CHAPTER 3

The Human Genome Project(s)

One of the consequences of excluding a great deal of the collaborative genome efforts that proliferated in the 1980s and 1990s from the success story of genomics has been the assumption that human genomics corresponded to a single initiative or entity. This assumption portrays the Human Genome Project as one international endeavour that started and ended at defined dates, presented a set of stable participants, and operated according to a predefined plan: the large-scale production of a reference sequence of the whole human genome. The narrative of a single human genome effort consolidated in June 2000, when a consortium of funders, sequencing centres and bioinformatics institutions from Europe, Asia and North America presented a first draft of the full sequence of *Homo sapiens* in a ceremony chaired by the US president, Bill Clinton, and attended remotely by UK Prime Minister, Tony Blair (Chap. 4). Before—and contemporaneous to—this announcement, a number of multinational genome initiatives to sequence yeast (*Saccharomyces cerevisiae*), the fruit fly (*Drosophila melanogaster*) and the thale cress (*Arabidopsis thaliana*) were unfolding with substantial leadership from the European Commission (Chap. 2).

The draft human genome sequence was published in the journal *Nature* in 2001. This article referred to the sequencing effort as the "Human Genome Project" and defined this project as an "international

M. García-Sancho, J. Lowe, *A History of Genomics across Species, Communities and Projects*, Medicine and Biomedical Sciences in Modern History, https://doi.org/10.1007/978-3-031-06130-1_3

collaboration" that had started in 1990 and was scheduled to conclude with the release of a more final sequence, which appeared in a follow-up publication, also in *Nature*, in 2004 (International Human Genome Sequencing Consortium, 2001, pp. 860 and 862; 2004). Since then, press coverage, popular literature and a substantial amount of academic scholarship have depicted a single, international Human Genome Project.[1] The depiction of the role of the European Commission (EC) as a funder and broker of genomic endeavours has tended to be restricted to yeast sequencing and presented as an antecedent to the Human Genome Project. As we discussed earlier in the book, this consideration of *S. cerevisiae* as a pilot or model platform for human sequencing aligns more with the US yeast genome effort than with the EC one. The EC, rather, selected yeast as an industrially-significant organism that would foster economic growth and scientific collaboration across its member-states (Chap. 2; see also Parolini, 2018).

In this chapter, we continue augmenting the historical landscape of genomics and de-centring it beyond the production of a human reference sequence. We start by arguing that instead of a monolithic Human Genome Project (with capitals H, G and P), a plethora of national and international human genome initiatives co-existed from the mid-1980s onwards with different rationales, spokespersons and funding regimes.[2] As late as 1996, the strategy of tackling the whole human genome via the concerted action of a handful of large-scale sequencing centres was not yet dominant. Contemporary historical accounts (e.g., Cook-Deegan, 1994,

[1] This terminology was already present in the 1990s and early 2000s, especially among scholars based in the USA and working on the socio-ethical implications of "the Human Genome Project" (Kevles & Hood, 1992; Sloan, 2000). As we show later in the chapter, this literature often conflated the international sequencing effort with the US national human genome project. Europe-based scholars have tended to be more nuanced and distinguish different initiatives and approaches to the human genome (Glasner & Rothman, 1998). Later sociological and historical investigations have continued to refer to the Human Genome Project, acknowledging the multiple genealogies behind the genesis of this term (Hilgartner, 2017; Stevens, 2013).

[2] For a survey of different human genome efforts written at the time they were developing, see McLaren (1991), Cook-Deegan (1994, Ch. 14) and the section "European contributions" from the Spring 1991 newsletter of the UK Human Genome Mapping Project: Nigel K. Spurr (ed.) *G-Nome News,* volume 6: 30-69, National Archives of the UK at Kew (London), Medical Research Council Series, file number FD7/2745. In this chapter, we focus on the national human genome projects of the USA and UK, and to a lesser extent on the EC's Human Genome Analysis Programme.

Part Three) document that only one national initiative unambiguously sought, from the onset, to produce a physical map and a reference sequence of the entire human genome: the joint programme of the USA's Department of Energy (DoE) and National Institutes of Health (NIH). Due to this, a widely accepted meaning of the capitalised phrase 'Human Genome Project' during most of the 1990s was just the US national effort, which itself adopted that name.

We designate the US national programme throughout this chapter as 'US-HGP'. It formally commenced in 1990, when some other national human genome projects were already underway, and had as a defining characteristic the concentration of NIH and DoE funding in a series of centres that specialised in various aspects of genomics, such as physical mapping, large-scale sequencing, bioinformatics or technology development (Hilgartner, 2017, Ch. 2 and pp. 91-110). Some of these centres already existed and were devoted to other types of research, such as medical genetics or, in the case of those supported by the DoE, the effects of radiation on DNA. Others were created *de novo* to comprehensively sequence the human genome and those of pilot organisms, such as yeast (Chap. 2). All of the centres were committed to the objective of full genome mapping and sequencing, a feature that distinguished the US-HGP from many other contemporary human genome programmes. As we highlight, a leading architect in the design of the new centres and advocate of their whole-genome approach was the Nobel Prize-winning molecular biologist—and co-discoverer of the double helical structure of DNA—James Watson, who led the NIH arm of the US-HGP until 1992.

Among national human genome programmes, the objective of mapping and sequencing the whole human genome was unique to the US-HGP, as was the prominence of a leader such as Watson. The main insights of this chapter stem from comparing the US programme with another, less well-known national initiative: the UK Human Genome Mapping Project (HGMP). Launched one year earlier, in 1989, and funded by the British Government through its Medical Research Council (MRC), the HGMP did not create large-scale sequencing centres. As with many other emergent human genome projects, its strategy aligned with the distributed, network approach that the EC was forging for the sequencing of yeast (Chap. 2).

The HGMP enabled the MRC to secure funds from the UK Treasury for a Directed Programme of grants specifically tailored to map and

sequence human DNA. The recipients of those grants were laboratories in the fields of human genetics and, especially, medical genetics. Those recipients and the ways they aimed to tackle the human genome were key differences between the British programme and the US-HGP. Rather than promoting a new breed of whole-genome-oriented practitioners, as the DoE and the NIH were fostering, the HGMP funded and coordinated research groups that kept working on specific parts of the human genome. In other words, the communities of genomicists that constituted each programme differed. Although the beneficiaries of HGMP grants collectively produced map and sequence data across the genome, they retained their individual identity as specialists in diseases or biological phenomena affecting only certain genome regions. Conversely, the specialism of the DoE and NIH-funded genomicists increasingly tended towards the large-scale mapping and sequencing of the entire human genome.

The HGMP beneficiaries used their grants to develop mapping and sequencing methods aimed at positioning, within human chromosomes, DNA fragments encompassing genes or gene markers associated with diseases, or any other biologically or medically relevant characteristic. They were assisted by a resource centre that the HGMP established as both a technological hub and a repository of the genomic data produced by the laboratories in receipt of Directed Programme funding (Balmer, 1998; Glasner, 1996). Apart from providing technical support and advice to the HGMP-funded laboratories, the resource centre pooled their mapping results and compiled them in databases.[3] It also conducted partial sequencing of the mapped DNA fragments, particularly the regions corresponding to genes that were thought to be involved in the genetic diseases that the HGMP laboratories investigated. This work was developed in collaboration with gene-specific sequencing groups sponsored by the Human Genome Analysis Programme, an initiative that the EC launched in 1990 that followed the distributed model it had just implemented for the yeast sequencing project (Chap. 2).

The HGMP Resource Centre differed from the US-HGP genome centres in two key aspects: (1) it fulfilled a service role and conducted mapping and sequencing work at the request—and based on the results—of the Directed Programme-funded laboratories rather than comprehensively

[3] As we see later in the book, the Resource Centre provided mapping tools and assistance to other communities throughout the 1990s, such as those involved in the mapping of the genome of the pig, *Sus scrofa* (Chap. 5).

sequencing at its own initiative; and (2) the map and sequence data it compiled represented only the areas of interest of the contributing laboratories and was thus not intended to be a *complete* representation of the whole human genome. As we argue, out of the HGMP, HGAP and other groupings of human and medical geneticists—such as the chromosome mapping workshops—a community of genomicists emerged, one that was larger and more diverse than the one working at the genome centres conducting the US-HGP.

This contrast enables us to conclude that a key factor distinguishing the US-HGP from the HGMP, and more generally from the distributed approach promoted by the EC, was in the assemblage of the research communities and funding regimes underlying each of them. In the case of the US-HGP, this assemblage embodied Watson's vision, his circle of influence and the joint funding provision of the DoE and NIH. Watson was a founder of molecular biology and, from the late-1960s onwards, had been instrumental in structuring this community from his position of director of Cold Spring Harbor Laboratory (CSHL). One of the pillars in this structuring process had been fostering the shared belief in the mechanistic action of genes and the community's commitment to detailed investigations of model organisms. It was hoped that a full molecular description of those organisms would unveil the role of genes in a myriad of biological processes.[4]

As we showed in the previous chapter, the worm *Caenorhabditis elegans* had become one such model organism. It was at CSHL where, in 1989, Watson met John Sulston, Alan Coulson and Robert Waterston, and instigated the start of the worm's sequencing project. The designated host institution of the project in the USA, Washington University, subsequently inaugurated a genome centre and undertook the sequencing of two other organisms: yeast and *H. sapiens*. This intensive and scaled-up approach differentiated the genome centres from the distributed model that the EC was promoting in its sequencing programmes (Chap. 2). The status of *C. elegans* and yeast as model organisms was one of the reasons that led Watson to regard them as suitable pilot platforms to inform the mapping and sequencing of the human genome. He did not hesitate in adopting the same genome centre model when the US-HGP—a self-contained,

[4] Other crucial figures in spreading the influence of molecular biology and promoting its mechanistic view of gene action were Sydney Brenner (discussed below) and co-elucidator of the DNA double helical structure, Francis Crick (Aicardi, 2016).

national initiative as opposed to the multi-country programmes of the EC—sponsored the comprehensive human genome sequencing effort. Under Watson's leadership, the US-HGP was the vehicle for producing a reference sequence from which the connections between genes and biological properties—implicating evolution, health and disease—could later be drawn.

The HGMP was also promoted by a founding figure of molecular biology: the proponent of *C. elegans* as a model organism, Sydney Brenner. Yet the MRC, partly due to the size of the UK relative to the USA, lacked the resources to launch a whole-genome initiative on its own. This led Brenner and the MRC to look at the communities of human and medical geneticists as possible allies to execute the project. Unlike molecular biologists, these communities were interested in variation rather than comprehensive standard descriptions as an entry point into investigating gene function. Consequently, their motivation to tackle the human genome was not achieving a complete reference sequence, but using the reference sequence data as a scaffold to aid in the determination of variants associated with diseases or evolutionary traits. Identifying and investigating variation, as opposed to establishing a canonical reference sequence, was thus a driving force behind the organisation of the HGMP and its indifference to adopting whole-genome approaches. From the viewpoint of the many laboratories supported by the HGMP Directed Programme, focusing on specific genome regions that could be compared with either other organisms or between patients suffering a genetic condition and non-sufferers was far more useful than mapping and sequencing the entire human genome.[5]

In what follows, we show that the differences in the funding systems, organisational models, communities and genomicists involved in the HGMP and US-HGP assemblages point to a diverse landscape. This diversity is difficult to grasp from a perspective that narrowly focuses historical inquiry on the human reference sequence published in 2001 and 2004. What is now associated with a single, coherent and successful Human Genome Project represents just one route through complex

[5] Elsewhere, we have characterised these different approaches through the categories of horizontal and vertical sequencing. Whereas horizontal sequencing would involve producing a one-dimensional reference genome, the vertical strategy would explore sequence variation in one specific genome region across individuals or different species, with the aim of augmenting clinical or evolutionary knowledge (García-Sancho, Leng et al., 2022).

historical terrain. Our ability to identify the web of pathways that criss-cross this terrain enables us to extend our historical interrogation from yeast to *H. sapiens*. The multiplicity of both parallel and interwoven lineages in the development of the HGMP and US-HGP indicates that the historical landscape was as heterogeneous in human as in non-human genomics. By looking both within and beyond human genomics, we can highlight the factors that led to the increasing prominence of the human reference genome. This enables us to assess its significance in a fresh light, while at the same time preventing it from narrowing our vision.

3.1 The Exception Rather Than the Rule

In 1988, Watson supplemented his CSHL directorship with a new role as associate director of the freshly-established NIH Office for Human Genome Research. He had held the CSHL position since 1968—15 years after co-elucidating the DNA double helix and 6 years after receiving the Nobel Prize—and transformed this institution into the most influential forum of molecular biology. CSHL held annual symposia in which the invitees, considered to be the international elite of molecular biologists, would discuss pressing scientific challenges. The 1986 symposium had been devoted to the *Molecular Biology of Homo sapiens* and became one of the first settings in which the feasibility of mapping and sequencing the human genome was assessed. The enormous size of the human genome—three billion DNA nucleotides compared to the 12 million of yeast and 100 million of *C. elegans*—made the viability and utility of the enterprise a matter of debate within and outside the CSHL meeting. In his 1988 CSHL director's report, Watson expressed concerns about his increased responsibilities and the stress of commuting. He considered, however, that the remit of the new NIH Office—implementing a national human genome programme in the USA—represented a one-time "opportunity". From this new position, Watson could let his scientific life "encompass a path from double helix to the three billion steps of the human genome".[6]

Watson's commitment to the sequencing of the human genome was shared by scientists and administrators at the DoE. However, rather than completing the molecular description of DNA—from ascertaining the

[6] "Director's report" in *Annual Report 1988—Cold Spring Harbor Laboratory:* 1-24, quote from p. 5. We thank Robert Cook-Deegan at Arizona State University for generously providing access to this record.

double helix to laying out its nucleotide sequence—what the DoE human genome advocates sought was to build on a longstanding tradition of investigating the genetic effects of radiation. This line of research had started after World War II, following the dropping of the atomic bombs and their devastating medical effects on local populations in Hiroshima and Nagasaki (Lenoir & Hays, 2000; Lindee, 1994). It had led to the reorientation of some of the personnel and research programmes of DoE-funded laboratories from physics to the life sciences. An example of this was Los Alamos National Laboratory, which after playing a leading role in the wartime race to develop the atomic bomb—it was the home of the flagship Manhattan Project—devoted a growing proportion of its mathematics expertise and computing resources to solve biological and medical problems.

Due to this, Los Alamos was chosen as the institution that would host the first centralised DNA sequence database in the USA—GenBank—in 1982 (Strasser, 2019, Ch. 5). A few months prior to the 1986 symposium at CSHL, DoE representatives organised a workshop in Santa Fe and subsequently announced a pioneering programme called the Human Genome Initiative.[7] As a result of this, the biomedical lines of research at two other DoE-sponsored institutions, the Lawrence Berkeley and the Lawrence Livermore National Laboratories, were strengthened and largely channelled towards technology development and genome-wide mapping and sequencing of human DNA. That the DoE network of national laboratories was equipped with personnel and infrastructures to conduct big science endeavours was a competitive advantage that favoured their early leadership in the incipient human genome work in the USA.[8]

How the DoE initiative converged with the NIH effort has been amply described in the literature (Cook-Deegan, 1994, Part Three; Hilgartner,

[7] The CSHL and Santa Fe meetings had been preceded by a workshop convened in 1985 by Robert Sinsheimer, a molecular biologist who was then chancellor of the University of California at Santa Cruz. Its participants cited multi-million dollar grants that the University had been awarded in the areas of particle physics and space science to argue for the necessity of a similar NIH investment in a human genome programme. No significant NIH move occurred until three years later (Sinsheimer, 1989).

[8] While the DoE's advocacy and pursuit of human genomics was motivated by the tracking of heritable radiogenic genetic mutations, they also needed to find new purposes for their national laboratories in the light of arms reduction treaties. In genomics, key DoE figures saw the potential for a big-enough science that would take the place of weapons development and make use of expensively-assembled facilities. This prompted the acerbic comment by David Botstein that the DoE's plans constituted a program for unemployed bomb-makers (Cook-Deegan, 1994, pp. 96–100, quote from p. 98).

2017, pp. 91-110). In 1988, two reports issued by the US National Academy of Sciences and the Office of Technology Assessment recommended a single national initiative that would initially focus on physical mapping and improving the existing instrumentation to create a platform for sequencing the human genome in the longer term. This led the DoE and NIH to merge their endeavours into the US-HGP, a 15-year programme that was launched in 1990 with a three billion dollar budget that was contributed towards by both agencies, the former through the Office of Health and Environmental Research and the latter through the National Center for Human Genome Research, an expanded version of Watson's Office that was later renamed as the National Human Genome Research Institute (NHGRI).[9] The explicit goal of the US-HGP was to produce a physical map and reference sequence of the whole human genome by 2005.

What we want to stress concerning the history of the US-HGP is how this initiative, and other contemporary human genome projects, disrupted the funding and organisational regimes of biomedicine. This disruptive effect has already been noted by scholars who have investigated the impact of big science and data-intensive approaches on different areas of contemporary biological and medical research (Leonelli, 2016; Stevens, 2013; Vermeulen, 2016).[10] With regard to genomics, Stephen Hilgartner has argued that it propelled a new "knowledge-control regime" that was distinct from existing disciplines, such as molecular biology. This regime constituted new categories of "agents, spaces, objects and relationships", and

[9] The National Center for Human Genome Research was established in 1989 and renamed as the NHGRI in 1997. Also in 1997, the DoE changed the name of its office to Biological and Environmental Research. Given that in subsequent chapters we refer to events that occurred under the new names, we use NHGRI throughout the book for ease of reading. Watson was appointed director of the National Center for Human Genome Research and remained in this position until his resignation in 1992. Elke Jordan, who had acted as day-to-day director of the NIH Office—while Watson was part-time associate director—became deputy director of the National Center until 2002, overseeing its transition to NHGRI and the early years of that institution.

[10] The existence and impact of data-intensive endeavours is not exclusive to the twentieth and twenty-first centuries. Historians have documented how expeditions to Asia, Africa and the Americas in the early-modern period led to the introduction of large amounts of new knowledge in Europe, and a feeling of "information overload" that was crucial for the emergence of natural history (Müller-Wille & Charmantier, 2012; Rosenberg, 2003). Building on this, Bruno Strasser (2019) has forcefully argued that the perspective of obtaining new knowledge through the collection, compilation and comparison of specimens and data about them has always existed in the life sciences and interacted in different ways with more experimental approaches.

allocated to them "entitlements and burdens" that led to novel ways of conceiving and disseminating knowledge (Hilgartner, 2017, p. 9).

Hilgartner's empirical work has focused on the US-HGP as an exemplar of new players—the genome centres—and new rules for processing, storing and sharing the data they produced. Crucially, the emergence of the knowledge-control regime of genomics was neither immediate nor uniform. It occurred gradually throughout the 1990s, with more intensity in some parts of the world than in others. The rest of this chapter emphasises the gradualism of the transformation within the US-HGP, and how other human genome programmes adopted different knowledge-control regimes. Some of these alternatives to the US-HGP, we argue, never converged with what Watson and his DoE colleagues advanced.

A challenge that the NHGRI faced was in transforming the funding culture of the NIH into a system that would enable large-scale mapping and sequencing. Like many other biomedical funders, NIH managers and administrators were used to issuing competitive calls for proposals and awarding grants across relatively large numbers of laboratories, following peer review of their applications. This differed from the DoE model, which rather than running a responsive grant mechanism would distribute their budget among a narrower cohort of recipients: its network of national laboratories. The DoE funding system had allowed the creation of a number of genome centres that prioritised the production of map and sequence data via the development of high-throughput technologies and the deployment of industrial modes of production. These genome centres were based in some of the DoE laboratories and had begun operating during the preceding Human Genome Initiative. Although Watson could not exempt the NHGRI from the NIH grant-award system, he established different, specific criteria when distributing US-HGP funds with the aim of fostering a similar type of operation to the DoE one.

The main criterion for NHGRI grants was whether the applicants and their home institutions could contribute to the establishment of a solid base of whole-genome mapping and sequencing centres. With this, Watson sought to avoid what he labelled the "cottage industry" approach, which he attributed to the sequencing of microorganisms (Watson, 1990, p. 45). This approach consisted in the formation of large inclusive consortia and required the distribution of resources as widely as possible among the communities working on the organisms to be characterised. Watson's attribution of "cottage industry" was

initially aimed at the sequencing of the bacterium *Escherichia coli*, but as the 1990s progressed, the EC's Yeast Genome Sequencing Project emerged as the most widely cited example of cottage industry genomics (e.g. Palca & Roberts, 1992, p. 957).

For Watson, the cottage industry approach presented several logistical problems when applied to larger genomes. Instead, the NHGRI sought to gradually form a small set of funding recipients with industrial mapping and sequencing capacities that were not necessarily interested in conducting research using the resulting data. This change of ethos, however, did not become fully implemented in the USA until the mid-to-late 1990s, partly due to the resistances it encountered among some quarters of the genetics community.[11]

During the early days of the US-HGP, the NHGRI administrator in charge of distributing genome mapping grants was Jane Peterson. She worked hard to persuade laboratories equipped with the appropriate technologies and expertise to broaden the genome areas they would tackle. Some of these laboratories featured long-established teams of medical geneticists that had historically focused on smaller regions of human chromosomes encompassing genes or genetic markers connected to diseases.[12] Examples of this were Victor McKusick and Frank Ruddle's groups, at Johns Hopkins University and Yale University respectively. These two scientists (Fig. 3.1) had pioneered the chromosome mapping workshops, forums at which geneticists from all over the world shared their mapping results.

Started in 1973 and continued annually or biennially until the release of the human reference genome, these workshops produced human genome maps with increasing numbers of genes and markers on them, and at improved resolution (Fig. 3.2).[13] They achieved this through the collation of multiple partial results: those reported by individual genetics

[11] See, for instance, Ayala (1987); Baltimore (1987). Some commentators, including reputed biomedical scientists, argued that the potential outputs of the US-HGP, in the form of a full human genome map and sequence, did not justify an expenditure that would curtail other areas of life science research.

[12] Jane Peterson, interview with Miguel García-Sancho, National Human Genome Research Institute (Bethesda, Maryland), November 2018.

[13] Until 1991, such meetings were called Human Gene Mapping Workshops. These were subsequently replaced by Single Chromosome Workshops under the auspices of the Human Genome Organisation (HUGO; see Chap. 4).

Fig. 3.1 Victor McKusick (left photograph, second seated from left) with fellow medical geneticist P. S. Gerald; and Frank Ruddle (right photograph, standing wearing a white shirt) surrounded by, among others, G. J. Darlington and R. S. Kucherlapati. They were all attending the first chromosome workshop, held at Yale University in 1973. Both pictures from: New Haven Conference (1974, pp. 209 and 211); copyright © 1974 Karger Publishers, Basel, Switzerland

groups working on a specific disease or set of diseases at given chromosomal locations. By collectively gathering and pooling these results, the workshops gradually covered broader areas of the chromosomes and populated them with an increased number of landmarks (Jones & Tansey, 2015). In 1987, building on the success and consolidation of this model, McKusick and Ruddle co-founded *Genomics,* a journal devoted to the publication of mapping results (Kuska, 1998; Powell et al., 2007, pp. 13ff). Yet in order to achieve the US-HGP goals, the NHGRI needed to fund institutions—rather than collectives—whose mapping went well beyond the contributions to the chromosome workshops or the results published in the articles of *Genomics.*

On the sequencing front, the NHGRI initially funded a small number of individual grants aimed at model organisms with relatively small genomes, such as *E. coli*, *C. elegans*, and *D. melanogaster*, as well as a number of yeast (*S. cerevisiae*) chromosomes (Chap. 2). Some, but not all, of these grants were among the first set of genome centre grants funded in 1990. Strategically, not only were those grants intended to contribute towards the completion of the sequences of their target organisms but, more importantly in the long term, to act as platforms for technology development and the creation of the infrastructures for the establishment of sequencing centres. In 1996, the NHGRI awarded a set of six grants as pilots for human genome sequencing; these projects had a minimum

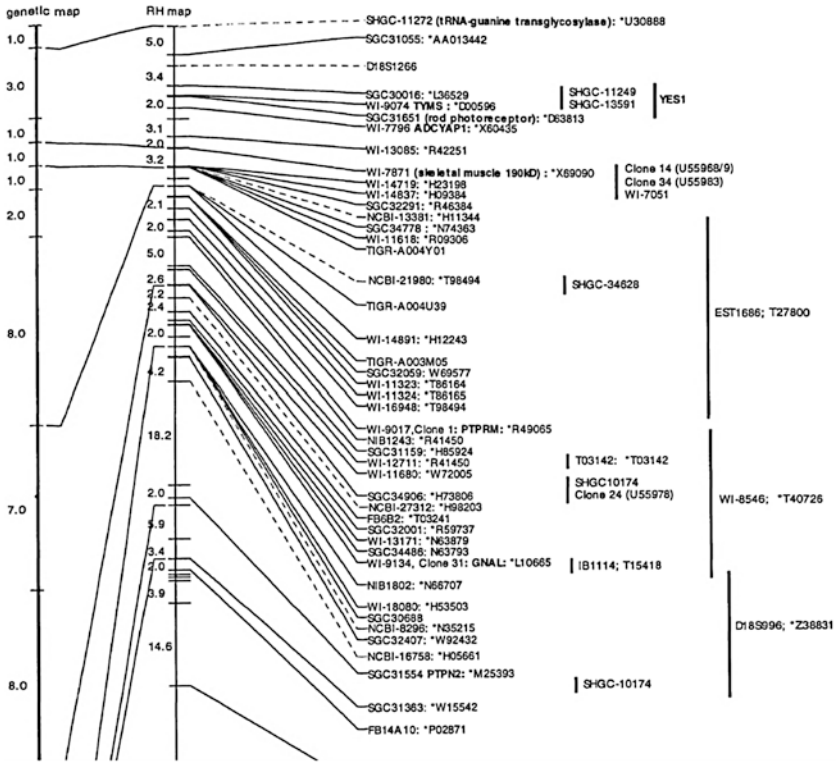

Fig. 3.2 Part of the genetic linkage map and physical map of human chromosome 18, as reported in the Fourth International Workshop devoted to its mapping, held in Boston (USA) in 1996. The genetic linkage map is displayed on the left of the picture and labelled as "Genetic map", with the physical "RH map" arrayed next to it (Radiation Hybrid—RH—maps are a form of physical map). From: Silverman et al. (1996), p. 119; copyright © 1996 Karger Publishers, Basel, Switzerland

target of sequencing one Megabase (one million nucleotides, or bases) of human DNA. With the information and experience gained in these pilots, the NHGRI scaled up its sequencing programme in 1999 with the funding of three Genome Sequencing Centers at Washington University, Baylor College of Medicine and the Whitehead Institute (of the Massachusetts Institute of Technology and Harvard University). At all

three of these sites, the sequencing centres were outgrowths of previously funded genome centres and pilot sequencing projects.[14]

A defining characteristic of those sequencing centres was that their funding and organisation prioritised the completion of their target genomes over any other scientific or medical objective, including the mapping of genes or markers associated with diseases. This form of operation was difficult to deploy beyond the USA. For example, in most European countries, governments had neither the resources nor the motivation to create specific grants for large-scale genome mapping and sequencing at dedicated centres. Private and charitable funds, by contrast, had fewer constraints and could be more easily channelled to a particular enterprise or group, as opposed to having to support a wider scientific community. This was the case for the Wellcome Trust, a British charity that teamed up with the MRC in 1992 to create the Sanger Institute, an institution that substantially contributed to the completion of the yeast, *C. elegans,* human and pig genomes (Chaps. 4 and 5). Another example of a charitably-funded genome centre was Généthon, supported by AFM-Téléthon, the French Muscular Dystrophy Association. Established in 1990, this institution was devoted to comprehensive mapping and quickly became a world leader in the production of genetic and physical maps encompassing the

[14] Mark Guyer, director of the NHGRI's extramural (grant-funding) programme; personal communication with Miguel García-Sancho, National Human Genome Research Institute (Bethesda, Maryland), November 2018. On the history of the Center for Genome Research at the Whitehead Institute, which became part of the Broad Institute, see García-Sancho, Leng et al., 2022. Its leader, Eric Lander, had a vision of genome mapping and sequencing that differed from the traditional ways of working of medical geneticists. The group behind the Human Genome Sequencing Center at the Baylor College of Medicine, by contrast, was furnished with a strong tradition in medical genetics research. See: Jim Lupski, "Applications of sequencing in clinical genetics", presentation delivered at *The Evolution of Sequencing Technology: A Half-Century of Progress* meeting, organised by the Genentech Center and Cold Spring Harbor Laboratory Archives in Long Island, 16th–19th July 2015. Available at: http://library.cshl.edu/Meetings/sequencing/video-pages/Lupski.php (last accessed 14th December 2022). Most scientists in this group have had double affiliations, also belonging to the Department of Molecular Genetics of Baylor College. This enables them to perform large-scale sequencing at the Genome Center alongside medical genetics research at the Department of Molecular Genetics.

entire human genome.[15] Généthon combined whole-genome work at its own initiative with a service role, attending to mapping requests from the French medical genetics community. This service role differentiated it from the US sequencing centres (Jordan, 1993, pp. 131ff; Kaufmann, 2004).

In spite of their influence, Généthon and the Sanger Institute were exceptional cases outside the USA. The US-HGP was rather unique in its commitment to full human genome mapping and sequencing when compared to programmes introduced by other governments, especially in Europe.[16] Those other programmes did not distinguish the human genome work they sponsored as sharply from medical genetics research as the US-HGP did. For this reason, they refrained from focusing on producing a reference map and sequence of the full human genome and were closer to the distributed, networked organisation that the EC was implementing in its sequencing projects. This distributed form of organisation was more suitable for fostering communication and tailoring the genome work to the regions of interest of the local medical genetics communities. The British HGMP was one of the earliest examples of this way of approaching the human genome.

[15] As we have shown elsewhere, Généthon was the institution that submitted the largest volume of human DNA sequence data to public repositories prior to 1996, well above any other laboratory, including the Sanger Institute and the US-HGP genome centres (Garcia-Sancho, Leng et al., 2022, Table 1, p. 334). This sequencing, however, was conducted to enable mapping work rather than to comprehensively characterise the human genome. In 1996, the publicly-funded French atomic energy commission (Commissariat à l'énergie atomique; CEA) created Genoscope, a sequencing centre that contributed to the elucidation of the human reference genome, albeit to a lesser extent than its US and British counterparts (Ramillon, 2007). The French CEA was also involved in early mapping and sequencing work on the pig *S. scrofa,* as we see later in the book (Chap. 5).

[16] Some Asian programmes pursued whole human genome sequencing. During the 1980s, Japan invested heavily in the automation of sequencing techniques and deployed an ambitious human genome project (Fujimura, 2000; Yoshikawa, 1990; Cook-Deegan, 1994, Ch. 15). Yet the Japanese DNA sequencing machines were never as popular as those manufactured in the USA and Europe, and Japanese institutions performed below the British and US genome centres, despite being involved in the human reference sequence. China created high-throughput sequencing centres—namely the Beijing Genomics Institute—but only joined the production of the human reference genome in the late-1990s (Wang et al., 2021). In the Americas, Canada created human genome programmes that were more in line with the HGMP and the EC (Dusyk, 2007).

3.2 The UK Human Genome Mapping Project

In 1989, one year before the launch of the US-HGP, the British Government authorised the release of 11 million pounds to fund the HGMP, a three-year programme that would be managed by the MRC.[17] The key proponents of this initiative were Brenner, a senior scientist who had just left the Laboratory of Molecular Biology of Cambridge (LMB) after a successful 30-year tenure, and Walter Bodmer, a reputed geneticist who coordinated the research laboratories of the medical charity Imperial Cancer Research Fund (ICRF).[18] Keith Peters, a practising physician with ample experience in teaching and researching immunology at London's Hammersmith Hospital, had presented the HGMP proposal on Brenner and Bodmer's behalf to the Advisory Committee on Science and Technology (ACOST). This body directly reported to the UK Prime Minister—in this case Margaret Thatcher—on projects that were likely to generate impact and required rapid funding. It approved the HGMP on Peters' recommendation and transferred the funds in less than one year (Balmer, 1996).

The prime mover behind the HGMP was Brenner. He had moved to Cambridge (UK) in 1956 to begin his research career, having recently concluded his PhD. Watson had also moved to Cambridge at the same stage in his career and returned to the USA the same year Brenner arrived in the UK. Brenner became the main collaborator of physicist-turned-biologist Francis Crick, who had successfully worked out the structure of DNA with Watson. Up to the early-1960s, Crick, Brenner and Watson focused on what became known as the coding problem: how the order of the nucleotides comprising DNA affects the synthesis of specific proteins that are responsible for most of the structural and functional aspects of the living cell (de Chadarevian, 2002, Part II; see also Kay, 2000).

In 1962, the same year Watson and Crick were awarded the Nobel Prize, the LMB was founded as an MRC-supported institution that would host an increasingly influential group of biologists in Cambridge. Crick became the director of the LMB Division of Molecular Genetics and

[17] According to the UK Retail Price Index measure of inflation, the equivalent sum as of November 2022 would be about 26.7 million pounds. https://www.bankofengland.co.uk/monetary-policy/inflation/inflation-calculator (last accessed 14th December 2022).

[18] In 2002, ICRF merged with The Cancer Research Campaign to form Cancer Research UK. A substantial part of its laboratories have now been amalgamated into the Francis Crick Institute in London, see https://www.crick.ac.uk/about-us/our-history (last accessed 14th December 2022).

Brenner started a long-term line of research, adopting the nematode worm *C. elegans* as a model to investigate the genetics of development and behaviour. This enterprise sought a detailed description of the worm's neuron circuitry, as well as its development from embryo to adult, with the hope of finding the "programme" that connected brain activity and cell differentiation to particular *C. elegans* genes.[19] The project included crossing experiments in which Brenner attempted to produce mutant worms and identify specific genes associated with variation in properties such as size or mode of movement, as geneticists had done with the fruit fly *Drosophila* and other organisms. Brenner also recruited more junior associates that would carefully detail the fates of every single cell throughout the *C. elegans* life cycle—its cell lineages—and the position and synaptic connections of each neuron in its brain.[20] To this end, John Sulston joined the LMB in 1969 to chart the multiple divisions of cells during the worm's embryonic and post-embryonic development (de Chadarevian, 1998).

By the time Brenner first proposed to map the human genome, in 1986, the worm project was experiencing a profound transformation. The description of cell lineages and brain connectivity had been completed by the early-1980s and a project to construct a physical map of its genome had started under the leadership of Sulston and Alan Coulson (Fig. 3.3). Coulson was a research assistant who joined the team after working at another LMB division on the development of early DNA sequencing techniques. Brenner, however, was becoming increasingly sceptical about the possibility of matching the detailed information his team had gathered

[19] See Brenner (1973, p. 271; 1974, p. 71). This language of programmes, circuitry and information flows had been mobilised by cybernetics after World War II and imported into molecular biology by various researchers, among them François Jacob and Jacques Monod at the Pasteur Institute in Paris. During the early-1970s, Jacob and Monod published popular accounts that further spread the use of cybernetic vocabulary in biology (Kay, 2000; Rheinberger, 2006). The Pasteur Institute played a major role in the EC's yeast genome effort (Dujon, 2019) and Mark Johnston, one of the leading yeast sequencing scientists in the USA, started his career with the *S. cerevisiae* GAL system (Chap. 2), a closely related gene expression system to the one Jacob and Monod had explored back in the 1960s.

[20] On the deeper history of cell lineage research, reaching back to the late-nineteenth and early-twentieth century, see Guralnick (2002); Lowe (2016); and Maienschein (1978, 1990). Concerning cell lineage research on *C. elegans*, see de Chadarevian (1998) and Jiang (2013). The ability to trace the lineages of adult cells back to cell divisions earlier in development—and therefore the fates of those earlier cells and divisions—provides the basis for precise experimental intervention, for example being able to assess changes wrought on the process of development and resultant outcomes by a mutation in a gene or genes.

Fig. 3.3 Left, Sydney Brenner with co-discoverer of the double helical structure of DNA, Francis Crick, at the Laboratory of Molecular Biology of Cambridge in 1962. Right, John Sulston holding a section of the physical map of *C. elegans* around 1985 (pictures of the nematode worm are pinned to the wall behind him). Copyright of left image: Hans Boye/MRC Laboratory of Molecular Biology. Copyright of right image: MRC Laboratory of Molecular Biology. Both reproduced with permission

about cell divisions and synaptic transmissions in *C. elegans* to the genes Sulston and Coulson would identify in their map, given the complexity of developmental processes in multicellular organisms (Lewin, 1984).

Partly because of this, in the same year of his human genome map proposal, Brenner left the LMB and established a Molecular Genetics Unit that, despite being also supported by the MRC, was part of the School of Clinical Medicine of the University of Cambridge. In this Unit, Brenner continued some work on the genetics of *C. elegans* but left the physical map to Sulston and Coulson, who remained at the LMB. The other lines of research in Brenner's Unit were the development of genome mapping technologies and "certain aspects of gene evolution".[21]

[21] Anonymous (1986) "Extract of minutes of the Council meeting held on Thursday 17th July 1986—Molecular genetics: proposal from Dr S. Brenner (MRC Laboratory of Molecular Biology) for a new Unit under his direction (86/C616; file E243/130)", National Archives of the UK at Kew (London), Medical Research Council Series, file number FD12/1191, quote from unnumbered page. Brenner had also concluded, by 1986, his tenure as director of the LMB and was approaching retirement age. To some especially distinguished scientists reaching this career stage, the MRC offered to create a more personally-managed research unit for them.

Brenner's proposal was entitled "A physical map of the human genome" and it was submitted in November 1986 to the Cell Board, the body of the MRC that funded genetics research. In his case for support, he argued that it was by then "not clear" whether the resources needed for a "central facility" to sequence the entire human genome "would ever be made available". This led Brenner to advocate for the construction of a physical map not only as a "first necessary step towards the grander sequencing proposal, but also for the more immediate benefits" it could bring "to medical research and practice". Brenner's vision started with a laboratory that would "carry out" the mapping programme and "act as the reference centre for human genetics". A "central concept" of his strategy was to establish "cooperative links and not enter into competition with individual research projects". In this regard, Sulston and Coulson's ongoing physical map of *C. elegans* provided a "useful benchmark" for Brenner's intended human mapping enterprise.[22]

At the time of this proposal, Brenner was serving on the committee of the National Academy of Sciences that advised the US Government on the plausibility and best strategy for conducting a human genome project. By late 1986, the discussions were still nascent and the model of tackling the entire human genome at dedicated and comprehensive mapping and sequencing centres had not yet attained majority support. Nevertheless, this comprehensive and concentrated strategy was gaining momentum in the USA. The physical mapping exercise that Brenner envisaged for the UK and the reference laboratory that would execute it differed in many respects with what became the US-HGP.

First, and contrarily to Watson, who also served in the committee, Brenner did not support a whole-genome operation. For Brenner, the size of the human genome—30 times bigger than *C. elegans*—meant that a comprehensive mapping and sequencing initiative would yield a substantial volume of data that would not correspond to genes. Biomedical scientists were well aware that only a small fraction of human DNA constituted genic regions, i.e., those directly involved in the synthesis of proteins. By the mid-to-late 1980s and early-1990s, a large proportion of those scientists—especially within the human and medical genetics communities—regarded the remainder of the genome as 'junk DNA': repetitive sequences

[22] S. Brenner (1986) "A physical map of the human genome", National Archives of the UK at Kew (London), Medical Research Council Series, file number FD23/3441, quotes from pp. 1 and 2.

that were expected to be non-functional.[23] Based on this common wisdom, Brenner argued that mapping and sequencing the entire human genome was not a worthwhile enterprise (Brenner, 1990). He, however, maintained his commitment to detailed descriptions of organisms that, due to their simpler developmental processes, could be used to model the molecular basis of life properties.[24]

Secondly, the reference laboratory that would channel Brenner's genome project was conceived to operate at the behest of human and medical geneticists. This was largely due to the framing of his proposal against the background of the ongoing physical mapping of *C. elegans*. Since 1983, Sulston and Coulson had mapped ever-increasing areas of the worm's genome by fulfilling requests of laboratories working on specific *C. elegans* genes. This had been mutually beneficial and ensured that the mappers were regarded as important, foundational members of the *C. elegans* research community: Sulston and Coulson crucially contributed to the objectives of this community, while increasing the resolution of their physical map (García-Sancho, 2012). The genome centres that Watson established for the US-HGP lacked this community service role: they mapped and sequenced comprehensively, at their own initiative rather than addressing requests from other laboratories. Although the genomes of *C. elegans* and *S. cerevisiae* were part of the remit of these large-scale centres, the US-HGP approached the mapping and sequencing of both

[23] Biomedical research communities who were less focused on human genes and their role in disease were more mindful of the importance of non-genic regions. This was the case for developmental biologists, with whom Sulston and Coulson collaborated during the mapping of *C. elegans* and for whom some non-genic DNA exerted a key role in inhibiting or activating the mechanisms of protein synthesis (Chap. 4). Up to the mid-to-late 1990s, it was believed that the human genome was formed of around 100,000 genes. Following the publication of the first draft of the reference sequence (2001), this figure was reduced to 20,000 to 25,000 and the estimated percentage of protein-coding regions lowered to 1.5% of the DNA in the human genome. On the origins of the term 'junk DNA', see: https://judgestarling.tumblr.com/post/64504735261/the-origin-of-junk-dna-a-historical-whodunnit (last accessed 14th December 2022).

[24] One of the flagship projects of Brenner's newly-established Molecular Genetics Unit was the full sequencing of *Fugu*, a pufferfish whose genome is characterised by a high-density of genic regions and a lack of repetitive sequences. Using comparative approaches, Brenner believed that *Fugu's* genome would provide insights concerning the sequence and protein synthesis mechanisms of human genes with a fraction of the effort of tackling the human genome in its entirety (Venkatesh, 2019).

organisms as a means of easing the path to human genome work rather than engaging with the research necessities of worm and yeast biologists.[25]

Thirdly, and as a consequence of the above, Brenner's project sought to involve the existing human and medical genetics groups rather than creating a new community of genome centres and specialist genomicists. After receiving Brenner's proposal, the MRC sounded out the opinion of reputed scientists and institutions in search of arguments for approval or rejection, as well as possible sources of co-funding. One of Brenner's first allies was Bodmer, who belonged to a group of geneticists that in the 1960s and 1970s had pioneered the mapping of a region of the human genome called the Human Leukocyte Antigen system (HLA).[26] This region contains densely-packed and hypervariable genes implicated in the immune response to infection; the variability of many of these genes aided their mapping (Löwy, 1987; see also Heeney, 2021). In his role of director of research at ICRF, which he took up in 1979, Bodmer equipped the charity's laboratories with cutting-edge DNA mapping and sequencing technologies (Weston, 2014, esp. Chs. 2-4). Another supporter of Brenner's proposal was Peters, who in 1987 moved from Hammersmith Hospital to the University of Cambridge due to his appointment as Regius Professor of Physic and Dean of the School of Clinical Medicine. From this position, he oversaw the establishment of Brenner's Unit in the school and saw human genome mapping as an opportunity to connect genetics research with medical goals.[27]

Peters had also become life sciences adviser in ACOST and suggested this committee—directly reporting to the Prime Minister's Office—as a potential source through which the MRC could obtain the necessary

[25] The genome centre at Washington University was headed by Sulston and Coulson's *C. elegans* collaborator, Robert Waterston. As we show in the next chapter, this collaboration and Watson's intervention were crucial for the redefinition of the worm genome effort from *à la carte* mapping to comprehensive, large-scale sequencing, and for Sulston and Coulson's institutional migration to a UK-based genome centre (Chap. 4).

[26] Another scientist who was heavily involved in the mapping of the HLA region was French immunologist Jean Dausset. In 1984, following the award of the Nobel Prize, Dausset established the Centre for the Study of Human Polymorphism (Centre d'Etude du Polymorphisme Humain), from which Généthon was created. One of Dausset's associates, Daniel Cohen, led the human genome work at Généthon and collaborated with pig geneticists in the mapping of the equivalent swine region: the SLA (Chap. 5).

[27] Keith Peters, two-part interview with Miguel García-Sancho: in person (October 2013) and by telephone (December 2013). Peters had previously attempted to persuade Brenner to move to Hammersmith Hospital.

funding for the human mapping project. In 1988, he formally endorsed Brenner's proposal and presented it to an audience that included Thatcher and her chief scientific advisor. He emphasised his experience as a practising physician and argued that the resulting physical map would become "the central tool for basic and applied research in the medical sciences".[28] ACOST agreed to support the initiative, which was subsequently named as the HGMP. This support materialised in an extra 11 million pounds that the Treasury transferred to the MRC as an earmarked fund to be exclusively spent in a Directed Programme of grants and a Resource Centre for human genome mapping. The funding was for a three-year period (April 1989 to April 1992) subject to extension following a progress review.

From its inception, the HGMP sought to build an identity that distinguished it from other human genome projects, especially the one that was already set to start in the USA. The US National Academy of Sciences had issued its report a few months before Peters' presentation to ACOST and, by 1989, the NIH and DoE's agreement to join forces in the US-HGP was being ironed out. Given the extraordinary budget and timeframe of the US effort—three billion dollars over 15 years—an early concern for the HGMP was how to make a differentiated contribution with a fraction of the money and a much more limited time horizon.

Tony Vickers, the HGMP manager, argued in his first report to the MRC in 1991 that in the UK there was "no individual enthusiasm" for becoming involved "in mega-sequencing", a task that was "unlikely to yield rewards to compensate workers for the drudgery involved". The British biomedical community, however, had "substantial strengths" in "many fields of genetics" where human genome mapping offered "promise of immediate and substantial pay-off". This short-term pay-off had somehow been "left aside" by the US-HGP with its focus on comprehensive, large-scale work at genome centres that were distant from the communities that would use the map and sequence data. The HGMP sought to take advantage of the prompt exploitation of results by involving the human and medical genetics communities in the mapping exercise.[29]

[28] K. Peters (1988) "Mapping and sequencing the human genome", typescript of presentation to Margaret Thatcher with additional manuscript notes (courtesy of Keith Peters), quote from p. 1. See also National Archives of the UK at Kew (London), Medical Research Council Series, file FD23/3442, and Bodleian Library (Oxford), Papers and Correspondence of Sir Walter Bodmer, file MS. Bodmer 1304.

[29] T. Vickers (1991) "The UK Human Genome Mapping Project: project manager's report", p. 4 (courtesy of Tony Vickers).

Consequently, the research grants awarded by the Directed Programme supported groups that were either developing mapping and sequencing technologies, creating shared resources to aid in these operations, or focusing on chromosomal regions connected to various types of genetic conditions, among them disorders affecting blood (haemophilia), mental health (aneuploidy syndromes) and muscular mobility (myotonic dystrophy). Of the five institutions in receipt of the largest amount of funding (Fig. 3.4), four of them investigated different aspects of medical genetics: ICRF, the Human Genetics Unit of the University of Edinburgh, the Institute of Molecular Medicine at John Radcliffe Hospital in Oxford and Guy's Hospital in London.[30]

The outcomes of the Directed Programme grants were delivered to the Resource Centre. This institution was housed in the Clinical Research Centre, a unit that the MRC had established in 1970 at Northwick Park Hospital (in northwest London) to foster collaboration between biomedical research and clinical practice. The Resource Centre was organised into two divisions that were headed by a biological manager (Ross Sibson) and a computing manager (Martin Bishop). Their duties involved assisting HGMP awardees in various capacities, from conducting mapping and sequencing work on request, to providing punctual support through their advanced technology and expertise (Balmer, 1998; Glasner, 1996). To do this, Sibson and Bishop's teams liaised with the so-called "user community", addressed their feedback and ensured access to the shared resources. They also collated the map and sequence data coming from the grant-supported laboratories.[31]

[30] T. Vickers (1991) "The UK Human Genome Mapping Project: project manager's report", Appendix, pp. 30ff; and T. Vickers (1992) "MRC Review of the UK Human Genome Mapping Project: Project Manager's Report", Annex 1, pp. 100ff (both courtesy of Tony Vickers). The contributions of ICRF, Guy's Hospital and the Institute of Molecular Medicine to the HGMP are detailed below. The Human Genetics Unit in Edinburgh housed a group with a longstanding tradition of cytogenetic mapping, based on chromosomal images that allowed the detection of malformations and helped diagnose diseases at the city's Western General Hospital (de Chadarevian, 2020). The institution with the largest share of HGMP funding was the LMB in Cambridge, although more than half of this support was awarded to Sulston and Coulson's *C. elegans* sequencing project (see the caption to Fig. 3.4).

[31] Ross Sibson, interview with Miguel García-Sancho, Royal Liverpool University Hospital, March 2014, see also: "The concept of a user", in T. Vickers (1991) "The UK Human Genome Mapping Project: project manager's report", pp. 14-16 (courtesy of Tony Vickers). The Resource Centre hosted a sociological investigation of users' experiences that was conducted by Peter Glasner, Harry Rothman and Wan Ching Yee, all of them social scientists who were working on the HGMP. The study was supported by funds that the different human genome programmes devoted to ethical, legal and social aspects of genomics research (Glasner et al., 1998).

PLACE	INSTITUTION OR DEPARTMENT	NO	VALUE
			(£KS)
BELFAST	MEDICAL GENETICS	1	5
CAMBRIDGE	GENETICS	2	692
	MEDICINE	1	83
	MRC LMB	7	2,011
	MRC MOLECULAR GENETICS UNIT	2	710
	PATHOLOGY	8	580
CARDIFF	GENETICS	1	44
DUNDEE	BIOCHEMICAL MEDICINE	1	35
EDINBURGH	CENTRE FOR GENETICS RESEARCH	1	106
	MRC HUMAN GENETICS UNIT	12	1,465
	BIOCOMPUTING RESEARCH UNIT	1	76
	MOLECULAR BIOLOGY	1	19
GLASGOW	GENETICS	1	28
	MEDICAL GENETICS	1	8
HARWELL	MRC RADIOBIOLOGY UNIT	7	635
LEICESTER	GENETICS	2	166
LONDON	GUY'S	6	1,015
	MRC HBGU AND GALTON LAB UCL	9	641
	IMPERIAL CANCER RESEARCH FUND	8	1,565
	IMPERIAL COLLEGE	4	399
	INSTITUTE OF CANCER RESEARCH	1	2
	INSTITUTE OF CHILD HEALTH	2	22
	MRC LEUKAEMIA UNIT	1	104
	NIMR	1	79
	ST MARY'S	6	747
	UNIVERSITY COLLEGE/MIDDLESEX	1	58
MANCHESTER	CELL AND STRUCTURAL BIOLOGY	1	89
	MEDICAL GENETICS	1	182
OXFORD	BIOCHEMISTRY	6	741
	DYSON-PERRINS (CHEMISTRY)	1	77
	REGIONAL GENETICS CENTRE	1	65
	IMM AND SURGERY	8	1079
	PATHOLOGY	2	125
SALISBURY	WESSEX REGIONAL GENRTICS CENTRE	1	49
SOUTHAMPTON	BIOCHEMISTRY	1	69
(TOTALS)		108	14,349

Fig. 3.4 The level of grant support per institution that the UK Human Genome Mapping Project had awarded by 1992, in thousands of pounds. Of the overall £2,011,000 that the LMB (Laboratory of Molecular Biology) received, just

(*continued overleaf*)

By 1991, a probe bank and a library of Yeast Artificial Chromosomes (YACs) were being transferred from their originators—all of them Directed Programme awardees—to the Resource Centre. The probe bank had been compiled by Nigel Spurr, a researcher at Clare Hall Laboratories, one of the ICRF divisions that Bodmer had equipped with the latest mapping and sequencing instruments during the 1980s (Weston, 2014, Ch. 4). It consisted of a series of DNA fragments whose known sequence enabled screening and the detection of specific chromosomal locations. The YAC library was a collection of human DNA fragments inserted in yeast cells and kept under controlled conditions in cultures. It was used as a source for chromosome mapping and derived from a collaboration between David Bentley at Guy's Hospital in London and Kay Davies at John Radcliffe Hospital's Institute for Molecular Medicine. Both scientists were renowned for applying genetics research to medical problems—presented by the patients of their home hospitals—and were regular recipients of HGMP funding.[32]

Fig. 3.4 over half (£1,150,000) went to co-fund the start of Sulston and Coulson's *C. elegans* sequencing project. The *C. elegans* grant was an outlier in the funding policies of the HGMP and, as such, is further examined in Chap. 4. Source: "Table 1: Distribution of HGMP awards (numbers and volume) amongst centres" in T. Vickers (1992) "MRC Review of the UK Human Genome Mapping Project: Project Manager's Report", p. 13. Report courtesy of Tony Vickers; Table 1 reproduced by kind permission of the Medical Research Council, as part of UK Research and Innovation

[32] N. Spurr (1990) "UK DNA probe bank: how it will function", *G-Nome News: the newsletter from the UK Human Genome Mapping Project,* number 3—February 1990, pp. 8-9, available online at https://groups.google.com/g/bionet.molbio.genome-program/c/dLgdQ83qTWY/m/9eqUC4Ic3DMJ (last accessed 14th December 2022). D. Bentley and K. Davies (1990) "Yeast artificial chromosome resources and genome mapping", *G-Nome News: the newsletter from the UK Human Genome Mapping Project,* number 7—Summer 1991, pp. 17-20, National Archives of the UK in Kew (London), Medical Research Council Series, file number FD7/2745. Only a few years later, in 1993, Bentley was chosen by Sulston to lead the mapping and sequencing of whole human chromosomes at the Sanger Institute, a UK-based genome centre that the MRC and the Wellcome Trust established (Chap. 4).

On top of housing and managing these shared tools, the Resource Centre started an in-house sequencing programme using complementary DNA (cDNA) methods. These methods allowed researchers to sequence only the DNA that is transcribed to produce messenger RNA, a vital step in protein synthesis. They therefore enabled the capturing of protein-coding genes in the DNA. The HGMP Directed Programme Committee decided, in 1990, that Sibson's division would apply this technique to "tissue"-specific and "developmental stage"-specific DNA, as well as the mapped fragments that the Resource Centre compiled from grant-awarded laboratories. This approach would produce "cDNA markers" that, combined with the ongoing physical map, would become "a valuable tool for researchers in human genetic disease". The cDNA component was adopted as a "strategy" aimed at yielding sequence information "in a relatively short time span", thus being "more practicable than mega sequencing of the human genome". It was regarded as a "flagship for the UK" and "essential" for achieving "international credibility" and taking "the lead" among the competing genome efforts.[33]

This mode of operation meant that the HGMP pursued a similar strategic approach to the EC's genome programmes. As the EC was doing for yeast and *H. sapiens* (Chap. 2), the MRC sought to involve existing genetics research laboratories in its genome project and distribute the HGMP grants among them as inclusively as possible.[34] This differed from the more selective funding regime of the US-HGP and the wider distance between the large-scale genome centres and genetics research institutions. More fundamentally, the two genome projects differed in their overall

[33] Anonymous: "Directed Programme Strategy Meeting held on 7th March 1990: discussion and development of a strategy by the Directed Programme Committee", National Archives of the UK at Kew (London), Medical Research Council Series, file number FD 7/2749, quotes from pp. 3 and 4. The main advocate for the cDNA strategy was Edwin Southern, a prominent HGMP grant awardee at the Oxford University Department of Biochemistry and inventor of the Southern Blot, a technique that allowed the probing of DNA fragments and the detection of certain variants among them, including mutations connected to diseases. He was supported by Duncan McGeoch, a genetic virologist based in the University of Glasgow and member of the HGMP Directed Programme Committee.

[34] The networks resulting from this distributed funding often overlapped, as in the HGMP Resource Centre participation in an international cDNA consortium sponsored by the EC's Human Genome Analysis Programme: T. Vickers (1991) "The UK Human Genome Mapping Project: project manager's report", p. 24 (courtesy of Tony Vickers). See also Chap. 2 on Horst Domdey and Brigitte Obermaier's Munich-based contribution to both international cDNA consortia and the EC's Yeast Genome Sequencing Project.

goals: whereas the US-HGP aimed for a reference sequence of the whole human genome—something that its much larger budget and timespan allowed—the HGMP restricted its remit to the genome regions on which its *user communities* were working. These human and medical geneticist users would develop catalogues of variation from the resulting mapped and sequenced regions.

3.3 Reference Sequence vs Catalogues of Variation

Historically, the production of a reference sequence of the whole human genome was not an objective of the human and medical genetics communities. These communities had indeed engaged in the mapping of the human genome and had done so at an increasing scale since the start of the chromosome workshops, in 1973. However, they had always limited the scope of their efforts to the regions of interest to the genome mappers: geneticists studying specific diseases or biological traits who pooled their results on the chromosomal locations of genes or genetic markers with other community members. The HGMP and the EC's Human Genome Analysis Programme (HGAP) had built on this collective endeavour and sought to foster it with ring-fenced funding, international networking and resource centres that provided technical assistance and shared mapping technologies, as well as cDNA sequence data. Yet, as the support of these programmes was tailored to human and medical geneticists, the mapping and sequencing results were constrained to the genes and markers they were pursuing, rather than covering the entire human genome.

Human and medical geneticists would deem these genes and markers to be mapped at sufficient resolution when they could be assigned to a precise DNA fragment. Once this happened, the fragment would often be sequenced and compared with equivalent genome regions. These comparisons were made between humans and closely-related non-human species, or between healthy individuals and patients suffering the condition with which the gene or marker was associated. The mapping and sequencing processes combined collaboration—at chromosome workshops and more specific groupings, often deploying cDNA techniques—with competition for being the first to determine the chromosome locus or sequence of a gene or marker. A source of inter-species comparison was the growing number of databases with map and sequence information from simpler

organisms, such as *S. cerevisiae* or *C. elegans,* that were being compiled through either their own specific programmes or as a result of funding from human genome efforts. In this regard, both the HGMP and HGAP supported the consolidation of mouse data repositories, an organism evolutionarily much closer to *H. sapiens* than yeast or a worm, and from which both medical and developmental inferences could be made.[35] To access data from patients, medical geneticists created consortia—some of them also sponsored by the HGAP (Table 3.1)—that enabled them to uncover genes involved in diseases and compile catalogues of genetic variants associated with the conditions.[36]

These catalogues of variation were often curated at hospitals with strong genetics departments. They formed repositories to which the rest of the community could contribute data, and from which they could access it. The HGMP Resource Centre and other similar central facilities that the HGAP developed shared this philosophy through the community-built and collectively-accessible probe banks, YAC libraries and map and sequence databases they offered to their users.[37] These shared resources were themselves the product of collaborative projects that the resource centres and genetics research laboratories jointly undertook with funding from the HGMP or HGAP (Table 3.1).

In the mid-1960s, before the arrival of DNA sequencing techniques, McKusick had pioneered these types of collections in *Mendelian*

[35] See the section "Mouse genetics", in T. Vickers (1992) "MRC Review of the UK Human Genome Mapping Project: Project Manager's Report", pp. 7-8 (courtesy of Tony Vickers). The mouse *Mus musculus* is furnished with a longstanding history of use as a model by both human and animal geneticists, due to its tractability in the laboratory and close evolutionary relatedness to larger mammals such as humans (García-Sancho & Myelnikov, 2019; Rader, 2004).

[36] While map and sequence data tended to be published, the methods to detect variants were often patented, so pharmaceutical companies could obtain licenses and produce diagnostic tests. Due to this, some human and medical geneticists approached the open access agenda for sequence data—that the US-HGP and other genome centres worldwide implemented during the mid-to-late 1990s—with reservations (Chap. 4).

[37] As Table 3.1 shows, the HGAP supported the establishment of three resource centres, two of them providing different libraries of human DNA and one acting as a centralised data platform. The European Data Resource was based in the Centre for Cancer Research at Heidelberg and the repositories of DNA libraries were housed at ICRF and Centre for the Study of Human Polymorphism, the institution from which Généthon emerged as a large-scale mapping centre in France (see note 26). HGAP funding also enabled the HGMP Resource Centre in Britain to host a database with international contributions of cDNA sequences.

Table 3.1 An example of a consortium of institutions pursuing medical genetics goals: the European Gene Mapping Project (EUROGEM), supported by the European Commission's Human Genome Analysis Programme (HGAP). The consortium included institutions involved in genome mapping activities and resource centres. None of these institutions participated in the determination of the human reference sequence nor in the whole-genome physical mapping that aided the sequencing (compare with Chap. 4, Table 4.1). Elaborated by Miguel García-Sancho and Jarmo de Vries, from data collected by Hallen and Klepsch (1995, esp. p. 20)

Institution	*Role in European Gene Mapping Project (EUROGEM)*
University of Marburg (Germany)	Mapping institution
University of Kiel (Germany)	Mapping institution
Cancer Research Centre at Heidelberg (Germany)	Resource centre (centralised data facility in Europe)
University of Aarhus (Denmark)	Mapping institution
Hospital Ramon y Cajal (Spain)	Mapping institution
Hospital de la Santa Creu i Sant Pau (Spain)	Mapping institution
Laboratoire de Genetique Moleculaire at Vert le Petit (France)	Mapping institution
Pasteur Institute (France)	Mapping institution
Centre for the Study of Human Polymorphism (France)	Mapping institution and resource centre (shared DNA libraries)
Institute of Molecular Biology and Biotechnology at Heraklion (Greece)	Mapping institution
University of Cagliari (Italy)	Mapping institution
University of Rome (Italy)	Mapping institution
University of Dublin (Ireland)	Mapping institution
University College Cork (Ireland)	Mapping institution
University of Leiden (Netherlands)	Mapping institution
University of Groningen (Netherlands)	Mapping institution
Universidade Nova de Lisboa (Portugal)	Mapping institution
MRC Human Genetics Unit at Edinburgh (UK)	Mapping institution
University of Cambridge (UK)	Mapping institution
University College London (UK)	Mapping institution
University of Wales (UK)	Mapping institution
St Mary's Hospital Medical School (UK)	Mapping institution
Imperial Cancer Research Fund (UK)	Resource centre (shared DNA libraries)
HGMP Resource Centre (UK)	Resource centre (cDNA sequence data bank)

Inheritance in Man, a catalogue of annotated chromosome maps that was first published as a series of printed volumes and later as an electronic database (*Online Mendelian Inheritance in Man*). Both the volume series and database incorporated updates with new data stemming from the chromosome mapping workshops and other disease-specific consortia, as well as clinical information about the underlying genetic conditions (Lindee, 2005, Ch. 3; Hogan, 2016, Ch. 3).

With the growth and development of physical mapping and sequencing techniques across the genetics community from the late-1980s onwards, both the workshops and variation catalogues became more specific: the former devoted to single chromosomes and the latter to individual diseases. An early example of this followed from the mapping of the cystic fibrosis gene in 1989, the first condition to be assigned to a physical location, in this case in human chromosome 7. One of the mapping scientists, Lap-Chee Tsui, was subsequently appointed as co-convenor of the chromosome 7 mapping workshops. Tsui also established the Cystic Fibrosis Genetic Analysis Consortium and coordinated the compilation of sequence variants connected to different forms of the disease that were determined by researchers all around the world. The results were gathered in a database that is still active at the University of Toronto Hospital for Sick Children—Tsui's home institution until 2004—and used to diagnose the condition.[38]

During the mid-to-late 1990s, Tsui's endeavour developed into a map encompassing the whole of chromosome 7. A younger member of the Toronto team, Stephen Scherer, built on the networks around the cystic fibrosis consortium and chromosome workshops to create a growing map with assignments associated with other conditions and *loci*. Scherer's collaborators included both medical geneticists and institutions working on the comprehensive mapping and sequencing of chromosome 7, among them the genome centre at Washington University. Yet the objective of Scherer's map was not to serve as a platform for the sequencing of the entire chromosome. Rather than pursuing a single reference sequence—as Washington University and the other genome centres did—Scherer and his fellow medical geneticists sought a way of detecting, mapping and cataloguing variation. Their map was a means of obtaining a set of ordered DNA fragments, some of which could be compared to data derived from patients. That way, differences in both fragment size and pattern, or

[38] See http://www.genet.sickkids.on.ca/ (last accessed 14th December 2022).

underlying DNA sequence, could be connected to particular conditions and assigned to specific chromosomal locations.[39]

The pursuit of variation by medical geneticists contrasted with other communities working on non-human organisms. Compared to the HGAP, the EC used a different strategy for yeast and sought a full reference sequence of its genome (Chap. 2). Apart from the extreme discrepancies in genome size, this divergent strategy was due to the aims and necessities of yeast geneticists, biochemists and cell biologists being distinct from those of the communities working on human DNA. While human and medical geneticists were interested in sequence differences underlying disease or other traits, the consortium of laboratories that undertook the EC's Yeast Genome Sequencing Project aimed to use this organism to model the functioning of the eukaryotic cell. Each community, therefore, approached its target genome in a different fashion. In the case of the human genome, the focus was on comparing specific regions—those where genes were located—across either different species or hospital patients versus controls. In the case of yeast, the laboratories in charge of the sequencing project used this organism as a "wild type" (Holmes, 2017) and pursued a standardised description of its genome, in order to relate the sequence data to functional aspects of cell genetics and metabolism. For this reason, they targeted a specific strain—S288C of *S. cerevisiae*—as representative of the yeast species as a whole and did not address variants until the full reference sequence was completed (Szymanski et al., 2019).

Similarly, within the history of molecular biology, substantial efforts had been devoted to achieve comprehensive descriptions of "exemplary" model organisms: viruses and bacteria first and further unicellular and multicellular organisms from the 1970s onwards (quote from Strasser & de Chadarevian, 2011; see also: Creager, 2002; Kay, 1993; Ankeny & Leonelli, 2020). The hope was that, as with the S288C strain of *S. cerevisiae*, those organisms would enable researchers to connect genes to different biological mechanisms and processes, and their effects. This, therefore, paralleled the goal of Brenner's *C. elegans* project, and Sulston's mapping and sequencing of the full genome of the worm. Like the yeast communities, molecular biologists would use the exemplary descriptions and descriptive models (Ankeny, 2000) as the basis of comparative practices.

[39] Elsewhere, we have referred to these two different approaches as horizontal sequencing—determining a single sequence representative of the entire human genome—and vertical sequencing: finding multiple variants in a specific genome region (García-Sancho, Leng et al., 2022).

Unlike *S. cerevisiae,* however, the reference sequence of *C. elegans* could not be traced to a specific population.[40]

Brenner considered the human genome to be too large and complex for an equivalent description to that being pursued for *C. elegans,* and so aligned with the human and medical genetics communities through the proposal of the HGMP. Yet, on the other side of the Atlantic, Watson found in the US-HGP the timeframe and resources needed to export the exemplary descriptive approach to the human genome. His genome centre model sought to fully describe the human genome as a standard or wild type, by producing a reference sequence rather than selectively tackling and comparing regions, as human and medical geneticists had traditionally done. This is what has led Hilgartner to identify Watson with a "vanguard" that consolidated genomics as an independent field, one that could be distinguished from other life sciences disciplines (Hilgartner, 2017).[41] In this differentiation, however, the large-scale centres that produced the reference sequence became both separated and distant from the genetics laboratories that would use the data and that were often involved in other forms of conducting genomics, more aligned with the approaches of the HGMP and the EC's programmes.[42]

The US-HGP dominates the historiography of genomics. As we have argued, however, its model of organisation was the exception rather than the rule during the formative years of genomics research. In the previous chapter, we conveyed the heterogeneous array of institutions, genomicists and organisational models involved in yeast genome sequencing. In this chapter, we have documented the diversity that also characterised human genomics. Taken together, both chapters show that the model of the US-HGP—with its large-scale centres and comprehensive sequencing regime—falls short in representing not only the history of genomics but also of the more specific subfield of human genomics (Fig. 3.5).

[40] Brenner specifically bred the *C. elegans* variant that was used in the descriptions of the cell lineages and neuron connectivity through complex genetic crossing experiments (Brenner, 1974). Yet this variant was never labelled or attributed to a specific strain, as in the case of yeast. In *H. sapiens*, the next chapter details the protocol by which the DNA to be included in the reference sequence was chosen, and later in the book we compare this process with the production of the yeast and pig reference genomes (Chaps. 4 and 5).

[41] In a similar vein, Michael Fortun has argued that genomics is nothing else than genetics research imbued with high-throughput technologies at accelerated speed (Fortun, 1999, esp. pp. 26-27).

[42] Elsewhere, we have identified different degrees of sequence production—from more proximate to more distal—and argued that there is considerable diversity and gradation across institutions that are outside the confines of the large-scale centre model (García-Sancho, Lowe et al., 2022).

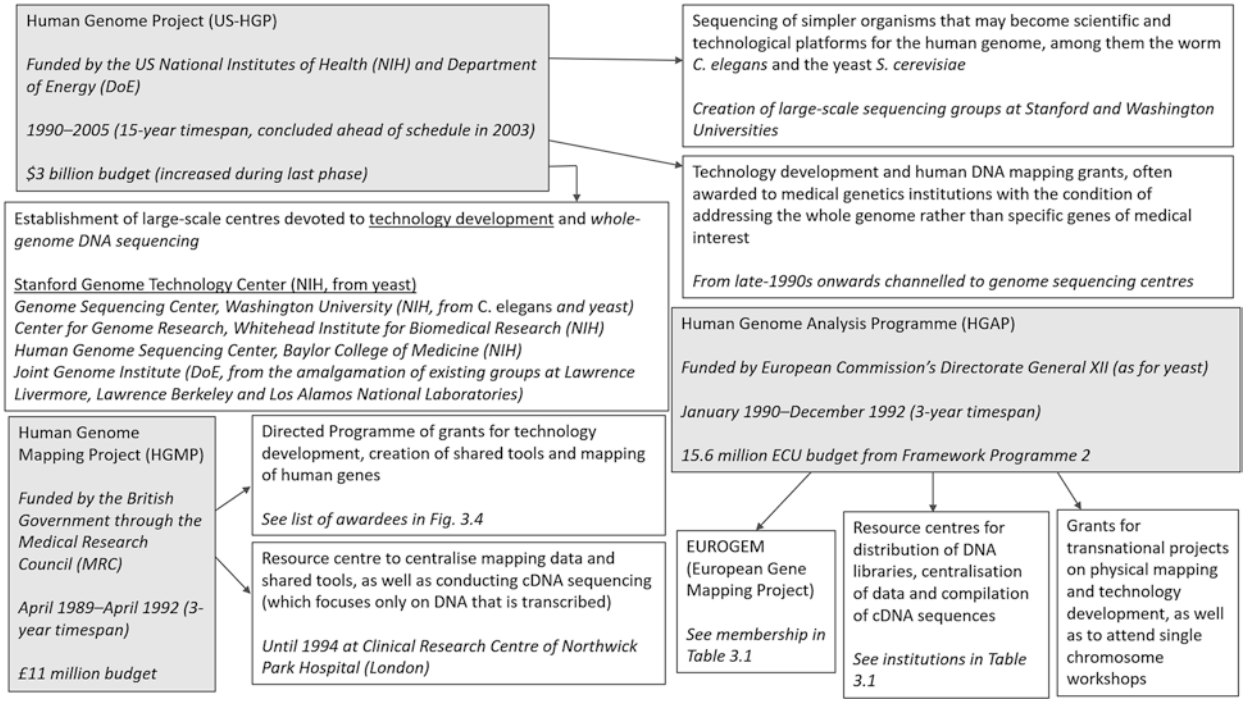

Fig. 3.5 An outline representation of the US Human Genome Project, UK Human Genome Mapping Project and European Commission's Human Genome Analysis Programme. Only aspects that have been discussed in the chapter are included and there are some notable absences, such as the programmes on ethical, legal and social aspects of genomics research that the three initiatives supported. Elaborated by both authors. For a larger version of this figure that can be zoomed in and out, see https://www.pure.ed.ac.uk/ws/portalfiles/portal/314798928/Fig_3_5_increased_final.pdf

In the next chapter, we identify the factors that led to a growing concentration of institutions and productive capacity during the determination of the human reference sequence. The transition of the *C. elegans* project from mapping to sequencing—along with the rise of the Wellcome Trust as an influential, proactive funder—spread the genome centre model beyond the USA and made it dominant in human reference genomics towards the mid-to-late 1990s. This process, we argue, not only affected scientific practice and organisation: it also occluded other historical trajectories in favour of the canonical winners' story based on the US-HGP.

References

Aicardi, C. (2016). Francis Crick, cross-worlds influencer: A narrative model to historicize big bioscience. *Studies in History and Philosophy of Biological and Biomedical Sciences, 55*, 83–95.

Ankeny, R. (2000). Fashioning descriptive models in biology: Of worms and wiring diagrams. *Philosophy of Science, 67*, S260–S272.

Ankeny, R., & Leonelli, S. (2020). *Model organisms*. Cambridge University Press.

Ayala, F. J. (1987). Two frontiers of human biology: What the sequence won't tell us. *Issues in Science and Technology, 3*(3), 51–56.

Balmer, B. (1996). The political cartography of the Human Genome Project. *Perspectives on Science, 4*(3), 249–282.

Balmer, B. (1998). Transitional science and the Human Genome Mapping Project Resource Centre. In P. Glasner & H. Rothman (Eds.), *Genetic imaginations: Ethical, social and legal issues in human genome research* (pp. 7–19). Ashgate.

Baltimore, D. (1987). A small-science approach. *Issues in Science and Technology, 3*(3), 48–50.

Brenner, S. (1973). The genetics of behaviour. *British Medical Bulletin, 29*(3), 269–271.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics, 77*(1), 71–94.

Brenner, S. (1990). The human genome: The nature of the enterprise. *CIBA Foundation Symposia, 149*, 6–17.

Cook-Deegan, R. M. (1994). *The gene wars: Science, politics, and the human genome*. WW Norton & Company.

Creager, A. (2002). *The life of a virus: Tobacco mosaic virus as an experimental model, 1930–1965*. University of Chicago Press.

de Chadarevian, S. (1998). Of worms and programmes: *Caenorhabditis elegans* and the study of development. *Studies in History and Philosophy of Biological and Biomedical Sciences, 29*(1), 81–105.

de Chadarevian, S. (2002). *Designs for Life: Molecular biology after World War II*. Cambridge University Press.

de Chadarevian, S. (2020). *Heredity under the microscope: Chromosomes and the study of the human genome*. University of Chicago Press.

Dujon, B. (2019). My route to the intimacy of genomes. *FEMS Yeast Research, 19*(3), 1.

Dusyk, N. (2007). The political moral economies of science: A case study of genomics in Canada and the United Kingdom. *Health Law Review, 15*(3), 3–5.

Fortun, M. (1999). Projecting speed genomics. In M. Fortun & E. Mendelsohn (Eds.), *The practices of human genetics* (pp. 25–48). Kluwer.

Fujimura, J. (2000). Transnational genomics: Transgressing the boundary between the modern West and the premodern East. In D. Reid & S. Traweek (Eds.), *Doing Science + Culture* (pp. 71–92). Routledge.

García-Sancho, M. (2012). From the genetic to the computer program: The historicity of 'data' and 'computation' in the investigations on the nematode worm *C. elegans* (1963–1998). *Studies in History and Philosophy of Biological and Biomedical Sciences, 43*(1), 16–28.

García-Sancho, M., Leng, R., Viry, G., Wong, M., Vermeulen, N., & Lowe, J. (2022). The Human Genome Project as a singular episode in the history of genomics. *Historical Studies in the Natural Sciences*, 52(3), 320–360.

García-Sancho, M., Lowe, J., Viry, G., Leng, R., Wong, M., & Vermeulen, N. (2022). Yeast sequencing: 'Network' genomics and institutional bridges. *Historical Studies in the Natural Sciences*, *52*(3), 361–400.

García-Sancho, M., & Myelnikov, D. (2019). Between mice and sheep: Biotechnology, agricultural science and animal models in late-twentieth century Edinburgh. *Studies in History and Philosophy of Biological and Biomedical Sciences, 75*, 24–33.

Glasner, P. (1996). From community to 'collaboratory'? The Human Genome Mapping Project and the changing culture of science. *Science and Public Policy, 23*(2), 109–116.

Glasner, P., & Rothman, H. (Eds.). (1998). *Genetic imaginations: Ethical, legal and social issues in human genome research.* Routledge.

Glasner, P., Rothman, H., & Yee, W. C. (1998). The UK Human Genome Mapping Project Resource Centre: A user analysis. In P. Wheale, R. von Schomberg, & P. Glasner (Eds.), *The social management of genetic engineering* (pp. 63–75). Routledge.

Guralnick, R. (2002). A recapitulation of the rise and fall of the cell lineage research program: The evolutionary-developmental relationship of cleavage to homology, body plans and life history. *Journal of the History of Biology, 35*(3), 537–567.

Hallen, M., & Klepsch, A. (Eds.). (1995). *Human Genome Analysis Programme.* IOS Press.

Heeney, C. (2021). Problems and promises: How to tell the story of a Genome-Wide Association Study. *Studies in History and Philosophy of Science, 89*, 1–10.

Hilgartner, S. (2017). *Reordering life: Knowledge and control in the genomics revolution.* MIT Press.

Hogan, A. J. (2016). *Life histories of genetic disease: Patterns and prevention in postwar medical genetics.* Johns Hopkins University Press.

Holmes, T. (2017). The wild type as concept and in experimental practice: A history of its role in classical genetics and evolutionary theory. *Studies in History and Philosophy of Biological and Biomedical Sciences, 63*, 15–27.

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860–921.

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature, 431*(7011), 931–945.

Jiang, L. (2013). *Degeneration in miniature: History of cell death and aging research in the twentieth century*. Arizona State University.

Jones, E. M., & Tansey, E. M. (Eds.). (2015). *Human Gene Mapping Workshops c.1973–c.1991: The transcript of a Witness Seminar held by the History of Modern Biomedicine Research Group*. Queen Mary University and Wellcome Trust.

Jordan, B. (1993). *Travelling around the human genome: An in situ investigation*. INSERM.

Kaufmann, A. (2004). Mapping the human genome at Généthon laboratory: The French Muscular Dystrophy Association and the politics of the gene. In J. P. Gaudilliere & H. J. Rheinberger (Eds.), *From molecular genetics to genomics: The mapping cultures of twentieth century genetics* (pp. 147–175). Routledge.

Kay, L. E. (1993). *The molecular vision of life: Caltech, the Rockefeller Foundation, and the rise of the new biology*. Oxford University Press.

Kay, L. E. (2000). *Who wrote the Book of Life: A history of the genetic code*. Stanford University Press.

Kevles, D., & Hood, L. (Eds.). (1992). *The code of codes: Scientific and social issues in the Human Genome Project*. Harvard University Press.

Kuska, B. (1998). Beer, Bethesda, and biology: How "genomics" came into being. *Journal of the National Cancer Institute, 90*(2), 93.

Lenoir, T., & Hays, M. (2000). The Manhattan Project for biomedicine. In P. R. Sloan (Ed.), *Controlling our destinies. historical, philosophical, ethical, and theological perspectives on the Human Genome Project* (pp. 29–62). University of Notre Dame Press.

Leonelli, S. (2016). *Data-centric biology: A philosophical study*. University of Chicago Press.

Lewin, R. (1984). Why is development so illogical? *Science, 224*(4655), 1327–1329.

Lindee, M. S. (1994). *Suffering made real: American science and the survivors at Hiroshima*. University of Chicago Press.

Lindee, M. S. (2005). *Moments of truth in genetic medicine*. Johns Hopkins University Press.

Lowe, J. W. E. (2016). Normal development and experimental embryology: Edmund Beecher Wilson and *Amphioxus*. *Studies in History and Philosophy of Biological and Biomedical Sciences, 57*, 44–59.

Löwy, I. (1987). The impact of medical practice on biomedical research: The case of human leucocyte antigens studies. *Minerva, 25*(1-2), 171–200.
Maienschein, J. (1978). Cell lineage, ancestral reminiscence, and the biogenetic law. *Journal of the History of Biology, 11*(1), 129–158.
Maienschein, J. (1990). Cell theory and development. In R. C. Olby, G. N. Cantor, J. R. R. Christie, & M. J. S. House (Eds.), *Companion to the history of modern science* (pp. 357–373). Routledge.
McLaren, D. (1991). *Human genome research: A review of European and international contributions*. Medical Research Council.
Müller-Wille, S., & Charmantier, I. (2012). Natural history and information overload: The case of Linnaeus. *Studies in History and Philosophy of Biological and Biomedical Sciences, 43*(1), 4–15.
New Haven Conference (1974). First International Workshop on Human Gene Mapping. 1974. *Cytogenetics and Cell Genetics, 13*(1–2), 1–216.
Palca, J., & Roberts, L. (1992). The Genome Project: Life after Watson. *Science, 256*(5059), 956–958.
Parolini, G. (2018). Building human and industrial capacity in European biotechnology: The Yeast Genome Sequencing Project (1989–1996). Technical University of Berlin preprint series. Retrieved December 14, 2022, from https://depositonce.tu-berlin.de//handle/11303/7470 (last accessed December 14 2022).
Powell, A., O'Malley, M. A., Muller-Wille, S., Calvert, J., & Dupré, J. (2007). Disciplinary baptisms: A comparison of the naming stories of genetics, molecular biology, genomics, and systems biology. *History and Philosophy of the Life Sciences, 29*(1), 5–32.
Rader, K. A. (2004). *Making mice: Standardizing animals for American biomedical research, 1900-1955*. Princeton University Press.
Ramillon, V. (2007). Le deux génomiques. Mobiliser, organiser, produire: du séquençage à la mesure de l'expression des gènes. PhD dissertation, École des Hautes Études en Sciences Sociales.
Rheinberger, H.-J. (2006). The notions of regulation, information, and language in the writings of François Jacob. *Biological Theory, 1*(3), 261–267.
Rosenberg, D. (2003). Early modern information overload. *Journal of the History of Ideas, 64*(1), 1–9.
Silverman, G. A., Overhauser, J., Gerken, S., Aburomia, R., O'Connell, P., Krauter, K. S., et al. (1996). Report of the fourth international workshop on human chromosome 18 mapping 1996. *Cytogenetics and Cell Genetics, 75*, 111–131.
Sinsheimer, R. L. (1989). The Santa Cruz Workshop—May 1985. *Genomics, 5*(4), 954–956.
Sloan, P. R. (Ed.). (2000). *Controlling our destinies: Historical, philosophical, ethical, and theological perspectives on the Human Genome Project*. University of Notre Dame Press.

Stevens, H. (2013). *Life out of sequence*. University of Chicago Press.
Strasser, B. J. (2019). *Collecting experiments: Making big data biology*. University of Chicago Press.
Strasser, B. J., & de Chadarevian, S. (2011). The comparative and the exemplary: Revisiting the early history of molecular biology. *History of Science, XLIX*, 317–336.
Szymanski, E., Vermeulen, N., & Wong, M. (2019). Yeast: One cell, one reference sequence, many genomes? *New Genetics and Society, 38*(4), 430–450.
Venkatesh, B. (2019). Sydney Brenner—A personal perspective. *Genome Research, 29*(6), vii–ix.
Vermeulen, N. (2016). Big Biology: Supersizing science during the emergence of the 21st century. *NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin, 24*(2), 195–223.
Wang, K., Shen, X., & Williams, R. (2021). Sequencing BGI: The evolution of expertise and research organisation in the world's leading gene sequencing facility. *New Genetics and Society, 40*(3), 305–330.
Watson, J. D. (1990). The Human Genome Project: Past, present, and future. *Science, 248*(4951), 44–49.
Weston, K. (2014). *Blue skies and bench space: Adventures in cancer research*. Cold Spring Harbor Laboratory Press.
Yoshikawa, A. (1990). Japanese biotechnology: Government, corporations, and technology transfer. *The Journal of Technology Transfer, 15*(1–2), 53–60.