. "An Overview of the Human Genome Project" (<http://genome.ucsc.edu>)
27. "Ensembl Genome Browser" (<http://www.ensembl.org>) *ensembl.org*.
28. Mardis, E. (2008). "The impact of next-generation sequencing technology on genetics" *Trends in Genetics* **24** (3): 133–41. [doi:10.1016/j.tig.2007.12.007](https://doi.org/10.1016/j.tig.2007.12.007) (<https://doi.org/10.1016/j.tig.2007.12.007>) PMID [18262675](https://pubmed.ncbi.nlm.nih.gov/18262675/) (<https://pubmed.ncbi.nlm.nih.gov/18262675/>)
29. Mihaela Pertea & Steven Salzberg (2010). "Between a chicken and a grape: estimating the number of human genes" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2898077>). *Genome Biology*. **11**: 206. [doi:10.1186/gb-2010-11-5-206](https://doi.org/10.1186/gb-2010-11-5-206) (<https://doi.org/10.1186/gb-2010-11-5-206>) PMC [2898077](https://pubmed.ncbi.nlm.nih.gov/20441615/) (<https://pubmed.ncbi.nlm.nih.gov/20441615/>) PMID [20441615](https://pubmed.ncbi.nlm.nih.gov/20441615/) (<https://pubmed.ncbi.nlm.nih.gov/20441615/>)
30. Venter, JC; et al. (2001). "The sequence of the human genome" *Science*. **291**: 1304–1351. [Bibcode:2001Sci...291.1304V](http://adsabs.harvard.edu/abs/2001Sci...291.1304V) (<http://adsabs.harvard.edu/abs/2001Sci...291.1304V>) [doi:10.1126/science.1058040](https://doi.org/10.1126/science.1058040) (<https://doi.org/10.1126/science.1058040>) PMID [11181995](https://pubmed.ncbi.nlm.nih.gov/11181995/) (<https://pubmed.ncbi.nlm.nih.gov/11181995/>).
31. International Human Genome Sequencing Consortium (IHGSC) (2004). "Finishing the euchromatic sequence of the human genome" (<http://www.nature.com/nature/journal/v431/n7011/full/nature03001.html>) *Nature*. **431** (7011): 931–945. [Bibcode:2004Natur.431..931H](http://adsabs.harvard.edu/abs/2004Natur.431..931H) (<http://adsabs.harvard.edu/abs/2004Natur.431..931H>). [doi:10.1038/nature03001](https://doi.org/10.1038/nature03001) (<https://doi.org/10.1038/nature03001>) PMID [15496913](https://pubmed.ncbi.nlm.nih.gov/15496913/) (<https://pubmed.ncbi.nlm.nih.gov/15496913/>).
32. Spencer, Geoff (20 December 2004). "International Human Genome Sequencing Consortium Describes Finished Human Genome Sequence" (<http://www.genome.gov/12513430>) *NIH News Release*. National Institutes of Health.

33. Bryant, J. A (2007). *Design and information in biology: From molecules to systems* (<https://books.google.com/?id=yzFFAQAIAAJ>). p. 108. ISBN 9781853128530 "...brought to light about 1200 protein families. Only 94 protein families, or 7%, appear to be vertebrate specific"
34. Wellcome Sanger Institute. "The Human Genome Project: a new reality" (<https://web.archive.org/web/20130801231124/http://www.sanger.ac.uk/about/history/hgp/>) Wellcome Trust Sanger Institute, Genome Research Limited. Archived from the original (<http://www.sanger.ac.uk/about/history/hgp/>) on 2013-08-01 Retrieved 1 August 2013.
35. Naidoo N; Pawitan Y, Soong R; Cooper DN; Ku CS (2011). "Human genetics and genomics a decade after the release of the draft sequence of the human genome" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3525251>). *Hum Genomics*. **5** (6): 577–622. doi:10.1186/1479-7364-5-6-577 (<https://doi.org/10.1186%2F1479-7364-5-6-577>) PMC 3525251 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3525251>) PMID 22155605 (<https://www.ncbi.nlm.nih.gov/pubmed/22155605>)
36. "Celera: A Unique Approach to Genome Sequencing" (<http://www.ocf.berkeley.edu/~edy/genome/celera.html>) *ocf.berkeley.edu*. Biocomputing. 2006 Retrieved 1 August 2013.
37. Davidson College (2002). "Sequencing Whole Genomes: Hierarchical Shotgun Sequencing. Shotgun Sequencing" (<http://www.bio.davidson.edu/courses/genomics/method/shotgun.html>) *bio.davidson.edu*. Department of Biology Davidson College Retrieved 1 August 2013.
38. Human Genome Project Information Archive (2013). "U.S. & International HGP Research Sites" (http://web.ornl.gov/sci/techresources/Human_Genome/research/centers.shtml) U.S. Department of Energy & Human Genome Project Retrieved 1 August 2013.
39. Vizzini, Casimiro (March 19, 2015). "The Human Variome Project: Global Coordination in Data Sharing" (<http://www.sciencediplomacy.org/article/2015/human-varome-project>). *Science & Diplomacy*. **4** (1).
40. Venter, J. C.; Adams, M. D.; Myers, E. W; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A. (2001-02-16). "The sequence of the human genome" *Science*. **291** (5507): 1304–1351. Bibcode:2001Sci...291.1304V (<http://adsabs.harvard.edu/abs/2001Sci...291.1304V>) doi:10.1126/science.1058040 (<https://doi.org/10.1126%2Fscience.1058040>) ISSN 0036-8075 (<https://www.worldcat.org/issn/0036-8075>) PMID 11181995 (<https://www.ncbi.nlm.nih.gov/pubmed/11181995>).
41. Roach JC; Boysen C; Wang K; Hood L (1995). "Pairwise end sequencing: a unified approach to genomic mapping and sequencing" (<http://linkinghub.elsevier.com/retrieve/pii/088875439580219C>) *Genomics*. **26** (2): 345–353. doi:10.1016/0888-7543(95)80219-C (<https://doi.org/10.1016%2F0888-7543%2895%2980219-C>) PMID 7601461 (<https://www.ncbi.nlm.nih.gov/pubmed/7601461>)
42. Center for Biomolecular Science & Engineering "The Human Genome Project Race" (http://www.cbse.ucsc.edu/research/hgp_race) Center for Biomolecular Science and Engineering Retrieved 1 August 2013.
43. International Human Genome Sequencing Consortium (2001). "Initial sequencing and analysis of the human genome" (<http://www.nature.com/nature/journal/v409/n6822/pdf/409860a0.pdf>) (PDF). *Nature*. **409** (6822): 860–921. Bibcode:2001Natur.409..860L (<http://adsabs.harvard.edu/abs/2001Natur.409..860L>). doi:10.1038/35057062 (<https://doi.org/10.1038%2F35057062>) PMID 11237011 (<https://www.ncbi.nlm.nih.gov/pubmed/11237011>).
44. Venter, JC; et al. (2001). "The sequence of the human genome" (<http://www.sciencemag.org/cgi/reprint/291/507/1304.pdf>) (PDF). *Science*. **291** (5507): 1304–1351. Bibcode:2001Sci...291.1304V (<http://adsabs.harvard.edu/abs/2001Sci...291.1304V>) doi:10.1126/science.1058040 (<https://doi.org/10.1126%2Fscience.1058040>) PMID 11181995 (<https://www.ncbi.nlm.nih.gov/pubmed/11181995>)
45. Osoegawa, Kazutoyo; Mammosser AG; Wu, C; Frengen, E; Zeng, C; Catanese, JJ; De Jong, PJ (2001) "A Bacterial Artificial Chromosome Library for Sequencing the Complete Human Genome" (<http://www.genome.org/cgi/content/full/11/3/483>). *Genome Research* **11** (3): 483–96. doi:10.1101/gr.169601 (<https://doi.org/10.1101%2Fgr169601>). PMC 311044 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC311044>) PMID 11230172 (<https://www.ncbi.nlm.nih.gov/pubmed/11230172>)
46. Tuzun, E; et al. (2005). "Fine-scale structural variation of the human genome" (<http://www.nature.com/ng/journal/v37/n7/full/ng1562.html>) *Nature Genetics*. **37** (7): 727–737. doi:10.1038/ng1562 (<https://doi.org/10.1038%2Fng1562>) PMID 15895083 (<https://www.ncbi.nlm.nih.gov/pubmed/15895083>).
47. Kennedy D (2002). "Not wicked, perhaps, but tacky" *Science*. **297** (5585): 1237. doi:10.1126/science.297.5585.1237 (<https://doi.org/10.1126%2Fscience.297.5585.1237>) PMID 12193755 (<https://www.ncbi.nlm.nih.gov/pubmed/12193755>)
48. Venter D (2003). "A Part of the Human Genome Sequence". *Science*. **299** (5610): 1183–4. doi:10.1126/science.299.5610.1183 (<https://doi.org/10.1126%2Fscience.299.5610.1183>) PMID 12595674 (<https://www.ncbi.nlm.nih.gov/pubmed/12595674>)

49. Wadman, Meredith (2008-04-16). "James Watson's genome sequenced at highspeed" (<http://www.nature.com/news/2008/080416/full/452788b.html>) *Nature News*. **452** (7189): 788–788. Bibcode:2008Natur.452R...W (<http://adsabs.harvard.edu/abs/2008Natur452R...W>). doi:10.1038/452788b (<https://doi.org/10.1038%2F452788b>)
50. Gonzaga-Jauregui C; Lupski JR; Gibbs RA (2012). "Human genome sequencing in health and disease" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3656720>). *Annu Rev Med*. **63** (1): 35–61. doi:10.1146/annurev-med-051010-162644 (<https://doi.org/10.1146%2Fannurev-med-051010-162644>) PMC 3656720 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3656720>) PMID 22248320 (<https://www.ncbi.nlm.nih.gov/pubmed/22248320>).
51. Snyder M, Du J; Gerstein M (2012). "Personal genome sequencing: current approaches and challenges" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2827837>). *Genes Dev*. **24** (5): 423–431. doi:10.1101/gad.1864110 (<https://doi.org/10.1101%2Fgad.1864110>) PMC 2827837 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2827837>) PMID 20194435 (<https://www.ncbi.nlm.nih.gov/pubmed/20194435>).
52. Lander ES (2011). "Initial impact of the sequencing of the human genome" *Nature*. **470** (7333): 187–197. Bibcode:2011Natur.470..187L (<http://adsabs.harvard.edu/abs/2011Natur.470..187L>). doi:10.1038/nature09792 (<https://doi.org/10.1038%2Fnature09792>) PMID 21307931 (<https://www.ncbi.nlm.nih.gov/pubmed/21307931>).
53. Peng, J; Sun, E; Nevo, D (2011). "Domestication Evolution, Genetics And Genomics In Wheat" *Molecular Breeding* **28** (3): 281–301. doi:10.1007/s11032-011-9608-4 (<https://doi.org/10.1007%2Fs11032-011-9608-4>)
54. Greely, Henry (1992). *The Code of Codes: Scientific and Social Issues in the Human Genome Project* Cambridge, Massachusetts: Harvard University Press. pp. 264–65 ISBN 0-674-13646-2
55. US Department of Health and Human Services: "Understanding Health Information Privacy" (<https://www.hhs.gov/oc/privacy/hipaa/understanding/index.html>)
56. Genetics Home Reference (2013). "What were some of the ethical, legal, and social implications addressed by the Human Genome Project?" (<http://ghr.nlm.nih.gov/handbook/hgp/elsi>) *ghr.nlm.nih.gov*. Retrieved 1 August 2013.
57. "ELSI Research Program Fact Sheet - National Human Genome Research Institute (NHGRI)" (<https://www.genome.gov/10002329/elsi-research-program-fact-sheet>) *www.genome.gov*. Retrieved 2016-09-27.
58. Rheinberger, H.J. (2000). *Living and Working with the New Medical Technologies*. Cambridge: Cambridge University Press. p. 20.

Further reading

- McElheny, Victor K. (2010). *Drawing the Map of Life: Inside the Human Genome Project* Basic Books. ISBN 978-0-465-03260-0. 361 pages. Examines the intellectual origins, history and motivations of the project to map the human genome; draws on interviews with key figures.
- Collins, F. (2006). *The Language of God: A Scientist Presents Evidence for Belief* Free Press. ISBN 0-7432-8639-1 OCLC 65978711.
- Venter, J. Craig (October 18, 2007). *A Life Decoded: My Genome: My Life* New York, New York: Viking Adult. ISBN 0-670-06358-4. OCLC 165048736.
- Cook-Deegan, Robert (1994). *The Gene Wars: Science, Politics, and the Human Genome*. New York: W W Norton.

External links

- **Human Genome Project**— official information page
- National Human Genome Research Institute (NHGRI) NHGRI led the National Institutes of Health's contribution to the International Human Genome Project. This project, which had as its primary goal the sequencing of the three thousand million base pairs that make up human genome, was successfully completed in April 2003.
- Human Genome News Published from 1989 to 2002 by the US Department of Energy this newsletter was a major communications method for coordination of the Human Genome Project. Complete online archives are available.
- The HGP information pages Department of Energy's portal to the international Human Genome Project, Microbial Genome Program, and Genomics:GTL systems biology for energy and environment
- yourgenome.org: The Sanger Institute public information page has general and detailed primers on DNA, genes and genomes, the Human Genome Project and science spotlights.
- Ensembl project, an automated annotation system and browser for the human genome
- UCSC genome browser, This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides a portal to the ENCODE project.
- Nature magazine's human genome gateway including the HGP's paper on the draft genome sequence

- [Wellcome Trust Human Genome website](#)A free resource allowing you to explore the human genome, your health and your future.
- [Learning about the Human Genome. Part 1: Challenge to Science Educators. ERIC Digest.](#)
- [Learning about the Human Genome. Part 2: Resources for Science Educators. ERIC Digest.](#)
- [Patenting Life](#) by Merrill Goozner
- [Prepared Statement of Craig Venter of Celera](#) Venter discusses Celera's progress in deciphering the human genome sequence and its relationship to healthcare and to the federally funded Human Genome Project.
- [Cracking the Code of Life](#) Companion website to 2-hour NOVA program documenting the race to decode the genome, including the entire program hosted in 16 parts in either [QuickTime](#) or [RealPlayer](#) format.
- Lone Dog L (1999). "Whose genes are they? The Human Genome Diversity Project" *J Health Soc Policy*. **10** (4): 51–66. doi:10.1300/J045v10n04_04 PMID 10538186.
- [Bioethics Research Library](#) Numerous original documents at Georgetown University

Works by archive

- [Works by Human Genome Project at Project Gutenberg](#)
 - [Project Gutenberg](#) hosts e-texts for Human Genome Project, titled *Human Genome Project, Chromosome Number #* (# denotes 01-22, X and Y). This information is raw sequence, released in November 2002; access to entry pages with download links is available through <https://www.gutenberg.org/etext/3501> for Chromosome 1 sequentially to <https://www.gutenberg.org/etext/3524> for the Y Chromosome. Note that this sequence might not be considered definitive due to ongoing revisions and refinements. In addition to the chromosome files, there is a [supplementary information file](#) dated March 2004 which contains additional sequence information.
- [Works by or about Human Genome Project at Internet Archive](#)

Retrieved from 'https://en.wikipedia.org/w/index.php?title=Human_Genome_Project&oldid=831473840

This page was last edited on 20 March 2018, at 19:44.

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.