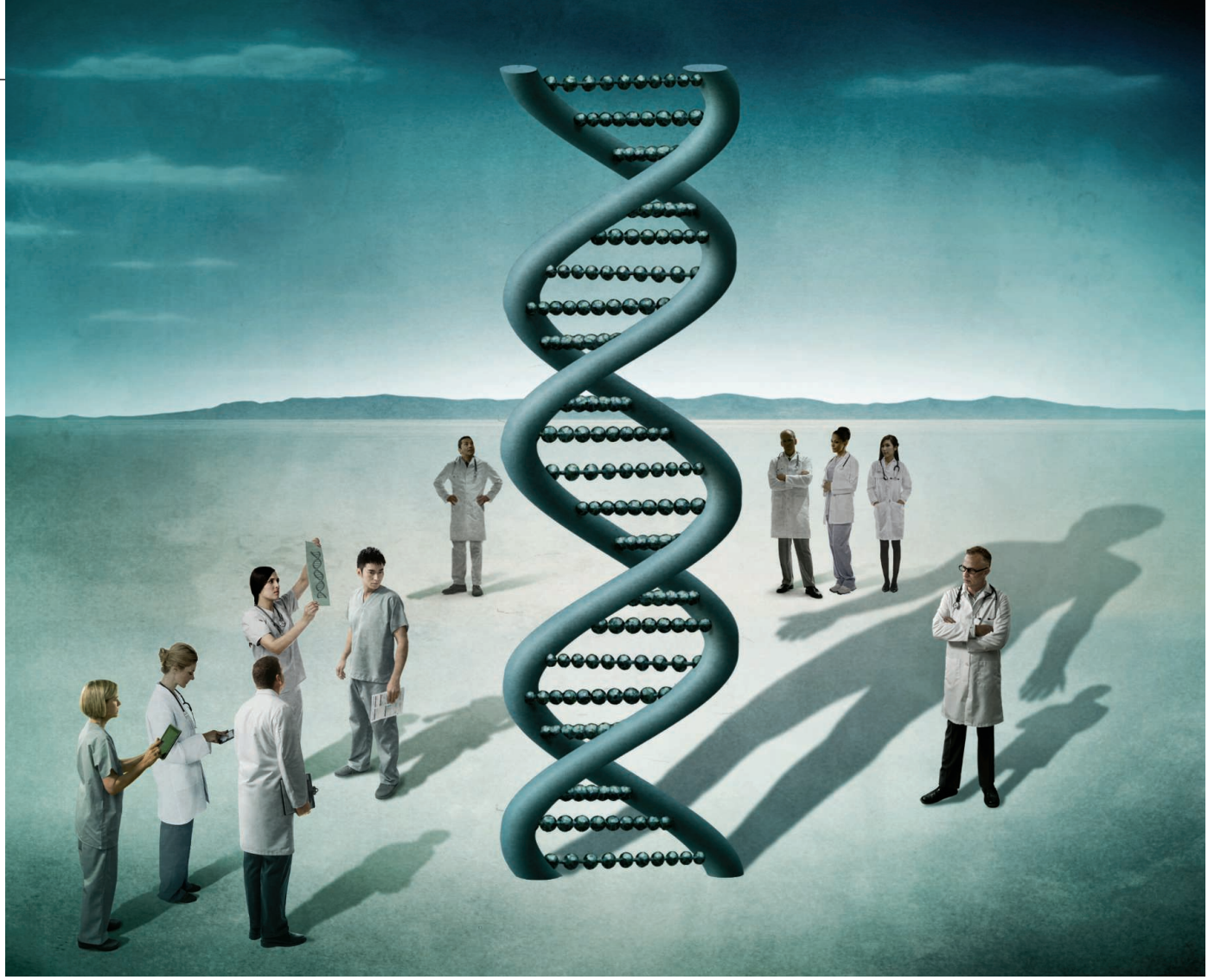


IMAGE: WILLIAM DUKE

TRUST ME, I'M A MEDICAL RESEARCHER

Scientists can no longer guarantee patients' privacy. They're looking for new ways to build trust

By **Jennifer Couzin-Frankel**

In an Oxford, U.K., suburb, a short distance from the track where Roger Bannister ran the world's first 4-minute mile, a quiet revolution is under way. One hundred and twenty-six people and counting, all suffering from a rare rheumatologic disease or the parent of an affected child, are involved in a research project on the disorders. But rather than donating a few samples, filling out a questionnaire, and hoping something useful will come of it one day, these subjects are deeply invested in the research. They are contributors with a voice.

One is Elaine Rush, a 53-year-old from outside Southampton. She was born with the brittle bone disease osteogenesis imperfecta and has had, in her words, "only around 25 fractures in all—quite low compared to many." Once not expected to live past the age of 5, Rush uses a wheelchair and has battled heart and lung problems associated with the disease. An eager partner in the quest to advance science, she now dials in to Skype calls every other month with one or more researchers and offers advice. When they were struggling with recruitment, Rush advocated posting on Facebook, where patients find each other. The study leaders are looking to follow her suggestion.

RUDY, as this project to study rare diseases of the bones, joints, and muscles is called, represents a new kind of bargain between researchers and subjects in response to dwindling expectations of privacy. Until quite recently, a volunteer might have

offered DNA or tissue to a single research group at a nearby university. Today many samples are banked, sequenced, and shared with potentially thousands of researchers. That allows for bigger studies with more statistical muscle, but it also makes it more difficult to keep patients' data private. It's now widely accepted that if someone can read your DNA, they might figure out who you are, either now or in the future, as technology marches ahead. The promise long made to participants—that their identity is stored in an unbreakable vault—no longer holds.

"Patients are scared about access to 'my data,' " says one of RUDY's leaders, Kassim Javaid, a balding, bespectacled University of Oxford rheumatologist based at the university's Nuffield Orthopaedic Centre. Offering them many layers of control, as RUDY does, "is a possible solution," he believes.

Rush and other participants in RUDY—which is co-funded by the United Kingdom's National Institute for Health Research—can decide whether their blood, their scans, and their medical histories can be shared with researchers at, say, a lab elsewhere in Europe or in the United States. They will be able to log on to a clinical trial Web page to learn whether one of their tissue samples has been flagged—an indication that a researcher somewhere is studying it.

Throughout biomedical research, the advent of large repositories of DNA and tissue samples has forced researchers and ethicists to rethink their relationship with the volunteers who make their work possible. "Twenty years ago, people consented to do experiments based on trust and a handshake," says Jamie Heywood, the co-founder of PatientsLikeMe in Cambridge, Massachusetts, a company that provides a platform for people with different diseases to share their health data. Now, patients shake hands with faceless others around the world. And in return for sharing their DNA far and wide—and potentially shelving their privacy—participants want a louder voice in research, and transparency about how it's conducted.

THAT THE END OF GENOMIC PRIVACY has arrived became clear 2 years ago, when a young human geneticist now at Columbia University, Yaniv Erlich, published a startling paper in *Science* (18 January 2013, p. 321) that confirmed the worries of many in the field. Erlich and his colleagues showed that it was possible to identify a man based on a partial DNA sequence of his Y chromosome, his age, and his U.S. state of residence—the type of basic information that researchers commonly post in DNA databases widely accessible to their community. By combining these snippets of information with what he found for others in the same family on popular genealogy databases—where more than 100,000 people have already posted DNA markers—Erlich could not only identify the donors of the DNA, but also their family members as far as second cousins once removed. "I don't even know my second cousins," he says.

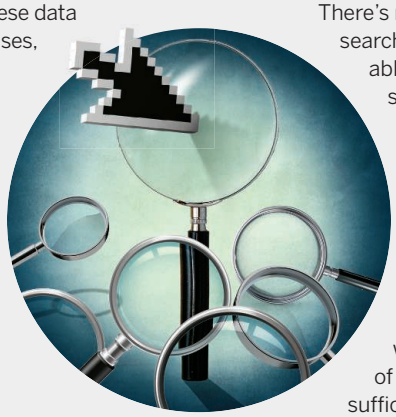
Erlich hastens to point out that DNA can be anonymized, for instance by scrambling or deleting nucleotides containing sensitive information. That technique can render the information largely useless for research, however, and doesn't even always protect the donor. When James Watson, a co-discoverer of DNA's double helix, had his genome sequenced and published in *Nature* in 2008, he requested that his *APOE* gene—which can reveal a predisposition to Alzheimer's disease—be left out. But as three geneticists politely pointed out later that year in the *European Journal of Human Genetics*,

THE PRIVACY ARMS RACE

Camouflaging searches in a sea of fake queries

By **Jia You**

From health questions to shopping habits, your Web search history contains some of the most personal information that you reveal online. Search engine giants such as Google and Bing carefully log these data and save them in databases, where they might be shared with advertisers and the government.



Privacy-conscious users can switch to anonymous search engines such as DuckDuckGo, which doesn't log a user's IP address, identifying information, or search history. DuckDuckGo alone processes about 7 million direct queries a day; traffic spiked after the 2013 revelations about the National Security Agency's snooping. But these services don't match the speed and convenience that Google offers. For consumers who want to continue using their favorite search services but with added protection, researchers at New York University in New York City have developed a browser extension that produces dummy search requests that drown out a user's real queries, thwarting any attempt to profile them.

The software, known as TrackMeNot—which can be downloaded as a Firefox or Chrome extension—creates the fake search queries by harvesting phrases from RSS feeds from popular websites such as *The New York Times*. Dummies such as "George Clooney" and "Amtrak" are sent to a search engine in the background while consumers use their browsers as usual. You can customize the RSS feed to control the content of the decoys and pick which search engines to target. To make the dummies more believable, the algorithm automatically updates the search terms and even simulates clicks on links displayed on the results pages. It can also schedule fake queries primarily when users are actually searching.

There's no guarantee that search engines wouldn't be able to separate the fake searches from the real ones, but it could cost them considerable resources to do so, says privacy expert Helen Nissenbaum, who co-developed the software. She hopes the project will serve as a proof of concept and garner sufficient users—it has more than 60,000 so far—to pressure search engine companies into meaningful dialogues on their privacy policies.

The software may not be much help to users who look up sensitive terms monitored by governments—those related to political opposition, for instance—as it doesn't hide a user's real queries. For those users, computer scientists at Purdue University have developed an algorithm that not only sends out fake queries, but also hides a user's real interests by substituting real queries with phrases related to the same topics. The downside: The results become less relevant, forcing users to go through multiple pages of results to find the link they need. ■

IMAGE: WILLIAM DUKE

genetic knowledge had advanced sufficiently to impute Watson's *APOE* status based on patterns in nearby DNA.

Cryptographers are still exploring how to better protect DNA, and many agree it's important to continue that work. But Erlich's energies have shifted elsewhere. About a year ago, he and about 30 others convened at Cold Spring Harbor Laboratory to consider alternatives to privacy in research. They came back to that decades-old handshake and contemplated how to adapt it for 21st century science.

Erlich was inspired in part by recent Internet phenomena where trust is a guiding force, such as Uber, which runs an online car-sharing service that matches drivers with passengers. "Uber takes two individuals that don't know each other," he says. "I'm getting into someone else's car; he could chop me to pieces." Airbnb, where people offer a room or their entire home for rent to complete strangers, is another example. These websites hold users accountable with reviews, profiles, and extensive documentation. This openness appears to build trust, Erlich says, and he thinks the same strategy can be applied to genetics and biomedical research.

Research volunteers have always valued trust and transparency. In 2007, Alan Westin, a legal scholar who studied privacy and who died in 2013, conducted a survey of almost 2400 people for the Institute of Medicine. He found that respondents were less preoccupied with whether researchers knew who they were than with knowing what was happening to their medical information. Among those surveyed, 81% were not happy to have researchers parsing even so-called de-identified health data without their consent.

"They are not hung up on privacy so much as autonomy," says Mark Rothstein, a law professor at the University of Louisville in Kentucky. "Let's assume that you've de-identified, anonymized, and nobody can figure out who it is—but if people think you've used that information without their permission, they're still going to be very angry."

U.S. regulation is adapting to that sentiment. In August, the National Institutes of Health announced that, starting this month, it expects researchers to obtain informed consent from participants if their DNA, cell lines, tissue, or any other de-identified biological material will be used for research at any point in the future. "Part of governance is transparency," says Bartha Maria Knoppers, who studies law and genetics at McGill University in Montreal, Canada, and is a member of a consortium called the Global Alliance for Genomics and Health, which is looking for new methods to share data more openly and responsibly. "You put in place a process of oversight and a mechanism to ensure that what you tell me is going to happen to my data is what is going to happen to my data."

Knoppers and many others point out that patients often want to share their DNA in the name of advancing research—and that fears of being identified through DNA may be overblown. Some databases ban researchers from re-identifying volunteers. There have been no breaches yet—or at least none that anyone knows about. Gail Jarvik, a medical geneticist at the University of Washington, Seattle, believes that most scientists don't care who handed over a blood or tissue sample. "Why identify them?" she asks.

A HANDFUL OF EXPERIMENTS are now testing how to better inform volunteers about what's happening to their data. PatientsLikeMe has recruited 300,000 people with more than 2300 different diseases. Participants share their health data, analyze how they're faring in clinical trials, support each other, and help researchers and drug companies answer existing scientific questions and pose new ones. Heywood founded the nonprofit in 2004 while his younger brother Stephen was suffering from amyotrophic lateral sclerosis. (Stephen died 2 years later.) Jamie Heywood's philosophy is that if people are "understanding and enthusiastic participants," they will agree that sharing widely will maximize the value of their DNA and other health information to the community—even if this offers them less privacy. The company keeps in touch with participants with a blog, social media, and regular e-mails.

> ***"I personally don't feel that there's anything I need to hide."***
>
> **Elaine Rush**, trial participant

The Personal Genome Project (PGP) at Harvard Medical School in Boston, founded by geneticist George Church, goes even further: It asks participants to share their DNA sequences and health histories online for everyone to see. Almost 4000 have signed up so far; last month, PGP launched a "real name" option, whereby they can post their identity. "Many participants do mention altruistic reasons: sharing data publicly in order to promote our collective knowledge," says Jeantine Lunshof, a Harvard ethicist on the project. "That seems to outweigh potential drawbacks." To make sure participants understand its ramifications, PGP asks a series of hard questions. One example, says Heywood, who participates in PGP but flunked the test the first time: "If I commit a crime, could the DNA in this bank be used to identify me?" (The answer is yes.)

Javaid of RUDY hopes that his strategy, called dynamic consent because consent is a continual process, will change how patients think about research. Patients can choose which portions of the study to complete—questionnaires, or the sharing of scans, for example—and also whether to restrict their data to RUDY investigators or allow them to be more broadly distributed. If someone says, "don't use my DNA, don't use my blood … we archive the samples so no one else can use them," Javaid says, but they're preserved in case the patient changes his or her mind.

Rush, the participant with brittle bone disease, has given broad consent, along with nearly all of RUDY's early adopters. "I personally don't feel that there's anything I need to hide," Rush says, but she recognizes that not everyone feels the same way. She has also reached out to fellow patients. "The fact that they can opt out of some things" has helped her explain the study to those who might hesitate to sign on.

As for Erlich, he's ready to borrow more ideas from Uber and Airbnb. In November, he and nine other attendees at the Cold Spring Harbor meeting published a paper in *PLOS Biology* outlining a "trust-centric framework … that rewards good behavior, deters malicious behavior, and punishes noncompliance." Like people griping online about a driver's body odor or praising the free coffee and snacks in their vacation home, patients could write reviews about the researchers they have worked with. The system could include trusted mediators to engage both researchers and participants, and automated auditing of how study data are used. Perhaps, Erlich speculates, the visibility of researchers and their reputation would climb when they received accolades from peers or high marks from patients for returning results and raw data.

Of course, trust is difficult to build and easy to squander. Last year, Uber itself received a failing grade from the Better Business Bureau after a deluge of customer complaints, and the company has been accused of exaggerating how carefully it vets its drivers. A similar breakdown could transpire in scientific projects.

RUDY has won Rush's trust. "The RUDY researchers are reputable," she says. "They wouldn't be sharing with [just] anyone." As the study plods on in the months and years ahead, its success will depend on upholding that confidence. ■