Science

A History of the Huma

1953

▶ (April) **James Watson** and **Francis Crick** discover the double helical structure of DNA (*Nature*).

A. BARRINGTON BROWN/SCIENCE PHOTO LIBRARY

1972

(October) **Paul Berg** and co-workers create the first recombinant DNA molecule (*PNAS*).

1977

▼ **Allan Maxam** and **Walter Gilbert** (below) at Harvard University and **Frederick Sanger** at the U.K. Medical Research Council (MRC) independently develop methods for sequencing DNA (*PNAS*, February; *PNAS*, December).

JANE REED/HARVARD U.

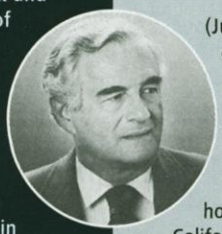
1980

(May) **David Botstein** of the Massachusetts Institute of Technology, **Ronald Davis** of Stanford University, and **Mark Skolnick** and **Ray White** of the University of Utah propose a method to map the entire human genome based on RFLPs (*American Journal of Human Genetics*).

1982

▼ **Akiyoshi Wada**, now at RIKEN in Japan, proposes automated sequencing and gets support to build robots with help from Hitachi.

DENNIS NORMILE



VICTOR MCKUSICK

1984

(May) **Charles Cantor** and **David Schwartz** of Columbia University develop pulsed field electrophoresis (*Cell*).

(July) MRC scientists decipher the complete DNA sequence of the Epstein-Barr virus, 170 kb (*Nature*).

1985

◀ (May) **Robert Sinsheimer** hosts a meeting at the University of California (UC), Santa Cruz, to discuss the feasibility of sequencing the human genome.

(December) **Kary Mullis** and colleagues at Cetus Corp. develop PCR, a technique to replicate vast amounts of DNA (*Science*).

COLD SPRING HARBOR LAB

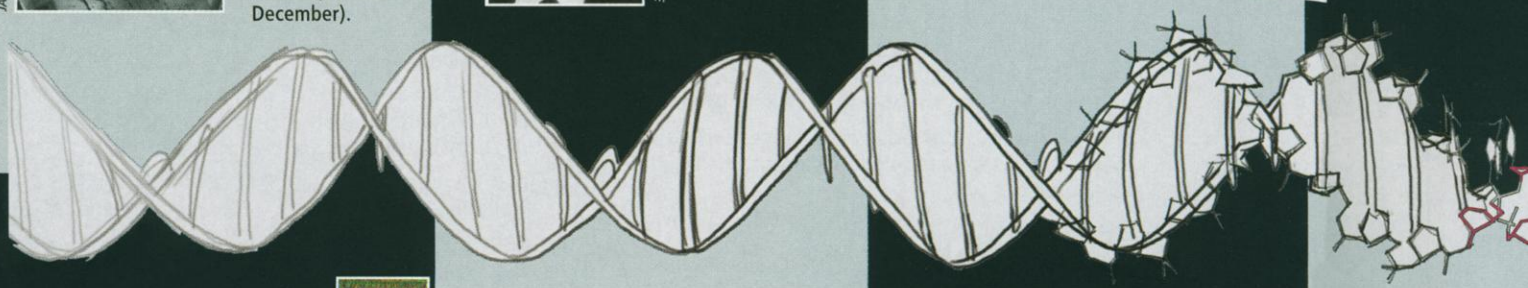
1986

(February) **Sydney Bre** the European Union to take a concerted progr map and sequence the genome; **Brenner** also small genome initiative

(March) The U.S. **Depa** Energy (**DOE**) hosts a m Santa Fe New Mexico, cuss plans to sequence human genome.

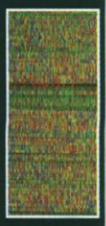
(March) **Renato Dulbe** Institute promot human genome i

◀ (June) Merits o project are hotly d ing at Cold Spring in New York state, Biology of *Homo sa*



1990 cont'd

(October) **David Lipman**, **Eugene Myers**, and colleagues at the National Center for Biotechnology Information (NCBI) publish the BLAST algorithm for aligning sequences (*Journal of Molecular Biology*).



SANGER CENTRE/WELLCOME PHOTO LIBRARY

1991

(June) NIH biologist **J. Craig Venter** announces a strategy to find expressed genes, using ESTs (*Science*). A fight erupts at a congressional hearing 1 month later, when **Venter** reveals that NIH is filing patent applications on thousands of these partial genes.

DAVID SIERENVISUALS UNLIMITED

◀ (October) The Japanese rice genome sequencing effort begins.

(December) **Edward Uberbacher** of Oak Ridge National Laboratory in Tennessee develops GRAIL, the first of many gene-finding programs (*PNAS*).

1992

(April) After a dispute with then–NIH director **Bernadine Healy** over patenting partial genes, **Watson** resigns as head of NCHGR.

(June) **Venter** leaves NIH to set up The Institute for Genomic Research (TIGR), a nonprofit in Rockville, Maryland. **William Haseltine** heads its sister company, Human Genome Sciences, to commercialize TIGR products.

(July) Britain's Wellcome Trust enters the HGP with $95 million.

(September) **Mel Simon** of Caltech and colleagues develop BACs for cloning (*PNAS*).

▶ (October) U.S. and French teams complete the first physical maps of chromosomes: **David Page** of the Whitehead Institute and colleagues (above) map the Y chromosome (*Science*); **Daniel Cohen** of the Centre d'Etude du Polymorphisme Humain (CEPH) and Généthon and colleagues map chromosome 21 (*Nature*).

SAM OGDEN

(December) After lengthy debate, NIH and DOE release guidelines on sharing data and resources, encouraging rapid sharing and enabling researchers to keep data private for 6 months.

U.S. and French teams complete genetic maps of mouse and human: mouse, average marker spacing 4.3 cM, **Eric Lander** and colleagues at Whitehead (*Genetics*, June); human, average marker spacing 5 cM, **Jean Weissenbach** and colleagues at CEPH (*Nature*, October).

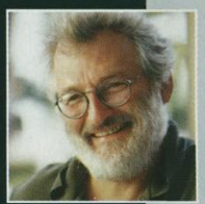
1993

(April) **Francis Collins** of the University of Michigan is named director of NCHGR.

(October) NIH and DOE publish a revised plan for 1993–98. The goals include sequencing 80 Mb of DNA by the end of 1998 and completing the human genome by 2005.

WELLCOME TRUST PHOTO LIBRARY

▶ (October) The Wellcome Trust and MRC open the Sanger Centre at Hinxton Hall, south of Cambridge, U.K. Led by **John Sulston**, the center becomes one of the major sequencing labs in the international consortium.



(October) The GenBar moves from Los Alam NIH's and DOE's tussl

1994

(September) **Jeffrey M** of Iowa, **Cohen** of Gén publish a complete ge the human genome, w spacing of 0.7 cM (*Scie*

1995

(May to August) **Rich** colleagues at UC Ber develop improved se (*PNAS*, May); **Michae Fuller** at Amersham mostable polymerase

▶ (July) **Venter** and **Fraser** of TIGR and **H Smith** of Johns Hopk lish the first sequenc free-living organ *Haemophilus infl* 1.8 Mb (*Science*)

(September) The ment funds seve groups for a tot over 5 years: To University of To University.

uman Genome Project

ydney Brenner of MRC urges
n Union to under-
erted program to
quence the human
enner also starts a
e initiative at MRC.

U.S. Department of
) hosts a meeting in
ew Mexico, to dis-
o sequence the
ome.

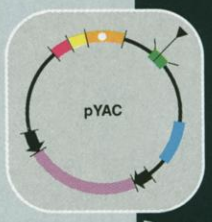
ato Dulbecco of the Salk
te promotes sequencing the
genome in a paper (*Science*).

e) Merits of a human genome
are hotly debated at a meet-
old Spring Harbor Laboratory
York state, "The Molecular
of *Homo sapiens*."

ROBERT FAZ/CALTECH

▼ (June) **Leroy Hood** (below) and **Lloyd Smith** of the California Institute of Technology (Caltech) and colleagues announce the first automated DNA sequencing machine (*Nature*).

(September) **Charles DeLisi** begins genome studies at DOE, reallocating $5.3 million from the fiscal year 1987 budget.



BURKE ET AL.

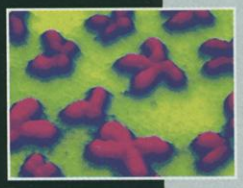
1987

(February) **Walter Gilbert** resigns from the U.S. National Research Council (NRC) genome panel and announces plans to start Genome Corp., with the goal of sequencing and copyrighting the human genome and selling data for profit.

(April) An advisory panel suggests that DOE should spend $1 billion on mapping and sequencing the human genome over the next 7 years—and that DOE should lead the U.S. effort. DOE's Human Genome Initiative begins.

▼ (May) **David Burke, Maynard Olson,** and **George Carle** of Washington University in St. Louis develop YACs for cloning, increasing insert size 10-fold (*Science*).

(October) **Helen Donis-Keller** and colleagues at Collaborative Research Inc. publish the "first" genetic map with 403 markers, sparking a fight over credit and priority (*Cell*).



T.J. MCMASTER/WELLCOME PHOTO LIBRARY

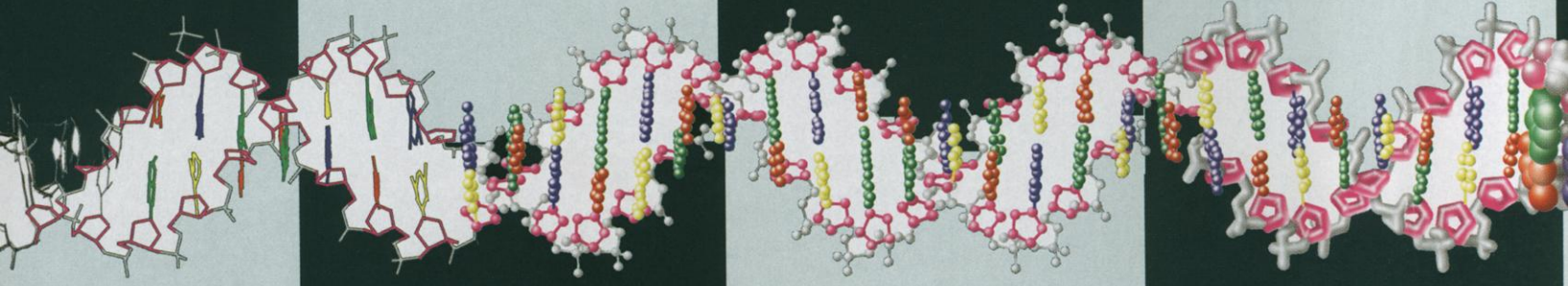
(October) DuPont scientists develop a system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides (*Science*).

Applied Biosystems Inc. puts the first automated sequencing machine, based on **Hood**'s technology, on the market.

1988

(February) In a pivotal report, the NRC endorses the Human Genome Project (HGP), calling for a phased approach and a rapid scale-up to $200 million a year of new money.

(March) Prompted by advisers at a meeting in Reston, Virginia, **James Wyngaarden**, then director of the National Institutes of Health (NIH), decides that the agency should be a major player in the HGP, effectively seizing the lead from DOE.



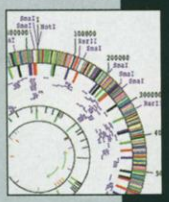
The GenBank database officially
n Los Alamos to NCBI, ending
DOE's tussle over control.

') **Jeffrey Murray** of the University
hen of Généthon, and colleagues
omplete genetic linkage map of
genome, with an average marker
0.7 cM (*Science*).

ugust) **Richard Mathies** and
at UC Berkeley and Amersham
proved sequencing dyes
y); **Michael Reeve** and **Carl**
Amersham develop ther-
polymerase (*Nature*, August).

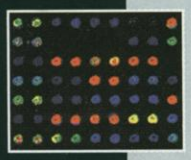
enter and **Claire**
TIGR and **Hamilton**
ohns Hopkins pub-
st sequence of a
ving organism,
ophilus influenzae,
b (*Science*).

ember) The Japanese govern-
funds several sequencing
s for a total of $15.9 million
5 years: Tokai University,
rsity of Tokyo, and Keio



THE INSTITUTE FOR GENOMIC RESEARCH

▶ (October) **Patrick Brown** of Stanford and colleagues publish first paper using a printed glass microarray of complementary DNA (cDNA) probes (*Science*).



SCHENA ET AL.

▼ (December) Researchers at Whitehead and Généthon (led by **Lander** and **Thomas Hudson** at Whitehead) publish a physical map of the human genome containing 15,000 markers (*Science*).

1996

(February) At a meeting in Bermuda funded by the Wellcome Trust, international HGP partners agree to release sequence data into public databases within 24 hours.

(April) NIH funds six groups to attempt large-scale sequencing of the human genome.

(April) Affymetrix makes DNA chips commercially available.

(September) DOE initiates six pilot projects, funded at $5 million total, to sequence the ends of BAC clones.

(October) An international consortium publicly releases the complete genome sequence of the yeast *S. cerevisiae* (*Science*).

(November) **Yoshihide Hayashizaki**'s group at RIKEN completes the first set of full-length mouse cDNAs.

1997

(January) NCHGR is promoted to the National Human Genome Research Institute; DOE creates the Joint Genome Institute.

BLATTNER ET AL.

▼ (September) **Fred Blattner, Guy Plunkett,** and University of Wisconsin, Madison, colleagues complete the DNA sequence of *E. coli*, 5 Mb (*Science*).

(September) Molecular Dynamics introduces the MegaBACE, a capillary sequencing machine.

1998

(January) NIH announces a new project to find SNPs.

(February) Representatives of Japan, the U.S., the E.U., China, and South Korea meet in Tsukuba, Japan, to establish guidelines for an international collaboration to sequence the rice genome.

JOANNA GREEN

◀ (March) **Phil Green** (left) and **Brent Ewing** of Washington University and colleagues publish a program called phred for automatically interpreting sequencer data (*Genetic Research*). Both phred and its sister program phrap (used for assembling sequences) had been in wide use since 1995.

(May) PE Biosystems Inc. introduces the PE Prism 3700 capillary sequencing machine.

(May) **Venter** announces a new company named Celera and declares that it will sequence the human genome within 3 years for $300 million.

(May) In response, the Wellcome Trust doubles its support for the HGP to $330 million, taking on responsibility for one-third of the sequencing.

nt scientists develop a
d DNA sequencing with
in-terminating dideoxynu-
ıce).

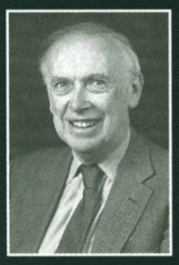
ems Inc. puts the first auto-
ıg machine, based on
ogy, on the market.

pivotal report, the NRC
uman Genome Project
or a phased approach and
p to $200 million a year of

n) Prompted by advisers at a
ng in Reston, Virginia, **James**
aarden, then director of the
al Institutes of Health (NIH),
s that the agency should be
r player in the HGP, effec-
seizing the lead from DOE.

(June) The first annual genome meeting is held at Cold Spring Harbor Laboratory.

▶ (September) NIH establishes the Office of Human Genome Research and snags **Watson** as its head. Watson declares that 3% of the genome budget should be devoted to studies of social and ethical issues.



MARGO BENNETT/COLD SPRING HARBOR LAB

(October) NIH and DOE sign a memorandum of understanding and agree to collaborate on the HGP.

1989

(January) **Norton Zinder** of Rockefeller University chairs the first program advisory committee meeting for the HGP.

(September) **Olson**, **Hood**, **Botstein**, and **Cantor** outline a new mapping strategy, using STSs (*Science*).

(September) DOE and NIH start a joint committee on the ethical, legal, and social implications of the HGP.

(October) NIH office is elevated to the National Center for Human Genome Research (NCHGR), with grant-awarding authority.

1990

▶ Three groups develop capillary electrophoresis, one team led by **Lloyd Smith** (*Nucleic Acids Research*, August), the second by **Barry Karger** (*Analytical Chemistry*, January), and the third by **Norman Dovichi** (*Journal of Chromatography*, September).

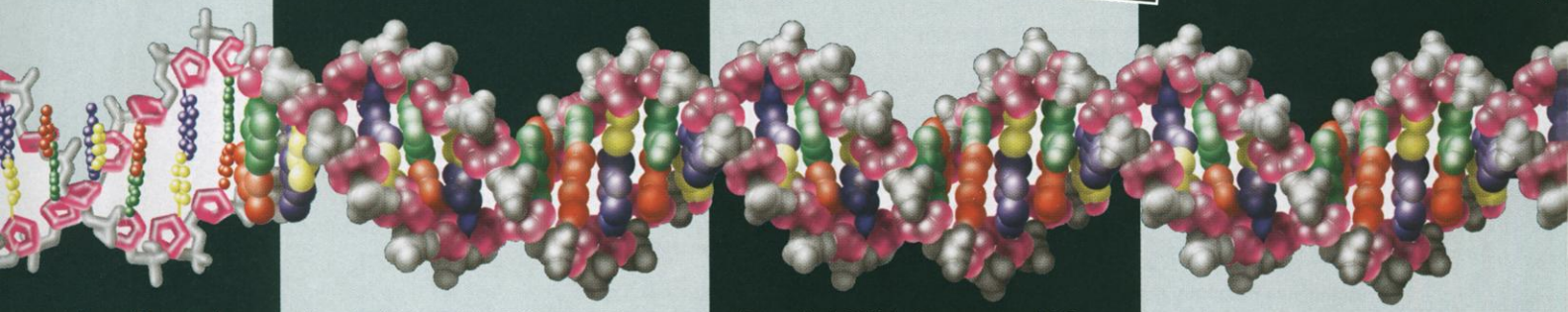
JAMES KING-HOLMES/SCIENCE PHOTO LIBRARY

(April) NIH and DOE publish a 5-year plan. Goals include a complete genetic map, a physical map with markers every 100 kb, and sequencing of an aggregate of 20 Mb of DNA in model organisms by 2005.

(E. COLI) K.G. MURTI/VISUALS UNLIMITED; (C. ELEGANS) JAMES HOLMES KING/VISUALS UNLIMITED; (YEAST) PHOTO RESEARCHERS, INC.

▲ (August) NIH begins large-scale sequencing trials on four model organisms: *Mycoplasma capricolum*, *Escherichia coli*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae*. Each research group agrees to sequence 3 Mb at 75 cents a base within 3 years.

(October) NIH and DOE restart the clock, declaring 1 October the official beginning of the HGP.



presentatives of Japan, the
China, and South Korea
ba, Japan, to establish
an international collabora-
nce the rice genome.

◀ (March) **Phil Green** (left) and **Brent Ewing** of Washington University and colleagues publish a program called phred for automatically interpreting sequencer data (*Genetic Research*). Both phred and its sister program phrap embling sequences) had been ince 1995.

systems Inc. introduces the
0 capillary sequencing

r announces a new com-
Celera and declares that
ce the human genome
s for $300 million.

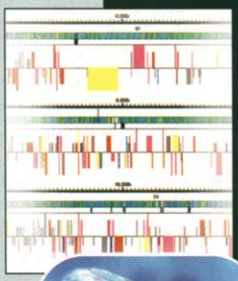
onse, the Wellcome
its support for the HGP to
taking on responsibility for
the sequencing.

(October) NIH and DOE throw HGP into overdrive with a new goal of creating a "working draft" of the human genome by 2001, and they move the completion date for the finished draft from 2005 to 2003.

(December) **Sulston** of the Sanger Centre and **Robert Waterston** of Washington University and colleagues complete the genomic sequence of *C. elegans* (*Science*).

1999

(March) NIH again moves up the completion date for the rough draft, to spring 2000. Large-scale sequencing efforts are concentrated in centers at Whitehead, Washington University, Baylor, Sanger, and DOE's Joint Genome Institute.



(SEQUENCE) ADAMS ET AL.; (FLY) DWIGHT KUHN PHOTOGRAPHY

(April) Ten companies and the Wellcome Trust launch the SNP consortium, with plans to publicly release data quarterly.

(September) NIH launches a project to sequence the mouse genome, devoting $130 million over 3 years.

(December) British, Japanese, and U.S. researchers complete the first sequence of a human chromosome, number 22 (*Nature*).

2000

◀ (March) Celera and academic collaborators sequence the 180-Mb genome of the fruit fly *Drosophila melanogaster*, the largest genome yet sequenced and a validation of **Venter**'s controversial whole-genome shotgun method (*Science*).

(March) Because of disagreement over a data-release policy, plans for HGP and Celera to collaborate disintegrate amid considerable sniping.

(May) HGP consortium led by German and Japanese researchers publishes the complete sequence of chromosome 21 (*Nature*).

(June) At a White House ceremony, HGP and Celera jointly announce working drafts of the human genome sequence, declare their feud at an end, and promise simultaneous publication.

▶ (October) DOE and MRC launch a collaborative project to sequence the genome of the puffer fish, *Fugu rubripes*, by March 2001.

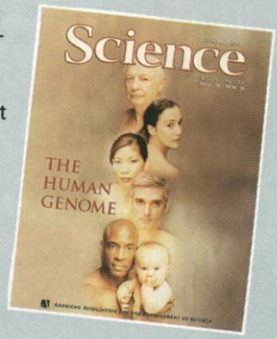
JOSHUA MYLNE; DOUG WECHSLER

◀ (December) An international consortium completes the sequencing of the first plant, *Arabidopsis thaliana*, 125 Mb.

(December) HGP and Celera's plans for joint publication in *Science* collapse; HGP sends its paper to *Nature*.

2001

▶ (February) The HGP consortium publishes its working draft in *Nature* (15 February), and Celera publishes its draft in *Science* (16 February).



The Human Genome Project ... In their own words

From the outset, the proposal to map and sequence the human genome has sparked controversy and evoked strong emotions. The following quotes capture how the debate has shifted over the years.

EARLY DEBATES

"It endangers all of us, especially the young researchers." David Botstein, *Science*, 27 June 1986

"The idea is gathering momentum. I shiver at the thought." David Baltimore, *Science*, 27 June 1986

"The idea of trudging through the genome sequence by sequence does not command wide and enthusiastic support in the U.K." Sydney Brenner, *Science*, 8 August 1986

"The total human sequence is the grail of human genetics." Walter Gilbert, *Science*, 27 June 1986

"It is clearly no longer a question of whether the project ought to be done, but of how fast it will be done." Russell Doolittle, *Science*, 13 February 1987

"I'm surprised consenting adults have been caught in public talking about it [sequencing the genome]. ... It makes no sense." Robert Weinberg, *New Scientist*, 5 March 1987

"The sequence of the human genome would be perhaps the most powerful tool ever developed to explore the mysteries of human development and disease." Leroy Hood, *Issues in Science and Technology*, Spring 1987

WALTER GILBERT DECLARES HE WILL COPYRIGHT AND SELL DNA DATA

"The idea of the company is to be a service to the biotech and pharmaceutical industries and to the research community. ... [The sequence data] would be made available to everyone—for a price." Walter Gilbert, *Science*, 24 July 1987

"This information is so important that it cannot be proprietary." C. Thomas Caskey, *Science*, 24 July 1987

"If a company behaves in what scientists believe is a socially responsible manner, they can't make a profit." Robert Cook-Deegan, *Science*, 24 July 1987

THE PUBLICATION OF THE "FIRST" GENETIC MAP

"What they have accomplished is important. ... But it is not what we believe should be properly called a map. ... We would never have dreamed of making such a publication with our data set, which is substantially larger than theirs, because we still have significant gaps." Ray White, *Science*, 6 November 1987

"A map is a map. Our map has holes, we make no bones about it. ... It is not Ray White's ideal, but so what?" Helen Donis-Keller, *Science*, 6 November 1987

"It's a real shame that the only two groups in the world who are doing this haven't communicated and shared probes." Leroy Hood, *Science*, 6 November 1987

SUPPORT BUILDS

"You can't be against getting this information; it is too fundamental." Charles Cantor, *Science*, 12 February 1988

"The argument against DOE is that while they talk about peer review, it is not clear that they do it. ... [About NIH,] you can't have a lead agency that doesn't want to do it." Bruce Alberts, *Science*, 12 February 1988

PATENT SKIRMISHES

"I am horrified." James Watson, *Science*, 11 October 1991, on NIH's plans to patent J. Craig Venter's partial genes

"There is no coherent government policy [on gene patents] and we need one—quick—since the sequence is just pouring out. It would be a big mistake to leave this one to the lawyers." David Galas, *Science*, 11 October 1991

VENTER ANNOUNCES CELERA

"It strikes me that this is a cream-skimming approach. It's clearly an attempt to short-circuit the hard problems and defer them to the [research] community at a very substantial cost." Robert Waterston, *Science*, 15 May 1998

"I think it's great." David Cox, *Science*, 15 May 1998

"Every time we talk, we move [the deadline] up." Robert Waterston, *Science*, 19 March 1999, on the new goal to produce a rough draft

"The scientific community thinks this is just a business project, and the business community thinks it's just a science project." J. Craig Venter, *Science*, 18 June 1999

"Why should I play by their rules when I am not getting a cent of federal money? Let me get this straight. I am being criticized for doing the work and giving it away free, but not giving it away fast enough?" J. Craig Venter, interview with L. Roberts, 2 September 1999

THE DRAFT NEARS COMPLETION

"The change is so fundamental it is hard for even scientists to grasp." Maynard Olson, interview with L. Roberts, 16 November 1999

"Ten, 15 years from now, nobody is going to care about all this fuss and bother. They're going to care that we got the ... human sequence done. ... And all this back and forthing over who did what and what strategy was used and which money was public and which was private is probably going to sink below the radar screen. And hallelujah." Francis Collins, interview with L. Roberts, 19 August 1999

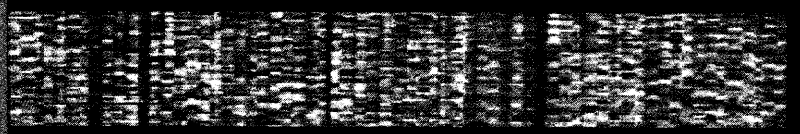
"We've called the human genome the blueprint, the Holy Grail, all sorts of things. It's a parts list. If I gave you the parts list for the Boeing 777 and it has 100,000 parts, I don't think you could screw it together, and you certainly wouldn't understand why it flew." Eric Lander, Millennium Evening at the White House, 14 October 1999

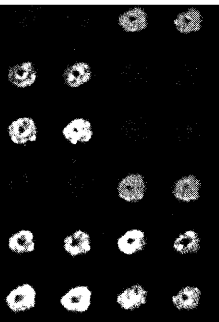
"Free will will not go out of style once the sequence is done." Francis Collins, interview with L. Roberts, 11 November 1999

"The prevailing view is that the genome is going to revolutionize biology, but in some ways, it's overhyped. In the end, the real insights are coming from individuals studying one gene at a time in real depth." Gerald Rubin, interview with E. Pennisi, May 2000

"If there is anything worth doing twice, it's the human genome." David Haussler, interview with E. Pennisi, July 2000

"Biology will never be the same." John Sulston, interview with E. Pennisi, February 2000

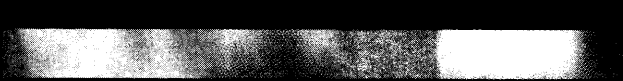


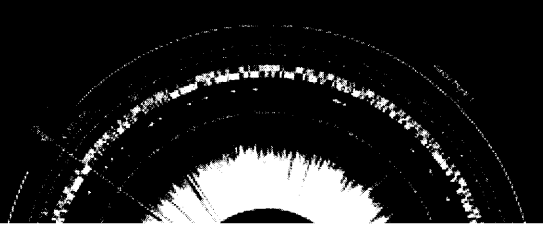


1953

A History of the Human Genome Project 2001

Special News Section Editor: Leslie Roberts
Research: R. John Davenport, Elizabeth Pennisi, Eliot Marshall, Leslie Roberts
Design: Joshua Moglia, Alan Stonebraker, C. Faber Smith
Illustration: Cameron Slayden
Photo research: Leslie Blizard
Production: Tara Kelly
Copy editor: Linda Felaco
Glossary: Laura Helmuth





A GENOME GLOSSARY

Science

Allele Alternative versions of a gene or other segment of a chromosome

Alternative splicing Different ways of combining a gene's exons to make variants of the complete protein

Amplification Repeated copying of a piece of DNA

Annotate Identify the locations and coding regions of genes in a genome and determine what they do

Assembly Putting sequenced fragments of DNA into their correct chromosomal positions

BAC Bacterial artificial chromosome: bacterial DNA spliced with a medium-sized fragment of a genome (100 to 300 kb) to be amplified in bacteria and sequenced

Bioinformatics The study of genetic and other biological information using computer and statistical techniques

BLAST A computer program that identifies homologous genes in different organisms, such as human, fruit fly, or nematode

Centimorgan (cM) A unit of genetic distance, determined by how frequently two genes on the same chromosome are inherited together

Centromere The difficult-to-sequence central region of a chromosome

Coding DNA Sequences transcribed into protein structures; also called exons

Contig Contiguous sequence of DNA created by assembling overlapping sequenced fragments of a chromosome (whether natural or artificial, as in BACs)

Cosmid DNA from a bacterial virus spliced with a small fragment of a genome (45 kb or less) to be amplified and sequenced

Directed sequencing Successively sequencing DNA from adjacent stretches of chromosome

Draft sequence Sequence with lower accuracy than a finished sequence; some segments are missing or in the wrong order or orientation

EST Expressed sequence tag: a unique stretch of DNA within a coding region of a gene; useful for identifying full-length genes and as a landmark for mapping

Exon Region of a gene's DNA that encodes a portion of its protein; exons are interspersed with noncoding introns

Finished sequence Sequence in which bases are identified to an accuracy of no more than 1 error in 10,000 and are placed in the right order and orientation along a chromosome with almost no gaps

FISH Fluorescence in situ hybridization: a method for pinpointing the location of a piece of DNA sequence on a chromosome

Functional genomics The study of genomes to determine the biological function of all the genes and their products

Gene expression Conversion of the information encoded in a gene first to messenger RNA and then to a protein

Gene prediction Predictions of possible genes made by a computer program based on how well a stretch of DNA sequence matches known gene sequences

Genetic linkage map A map of the relative positions of genes and other regions on a chromosome, determined by how often loci are inherited together

Genome The entire chromosomal genetic material of an organism

Genomics The comprehensive study of whole sets of genes and their interactions rather than single genes or proteins

Homologous genes Genes with similar structures and functions

Intron Region of a gene's DNA that is not translated into a protein

Junk DNA Stretches of DNA that do not code for genes; most of the genome consists of so-called junk DNA

Kilobase (kb) Unit of DNA length equal to 1000 bases

Locus Chromosomal location of a gene or other piece of DNA

Megabase (Mb) Unit of DNA length equal to 1 million bases

PCR Polymerase chain reaction: a technique for amplifying a piece of DNA quickly and cheaply

Physical map A map of the locations of identifiable markers spaced along the chromosomes; a physical map may also be a set of overlapping clones

Plasmid Loop of bacterial DNA that replicates independently of the chromosomes; artificial plasmids can be inserted into bacteria to amplify DNA for sequencing

Polymorphism A variation in DNA sequence within a population

Proteome The full complement of proteins produced by a particular genome

Proteomics The study of the full set of proteins encoded by a genome

Pseudogene A sequence of DNA similar to a gene but nonfunctional; probably the remnant of a once-functional gene that accumulated mutations

Regulatory region A segment of DNA that controls whether a gene will be expressed and to what degree

Repetitive DNA Sequences of varying lengths that occur in multiple copies in the genome; it represents much of the genome

Restriction enzyme An enzyme that cuts DNA at specific sequences of base pairs

RFLP Restriction fragment length polymorphism: genetic variation in the length of DNA fragments produced by restriction enzymes; useful as markers on maps

Scaffold A series of contigs that are in the right order but are not necessarily connected in one continuous stretch of sequence

Shotgun sequencing Breaking DNA into many small pieces, sequencing the pieces, and assembling the fragments

SNP Single-nucleotide polymorphism: common, single-base-pair variations in DNA

Structural genomics The effort to determine the 3D structures of large numbers of proteins using both experimental techniques and computer simulation

STS Sequence tagged site: a unique stretch of DNA whose location is known; serves as a landmark for mapping and assembly

Telomere The free end of a chromosome

Transcription factor A protein that binds to regulatory regions and controls gene expression

Transcriptome The full complement of activated genes, or mRNAs or transcripts, in a particular tissue at a particular time

YAC Yeast artificial chromosome: yeast DNA spliced with a large fragment of a genome (up to 1000 kb) to be amplified in yeast cells and sequenced