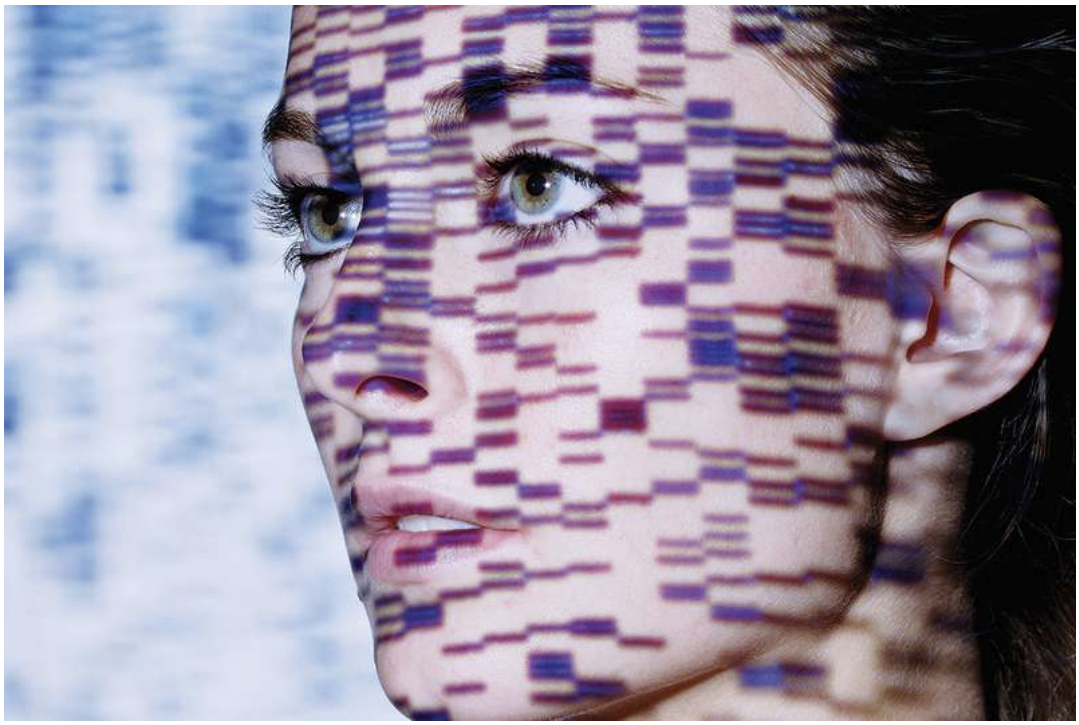
Your genetic data can be exploited without you ever knowing about it

Your genome literally identifies you, but researchers and genetic firms keep saying that DNA data is anonymous. It's a privacy scandal waiting to happen

Technology 27 February 2019



Tim Flach/Getty

By Chelsea Whyte

EVERY person in the world is issued with a unique code before they are even born. Governments, insurance firms and indeed pretty much anyone can use this code to catalogue us throughout our entire lives. This isn't a sci-fi dystopia â€“ it is just genetics.

While your genome doesn't explicitly record your name, address or other identifying information, the rise of consumer genetics services means these details are increasingly being linked to your DNA. Once that happens, the gene genie is out of the bottle.

Yet many genetics firms and researchers continue to insist that your genome isn't personally identifiable information, despite it literally identifying you. A recent revision of US ethical guidelines, for example, has continued this fiction, ignoring multiple objections. So should we be concerned about our lack of genetic anonymity?

The mistaken idea that medical information can be anonymised isn't new. In the mid-1990s, the Massachusetts Group Insurance Commission, which provides healthcare to state employees, decided to make all medical records available for research. The state governor at the time, William Weld, assured the public that the records would be stripped of personally identifiable information.

Then he fell ill, visited a hospital, and a computer science graduate student at the Massachusetts Institute of Technology took the opportunity to show him how easily identifiable his own records were. Latanya Sweeney, now the director of the data privacy lab at Harvard University, used Weld's zip code, birth date and gender to search hospital records on the day he visited. She got an exact match, and sent the governor his medical records in the mail, showcasing the limitations of so-called anonymised data.

â€œDNA made available to researchers doesn't have to be protected and can be used without consentâ€

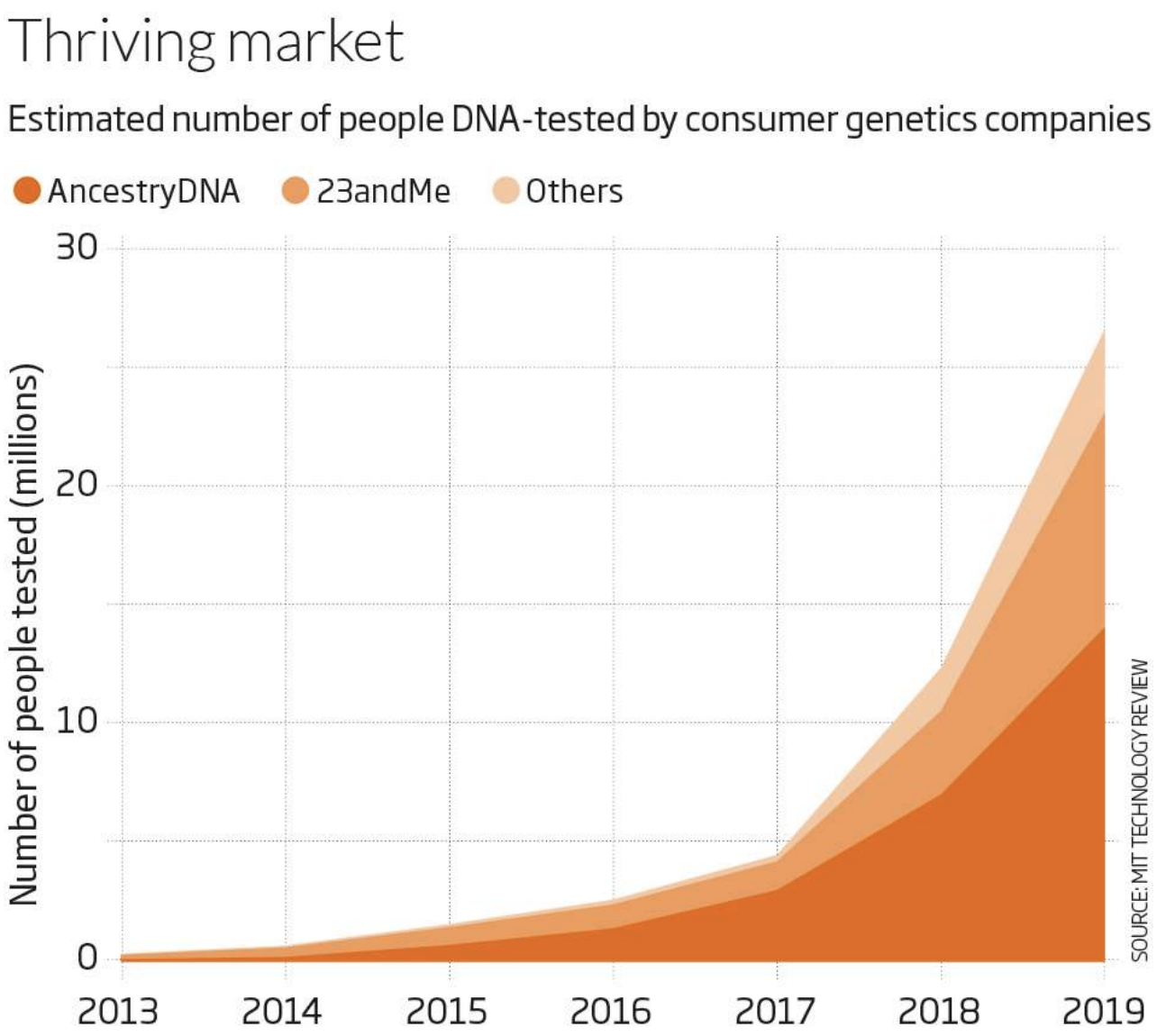
In the US, the privacy of medical data is protected under the Health Insurance Portability and Accountability Act (HIPAA), which lays out 18 identifiers that must be removed before medical data can be stored in an open database for, say, research purposes. This covers obvious things like names, addresses and health insurance account numbers. It also includes some biometric markers, such as fingerprints and voice patterns. But it doesn't include DNA.

In other words, any DNA made available to the research community or to the public in databases doesn't have to be protected. It can therefore be used without consent for research or other purposes.

Law enforcement agencies have taken advantage of this lately, trying a kind of genomic triangulation to find the perpetrators of unsolved murders. This technique has been used to find dozens of suspects by matching DNA left at a crime scene to that on genetic ancestry websites and tracing people through any listed relatives.

The use of this kind of DNA searching has exploded along with the rise of consumer genetic kits. A recent article in *MIT Technology*

Review used public data to estimate that 26 million people around the world have used such kits, sending in a swab of their saliva to one of several genetics companies such as Ancestry DNA and 23 and Me (see â€œGraphâ€).



Even if you haven't taken one of these tests yourself, parts of your DNA are likely to be out there. "We've shown in our research that if we have a database of 2 per cent of a population, then virtually everyone is traceable," says Yaniv Erlich, chief science officer at genetic ancestry company MyHeritage. That is because the DNA of even distant relatives can be linked back to you.

Erlich and his colleagues demonstrated as much in a 2013 study. They used patterns found in DNA called short tandem repeats to search through public genetic databases, and showed that they could discover the surname of the person to whom the DNA belonged. "With enough time and effort, I can get to you," he says.

Myth of anonymity

Stripping records of information like names, addresses and social security numbers was once enough to keep it from being identifiable, but that changed about 20 years ago.

"There was this notion that was useful for decades, that if you redact certain types of information, it becomes quite hard to trace back records. And it actually worked quite well," says Erlich. "But as we got into the era of big data and large-scale internet resources, it became true that it's hard to anonymise any big data."

The myth of genetic anonymity persists, however, because it is useful. It gives researchers access to a wealth of information without having to seek informed consent.

Research of human subjects in the US is governed by the Common Rule, which applies to all federally funded research. This rule is rewritten periodically to bring it in line with current ethical standards and take into account new technology. This happened in January, but the rulebook still doesn't count DNA as identifiable information. "Many people wrote opinions saying that DNA is identifiable and that we should treat it this way," says Erlich. Instead, the new language explicitly says DNA isn't identifiable.

There are clear benefits to allowing this, because it is a good way of sampling the entire population. For example, if you have blood drawn at the doctor's office and there is a bit left over after your tests are done, it could be stripped of identifiers and put into a

repository where it can be used for research without you ever knowing about it. But increasingly, people want control over the use of their data.

In the European Union, the recent General Data Protection Regulation (GDPR) aims to give people that power, but it doesn't apply to "anonymous data", which includes DNA.

"GDPR is of such a general nature that it couldn't possibly address the peculiarities of genetic data," says KÃ¤rt Pormeister at the University of Tartu in Estonia. "The fact that it's shared in significant part with your relatives â€" you don't see that with other types of data."

â€œEven if you haven't taken a genetic test yourself, parts of your DNA are likely to be out thereâ€

Protecting genetic information stored by consumer genetics companies, rather than medical researchers, is even more complex. "They're doing a lot of genetic testing, sequencing, screening and sharing of data. They aren't covered by HIPAA, and in their capacity as consumer-facing, profit-driven companies, they're not covered by research protections. So, they kind of fall outside the basic privacy provisions," says Natalie Ram at the University of Baltimore, Maryland. What does constrain them is their terms and conditions, but these can be unilaterally changed. For example, police were able to upload DNA to GEDmatch, a genetic genealogy database, to look for suspects, without users knowing this was possible. GEDmatch updated its terms of service after the fact.

All this means the cat is firmly out of the bag when it comes to genetic anonymity. "[These databases] create the largest genetic surveillance apparatus for US individuals that has ever been established," says Erlich.

Maybe the best approach is to simply make that fact clear, says Jeantine Lunshof at MIT. "When you generate DNA data, it's out there and you can't get it back," she says.

As ethics consultant for the Personal Genome Project, which aims to collect and publish genomic data for 100,000 people, Lunshof is putting this into practice. Participants are made aware that their data will be fully public and could be used for any kind of research, even something they might not approve of like biological weapons research.

This kind of open consent model is important in a clinical or research setting, because it may not be possible to explain all the possible ways genetic data could be used in the future, says Sandra Lee at Stanford University in California. But the rules of the game change when we introduce commercial genetics companies. "When somebody is a patient interacting with a physician, they are operating with a set of ethical expectations," she says. "When you shift that to the marketplace, those aren't in place. That's worrisome."

Should governments gather DNA?

Some US states are trying to limit the use of genetic data, while others want to amass large databases.

In Arizona, a bill to create a statewide DNA database elicited criticism when it was announced on 19 February. It required people who work or volunteer for the state to submit DNA, along with anyone applying to serve as a foster parent or get a driving licence.

The bill may have been a response to a recent criminal case where a healthcare facility worker was traced through his DNA and charged with impregnating a patient who was incapacitated. Following a backlash, the proposal has since been amended to limit DNA collection to healthcare workers who directly care for patients.

In Maryland, a bill aims to stop police from searching public genetic databases to hunt down criminals, seen by some as an invasion of the privacy guaranteed by the US constitution. Law enforcement lobbyists argue that limiting police capabilities won't benefit the public, particularly because it isn't illegal for a citizen to take a DNA sample found at a crime scene and upload it to a genetic genealogy database, says Yaniv Erlich at genetic ancestry company MyHeritage. "To say that police cannot do something an ordinary citizen can do is unusual."