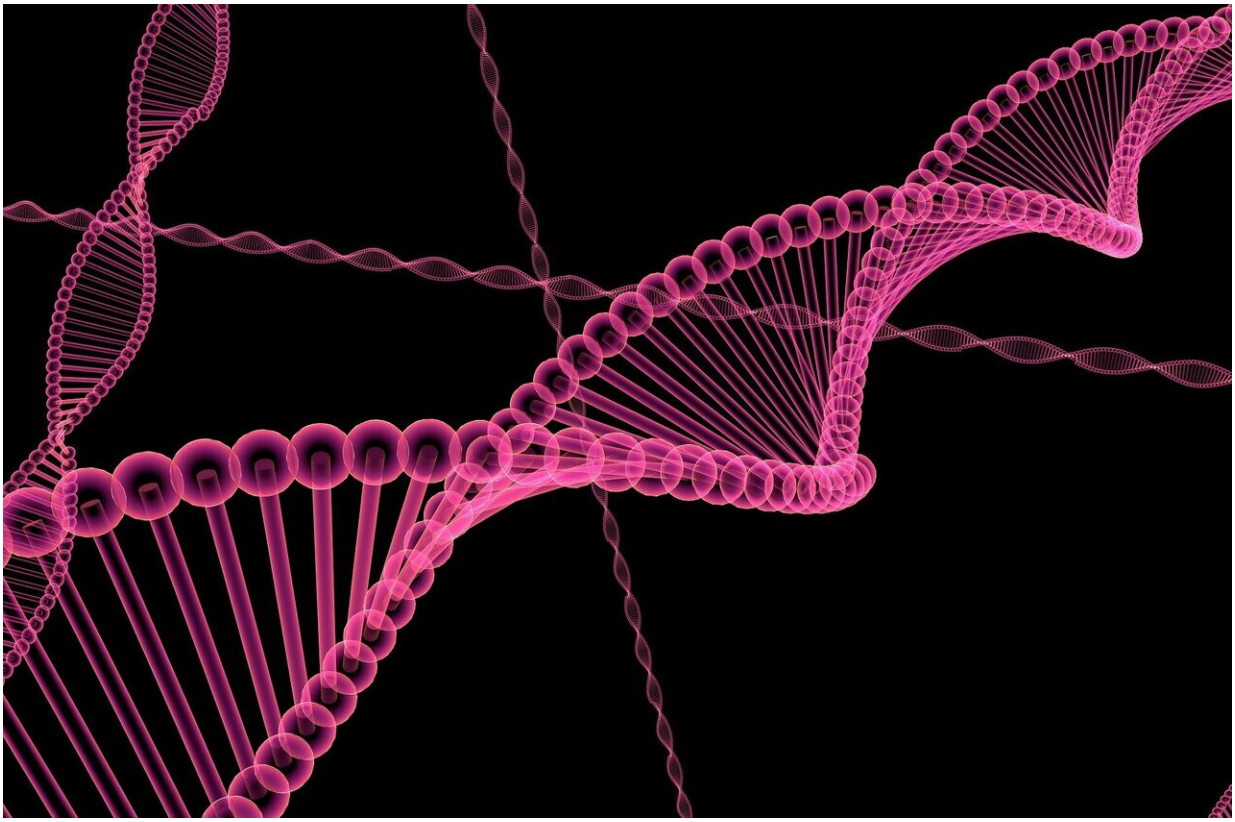
MedicalXpress

Hobbyist DNA services may be open to genetic hacking

January 7 2020, by Andy Fell



Credit: CC0 Public Domain

Online services that allow users to upload their genetic information, research genealogy and find lost relatives may be vulnerable to a sort of genetic hacking, according to two geneticists at the University of

California, Davis. A paper describing the work is published Jan. 7 in the journal *eLife*.

With the growth of home DNA testing, online services such as GEDMatch, MyHeritage and FamilyTreeDNA have become popular places for people to upload their genetic information and research their genealogy. They have also been used by law enforcement to find criminal suspects through a DNA match with relatives.

But according to Professor Graham Coop and postdoctoral researcher Michael 'Doc' Edge at the UC Davis Department of Evolution and Ecology, someone with a bit of expertise in genetics and computing could design and upload DNA sequences that extract far more from these databases than some lost cousins. It may be possible for an attacker to pull out the genetic information of most people in a database or to identify people with specific genetic traits such as Alzheimer's Disease.

Coop and Edge notified the database companies of the problem in July, 2019 to allow them time to put countermeasures in place before publishing a preprint in October.

"People are giving up more information than they think they are," when they upload to these publicly accessible sites, Coop said. And unlike credit card information, you can't just cancel your old genome and get a new one.

The problems do not affect for-profit DNA sequencing companies such as 23andMe, Coop said. You have to submit your DNA as a saliva sample to get access to their genetic data. The public databases, however, allow anyone to upload DNA sequences and search for other users with matching sequences.

Identical by state and descent

These sites work by using software to compare DNA sequences uploaded by users with sequences already in their database. Your genome is a mosaic of pieces inherited from your ancestors. Bigger pieces, or tiles in the mosaic, come from recent ancestors. As generations pass, matching sequences get chopped into smaller pieces. So if you share large chunks of DNA sequence with someone else, it's likely you share a recent ancestor.

Coop and Edge found three approaches to attacking these databases. They call these methods IBS (identical by sequence) tiling, IBS probing and IBS baiting.

Their tests primarily used a public collection of human DNA sequences available for research. They also carried out a proof of concept test in the GEDMatch database but without interacting with other users' DNA data.

In IBS tiling, an attacker uploads several genomes found in public research databases and keeps track of which ones match with other genomes in the database, and where. If they can find enough matching tiles, they can put together most of someone's genome.

IBS probing can be used to hunt for people who carry a specific genetic variant—for example, a gene tied to Alzheimer's disease. To do this, the attacker creates a fake genome with a DNA sequence that isn't likely to match anyone, except for one small section that will match the gene of interest. Matches from the database are likely to be people with this genetic variant.

Finally, IBS baiting relies on tricking one class of algorithms used to identify relatives. (Not all databases use this type of algorithm, though). Coop and Edge calculate that with as few as 100 uploaded DNA sequences, an attacker could use this method to obtain most of the

genomic information in a database.

Coop and Edge carried out a proof-of-concept test with the GEDMatch database in December 2019. Working with only with DNA sequences they had uploaded and using GEDMatch's 'research mode' so as not to interact with other users' data, they showed that IBS baiting could be used to identify specific genetic variants (single nucleotide polymorphisms, or SNPs) in the database.

All three attacks could be carried out by someone with knowledge of genetics and computing, such as a graduate student or serious hobbyist, but "the good news is that it's quite preventable," Edge said.

Coop and Edge's paper sets out a series of steps direct-to-consumer genetics services could take to block these attacks. While they have already shared the information with the leading services, they have had a 'varied' response, Coop said.

Using these services necessarily involves giving up personal information, and millions of people seem willing to do that in exchange for researching family history or other personal uses. But users should be more aware of exactly how much information they might be giving up when they access these services.

"We would like (the services) to clarify their vulnerabilities and how they're addressing them," Coop said.

More information: Michael D Edge et al, Attacks on genetic privacy via uploads to genealogical databases, *eLife* (2020). DOI: 10.7554/eLife.51810

Provided by UC Davis