ANNUAL REVIEWS

*Annual Review of Genomics and Human Genetics*

# The Genomic Commons

Jorge L. Contreras[1] and Bartha M. Knoppers[2]

[1]S.J. Quinney College of Law and School of Medicine, University of Utah, Salt Lake City, Utah 84112, USA; email: jorge.contreras@law.utah.edu

[2]Centre of Genomics and Policy and Department of Medicine, McGill University, Montreal, Quebec H3A 0G1, Canada; email: bartha.knoppers@mcgill.ca

Annu. Rev. Genom. Hum. Genet. 2018. 19:429–53

First published as a Review in Advance on January 25, 2018

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

https://doi.org/10.1146/annurev-genom-083117-021552

Copyright © 2018 by Annual Reviews. All rights reserved

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

## Keywords

genome, commons, data sharing, data release, Bermuda Principles, Human Genome Project, law, ethics, informed consent, privacy, GA4GH

## Abstract

Over its 30 or so years of existence, the genomic commons—the worldwide collection of publicly accessible repositories of human and nonhuman genomic data—has enjoyed remarkable, perhaps unprecedented, success. Thanks to the rapid public data release policies initiated by the Human Genome Project, free access to a vast array of scientific data is now the norm, not only in genomics, but in scientific disciplines of all descriptions. And far from being a monolithic creation of bureaucratic fiat, the genomic commons is an exemplar of polycentric, multistakeholder governance. But like all dynamic and rapidly evolving systems, the genomic commons is not without its challenges. Issues involving scientific priority, intellectual property, individual privacy, and informed consent, in an environment of data sets of exponentially expanding size and complexity, must be addressed in the near term. In this review, we describe the characteristics and unique history of the genomic commons, then address some of the trends, challenges, and opportunities that we envision for this valuable public resource in the years to come.

## INTRODUCTION

Publicly accessible repositories around the world today offer researchers access to a vast quantity of human and nonhuman genomic data. According to the Joint Genome Institute's Genomes OnLine Database (GOLD), data from more than 30,000 sequencing projects around the world, representing more than 280,000 different organisms, were available to researchers as of October 2017 (65, 88). With respect to human genomic data, projects of increasingly ambitious scope have continued to contribute to this public data trove, ranging from the Human Genome Project (HGP), which in 2001 released sequence data derived from a handful of individuals (62), to the 1000 Genomes Project, which was completed in 2012 (1); the United Kingdom's 100,000 Genomes Project (47); and the US Precision Medicine Initiative and *All of Us* Research Program launched in 2016, the latter of which is seeking to collect and release data from approximately one million Americans (97). This vast global aggregation of genomic data has been termed the genome commons or genomic commons (14, 15, 77).[1]

The terminology of the commons evokes two related but distinct bodies of literature: that of open access and the public domain (8, 9, 27, 86, 110) and the more complex construct of "common pool resources" pioneered by Elinor Ostrom and other new institutional economic theorists (8, 104).[2] The public availability of genomic data has yielded advances in medical genetics, molecular biology, and bioinformatics; has reduced research costs; and has enabled greater reproducibility of research results, all of which have contributed to the acceleration of scientific discovery and biomedical research (1, 12, 131).

But the existence of this valuable public resource was by no means assured when the HGP was initiated in the late 1980s. In fact, it was widely assumed that the majority of genomic data would be placed in proprietary databases, protected by intellectual property or confidentiality restrictions, and made available only under costly subscription agreements (83). The fact that the genomic commons is today a global, public resource owes much to a 1996 accord reached by scientific leaders and policy makers in Bermuda. The groundbreaking Bermuda Principles (61) required that all DNA sequence data generated by the HGP be released to the public a mere 24 hours after generation, a stark contrast to the typical release of data following publication of the associated analysis, usually years after the generation of the data (14, 15). The Bermuda Principles arose from an early recognition that rapid and efficient data sharing would be necessary to coordinate activity among the geographically dispersed laboratories working on the massive project. But site coordination was not the only factor motivating the unorthodox rapid release requirements of the Bermuda Principles. More importantly, this approach emerged from the conviction among project leaders, many of whom were trained in a culture of liberal exchange of ideas, research materials, and data, that the rapid release of genomic data—originally data from laboratories studying nematodes and other model organisms—was necessary for the advancement of scientific research and the public good (62, 83).

The Bermuda Principles continue to shape the data release practices of the genomics community today and have established rapid prepublication data release as the norm in this and other fields (69). Advances in science and technology, however, together with challenging ethical and legal issues, have given rise to policy considerations not foreseen during the HGP. Among these

[1] Another variant refers to the "genetic commons" (7).

[2] Other current and proposed aggregations of scientific data have adopted this commons designation; there are a proteome commons (56), a microbial research commons (113), a global crop commons (113), a generally applicable science commons (132), and ambitious plans for a cross-cutting medical information commons (28, 36). A similar large-scale project in the area of cytology is the proposed international Human Cell Atlas (2).

are the need to protect human subject data, even at the genomic level, and the desire of scientists who generate large data sets to analyze and publish their research before others make use of it. The emergence of these considerations has led to an evolution of genomics data release policies and norms that are more restrictive and complex than those of the HGP but nevertheless preserve the fundamental shared nature of the genomic commons. In this respect, the genomic commons resembles a managed common pool resource of the type elucidated by Ostrom, rather than the simpler public-domain/public-goods model that is often associated with basic scientific research (17, 111).

In this review, we summarize the principal formal characteristics of the genomic commons, organized according to the categories laid out in Ostrom's Institutional Analysis and Development (IAD) framework (104), which was originally created to assess and compare tangible common pool resources such as fisheries and pastures, but which has been applied more recently by Ostrom & Hess (106) and Madison et al. (79) to shared intangible resources. The characteristics considered in this framework include the nature of the shared resource, its stakeholders, and its rules in use. Our focus on rules in use encompasses the shared governance structure of the commons, the manner in which individual privacy concerns are addressed, and the role played by exogenous laws and regulations. The IAD analytical framework is useful both in assessing how the genomic commons has changed over time in response to legal, social, and commercial pressures and in comparing it to other sets of shared intangible resources. We believe that the insights gained by studying the innovative governance, access, and usage policies of the genomic commons can help future planners develop, and anticipate issues that may arise in connection with, new commons of scientific information.

## RESOURCE CHARACTERISTICS

### What Is the Genomic Commons?

There is no single, generally accepted definition of the genomic commons. For the purposes of this review, we limit our discussion to the worldwide collection of genomic data that is generally available for public use. The types of data encompassed by this definition include germline and somatic DNA and RNA sequences, haplotype and single-nucleotide polymorphism (SNP) data, allele frequencies, variant data, epigenetic data, genotype–phenotype associations, annotations, and related phenotypic, genealogical, environmental, and clinical data. Because our focus is on genomics, which itself is a vast area with a unique history and place in the scientific world, we do not address data pertaining primarily to other biological systems or structures, including other "-omics" fields (proteomics, transcriptomics, and metabolomics), or to biological systems operating at a level higher than DNA (cellular or organ systems).

The data that make up the genomic commons are not stored in a single repository but are instead distributed among public and private databases around the world that operate, by and large, according to a consistent set of principles regarding openness and access (see the section titled Rules in Use). The principal databases for the deposit of genomic sequence data are GenBank, which is administered by the National Center for Biotechnology Information (NCBI) [a division of the National Library of Medicine within the National Institutes of Health (NIH)], the European Molecular Biology Library (EMBL), and the DNA Data Bank of Japan. NCBI also maintains the RefSeq database, which consolidates and annotates much of the sequence data found in GenBank. In addition to these repositories, researchers have experimented with new approaches to the storage, handling, and public release of large genomic data sets, including the use of commercial cloud services (2, 95, 118). And while genomic data repositories have

traditionally been maintained as stand-alone resources, recent initiatives have sought to connect independent repositories through federated systems, including tools for searching, integrating, analyzing, and providing other value-added features (23, 49, 111). For example, the Genome Aggregation Database (gnomAD) [previously known as the Exome Aggregation Consortium (ExAC) database] hosted by the Broad Institute aggregates exome and genome sequence data from a variety of large-scale sequencing projects, including 1000 Genomes and The Cancer Genome Atlas (TCGA) (67, 68). And the National Cancer Institute's Genomic Data Commons links multiple cancer genomic projects across the institute in a unified data repository and makes available data analysis, visualization, and exploration tools to enable interaction with the affiliated data (90).

In addition to DNA sequence data, genomic studies generate data relating to associations between particular genetic markers and disease risk and other physiological traits. The compilation of disease-specific genetic data dates to the early days of medical genetics, when particular (and hard-won) genetic markers were associated with diseases such as cystic fibrosis and breast and ovarian cancer. Databases such as CFTR2 (**https://cftr2.org**) and the BRCA Exchange (**https://ucscgenomics.soe.ucsc.edu/brca-exchange**) continue to catalog genetic variants relating to these diseases. More broadly, NCBI's ClinVar (**https://www.ncbi.nlm.nih.gov/clinvar**) collects thousands of clinically significant genetic variants in a public archive (75).

Beyond the association of particular genetic variants with specific diseases, large-scale genome-wide association studies (GWASs) now enable the association of multiple genetic variants with disease risk and health conditions. This type of data, which is more complex to record, search, and correlate than raw sequence data or disease-specific variant data, can be housed in databases such as the Database of Genotypes and Phenotypes (dbGaP), also operated by the National Library of Medicine. dbGaP accommodates a large range of phenotypic data, which include elements such as deidentified subject age, ethnicity, weight, demographics, exposure, disease state, and behavioral factors, as well as study documentation and statistical results. Given the potential sensitivity of phenotypic data, dbGaP allows access to data on two levels: open and controlled. Open data access is available to the general public via the Internet and includes nonsensitive summary data, generally in aggregated and anonymized form. Data from the controlled portion of dbGaP may be accessed only under conditions specified by the data generator, often requiring certification of the user's identity and research purpose. A third intermediary and automated tier for access to minimal-risk data is under development (38).

Because our focus is on data that are, at least to some degree, accessible to and usable by researchers around the world, we do not consider proprietary data held by private-sector firms to constitute part of the genomic commons. As noted above, the commons construct contemplates a degree of shared use, access, and governance of a common resource, which is lacking with respect to private data resources. This being said, several private-sector initiatives have contributed substantial data to the genomic commons, and we discuss these below.

We define the genomic commons specifically as an aggregation of shared genomic data, and exclude from this definition physical samples and specimens of all kinds (seeds, cell lines, blood, tissue, and DNA). While it is understood that genomic data must be derived from physical DNA, the rules, policies, and norms governing the operation of and access to repositories (biobanks) of physical samples (both human and nonhuman) are significantly different and more restrictive than those governing data per se (52, 103, 111, 112). These differences are material enough that we consider biobanks and other repositories of physical material to occupy a separate place in the research ecosystem than genomic data repositories. Accordingly, we do not address issues pertaining to biobanks in this review.

It is also worth noting that our discussion focuses primarily on human genomic data. While we do not expressly exclude nonhuman data from the genomic commons [and we acknowledge

that the sheer quantity of nonhuman genomic data—e.g., for plants, model organisms, and the microbiome—vastly exceeds the quantity of available human genomic data (65, 88)], legal and ethical issues concerning human genomic data have driven many key policy and governance issues. Thus, with a few exceptions, we focus on the manner in which genomic data policies have been shaped by issues arising from human subject data, as data about nonhuman organisms do not raise the issues of informed consent and privacy typically associated with human data.

### Background Legal Environment

Madison et al. (79) emphasized that an understanding of the natural environment in which a knowledge commons exists is critical to understanding the attributes and operation of that commons. In the case of intangible resources, this natural environment includes the legal rules that govern rights and permissions with respect to the elements of the common resource. The genomic commons presents a complex picture, as it embodies both biomedical discoveries, which are typically addressed via the patent system, and large aggregations of data, which are typically addressed via access restrictions, contractual obligations, and copyright rules. In addition, the storage, transmission, and use of human subject data are regulated under a wide variety of national laws. We discuss each of these issues below in terms of the governance policies developed to address them within the genomic commons.

## STAKEHOLDERS

Much of the literature concerning common resource governance seeks to describe and assess the attributes of the communities that share common resources. This approach is equally applicable to the genomic commons. While genomic data release policies are typically drafted and adopted by funding agencies, the NIH in particular has given substantial deference to the views and opinions of the scientific community and also seeks to represent the interests of the general public. Likewise, responsibility for the creation, use, maintenance, and curation of the genomic commons is shared among different groups and is not directed by government alone. Thus, the development of genomic data release policies has been a polycentric process of negotiation and compromise. As described below, the principal stakeholder communities relevant to the genomic commons, both initially and as it has evolved over time, include funders, data generators, data users, scientific leaders, data intermediaries, individuals who are the sources of genomic data (data subjects), and the public.

### Funders and the State

The NIH and the US Department of Energy, the principal funders of the HGP, together with their counterparts at the Wellcome Trust in the United Kingdom, exerted significant influence over the technical and policy direction of the HGP. Major funders continue to play a crucial role in genomic research, although today smaller-scale and disease-specific studies are frequently conducted in the private sector without significant governmental or philanthropic support. The role of state agencies in genomic research has also evolved since the HGP. Governmental organizations such as the National Library of Medicine now play a key role in maintaining and curating data that are generated by both NIH-funded and other researchers (20, 77). Likewise, as interactions and collaboration among state and private-sector actors have continued to evolve, the role of the state in guiding and influencing the direction of genomic research, including private-sector research, has become more complex and multifaceted (20). In addition, a range of international and intergovernmental agencies have begun to play an increasingly prominent role in the governance and oversight of genomic data and other resources (112). Thus, in some cases, multiple

funding sources and governmental agencies have a significant role in developing policy surrounding genomic research and data use.

## Data Generators

Prior to the HGP, genetic research was conducted in hundreds of laboratories around the world and funded primarily by small grants directed toward the investigation of specific genetically linked diseases such as cystic fibrosis, Huntington's disease, and breast cancer. The HGP, by contrast, approached mapping the human genome as a campaign of large-scale data production. The intensity of this work, the equipment required to perform it, and the degree of specialization that it demanded led to the creation of a new breed of scientist: one whose principal research aim was the generation of data rather than the development and testing of scientific hypotheses. Unlike most other researchers, data-generating researchers face challenges when it comes to publishing their work in traditional scientific journals, as the creation of large data sets has not traditionally been viewed as meriting the same degree of recognition as traditional hypothesis-driven research (15, 121). Researchers from major sequencing centers have made significant contributions to the development of genomic data policy over the years.

## Data Users

Prior to the completion of the HGP, researchers studying a particular genetic disease devoted substantial time and effort to isolating and sequencing one or a few relevant genes—work that often would take years of painstaking trial-and-error experimentation. The data generated by the HGP and subsequent projects eliminated the need for researchers to conduct much of this groundwork. Unlike the original close-knit community of data generators at large-scale sequencing centers, the population of genomic data users no longer forms a coherent community. Today, users of genomic data include researchers across the world in nearly every biological discipline, encompassing basic discovery-oriented research, drug development, and clinical care, and now extending to more distant areas, such as agriculture and environmental science. Furthermore, the users of genomic data are no longer limited to academic or corporate researchers, but have expanded to include clinicians, genetic counselors, and patients themselves. The growth of this diverse community of data users and its divergence from the more tightly knit community of data generators have had a significant impact on policies for the release and use of genomic data.

## Scientific Leaders

Many of the individuals involved in the planning and execution stages of the HGP were world-renowned researchers, including several Nobel laureates. This early involvement by preeminent scientists was critical to the HGP's success and gave the group's decisions a gravitas that was important in attracting both public and political support (25, 83). It also engendered among the project's leadership a sense of public stewardship that contributed to the community-spirited character of several HGP policies (128). This public spirit appears to have developed independently of the role that such individuals may have played as funders, data generators, or data users. As such, the role of scientific leaders in genomic policy development is worth considering separately from these other categories.

## Data Intermediaries

The researchers that generate genomic data seldom have effective means to make it broadly available or accessible to others. In most cases, researchers rely on data intermediaries, whether

these are scientific journals that publish their analyses or centralized database managers that host large quantities of raw data. Data intermediaries may operate as either commercial entities (as in the case of commercial publishers and paid database services) or nonprofit/governmental entities (such as the National Library of Medicine and open access journals, such as those published by the Public Library of Science). Not surprisingly, the interests of commercial and noncommercial data intermediaries often differ, most notably in terms of pricing access to information. Nevertheless, these stakeholders also share several common motivations, including the desire to disseminate information in ways that are effective, secure, and accurate, and the need to maintain some level of financial sustainability. Recently, the critical role of scientific journals in the creation and sustainability of the genomic commons has been recognized, particularly with respect to the need to offer meaningful and career-enhancing publication opportunities to data-generating researchers (63, 121).

## Data Subjects

Because the goal of the HGP was to generate a baseline map of the human genome without regard to the particular physiological and pathological traits associated with genetic variation among individuals, the data generated by the HGP were anonymous and retained no association with the individuals whose DNA was sequenced (93). In later projects, however, and particularly with the commencement of large-scale GWASs, concerns emerged about the potential identification of human subjects from their genomic data, with associated concerns regarding privacy and consent (121) (discussed below in the section titled Privacy and Data Protection). Because a GWAS seeks to associate genotypic information (e.g., genetic markers) with disease risk, it is necessary to correlate information regarding donor demographics, disease state, and treatment with genomic data. The prospect of releasing such clinical and phenotypic data to the public led to the recognition of human data subjects as important stakeholders in the genomic data commons (63, 103).

## The Public

In addition to individuals who contribute their own DNA and other information to the genomic commons, the general public is an important stakeholder in genomic research. First, members of the public who are affected, directly or indirectly, by genetic disorders often support or participate in advocacy and disease interest groups. The members of these groups often possess a high degree of familiarity with the scientific literature and have both the motivation and the financial wherewithal to fund research and advocate for policy changes (76, 120). Members of the public beyond health advocacy groups have also taken an interest in, and begun to play an active role in, genomic research and the data-sharing practices of genomics researchers (54, 81, 120). Public engagement in health-related science, particularly genomics, has actively been fostered by governmental programs such as the Patient-Centered Outcomes Research Institute, established in 2010 (115), and public interest in genetics and genomics has been substantially bolstered by the growth of direct-to-consumer genomic information services such as 23andMe and Ancestry DNA (21). And, at the broadest level, government-sponsored research is largely taxpayer funded, giving taxpayers and the citizenry as a whole a legitimate interest in the direction and results of government-funded genomics research. For all of these reasons, the public has become an important stakeholder in both the creation and use of the genomic commons.

## RULES IN USE

Under the IAD framework, the rules in use or governance structure of a commons system constitutes its third primary attribute. When considering physical resource commons, rules in use typically allocate access and usage rights with respect to a preexisting resource, and while such rules necessarily affect its sustainability and the rate at which it is depleted (and replenished), they do not create or define it. Commons of intangible goods such as knowledge, however, are not depleted by use and can accommodate a virtually unlimited number of users. This is not to say, however, that knowledge commons do not present substantial governance issues. As observed by Madison et al. (79), the rules governing a knowledge commons dictate its very nature, including, in the case of the genomic commons, features such as its size and content, the speed at which data are deposited in it, and when and how it can be accessed and used (17). These rules in use include both formal (de jure) rules and informal (but often forceful) norms that govern members' behavior. In the case of the genomic commons, the formal rules established at the outset of the HGP were strongly influenced by the norms of the scientific community at the time (see below).

Although the formal rules applicable to NIH-funded genomics projects are ultimately promulgated as official agency policies, the environment in which they are developed resembles the polycentric governance systems identified by Ostrom more than top-down rulemaking models. As observed by Ostrom, "[p]olicy changes are experiments based on more or less informed expectations about potential outcomes" (105, p. 243). Thus, the rules of the genomic commons have evolved over numerous iterations to address the concerns, experiences, and preferences of the diverse stakeholder groups involved in genomics research.

It is also worth recalling that the institutional rules and structures surrounding genomic data were created during the early days of the HGP, when this resource was several orders of magnitude smaller than it is today. It is doubtful that any shared resource described in the literature of traditional common pool resources has grown at a rate even close to that of the genomic commons. Considered in this light, the rules and norms established by the HGP, which continue to shape policy today, have exhibited remarkable resilience.

### Governance: Data Sharing and Access

As noted above, one of the key principles motivating the creation of genomic data release policies under the HGP was the desire that genomic data be shared as broadly and rapidly as possible. The strong desire for data accessibility and sharing has continued to influence the rules governing the genomic commons today.

**The Bermuda Principles and the data-sharing legacy.** In 1996, the Bermuda Principles established that all DNA sequence information from the HGP should be made "freely available and in the public domain in order to encourage research and development and to maximise its benefit to society" (61). Most importantly, the Bermuda Principles required that these data be released to public databases a mere 24 hours after generation. These rapid prepublication data release requirements reflected the strong open science norms of the model organism researchers who formed the core of the HGP planning group (83, 100, 111). They also echoed a sense among the leaders of the project that human genomic data possessed a special and unique character, expressed in the National Research Council's 1988 recommendation that all HGP data be "provided in an accessible form to the general research community worldwide" (99, p. 8). In addition to these ideological underpinnings, there were practical reasons for the HGP's aggressive data release requirements. First, they facilitated project coordination by requiring sequencing centers to utilize comparable data quality measures and reporting standards and enabling them to avoid

duplication of effort and optimize their respective tasks. Second, it was believed that rapid data release was the best way to maximize scientific advancement, thereby accelerating health-related discoveries and justifying the enormous cost of the HGP (15, 83).

Despite these practical and societal benefits, the rapid data release requirements of Bermuda also effectively eliminated the ability of data generators to publish their analyses and conclusions before others could access the data they had produced (15, 121). For the HGP sequencing centers, which were paid handsomely for their work and which earned global reputations as the leaders in the emerging field of large-scale DNA sequencing, this may not have been a serious drawback. But once the HGP ended its work in the early 2000s, researchers in fields as varied as oncology, virology, and microbiology began to undertake genomic research projects and were less sanguine about their inability to analyze and publish papers regarding the data they had generated (15, 26).

In 2003, the Wellcome Trust convened a summit of stakeholders in Fort Lauderdale, Florida, to revisit rapid data release requirements in the post-HGP world. While the Fort Lauderdale participants "enthusiastically reaffirmed" the 1996 Bermuda Principles (129, p. 2), they also expressed reservations about extending rapid data release requirements to all areas of genetic research. The resulting report thus distinguished between HGP-like community resource projects (CRPs) that were "specifically devised and implemented to create a set of data, reagents or other material whose primary utility will be as a resource for the broad scientific community" (129, p. 2) and hypothesis-driven research, in which the goal is to answer a particular scientific question through the analysis of experimental data and in which success is typically measured by the degree to which the question is answered rather than by the size or completeness of a particular data set. Researchers engaged primarily in hypothesis-driven research generally resisted giving data away before their conclusions were published, lest a competing group scoop them using their own data. Accordingly, the Fort Lauderdale participants agreed that, while the 24-hour rapid release rules of Bermuda would continue to apply to CRPs, they would not apply to research projects other than CRPs (129).

In the succeeding years, several large-scale genomic research projects were launched with increasingly sophisticated rules regarding data release and sharing. For example, the Genetic Association Information Network (GAIN) arose in 2006 as a public–private collaboration among the Foundation for the National Institutes of Health (FNIH), the NIH, and several private companies to conduct GWASs on six common diseases (45). Researchers were required to sign agreements calling for immediate release of GAIN data. Yet the data, which were deposited in the NIH's dbGaP database, were segregated into open and controlled access portions to protect the privacy of research subjects. Researchers wishing to access GAIN data from the controlled portion of the database needed to obtain approval from the GAIN data access committee (DAC) (45) (see discussion below). Once approved, researchers were required to agree to keep the data secure, use it only for approved research purposes, refrain from patenting the data or conclusions drawn directly from it, acknowledge data generators in their publications, and refrain from attempting to identify individual data subjects (46). In addition, GAIN adopted the first major genomic data release policy to introduce a temporal restriction on the use of shared data. Thus, in order to secure an exclusive period of use and analysis for the GAIN participants, data users were prohibited from publishing and making presentations based on GAIN data for an embargo period generally in the range of 6–12 months (45). Similar restrictions and procedures were included in the NIH's 2007 policy covering GWASs (15, 94) and the policies of the Wellcome Trust Case Control Consortium discussed below (130).

The NIH's most recent policy addressing genomic data release, the 2014 Genomic Data Sharing policy (96), which applies to all NIH-funded projects generating genomic data, allows

data generators to retain human genomic data for up to six months while they analyze it, and to retain nonhuman genomic data until the time that results based on the data are published. This latest policy, with its built-in data-holding periods, represents a retreat from the rapid data release requirements established by the Bermuda Principles. The compromise evidenced by this policy may have been necessitated by the continued broadening of genomic research into fields in which researchers are unfamiliar (and possibly uncomfortable) with the rapid data release norms that originated in the early genomics community (18).

**Private-sector projects.** In addition to the HGP and other public-sector sequencing efforts, several private-sector initiatives have made substantial contributions to the genomic commons, many with data release policies informed by the principles established in Bermuda and Fort Lauderdale. These private-sector initiatives were important, as they both reacted to and were closely observed by the publicly funded projects that operated alongside them. While many private-sector research efforts have been undertaken within the highly proprietary environments of pharmaceutical and biotechnology companies, the existence of privately funded activities that contribute to the public genomic commons suggests that the common resource structure established by the NIH and its publicly funded projects has taken hold as an accepted mode of organizing genomic research, even in the private sector.

The first notable private-sector contribution to the genomic commons occurred in 1994, when pharmaceutical giant Merck began to release to the public a large number of expressed sequence tags (ESTs) (85). By 1998, the so-called Merck Gene Index included more than 800,000 of these ESTs, which were also deposited in the public GenBank. It is believed that Merck chose to release these potentially valuable assets due to a combination of philanthropic intent and corporate self-interest (i.e., preempting patenting of ESTs by biotech companies, several of which had already announced business plans that involved the patenting and licensing of genetic data to the pharmaceutical industry) (15, 109).

Similar strategies were followed by the SNP Consortium in 1999 (57) and the International Serious Adverse Events Consortium (iSAEC) in 2007 (22), each of which raised funding from the pharmaceutical sector to support the discovery of genomic markers, which were then made freely available to the public. Unlike the Merck Gene Index, these efforts represented collaborations among multiple private firms and were organized as nonprofit, membership-based corporations governed by boards comprising representatives from each of their funding companies (24). The research funded by these consortia was carried out largely under contract by academic research centers. Like the other policies discussed in this section, those adopted by the SNP Consortium and iSAEC imposed various security, research, and nonpatenting restrictions. They also secured for data-generating researchers a period of exclusivity (up to 12 months) during which they had exclusive access to the data for analysis and the preparation of publications (15). The fact that these commons-based organizational structures emerged in the private sector is a testament to the unusually strong shared community norms within the genomics research community, even among researchers employed by industry.

A different approach was taken by the Personal Genome Project (PGP) led by Harvard researcher George Church (**http://www.personalgenomes.org**). The PGP, launched in 2008 with substantial fanfare, asks volunteers to submit DNA samples and accompanying phenotypic data. Researchers are then authorized to analyze the submitted samples and publish any resulting genomic information on the PGP website. All such data are released without restriction under the No Rights Reserved (CC0) Creative Commons copyright waiver, with a further waiver of all data subject privacy rights. The PGP approach differs markedly from that of the government and private-sector projects described above in that it dispenses entirely with any attempt to protect

or restrict the use of genomic data (13). As such, the PGP returns to the broad public-domain character of the Bermuda Principles and may signal, at least among some researchers and data subjects, a reconceptualization of genomic data as a public good.

**International governance of the commons.** At about the same time that the Bermuda Principles were adopted, there was a concomitant surge in international policy activity. Indeed, 1996 and 1997 were banner years for the genomic commons, as the Geneva-based Human Genome Organisation (HUGO); the United Nations Educational, Scientific, and Cultural Organization (UNESCO); and the Organisation for Economic Co-operation and Development (OECD) each produced a framework for the ethical and governance underpinnings of genetic research.

In 1996, HUGO's ethics committee adopted four principles that became the hallmark of all its statements over the next decade, including that "the human genome is part of the common heritage of humanity" (58). In 2002, this committee specifically addressed human genomic databases, considering them to be global public goods—that is, "goods. . .whose scope extends worldwide, are enjoyable by all with no groups excluded, and, when consumed by one individual are not depleted for others" (59, p. 208). Among other things, it encouraged the "free flow of data" and the fair and equitable distribution of benefits (59, p. 208).

Then, in 1997, UNESCO adopted the Universal Declaration on the Human Genome and Human Rights—the first international instrument to frame genetic research prospectively. Founding this declaration on the principle of human dignity, it considered the human genome in "a symbolic sense" to be "the heritage of humanity" (125, art. 1). Among the principles guiding research on the human genome should be making the "[b]enefits from advances in biology, genetics and medicine, concerning the human genome. . .available to all" (125, art. 12). This sentiment was echoed in the preamble of the 1997 Convention on Human Rights and Biomedicine of the Council of Europe, which maintained "that all humanity may enjoy the benefits of biology and medicine" (30). This convention has since been signed and ratified by 29 countries (33). It bears noting that these two 1997 instruments could be seen as instantiating the human right to share in scientific advancement and its benefits, as found in article 27 of the Universal Declaration of Human Rights (discussed below).

The complete sequence map of the human genome was published in 2003, which coincided with another phenomenon—the emergence of national efforts to develop population biobanks (52). deCODE in Iceland (1996), the Estonian Genome Project (2000), the UK Biobank (2002), and CARTaGENE in Quebec (2003), to name but a few, all launched longitudinal population cohorts (73). Also in 2003, the Public Population Project in Genomics and Society ($P^3G$) (107) built policies and tools to foster international data transfer compatibility, focusing on enabling both retrospective use of legacy data and samples and prospective harmonization across biobanks for a more efficient gain in statistical significance (10).

The recognition and more formal governance and management of these infrastructures were consolidated in the OECD's 2009 Guidelines on Human Biobanks and Genetic Research Databases (102). Biological samples and environmental, sociodemographic, and medical data are still collected in national initiatives to be shared as part of an infrastructure science that supports discovery science (114). The majority of such resources are largely epidemiological in nature and foresee ongoing access and use for more specific clinical or socioenvironmental studies.

The Council of Europe has remained active in the ethics and governance of the genomic commons, as witnessed by its 2016 Recommendation on Research on Biological Materials of Human Origin (31) and Recommendation on the Processing of Personal Health-Related Data for Insurance Purposes, Including Data Resulting from Genetic Tests (32). These recommendations

are largely a consolidation of a decade of self-regulatory professional guidance on the governance of biobanking and on genetic research in general.

The 2014 launch of the Global Alliance for Genomics and Health (GA4GH)—an organization dedicated to creating the tools and policies for international data sharing (50)—marks nearly 30 years of effort to build the genomic commons. What is unique about this recent initiative is that its data-sharing ethos is founded on the goal of realizing the hitherto dormant legal human right of all individuals "to share in scientific advancement and its benefits," as recognized by the 1947 United Nations Universal Declaration of Human Rights (123, art. 27). This largely dormant right traces its origins not only to the Universal Declaration of Human Rights but also to the 1966 International Covenant on Economic, Social, and Cultural Rights (124, art. 15), the latter of which has been signed and ratified by 165 countries and thus affects domestic national law. GA4GH provides not only a framework (71) but also a data-sharing lexicon and policies and procedures on (*a*) consent (including ethics waivers under certain conditions for the use of legacy data); (*b*) privacy and security, stressing the use of a proportionate risk–benefit ratio focused on the real risks of data-intensive research; and (*c*) accountability, where the responsibilities of stakeholders are addressed, especially those related to actively sharing consented-to data. In 2017, GA4GH's international ethics review recognition policy (37, 50) was also completed, with the goal of providing the core elements to enable mutual recognition between institutional review boards and ethics review boards involved in multijurisdictional data-intensive research. Later that year, GA4GH entered into formal collaborations with 15 international genomic data initiatives (51). Practical procedural guidance is under development, but there is no doubt that multiple and contradictory ethics reviews (like access requirements) impede data sharing (37).

## Privacy and Data Protection

Safeguarding identifiable human subject data contained within or discernible from genomic data sets, together with the need to comply with a bewildering range of data and health record privacy regulations around the world, represents one of the greatest challenges facing the generators and users of human genomic data. Perhaps the most important instrument with worldwide implications for data sharing will be the European General Data Protection Regulation (GDPR), which came into effect in 2018 (42). Data privacy and security, sharing, and access across Europe will be subject to the GDPR, but it will affect data sharing beyond Europe as well.

The GDPR replaces the EU Data Protection Directive of 1995. It will be directly applicable to all European Union member states, and will override national data protection legislation in Europe. It affects all personal data, including genetic and health data. It applies only to personal data, not to anonymized or anonymous (that is, nonpersonal) data. Anonymized or anonymous data are defined as "information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable" (42).[3] Anonymization means that the individual can no longer withdraw or receive results or be recontacted, as no further communication is possible. The GDPR promotes pseudonymization—that is, the coding of samples and data.

Another interesting feature of the GDPR is the possibility of developing European codes of conduct or certification (42, art. 40). Codes of conduct approved by the European Data Protection Board can establish legally binding standards based on expert knowledge and accountability in a

[3] As technology and methods have improved and the quantity of genomic and related genealogical and phenotypic information that has entered the commons has dramatically increased, many have become convinced that "the end of genomic privacy has arrived" and that genetic data cannot definitively be deidentified (35).

specific field, such as genomic databases, thus creating practical, self-regulatory benchmarks to demonstrate compliance with the GDPR.

In the United States, personal health records are subject to the Privacy Rule of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) (Pub. L. No. 104-191, 110 Stat. 1936, codified as amended in scattered sections of 18, 26, 29, and 42 U.S.C.), and research on human subjects, including research on human genomic data, is subject to the so-called Common Rule (45 C.F.R. 46, Part A). Both of these regulatory regimes are complex, with sometimes overlapping and inconsistent application (43), and interact with state privacy laws in a manner that has been unpredictable and, some have argued, inadequate to address the practicalities and requirements of large-scale genomic research (19).

**Data access committees.** A key development in the organizational framework for access to and use of genomic data arose from biobanking infrastructures in the form of institutional DACs. DACs are chartered to review requests for data and samples from outside researchers. DACs were established by the NIH, for example, to oversee access to controlled data from dbGaP (108), as well as within large international consortia such as the Wellcome Trust Case Control Consortium (130) and the International Cancer Genome Consortium in 2008 (**http://icgc.org**). The DAC structure of the International Cancer Genome Consortium served as a model for the policies and procedures of others that followed [e.g., the International Rare Diseases Research Consortium (**http://www.irdirc.org**), the International Human Epigenome Consortium (**http://ihec-epigenomes.org**), and the Human Heredity and Health in Africa Initiative (**http://h3africa.org**)]. DACs have also been employed by private-sector genomic data consortia, such as iSAEC (22).

What characterizes DACs is their organizational independence from the consortium or biobank/database and the inclusion of individuals who have information technology and other relevant expertise. Like multidisciplinary institutional review boards and ethics review boards that approve research protocols, DACs serve as gatekeepers for secured and approved data access. Applicants must demonstrate plans for data confidentiality and security and show how the proposed research will meet the objectives of the biobank/database. If return-of-results options are part of the proposed protocol, then these plans are also scrutinized by the DACs (116).

**The cloud.** There is increasing recognition among policy makers of the need to create open clouds for data-sharing purposes, as evidenced by the first summit of the European Open Science Cloud in June 2017 (40). Also in June 2017, the NIH reiterated its commitment to achieving a "comprehensive vision for an interoperable, FAIR (Findable, Accessible, Interoperable and Reusable) compliant, multi-cloud NIH Data Commons founded on open source and open standards" (98, p. 2).

The gradual availability of cloud computing to handle big data, particularly health and genetic data, has brought with it new regulatory challenges (118). The 2014 GA4GH framework (50, 71) led the way in what was to be a series of proposed normative data use and protection guidance. However, it is the GDPR's entry into legal force in 2018 that will undoubtedly have the greatest impact on the future of the genomic commons and the use of the cloud for storage and computing.

## Consent

As noted above, most legal and ethical systems today require that an individual's informed consent be obtained before research is conducted on that individual. This requirement has also been applied to research using data obtained from an individual. In the case of research concerning genomic

data, some regulations, such as the HIPAA Privacy Rule and the Common Rule in the United States, allow some forms of research to be conducted without the necessity to obtain consent when data have been adequately deidentified.[4]

In 2016, the Council for International Organizations of Medical Sciences and the World Health Organization issued their long-awaited International Ethical Guidelines for Health-Related Research Involving Humans (34). These guidelines endorsed not only broad consent in such research but also an informed opt-out procedure for data collected in the context of routine clinical care (34, guideline 12). The same position holds for the World Medical Association's 2016 Declaration of Taipei, which also affirms that "[h]ealth research represents a common good" (132, art. 5).

The most contentious aspect of population biobanking was the use of a broad consent for the future unspecified use of data and samples for research, subject to ongoing ethics approval and governance (70). Since the agreement to participate in research includes long-term follow-up and recontact, it would seem that such broad consent is fit for the purpose and so should be considered legally material and specific to the very nature of such resources. Furthermore, today's data-intensive science reveals a normative shift from a sole emphasis on individual consent and privacy concerns to the inclusion of data management and governance with accompanying privacy and security mechanisms (72).

The GDPR also holds that consent can be provided for "certain areas of scientific research when in keeping with recognised ethical standards for scientific research" (42), thereby permitting broad consent. Broad consent for data-based genomic research is also prescribed under the 2014 NIH Genomic Data Sharing policy (18, 96).

The shift in the consent-privacy rationale and shifting parameters of protection are visible in the effort to bridge the genomic commons with the to-be-created medical information commons (26, 28, 36, 101). The aim is to achieve personalized medicine, where interoperability between the research and medical care sectors is key. Common consent templates would facilitate this and accelerate access (82).

It is now safe to say that, in spite of a decade of debate on the legality of broad consent, internationally it is an accepted approach for biobanks and associated databases (113) as well as for data-intensive science generally. It remains to be seen whether it will be used in the context of emerging learning health systems using banked residual samples and electronic health record data in the context of medical care. In fact, a notification system with an opt-out might be more appropriate in the clinical setting.

## Proprietary Rights

Intellectual property protection has figured prominently in the ongoing debate over access to scientific data, particularly the genomic commons. Intellectual property can have two principal effects on an information commons: It can prevent the entry of data into the commons, and it can limit the ability of users to utilize data that are already in the commons. Below, we discuss the principal forms of intellectual property that impact the genomic commons.

**Trade secrets.** The effect of trade secret protection is the most straightforward to analyze. Scientific work that is sponsored by industry is often subject to written confidentiality obligations or other contractual restrictions that explicitly prevent researchers from sharing the resulting data

[4]But see footnote 3 above, questioning the possibility of true deidentification of genomic data.

and materials with others and, in some cases, delay or even prohibit the publication of their results (15, 89). With such restrictions in place, data cannot, practically speaking, enter the commons.

**Copyright.** Copyright law, unlike trade secret law, does not have as its end the concealment of information. Rather, copyright law grants the owner of a copyrighted work certain exclusive rights to exploit that work, including the exclusive rights to display, distribute, and reproduce it. In the context of an information commons, copyright principles are typically raised when discussing limitations on access to scientific data that have already been published. That is, even though scientific facts and conclusions are not themselves copyrightable, the articles in which they are presented (including the text and any illustrations) are subject to copyright and thus controlled by the publishers of the journals carrying those articles. In some cases, even data that might otherwise be in the public domain (such as mapping and geographic data developed under a contract with the US government) may be stored in proprietary databases that are accessible only by paid subscribers (89, 110, 111). In several areas, the privatization of governmental data is proceeding rapidly, leading to fears that increasing amounts of data will become "enclosed" (evoking the historical fencing-off of commonly held land) and thereby unavailable for public use (9).

**Database and data protection.** Databases that contain genomic data may be protected under various legal regimes. In European Union member states and other countries, databases with commercial value have legally protected status (41). While legal protection for databases does not exist per se in the United States, access to data that are contained in electronic databases can be controlled by a database operator via technical means, such as password-restricted access, as well as by limitations built into contractual access agreements (110, 111). Thus, while data themselves may not be subject to legal protection, circumvention of such technical protection or contractual measures can be challenged under a number of legal theories, and is explicitly addressed in the Digital Millennium Copyright Act [Pub. L. No. 105-304 (Oct. 28, 1998), codified at 17 U.S.C. Sec. 101 et seq.]. In this way, scientific information that might otherwise be in the public domain can become encumbered when compiled in proprietary databases (110, 111). Such restrictions were adopted by Celera Genomics when it announced its intention to sequence the human genome in competition with the publicly funded HGP and offer the resulting data to commercial users pursuant to license agreements. A similar approach has been used by Myriad Genetics, which maintains a proprietary database that includes tens of thousands of *BRCA1/2* variants bearing on cancer susceptibility but is subject to usage restrictions and trade secret protection implemented through a click-through agreement on the portal (53). The threat of propertization of the genome in this manner has fueled continuing public support for publicly accessible genomic data resources.

**Patents.** Perhaps the most hotly debated intellectual property issue surrounding the genomic commons is the extent to which patents may hinder use of the information contained in the commons. There are two aspects to this debate. The first is largely exogenous to the commons itself: To what extent can genomic data, or the uses thereof, be patented? The second implicates the governance of stakeholders within the commons: How should patent holders behave with respect to the enforcement and licensing of their patents covering aspects of the commons?

With respect to the first question, patents on molecular biology discoveries were first issued in the United States beginning in the 1970s (117), but it was not until the early 1990s that patents were first sought on human DNA. In 1991, J. Craig Venter, then a researcher at the NIH, filed patent applications claiming 337 short cDNA sequences known as ESTs. These filings resulted in a public outcry, which eventually led the NIH to reverse its position on gene patenting

(4, 25, 117). The NIH's position, which now reflects that of the US Patent and Trademark Office, is that raw genomic sequence data, without known function, fail to meet the utility requirement under patent law and are thus not patentable (126).

Another blow to the patenting of genetic material came via two US Supreme Court decisions in 2012 and 2013. In *Mayo Collaborative Services v. Prometheus Laboratories, Inc.* (84), the court held that medical treatment decisions informed by human physiological reactions are mere "mental steps" that are ineligible for patent protection; in *Association for Molecular Pathology v. Myriad Genetics* (6), the court held that genomic DNA, even when purified and isolated, is a "product of nature" that is ineligible for patent protection, but at the same time confirmed that human DNA constructs such as cDNA are patent eligible.[5] These cases have had a significant effect on the issuance of new patents on genomic discoveries, though the law continues to develop in this area, and the private sector has undertaken substantial lobbying efforts to reverse these decisions, both in the United States and elsewhere (26, 117).

Even with changes in the law that currently disfavor patenting of genetic discoveries, concern has persisted over the patenting of DNA-based discoveries. In 1998, Heller & Eisenberg (55) postulated that a surfeit of broadly held patents covering different genes and other biological systems could impede biomedical research, thus creating a patent thicket or "anticommons." Adding to the sense of alarm were reports from the mid-2000s that 20% or more of the human genome was already covered by patents (64). Other scholars echoed these concerns (5), even in the wake of patent-limiting decisions such as *Mayo* and *Myriad* (74, 117, 127). The potential problem for researchers is exacerbated by the lack of an effective research exemption under the laws of many countries, including the United States, which might allow academic research to continue notwithstanding the ability of patent holders to prevent commercial exploitation of patented technologies (78).

In response to these concerns, several of the early funders of the HGP took direct action to lessen the impact of patents on this shared resource. Notably, the rapid data release requirements of the Bermuda Principles were intended to limit the ability of researchers to patent sequence data generated by the HGP. That is, requiring data generators to release the data immediately upon their creation limited the researchers' ability to seek patents on inventions arising from those data and also made those data available as prior art to defeat third-party patent filings (15). This approach, though praised by many, was also criticized by those who believed that the NIH's adoption of this antipatenting approach contravened the requirements of the Bayh-Dole Act of 1980 (35 U.S.C. § 200–12), which expressly allows the patenting of federally funded inventions for the benefit of the US economy. In response to this criticism, the NIH's 1996 policy adopting the Bermuda Principles gives a nod to the Bayh-Dole Act, acknowledging that recipients of NIH funding have the right to patent inventions that "reveal convincing evidence for utility" (91). But in the same breath, the NIH cautions that it "will monitor grantee activity in this area to learn whether or not attempts are being made to patent large blocks of primary human genomic DNA sequence" (91). The consequences of doing so, however, are left unstated. The NIH's approach has thus been one of norm setting rather than legally enforceable regulation, an approach that it has repeated throughout the evolution of the genomic commons.

Nevertheless, over the past decade, the NIH's earlier aversion to patents seems to have softened. In its 2007 GWAS policy, the NIH merely expressed a "hope" that "genotype-phenotype associations identified through NIH-supported and NIH-maintained GWAS datasets and their

[5] There is some indication that lower courts have applied *Myriad* more broadly than the Supreme Court intended, in order to invalidate even patents covering human-made DNA constructs (3, 119).

obvious implications will remain available to all investigators, unencumbered by intellectual property claims" and stated that "[t]he filing of patent applications and/or the enforcement of resultant patents in a manner that might restrict use of NIH-supported genotype-phenotype data could diminish the potential public benefit they could provide" (94). However, in an effort to show some support for patent seekers, the policy also "encourages patenting of technology suitable for subsequent private investment that may lead to the development of products that address public needs" (94). Finally, in its 2014 Genomic Data Sharing policy (96), the NIH appears largely to have discarded the patent-deterring early data release mandate that it championed in the Bermuda Principles, permitting data generators to sequester human data for periods of up to six months and nonhuman data until the time of publication (18). The agency, perhaps too optimistically, cites the *Myriad* decision as sufficient protection against DNA sequences becoming encumbered by patents (18, 96).

US funding agencies were not the only ones to address patenting activity with respect to genomic research. Genome Canada adopted its first formal data release policy in 2005 (48). While acknowledging the Fort Lauderdale principles, the Canadian policy does not adopt the 24-hour release requirement of the Bermuda Principles. With respect to patents, Genome Canada "recognizes the need to protect patentable and other proprietary data" (48, p. 1) and thus requires that data generators release data following publication or the filing of a patent application, whichever occurs first.

Since the beginning of the HGP, the UK-based Wellcome Trust has supported genomic research, both through grants and through its Sanger Institute in Cambridge, England, a leading sequencing center. In 2006, the Wellcome Trust funded a large-scale GWAS of seven complex human diseases by more than 50 research groups across the United Kingdom (the Wellcome Trust Case Control Consortium) (130). The study generated a large quantity of data, including aggregated and individual-level genotypic and phenotypic information. Most of these data were released to the public in accordance with the Fort Lauderdale principles, and the project designated itself as a CRP. The consortium required prospective data users to apply to the consortium's DAC and sign a written data access agreement. Access to data was granted only to qualified investigators for "appropriate use" (§4 in the agreement). The data access agreement requires security, acknowledgment, transfer, and use restrictions comparable to those found in the GAIN policy and other contemporary policies, as well as restrictions that are specific to the study samples. The agreement does not, however, contain any specific embargo on publication or any restriction on patenting activity.

In 2002, at the tail end of the HGP, an international group of researchers and funders undertook the development of a haplotype map of the human genome through the International HapMap Project (60). This project's data release policy was based on the Fort Lauderdale principles, and the project designated itself as a CRP. The project also took several affirmative steps to ensure that patents would not be filed on the results of its research by either data generators or data users. Specifically, each user of HapMap data (including data generators) was expressly prohibited from restricting access to the HapMap database and, in particular, from filing patent applications on the data generated by the project. The project also took the legal position that raw SNP and haplotype data lack specific utility and are therefore ineligible for patent protection (60). These policy provisions were viewed with admiration by many, including policy makers at the NIH, who lauded the International HapMap Project's success at deterring "parasitic patents" (92).

Similar patent-deterrence strategies have been adopted by industry-led consortia, such as the SNP Consortium and iSAEC. Each of these organizations implemented a protective patenting strategy whereby the findings of the research program were submitted to the US Patent and

Trademark Office as patent applications, and the applications were eventually allowed to lapse prior to patent issuance. This approach inserted the resulting research findings into the US patent database at an early date, thereby defeating any ability to patent by individual research institutions and acting as prior art against potential third-party patents claiming the same discoveries (15, 22, 57). In these cases, note that the private-sector research funders preferred to release data free from patent encumbrances not out of a public interest in the free availability of data, but in order to avoid capture of valuable research tools by other private-sector firms—in other words, they preferred that the results be free to all rather than owned by someone else.

## THEMES AND CONCLUSIONS

Over its 30 or so years of existence, the genomic commons has enjoyed remarkable success. And thanks to the rapid public data release program initiated by the Bermuda Principles and subsequent norm-setting policies, free access to a vast array of scientific data is now the norm, not only in genomics but in scientific disciplines of many descriptions. And far from being a monolithic creation of bureaucratic fiat, the genomic commons is an exemplar of polycentric, multistakeholder governance.

But like all dynamic and rapidly evolving systems, the genomic commons faces substantial challenges. As discussed in this review, issues involving scientific priority, intellectual property, individual privacy, and informed consent, in an environment of data sets of exponentially expanding size and complexity, must be addressed in the near term. Below, we summarize some of the trends, challenges, and opportunities that we envision for the genomic commons in the years to come.

### Genomic Data to Health Data

Genomics is no longer a scientific island. Genomic data today are being used in nearly every branch of biomedical research. As different data types become increasingly integrated, the genomic commons will likely merge with or become linked to more comprehensive collections of biomedical, sociological, environmental, and other data, as well as with repositories of physical samples and materials. The world is moving toward what has been termed the medical information commons (26, 28, 36, 101), of which genomic data will be only a small, albeit significant, component. Yet researchers in fields beyond genomics, who did not grow up with the data-sharing norms instilled by the Bermuda Principles, may not be as willing to surrender their data for the public good (39). It remains to be seen how the liberal data access and usage policies that have evolved in the sphere of genomic data will be adapted, or subsumed, by the policies governing these larger resources.

### The Need for Interoperability

Given their origins in the multisite HGP, the data contained in the genomic commons are largely stored and accessed using a set of common data formats that are observed around the world (16). This degree of standardization has facilitated the use and growth of the genomic commons. Unfortunately, other forms of biological, medical, and health data are not nearly so uniform, and data incompatibility is a serious impediment to data sharing in areas outside of genomics (29). As genomic data become increasingly integrated into broader health-care data sets, the need for consistent and uniformly applied data standards will only increase.

In addition to data interoperability, transborder data sharing is impeded by a wide range of inconsistent and often incompatible legal regimes governing the use and transfer of human health

information. To facilitate the sharing of health data and international scientific collaboration on pressing health issues, planners will need to ensure that data-sharing plans and protocols comply with applicable national data privacy, security, and consent regulations (23, 113, 119).

## Science and Human Rights

As discussed above, the animating principle behind many recent international governance frameworks for genomic data sharing, including the GA4GH framework, is grounded in the human right to benefit from the fruits of scientific research. This emphasis, of course, represents a significant departure from the technocratic orientation of the HGP planners, who promoted broad data sharing in the service of scientific efficiency and advancement. Yet it appears that this human rights orientation is gaining significant purchase, particularly in countries with universal healthcare systems (87). As modern societies become increasingly heterogeneous, both research and medical data will have to represent "a nation of nations," thereby requiring more, and not less, international data sharing (80).

## Proportionality of Sharing Versus Privacy

As discussed above, concerns about data privacy have driven many of the governance and design features of elements in the genomic commons that contain data derived from individuals. Yet an excessive emphasis on the protection of individual privacy, in a world where complete anonymity might be impossible, could weigh the balance too far in one direction, precluding collective benefits from less restricted uses of data. There is thus a growing sense that privacy concerns should not always outweigh the legitimate and socially beneficial sharing of scientific data (66). As Wright et al. (133) have argued, genomic data sharing should be guided by a principle of proportionality, in which what is shared is weighed against the scope of the sharing rather than hypothetical what-if scenarios. In a similar vein, the second UK National Health Service Caldicott report recognizes that "[t]he duty to share information can be as important as the duty to protect patient confidentiality" (11, p. 119). This position was affirmed in the 2016 UK National Health Service annual report, which emphasized the need for a new social contract based on genomic solidarity (122, chap. 16).

## Diminishing Concerns over Intellectual Property

From the early days of the HGP, NIH policy makers and scientific leaders expressed strong aversion to the encumbrance of genomic information with patents (as evidenced by the EST patenting debate) and database access restrictions (exemplified by the HGP's open science approach, as contrasted with the proprietary approach taken, for example, by Celera Genomics). Through a combination of regulatory and judicial decisions, coupled with normalized institutional and funder policies discouraging the encumbrance of genomic data sets with intellectual property protection, intellectual property—or at least patenting of genes—appears to have become less of a concern among policy makers and researchers. This being said, the scientific community should not become complacent regarding intellectual property and its potential to limit access to and use of common resources. While the pendulum of intellectual property protection may have swung toward a more relaxed and sharing-friendly position, there is no guarantee that this will continue indefinitely. Thus, planners should give attention to intellectual property details in new data-sharing policies, particularly as the genomic commons migrates and expands into a more comprehensive medical research commons.

## DISCLOSURE STATEMENT

J.L.C. is a member of the NIH Council of Councils and served as legal counsel to the SNP Consortium and International Serious Adverse Events Consortium.

## ACKNOWLEDGMENTS

J.L.C. gratefully acknowledges support from the Huntsman Cancer Institute/Huntsman Cancer Foundation and the Albert and Elaine Borchard Fund for Faculty Excellence at the University of Utah. B.M.K. gratefully acknowledges support from the Canada Research Chair in Law and Medicine and the P$^3$G Can-SHARE project (jointly funded by Genome Canada and the Canadian Institutes for Health Research). The authors are grateful to Robert Cook-Deegan for his helpful comments and suggestions throughout this article and for his foundational work in the field.

## LITERATURE CITED

1. 1000 Genomes Proj. Consort. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65
2. Abbasi J. 2017. An international Human Cell Atlas consortium takes shape. *J. Am. Med. Assoc.* 318:685–86
3. Aboy M, Liddell K, Liddicoat J, Crespo C. 2016. *Myriad*'s impact on gene patents. *Nat. Biotechnol.* 34:1119–23
4. Anderson C. 1994. NIH drops bid for gene patents. *Science* 263:909–10
5. Andrews LB. 2002. Genes and patent policy: rethinking intellectual property rights. *Nat. Rev. Genet.* 3:802–8
6. *Assoc. Mol. Pathol. v. Myriad Genet*, 133 S. Ct. 2107 (2013)
7. Barnett J. 2000. Cultivating the genetic commons: imperfect patent protection and the network model of innovation. *San Diego Law Rev*. 37:987–1067
8. Benkler Y. 2014. Between Spanish huertas and the open road: a tale of two commons? See Ref. 44, pp. 69–98
9. Boyle J. 2008. *The Public Domain: Enclosing the Commons of the Mind*. New Haven, CT: Yale Univ. Press
10. Burton PR, Fortier I, Knoppers BM. 2009. The global emergence of epidemiological biobanks: opportunities and challenges. In *Human Genome Epidemiology: Building the Evidence for Using Genetic Information to Improve Health and Prevent Disease*, ed. M Khoury, S Bedrosian, M Gwinn, J Higgins, J Ioannidis, J Little, pp. 77–99. Oxford, UK: Oxford Univ. Press. 2nd ed.
11. Caldicott F. 2013. *Information: To Share or Not to Share? The Information Governance Review*. London: UK Dep. Health
12. Collins F. 2010. Has the revolution arrived? *Nature* 464:674–75
13. Conley JM, Doerr AK, Vorhaus DB. 2010. Enabling responsible public genomics. *Health Matrix* 20:325–85
14. Contreras JL. 2010. Prepublication data release, latency and genome commons. *Science* 329:393–94
15. Contreras JL. 2011. Bermuda's legacy: policy, patents, and the design of the genome commons. *Minn. J. Law Sci. Technol.* 12:61–125
16. Contreras JL. 2013. Technical standards and bioinformatics. In *Bioinformatics Law: Legal Issues for Computational Biology in the Post-Genome Era*, ed. JL Contreras, AJ Cuticchia, pp. 113–36. Chicago: ABA
17. Contreras JL. 2014. Constructing the genome commons. See Ref. 44, pp. 99–136
18. Contreras JL. 2015. NIH's genomic data sharing policy: timing and tradeoffs. *Trends Genet*. 31:55–57
19. Contreras JL. 2016. Genetic property. *Georgetown Law J.* 105:1–54
20. Contreras JL. 2017. Leviathan in the commons: biomedical data and the state. In *Governing Medical Knowledge Commons*, ed. KJ Strandburg, BM Frischmann, MJ Madison, pp. 19–45. Cambridge, UK: Cambridge Univ. Press
21. Contreras JL, Deshmukh VG. 2017. Development of the personal genomics industry. In *Genetics, Ethics and Education*, ed. S Bouregy, EL Grigorenko, SR Latham, M Tan, pp. 284–308. Cambridge, UK: Cambridge Univ. Press

22. Contreras JL, Floratos A, Holden AH. 2013. The International Serious Adverse Events Consortium's data sharing model. *Nat. Biotechnol.* 31:17–19
23. Contreras JL, Reichman JH. 2015. Sharing by design: data and decentralized commons: overcoming legal and policy obstacles. *Science* 350:1312–14
24. Contreras JL, Vertinsky L. 2016. Pre-competition. *N.C. Law Rev.* 95:67–131
25. Cook-Deegan R. 1994. *The Gene Wars: Science, Politics, and the Human Genome*. New York: Norton
26. Cook-Deegan R, Ankeny RA, Maxson Jones K. 2017. Sharing data to build a medical information commons: from Bermuda to the Global Alliance. *Annu. Rev. Genom. Hum. Genet.* 18:389–415
27. Cook-Deegan R, Dedeurwaerdere T. 2006. The science commons in life science research: structure, function, and value of access to genetic diversity. *Int. Soc. Sci. J.* 58:299–317
28. Cook-Deegan R, McGuire AL. 2017. Moving beyond Bermuda: sharing data to build a medical information commons. *Genome Res.* 27:897–901
29. Counc. Can. Acad. 2015. *Accessing health and health-related data in Canada: the expert panel on timely access to health and social data for health research and health system innovation*. Rep., Counc. Can. Acad., Ottawa, Can.
30. Counc. Eur. 1997. *Convention for the Protection of Human Rights and Dignity of the Human Being with Regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine*. Eur. Treaty Ser. 164, Counc. Eur., Strasbourg, Fr. **https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/090000168007cf98**
31. Counc. Eur. 2016. *Recommendation CM/Rec(2016)6 of the Committee of Ministers to member States on research on biological materials of human origin*. Recomm., Counc. Eur., Strasbourg, Fr. **https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=090000168064e8ff**
32. Counc. Eur. 2016. *Recommendation CM/Rec(2016)8 of the Committee of Ministers to the member States on the processing of personal health-related data for insurance purposes, including data resulting from genetic tests*. Recomm., Counc. Eur., Strasbourg, Fr. **https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016806b2c5f**
33. Counc. Eur. 2017. *Chart of signatures and ratifications of Treaty 164*. Eur. Treaty Ser. 164, Counc. Eur., Strasbourg, Fr. **https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/164/signatures**
34. Counc. Int. Organ. Med. Sci. (CIOMS), World Health Organ. (WHO). 2016. *International Ethical Guidelines for Health-Related Research Involving Humans*. Geneva: CIOMS
35. Couzin-Frankel J. 2015. Trust me, I'm a medical researcher. *Science* 347:501–2
36. Deverka PA, Majumder MA, Villanueva AG, Anderson M, Bakker AC, et al. 2017. Creating a data resource: What will it take to build a medical information commons? *Genome Med.* 9:84
37. Dove ES, Townend D, Meslin EM, Bobrow M, Littler K, et al. 2016. Ethics review for international data-intensive research. *Science* 351:1399–400
38. Dyke SOM, Kirby E, Shabani M, Thorogood A, Kato K, Knoppers BM. 2016. Registered access: a 'Triple-A' approach. *Eur. J. Hum. Genet.* 24:1676–80
39. Editorial. 2017. Empty rhetoric over data sharing slows science. *Nature* 546:327
40. Eur. Comm. 2017. *European Open Science Cloud Summit*. **https://ec.europa.eu/research/index.cfm?eventcode=44D86060-FBA1-1BD1-9355822B162BB0EE&pg=events**
41. Eur. Union. 1996. *Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases*. Dir. 96/9/EC, Eur. Union
42. Eur. Union. 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. Regul. 2016/679, Eur. Union. **http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679**
43. Evans BJ. 2012. Much ado about data ownership. *Harv. J. Law Technol.* 25:69–130
44. Frischmann BM, Madison MJ, Strandburg KJ, eds. 2014. *Governing Knowledge Commons*. New York: Oxford Univ. Press
45. GAIN Collab. Res. Group. 2007. New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat. Genet.* 39:1045–51

46. Genet. Assoc. Inf. Netw. (GAIN). 2010. *Data use certification agreement*. **https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?view_pdf&stacc=phs000021.v1.p1**
47. Genom. Engl. 2017. *The 100,000 Genomes Project*. **https://www.genomicsengland.co.uk/the-100000-genomes-project**
48. Genome Can. 2008. *Data release and resource sharing*. **http://www.genomecanada.ca/medias/PDF/EN/DataReleaseandResourceSharingPolicy.pdf**
49. Glob. Alliance Genom. Health (GA4GH). 2016. A federated ecosystem for sharing genomic, clinical data. *Science* 352:1278–80
50. Glob. Alliance Genom. Health (GA4GH). 2017. *About us*. **https://www.ga4gh.org/aboutus**
51. Glob. Alliance Genom. Health (GA4GH). 2017. *GA4GH strikes formal collaborations with 15 international genomic data initiatives*. **https://www.ga4gh.org/news/sAhZCeJjS96QHhVPIYwwWA.article**
52. Greely HT. 2007. The uneasy ethical and legal underpinnings of large-scale genomic biobanks. *Annu. Rev. Genom. Hum. Genet.* 8:343–64
53. Guerrini CJ, McGuire AL, Majumder MA. 2017. Myriad take two: Can genomic databases remain secret? *Science* 356:586–87
54. Haga SB, O'Daniel J. 2011. Public perspectives regarding data-sharing practices in genomics research. *Public Health Genom.* 14:319–24
55. Heller MA, Eisenberg RS. 1998. Can patents deter innovation? The anticommons in biomedical research. *Science* 280:698–701
56. Hill JA, Smith BE, Papoulias PG, Andrews PC. 2010. ProteomeCommons.org collaborative annotation and project management resources integrated with the tranche repository. *J. Proteome Res.* 9:2809–11
57. Holden A. 2002. The SNP Consortium: summary of a private consortium effort to develop an applied map of the human genome. *BioTechniques* 32:22–26
58. Hum. Genome Organ. (HUGO). 1996. *Statement on the principled conduct of genetics research*. Rep., HUGO, Geneva, Switz.
59. Hum. Genome Organ. (HUGO). 2003. Statement on human genomic databases, December 2002. *J. Int. Bioeth.* 14:207–10
60. Int. HapMap Consort. 2003. The International HapMap Project. *Nature* 426:789–96
61. Int. Hum. Genome Seq. Consort. 1996. *Summary of principles agreed upon at the first international strategy meeting on human genome sequencing (Bermuda, 25–28 February 1996) as reported by HUGO*. **http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml**
62. Int. Hum. Genome Seq. Consort. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
63. Jasny BR. 2013. Realities of data sharing using the genome wars as case study - an historical perspective and commentary. *EJP Data Sci.* 2:1
64. Jensen K, Murray F. 2005. Intellectual property landscape of the human genome. *Science* 310:239–40
65. Joint Genome Inst. 2017. *Genomes OnLine Database*. Retrieved Oct. 17. **https://gold.jgi.doe.gov**
66. Jones KH, Laurie G, Stevens L, Dobbs C, Ford DV, Lea N. 2017. The other side of the coin: harm due to the non-use of health-related data. *Int. J. Med. Inform.* 97:43–51
67. Karczewski KJ, Francioli L. 2017. The genome Aggregation Database (gnomAD). *MacArthur Lab*, Feb. 27. **https://macarthurlab.org/2017/02/27/the-genome-aggregation-database-gnomad**
68. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, et al. 2017. The ExAC browser: displaying reference data information from over 60,000 exomes. *Nucleic Acids Res.* 45:D840–45
69. Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P. 2009. Data sharing in genomics—re-shaping scientific practice. *Nat. Rev. Genet.* 10:331–35
70. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. 2014. Dynamic consent: a patient interface for twenty-first century research networks. *Eur. J. Hum. Genet.* 23:141–46
71. Knoppers BM. 2014. Framework for responsible sharing of genomic and health-related data. *HUGO J.* 8:3
72. Knoppers BM, Thorogood A. 2017. Ethics and big data in health. *Curr. Opin. Syst. Biol.* 4:53–57
73. Knoppers BM, Zawati MH, Kirby ES. 2012. Sampling populations of humans across the world: ELSI issues. *Annu. Rev. Genom. Hum. Genet.* 13:395–413

74. Laakmann AB. 2015. The new genomic semicommons. *UC Irvine Law Rev*. 5:1001–39
75. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, et al. 2016. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 44:D862–68
76. Lee P. 2009. Toward a distributive commons in patent law. *Wisc. Law Rev*. 2009:917–1016
77. Lee P. 2017. Centralization, fragmentation, and replication in the genomic data commons. In *Governing Medical Knowledge Commons*, ed. KJ Strandburg, BM Frischmann, MJ Madison, pp. 46–73. Cambridge, UK: Cambridge Univ. Press
78. *Madey v. Duke Univ*., 307 F.3d 1351 (Fed. Cir.) (2002)
79. Madison MJ, Frischmann BM, Strandburg KJ. 2010. Constructing commons in the cultural environment. *Cornell Law Rev*. 95:657–709
80. Majumder MA, Cook-Deegan R, McGuire AL. 2016. Beyond our borders? Public resistance to global genomic data sharing. *PLOS Biol*. 14:e2000206
81. Mandl KD, Bourgeois FT. 2017. The evolution of patient diagnosis: from art to digital data-driven science. *J. Am. Med. Assoc.* 318:1859–60
82. Manolio TA, Fowler DM, Starita LM, Haendel MA, MacArthur DG, et al. 2017. Bedside back to bench: building bridges between basic and clinical genomic research. *Cell* 169:6–12
83. 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
84. *Mayo Collab. Serv. v. Prometheus Lab., Inc*., 132 S. Ct. 1289 (2012)
85. Merck. 1995. *First installment of Merck Gene Index data released to public databases: cooperative effort promises to speed scientific understanding of the human genome*. Press Release, Feb. 10, Merck, Whitehouse Station, NJ. **http://www.bio.net/bionet/mm/bionews/1995-February/001794.html**
86. Merges RP. 1996. Property rights theory and the commons: the case of scientific research. In *Scientific Innovation, Philosophy, and Public Policy*, ed. EF Paul, FD Miller Jr., J Paul, pp. 145–167. Cambridge, UK: Cambridge Univ. Press
87. Meslin EM, Ibrahim Garba I. 2011. Biobanking and public health: is a human rights approach the tie that binds? *Hum. Genet.* 130:451–63
88. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Verezemska O, et al. 2017. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res*. 45:D446–56
89. Natl. Acad. Sci. (NAS). 2009. *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. Washington, DC: Natl. Acad. Press
90. Natl. Cancer Inst. (NCI). 2017. *The next generation cancer knowledge network*. **https://gdc.cancer.gov**
91. Natl. Hum. Genome Res. Inst. (NHGRI). 1996. *NHGRI policy regarding intellectual property of human genomic sequence*. **http://www.genome.gov/10000926**
92. Natl. Hum. Genome Res. Inst. (NHGRI). 2003. *ENCODE Project data release policy (2003–2007)*. **http://www.genome.gov/12513440**
93. Natl. Hum. Genome Res. Inst. (NHGRI). 2010. *The Human Genome Project completion: frequently asked questions*. **http://www.genome.gov/11006943**
94. Natl. Inst. Health (NIH). 2007. *Policy for sharing of data obtained in NIH supported or conducted genome-wide association studies (GWAS)*. Not. NOT-OD-07-088, NIH, Bethesda, MD. **https://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html**
95. Natl. Inst. Health (NIH). 2012. *1000 Genomes Project data available on Amazon Cloud*. News Release, Mar. 29, NIH, Bethesda, MD. **http://www.nih.gov/news/health/mar2012/nhgri-29.htm**
96. Natl. Inst. Health (NIH). 2014. *Final NIH Genomic Data Sharing policy*. 79 Fed. Reg. 51345–54
97. Natl. Inst. Health (NIH). 2017. *About the* All of Us *Research Program*. **https://allofus.nih.gov/about/about-all-us-research-program**
98. Natl. Inst. Health (NIH). 2017. *NIH data commons pilot phase*. Fund. Announc. RM-17-026, NIH, Bethesda, MD. **https://commonfund.nih.gov/sites/default/files/RM-17-026_CommonsPilotPhase.pdf**
99. Natl. Res. Counc. (NRC). 1988. *Mapping and Sequencing the Human Genome*. Washington, DC: Natl. Acad. Press

100. Natl. Res. Counc. (NRC). 2006. *Reaping the Benefits of Genomic and Proteomic Research*. Washington, DC: Natl. Acad. Press
101. Natl. Res. Counc. (NRC). 2011. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington, DC: Natl. Acad. Press
102. Organ. Econ. Coop. Dev. (OECD). 2009. *OECD guidelines on human biobanks and genetic research databases*. Rep., OECD, Paris
103. Ossorio PN. 2011. Bodies of data: genomic data and bioscience data sharing. *Soc. Res.* 78:907–32
104. Ostrom E. 1990. *Governing the Commons*. Cambridge, UK: Cambridge Univ. Press
105. Ostrom E. 2005. *Understanding Institutional Diversity*. Princeton, NJ: Princeton Univ. Press
106. Ostrom E, Hess C, eds. 2007. *Understanding Knowledge as a Commons: From Theory to Practice*. Cambridge, MA: MIT Press
107. P3G Consort., Church G, Heeney C, Hawkins N, de Vries J, et al. 2009. Public access to genome-wide data: five views on balancing research with privacy and protection. *PLOS Genet.* 5:e1000665
108. Paltoo DN, Rodriguez LL, Feolo M, Gillanders E, Ramos EM, et al. 2014. Data use under the NIH GWAS Data Sharing Policy and future directions. *Nat. Genet.* 46:934–38
109. Rai AK. 1999. Regulating scientific research: intellectual property rights and the norms of science. *Northwest. Univ. Law Rev.* 94:77–152
110. Reichman JH, Uhlir PF. 2003. A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment. *Law Contemp. Probl.* 66:315–462
111. Reichman JH, Uhlir PF, Dedeurwaerdere T. 2016. *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons*. Cambridge, UK: Cambridge Univ. Press
112. Rhodes C. 2013. *Governance of Genetic Resources: A Guide to Navigating the Complex Global Landscape*. Cheltenham, UK: Edward Elgar
113. Rothstein M, Knoppers BM. 2015. Harmonizing privacy laws to enable international biobank research. *J. Law Med. Ethics* 43:673–74
114. Schofield PN, Epping J, Huala E, Hrabe de Angelis M, Harvey M, et al. 2010. Sustaining the data and bioresource commons. *Science* 330:592–93
115. Selby JV, Beal AC, Frank L. 2012. The Patient-Centered Outcomes Research Institute (PCORI) national priorities for research and initial research agenda. *J. Am. Med. Assoc.* 307:1583–84
116. Shabani M, Dove ES, Murtagh M, Knoppers BM, Borry P. 2017. Oversight of genomic data sharing: what roles for ethics and data access committees? *Biopreserv. Biobank.* 15:469–74
117. Sherkow JS, Greely HT. 2015. The history of patenting genetic material. *Annu. Rev. Genet.* 49:161–82
118. Stein LD, Knoppers BM, Campbell P, Getz G, Korbel JO. 2015. Data analysis: creating a cloud commons. *Nature* 523:149–51
119. Tasse AM, Kirby E, Fortier I. 2016. Developing an ethical and legal interoperability assessment process for retrospective studies. *Biopreserv. Biobank.* 14:249–55
120. Terry SF, Terry PF, Rauen KA, Uitto J, Bercovitich LG. 2007. Advocacy groups as research organizations: the PXE international example. *Nat. Rev. Genet.* 8:157–64
121. Tor. Int. Data Release Workshop Authors. 2009. Prepublication data sharing. *Nature* 461:168–70
122. UK Dep. Health. 2017. *Annual report of the Chief Medical Officer 2016: Generation Genome*. Rep., UK Dep. Health, London. **https://www.gov.uk/government/publications/chief-medical-officer-annual-report-2016-generation-genome**
123. UN. 1947. *Universal Declaration of Human Rights*. Declar., UN, New York. **http://www.un.org/en/universal-declaration-human-rights**
124. UN. 1966. *International Covenant on Economic, Social and Cultural Rights*. Treaty 14531, UN, New York. **https://treaties.un.org/doc/Publication/UNTS/Volume%20993/volume-993-I-14531-English.pdf**
125. UN Sci. Educ. Cult. Organ. (UNESCO). 1997. *The Universal Declaration on the Human Genome and Human Rights*. Declar., UNESCO, Paris. **http://portal.unesco.org/en/ev.php-URL_ID=13177&URL_DO=DO_TOPIC&URL_SECTION=201.html**
126. US Pat. Trademark Off. 1999. *Revised utility examination guidelines; request for comments*. 64 Fed. Reg. 71440–42

127. Van Overwalle G. 2014. Governing genomic data: plea for an "open commons." See Ref. 44, pp. 137–54
128. Watson JD. 1997. Genes and politics. *J. Mol. Med.* 75:624–36
129. Wellcome Trust. 2003. *Sharing data from large-scale biological research projects: a system of tripartite responsibility*. Rep., Wellcome Trust, London. **http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf**
130. Wellcome Trust Case Control Consort. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–78
131. Williams HL. 2013. Intellectual property rights and innovation: evidence from the human genome. *J. Political Econ.* 121:1–27
132. World Med. Assoc. 2016. *WMA Declaration of Taipei on ethical considerations regarding health databases and biobanks*. **https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks**
133. Wright CF, Hurles ME, Firth HV. 2016. Principle of proportionality in genomic data sharing. *Nat. Rev. Genet.* 17:1–2