

COMMENT

SUPERCOMPUTING Approximate processing will advance modelling **p.32**

AUTUMN BOOKS Grande dame of food politics takes on the soft-drinks industry **p.34**



CLIMATE Geoengineering debated, with erudition and poetry **p.38**

EARTH Marking the centenary of the book that proposed continental drift **p.43**

TOP: BOTTOM, RIGHT: COLD SPRING HARBOR LAB. LIBRARY & ARCHIVES; BOTTOM LEFT: ERIC GREEN



1989, Pre-HGP Banbury meeting. Francis Collins & James Watson in top row.



1990, Washington University School of Medicine. Madynard Olsen & Eric Green.



1997, HGP meeting at Cold Spring Harbor. E. Green, R. Myers, J. Witkowski & R. Gibbs.

Twenty-five years of big biology

The Human Genome Project, which launched a quarter of a century ago this week, still holds lessons for the consortium-based science it ushered in, say **Eric D. Green, James D. Watson and Francis S. Collins.**

Twenty-five years ago, the newly created US National Center for Human Genome Research (now the National Human Genome Research Institute; NHGRI), which the three of us have each directed, joined forces with US and international partners to launch the Human Genome Project (HGP). What happened next represents one of the most historically significant scientific endeavours: a 13-year quest to sequence all three billion base pairs of the human genome.

Even just a few years ago, discussions surrounding the HGP focused mainly on what insights the project had brought or would bring to our understanding of human disease. Only now is it clear that, as well as dramatically accelerating biomedical research, the HGP initiated a new way of doing science.

As biology's first large-scale project, the HGP paved the way for numerous consortium-based research ventures. The NHGRI alone has been involved in launching more than 25 such projects since 2000. These have presented new challenges to biomedical research — demanding, for instance, that diverse groups from different countries and disciplines come together to share and analyse vast data sets.

It is easy for young researchers to forget that many of the problems they are trying to solve today had not even been thought about by their predecessors a quarter of a century ago. Equally easy to lose sight of are the insights that the HGP still offers to those pursuing big science projects. In fact, we think that the success of today's consortium-based science depends on six key lessons from the HGP.

SIX LESSONS

Embrace partnerships. By necessity, the HGP broke the mould of individual researchers toiling away in isolation to answer a small set of scientific questions. It also ran against the grain of hypothesis-driven research, focusing instead on the discovery of fundamental information that would inform many follow-on investigations.

The HGP brought together more than 2,000 researchers from many countries, disciplines and levels of seniority, with subgroups answering to different funding agencies. Success stemmed from: strong ▶

► leadership from the funders; the shared sense of the importance of the task; and the willingness of the researchers involved to cede individual achievements for the collective good¹.

Many consortium-based genomics projects followed. Among them are the 1000 Genomes Project, which is cataloguing sequence variants in the human genome (see pages 68 and 75), The Cancer Genome Atlas, which is characterizing the mutations responsible for cancer, and the Human Microbiome Project, which uses genome sequencing, among other techniques, to study microbial communities.

A frequent barrier to consortium-based science is the unwillingness of participants to embrace new partnerships. But various efforts — combined with the increasing realization that pooling data and resources can benefit everyone — are dismantling old norms.

Until recently, for instance, African genetics and genomics researchers collaborated most often with US or European scientists, and seemed less inclined to partner with other African researchers. A key objective of the Human Heredity and Health in Africa (H3Africa) initiative², which aims to enhance genomics research in Africa, has been to foster collaborations within Africa. The initial set of grants awarded by the US National Institutes of Health (NIH) and Britain's Wellcome Trust for the project in 2012 and 2013 established 29 collaborations involving 24 African countries; those numbers have since increased. H3ABioNet, a bioinformatics network that aims to facilitate the sharing of expertise, infrastructure and tools for analysing data across Africa, now involves 32 research groups in 15 countries.

Maximize data sharing. The HGP changed the norms around data sharing in biomedical research. Once large amounts of genome mapping and sequence data began to be generated, momentum quickly grew for establishing policies that shortened the time between the generation and release of data. These efforts culminated in adoption of the Bermuda Principles in 1996, when the heads of the major groups involved in the project agreed to submit genome-sequence assemblies above a certain size to a public database within 24 hours of generating them.

Such efforts have been built on in the years since. The principles were extended by the Fort Lauderdale Agreement in 2003. And in 2008, the NIH expanded its data-sharing expectations to include genome-wide association studies — analyses of common genomic variants in hundreds or thousands of people conducted to reveal variants associated with some trait of interest. In 2014, it started implementing an expansive Genomic Data Sharing Policy, which requires that almost all



Early days: a DNA-sequencing lab in 1994.

large-scale genomic data generated or analysed using NIH funds are shared.

Widespread sharing of data is throwing up new challenges. These include the computational and logistical difficulties of analysing and moving vast data sets; and in the case of human data (especially genomic and clinical), the problem of how to protect the privacy of research participants. Various initiatives are being pursued to address these problems.

The need for robust and powerful computing platforms is leading to rapid growth in the use of cloud computing in biomedical research, for instance. New resources are being proposed, such as a 'data commons' to house published and unpublished data³. And the Global Alliance for Genomics and Health, an international coalition established in 2013, is preparing an international Framework for Responsible Sharing of Genomic and Health-related Data⁴. This will take into account legal, ethical and technical considerations.

Plan for data analysis. Planning for the HGP had its flaws. In retrospect, one area that received insufficient attention early on was data analysis. The first

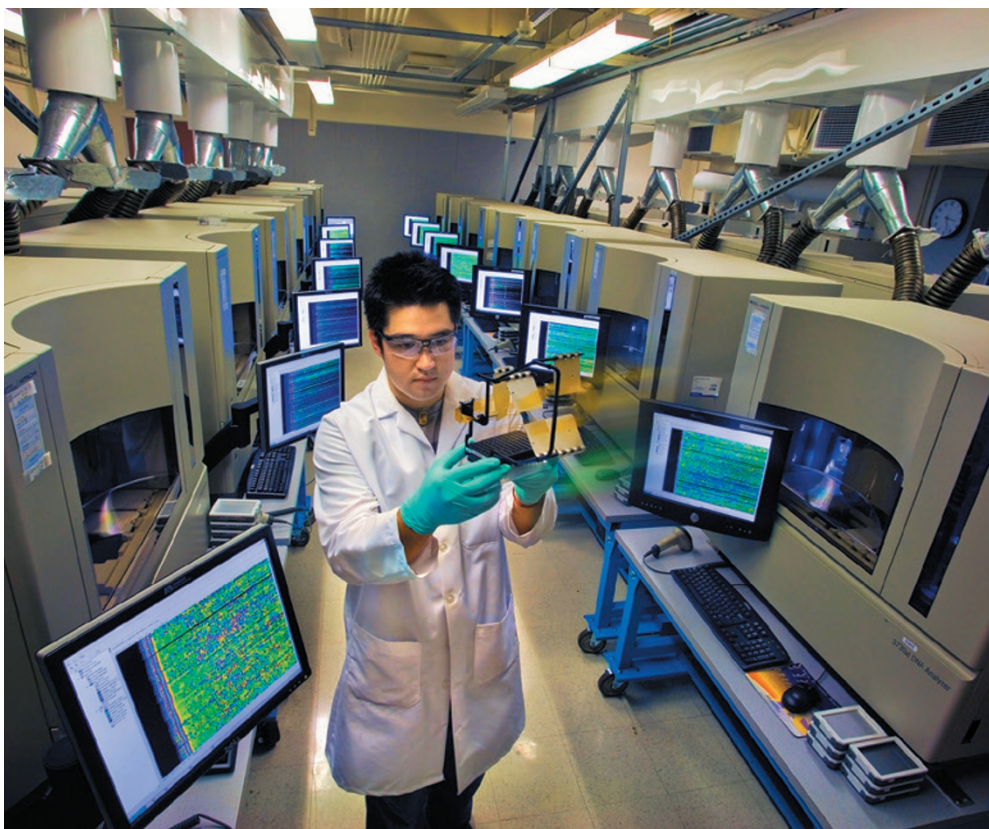
human genome sequence was produced in a piecemeal fashion. And to generate a contiguous sequence for each chromosome, thousands of individually assembled sequence segments (each around 100–300 kilobases) had to be stitched together computationally. The need for such a computational process (which turned out to be technically challenging) became apparent relatively late in the project. Through the heroic efforts of a small group of bioinformaticians, this task was accomplished in a matter of months. More care in planning would have made the endeavour much less stressful.

In recent years, several genomics projects (such as the 1000 Genomes Project and The Cancer Genome Atlas) have demonstrated how the early design of plans for data analysis can inform strategies for data generation. More recently, planning for the US Precision Medicine Initiative⁵ included considerable discussion about how best to merge and analyse the anticipated myriad data types — from electronic health records and genomic analyses to information from environmental monitors and wearable body sensors.

Prioritize technology development. In October 1990, the HGP participants pressed ahead, fully aware that the tools and methods for mapping and sequencing the human genome would need

“Waiting for absolute clarity about how the ultimate goals will be achieved risks missing opportunities.”

LEFT: HANK MORGAN/SPL; RIGHT: LAWRENCE BERKELEY NAT'L LAB/SPL



By 2006, DNA sequencing required much less manpower.

to be developed as part of the larger programme. In fact, the project catalysed the development of numerous crucial genomic technologies, and led to substantial innovations in molecular biology, chemistry, physics, robotics and computation, as well as to strategies for using tools and methods in innovative ways. In some cases, multiple incremental improvements were cobbled together to yield revolutionary advances, such as the capillary-based DNA sequencing instruments that were ultimately used to generate the first human genome sequence.

The need to foster technical innovations from the start is similarly crucial for today's large-scale projects. One effort leading the way in this respect is the US Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative⁶. With the overarching goal of revolutionizing our understanding of the human brain, the programme will focus initially on developing a new generation of tools for defining all the cell types in the brain, building maps of their connections, and recording signals from circuits that can be correlated with functions and behaviours.

Address the societal implications of advances. The founders of the HGP recognized that the information gained from mapping and sequencing the human genome could have profound implications

for society. The HGP thus became the first large-scale research project to include a component dedicated to examining broader societal issues, such as how to protect people's privacy and prevent discrimination. This arm of the project — known as ELSI (ethical, legal and social implications) research — was supported by about 5% of the NIH budget for the HGP⁷. It was the largest ever investment in bioethics research.

Societal and ethical considerations attend many of today's cutting-edge pursuits. High-profile examples include the use of the CRISPR/Cas9 gene-editing tool to alter the genomes of humans and other species, and the fast-tracking of clinical trial design for the rapid study of potential treatments during infectious outbreaks. Unfortunately, most consortium-based projects do not include a dedicated bioethics research programme as the HGP did. We think that as new large initiatives are launched, such programmes should be a key component.

Be audacious yet flexible. The goals of the HGP were bold. Given the lack of clarity on how exactly the human genome would be mapped and eventually sequenced, it was not surprising that the effort was viewed with some scepticism.

We believe that key to the HGP's success was the continued open-mindedness of the scientific leaders, and the

regular pauses they took to take stock. The initial five-year plan for the HGP was updated with revised plans in 1993 and in 1998. Individual HGP elements were regularly refined⁸.

Large projects with daring goals can prosper as long as overall objectives are grounded in explicit milestones, quality metrics and assessments. They also need a willingness to iterate plans as needed. Waiting for absolute clarity about how the ultimate goals will be achieved risks missing opportunities that present themselves only after researchers start work. This formula has become the norm for several large-scale projects, among them the BRAIN Initiative and the Precision Medicine Initiative.

GAME CHANGER

In the early 1990s — whether it was while leading the NIH's effort in the HGP (J.D.W. and F.S.C.) or working on the front line of the project (E.D.G.) — none of us foresaw that a major legacy of the HGP would be a new way of doing science.

During their careers, today's graduate students will probably witness and facilitate the unravelling of the molecular mechanisms for thousands of diseases, a revolution in cancer diagnosis and treatment, the maturing of microbiome science, the routine use of stem-cell therapies, and other spectacular biomedical advances.

The story of the HGP provides a valuable reminder that some of these advances will almost certainly trigger fundamental changes in the way that research is done — as well as a reminder of the importance of accepting and celebrating those changes. ■

Eric D. Green is director of the US National Human Genome Research Institute at the US National Institutes of Health, Bethesda, Maryland, USA. **James D. Watson** is chancellor emeritus at the Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA, and former director of the US National Center for Human Genome Research. **Francis S. Collins** is director of the US National Institutes of Health, Bethesda, Maryland, USA, and former director of the US National Human Genome Research Institute.

e-mails: egreen@nhgri.nih.gov; collinsf@mail.nih.gov

- Collins, F. S. *et al.* *Science* **300**, 286–290 (2003).
- H3Africa Consortium. *Science* **344**, 1346–1348 (2014).
- Stein, L. D. *et al.* *Nature* **523**, 149–150 (2015).
- Knoppers, B. M. *HUGO J.* **8**, 3 (2014).
- Collins, F. S. & Varmus, H. N. *Engl. J. Med.* **372**, 293–295 (2015).
- Insel, T. R. *et al.* *Science* **340**, 687–688 (2013).
- McEwen, J. E. *et al.* *Annu. Rev. Genomics Hum. Genet.* **15**, 481–505 (2014).
- Green, E. D. in *The Metabolic and Molecular Bases of Inherited Disease* 8th Edn (eds Scriver, C. R. *et al.*) 259–298 (McGraw-Hill, 2001).