

Benefits and Risks of Sharing Genomic Information

Dhikshitha Balaji¹ and Sharon F. Terry²

GENETIC TESTS CAN result in abundant data that must be managed, interpreted, and afforded the appropriate protections (McGuire *et al.*, 2008). Because of the numerous potential uses of these data, ethical use requires engaged and representative governance and oversight. The movement to publish these data in large genomic databases to make it broadly accessible is generally widely supported. It is critical to weigh the benefits and the possible dangers of sharing personal data on large genomic databases.

Inappropriate use of genomic data poses very specific risks since it can be used to identify an individual. Because each person can be identified by the variations in their genome, even databases of deidentified data can be used, in combination with other databases, to reidentify individuals (Erich and Narayanan, 2014; Bustamante and Shringarpure, 2015). Especially as the technology progresses, correlations between phenotype and genotype will be more easily ascertained and associated with an individual. In fact, in a “vulnerability research” experiment, a team at MIT’s Whitehead Institute was able to identify 50 individuals who had participated in genomic studies by their full name using only a computer, Internet access, and public resources. Other studies have shown that an individual can be identified even if a distant relative submits DNA (Fearer, 2013). In another example, Harvard Medical School’s Personal Genome Project (PGP) is based on the premise that guaranteeing privacy is impossible. When recruiting volunteers, PGP informs them of the benefits and risks of participating, which can be potentially dramatic. PGP publishes participants’ demographic and clinical information, their genomic sequences, and their names and headshots if they wish. George Church, the geneticist in charge, calls this open consent (Kupersmith, 2013).

There is a strong argument for collecting and/or archiving genomic data in large databases. It is widely believed that “big” data will propel research forward. No single organization or laboratory will collect sets of data that will be large enough to truly accelerate the science that is critical to understand the genome. This is largely due to practical and financial barriers. Thus, it is important that these scarce resources are used responsibly (National Institutes of Health, 2015).

There are challenges to these large databases. It is not possible, and it may not be best, to aggregate these data in a single large centralized repository. Therefore, federated

models are used and/or proposed by such entities as the Global Alliance for Genomes and Health. There is a significant challenge in the federated model: it is difficult to ascertain whether data are duplicative. There are two obvious ways to mitigate this problem. One would be to give everyone a unique identifier from a central generator of some sort, for example, the National Institutes of Health’s Global Unique Identifier (National Institutes of Health, 2015). Another would be to put the participant at the center, much like monetary banking, and allow the participant to control access to their health information in the context of their needs (Terry, 2013). Combinations of these models also offer interesting solutions.

Access to large sets of data will likely enable researchers to predict who might develop a condition and then appropriately personalize treatment (Collins and Varmus, 2015). Data from large populations increase the probability of finding the genetic correlation between genotype and phenotype. If reasonable privacy and security protections are in place and databases maintain transparency, the risks are manageable in the face of the benefits gained through research.

Today, there is a plethora of privacy and security policies for various databases. For example, in Estonia, a government project is creating a database that includes genetic information aiming to involve three quarters of the country’s population. It will be used in large-scale association studies (Frank, 2015). In another example, Kaiser Permanente institutional review board oversees their collection (Kaiser Permanente, 2015). Mayo Clinic’s Biobank privacy policy says that samples will not be stored with a name, address, birth data, social security number, or Mayo Clinic number and also comments that in the case of reidentification, the Genetic Information Nondiscrimination Act (GINA) of Library of Congress (2008) offers protection (Mayo Clinic, 2015).

The diversity of protections and rules is complex. Ultimately, privacy and security cannot be guaranteed. Despite these risks, large-scale sharing of genomic information and associated clinical information is essential to accelerate biomedical research. President Barack Obama announced the Precision Medicine Initiative (PMI) at the State of the Union address in 2015. Throughout 2015, the NIH, White House, and FDA have worked to flesh out what this effort will entail. A report was issued in September 2015 (National Institutes of Health, 2015). The PMI calls for a cohort of more than 1

¹Yale University, New Haven, Connecticut.

²Genetic Alliance, Washington, District of Columbia.

million people, many of whom will be sequenced and all of whom will contribute health information, data from wearables, and environmental data. This cohort will form the foundation for a longitudinal study with the hopes of answering many questions. Critical to all of this is the principle that people should be treated as partners. There must be a high degree of authentic engagement and transparency. The stakes are high for those who suffer. PMI and other large cohorts that collect genomic data are critical to alleviating this suffering.

References

- Collins FS, Varmus H (2015) A new initiative on precision medicine. *N Engl J Med* 372:793–795.
- Erlich Y, Narayanan A (2014) Routes for breaching and protecting genetic privacy. *Nat Rev Genet* 15:409–421.
- Fearer M (2013) Scientists expose new vulnerabilities in the security of personal genetic information. In: RSS News. Whitehead Institute for Biomedical Research, January 17, 2013
- Frank L (2015) Give and take-Estonia's new model for a National Gene Bank. In: Give and Take—Estonia's New Model for a National Gene Bank. Genome News Network, October 6, 2000.
- Kaiser Permanente (2015) Kaiser Permanente Division of Research. In: Privacy and Confidentiality. Kaiser Permanente, Oakland, CA.
- Kupersmith J (2013) The privacy conundrum and genomic research: re-identification and other concerns. In: Health Affairs. Project HOPE, September 11, 2013.
- Library of Congress (2008) Genetic Information Discrimination Act of 2008. In: Bill Summary & Status. The Library of Congress, Washington, DC.
- Mayo Clinic (2015) Issues in Genomic Medicine and Research. Center for Individualized Medicine. Mayo Clinic, Rochester, MN.
- McGuire A, Fisher R, Cusenza P, *et al.* (2008) Confidentiality, Privacy, and Security of Genetic and Genomic Test Information in Electronic Health Records: Points to Consider. Nature.com. Nature Publishing Group, London, United Kingdom.
- National Institutes of Health (2015) PMI Working Group—Precision Medicine Initiative—National Institutes of Health (NIH). U.S National Library of Medicine.
- Shringarpure SS, Bustamante CD (2015) Privacy risks from genomic data-sharing beacons. *Am J Hum Genet* 97:631–646.
- Terry SF, Shelton R, Biggers G, *et al.* (2013) The haystack is made of needles. *Genet Test Mol Biomarkers* 17:175–177.

Address correspondence to:

Sharon F. Terry, MA

Genetic Alliance

4301 Connecticut Avenue, NW

Washington, DC 20008

E-mail: sterry@geneticalliance.org