DNA for the greater good: Should the police have access to consumer DNA databases?

In the spring of 2018, the capture of the Golden State Killer using a consumer DNA database catapulted the issue of genetic privacy into the headlines. A year later, a second case has pushed genetic privacy to the precipice of a slippery slope as the mothership of DNA databases involved in both cases, GEDmatch, has changed its Terms of Service to give users more control over accessibility of their data to law enforcement.

But will increased privacy control slow the momentum in using DNA to catch criminals? The new forensic technology is cracking a case a week now, turning cold cases red hot.

Joseph James DeAngelo, the alleged Golden State Killer. Image: AP

The FBI works with genetic genealogists at Parabon NanoLabs, which for many cases uses GEDmatch, which is free. These experts combine DNA information with traditional resources like historical accounts, diaries, and census data to identify individuals.

The Utah case

This spring's flashpoint centered on use of GEDmatch to break an assault case in Utah. Up until then, the 55 crimes that Parabon Nanolabs had solved using DNA data had all been sexual assault or murder.

GEDmatch, the brainchild of retired businessman Curtis Rogers and transportation engineer John Olson, began in 2010 as a place for people to upload their DNA data from the testing companies, such as AncestryDNA, 23andMe, FamilyTreeDNA and MyHeritage. Early users were mostly adoptees searching for biological relatives or people just intrigued with fleshing out their family trees.

As a clearinghouse of sorts, GEDmatch breaks down the silos of proprietary company databases. That's why I use it: 23andMe and AncestryDNA give me different, if overlapping, lists of relatives.

Use by law enforcement, however, is not what the GEDmatch founders intended. But Olson and Rodgers reportedly felt the recent assault case in Utah was so heinous that they wanted to help before the attacker harmed others. So they opened access to GEDmatch to investigators, without alerting users.

Oops.

No one is minimizing the violence of the crime.

On November 17, 2018, a 71-year-old woman was practicing the organ alone in a church in Centerville, Utah, when someone threw a rock through a window, apparently cutting himself while jumping in, then attacked and strangled her. She survived, but with injuries.

A police detective sent blood on the rock to the FBI, but investigators found no hits in the forensic DNA database. So an FBI detective contacted Parabon Nanolabs, which initially refused to take the case, citing the history of use only for rape or murder.

Image axd

Image not found or type unknown

Parabon's chief genetic genealogist CeCe Moore (who is also on PBS's Finding Your Roots) and her team zeroed in on a great uncle of the suspect, using GEDmatch. His 17-year-old grand-nephew lived in Centerville, the scene of the crime. On April 11, DNA extracted from a milk container and juice box that the teen had tossed matched the blood on the rock from the crime scene.

Bingo.

Repercussions

The social media uproar over the use of DNA data to solve the Utah case prompted changes in GEDmatch's Terms of Service, the changes helpfully in red on the website. Law enforcement can now use DNA data "to identify a perpetrator of a violent crime against another individual, where 'violent crime' is defined as murder, nonnegligent manslaughter, aggravated rape, robbery, or aggravated assault" and "to identify remains." *Without permission.*

Users must now go to the GEDmatch website for updates – they'll no longer be emailed, because that's a way for info to get out.

The change in GEDmatch's Terms of Service instantly opted-out 1.25 million DNA "kits." Users must now opt-in to allow their information to be compared to any entries in the GEDmatch database, including for forensic purposes.

But Blaine Bettinger writes at the *The Genetic Genealogist* that opting out isn't enough protection for a

GEDmatch user, because law enforcement can still use DNA data to search for relatives, which isn't the same as directly comparing it to a suspect's sample. Instead, Bettinger urges, a concerned user should delete GEDmatch data or designate it for research – or for newbies, not upload it in the first place. (GEDmatch converts data to a form accessible to software but it's not a foolproof encryption.)

Bettinger nails the underlying issue: informed consent versus the ends-justifies-the-means mindset.

gina

Image not found or type unknown

I opted in. My non-scientific perusal of postings on a large, closed *Facebook* group for GEDmatch users revealed that most support making their data available to law enforcement. But so far only about 20,000 of the 1.2 million GEDmatch users have opted in, Rogers told CNN.

The most common reason cited by the objectors was that GEDmatch violated their original Terms of Service. True. Some people imagined scenarios: denial of health insurance based on disease susceptibilities (although the nature of the DNA data, 700,000 or so SNPs, doesn't reveal traits or actionable medical information), or an uptick in crooks intentionally contaminating crime scenes with other DNA to divert investigators. But anyone who watches *Law and Order* knows that already happens.

One person alarmed at the changed Terms of Service is CeCe Moore, whose name is synonymous with genetic genealogy.

In the wake of the Utah case that led to rewriting the script at GEDmatch, Moore told CNN that "people will die" as a result of users yanking their data, insisting that she wasn't kidding. With fewer distant cousins to triangulate back to a suspect who can then be matched to crime scene evidence, forensic genetic genealogy will be "crippled," she said.

FamilyTreeDNA spilled first

The kerfuffle at GEDmatch is déjà vu all over again for consumers who tested with FamilyTreeDNA, which isn't free, like GEDmatch. On January 31, *Buzzfeed News* broke the story of FamilyTree providing access to their consumer DNA database to the FBI, to help solve a few cold cases.

Customers weren't happy with the unannounced breach in privacy.

Company president Bennett Greenspan emailed them: "I am genuinely sorry for not having handled our communications with you as we should have. We've received an incredible amount of support from those of you who believe this is an opportunity for honest, law-abiding citizens to help catch bad guys and bring closure to devastated families."

FamilyTreeDNA users could have disabled the “matching” function to block FBI access to their data, if they somehow knew about the change before it hit the media, by checking the updated Terms of Service. They weren’t otherwise notified.

elephant head on

Image not found or type unknown
Image: Shutterstock

The elephant in the room – all of us

You don’t need to have spit in a tube and or have your DNA markers in a cloud for law enforcement to find you. Genetic genealogists may start with shared DNA hunks (explained here), but they use other types of information to flesh out and ultimately reveal identities, even of those who’ve never tested.

Anyone who’s used a DNA ancestry site to try to find relatives knows that not everyone who is outed was actually tested. For example, I discovered a few months ago that I have six half-siblings, thanks to a sperm donor I knew nothing about. But two of them never had their DNA tested and don’t know that the rest of us exist. Their adult children matched to our insta-family, but don’t want their parents to know.

My new half-niece did some traditional genealogical sleuthing to identify two branches of an extended family wherein lies my mystery sperm donor. The members of one look a lot like the 8 of us siblings, thanks to glimpses on *Facebook* (enough of us friended each other to see, and anyway you can see photos without friending someone).

The piecing together of my family from second and third cousins is somewhat like what law enforcement did to track the Golden State Killer. And GEDmatch provides a nice convenient list of cousins in descending order of degree of relatedness, with email addresses. Many people, however, use pseudonyms.

Where is all of this headed?

At the rate that people are submitting DNA for testing, someday we'll all be identifiable by our genetic material. A recent study published in *Science* from Yaniv Erlich from MyHeritage and colleagues deduced that 60 percent of US citizens of European ancestry can already be outed without having taken a DNAtest, thanks to all those third cousins who did.

And it's getting hard to keep up with the swelling databases.

An estimate from *The DNA Geek* circa October 2018 is already dated. More recent figures indicate, in addition to the 1.2 million GEDmatch users, AncestryDNA recently topped 15 million and 23andMe exceeds 10 million. FamilyTreeDNA claims two million users and MyHeritage 2.5 million.

c b d a b f cd ed

Image not found or type unknown

By early 2019, more than 26 million people had taken DNA ancestry tests, according to *MIT Tech Review*. That number is projected to reach more than 100 million within two years. And while the pool is white Euro-centric for now, that's changing.

The companies are diversifying their reference populations by offering free kits to researchers working with underrepresented groups. The dynamic nature of the reference populations is why some users are frustrated to log in and suddenly find their place in the world altered. Identifying family members uses different calculations, but I think with time these will come to embrace more ethnicities too.

Crime severity is an important factor in consumers' evaluation of DNA data transparency, according to a study in *PLOS Biology* from Amy L. McGuire, of the center for medical ethics and health policy at Baylor College of Medicine. Of 1,587 respondents 80 percent supported use for violent crime, 79 percent if against children, 77 percent if a person is missing, but only 39 percent in cases of nonviolent crimes.

From social media responses to the FamilyTreeDNA and GEDmatch breaches of privacy, it's clear that the issue of genetic privacy for consumer DNA information to serve the greater good is igniting.

Summarizes Debbie Kennett, from the department of genetics, evolution and environment at University College, London, in “Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes,“ just published in *Forensic Science International*, “There is an urgent need for forensic scientists, bioethicists, law enforcement agencies, genetic genealogists and other interested parties to work together to produce international guidelines and policies to ensure that the techniques are used responsibly and effectively.”

Stay tuned.

Ricki Lewis is the GLP’s senior contributing writer focusing on gene therapy and gene editing. She has a PhD in genetics and is a genetic counselor, science writer and author of The Forever Fix: Gene Therapy and the Boy Who Saved It, the only popular book about gene therapy. BIO. Follow her at her website or Twitter @rickilewis