Genetic privacy needs a more nuanced approach

Because confidentiality of health data cannot be guaranteed, people should consider both the risks and advantages of sharing them, argues **Misha Angrist**.

The US National Institutes of Health has warned that research is at a "crucial juncture". Bioethicists are fretting. Scientists are anxious. And all because an article in *Science* last month raised doubts about the privacy of volunteers who hand over their genetic data (M. Gymrek *et al. Science* **339,** 321–324; 2013). "Oh my God, we really did this," said Yaniv Erlich of the Massachusetts Institute of Technology in Cambridge to *The New York Times*, after his group managed to cross-reference information from public databases to put names to samples of DNA donated to research. One can imagine law enforcement salivating at the prospect of turning a bloodstain into a name and address.

Yet what the scientists did is not shocking or all that new. The DNA re-identification bogeyman has lurked at the door for years. The warning signs were there in 2005 when a precocious 15-year-old boy called Ryan Kramer found his sperm-donor father. Just as Erlich and his colleagues would do years later, Kramer used a combination of Y-chromosome data — his own in this case — and genealogical searching of public records to track down a donor dad who had almost certainly been promised anonymity by the sperm bank.

No responsible scientist can guarantee absolute privacy. Researchers know this and many volunteers accept it, yet official discussion of the issue remains firmly rooted in the twentieth century. Like whales and rainforests, research participants are viewed only as helpless things that must be protected. I suspect that much of the present hand-wringing has less to do with the welfare of these people and more to do with protecting researchers and their institutions from legal action.

Because information about individuals' health can be used to discriminate against them, the privacy provisions of the US Health Insurance Portability and Accountability Act of 1996 (HIPAA) were revised in 2003 to create a category of protected health information that can be used or disclosed only under certain conditions.

Although genetic data are considered protected health information under the HIPAA, many of the protections disappear when the information is 'de-identified' — that is, the 18 identifiers specified in the act (including names, addresses, birthdates and the like) are removed. And because genetic information is not one of those 18 identifiers, it does not need to be removed from health records to follow the letter of HIPAA privacy. If researchers do not know who you are, and cannot easily find out, then their obligations to you diminish by orders of magnitude. Furthermore, their protocols are less likely to need full review by an institutional review board; their grant applications become less onerous; and their technology costs go down.

One can see, then, how the Kramers of the world pose a problem not just to sperm donors, but also to biomedical research. What if the absence of the 18 identifiers isn't enough to protect someone's identity?

A few weeks ago, the US Department of Health and Human Services had the perfect opportunity to address this issue when it released its 563-page reboot of the HIPAA. But although it addressed genetic information explicitly, the de-identification criteria were summarily brushed off in a single sentence on page 416: "The Privacy Rule's de-identification standard is outside the scope of this rulemaking."

The risks of re-identification from genomic data sources were partly responsible for the launch of the Personal Genome Project (about which I have written a book and on whose unpaid board of directors I serve). The project's approach has been to eschew any promises of privacy and confidentiality. To date, it has more than 2,000 participants, all of whom have agreed to make public, and potentially identifiable, any genomic, medical, environmental and trait data collected about them during the study. I am one of them.

Such open consent is not for everyone. Many of the risks — from identity theft to being framed for crimes — are clear. So why would anyone enrol?

Fairness, for one: I can, if I want, access my sequence and other 'omic' data at any time, day or night. So, too, can a poorly funded geneticist in a tiny lab in Slovenia or Kenya. My data are not privy only to the select few running the study.

Second, research will work better if scientists have more information about the people they study. If an investigator wants to study the genome of someone with an anxiety disorder, ear pits and male pattern baldness, he or she is free to look me up. If someone is interested in induced pluripotent stem cells from a human male, mine are available from the Coriell Institute for Medical Research in Camden, New Jersey. If we agree that part of the mission of biomedical science is to understand the relationship between genotype and phenotype, it is surely helpful to have access to a cohort's unredacted phenotypes before its members die (at which point they are no longer considered 'human subjects' in the eyes of the government).

IT IS SURELY **HELPFUL** TO HAVE ACCESS TO A COHORT'S **UNREDACTED** PHENOTYPES BEFORE ITS MEMBERS DIE.

Third, some genomic information is going to be medically useful. A few months ago, Bloomberg News reporter and Personal Genome Project participant John Lauerman learned that he was predisposed to a rare blood disorder, signs of which he can keep watch for. Finally, as Erlich and Kramer have shown, de-identification is increasingly difficult. Privacy and confidentiality are important principles. But being identifiable has some benefits, and being anonymous has some costs; science will be better off when it acknowledges this reality. ■

➲ NATURE.COM
Discuss this article online at:
go.nature.com/uvx1dx

Misha Angrist *is an assistant professor at the Duke University Institute for Genome Sciences and Policy in North Carolina and author of* Here is a Human Being: At the Dawn of Personal Genomics.
e-mail: misha.angrist@duke.edu