# The Illusion of Inclusion — The "All of Us" Research Program and Indigenous Peoples' DNA

Keolu Fox, Ph.D.

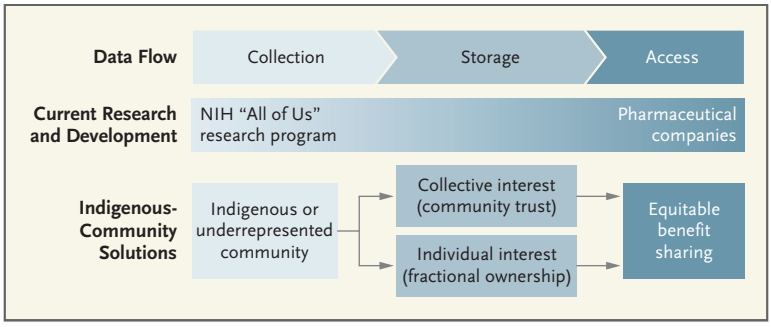
Raw data, including digital sequence information derived from human genomes, have in recent years emerged as a top global commodity. This shift is so new that experts are still evaluating what such information is worth in a global market. In 2018, the direct-to-consumer genetic-testing company 23andMe sold access to its database containing digital sequence information from approximately 5 million people to GlaxoSmithKline for $300 million. Earlier this year, 23andMe partnered with Almirall, a Spanish drug company that is using the information to develop a new antiinflammatory drug for autoimmune disorders. This move marks the first time that 23andMe has signed a deal to license a drug for development.

Eighty-eight percent of people included in large-scale studies of human genetic variation are of European ancestry, as are the majority of participants in clinical trials.[1] Corporations such as Geisinger Health System, Regeneron Pharmaceuticals, AncestryDNA, and 23andMe have already mined genomic databases for the strongest genotype–phenotype associations. For the field to advance, a new approach is needed. There are many potential ways to improve existing databases, including "deep phenotyping," which involves collecting precise measurements from blood panels, questionnaires, cognitive surveys, and other tests administered to research participants. But this approach is costly and physiologically and mentally burdensome for participants. Another approach is to expand existing biobanks by adding genetic information from populations whose genomes have not yet been sequenced — information that may offer opportunities for discovering globally rare but locally common population-specific variants, which could be useful for identifying new potential drug targets.

Many Indigenous populations have been geographically isolated for tens of thousands of years. Over time, these populations have developed adaptations to their environments that have left specific variant signatures in their genomes. As a result, the genomes of Indigenous peoples are a treasure trove of unexplored variation. Some of this variation will inevitably be identified by programs like the National Institutes of Health (NIH) "All of Us" research program. NIH leaders have committed to the idea that at least 50% of this program's participants should be members of underrepresented minority populations, including U.S. Indigenous communities (Native Americans, Alaskan Natives, and Native Hawaiians), a decision that explicitly connects diversity with the program's goal of promoting equal enjoyment of the future benefits of precision medicine.

But there are reasons to believe that this promise may be an illusion. Previous government-funded, large-scale human genome sequencing efforts, such as the Human Genome Diversity Project, the International HapMap Project, and the 1000 Genomes Project, provide examples of the ways in which open-source data have been commodified in the past. These initiatives, which promised unrestricted, open access to data on population-specific biomarkers, ultimately enabled the generation of nearly a billion dollars' worth of profits by pharmaceutical and ancestry-testing companies. If the All of Us program uses the same unrestricted data-access and sharing protocols, there will be no built-in mechanisms to protect against the commodification of Indigenous peoples' DNA.

Many communities have participated in large-scale studies of human genetic variation as part of drug-development efforts. For example, Vertex Pharmaceuticals' next-generation cystic fibrosis cocktail, Trikafta (elexacaftor/tezacaftor/ivacaftor and ivacaftor), was developed using digital sequence information from patients with cystic fibrosis and funding from the Cystic Fibrosis Foundation. Loci identified in persons of African ancestry led to the development of PCSK9 inhibitors.[2] Most recently, Regeneron scientists identified mutations in the *B4GALT1* gene that are associated with lower cholesterol levels and lower fibrinogen levels in members of




**Approaches to Equitable Benefit Sharing.**

Two systems that highlight equitable benefit sharing are collective-interest models (i.e., community trusts) and individual-interest models (i.e., fractional ownership of stock or a shareholder model).

an Amish community, a discovery that could potentially yield a new class of cholesterol drugs.[3]

Although participants in these studies may benefit from the development of new treatments, it's not clear that any of the drugs developed using insights gained from research on these populations led or will lead to direct benefits for those communities in the form of subsidized medications, royalties, or intellectual-property rights. What's more, the prices of such drugs often put them out of reach for some patients in the study populations; for example, Vertex's cystic fibrosis drug was priced at more than $300,000 per year. The Common Rule, which guides human-subjects research, makes it clear that research shouldn't be conducted on groups that won't benefit from the results, and lack of affordability of new treatments constitutes a breach of that principle.

Indigenous peoples are legitimately concerned about the potential for commodification of drugs derived from research on their genomes, and as a consequence, they are sometimes reluctant to participate in genomics research. All of Us investigators are interested in recruiting participants from native groups, but given the fraught history of genetic studies involving Indigenous peoples — including the example of *Havasupai v. Arizona State University*, in which the tribe successfully sued the university for improperly using its members' blood samples — tribal communities continue to be wary about participating in the NIH's newest endeavor.[4]

Commodification of data and policies permitting unrestricted access to them extend histories of marginalization and disempower Indigenous people from making decisions about how and under what circumstances their data can be used. If All of Us investigators want to recruit Indigenous-community members, they could include Indigenous people in the development of policies concerning data access, data use, and intellectual property. Advocates for the emerging concept of "Indigenous data sovereignty" have called for greater participation of Indigenous people in the governance of data and biologic samples and for the use of digital tools that define sample usage rights in order to increase transparency and integrity in the use of digital sequence information.[5]

One way to facilitate a paradigm shift toward equitable benefit sharing would be to ensure that Indigenous people have control of data from Indigenous populations, including digital sequence information. Two approaches for achieving this control have been used (see figure): individual-interest models (also known as shareholder models, which involve fractional ownership of stock) and collective-interest models (which involve community trusts). LunaDNA, a community-owned platform for biomedical research, is an example of the fractional-ownership model. This public-benefit corporation distributes proceeds from the platform back to people who share their DNA for research. Community trusts, which not only provide subsidized access to drugs but also reinvest in communities that participate in genomic research, can also be established in partnership with both the NIH and pharmaceutical companies.

There is precedent for community-partner–based benefit sharing in several industries. In November 2019, the South African government announced that, in accordance with the Nagoya Protocol, the San and Khoi peoples of southern Africa would share in the profits of the lucrative rooibos-tea industry. The Lucara Diamond Corporation, which operates in Botswana, recently began allocating 5% of all retail sales for community-based benefit sharing. As a result, the Botswanan government is beginning to use diamond royalties to fund infrastructure, health care, and education initiatives. In the pharmaceutical industry, startups such as Variant Bio are exploring long-term benefit-sharing models, under which royalties from drugs

developed using community-specific digital sequence information will be shared by means of a collective-interest model.

As we chart the future of predictive and preventive medicine, equity and inclusiveness should guide the trajectory of innovation in the United States. Just as the public is taking an increasingly critical look at extractive industries to ask what is environmentally sustainable, when data are the commodifiable resource at issue, we need to ask questions: What is socially and culturally sustainable? Who has access to digital sequence information? What might partnerships that rely on sharing valuable data sets look like? How should benefits be shared? And how do we develop drugs in a way that respects the contributions of various communities and encourages investment and capacity building for marginalized groups, while still providing financial incentives for drug development and commercial research and development?

The success of the All of Us program will depend on answering these questions. As new technologies converge in the field of human-driven therapeutics, opportunities for developing blockbuster drugs using information from studies of human genetic variation will increase exponentially. It will take equitable innovation in this area to ensure that the benefits truly reach "all of us."

Disclosure forms provided by the author are available at NEJM.org.

From the Department of Anthropology, the Global Health Program, and the Indigenous Futures Lab, University of California at San Diego, San Diego.

1. Mills MC, Rahal C. A scientometric review of genome-wide association studies. Commun Biol 2019;2:9.
2. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. N Engl J Med 2006;354:1264-72.
3. Montasser ME, Van Hout CV, McFarland R, et al. Genetic and functional evidence relates a missense variant in *B4GALT1* to lower LDL-C and fibrinogen. August 10, 2019 (https://www.biorxiv.org/content/10.1101/721704v1). preprint.
4. Garrison NA, Hudson M, Ballantyne LL, et al. Genomic research through an indigenous lens: understanding the expectations. Annu Rev Genomics Hum Genet 2019;20:495-517.
5. Hudson M, Garrison NA, Sterling R, et al. Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. Nat Rev Genet 2020;21:377-84.

DOI: 10.1056/NEJMp1915987