

Annual Review of Genomics and Human Genetics
**Sharing Data to Build a Medical
 Information Commons: From
 Bermuda to the Global Alliance**

Robert Cook-Deegan,¹ Rachel A. Ankeny,²
 and Kathryn Maxson Jones³

¹School for the Future of Innovation in Society, Arizona State University, Washington, DC 20009; email: bcd@asu.edu

²School of Humanities, University of Adelaide, Adelaide, South Australia 5005, Australia

³Program in History of Science, Department of History, Princeton University, Princeton, New Jersey 08544



**ANNUAL
REVIEWS Further**

Click here to view this article's online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

Annu. Rev. Genom. Hum. Genet. 2017.
 18:389–415

First published as a Review in Advance on April
 17, 2017

The *Annual Review of Genomics and Human Genetics*
 is online at genom.annualreviews.org

<https://doi.org/10.1146/annurev-genom-083115-022515>

Copyright © 2017 Robert Cook-Deegan et al.
 This work is licensed under a Creative Commons
 Attribution 4.0 International License, which
 permits unrestricted use, distribution, and
 reproduction in any medium, provided the original
 author and source are credited. See credit lines of
 images or other third party material in this article
 for license information.



Keywords

patents, data sharing, sociology of science, model organisms, knowledge commons, science policy

Abstract

The Human Genome Project modeled its open science ethos on nematode biology, most famously through daily release of DNA sequence data based on the 1996 Bermuda Principles. That open science philosophy persists, but daily, unfettered release of data has had to adapt to constraints occasioned by the use of data from individual people, broader use of data not only by scientists but also by clinicians and individuals, the global reach of genomic applications and diverse national privacy and research ethics laws, and the rising prominence of a diverse commercial genomics sector. The Global Alliance for Genomics and Health was established to enable the data sharing that is essential for making meaning of genomic variation. Data-sharing policies and practices will continue to evolve as researchers, health professionals, and individuals strive to construct a global medical and scientific information commons.

INTRODUCTION

Genomics has data in its DNA. The term genomics took root in 1987 when Victor McKusick and Frank Ruddle borrowed Tom Roderick's neologism to launch a new journal (98). Genomics has since become a field, or at least an approach to biology and biomedical research. It generally describes a style of science and application that analyzes many genes (and genomic elements) at once, rather than one at a time; intensive use of instruments for mapping and sequencing nucleic acids; generation and utilization of large data sets, including DNA sequences, their underlying genomic locations, and functional analyses of genes and proteins; and computation. The extension from more traditional genetics entailed greater scale and implied a need to share data, because no one laboratory could fully make sense of the deluge of data being generated by mapping and sequencing the genomes of humans and other organisms.

The Human Genome Project (HGP) was conceived in 1985, when Robert Sinsheimer, Renato Dulbecco, and Charles DeLisi independently realized that it would be useful to have a reference sequence of the human genome as a tool for research and application (39, 96, 112). DeLisi had the authority to fund a research initiative within the US Department of Energy (DOE). As vigorous debate proceeded through 1988, the goals of the project and affiliated efforts were broadened to include—beyond creating a reference sequence of the human genome—mapping and sequencing the genomes of several model organisms: *Escherichia coli* (a bacterium), *Arabidopsis thaliana* (a plant), *Saccharomyces cerevisiae* (a yeast, with several other species added later), *Caenorhabditis elegans* (a nematode), *Drosophila melanogaster* (the common fruit fly), and *Mus musculus* (the house mouse, as the mammalian model). The envisioned project included research to advance—by increasing speed, reducing costs, and improving accuracy—the instruments and methods for DNA mapping and sequencing, as well as computational methods and algorithms. Finally, the project also incorporated research on the ethical, legal, and social implications (ELSI) of genomic advances, a novel and distinctive program that set a precedent for initiatives that had foreseeable impacts well beyond the technical community (172). It appears today that the greatest challenges of data sharing (as discussed below) are indeed in law, ethics, and policy. The technical challenges are daunting; the ethical, legal, and social complexities are even more so.

The HGP left many legacies. The most obvious is a reference sequence of the human genome that continues to be refined. Yet the project also drove the creation of new technologies and instruments. A major avenue has been the Advanced DNA Sequencing Technology Development Program, for many years directed by Jeff Schloss of the National Human Genome Research Institute (NHGRI), which helped conceive and develop many of the technologies that gave rise to the hyper-Moore's-curve acceleration in DNA sequencing speed and plummeting costs from 2004 to 2014 (48, 72). This program supported new mathematical approaches to the bioinformatic analysis of data, and the NHGRI became known as an institution that could manage large-scale, technology- and data-intensive research programs that required more coordination than could be provided by the distribution of individual peer-reviewed grants. One of the signature features of the NHGRI has always been an open science ethos associated with the prepublication sharing of data (37, 64).

The HGP reached its goal of generating a human genome reference sequence earlier than predicted, in the period from 2000 to 2003. The initial milestone toward completion was achieved on June 26, 2000, when a draft human genome sequence was first announced by US president Bill Clinton and UK prime minister Tony Blair (96, 143). Articles describing the HGP's public genome assembly and the genome assembly produced by a private company, Celera Genomics, were then published one day apart in *Nature* and *Science*, respectively, in February 2001 (89, 168). This initiated a cascade of publications on sequences of individual human chromosomes that

culminated in April 2003, marking the 50th anniversary of the canonical publication describing the double-helical structure of DNA by James Watson and Francis Crick on April 25, 1953 (33, 83, 105, 173, 174).

The 2001 publications represented two scientific strategies and two modes of data management. The publicly funded HGP sequence—authored by a group of laboratories from the United States, the United Kingdom, France, Japan, Germany, and China under the banner of the International Human Genome Sequencing Consortium—represented the work of a global coalition. It entailed extensive and systematic data sharing, characterized perhaps most distinctly by the daily release of new, publicly funded DNA sequences into the public domain. Data from the HGP came primarily from high-throughput sequencing laboratories in the six partner countries, with leadership for overall assembly centered at the University of California, Santa Cruz, and databases at the US National Center for Biotechnology Information and the European Bioinformatics Institute. This public hierarchical sequence published in *Nature* depended on genetic and physical maps, which situated DNA markers based on their chromosomal locations, and sequencing coordinated by chromosomal region, with assignments distributed to the various global partners (26, 161, 171). *Science*, by contrast, published a human genome sequence assembled by shotgun methods, characterized by Celera's vertical organization that integrated data generation and computation within a single laboratory and also incorporated the public HGP data (169, 170).

Access was open and free for the HGP-produced data, but there were restrictions for the data generated by Celera Genomics. Celera's data were available for a fee to subscribers, for free for noncommercial use in parcels of up to 1 Mb, or by access agreement with the company (178). Divergent access policies were thus present even during the gestational phase of generating the human genome reference sequence, and controversy over access was intense (3, 69, 84, 101, 115). It is not an exaggeration to say that debates about sharing have been woven into the fabric of genomics itself, an inextricable part of the new field from its very beginning. In 2003, a US National Research Council committee was appointed to make recommendations about responsibilities accompanying publication, partly in reaction to the divergent approaches adopted by the HGP and Celera. Chaired by Nobel laureate Tom Cech, this committee laid out the Uniform Principle for Sharing Integral Data and Materials Expeditiously (UPSIDE), and various other committees and reports on data sharing have followed since (28, 42, 104–108). The strongest embodiment of the HGP's open science ethos, however, was not sharing sequences at publication, but rather doing so long before, even as the data were being generated.

THE EMERGENCE OF THE BERMUDA PRINCIPLES FOR PREPUBLICATION DATA SHARING

By early 1996, the HGP had been under way for nearly six years. A workable human genetic linkage map was available, and physical maps of cloned DNA (associated with the bacterial and yeast clones that housed them) were available for most regions of human chromosomes (80, 91, 100, 114, 140). Reference genome sequences for several yeast chromosomes were published, and a complete yeast genomic sequence was in the offing, with a network of European laboratories leading the way (7, 61). Progress toward a complete 19,000-gene, 97-Mb genome sequence of *C. elegans* bred confidence that it would soon be completed, as indeed it was in 1998 (24).

However, the initial (and centerpiece) goal of the HGP—to produce a reference sequence of the human genome—was more distant, although it was visible over the horizon. The Wellcome Trust—the British biomedical research charity that by 2003 had funded one-third of the public HGP—had already increased its funding substantially, focusing on human genome sequencing and basing its efforts at the new Sanger Centre (founded in 1993) near the University of Cambridge

(45, 56). The US National Center for Human Genome Research (NCHGR, which would become the NHGRI in 1997) announced in 1995 a competition for high-throughput sequencing of parts of the human genome and reviewed center-grant proposals later that year (91). The DOE, moreover, was constructing the Joint Genome Institute to house its genome-sequencing operations, which would open in 1997 (87). By early 1996, the HGP was transitioning from mapping to large-scale human genome sequencing, with contributions from five national partners. China would become the sixth in 1999 (130). The stage was set for a hard push to generate the human genome reference sequence.

It was quickly becoming apparent that robust coordination was required. Even after earlier mapping efforts, sequencing—of both the human genome and other large model genomes—would generate more data than any other discrete effort in the history of biology (112). This had been predicted in the HGP's founding reports and by 1996 was becoming a tangible reality. More specifically, the pragmatic tasks at hand included (*a*) deciding which human chromosomal regions to assign to each group, lest several centers sequence the same juicy bits and waste time and effort; (*b*) specifying targets for data quality as the sequencing goals were set; and (*c*) verifying sequencing outputs from individual laboratories, to measure progress and to help justify the project's large public investment.

As the US National Institutes of Health (NIH) large-scale genome-sequencing grants were about to commence, Michael Morgan at the Wellcome Trust and NCHGR director Francis Collins decided to hold an organizational meeting among those funded (or expected to be funded) to plan the launch phase of large-scale human genome sequencing (96). They looked for a relatively neutral venue, one that would not be perceived as dominated by the United States (151). They settled on Bermuda, which was accessible from all of the HGP centers and appropriately located in the mid-Atlantic between Europe and North America. The meeting took place at the so-called Pink Palace, the Princess Hotel in Bermuda, from February 26 to 28, 1996 (**Figure 1**).



Figure 1

Participants at the first Bermuda meeting, February 1996. Photograph courtesy of Richard Myers, HudsonAlpha Institute for Biotechnology, on file with Duke University Libraries (<http://hdl.handle.net/10161/7713>).

The weather was dreary, but the conference was not (96). The new NIH sequencing centers had been announced but had yet to receive their funding. The attendees—which included the Hinxton group, funded by the Wellcome Trust; the leaders of the DOE’s efforts; the leaders of the NIH-funded centers; and mappers, technology developers, and administrators from Japan, Germany, and France—were eager to get started but anxious about the long journey ahead. They were, frankly, not sure a human genome reference sequence was yet attainable (21, 30, 151).

In addition to the practical needs to allocate work and measure progress toward a human genome sequence, two intimately related issues loomed as the 1996 meeting was planned: patenting and data sharing. The patent issue surfaced when a scientist at the National Institute of Neurological Disorders and Stroke, J. Craig Venter, filed patent applications on short segments of DNA sequence that allowed for unique identification of sequences coding for genes in the human brain (2, 25, 135). Expressed sequence tags (ESTs) could be used to fish genes out of the genome by identifying unique sequences that are translated into protein. Genentech patent lawyer Max Hensley advised Reid Adler, the lawyer at the NIH’s Office of Technology Transfer, that the NIH should file for patents on Venter’s ESTs to ensure that they could be licensed for further development (39). Such patents could “protect” the DNA fragments and might be important to preserve private investments in characterizing the corresponding full-length genes and related proteins (5). These incentives might be important for developing drugs and biologics as treatments as well as genetic tests for neurological diseases. Controversy had erupted in July 1991, when at a Senate meeting Venter made public that the NIH had filed its first application and was planning others, making claims on more ESTs and the method for obtaining them (135). The EST-generating methods patent was later converted to a statutory invention registration, effectively preventing anyone from patenting it (39).

Patenting DNA molecules and methods had become common at universities and biotechnology and pharmaceutical companies, a practice that in the 1980s and 1990s often conflicted with the ideas of more traditionally minded academic biologists, who were not used to patenting their work (40, 142). One of the foremost patent scholars, Rebecca Eisenberg of the University of Michigan Law School, observed,

[T]he patent system rests on the premise that scientific progress will best be promoted by conferring exclusive rights in new discoveries, while the research scientific community has traditionally proceeded on the opposite assumption that science will advance most rapidly if the community enjoys free access to prior discoveries. (50, p. 742)

According to some, the NIH patent applications—which claimed both ESTs (gene fragments) and corresponding full-length genes [in the form of complementary DNAs (cDNAs)]—could block downstream research and development requiring the use of many genes in tandem. NIH director Bernadine Healy supported the patents, although the NCHGR director at the time, James Watson (of DNA structure fame) vigorously disagreed with them (73, 172). This became one of several bones of contention between Watson and Healy that culminated in Watson’s resignation as head of the NCHGR in spring 1992, leaving the door open for Healy to recruit Francis Collins, who assumed the leadership of the NIH’s genome sequencing efforts in 1993 (152).

Reflecting contemporary commercialization trends, the debate over gene patents echoed between the public and private sectors. As the NIH’s applications were pending, Venter left the NIH to direct The Institute for Genomic Research (TIGR), a private nonprofit research institute (39, 143). Some of the patent rights from TIGR would be assigned to a for-profit corporation, Human Genome Sciences, which itself began sequencing genes and genome fragments and filing for its own patents, while also drawing on TIGR’s output. Meanwhile, another small firm,

Incyte, had also become interested in sequencing ESTs and full-length cDNAs. Human Genome Sciences, Incyte, and several other companies were building business models around discovering and sequencing genes and genome fragments and patenting parts of the genome likely to contain sequences of keen biological interest and commercial value (113).

Concerns about the patent impediments to research that would be created by thickets of broad EST and cDNA patents on genes of unknown function haunted debates about genomics and patent policy (51). In 1993, the pharmaceutical giant Merck initiated a partnership with the head of the HGP center at Washington University in St. Louis, the *C. elegans* expert Robert Waterston (46, 179). The goal of the Merck Gene Index project was to sequence human ESTs and release them into the public domain with a minimal delay (usually of only 48 hours), with the logic that “making the EST data freely available should stimulate patentable inventions stemming from subsequent elucidation of the entire sequence, function and utility of each gene” (179, p. 118). One reason for this policy was a spirit of open science; another was to thwart patents on short DNA fragments by companies like Human Genome Sciences and Incyte; and a third was that, because the Gene Index was funded by a nonprofit unit of Merck, the company had to demonstrate that it did not have privileged access. Meanwhile, Harold Varmus soon became NIH director, appointed by Bill Clinton to replace Bernadine Healy. Varmus sought expert advice on the NIH’s EST patent applications, which had been initially rejected by the US Patent and Trademark Office in 1992 (53, 93). Rebecca Eisenberg and another legal scholar, Robert Merges from the University of California, Berkeley, drafted a detailed memo for Varmus noting that the NIH’s EST patent strategy made little sense, given that ESTs were primarily research tools (53). Varmus abandoned the applications. This was yet another turn along the tortuous path whereby genomics and the rules about how and when genomic data should be shared and commercialized were developing in tandem (75, 76, 85, 86).

The brouhaha over ESTs, coupled to the evolving controversies and worldwide negative press over the patenting of full-length genes like *BRCA*, colored the Bermuda meeting in 1996 (13, 62). The Bermuda attendees, who hailed from five nations, the European Molecular Biology Laboratory in Heidelberg, and the Brussels-based European Commission, had to contend with the pragmatic, scientific problems of sequencing the human genome on time and accurately while also addressing the more principled problem of how the HGP’s impending deluge of sequence data should be shared, utilized, and commercialized. For the majority of attendees at the 1996 meeting, it was not surprising that developing a clear-cut data-sharing policy was a chief agenda item from the start (103).

The specific historical roots of the daily online release of HGP-funded DNA sequences—under the policy that came to be known as the Bermuda Principles—are complex and contingent. So, too, was the process by which this radical policy, which ran against the norm in most biomedical research of releasing data at publication, was ratified within the HGP and justified as the project proceeded. In a forthcoming historical article (95), we enumerate and assess these details at length. Several points about precedents and the measured agreement that HGP participants reached are relevant here.

The Bermuda Principles filled a policy lacuna, replacing a set of guidelines that had previously applied only to the NIH and DOE. The HGP’s founding reports, produced in 1988 by the US National Research Council and the congressional Office of Technology Assessment, were notoriously vague about data sharing, noting only that data and materials must be shared rapidly for coordination and quality control, and admitting that this might create conflicts with commercialization (23, 109, 112). The 1990 joint plan for the NIH and DOE echoed this message but similarly failed to provide a timeline for sharing among collaborators (26, 50, 66, 123, 145, 147, 161, 172). By the early 1990s, mapping and sequencing technology development were proceeding

impressively around the globe, not only in the United States but also in Great Britain, France, Japan, and several other nations (23, 109, 112). Yet despite the founding of the Genome Data Base (an electronic medium for sharing mapping data) in 1993, international coordination of human genome mapping was disorganized, nucleated around annual meetings and single chromosomes, and complicated by competition and secrecy (20). A 1992 NIH and DOE policy had required depositing data from mapping (to the Genome Data Base) and sequencing (to GenBank) efforts within six months of generation (20). Aside from the 48-hour data-sharing policy of the Merck Gene Index project, however, before the 1996 Bermuda Principles the NIH-DOE regulation was probably the most specific policy precedent for data sharing in genomics. Standard practice for GenBank, for instance, was to share unpublished sequences concurrently with or after the publication of accompanying journal articles, but not before (11, 145, 150).

The Bermuda Principles extended the 1992 policy, strongly recommending the daily sharing of all HGP-generated DNA sequences of 1 kb or longer through GenBank, the European Molecular Biology Laboratory data bank, or the DNA Databank of Japan (103). For the first time, the HGP had a project-wide policy, designed to unite all of the international contributors rather than just those funded by the NIH or the DOE. This facilitated the development of the Human Sequencing and Mapping Index, a website that linked laboratory webpages to GenBank and allowed globally distributed centers to declare genomic regions for sequencing and avoid duplication (18). Especially in the United States, the Bermuda Principles also helped to enforce quality standards (set in 1996 at 99.99% sequence accuracy) and output commitments, providing a means of checking whether the heavily funded centers were delivering on their sequence-generation promises. In 2001, Elliot Marshall of *Science* called the principles “community spirit, with teeth,” and for good reason: Centers that did not produce their sequences or failed to meet quality standards could lose their competitive funding and potentially their places within the prestigious HGP (92).

Daily sharing, however, was not necessarily required for these tasks, and the clearest historical test case for this policy was a perhaps unlikely source: the *C. elegans* research community. John Sulston and Robert Waterston, among many others in this network, had adopted daily (or as close to daily as possible) sharing in the mapping and early sequencing of the *C. elegans* genome beginning in the 1970s, a practice that was further facilitated in the 1980s by the rise of networked computing (11, 148, 151). By 1995, Sulston and Waterston (funded by the NIH, the Wellcome Trust, and the UK Medical Research Council) had done more large-scale genome sequencing (first in *C. elegans* and later in early human efforts) than anyone, and had become HGP leaders via connections to Francis Collins, James Watson, Maynard Olson, and other power players in the field (91, 180). At the first Bermuda meeting, Sulston and Waterston co-chaired the final session, on data release policies, which led to the first draft of the Bermuda Principles (151) (**Figure 2**). The statement from that session and a later NCHGR press release read,

It was agreed that all human genomic sequence information, generated by centres for large-scale human sequencing, should be freely available and in the public domain in order to encourage further research and development, and to maximise its benefit to society. (176; see also 102, 103).

Daily data release, however, was for some a bitter pill to swallow. The adoption of this policy in the HGP was driven by Sulston and Waterston, as the *C. elegans* genome-sequencing leaders, and strongly supported by leaders of the two foremost funders: the NIH and the Wellcome Trust. Debates raged about whether daily data release would hurt HGP data quality (4, 17, 140). Perhaps the most significant hurdle for daily data release, however, was its apparent incompatibility with commercialization. The US Bayh-Dole Act allowed universities and businesses the first right to title on inventions funded by government grants, and a German policy allowed HGP investigators

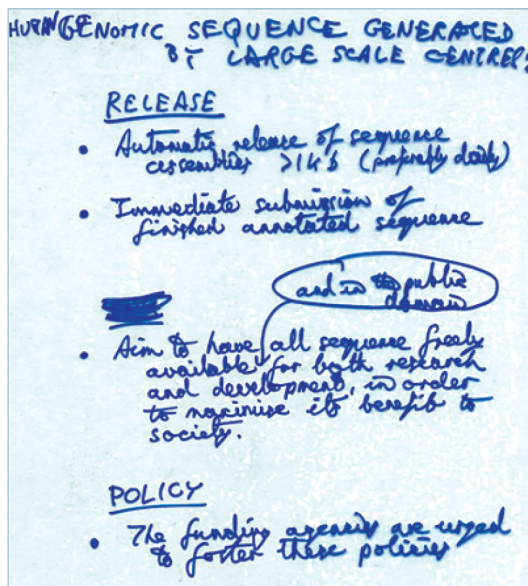


Figure 2

The white board on which John Sulston scribbled the Bermuda Principles at the 1996 meeting's final session. Robert Waterston was leading the discussion, and there was an informal vote to adopt the statement. Photograph courtesy of Richard Myers, HudsonAlpha Institute for Biotechnology, on file with Duke University Libraries (<http://hdl.handle.net/10161/7721>).

three months' time before data release to apply for patents drawing on HGP sequences, including patents on genes (40, 160). In the United States, daily release did not prevent patents on genes of known function, but it did block patents of the kind that Human Genome Sciences and Incyte were seeking. In Europe and other non-US patent jurisdictions, the Bermuda Principles indeed endangered gene patents, because unlike in the United States, there was no lag between data release into the public domain and the time when investigators could still apply for patents drawing upon it. [Until the America Invents Act somewhat changed the rules in 2010, the grace period in the United States was one year (40, 142).] Heidi Williams has since argued convincingly that the patent and database restrictions placed on Celera Genomics' sequence data led to a 20–30% reduction in downstream innovation and development in diagnostics, relative to genes sequenced first by the open and public HGP (178). But in 1996, as today, the HGP policy of daily sharing was just one stance among many, in a spectrum of uncertainty about how best to foster progress in genomics and its applications.

By early 1998, the Bermuda Principles had become the official data-sharing policy of the HGP, a condition of the large-scale genome-sequencing grants in all participating countries. In 1996, the NIH put out a statement that it opposed patents on “raw human genomic DNA sequence, in the absence of additional demonstrated biological information,” but because of the Bayh-Dole Act, it could only discourage this practice—as it surely did, with the suggestion that it would remove funding if such patents were filed—and could not prevent DNA sequence patents outright (102). A series of warning letters from the NIH, the DOE, and the Wellcome Trust helped shift the conflicting policy in Germany and another incompatible policy in Japan, moves that prevented a sharing delay from also materializing in France (14, 31, 32). American and British leaders were flexing their muscles here. The threat of removing other national contributors from the HGP if

they did not agree to daily sharing held real weight, as the NIH and the Wellcome Trust could effectively sequence the human genome without collaborators.

The Bermuda Principles continued to exert considerable influence, especially as scientists and administrators amended them to reflect changes in data and practice (23). Two more meetings were convened in Bermuda in February of 1997 and 1998. These were intended not only to address evolving scientific issues with the HGP, but also to revisit data-sharing policies and enact policy shifts as attendees saw fit. By the 1998 meeting, the principles had been revised to include a new daily release trigger of 2-kb rather than 1-kb stretches of DNA and had been extended first to the mouse genome sequence and later to all model organism sequences produced under the aegis of the HGP (65, 177). In 2000, the NHGRI extended the principles again to include several new kinds of data, including those generated during the finishing phase of the draft genome sequence and through whole-genome shotgun sequencing (104).

AFTER BERMUDA: BROADENING OPEN SCIENCE

The Bermuda Principles applied to a small group of laboratories that pulled together to produce a human genome reference sequence. Even as progress toward this goal accelerated in the late 1990s, it became apparent that studying genome variations—that is, individuals' deviations from the reference sequence—and linking genomic data to other kinds of data were essential to making genome sequences meaningful in terms of health outcomes, environmental exposures, genealogy, and family history. By 2000, it was clear that linkage among databases, through systematic data sharing, would drive the science and its applications.

The data and users were far more heterogeneous than those for high-throughput genome sequencing, yet sharing was still paramount. As microarray technology became pervasive after 1996, identifying and cataloging single-nucleotide variants [that is, single-nucleotide polymorphisms (SNPs)] became possible. Fears of hoarding and patent thickets led to the creation of the SNP Consortium, a novel effort to prevent SNP patents (77, 153). This consortium was a public-private partnership that applied for patents on SNPs in profusion to establish formal legal and scientific priority. It then abandoned the applications, releasing data to the public domain unencumbered by patent rights. Because SNPs were tools in agriculture, pharmaceuticals, biotechnology, and elsewhere, industry supported the consortium. Academics were involved because they wanted to use the tools, but commercial manufacturers such as Affymetrix and Illumina manufactured most SNP microarrays themselves. Although these manufacturers certainly believed in and held many patents, they nonetheless foresaw a hopelessly dense patent thicket based on DNA sequences and genomic variants and hoped to avoid it. Microarrays routinely included hundreds of thousands of DNA fragments as probes. If SNPs and ESTs were patented, would it not take hundreds of licenses to build a chip? Would anyone be able to afford it? Access to data, including the ability to use DNA molecules that included SNPs, was essential.

FORT LAUDERDALE AND BEYOND

In 2003, as the HGP was nearing completion, the Wellcome Trust convened a meeting in Fort Lauderdale, Florida, to define rules for biological infrastructure projects. The meeting produced the first of several statements supporting the open science ethos for community resource projects, the likes of which dominate genomics today (175). The meeting focused on the value of prepublication data sharing but relaxed some of the Bermuda Principles' provisions. Like at the 1998 Bermuda meeting, Fort Lauderdale attendees stipulated daily sharing of genome sequences longer than 2 kb for any organism. This was very much in the spirit of the Bermuda Principles, but the Fort Lauderdale statement also mandated that data generators be credited when data were used.

Finally, it acknowledged that hypothesis-driven research did not have the same sharing obligations as those focused on producing community resources, such as sequencing consortiums, the Merck Gene Index project, and the SNP Consortium. This recognition of a need for scientific attribution and credit grew from the diversity of data generators and users. Arias et al. (12) have beautifully summarized the evolution of data-sharing policies that derived from the Bermuda Principles.

Open science continued to infuse the many projects that were developed by the NHGRI or in which the NHGRI was a partner. These were highly diverse, but many were indeed community resource efforts. The International HapMap Project aimed to identify common genomic variants and characterize them from global populations, enabling detailed elucidation of the coinheritance of DNA markers (35, 82). HapMap policies discouraged patenting to such a vigorous degree that Rebecca Eisenberg questioned their wisdom and enforceability (52). The 1000 Genomes Project expanded on the International HapMap Project, including genome sequences from a larger sample of populations, and intended to identify less common genomic variants (29). Another contemporaneous project, the Encyclopedia of DNA Elements (ENCODE), was intended to establish the functions of DNA elements, including regulatory sequences, enhancers, promoters, and operators (36, 54, 106). Because these elements were by definition functional, however, they might also have practical utility and be proper subjects for patenting. The 2003 ENCODE pilot project developed a policy that explicitly acknowledged this, in another deviation from the Bermuda Principles driven by the nature of the work.

As new genomics tools enabled genome-wide studies, the Genetic Association Information Network (GAIN) was formed (55). GAIN built on tools made possible by the HapMap and SNP microarrays, probing hundreds of thousands of common genomic variants in large population studies to identify genomic regions associated with diseases and specific traits. It was intended to ensure rigor, facilitate collaboration, forge academic-industry partnerships, encourage sharing, promote publication, and prevent premature intellectual property claims. The Foundation for the National Institutes of Health, a nonprofit, nongovernment organization, led the effort and invited applications for genome-wide studies. An elaborate peer-review process assessed the technical merits of the proposals and ensured that study design, computational methods, and data sharing complied with guidelines. GAIN had its own set of principles, the first of which was to make results “immediately available for research use by any interested and qualified investigator or organization” (55, p. 1048). Yet because the data were associated with specific individuals, this was qualified “within the limits of providing appropriate protection of research participants” (55, p. 1048). The GAIN principles imposed a duty on users to respect the confidentiality of study participants and to ensure that uses fell within the terms of their informed consent. Another principle was to acknowledge data sources for any reuse. Finally, the original contributors would have a nine-month period during which they alone could submit abstracts or papers based on their data. This embargo period was a marked divergence from the Bermuda Principles, driven yet again by the needs of a different community of producers and users. The laboratories generating the data were disparate in size, type, geography, and purpose, and those differences needed to be accommodated.

Database structures also had to reflect the new realities. The US National Center for Biotechnology Information’s Database of Genotypes and Phenotypes (dbGaP) was established in 2006 (90). Instead of one unitary bank (i.e., GenBank) for sequences, dbGaP had two tiers: a public layer that was freely available to all, and a large data set that might contain private or identifiable information, which in turn required a data access committee to act as a gatekeeper and make sure that users had good reasons for access. Users needed to agree to protect privacy and confidentiality and follow rigorous data security practices. Moreover, many studies needed to be approved by an institutional review board to ensure compliance with ethics rules. The purpose of dbGaP was to enable research through broad access, but the involvement of possibly identifiable research

subjects who had not agreed to post their data on the Internet necessitated new layers of review. The NIH's 2007 policy on genome-wide association studies mandated that NIH-funded studies deposit data into dbGaP (107). This was open science, but the database had rules that were not as simple as free and open access.

The NIH's policy for data access was formally modified in 2014 (37, 108). The 2007 policy applied only to human genome-wide association studies; in 2014, the policy was generalized to cover all NIH institutes and all organisms. It permitted a validation embargo of the data for several months to ensure quality, and dbGaP could hold data in a protected private status for six months. After six months, however, human-derived data would be freely available, and data from nonhumans would be freely available upon publication. The new policy also made it easier for the NIH to continue to tweak parts of the policy in the future, without a formal review of the policy in its entirety. Finally, it stipulated that prospectively, studies should obtain broad consent for deposit and data use as well as approval from an institutional review board for uses of data contributed from past studies, to ensure compliance with the informed consent at the time the data were gathered.

Prepublication data sharing soon broadened beyond genomics, first to proteomics at a meeting in Amsterdam in 2008 and then to other data sets at a workshop in Toronto in 2009 (136, 137, 155). The Toronto statement applied to all data sets that were large scale, broadly useful, and infrastructural (i.e., intended to produce reference data sets) and had community buy-in. This meant respecting the interests of data generators to publish “first global analyses of their data set[s]” (155, p. 169), citing the sources of data, and contacting the data generators if those reusing data might reach publication before the data generators and make significant findings public before the originators.

These iterations and adaptations of data-sharing norms and policies were efforts to keep data open while also respecting the rights and interests of those contributing data (participants) and those generating the data (researchers). Different contexts and uses led to databases with layers of process and rules. Some projects, such as the Personal Genome Project and Open Humans, had an informed consent process that enabled sharing of one's genome sequence on the Internet with no restrictions, but these were for “information altruists” who did not fear misuse of their data, or who at least believed that the benefits of open access were greater than the risks (9, 10, 28, 117, 139). Most data and studies, however, came from projects that entailed narrower informed consent, which generally promised efforts to keep the data secure and prevent identification of individual research participants.

BUILDING A MEDICAL INFORMATION COMMONS: THEORY

The Bermuda Principles and its successor statements were practical efforts to set rules and build infrastructure for increasingly diverse research uses. Generally, these rules were crafted while projects were being designed and carried out, and thus they were generated for and by a community—later including both research and clinical care users—hoping to draw on a medical information commons. A 2011 report from the US National Research Council, *Toward Precision Medicine* (110), laid out a vision of layers of data oriented around individuals to enable an evolving health care and public health system based on biomedical research (21, 35, 86, 110, 148, 150, 151). The recommendations centered on building informational infrastructure, calling for an

“Information Commons” in which data on large populations of patients become broadly available for research use and a “Knowledge Network” that adds value to these data by highlighting their interconnectedness and integrating them with evolving knowledge of fundamental biological processes. (110, p. 4)

This language explicitly invoked a “commons” and thus Garrett Hardin’s classic essay from 1968 (67, 68). Hardin’s presidential address to the American Association for the Advancement of Science, on which this essay was based, discussed problems that had no technical solution and required collective action to solve. Hardin was concerned mainly with militarization and rising population, but the iconic example he used was drawn from an 1833 tragedy of the commons described by William Forster Lloyd. In a pasture open to many herdsmen, each would want to have his cattle feed as much as possible, but if each did so, overgrazing would deplete the pasture. Hardin pointed to the inadequacy of the invisible hand of a market in this situation; solving such problems requires agreement or coercion, preferably according to consensual rules. His examples were laws (against bank robbery), taxes, and other exercises of state power. This left the solution options as binary: the market (no rules, free choices, and creation of property rights) or the Leviathan (the state).

Elinor Ostrom and others have also studied the tragedy of the commons, but they have expanded the potential solution set by observing that real-world communities often craft rules that prevent the depletion of resources or orchestrate services requiring collective action. Ostrom’s earlier work focused on natural resource depletion (123, 125). She examined how some communities prevented overfishing while others depleted fish stocks for lack of a viable commons. She also studied collective interests in urban policing and managing water resources. In her later work, Ostrom and others extended their thinking to a knowledge commons, a concept closely aligned to a medical information commons (74). Theoretically, the biggest difference between a natural resource commons and a knowledge (or data) commons is that natural resources, like biorepositories of samples, can be depleted. Data and knowledge, by contrast, are not depleted no matter how many people use them.

Indeed, data and knowledge can sometimes achieve network efficiencies that expand dramatically as the number of users increases, according to Metcalfe’s law (167). Ostrom and her collaborators devoted much of their attention to institutional arrangements that enabled a viable knowledge commons to form, as well as how to govern it (121, 122, 124, 129). Trust and reciprocity are central themes, and institutional analysis and development were the framework that Ostrom proposed for addressing problems of collective action. In 2010, she concluded that “rules related to the production of generally accessible data” include the following:

- Who must deposit their data?
- How soon after production and authentication of data do researchers have to deposit the data?
- How long should the embargo last?
- How should conformance to the rules be monitored?
- How many researchers are involved in producing and analyzing the particular kind of data?
- Should an infraction be made public in order to tarnish the reputation of the infringer? (124, p. 813)

These questions address who is and is not a member of the community of contributors and users, how to formulate contributions and use rules, and governance and enforcement procedures. The successive statements about genomic data sharing, beginning with the Bermuda Principles in 1996, were all implicit or explicit responses to questions raised by Ostrom. Contreras (38) drew directly on this theory in his work on genomics, and Strandburg et al. (149) extended theories of the commons into data and knowledge domains.

Commons theory is not the only framework for open science, however. The Global Alliance for Genomics and Health, discussed in the next section, appeals to article 27 of the United Nations’ Universal Declaration of Human Rights, which includes the right “to share in scientific advancement and its benefits” (159). The value of sharing data globally is reiterated in a

series of declarations by the United Nations Educational, Scientific, and Cultural Organization, recommendations by the Council of Europe, and guidelines from the Organization for Economic Cooperation and Development (41, 107, 118–120, 156–158). Much of bioethics has centered on protecting research participants from risk, but the Global Alliance also focuses on the right to benefit, including through data sharing, as a means to advance science and improve medicine—the very purpose for which many people contribute samples and information to medical research. Yet merely stating the right to benefit and implementing that right in the real world are different things; moving from aspiration to realization is a constant struggle. Future statements will no doubt follow as international data sharing continues to take shape.

THE EMERGENCE OF THE GLOBAL ALLIANCE FOR GENOMICS AND HEALTH

Although the need to link data of many types, housed in many parts of the world, is obvious, the structure of a global commons of genomic and other data raises many practical problems. Some are technical: a need for application interfaces, standard formats, and interoperable data systems. But the legal and social challenges are even more formidable.

In January 2013, more than 50 individuals from 8 countries—roughly comparable to the first Bermuda meeting in size and national representation—met in New York City to develop some standards and articulate the need for genomic infrastructure. The results were summarized in a June 2013 white paper describing the need for collective international action and the need to avoid “a hodge-podge of balkanized systems—as developed in the U.S. for electronic medical records—a system that inhibits learning and improving health care” (57, p. 1). Attendees proposed a global alliance, which came to be known as the Global Alliance for Genomics and Health. By 2016, the Global Alliance comprised 800 people from 400 organizations in 70 countries (59). This growth alone indicates how much more complicated it will be to achieve data sharing among hundreds of diverse stakeholders, as compared with the organizational challenge when HGP leaders met in Bermuda, which involved roughly 50 people from only 5 countries.

Beacon and Matchmaker

As the Global Alliance was being formed, efforts converged on three pilot projects, and in 2016 a fourth was added. The three initial projects were Beacon, Matchmaker, and the *BRCA* Challenge. The Beacon project queried a network of genomic databases to see whether they harbored information about particular genomic variants. Those who came upon a variant in research or clinical practice could send a single query and find out which participating databases had relevant information, anywhere in the world. As of summer 2016, 25 institutions and 250 data sets were participating in Beacon (60). The Matchmaker Exchange was another data-brokering effort for those studying rare disorders, enabling users to find phenotype and genotype data pertinent to a genomic-clinical profile, again allowing a query of participating databases. The October 2015 issue of *Human Mutation* featured 16 articles from the Matchmaker demonstration project (128). Both Beacon and Matchmaker point to the places where data can be found, while the data themselves remain where they are, and users can seek access.

Case Study: Sharing Data About *BRCA* Variants

The *BRCA* Challenge, in contrast to Beacon and Matchmaker, entailed a further step for characterizing and curating variants in two of the most studied and clinically significant human genes, *BRCA1* and *BRCA2*. The *BRCA* Challenge was designed to pool publicly available data. The intent

was to build out from these two well-studied genes of obvious clinical significance, establishing a precedent for expansion into other genes. The resulting database, *BRCA* Exchange, has three tiers. The top tier is fully public and lists variants interpreted by the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA), an international consortium founded at a May 2009 meeting in Amsterdam (147). ENIGMA is the expert vetting committee for *BRCA* Exchange variant interpretation. The next layer is a research data set with links to the evidence base, including conflicting interpretations and unvetted reports, with pointers to other databases containing further information. The third layer, which was still under construction at the end of 2016, will contain case-level data that might be linked to identifiable individuals, thus requiring higher levels of security and a gatekeeper to ensure compliance with informed consent and prevent misuse or unauthorized reidentification. *BRCA* Exchange shares data extensively with ClinVar, the Leiden Open Variation Database (LOVD), and other genomic variant databases.

The need for the *BRCA* Challenge was an artifact of the history of testing for inherited risk of breast and ovarian cancer. Mary-Claire King's team found linkage to a putative risk gene in families in 1990, showing there was a mutated gene to be found—the first and most important step in a cascade of events to follow (66). This set off an intense race to identify, clone, and sequence the gene likely associated with cancer in high-risk families (8, 12, 44). *BRCA1* was identified in 1994 by a team that was led by Mark Skolnick at the University of Utah and was also associated with the genomic startup Myriad Genetics (62, 99). Linkage to human chromosome 13 was found in 1994 by a team in Great Britain, and *BRCA2* was cloned and sequenced in 1995 (181, 182). Two decades later, *BRCA1* and *BRCA2* remain the two genes most commonly mutated in families with inherited risk of breast and ovarian cancer, although *BRCA* mutations are found in other cancers, and mutations in another two dozen genes are also associated with carcinomas of the ovaries and breasts (albeit at a much lower frequency).

Both genes were patented, and the story is complicated (27, 40, 44, 62, 127). The first *BRCA1* patent was granted to OncorMed in summer 1997, and OncorMed sued Myriad for patent infringement. Myriad countersued the day after it received its first patent that December. The companies settled out of court, with OncorMed agreeing to exit the *BRCA* testing market and assign its patents to Myriad. (The Myriad/Utah filing date was earlier, and OncorMed was apt to lose if the patent office initiated an interference proceeding, the administrative procedure for allocating patent rights in cases of disputed priority.) There were other *BRCA* patents, including one granted to the UK team that first published on *BRCA2*, but that sequence was chimeric, and the Myriad/Utah team filed a patent application on *BRCA2* that avoided the chimerism just a day before the UK team published in *Nature*. Myriad cleared the US market of competitors for commercial *BRCA* testing by sending notification letters to laboratories offering the test and even suing one such laboratory at the University of Pennsylvania. The laboratory quickly settled and withdrew from testing except for patients in the University of Pennsylvania health care system.

From 1998 through 2013, Myriad had a service monopoly on American *BRCA* testing, but this strategy did not work anywhere outside the United States. European patents were opposed and narrowed (62, 94, 127).

In Australia, in October 2002, Myriad was forced to license to Genetic Technologies to settle a patent infringement suit over use of intervening sequences. As a “gift to Australia,” Genetic Technologies permitted laboratories in the Australian regional health system to offer testing (116). A new CEO at Genetic Technologies, Michael Ohanessian, later threatened to revoke this gift, but a firestorm of criticism led the firm to replace Ohanessian and generated a flurry of Senate activity. The gift was restored (116, 166).

In Canada, Ontario's premier as well as the minister of health and long-term care refused to recognize Myriad's rights, and Myriad never sued in Canada, so most provinces continued to offer

BRCA testing (62). In Great Britain, the National Health Service offered *BRCA* testing regionally and largely ignored Myriad's patents (71, 127).

By 2013, Myriad had administered more than a million *BRCA* tests. Its US monopoly ended at the US Supreme Court when Myriad lost an epic patent battle against the American Civil Liberties Union (126, 144). The following month, in July 2013, Myriad filed the first of several lawsuits against seven competitors, all of which ended when the Court of Appeals for the Federal Circuit invalidated Myriad's patents in December 2014 (81). Myriad had dismissed the last of its suits by February 2015. Several new laboratories entered the *BRCA* testing market on June 13, 2013, the day of the Supreme Court decision; several more entered in the following months; and still more did so after the litigation dust settled.

Myriad shared its data on *BRCA* genetic variants until November 2004 and allowed selective access through its proprietary database to academic collaborators through 2006 (13). But it stopped depositing data at the NHGRI's locus-specific Breast Cancer Information Core, the largest repository of such data. It has since taken further steps to protect its trade secret database through click-through agreements on its website, precluding users from sharing data with third parties and explicitly claiming trade secrecy (data on file with R. Cook-Deegan). A decade of testing by the company has revealed thousands of *BRCA* genetic variants, but Myriad alone knows what these are. Although it publishes the names of its interpretive methods, Myriad neither shares them sufficiently for replication nor provides the underlying data, publishing in journals that do not require such disclosure (49, 165).

The Global Alliance's *BRCA* Challenge was intended to address the anomalous situation occasioned by Myriad, which had a 15-year US patent monopoly and also decided to treat data as trade secrets. To interpret *BRCA* variants, the rest of the world had to catch up, because Myriad did not make its data available and did not participate in inter-laboratory comparisons of genetic variants organized by the community. Myriad had many collaborations, but none that channeled their data systematically into publicly accessible databases. This is one major reason that the American College of Medical Genetics and Genomics issued a public statement condemning data hoarding and encouraging data sharing (1). The number of *BRCA* tests administered worldwide by other companies and laboratories is comparable to the number administered by Myriad. Moreover, even as the largest single testing laboratory, Myriad has limited access to genetic variants in Africa, Asia, Latin America, and other places where its pricing precludes broad testing, different founder mutations have taken root as a result of context-specific evolutionary pressures, and rare alleles will continue to crop up (**Figure 3**). In short, the data have not been shared, stored, curated, or interpreted in ways that can be used for clinical decisions.

The *BRCA* Challenge was announced as ClinVar and ClinGen were getting started. *BRCA* Exchange regularly shares data with ClinVar, and ClinVar regularly does variant comparisons with LOVD, the Associated Regional and University Pathologists (ARUP), large academic laboratories, and some commercial laboratories. *BRCA* Exchange is linking to more databases and characterizing more variants as more databases participate and as more data become available. Several of the laboratories that started *BRCA* testing—Ambry Genetics, Invitae, GeneDx, Illumina, and others—promoted the open science framework of ClinVar and ClinGen and contributed their data on variants, including different degrees of clinical phenotype information. Another response to Myriad's data-hoarding policy was the Sharing Clinical Results Project (SCRIP), an effort to secure the laboratory reports that Myriad sent back to ordering laboratories and physicians, and the Free the Data project to obtain the same information from individual women who knew their *BRCA* status, which was organized by the Genetic Alliance (16).

The two largest diagnostic firms in the United States, Quest Diagnostics and LabCorp, started offering *BRCA* testing in 2013. In May 2015, Quest announced that it was contributing its data to

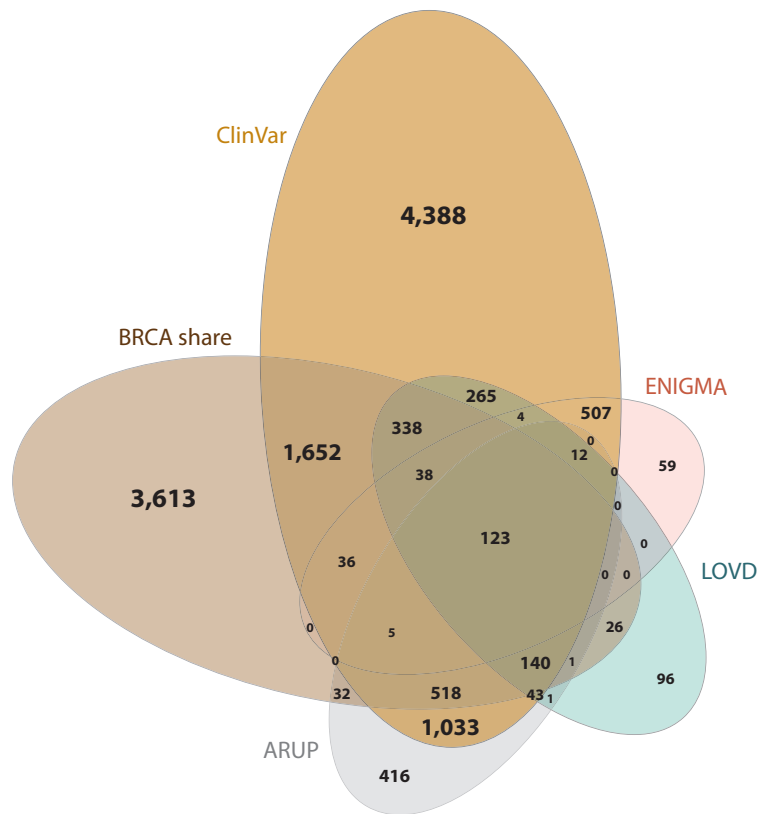


Figure 3

Venn diagram showing the number of *BRCA1* and *BRCA2* genetic variants associated with inherited risk of breast, ovarian, and other cancers across five databases (not drawn to scale). None of these databases, located in Europe and North America, includes the substantial numbers of rare variants yet to be discovered from Africa, Asia, Latin America, and elsewhere. The most common variants are present in all databases, but most rare variants are present in only one database, indicating the incomplete state of global data sharing and emphasizing the need to pool data. Abbreviations: ARUP, Associated Regional and University Pathologists; ENIGMA, Evidence-based Network for the Interpretation of Germline Mutant Alleles; LOVD, Leiden Open Variation Database. Data are combined from **Figures 2a** and **2b** in Reference 19, which was submitted in July 2016 and published in December 2016.

the Universal Mutation Database (UMD) in Paris, where it would be well curated and interpreted (19, 132). Quest proposed that commercial laboratories pay for access to the UMD data and contribute to it, whereas researchers would have free access. LabCorp joined that effort, which came to be known as BRCA Share. When, how, and how much data will flow into ClinVar and other freely available databases, where they can be used to guide clinical decisions worldwide, remain to be seen. ARUP Laboratories also established a *BRCA* database, which is likewise contributing to ClinVar. The spectrum of data sharing for *BRCA* variants thus ranges from proprietary data hoarding by Myriad to free and open sharing embodied by ClinVar, SCRP, and Free the Data, with intermediate models of research access and paid commercial storage and curation through BRCA Share and many national and regional databases.

The Cancer Gene Trust

The Global Alliance's newest demonstration project is the Cancer Gene Trust, which will first focus on sharing data about somatic cancer genomic variants (59). It is starting with somatic genomic variants in part for simplicity, until privacy issues with potentially identifiable germline data are resolved. One distinctive feature is packets of software that can move among linked servers, leaving the large repositories of data in place but allowing analyses and results to be returned to the user. The data stay in place; the analytical software migrates.

The Regulatory and Ethics Working Group

The Global Alliance's projects entail a substantial technical component, and much of the work has centered on developing application programming interfaces. One group addresses data security, a highly technical domain. Many of the challenges in building the Global Alliance, however, center on ethical issues, law, and policy, which are the domains of the Regulatory and Ethics Working Group. Indeed, the legal and social impediments to global data sharing are among the most challenging obstacles to establishing a medical information commons. One of the early efforts of the working group was to articulate its "Framework for Responsible Sharing of Genomic and Health-Related Data," a high-level agreement that has been translated into 13 languages (58). This framework emphasized the need to preserve the scientific value of data where possible, rather than anonymizing it and diminishing the ability to make inferences among diverse, individually oriented data types. Major issues include (a) the need to protect privacy and confidentiality and to honor informed consent agreements for data generated under diverse national laws, (b) the need for data security, (c) the accommodation of both clinical and research users, and (d) the fact that a much more complicated and diverse set of commercial firms are now involved in genomics research. We briefly review these issues below.

International data sharing must abide by national laws. Privacy and return of results to participants have been the subjects of several articles in the *Annual Review of Genomics and Human Genetics* (6, 63, 88, 97). One of the key findings from legal scholarship is that laws in different nations pertain to activities essential to data sharing. Branum & Wolf (22) recently reviewed the international law on return of results from genomic analysis. Another relevant body of law concerns privacy, confidentiality, and informed consent. Many nations have passed laws encouraging the engagement of local researchers with research done in their countries, both to foster economic development and to prevent "biocolonialism" and "helicopter research," wherein foreign researchers extract value but leave little of lasting value to the residents who are the source of data, violating the notion of reciprocity.

A symposium and two recent issues of the *Journal of Law, Medicine, and Ethics* resulted from a massive effort, led by Mark Rothstein and Bartha Knoppers, to review laws pertaining to sharing of samples and data (138, 139). Forty authors surveyed the law in 20 countries; the focus was on biobanks, but the articles also necessarily covered data sharing. Many countries have laws governing data security, privacy, and confidentiality, and many require governmental approval and/or sanction from a local ethics review board to export genetic data. Dove (47, p. 684) noted that full international harmonization of laws is unlikely, instead recommending efforts to cultivate "foundational responsible data-sharing principles in an overarching governance framework." Thorogood & Zawati (154) pointed out that, although incompatible national laws can be impediments to sharing, there is also virtue in a pluralism of approaches across nations and cultures.

National laws must be respected, but they are complicated and require national, regional, and sometimes local legal expertise to identify and interpret. Rothstein & Knoppers (138, p. 674) concluded,

[R]elevant laws differ widely among countries engaged in biobank-enabled research in terms of substance, procedure, and underlying public policies. The lack of international regulatory harmonization has been shown to impede data sharing for translational research in genomics and related fields. The daunting task is to identify and characterize the biobank structure and applicable standards in each country and then to devise possible ways to harmonize policies and laws to enable international biobank research while still giving effect to essential privacy protections.

Establishing the infrastructure for international data sharing—to reap the benefits of cataloging genomic variants around the globe—requires confronting drastically greater complexity than did organizing the sequences of mapped DNA segments from anonymized samples, the starting material for the human genome reference sequence. It is a long, choppy voyage from Bermuda.

Scholars are working assiduously to navigate these tumultuous waters. Vanderbilt's Center for Genetic Privacy and Identity in Community Settings, led by Ellen Wright Clayton and Bradley Malin, is a Center of Excellence in ELSI Research (146). Susan Wolf, Ellen Wright Clayton, and Frances Lawrenz are leading LawSeq, a prodigious accumulation of talent that is turning its attention to the legal foundations of translating genomics into clinical applications, focusing on US federal law (34). Finally, Amy McGuire and Robert Cook-Deegan are codirecting a grant, "Building the Medical Information Commons," also centered on American efforts (15).

Clinical laboratories generate as much (or more) information about genomic variants as do research laboratories. Since Bermuda, the flow of human genome data has shifted decisively from publicly funded research to commercial laboratory testing, to help individuals make better-informed decisions about medical care, ancestry, or other personal interests. The uses of genomic variation data are likewise embedded in both research and clinical care. The hundreds of databases that have nucleated around particular genes (e.g., *CFTR* and the Huntington disease locus) or medical conditions (e.g., epilepsy, Alzheimer disease, and various cancers) generally started with researchers contributing to the literature and depositing their data—sometimes in locus-specific databases and sometimes in more general databases, such as the Human Gene Mutation Database in Cardiff, the UMD in Paris, the LOVD in the Netherlands, and the many others maintained by the National Center for Biotechnology Information at the National Library of Medicine (e.g., GenBank, dbGaP, RefSeq, and ClinVar). These databases now contain evidence used to make clinical decisions, but (with the exception of ClinVar) they were generally intended for research, not to support clinical decisions. The information they contain needs to be validated before clinical use.

Data from commercial testing laboratories are used mainly for clinical inference, yet the flow of data into public databases is highly variable, depending in part on business models, history, and how hard it is to get data into them. ClinVar is unique in having been constructed from the beginning for clinical determinations, although it draws heavily from research databases (70). Some of the major contributors to the ClinVar database, through the panel of collaborators that constitute ClinGen, are commercial laboratories (128, 134). This trend toward data flowing primarily from clinical testing laboratories is bound to accelerate as genomic analysis is integrated into health care. The shift to clinical-grade databases will require systematic criteria and procedures for interpreting the clinical significance of genomic variants, storage and curation of the data, quality

control measures, and participation in proficiency testing, all of which are part of the nascent infrastructure for clinical genomics.

In most countries, genetic testing is incorporated into laboratory practices in national health systems. This is not the case in the United States, where federal regulation of genetic tests has been the subject of debates, books, and several reports from federal advisory committees since at least 1984 (78, 79, 111, 141). Laboratories in the United States are currently regulated by the Clinical Laboratories Improvement Amendments (CLIA) of 1988, through the Centers for Medicare and Medicaid Services. The College of American Pathologists also accredits clinical laboratories. The US Food and Drug Administration (FDA) floated draft guidance indicating its intent to regulate laboratory-developed genetic tests as medical devices in 2014, proposing to phase in such regulation over nine years (162). The FDA's proposed entry into regulation caused a kerfuffle and was opposed by many laboratories, their trade associations, and the Association for Molecular Pathology (131), and in November 2016 the FDA announced that it would back away from finalizing its guidance (133). The FDA then made a discussion paper publicly available to guide future deliberations (164). Although the FDA signaled its openness to discussion while not immediately contemplating strong regulation, the United States is just one jurisdiction, and the FDA is just one among many actors with a stake in clinical interpretation of genomic variants. Databases will be centrally important, and the need for independent verification of genomic interpretations used to guide clinical decisions will remain an important function that depends on accurate data and extensive data sharing. The FDA documents thus illuminate some of the challenges ahead in building the tools for clinical use and the need for regulatory-grade genomic databases (163). Indeed, clinical use requires far more formal oversight and regulation than creating and using data in research. The flow of data from commercial testing laboratories is likely to become the main source of new information about human genomic variation.

The diversity and importance of private firms are substantially greater now than during the Human Genome Project. Commercial genomics extends well beyond genetic testing. Even within genetic testing, it includes ancestry testing, personal genomic profiling with microarrays, exome sequencing or even whole-genome sequencing, and a panoply of tests ranging from single-gene tests to multigene panels. Genome sequencing has been used to study rare disorders and profile somatic mutations in cancers in order to make diagnoses and to guide choices about treatment and prevention. Other firms specialize in integrating genomic data into medical records or providing bioinformatics tools for genomic data analysis. Some are dedicated to genome sequencing as all or part of their business models. In a 2014 letter to the editor, Curnutte et al. (43) documented this diversity of commercial software, hardware, and services. A great deal of expertise in informatics and instrument manufacturing resides in the private sector, in companies of wildly different sizes, ages, financial health, and business models. Many are compatible with data sharing; some are not. The evolving pattern of sharing for *BRCA1* and *BRCA2* and the complexity of the genomic data commons illustrate the challenges ahead.

CONCLUSION

The Bermuda Principles for daily prepublication genomic data release set a strong foundation for other efforts to promote open science. They set a salutary precedent that enabled more rapid progress toward first assembling a reference sequence of the human genome and then interpreting the meaning of genomic variants in humans and other organisms. The initial community of approximately 50 people has morphed into a global endeavor involving hundreds of laboratories, ranging from pure research to clinical use, ancestry, and other applications.

As the community of data contributors and users and the types of data available have broadened, the idea of daily sharing has been modified and adapted to quite different contexts in order to permit informed consent of individuals to whom the data pertain, ensure data security, and address the considerable diversity in international ethical requirements and laws governing privacy, confidentiality, and cross-border flow of genetic samples and data. In addition to these challenges, commercial interests have both intensified and grown far more diverse. The spirit of open science, however, persists. The Bermuda Principles provided a foundation for current efforts to build a global medical information commons, but as this article has argued, it will be no simple task to ensure compliance with disparate national laws while also respecting the rights and interests of those contributing to and drawing from the commons. The wisdom of supporting legal scholarship and ethical analysis while the technology advances is thus vindicated, because expertise in ethics, law, and policy will be crucial to meeting the challenges ahead.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors' research was supported in part by an NHGRI grant to Baylor College of Medicine (R01 HG008918); an NHGRI grant to the Center for Public Genomics at Duke University, a former Center of Excellence in ELSI Research (P50HG-003391, principal investigator R.C.-D.); the Ewing Marion Kauffman Foundation (principal investigator R.C.-D.); and the United States Studies Centre at the University of Sydney (principal investigator R.A.A.). R.C.-D. is a senior fellow of FasterCures, a center of the nonprofit Milken Institute. All opinions expressed and errors remaining here are the responsibility of the authors.

LITERATURE CITED

1. ACMG Board Dir. 2017. Laboratory and clinical genomic data sharing is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics. *Genet. Med.* 19:721–22
2. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, et al. 1991. Complementary DNA sequencing: expressed sequence tags and the Human Genome Project. *Science* 252:1651–56
3. Adams MD, Sutton GG, Smith HO, Myers EW, Venter JC. 2003. The independence of our genome assemblies. *PNAS* 100:3025–26
4. Adams MD, Venter JC. 1996. Should non-peer-reviewed raw DNA sequence data release be forced on the scientific community? *Science* 274:534–36
5. Adler R. 1992. Genome research: fulfilling the public's expectations for knowledge and commercialization science. *Science* 257:908–14
6. Anderlik MR, Rothstein MA. 2001. Privacy and confidentiality of genetic information: what rules for the new science? *Annu. Rev. Genom. Hum. Genet.* 2:401–33
7. Anderson A. 1992. Yeast genome project: 300,000 and counting. *Science* 256:462
8. Angier N. 1994. Fierce competition marked fervid race for cancer gene. *New York Times*, Sept. 20, p. C3. <http://www.nytimes.com/1994/09/20/science/fierce-competition-marked-fervid-race-for-cancer-gene.html>
9. Angrist M. 2009. Eyes wide open: the Personal Genome Project, citizen science and veracity in informed consent. *Pers. Med.* 6:691–99
10. Angrist M. 2010. *Here Is a Human Being: At the Dawn of Personal Genomics*. New York: HarperCollins

11. Ankeny RA, Leonelli S. 2015. Valuing data in postgenomic biology: how data donation and curation practices challenge the scientific publication system. In *Postgenomics: Perspectives on Biology After the Genome*, ed. SS Richardson, H Stevens, pp. 126–49. Durham, NC: Duke Univ. Press
12. Arias JJ, Pham-Kanter G, Campbell EG. 2015. The growth and gaps of genetic data sharing policies in the United States. *J. Law Biosci.* 2:56–68
13. Baldwin AL, Cook-Deegan R. 2013. Constructing narratives of heroism and villainy: case study of Myriad’s BRACAnalysis[®] compared to Genentech’s Herceptin[®]. *Genome Med.* 5:8
14. Balter M. 1998. France’s sequencers aim to join the big league. *Science* 280:30–31
15. Baylor Coll. Med. 2016. *Information commons*. <https://www.bcm.edu/centers/medical-ethics-and-health-policy/research/elsi-research/information-commons>
16. Bean LJ, Tinker SW, da Silva C, Hegde MR. 2013. Free the data: one laboratory’s approach to knowledge-based genomic variant classification and preparation for EMR integration of genomic data. *Hum. Mutat.* 34:1183–88
17. Bentley DR. 1996. Genomic sequence information should be released immediately and freely in the public domain. *Science* 274:533–34
18. Bentley DR, Pruitt KD, Deloukas P, Schuler GD, Ostell J. 1998. Coordination of human genome sequencing via a consensus framework map. *Trends Genet.* 14:381–84
19. Bérout C, Letovsky SI, Braastad CD, Caputo SM, Beaudoux O, et al. 2016. BRCA Share: a collection of clinical BRCA gene variants. *Hum. Mutat.* 37:1318–28
20. Brandt KA. 1993. The GDB Human Genome Data Base: a source of integrated genetic mapping and disease data. *Bull. Med. Libr. Assoc.* 81:285–92
21. Branscomb E. 2011. Interview by Kathryn Maxson and Robert Cook-Deegan, Oct. 17. Duke Univ. Libr., Durham, NC. <http://hdl.handle.net/10161/7693>
22. Branum R, Wolf SM. 2015. International policies on sharing genomic research results with relatives: approaches to balancing privacy with access. *J. Law Med. Ethics* 43:576–93
23. Burris J, Cook-Deegan R, Alberts B. 1998. The Human Genome Project after a decade: policy issues. *Nat. Genet.* 20:333–35
24. *C. elegans* Seq. Consort. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282:2012–18
25. Calia KG. 1992. Patentability of expressed sequence tags: a study of the Venter application. *Biotechnol. Law Rep.* 11:540–57
26. Cantor CR. 1990. Orchestrating the Human Genome Project. *Science* 248:49–51
27. Cheon JY, Mozersky J, Cook-Deegan R. 2014. Variants of uncertain significance in *BRCA*: a harbinger of ethical and policy issues to come? *Genome Med.* 6:121
28. Church GM. 2005. The Personal Genome Project. *Mol. Syst. Biol.* 1:2005.0030
29. Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, et al. 2012. The 1000 Genomes Project: data management and community access. *Nat. Methods* 9:459–62
30. Collins F. 2012. Interview by Kathryn Maxson and Robert Cook-Deegan, Apr. 12. Duke Univ. Libr., Durham, NC. <http://hdl.handle.net/10161/7704>
31. Collins F, Patrinos A. 1997. Letter to Frank Laplace, March 21. Cent. Public Genom. Res. Files, Duke Univ. Libr., Durham, NC. <http://hdl.handle.net/10161/7727>
32. Collins F, Patrinos A, Morgan M. 1998. Letter to Kenichi Matsubara, March 9. Cent. Public Genom. Res. Files, Duke Univ. Libr., Durham, NC. <http://hdl.handle.net/10161/7737>
33. Collins FS, Green ED, Guttmacher AE, Guyer MS. 2003. A vision for the future of genomics research. *Nature* 422:835–47
34. Consort. Law Values Health Environ. Life Sci. 2016. *LawSeq: building a sound legal foundation for translating genomics into clinical application*. <https://consortium.umn.edu/research/lawseq-building-sound-legal-foundation-translating-genomics-clinical-application>
35. Contreras JL. 2011. Bermuda’s legacy: policy, patents, and the design of the genome commons. *Minn. J. Law Sci. Technol.* 12:61–125
36. Contreras JL. 2014. Constructing the genome commons. In *Governing Knowledge Commons*, ed. BM Frischmann, MJ Madison, KJ Strandburg, pp. 99–136. New York: Oxford Univ. Press

37. Contreras JL. 2015. NIH's genomic data sharing policy: timing and tradeoffs. *Trends Genet.* 31:55–57
38. Contreras JL. 2016. Genetic property. *Georgetown Univ. Law J.* 105:1–54
39. Cook-Deegan R. 1994. *The Gene Wars: Science, Politics, and the Human Genome*. New York: Norton
40. Cook-Deegan R, Heaney C. 2010. Patents in genomics and human genetics. *Annu. Rev. Genom. Hum. Genet.* 11:383–425
41. Counc. Eur. Comm. Minist. 2006. *Recommendation Rec(2006)4 of the Committee of Ministers to member states on research on biological materials of human origin*. https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805d84f0
42. Cozzarelli NR. 2004. UPSIDE: Uniform Principle for Sharing Integral Data and Materials Expediently. *PNAS* 101:3721–22
43. Curnutte MA, Frumovitz KL, Bollinger JM, McGuire AL, Kaufman DJ. 2014. Development of the clinical next-generation sequencing industry in a shifting policy climate. *Nat. Biotechnol.* 32:980–82
44. Davies K, White M. 1996. *Breakthrough: The Race to Find the Breast Cancer Gene*. New York: Wiley & Sons
45. de Chadarevian S. 2002. *Designs for Life: Molecular Biology After World War II*. Cambridge, UK: Cambridge Univ. Press
46. Dickson D. 1994. Merck to back 'public' sequencing. *Nature* 371:365
47. Dove ES. 2015. Biobanks, data sharing, and the drive for a global privacy governance framework. *J. Law Med. Ethics* 43:675–89
48. Editorial. 2014. How to get ahead: The success of the \$1,000 genome programme offers lessons for fostering innovation. *Nature* 507:273–74
49. Eggington JM, Bowles KR, Moyes K, Manley S, Esterling L, et al. 2014. A comprehensive laboratory-based program for classification of variants of uncertain significance in hereditary cancer genes. *Clin. Genet.* 86:229–37
50. Eisenberg RS. 1990. Patenting the human genome. *Emory Law J.* 39:721–45
51. Eisenberg RS. 1996. Intellectual property issues in genomics. *Trends Biotechnol.* 14:302–7
52. Eisenberg RS. 2006. Patents and data-sharing in public science. *Ind. Corp. Change* 15:1013–31
53. Eisenberg RS, Merges RP. 1995. Opinion letter as to the patentability of certain inventions associated with the identification of partial cDNA sequences. *AIPLA Q. J.* 23:1–52
54. ENCODE Proj. Consort. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306:636–40
55. GAIN Collab. Res. Group. 2007. New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat. Genet.* 39:1045–51
56. Garcia-Sancho M. 2012. *Biology, Computing, and the History of Molecular Sequencing: From Proteins to DNA, 1945–2000*. New York: Palgrave Macmillan
57. Glob. Alliance Genom. Health. 2013. *Creating a global alliance to enable responsible sharing of genomic and clinical data*. White Pap., Glob. Alliance Genom. Health, Toronto. <http://genomicsandhealth.org/white-paper-creating-global-alliance-enable-responsible-sharing-genomic-and-clinical-data-read-online>
58. Glob. Alliance Genom. Health. 2014. *Framework for responsible sharing of genomic and health-related data*. Framew. Doc., Glob. Alliance Genom. Health, Toronto. <https://genomicsandhealth.org/about-the-global-alliance/key-documents/framework-responsible-sharing-genomic-and-health-related-data>
59. Glob. Alliance Genom. Health. 2017. *About the Global Alliance*. <http://genomicsandhealth.org/about-global-alliance>
60. Glob. Alliance Genom. Health. 2017. *Beacon project*. <http://ga4gh.org/#/beacon>
61. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, et al. 1996. Life with 6000 genes. *Science* 274:546, 563–67
62. Gold ER, Carbone J. 2010. Myriad Genetics: in the eye of the policy storm. *Genet. Med.* 12:S39–70
63. Greely HT. 2007. The uneasy ethical and legal underpinnings of large-scale genomic biobanks. *Annu. Rev. Genom. Hum. Genet.* 8:343–64
64. Green ED, Watson JD, Collins FS. 2015. Human Genome Project: twenty-five years of big biology. *Nature* 526:29–31

65. Guyer M. 1998. Statement on the rapid release of genomic DNA sequence. *Genome Res.* 8:413
66. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, et al. 1990. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250:1684–89
67. Hardin G. 1968. The tragedy of the commons. *Science* 162:1243–48
68. Hardin G. 1998. Extensions of “The Tragedy of the Commons.” *Science* 280:682–83
69. Harris R, Dixon B. 2001. Genome jousting enlivens a sporting lull. *Curr. Biol.* 11:R202–3
70. Harrison SM, Riggs ER, Maglott DR, Lee JM, Azzariti DR, et al. 2016. Using ClinVar as a resource to support variant interpretation. *Curr. Protoc. Hum. Genet.* 89:1–23
71. Hawkins N. 2012. An exception to infringement for genetic testing—addressing patient access and divergence between law and practice. *Int. Rev. Intellect. Prop. Compet. Law* 43:641–61
72. Hayden EC. 2014. The \$1,000 genome. *Nature* 507:294–95
73. Healy B. 1992. Special report on gene patenting. *N. Engl. J. Med.* 327:664–68
74. Hess C, Ostrom E, eds. 2007. *Understanding Knowledge as a Commons: From Theory to Practice*. Cambridge, MA: MIT Press
75. Hilgartner S. 1998. Data access policy in genome research. In *Private Science: Biotechnology and the Rise of the Molecular Sciences*, ed. A Thackeray, pp. 202–18. Philadelphia: Univ. Pa. Press
76. Hilgartner S, Brandt-Rauf SI. 1994. Data access, ownership, and control. *Knowl. Creat. Diffus. Util.* 15:355–72
77. Holden AL. 2002. The SNP Consortium: summary of a private consortium effort to develop and applied map of the human genome. *BioTechniques* 32:S22–26
78. Holtzman NA. 1989. *Proceed with Caution: Predicting Genetic Risks in the Recombinant DNA Era*. Baltimore, MD: Johns Hopkins Univ. Press
79. Holtzman NA, Watson MS, eds. 1997. *Promoting safe and effective genetic testing in the United States: final report of the Task Force on Genetic Testing*. Rep., Natl. Inst. Health–Dep. Energy Work. Group Ethical Leg. Soc. Implic. Hum. Genome Res., Washington, DC
80. Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, et al. 1995. An STS-based map of the human genome. *Science* 270:1945–54
81. *In re BRCA1- and BRCA2-Based Hereditary Cancer Test Patent Litigation*, Case Nos. 2014-1361, 2014-1366 (Fed. Cir., Dec. 17, 2014)
82. Int. HapMap Consort. 2003. The International HapMap Project. *Nature* 426:789–96
83. Int. Hum. Genome Seq. Consort. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931–45
84. Istrail S, Sutton GG, Florea L, Halpern AL, Mobarry CM, et al. 2004. Whole-genome shotgun assembly and comparison of human genome assemblies. *PNAS* 101:1916–21
85. Jasanoff S. 2005. *Designs on Nature: Science and Democracy in Europe and the United States*. Princeton, NJ: Princeton Univ. Press
86. Jasanoff S, ed. 2011. *Reframing Rights: Bioconstitutionalism in the Genetic Age*. Cambridge, MA: MIT Press
87. Joint Genome Inst. 2016. *History*. <http://jgi.doe.gov/about-us/history>
88. Kaye J. 2012. The tension between data sharing and the protection of privacy in genomics research. *Annu. Rev. Genom. Hum. Genet.* 13:415–31
89. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
90. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, et al. 2007. The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* 39:1181–86
91. Marshall E. 1995. A strategy for sequencing the genome 5 years early. *Science* 267:783–84
92. Marshall E. 2001. Bermuda rules: community spirit, with teeth. *Science* 291:1192
93. Martinell J. 1992. Initial rejection of NIH patent application 07/837,195 (EST sequences and methods). *Biotechnol. Law Rep.* 11:578–96
94. Matthijs G. 2006. The European opposition against the BRCA gene patents. *Fam. Cancer* 5:95–102
95. Maxson Jones K, Ankeny RA, Cook-Deegan R. Forthcoming. The Bermuda Triangle: the politics, principles, and pragmatics of data sharing in the history of the Human Genome Project. *J. Hist. Biol.* Accepted, in review

96. McElheny VK. 2010. *Drawing the Map of Life: Inside the Human Genome Project*. New York: Basic
97. McGuire AL, Beskow LM. 2010. Informed consent in genomics and genetic research. *Annu. Rev. Genom. Hum. Genet.* 11:361–81
98. McKusick VA, Ruddle FH. 1987. A new discipline, a new name, a new journal. *Genomics* 1:1–2
99. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, et al. 1994. A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* 266:66–71
100. Murray JC, Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema T, et al. (Coop. Hum. Link. Cent.). 1994. A comprehensive human linkage map with centimorgan density. *Science* 265:2049–54
101. Myers EW, Sutton GG, Smith HO, Adams MD, Venter JC. 2002. On the sequencing and assembly of the human genome. *PNAS* 99:4145–46
102. Natl. Hum. Genome Res. Inst. (NHGRI). 1996. *NHGRI policy regarding intellectual property of human genomic sequence: policy on availability and patenting of human genome DNA sequence produced by NHGRI pilot projects (funded under RFA HG-95-005)*. Policy Doc., Apr. 9, NHGRI, Bethesda, MD. <https://www.genome.gov/10000926>
103. Natl. Hum. Genome Res. Inst. (NHGRI). 1996. Report of the International Strategy Meeting on Human Genome Sequencing held at the Princess Hotel, Southampton, Bermuda, on 25th–28th February 1996. Duke Univ. Libr., Durham, NC. <http://hdl.handle.net/10161/7716>
104. Natl. Hum. Genome Res. Inst. (NHGRI). 2000. *NHGRI policy for release and database deposition of sequence data*. Policy Doc., Dec. 21, NHGRI, Bethesda, MD. <https://www.genome.gov/10000910/policy-on-release-of-human-genomic-sequence-data-2000>
105. Natl. Hum. Genome Res. Inst. (NHGRI). 2003. *International consortium completes Human Genome Project*. News Release, Apr. 14, NHGRI, Bethesda, MD. <http://www.genome.gov/11006929>
106. Natl. Hum. Genome Res. Inst. (NHGRI). 2009. *ENCODE consortia data release, data use, and publication policies*. Policy Doc., NHGRI, Bethesda, MD. https://www.genome.gov/pages/research/encode/mod-encode_consortia_data_release_policy_revised_11-22-09.pdf
107. Natl. Inst. Health (NIH). 2007. *Policy for sharing of data obtained in NIH supported or conducted genome-wide association studies (GWAS)*. <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html>
108. Natl. Inst. Health (NIH). 2014. *NIH genomic data sharing policy*. Not. NOT-OD-14-124, NIH, Bethesda, MD. <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html>
109. Natl. Res. Council. 1988. *Mapping and Sequencing the Human Genome*. Washington, DC: Natl. Acad. Press
110. Natl. Res. Council. 2011. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington, DC: Natl. Acad. Press
111. Off. Technol. Assess. 1984. *Human gene therapy: background paper*. Rep. OTA-BP-BA-32, Off. Technol. Assess., Washington, DC
112. Off. Technol. Assess. 1988. *Mapping our genes—genome projects: how big, how fast?* Rep. OTA-BA-373, Off. Technol. Assess., Washington, DC
113. Off. Technol. Assess. 1994. *The Human Genome Project and patenting DNA sequences*. Rep., Off. Technol. Assess., Washington, DC. https://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/8116/HxDoc_BCD_OTAfulldraft.pdf
114. Olson MV. 1995. A time to sequence. *Science* 270:394–96
115. Olson MV. 2002. The Human Genome Project: a player’s perspective. *J. Mol. Biol.* 319:931–42
116. O’Neill G. 2009. Jacobson speaks: Genetic Technologies and BRCA testing. *LabOnline*, Apr. 3. <http://www.labonline.com.au/content/life-scientist/article/jacobson-speaks-genetic-technologies-and-brca-testing-1127244197>
117. Open Hum. 2016. *About Open Humans: contribute to research & citizen science!* <https://www.openhumans.org/about>
118. Organ. Econ. Coop. Dev. (OECD). 1980. *OECD guidelines on the protection of privacy and transborder flows of personal data*. Rep., OECD, Paris. <https://www.oecd.org/sti/ieconomy/oecdguidelinesontheProtectionofPrivacyandTransborderFlowsofPersonalData.htm>
119. Organ. Econ. Coop. Dev. (OECD). 2009. *Guidelines for human biobanks and genetic research databases (HBGRDs)*. Rep., OECD, Paris. <http://www.oecd.org/sti/biotechnology/hbgrd>

120. Organ. Econ. Coop. Dev. (OECD). 2013. *OECD work on privacy*. <http://oe.cd/privacy>
121. Ostrom AL. 2015. Service research priorities in a rapidly changing context. *J. Serv. Res.* 18:127–59
122. Ostrom E. 1986. An agenda for the study of institutions. *Public Choice* 48:3–25
123. Ostrom E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge Univ. Press
124. Ostrom E. 2010. The institutional analysis and development framework and the commons. *Cornell Law Rev.* 95:807–16
125. Ostrom E, Schroeder L, Wynne S, eds. 1993. *Institutional Incentives and Sustainable Development: Infrastructure Policies in Perspective*. Boulder, CO: Westview
126. Park SS. 2014. Gene patents and the public interest: litigating *Association for Molecular Pathology v. Myriad Genetics* and lessons moving forward. *N.C. J. Law Technol.* 15:519–36
127. Parthasarathy S. 2007. *Building Genetic Medicine: Breast Cancer, Technology, and the Comparative Politics of Health Care*. Cambridge, MA: MIT Press
128. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, et al. 2015. The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum. Mutat.* 36:915–21
129. Polski MM, Ostrom E. 1999. *An institutional framework for policy analysis and design*. Work. Pap. W98-27, Workshop Polit. Theory Policy Anal., Ind. Univ., Bloomington. <https://mason.gmu.edu/~mpolski/documents/PolskiOstromIAD.pdf>
130. Qiang B. 2004. Human genome research in China. *J. Mol. Med.* 82:214–22
131. Ray T. 2014. FDA draft LDT guidance slated for release in 60 days; industry readying recommendations, objections. *GenomeWeb*, Aug. 6. <https://www.genomeweb.com/clinical-genomics/fda-draft-ldt-guidance-slated-release-60-days-industry-readying-recommendations>
132. Ray T. 2015. With BRCA Share, Quest hopes to provide field with variant database to support clinical decisions. *GenomeWeb*, May 5. <https://www.genomeweb.com/business-news/brca-share-quest-hopes-provide-field-variant-database-support-clinical-decisions>
133. Ray T. 2016. FDA holding off on finalizing regulatory guidance for lab-developed tests. *GenomeWeb*, Nov. 18. <https://www.genomeweb.com/molecular-diagnostics/fda-holding-finalizing-regulatory-guidance-lab-developed-tests>
134. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, et al. 2015. ClinGen—the Clinical Genome Resource. *N. Engl. J. Med.* 372:2235–42
135. Roberts L. 1991. Genome patent fight erupts. *Science* 254:184–86
136. Rodriguez H. 2008. International summit on proteomics data release and sharing policy. *J. Proteome Res.* 7:4609
137. Rodriguez H, Snyder M, Uhlen M, Andrews P, Beavis R, et al. 2009. Recommendations from the 2008 international summit on proteomics data release and sharing policy: the Amsterdam principles. *J. Proteome Res.* 8:3689–92
138. Rothstein MA, Knoppers BM. 2015. Harmonizing privacy laws to enable international biobank research. *J. Law Med. Ethics* 43:673–74
139. Saulnier KM, Joly Y. 2016. Locating biobanks in the Canadian privacy maze. *J. Law Med. Ethics* 44:7–19
140. Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, et al. 1996. A gene map of the human genome. *Science* 274:540–46
141. Secr. Advis. Comm. Genet. Health Soc. (SACGHS). 2008. *U.S. system of oversight of genetic testing: a response to the charge of the Secretary of Health and Human Services*. Rep., SACGHS, US Dep. Health Hum. Serv., Bethesda, MD
142. Sherkow JS, Greely HT. 2015. The history of patenting genetic material. *Annu. Rev. Genet.* 49:161–82
143. Shreeve J. 2004. *The Genome War: How Craig Venter Tried to Capture the Code of Life and Save the World*. New York: Ballantine
144. Simoncelli T, Park S. 2015. Making the case against gene patents. *Perspect. Sci.* 23:106–45
145. Smith TF. 1990. The history of the genetic sequence databases. *Genomics* 6:701–7
146. Snyder B. 2016. *New center to study genomic privacy concerns*. News Release, May 19, Vanderbilt Univ., Nashville, TN. <https://news.vanderbilt.edu/2016/05/19/new-center-to-study-genomic-privacy-concerns>

147. Spurdle AB, Healey S, Devereau A, Hogervorst FB, Monteiro AN, et al. 2012. ENIGMA—Evidence-based Network for the Interpretation of Germline Mutant Alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in *BRCA1* and *BRCA2* genes. *Hum. Mutat.* 33:2–7
148. Stevens H. 2015. Networking biology: the origins of sequence-sharing practices in genomics. *Technol. Cult.* 56:839–67
149. Strandburg KJ, Madison MJ, Frischmann BM, eds. 2017. *Governing Medical Knowledge Commons*. New York: Cambridge Univ. Press
150. Strasser BJ. 2011. The experimenter’s museum: GenBank, natural history, and the moral economies of biomedicine. *Isis* 102:60–69
151. Sulston J, Waterston R. 2011. Interview by Kathryn Maxson, Robert Cook-Deegan, and Rachel Ankeny, Nov. 15. Duke Univ. Libr., Durham, NC. <http://hdl.handle.net/10161/7692>
152. Thompson L. 1993. Healy and Collins strike a deal. *Science* 259:22–23
153. Thorisson GA, Stein LD. 2003. The SNP Consortium website: past, present and future. *Nucleic Acids Res.* 31:124–27
154. Thorogood A, Zawati MH. 2015. International guidelines for privacy in genomic biobanking (or the unexpected virtue of pluralism). *J. Law Med. Ethics* 43:690–702
155. Toronto Int. Data Rel. Workshop Authors. 2009. Prepublication data sharing. *Nature* 461:168–70
156. UN Educ. Sci. Cult. Organ. 1997. *Universal declaration on the human genome and human rights*. http://portal.unesco.org/en/ev.php-URL_ID=13177&URL_DO=DO_TOPIC&URL_SECTION=201.html
157. UN Educ. Sci. Cult. Organ. 2003. *International declaration on human genetic data*. http://portal.unesco.org/en/ev.php-URL_ID=17720&URL_DO=DO_TOPIC&URL_SECTION=201.html
158. UN Educ. Sci. Cult. Organ. 2005. *Universal declaration on bioethics and human rights*. http://portal.unesco.org/en/ev.php-URL_ID=31058&URL_DO=DO_TOPIC&URL_SECTION=201.html
159. UN Gen. Assem. 1948. *The universal declaration of human rights*. <http://www.un.org/en/universal-declaration-human-rights>
160. Unterhuber R. 1997. Germany gets warning on access to sequence data. *Nature* 387:111
161. US Dep. Energy (DOE), Natl. Inst. Health (NIH). 1990. *Understanding our genetic inheritance: the U.S. Human Genome Project: the first five years: fiscal years 1991–1995*. DOE/ER-0452P, NIH Publ. 90-1590, US DOE and US Dep. Health Hum. Serv., Washington, DC
162. US Food Drug Admin. (FDA). 2014. *Draft guidance for industry, Food and Drug Administration staff, and clinical laboratories: framework for regulatory oversight of laboratory developed tests*. Draft Guid. Doc., US FDA, Silver Spring, MD. <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm416685.pdf>
163. US Food Drug Admin. (FDA). 2016. *Use of public human genetic variant databases to support clinical validity for next generation sequencing (NGS)-based in vitro diagnostics: draft guidance for stakeholders and Food and Drug Administration staff*. Draft Guid. Doc., US FDA, Silver Spring, MD. <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM509837.pdf>
164. US Food Drug Admin. (FDA). 2017. *Discussion paper on laboratory developed tests (LDTs)*. Discuss. Pap., US FDA, Silver Spring, MD. <http://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/LaboratoryDevelopedTests/UCM536965.pdf>
165. Vail PJ, Morris B, van Kan A, Burdett BC, Moyes K, et al. 2015. Comparison of locus-specific databases for *BRCA1* and *BRCA2* variants reveals disparity in variant classification within and among databases. *J. Community Genet.* 6:351–59
166. Van Zimmeren E, Nicol D, Gold R, Carbone J, Chandrasekharan S, et al. 2014. The BRCA patent controversies: an international review of patent disputes. In *Breast Cancer Gene Research and Medical Practices: Transnational Perspectives in the Time of BRCA*, ed. S Gibbon, G Joseph, J Mozersky, A zur Nieden, S Palfner, pp. 151–74. New York: Routledge
167. Varian HR, Farrell J, Shapiro C. 2005. *The Economics of Information Technology: An Introduction*. Cambridge, UK: Cambridge Univ. Press

168. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. *Science* 291:1304–51
169. Venter JC, Adams MD, Sutton GG, Kerlavage AR, Smith HO, Hunkapiller M. 1998. Shotgun sequencing of the human genome. *Science* 280:1540–42
170. Waterston RH, Lander ES, Sulston JE. 2002. On the sequencing of the human genome. *PNAS* 99:3712–16
171. Waterston RH, Sulston JE. 1998. The Human Genome Project: reaching the finish line. *Science* 282:53–54
172. Watson JD. 1990. The Human Genome Project: past, present, and future. *Science* 248:44–49
173. Watson JD, Crick FH. 1953. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171:737–38
174. Watson JD, Crick FH. 1953. The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* 18:123–31
175. Wellcome Trust. 2003. *Sharing data from large-scale biological research projects: a system of tripartite responsibility*. Rep., Wellcome Trust, London. <https://www.genome.gov/pages/research/wellcomereport0303.pdf>
176. Wellcome Trust. 2017. *Statement on genome data release*. <https://wellcome.ac.uk/funding/managing-grant/statement-genome-data-release>
177. Wellcome Trust, Natl. Inst. Health (NIH), US Dep. Energy (DOE), 164th Genome Technol. Comm., Jpn. Soc. Promot. Sci., UK Med. Res. Council. 1998. *Third International Strategy Meeting on Human Genome Sequencing, Hamilton Princess Hotel, Bermuda, 27th February–1st March 1998*. Cent. Public Genom. Res. Files, Duke Univ. Libr., Durham, NC. <http://hdl.handle.net/10161/7745>
178. Williams H. 2013. Intellectual property rights and innovation: evidence from the human genome. *J. Polit. Econ.* 121:1–27
179. Williamson AR. 1999. The Merck Gene Index project. *Drug Discov. Today* 4:115–22
180. Wilson R. 1999. How the worm was won: the *C. elegans* genome sequencing project. *Trends Genet.* 15:51–58
181. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, et al. 1995. Identification of the breast cancer susceptibility gene *BRCA2*. *Nature* 378:789–92
182. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, et al. 1994. Localization of a breast cancer susceptibility gene, *BRCA2*, to chromosome 13q12-13. *Science* 265:2088–90

Contents

The Clinic Is My Laboratory: Life as a Clinical Geneticist <i>Judith G. Hall</i>	1
A Conversation with Kurt and Rochelle Hirschhorn <i>Kurt Hirschhorn, Rochelle Hirschhorn, and Joel N. Hirschhorn</i>	31
Gene Regulatory Elements, Major Drivers of Human Disease <i>Sumantra Chatterjee and Nadav Abituv</i>	45
The Microbiome and Human Biology <i>Rob Knight, Chris Callewaert, Clarisse Marotz, Embriette R. Hyde, Justine W. Debelius, Daniel McDonald, and Mitchell L. Sogin</i>	65
Recent Advancements in DNA Damage–Transcription Crosstalk and High-Resolution Mapping of DNA Breaks <i>Valerio Vitelli, Alessandro Galbiati, Fabio Iannelli, Fabio Pessina, Sheetal Sharma, and Fabrizio d’Adda di Fagagna</i>	87
Cognitive Dysfunctions in Intellectual Disabilities: The Contributions of the Ras-MAPK and PI3K-AKT-mTOR Pathways <i>Sarah C. Borrie, Hilde Brems, Eric Legius, and Claudia Bagni</i>	115
Tailoring Medulloblastoma Treatment Through Genomics: Making a Change, One Subgroup at a Time <i>Borja L. Holgado, Ana Guerreiro Stucklin, Livia Garzia, Craig Daniels, and Michael D. Taylor</i>	143
The Yin and Yang of Autism Genetics: How Rare De Novo and Common Variations Affect Liability <i>Pauline Chaste, Kathryn Roeder, and Bernie Devlin</i>	167
Advances in Preimplantation Genetic Testing for Monogenic Disease and Aneuploidy <i>Nathan R. Treff and Rebekah S. Zimmerman</i>	189
Application of Panel-Based Tests for Inherited Risk of Cancer <i>Payal D. Shah and Katherine L. Nathanson</i>	201
Gene and Variant Annotation for Mendelian Disorders in the Era of Advanced Sequencing Technologies <i>Samya Chakravorty and Madhuri Hegde</i>	229

Recent Advances in Mitochondrial Disease <i>Lyndsey Craven, Charlotte L. Alston, Robert W. Taylor, and Doug M. Turnbull</i>	257
The Genetic Diversity of the Americas <i>Kaustubb Adbikari, Juan Camilo Chacón-Duque, Javier Mendoza-Revilla, Macarena Fuentes-Guajardo, and Andrés Ruiz-Linares</i>	277
On the Evolution of Lactase Persistence in Humans <i>Laure Séguérel and Céline Bon</i>	297
A Robust Framework for Microbial Archaeology <i>Christina Warinner, Alexander Herbig, Allison Mann, James A. Fellows Yates, Clemens L. Weiß, Hernán A. Burbano, Ludovic Orlando, and Johannes Krause</i>	321
Participatory Genomic Research: Ethical Issues from the Bottom Up to the Top Down <i>Heide Aungst, Jennifer R. Fishman, and Michelle L. McGowan</i>	357
Precisely Where Are We Going? Charting the New Terrain of Precision Prevention <i>Karen M. Meagher, Michelle L. McGowan, Richard A. Settersten Jr., Jennifer R. Fishman, and Eric T. Juengst</i>	369
Sharing Data to Build a Medical Information Commons: From Bermuda to the Global Alliance <i>Robert Cook-Deegan, Rachel A. Ankeny, and Kathryn Maxson Jones</i>	389

Errata

An online log of corrections to *Annual Review of Genomics and Human Genetics* articles may be found at <http://www.annualreviews.org/errata/genom>