

TIME

Millions of Americans Could Be Identified Using Consumer Genetic Databases—Even If They've Never Taken a DNA Test



Jose A. Bernat Bacete—Getty Images

By **JAMIE DUCHARME** Updated: October 13, 2018 10:38 AM ET | Originally published: October 12, 2018

TIME
Health

For more, visit [TIME Health](#).

Up to 60% of Americans of European descent could be identified using only a DNA sample, some basic personal information and consumer genetic databases—even if they’ve never done a genetic test themselves, **according to a new study published in *Science***.

Some **7 million Americans** have taken a direct-to-consumer DNA test, many of which promise insights about ancestry, wellness and **likelihood of developing certain chronic diseases** in exchange for a sample of spit. The proliferation of these tests has also given rise to crowdsourced online databases that allow people to anonymously upload their results for further analysis and ancestry research.

One of these sites, GEDmatch, **proved instrumental** in arresting the suspected Golden State Killer in April, **sparking questions about genetic privacy**. Those questions bubbled back up in July, **when consumer testing giant 23andMe announced** that it had sold a stake of its business to pharmaceutical company GlaxoSmithKline, so that GSK could use anonymized data in drug development.

In the new *Science* study, researchers wanted to see just how anonymous genetic data really is. Not very, they found: If investigators were presented with a random DNA sample from an American of European ancestry, they could in roughly 60% of cases use consumer genetic databases to find a third cousin or closer blood relative match. (Since fewer Americans of other heritages have participated in consumer genetic testing, it would be harder to identify these individuals.) Zeroing in on that relative match makes it significantly easier to identify an individual — and with some basic personal information, the search gets even simpler, the study says.

Narrowing the list of individuals to those within a 100-mile radius would exclude 57% of the possibilities, the study says. Estimating the target’s age, plus or minus five years, would cross off 91% of the remaining pool. And inferring the person’s biological sex would leave only 16 or 17 individuals, the

authors estimate — a short enough list, theoretically, to investigate them individually.

“This is the first time we’ve had some sort of thoughtful quantification for how easy it is to track any individual, whether they participated in these databases or not, through the people who have participated in these databases,” says Dr. Robert Green, a medical geneticist at Brigham and Women’s Hospital and a professor at Harvard Medical School, who was not involved in the research. “I think that’s something everybody should understand be aware of.”

The prospect of being tracked may seem scary. After all, genetic databases are feasibly open to abuse, Green says. Sperm donors or biological relatives who wished to remain anonymous could be identified, for example, or an individual’s medical privacy could be violated by exposing his predispositions to certain diseases. But Green says these risks need to be weighed against the significant benefits of genetic tracking.

“All of the crimes that are currently unsolved, which have DNA evidence, there’s now a pathway to trying to locate these perpetrators,” Green says, pointing to the **backlog of unanalyzed rape kits** currently sitting in storage across the country. “I would like to see it interpreted through this prism. We could do a huge amount of social good that way.”

Another new study, published Thursday in the journal *Cell*, adds to the debate. This paper looked at the possibility of identifying a person’s relatives by comparing national forensic databases, which include DNA samples from criminal offenders, with consumer genetic databases. The two typically rely on different genetic markers to identify an individual — but even still, the *Cell* paper found that around 30% of people in the forensic database could be linked to a parent or child in a consumer database, and about 35% could be linked to a sibling.

That possibility could make it even easier for law enforcement officials to track down suspects, but it also underscores how identifiable people are by their genetic information. Plus, the new study suggests that the DNA samples in the national forensic database are **more revealing than they are often described**.

A spokesperson for 23andMe, the biggest player in the consumer DNA testing space, emphasized that its private databases are secure, though the same can't necessarily be said for third-party sites.

“23andMe implements strong security and privacy protocols to protect our customers' data,” the spokesperson said in a statement. “However, 23andMe cannot guarantee the same level of protections on third party sites — for example, if a customer downloads their 23andMe raw data and uploads it to another website. Our Help Center, Privacy Statement, and in-product experience all advise customers that third parties may use personal information differently than we do under our Privacy Statement. We recommend that customers make such choices carefully and review the privacy statements of all other third parties involved.”

It's important that consumers understand these scenarios before they spit in a tube, but Green emphasizes that the issues they raise aren't so different from many realities of modern life.

“It is a legitimate concern for people who are privacy-minded,” Green says. “But I really think this needs to be put in the context of all the other tradeoffs we make in our daily life, like putting our credit card online for a purchase, or being on Facebook or other social media, where information we provide can be scraped and diced and analyzed. At the present time, I think the potential benefit, in terms of criminals who could be captured, far outweighs the vulnerability, in terms of privacy intrusions.”

Even one of the most common issues raised in the genetic privacy debate —

that family members who didn't choose to take a DNA test are pulled into the fray — has a social media parallel, Green says.

“The person who posts a lot of Facebook pictures generally posts them with other people,” he says. “As soon as facial recognition software gets better and better, any photograph is going to tell people who's in that photo.”

Of course, there are plenty of people who are equally uncomfortable with these privacy issues, particularly following **Facebook's Cambridge Analytica scandal** this year. But modern society has largely accepted these risks in continuing to weave social media, the Internet and now consumer genetics tests into everyday life.

According to the *Science* study, it would be possible to identify nearly anyone in a certain group using a genetic database that includes information about 2% of that population. For Americans of European descent, that's around 3 million people. And since **DNA test kits are only getting more popular**, the study says, that number is “foreseeable for some 3rd party websites in the near future.”

Write to Jamie Ducharme at jamie.ducharme@time.com.