

# A BROKEN CONTRACT



**L**ate in May, the direct-to-consumer gene-testing company 23andMe proudly announced the impending award of its first patent. The firm's research on Parkinson's disease, which used data from several thousand customers, had led to a patent on gene sequences that contribute to risk for the disease and might be used to predict its course. Anne Wojcicki, co-founder of the company, which is based in Mountain View, California, wrote in a blog post that the patent would help to move the work "from the realm of academic publishing to the world of impacting lives by preventing, treating or curing disease".

Some customers were less than enthusiastic. Holly Dunsworth, for example, posted a comment two days later, asking: "When we agreed to the terms of service and then when some of us consented to participate in research, were we consenting to that research being used to patent genes? What's the language that covers that use of our data? I can't find it."

The language is there, in both places. To be fair, the terms of service is a bear of a document — the kind one might quickly click past while installing software. But the consent form is compact and carefully worded, and approved by an independent review board to lay out clearly the risks and benefits of participating in research. "If 23andMe develops intellectual property and/or commercializes products or services, directly or indirectly, based on the results of this study, you will not receive any compensation," the document reads.

The example points to a broad problem in research on humans — that informed consent is often not very well informed (see 'Reading between the lines'). Protections for participants have been cobbled together in the wake of past controversies and have always been difficult to uphold. But they are proving even more problematic in the 'big data' era, in which biomedical scientists are gathering more information about more individuals than ever before. Many studies now include the collection of genetic data, and researchers can interrogate those

## AS RESEARCHERS FIND MORE USES FOR DATA, INFORMED CONSENT HAS BECOME A SOURCE OF CONFUSION. SOMETHING HAS TO CHANGE.

BY ERIKA CHECK HAYDEN

data in a growing number of ways. Several US states, including California, are considering laws that would curtail the way in which researchers, law-enforcement officials and private companies can use a person's DNA.

The research coordinators who develop consent forms cannot predict how such data might be used in the future, nor can they guarantee that the data will remain protected. Many people argue that participants should have more control over how their data are used, and efforts are afoot to give them that control. Researchers, meanwhile, often bristle at the added layers of bureaucracy wrought by the protections, which sometimes provide no real benefits to the participants. The result is a mess of opinions and procedures that sow confusion and risk deterring people from participating in research.

"A lot of times researchers will say, 'Why can't we just go back to the way it was?', which was basically that we take these samples and people do it for altruistic reasons and everything's lovely," says Sharon Terry, president of the patient-advocacy group Genetic Alliance in Washington DC. "That worked in a prior age. I don't think it works today."

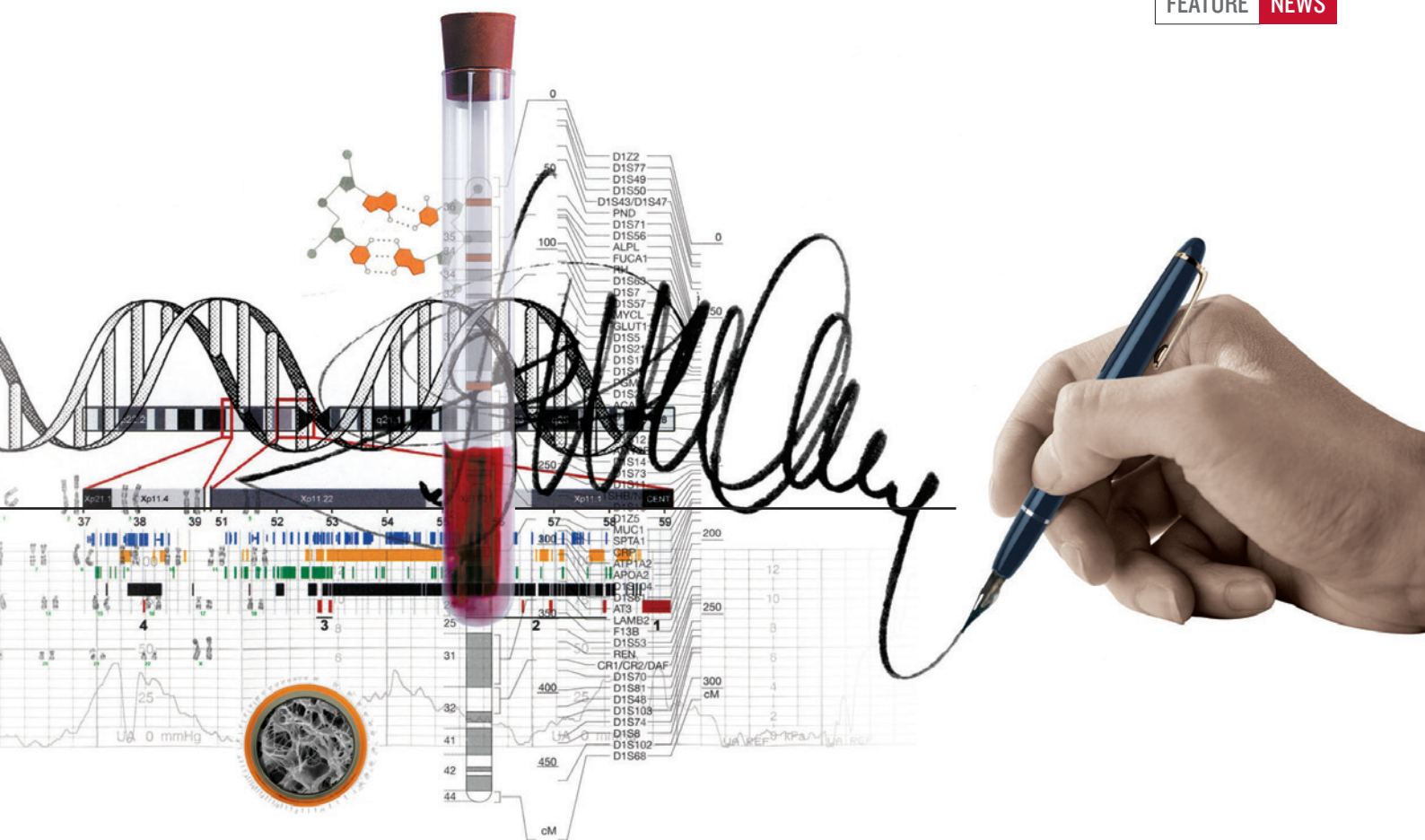
The concept of informed consent was first set out in the Nuremberg Code, a set of research-ethics principles adopted in the wake of revelations of torture by Nazi doctors during the Second World War. But in recent years, a series of mishaps over consent have

undermined support for research. In 2004, for example, scandal erupted in the United Kingdom after parents found out that from the late 1980s to the mid-1990s doctors and researchers had removed and stored organs and tissues from patients — including infants and children — without parental consent. New laws were passed that required explicit consent for such collections.

Then, in 2010, the Havasupai tribe of Arizona won a US\$700,000 settlement against Arizona State University in Phoenix. Individuals believed that they had provided blood for a study on the tribe's high rate of diabetes, but the samples had also been used in mental-illness research and population-genetics studies that called into question the tribe's beliefs about its origins. In the settlement, the university's board of regents said that it wanted to "remedy the wrong that was done".

The cases illustrate the divide between researchers and the public over what people need to know before agreeing to participate in research.

Many of the recent concerns over consent are driven by the rapid growth of genome analysis. Decades ago, researchers weren't able to glean much information from stored tissue; now, they can identify the donor, as well as his or her susceptibilities to many diseases. Researchers try to protect the genetic data through technological and legal mechanisms, but both approaches have weaknesses.



VIKTOR KOEN

It is not enough to strip out any information that would identify the donor, such as names and full health records, before the data are stored. In 2008, geneticists showed that they could easily identify individuals within pooled, anonymized data sets if they had a small amount of identified genetic information for reference (N. Homer *et al.* *PLoS Genet* 4, e1000167; 2008). And it may become possible to identify a person in a public database from other information collected during a study, such as data on ethnic background, location and medical factors unique to the study participants, or to predict a person's appearance from his or her DNA.

Even legal mechanisms have vulnerabilities. In 2004, Jane Costello, a social psychologist at Duke University in Durham, North Carolina, was forced to go to court to defend the confidentiality of patient records from the Great Smoky Mountains Study. The study, which is just going into its third decade, examines emotional and behavioural problems in a cohort of people who enrolled as adolescents. A participant in the study was testifying against her grandfather, John Trosper 'JT' Bradley, who had been accused of sexual abuse. JT's lawyers subpoenaed the granddaughter's records from the study in hope that the information would undermine her credibility as a witness.

It meant a major crisis of confidence for Costello. "I was telling 1,400-plus people every time we saw them that 'your data are absolutely

safe, and now I was in a position where I was told, 'No, that's not true,'" she says. After Costello's day in court, in August 2004, the records remained sealed, but mostly because the judge did not believe that they would exonerate JT. The result provided no clarity about patient protections.

#### BETTER MODELS FOR CONSENT

One solution is to keep genetic information separate from demographic data. The BioVU databank at Vanderbilt University Medical Center in Nashville, Tennessee, for instance, contains DNA samples from patients treated at the hospital — 143,939 people as of 11 June. The DNA is linked to health records in a second database, called a 'synthetic derivative', in which the data are anonymized and scrambled in ways that, its creators say, make it difficult for anyone to work back from the database to verify a patient's identity. Sample-collection dates are altered, for example, and some records are discarded at random, so that it is not possible to know that someone is in the database just because he or she was treated at the hospital. Even researchers who work with the data cannot determine whose data they are using. The databank expects to include as many as 200,000 individuals by 2014, making it one of the largest collections of linked genetic and health records in the world.

But when it comes to consent, BioVU takes a different approach from many other

programmes. Patients don't choose to participate; rather they are given the chance to opt out. Patients are asked to sign a 'consent to treat' form every year. It includes a box that they can tick to keep their DNA out of the database. That model helps BioVU to collect many more samples, and much more cheaply, than other projects can.

The opt-out model — which is used in only a few other places — troubles Misha Angrist, a genome policy analyst at Duke University, who says that it risks taking advantage of people when they are ill. "Even a routine visit to the clinic can be a vulnerable moment, and they're saying, 'Would you mind doing this for future generations, to help people just like you?'"

And legal challenges have shown the weaknesses of opt-out policies. Health officials are now destroying millions of blood samples taken from newborn babies in Texas and Minnesota because the families were not adequately informed that the samples, collected to screen for specific inherited disorders, would also be used in research.

Vanderbilt officials and researchers counter that they have run extensive public campaigns to ensure that people in Nashville are aware of BioVU and are comfortable with the way it works. They regularly consult a community advisory board about the project. And Vanderbilt's approach actually goes above and beyond what is required by federal law; because the synthetic derivative includes de-identified

## READING BETWEEN THE LINES

Despite the work that goes into making consent forms clear and detailed, some participants say they are confused by the wording.

## 5. What are the benefits and risks of participating?

.....If 23andMe develops intellectual property and/or commercializes products or services, directly or indirectly, based on the results of this study, you will not receive any compensation.

23andMe received a patent in May for its work on Parkinson's disease.

Some participants did not expect it to seek intellectual-property rights.

From 23andMe's research consent form [www.23andme.com/about/consent/](http://www.23andme.com/about/consent/)

2. I have been informed that the purpose of the research is to study the causes of behavioral/medical disorders.

Many participants in the Medical Genetics at Havasupai study in Arizona in the 1990s were unaware of how broad the research goals were and what kind of studies would be performed.

Courtesy of Pilar Ossorio

## Re-identification

We are quickly learning that with powerful computers and good mathematicians, it is increasingly possible to uniquely identify people inside large data sets ... Researchers will sign a contract in which they agree that, even if they were able to identify you, they won't do it.

In an attempt to be more transparent about privacy risks, the Consent to Research project's 'Portable Legal Consent' essentially states that although anonymity cannot be guaranteed, researchers must pledge to uphold it.

Courtesy of John Wilbanks Consent to Research

data, it doesn't legally require informed consent at all. Last July, the US Department of Health and Human Services signalled that it might be rethinking the rules that exempt de-identified data from the consent requirement, as part of a broad overhaul of research ethics regulations.

Irrespective of the outcome, obliterating patient identities has drawbacks. Researchers can't perform some types of research on the scrambled data. Because dates are changed, studies on the timing of influenza infections, for example, are impossible. And patients can't be told if the research has revealed that they carry individual genetic risks linked to disease.

## FULL DISCLOSURE

Returning study results to research participants has been another thorny issue for consent. Doctors might learn about genetic predispositions to disease that are separate from the ailments that led a patient to participate in the research in the first place, but it is not clear what they should do with this information.

UK researchers, for example, are forbidden from sharing genetic results with participants. But US research societies, such as the American College of Medical Genetics and Genomics in Bethesda, Maryland, are moving towards adopting standards that would encourage the practice for some types of findings, such as those that are medically relevant.

Some countries, such as Germany, Austria, Switzerland and Spain, are already feeding back such information. And some clinical sequencing programmes are considering offering patients 'tiered' consent, in which people can decide whether to be told about their data and how much they want to learn.

This is what Han Brunner, a geneticist at the Radboud University Nijmegen Medical Centre in the Netherlands, had hoped to do.

Last year, he began a project to sequence the exomes — the protein-coding regions of the genome — of 500 children and adults, looking for the genetic causes of intellectual disabilities, blindness, deafness and other disorders. Brunner proposed allowing participants to choose from three options: they could learn everything that researchers had divined about disease susceptibility; just information relevant to the disease for which their genomes were examined; or no information at all. Ethics reviewers shot down his proposal. "They said that in practice, it would be impossible for people to draw those lines, because people giving consent cannot foresee all the possible outcomes of the study," Brunner says. Instead, everyone participating in the studies must agree to learn all medically relevant information arising from the analysis of their genomes. As a consequence, Brunner recently had to tell the family of a child with a developmental disability that the child also has a genetic predisposition to colon cancer. Not all researchers endorse the idea of informing children about diseases that might affect them as adults. In this case, doctors recommended early screening, and Brunner says, "the family handled it very well; they said, 'This is not what we anticipated, but it's useful information'".

Many of the studies done now ask patients to give consent for research linked to particular investigators or diseases. But that means that researchers cannot pool data from separate studies to tackle different research questions. Many researchers say that the obvious solution is a broad consent document that gives researchers free rein with the data. But many non-scientists think participants should be able to control how their data are used, says lawyer Tim Caulfield of the University of Alberta in Calgary, Canada, who has surveyed patients about this idea. "There's an emerging consensus

within the research community about the need to adopt things like broad consent, but that hasn't translated out to the legal community or to the public," he says.

Another solution might be called 'radical honesty'. A US project called Consent to Research, which aims to provide a large pool of user-contributed genomic and health data, has devised what it calls a 'Portable Legal Consent', which allows anyone to upload information about him or herself, such as direct-to-consumer genetic results and lab tests ordered through medical providers, to an interface that strips the data of identifiers. It makes the data widely available to researchers under broad guidelines, but also requires data donors to go through a much more rigorous consent process than most studies do. The Portable Legal Consent specifically informs participants that researchers might be able to determine their identities, but that they are forbidden from doing so under the project's terms of use.

Such approaches could help scientists by giving them access to a trove of data with no restrictions on use. But the participant protections system that is in place might not be ready for such frank dialogues, says Angrist, who serves on one of Duke's institutional review boards.

While reviewing a research proposal for a large biobank, for example, Angrist suggested that the researchers send the participants an annual e-mail explaining how their samples were being used, and thanking them for donating their time and tissue. The review board voted this suggestion down after its chair argued that e-mailing the patients would create a problem in light of the Health Insurance Portability and Accountability Act (HIPAA) — the US law that guarantees the privacy of health records. "The irony is that the HIPAA is supposed to protect people, and what I was hearing was, 'We can't talk to people because we're too busy protecting them,'" Angrist says. "Institutions use informed consent to mitigate their own liability and to tell research participants about all the things they cannot have, and all the ways they can't be involved. It borders on farcical."

But as patient data become more precious to researchers, and as advocacy organizations become more involved in driving research agendas and in funding the work, such paternalistic attitudes will probably not survive, says Terry. She adds that technologies that allow research participants to control and track how researchers use their data will soon catch on. These approaches could benefit patients, who gain transparency and control over their data, and researchers, who gain access to richer data sets. "I think we're going to have to get to a place where consenting people becomes customizable easily through technology, and we're not there yet," Terry says. ■ SEE EDITORIAL P.293

**Erika Check Hayden** writes for Nature from San Francisco, California.