Check for updates

The Human Genome Project changed everything

Richard A. Gibbs

Thirty years on from the launch of the Human Genome Project, Richard Gibbs reflects on the promises that this voyage of discovery bore. Its success should be measured by how this project transformed the rules of research, the way of practising biological discovery and the ubiquitous digitization of biological science.

> ❝ it is simply inconceivable today that we would not have the genome at our fingertips ❞

The joint announcement of the release of the human 'draft' genome sequences occurred 20 years ago, at a ceremony in the White House. The first analyses by two groups, the publicly funded International Human Genome Project (HGP) Consortium and Celera Genomics, were published in *Nature*[1] and *Science*[2], respectively, shortly after. While the analyses were superficial by contemporary standards, this was nevertheless a milestone that provided exciting first glimpses into the entire human genome. The announcement was hailed as 'the end of the beginning' and a launch pad for a new era. After two decades, have the aspirational aims of the HGP been realized? Without doubt, the answer is yes; it is simply inconceivable today that we would not have the genome at our fingertips — as unimaginable, perhaps, as not having computers or the internet.

Critics cite a failure to meet the most outlandish visions as evidence that the HGP has not lived up to all promises. The project was initially conceived with fairly sober predictions, including the benefits of a complete cancer genome, advances in genetics and the development of improved technologies[3]. It was not until closer to the programme launch in 1990 and at milestones along the way that the rhetoric was loudly elevated to claims of revolutionizing biology, biotechnology, drug development and even society. A favourite prediction was the personalization of therapies and the liberation of drugs that otherwise were unusable, through identification of the few individuals with adverse responses. The mysteries of the architecture of common complex diseases were to be revealed and even behavioural traits might be solved. The predictions included the possibility to breed 'super babies' based on this new knowledge and, at the same time, perhaps even predict criminality[4]. In hindsight, there was plenty of hype that was shared with the media and the wider community. Critics are correct that the apex of these claims was not reached. The hyperbole that we look back on did not, however, come from the front line. It came from those who championed the programme, mindful of its long-term benefits. Thanks to them, they generated the enthusiasm to fund this transformative work.

Among those immersed in the delivery of the primary aims of the project, the mood was more measured. 'Basic' biologists wanted their favourite model organisms characterized so that human gene homologues could be identified. Clinical geneticists were fixated on discovery and genetic dissection of the molecular basis of inherited childhood disorders, while adult disease specialists sought answers to why some suffered common maladies, such as cardiovascular disease or cancer. Technologists recognized that this was the gateway to the new era of high-throughput, digital biology.

There were still lofty goals, and major contributors who were convinced of the imperative of completing the project shared core beliefs of the broad impact of a completed human sequence. All recognized that, for the first time, these studies would share a characteristic comprehensiveness that was an uncommon luxury in biology. For the first time, there would be knowledge on all genes, all diseases and all genetic variants. Participants recognized the power of broad data sharing and the legacy of the Bermuda Principles for future biology[5]. The organizational rigor required to manage the HGP was

Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA.

e-mail: agibbs@bcm.edu

https://doi.org/10.1038/
s41576-020-0275-3

The author

Richard Gibbs, AC, PhD is a human geneticist and the Founding Director of the Baylor College of Medicine Human Genome Sequencing Center (HGSC). He graduated from the University of Melbourne in Genetics and Radiation Biology and moved to Houston, TX, to study the molecular basis of genetic disease. He developed basic methods for DNA and mutation analysis and was an early contributor to the Human Genome Project (HGP), leading one of five sites that generated the majority of the sequence. Since the completion of the HGP, he has led multiple genome projects including the generation of the first personalized whole-genome diploid human sequences. His group pioneered the oligonucleotide exon-capture methods that are widely used today for whole-exome sequencing, and he is currently leading programmes for translation of genomic data into the clinic.

COMMENT

> *validation, contextualization, deployment and translation are all streamlined by the fruits of the HGP*

new for biology, and it was apparent that future programmes would benefit from HGP lessons in logistics. These ambitions were the backdrop for the knowledge of how difficult the task would be, without advanced computers, automated sequencing or any roadmap from a similar effort.

A 25-plus-year timetable

There was also a realistic insiders' view of likely post-HGP rates of progress and how difficult biological discovery can be, in the best of circumstances. The HGP was foundational and the project would lead to new ways to do things, but not all thought progress would be easy. The HGP took just 13 years, as after the 2000 announcement we all worked an extra 3 years to finish the 'essentially complete genome', and it is interesting to compare that period to other transitional milestones in biology. In 1987, the groups of Francis Collins and Lap-Chee Tsui discovered the gene that contains the variants that underlie cystic fibrosis[6]. That discovery (pre-HGP) was appropriately hailed as the first step towards a cure. In 2012, the first resulting drug to treat a subset of patients with cystic fibrosis was approved by the FDA. For Huntington disease, a similar time span was needed to go from gene discovery to a new treatment that is only now being tested[7]. The familial breast cancer gene is another example of the time between discovery and action; linkage to *BRCA1* was identified in the 1990s with initial hopes that isolating the gene underlying the 1% of cases that were familial would give insights into the vast majority of sufferers with sporadic disease. That connection was not obvious, and the complicated relationship between this gene, its germline and somatic variants, related genes and interacting proteins, and the consequences for cancer are still being unraveled[8]. A 25–30-year period between discovery and impact on health care is more the rule than the exception.

Parallel transformations

HGP participants trusted their own power to innovate but also hoped for other developments to leverage the programme. While the project unfolded, a revolution occurred in computation. In the late 1980s, the only computers in the laboratories of genomicists were the earliest PCs and Apple products. By 2000, we had all been connected by the internet, bandwidth was adequate to move the genome data, and adequate processing power was accessible. A strength of the HGP and its participants was that these parallel developments were rapidly incorporated into the framework of biology. Necessity speeds invention — and the need to manage copious amounts of digital genome data was the real driver of the growth of computational biology, ahead of the demands of physiologists or structural biologists. Most importantly, a generation of bioinformatics experts and computational biologists emerged who brought the genome data to the widest audiences.

The power of advances in genomics and computers was revealed in the spectacular series of post-HGP

projects that were of comparable scale. After multiple mammalian genome projects, programmes including the Haplotype Mapping (HapMap) Project[9], the 1000 Genomes Project[10] and The Cancer Genome Atlas (TCGA) progressively illustrated the advancement of knowledge by more sophisticated data sharing, comparison and analysis. As these and other projects unfolded, new constituencies were engaged and more scientists and clinicians became 'digital' and 'genomic'. The projects were emblematic of the advancement of scaling, digitization and sharing that was sparked by the HGP.

Some still tally the success of the HGP from lists of new drugs or therapies and argue that world-changing examples in biology, such as the spectacular advances of gene editing tools or the expansion of cancer therapeutics through targeted immunotherapy, are largely based on microbial, cellular and animal studies rather than genomics. This argument misses the point. These are among the myriad of discoveries that occurred in the backdrop of a new era. New ideas and primary discovery may still be the 'quiet conversation with nature' of the experimental biologist — but validation, contextualization, deployment and translation are all streamlined by the fruits of the HGP.

It is a vastly different world today in 2020, compared with 1990. Human genome sequences cost less than US$1,000 per genome, all trainees in experimental biology and genetics are pressed to be proficient in computer languages, and easy access to mountains of primary and derived data has come to be expected. As the recent coronavirus pandemic emerged, thousands of trainees, forced to remain out of the wet-lab, pivoted to computational studies; 30 years ago they would have been lost. The real fruits of the HGP lie in the contrast between the primitive state of digital biology in the late 1980s and the current ease with which all scholars can access, harness and analyse biological data.

1. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
3. Cook-Deegan, R. M. *The Gene Wars: Science, Politics, and the Human Genome* (W.W. Norton & Company, 1994).
4. Friedland, S. I. The criminal law implications of the Human Genome Project: reimagining a genetically oriented criminal justice system. *KY Law J.* **86**, 303–366 (1997).
5. Marshall, E. Bermuda rules: community spirit, with teeth. *Science* **291**, 1192 (2001).
6. Rommens, J. M. et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**, 1059–1065 (1989).
7. Dash, D. & Mestre, T. A. Therapeutic update on Huntington's disease: symptomatic treatments and emerging disease-modifying therapies. *Neurotherapeutics* https://doi.org/10.1007/s13311-020-00891-w (2020).
8. King, M.-C. "The race" to clone BRCA1. *Science* **343**, 1462–1465 (2014).
9. Gibbs, R. A. et al. The International HapMap Project. *Nature* **426**, 789–796 (2003).
10. Durbin, R. et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

Competing interests
The author declares no competing interests.