

# Every Patient a Subject

**When personalized medicine, genomic research, and privacy collide.**

By [Jennifer J. Kulynych](#) and [Hank Greely](#)



Advances in data science and information technology are eroding old assumptions about the anonymity of DNA specimens and genetic data.

DNA by Isak55/Shutterstock, hand with blood Praisang/Shutterstock.

Personalized medicine, the hoped-for use of the information in our genes to inform our medical care, may end up helping people live longer, healthier lives. Or it may not—the jury is still out. But one thing is certain: As our unique genomic data enter our medical records, researchers will be tempted to use that invaluable resource. The results may be good for science but bad for patients' privacy.

In 2013, reporter Carole Cadwalladr, writing for the [Guardian](#), described her encounter with the paradox of personalized medicine: Unlocking one's genetic code may feel empowering, but the implications can be frightening. Cadwalladr agreed to let Illumina, a company that makes and uses gene-reading machines, sequence her DNA and use her genome in research in connection with an upcoming conference.

At a conference she attended, Illumina gave all the participants party favors: iPads with copies of their own genomes. Cadwalladr was unnerved to realize that her unique genetic code was now stored by Illumina in the Amazon Cloud and could, like all digital data, be potentially hacked and leaked. But, she reminded herself, she had been told the risks and benefits and had made an informed choice to volunteer.

This choice is denied to many subjects of genomic research—a group that may one day soon include almost all of us. Cadwalladr told us she was “surprised to learn” that current norms for medical research permit a scientist who gets a sample of blood, tissue, or saliva to sequence and use that genome without the donor's specific consent, or even without her knowledge. The scientist then may share those genomic data with others, including a database maintained by the U.S. National Institutes of Health that's used by researchers and companies worldwide. This can all happen without any notice to the people whose DNA was sequenced. (In fact, if the study is federally funded, in some cases the scientist *must* share the information.) These practices are currently acceptable, as long as the genome is viewed as “de-identified”—meaning it isn't linked to obvious identifiers, such as names, addresses, or phone numbers, and it is not, in itself, considered identifiable.

That sounds reasonable, but “de-identification” is becoming only a reassuring myth. Subjects of genomic research should not confidently expect to remain anonymous. The *possibility* of “re-identifying” people from either their genomes or the health or demographic data connected with those genomes is real. The *probability* of re-identification is unclear but certainly growing, as the focus of genomic research shifts from the individual to the population, from small collections of DNA to vast electronic databases of genomic and health information.

Advances in data science and information technology are eroding old assumptions—and undermining researchers' promises—about the anonymity of DNA specimens and genetic data. Databases of identified DNA sequences are proliferating in law enforcement, government, and commercial direct-to-consumer genetic testing enterprises, especially in genetic genealogy. That growth is increasing the likelihood that anyone with access to such nonanonymous “reference” databases could use them to re-identify the person who provided a “de-identified” gene sequence. People with access could include amateur genetic genealogists but also hackers.

Similarly, information about a person's health conditions or demographic characteristics can be used for re-identification. How many 6-foot-2-inch-tall 62-year-old white men are there in a given state with white hair, an artificial left hip, type A positive blood, and a prescription for warfarin?

Newer re-identification risks will emerge as scientists learn to profile individuals using information encoded in the genome itself, such as ethnicity and eye color. Authors of a recent [study](#) published in *PLOS Genetics* described a method to use the genome and computerized rendering software to “computationally predict” 3-D models of individual “faces” of particular genomes; in a subsequent paper the authors describe how these techniques will be useful in criminal investigations.

Today medical ethicists, lawyers, and data scientists dispute whether de-identification remains a reliable means of privacy protection. One camp maintains that the risks of re-identification are overstated, creating a climate that impedes research unnecessarily; another group of experts, the “re-identification scientists,” counter by demonstrating repeatedly how they can [re-identify supposedly anonymous subjects in genomic research databases](#).

Yet to date, this conversation has been largely academic. Gene-sequencing technology is only now maturing into clinical use, and the number of people whose genomes have been sequenced for research in

the United States is relatively small compared with the total patient population. Though many of these research subjects contributed DNA before the advent of sequencing technology and are likely unaware that their genomes have been sequenced and shared, most did consent to participate in some form of medical research and provided DNA samples for this purpose. In theory, therefore, these subjects, like Carole Cadwalladr, all knew they were assuming new privacy risks by joining a study.

This is about to change, as gene sequencing moves from the research laboratory to the clinic— and we need to consider the consequences carefully. When the day arrives that each patient’s genome is sequenced routinely in the course of our medical care, all our genomic data will become part of, or linked to, our permanent, electronic medical records.

EMRs with gene-sequence information will be a treasure trove for genomic research on a populationwide scale, allowing researchers to forgo recruiting DNA donors in favor of obtaining genomic data directly from the EMR. The temptation to do good by doing research on this vast scale will be irresistible; the mushrooming literature on such genome-wide association studies shows that these very large studies may offer researchers enough statistical power to tease apart the complex interplay of genetic contributions to almost any health condition imaginable, from schizophrenia to diabetes.

Commonly accepted practices for records-based research, which don’t require patient consent, could eventually cause many of us to become the subjects of genomic research without our knowledge. As has already happened for many of the nearly 1 million [subjects in the NIH genomic database](#), our genomes might then be distributed to researchers worldwide, and we’d never know. That people who volunteered for specific studies have their genomes distributed across the world without their knowledge is bad enough. That this might happen to people who have sought medical care but have *not* volunteered for research would be worse.

Patients today generally don’t know when their medical records have been disclosed for research, or to whom—making it difficult to object. In the not-so-distant future, when medical records include our unique genomes, this status quo will be ethically unacceptable. To date, regulators have interpreted federal health privacy law to permit providers to treat whole genome sequence data as “de-identified” information subject to no ethical oversight or security precautions, even when genomes are combined with health histories and demographic data. Either this interpretation or the law should be changed.

The same EMRs that will make this research possible could also be used to record patients’ choices whether to participate in research, but that is not generally happening. If the research community truly believes that science must conscript patient genomes for public benefit, it should make that case openly, explaining how notice and consent will impose undue burdens on crucial research. Otherwise, do the right thing: Ask patients first.

*This article is part of [Future Tense](#), a collaboration among [Arizona State University](#), [New America](#), and [Slate](#). [Future Tense](#) explores the ways emerging technologies affect society, policy, and culture. To read more, visit the [Future Tense blog](#) and the [Future Tense home page](#). You can also [follow us on Twitter](#).*

Jennifer Kulynych is senior counsel at the Johns Hopkins Hospital and research counsel for Johns Hopkins Medicine. Hank Greely is the Deane F. and Kate Edelman Johnson professor of law and professor, by courtesy, of genetics at Stanford University, where he directs the Center for Law and the Biosciences.

*Together they are writing a white paper on this topic; their work is supported by the Greenwall Foundation. The views expressed are the authors’ alone and do not necessarily reflect positions of the Greenwall Foundation or of Johns Hopkins or Stanford University.*