



Credit: Rawpixel Ltd / Alamy Stock Photo

## 2000 MILESTONE 1

# The Human Genome Project

The signature aim of the Human Genome Project (HGP), which was launched in 1990, was to sequence the 3 billion bases of the human genome. Additional goals included the generation of physical and genetic maps of the human genome, as well as mapping and sequencing of key model organisms used in biomedical research.

In 1998, the HGP formally implemented the [Bermuda Principles](#), specifically the following: automatic release of sequence assemblies >1 kb, preferably within 24 h; immediate publication of finished annotated sequences; and making the entire sequence freely available in the public domain for both research and development in order to maximize its benefits to society. In exchange for the immediate online release of HGP-funded sequence data, research groups from the USA, UK, Japan, France, Germany and China conducting the sequencing retained the right to be the first to describe their complete datasets and to analyse their findings in peer-reviewed publications.

By insisting on the Bermuda Principles, the HGP sought to undermine the efforts of parties aiming to patent or commercialize human genomic sequences, which could restrict subsequent research efforts. In 1998, it was announced that a new company, later renamed Celera Genomics, would 'race' the publicly funded HGP to complete the sequencing of the human genome. Celera Genomics also intended to sell subscriptions to its database, release data quarterly, and obtain patents on genes and related technologies.

This new presence threatened the survival of the HGP (which by early

1998 had sequenced only a small fraction of the human genome), but after US President Clinton and UK Prime Minister Tony Blair jointly declared on 14 March 2000 that the human genome sequence "should be made freely available to scientists everywhere", the HGP and Celera Genomics brokered a deal leading to the simultaneous publication in February 2001 of two articles (by Venter et al. in *Science* and the International Human Genome Sequencing Consortium in *Nature*) describing the draft human genome sequence. The sequence included 26,588 protein-coding transcripts for which there was strong corroborating evidence and an additional ~12,000 computationally derived genes with mouse homologues or other weak evidence.

The HGP used a hierarchical shotgun sequencing approach, in which the genome was broken into ~150-kb segments and cloned into bacterial artificial chromosomes, before being matched to a genome-wide physical map comprising >96% of the euchromatic part of the human genome (~94% of the entire human genome). Selected bacterial artificial chromosomes were sequenced and finally reassembled to generate the draft sequence. By contrast, Celera Genomics used both HGP and their own private data in their whole-genome shotgun sequencing approach, which fragmented the genome into ~500-bp segments and subjected them to pairwise end sequencing (in which a given fragment is sequenced from both ends to produce a 'mate pair') to reconstruct the original sequence.

In 2003, a group of ~40 professionals working in genomics publicly

“ Today, the HGP remains notable for... revenue and paradigm shifts [it] generated ”

declared their support for the free and unrestricted use of genome-sequencing data by the scientific community before formal publication. This declaration, known as the [Fort Lauderdale Agreement](#), enshrined the collective responsibility of funding agencies, resource producers and users to maintain and expand a communal trove of genomic data. These principles were later implemented as policy by several funding agencies, notably the US National Institutes of Health (NIH), which today still mandates rapid data-sharing in its grant requirements. Collectively, these initiatives can be considered the forerunners of open-access publishing in biomedicine.

In 2004, the work of the HGP culminated in publication of a highly accurate (~1 error per 100,000 bases) human genome sequence that included ~99% of the euchromatic genome. The current version of the human reference genome, GRCh38.p13, comprises 3.27 billion nucleotides and 19,116 nuclear protein-coding genes.

Today, the HGP remains notable for an estimated US\$800 billion of revenue and paradigm shifts generated by this publicly funded 'big science' project. Offering a first view into the entire human genome, the HGP acted as a gateway to an era of high-throughput digital biology, ushering in rapid technological and computational developments and team-oriented research, the fruits of which continue to be felt across the clinical and life sciences.

Caroline Barranco,  
Nature Reviews Cross-Journal Team

**ORIGINAL ARTICLES** International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001) | Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304–1351 (2001)  
**FURTHER READING** International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004) | Maxson Jones, K., Ankeny, R. A. & Cook-Deegan, R. The Bermuda Triangle: the pragmatics, policies, and principles for data sharing in the history of the human genome project. *J. Hist. Biol.* **51**, 693–805 (2018)