Forensic Genetic Genealogy Searches

A PRIMER FOR DEFENSE ATTORNEYS AND POLICYMAKERS

ELECTRONIC FRONTIER FOUNDATION

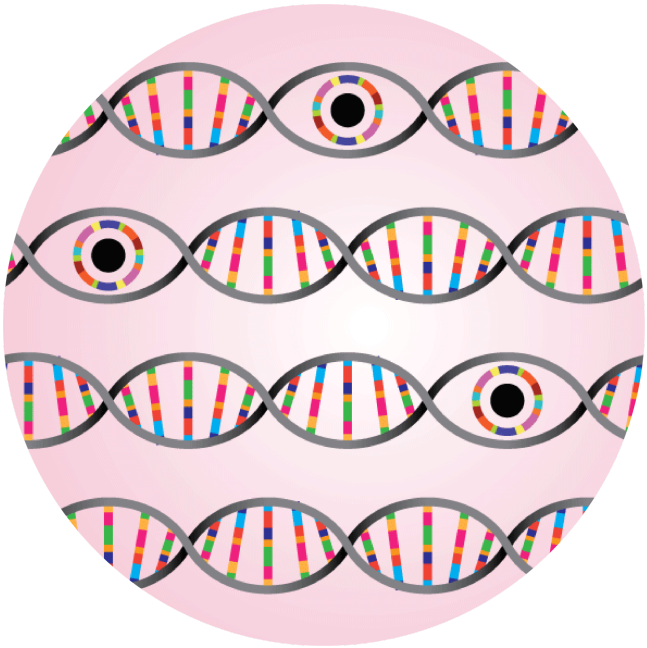
Author: Jennifer Lynch
Jennifer Lynch is the Surveillance Litigation Director at the Electronic Frontier Foundation. She leads EFF's legal work challenging government abuse of search and seizure technologies through the courts by filing lawsuits and amicus briefs on important issues at the intersection of technology and privacy.

Jennifer Lynch
Electronic Frontier Foundation
San Francisco, California
415-436-9333
Email: jlynch@eff.org
Website: www.eff.org/join
Twitter: @lynch_jen

A publication of the Electronic Frontier Foundation, 2023.

An earlier version of this article was published as "Forensic Genetic Genealogy Searches: What Defense Attorneys Need to Know" in *The Champion* (Nov. 2020), available online to members of the National Association of Criminal Defense Attorneys (NACDL).

View this report online:
https://www.eff.org/wp/forensic-genetic-genealogy-searches-what-defense-attorneys-need-know



Forensic Genetic Genealogy Searches

A PRIMER FOR DEFENSE ATTORNEYS AND POLICYMAKERS

JENNIFER LYNCH
Surveillance Litigation Director

July 2023

ELECTRONIC FRONTIER FOUNDATION

Contents

Introduction

A DNA sample contains a person's entire genetic makeup. This genetic information is deeply private. It can reveal intensely sensitive information about us, including our propensities for certain medical conditions, our ancestry, and our biological familial relationships. Raising the specter of eugenics and other discredited theories, some researchers have theorized that genetics could help explain human behaviors such as aggression, addiction, criminal tendencies, and even political views.[1] Private companies claim they can use our DNA to predict whether we are introverted or extroverted, averse to cilantro, excel at running, or are afraid of public speaking.[2] One company, regularly used by law enforcement in cold case investigations, even claims it can predict a person's physical facial appearance from their DNA, including "skin color, eye color, hair color, freckles, ancestry and face shape."[3]

DNA has been used in criminal cases since the late 1980s,[4] but DNA technology and research have advanced significantly since then. Where once, a useful forensic sample could only be obtained from blood, semen, or other bodily fluids, today, forensic investigators can detect, collect, and analyze trace amounts of DNA from objects merely touched by a person. DNA collection is now mandatory from those convicted of or arrested for many crimes, and the national CODIS DNA database, maintained by the FBI, contains nearly 16 million offender profiles and 5 million arrestee profiles.[5]

Despite these technological advances, some crimes continue to have no suspects. In an increasing number of these cases, police are turning to a relatively new investigative technique: forensic genetic genealogy (FGG). Through FGG, police access consumer-facing genetic genealogy websites to try to identify the source of crime scene DNA samples. These consumer sites allow people to upload their own genetic

---

[1] Erika Check Hayden, *Ethics: Taboo Genetics*, Nature (Oct. 2, 2013), http://www.nature.com/news/ethics-taboo-genetics-1.13858; Ricki Lewis, *Can DNA predict who might be a mass murderer?*, Genetic Literacy Project (Jan. 31, 2020), https://geneticliteracyproject.org/2020/01/31/can-dna-predict-who-might-be-a-mass-murderer/; Marta Zaraska, *The Genes of Left and Right*, Scientific American (May 1, 2016), https://www.scientificamerican.com/article/the-genes-of-left-and-right/; Sophie von Stumm & Robert Plomin, *Using DNA to predict intelligence*, Intelligence, Volume 86 (May–June 2021) https://doi.org/10.1016/j.intell.2021.101530.

[2] *Ancestry Launches a New Take on Genetic Traits*, Ancestry.com (Nov. 9, 2018), https://www.ancestry.com/corporate/blog/ancestry-launches-a-new-take-on-genetic-traits; AncestryDNA® Launches Introvert/Extrovert Traits Report, Ancestry.com (Mar. 10, 2022); https://www.ancestry.com/corporate/blog/ancestrydna-launches-introvertextrovert-traits-report; *Let's talk about Fear of Public Speaking & Genetics*, 23andMe, https://www.23andme.com/topics/traits/fear-of-public-speaking/ (last visited Oct. 26, 2022).

[3] Caitlin Curtis & James Hereward, *How Accurately Can Scientists Reconstruct A Person's Face From DNA?*, Smithsonian Magazine (May 4, 2018),https://www.smithsonianmag.com/innovation/how-accurately-can-scientists-reconstruct-persons-face-from-dna-180968951.

[4]https://www.theguardian.com/uk-news/2016/jun/07/killer-dna-evidence-genetic-profiling-criminal-investigation; https://www.smithsonianmag.com/smart-news/july-marks-the-25th-anniversary-of-the-first-use-of-dna-evidence-to-convict-a-killer-10509877/

[5] *See CODIS-NDIS Statistics*, FBI, https://le.fbi.gov/science-and-lab-resources/biometrics-and-fingerprints/codis/codis-ndis-statistics.

information and use the sites' proprietary algorithms to search through other users' data to identify long lost relatives and build a family tree. Several of these sites also allow police to search through the same data to try to find suspects. FGG has been gaining interest after one site, GEDmatch, was used to find the "Golden State Killer," a man responsible for a series of brutal rapes and murders that plagued California in the 1970s and 80s.[6] By the end of 2021, FGG had been used in well over 500 criminal cases.[7]

The Electronic Frontier Foundation, has worked on several cases where law enforcement used FGG, and this article is based on that work. This article will discuss how genetic genealogy works, law enforcement's use of FGG, and how FGG can both misidentify suspects and, in rare cases, help to clear wrongful prosecutions and exonerate wrongfully-convicted individuals. It will also discuss new statutory restrictions on the use of FGG and what policymakers need to know to place appropriate guardrails on these searches. Finally, it will discuss what defense attorneys should look for if FGG is used in their cases and how to challenge FGG as an unconstitutional search.

I. How Genetic Genealogy Searches Work

Genetic genealogy sites are run by private companies and offer to help people find long-lost relatives, learn more about their families and ancestors, and identify their own traits and health predispositions. There are two main types of consumer genetic databases—closed databases like Ancestry and 23andMe, where the company controls and can limit search results and direct access to other users' data, and open databases like GEDmatch, FamilyTreeDNA, and MyHeritage, which offer consumers much broader access and allow users to search their own genetic data against genetic information submitted by all other site users. As both Ancestry and 23andMe require warrants and limit law enforcement access, this article will focus on open databases.[8]

To use one of the open database sites, users provide extensive genetic data, either as a biological sample or an electronic file containing their raw genotyped DNA data, which the company converts into a genetic genealogy profile. The genetic genealogy site then uses a proprietary algorithm to compare that uploaded data to other users' data on the site, identifying relationships and estimating how close or distant those relationships may be (*e.g.*, a direct connection, like a parent, or a very distant connection, like a fifth cousin).[9] The sites measure these connections in "centimorgans" (cM), which are units

[6] https://www.nytimes.com/2018/04/26/us/golden-state-killer.html
[7] Michelle Taylor, *How Many Cases Have Been Solved with Forensic Genetic Genealogy?*, Forensic (March 3, 2023) https://www.forensicmag.com/594940-How-Many-Cases-Have-Been-Solved-with-Forensic-Genetic-Genealogy/; Dowdeswell, Tracey, *Forensic Genetic Genealogy Project* v. 2022, Mendeley Data, V1, (Feb. 22, 2023) doi: 10.17632/jcycgvhm96.1 https://data.mendeley.com/datasets/jcycgvhm96.
[8] Ancestry has pushed back on law enforcement requests to access their entire database. *See, e.g.,* Peter Aldhous, *A Court Tried To Force Ancestry.com To Open Up Its DNA Database To Police. The Company Said No*, Buzzfeed News (Feb. 3, 2020) https://www.buzzfeednews.com/article/peteraldhous/ancestry-dna-database-search-warrant.
[9] These identifications are possible because we share DNA with our biological relatives at predictable locations on our genome at somewhat predictable rates. People who are descended from the same ancestor, such as a great-great grandparent, will all have some of that ancestor's DNA. Because of this, even distant cousins who do not know each other will have some of the same DNA as one another.

of measurement designed to predict "genetic linkage" or "genetic distance" by measuring the probability that a section of DNA will be passed on to a descendant intact rather than split into separate segments.[10] Because we share longer stretches of DNA with closer biological relatives, a higher centimorgan number will, in general, mean two people are more closely related.

DNA recombination is random, however, so the amount of shared DNA can vary greatly, even for relatives of the same degree, like siblings. For example, parents and their children or siblings may share anywhere from 2,000-3,600 cMs, while first cousins may share anywhere from 425-1500 cMs.[11] Variation increases as the biological relationship becomes more distant (and therefore more recombination has occurred), such that about 10% of third cousins and 50% of fourth cousins share no detectable matching segments.[12] There are other complicating factors. For example, the amount of relatedness between two DNA samples does not necessarily, without more information, tell what that relationship is. Individuals that share 1,500 cMs could be first cousins, half-siblings, grandparent / grandchild, half uncles, great grandparent / great grandchild, or great uncles.[13] And for some populations where, for example, intermarriage has been common, there may be low genetic diversity, and even very distant relatives can share a significant amount of DNA[14]

Although sites like GEDmatch state they do not disclose a person's raw DNA to other users, they do allow users to see where, along each chromosome, their DNA may be similar to another user. Also, when users perform a one-to-many search on GEDmatch, each "match" includes the individual's name or alias, email address, and any family tree or haplogroup information they have chosen to share.[15]

By 2020, GEDmatch had about 1.45 million users.[16] While this may not seem like a lot of people compared to the U.S. population as a whole (it's the equivalent of about 0.5% of the U.S. adult population), research showed that 60% of white Americans could already be identified just from this small number of users.[17] This same research showed that once the number of GEDmatch's users reaches just 2% of the U.S. population, 90% of white Americans would be identifiable. And, in fact, a 2022 partnership between

[10] https://www.genome.gov/genetics-glossary/Centimorgan; https://support.ancestry.com/s/article/Measuring-Relatedness
[11] Ellen Greytac et al, *Genetic genealogy for cold case and active investigations*, Forensic Science International 299:103-113 (March 2019) https://www.researchgate.net/publication/332047092_Genetic_genealogy_for_cold_case_and_active_investigations.
[12] Id. (Third cousins share a great-great grandparent, while fourth cousins may only share a great-great-great grandparent).
[13] Greytac et al, *Genetic genealogy for cold case and active investigations*.
[14] Another complicating factor is that "biological family trees often do not match documented family trees due to misattributed paternity, unrecorded adoption, unknown parentage, etc., individuals in these situations are overrepresented in genetic genealogy databases" (Greytac)
[15] Haplogroups are different segments of DNA that are closely linked and tend to be inherited together. Genetic genealogy sites look at specific segments of DNA to predict which part of the world one's ancestors came from. Greytac et al, *Genetic genealogy for cold case and active investigations*.
[16]https://inhomelandsecurity.com/whats-the-fallout-from-the-gedmatch-genealogy-database-breach/.
[17] Jocelyn Kaiser, *We will find you: DNA search used to nab Golden State Killer can home in on about 60% of white Americans*, Science (Oct. 11, 2018), https://www.science.org/content/article/we-will-find-you-dna-search-used-nab-golden-state-killer-can-home-about-60-white (unlike criminal databases, genetic genealogy databases skew white).

GEDmatch and FamilyTreeDNA now allows law enforcement to "identify 90-95% of people to third cousin or closer or 60% of people to second cousin or closer," according to GEDmatch's CEO.[18]

II. Law Enforcement Use of Genetic Genealogy Services

A. Comparing Genetic Genealogy Searches with Traditional Criminal DNA Searches

Genetic genealogy searches differ from traditional criminal DNA searches in several respects. First, unlike CODIS, genetic genealogy databases are privately maintained and contain DNA data submitted by consumers who are not in any way connected to the criminal justice system. There are no laws that specifically regulate how these sites collect, handle, and share data. In contrast, DNA data entered into CODIS is subject to state and federal statutes that govern exactly how, when, and from whom it can be collected. Where DNA is collected upon arrest, there must be clearly-defined procedures for expunging DNA if a person is ultimately never charged or convicted. And even the labs that process DNA entered into CODIS are subject to federal regulations and licensing requirements.

Second, genetic genealogy requires significantly more—and more revealing—genetic data than the limited information collected and entered into CODIS. CODIS profiles typically consist of one or two alleles at each of the 13 to 20 loci that are part of CODIS's "Core Loci."[19] These short tandem repeat ("STR")[20] DNA markers are taken from "non-coding"[21] parts of the human genome. In contrast, genetic genealogy profiles are made up of more than half a million single nucleotide polymorphisms ("SNPs")[22] that span the entirety of the human genome. SNP profiles can reveal family members and distant ancestors as well as predict a person's propensity for various diseases like breast cancer and traits like addiction and drug response. Like STR DNA, SNPs can be inherited

[18] *The Democratization of Forensic Genetic Genealogy*, Forensic Magazine, at 9:00 (Aug. 22, 2022), https://soundcloud.com/forensic-magazine/the-democratization-of-forensic-genetic-genealogy.
[19] *See Frequently Asked Questions on CODIS and NDIS: 19. What are the CODIS core loci?*, FBI, https://www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet (noting CODIS profiles originally contained 13 DNA loci, but were increased to 20 in 2017).
[20] STR DNA typing is a widely-used forensic DNA technology that examines specific genetic locations on the non-sex chromosomes. A 'marker' is a genetic locus, or location.
[21] "Non-coding" means the DNA "does not provide instructions for making proteins." *What is noncoding DNA?*, National Library of Medicine - MedlinePlus, https://medlineplus.gov/genetics/understanding/basics/noncodingdna/. While scientists and police once believed this meant that the CODIS loci were nothing more than "junk DNA," researchers now know that it is possible to infer some medical conditions and phenotypic traits from CODIS alleles. See, e.g., Nicole Wyner, et al., *Forensic Autosomal Short Tandem Repeats and Their Potential Association With Phenotype*, Frontiers in Genetics (Aug. 6, 2020), https://www.frontiersin.org/articles/10.3389/fgene.2020.00884/full.
[22] SNPs are single-letter variants at a specific location in a section or sequence of a person's genome. They can influence health, disease, and physical traits and can indicate a person's ancestry, among other things. *Single-nucleotide polymorphisms (SNPs)*, Nat'l Hum. Genome Rsch. Inst., Nat'l Insts. Of Health (Oct. 20, 2022), https://www.genome.gov/genetics-glossary/Single-Nucleotide-Polymorphisms.

from and passed on by both males and females. This means that both CODIS and genetic genealogy databases can identify some biological familial relationships. However, given the extensive genetic information contained in SNP profiles, genetic genealogy sites are able to predict indirect and even very distant familial relationships. Currently, the CODIS loci, when combined with Y chromosome DNA ("Y-STR"),[23] can only identify father-son relationships with any kind of accuracy. Any attempted identification of more distant relatedness—even among siblings—is unreliable.[24]

A final significant difference between CODIS and genetic genealogy searches is in how each is generally intended to be used. CODIS searches are designed to identify the person behind the forensic sample, so the results of a search are generally one or zero positive matches.[25] In contrast, genetic genealogy searches are designed, not to identify a single individual, but to find as many of that person's potential biological relatives as possible. For this reason, initial search results could include hundreds or even thousands of people.

B. How Forensic Genetic Genealogy (FGG) Searches Work

Law enforcement FGG investigations tend to follow a typical pattern. Police investigators contract with an individual, frequently called a "genetic genealogist" (although there is currently no standardized training for this position), or a private company to conduct an FGG search. This individual or company may be employed by or independent from the genetic genealogy service hosting the database of users' DNA data.[26] The investigator provides the genetic genealogist either with a sample of DNA found at the crime scene or with data generated from that sample. The genetic genealogist uploads the DNA data, either under a pseudonym or via a private law enforcement account, to one or more of the genetic genealogy services like GEDmatch. The site's algorithm will identify biological relatives and predict how distant or close those relatives are to the forensic sample. This will result in a list of names and identifying information. Using this list, the genetic genealogist will conduct traditional genealogical research through public records, social media, and other sources to build a

[23] Y-STR is derived solely from the male sex-determining Y chromosome and can be used to identify patrilineal relationships and to identify the sex of an unknown DNA contributor. See, e.g., Manfred Kayser, *Forensic use of Y-chromosome DNA: a general overview*, 136 Human Genetics 621 (May 2017), https://pubmed.ncbi.nlm.nih.gov/28315050/. California, among other states, uses Y-STR data to confirm results of a CODIS-based familial search. S*ee Memorandum of Understanding: DOJ Familial Searching Protocol*, Cal. Dep't of the Att'y Gen., https://oag.ca.gov/sites/all/files/agweb/pdfs/bfs/fsc-mou-06072019.pdf (describing California's familial search requirements).

[24] Rori Rohlfs, et al., *The Influence of Relatives on the Efficiency and Error Rate of Familial Searching*, PLOS One (Aug. 14, 2013), http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0070495.

[25] Some states do try to use CODIS to identify partial matches. In that case, searches may produce several possible matches. FBI has stated that CODIS is not intended to be used for familial searching but does acknowledge partial match searching. *See* https://www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet; *see also* Natalie Ram, *Fortuity and Forensic Familial Identification*, 63 STAN. L. REV. 751 (2011) (questioning the distinction between "fortuitous" partial match searches and deliberate familial searching, which is also based on a partial match).

[26] The private company Parabon Nanolabs and two individual genetic genealogists, Barbara Rae Venter and Cece Moore, have been involved with the vast majority of FGG cases.

family tree and try to narrow down potential suspects. They will look for relatives who are around the same age as the perpetrator and who were likely to have been in the general location at the time the crime occurred. The genetic genealogist will compile a report that includes names of potential near and distant relatives of the forensic DNA sample with a prediction of relatedness. If the police have contracted with a company called Parabon Nanolabs for the FGG search, Parabon may also provide a computer-generated image that represents its prediction of what the suspect may look like based on their DNA.

Once officers have narrowed down their potential suspects, they then surveil those suspects until they can obtain a discarded item like a straw or beer bottle that may contain the individual's DNA. Officers collect a DNA sample from the item to determine if it matches the DNA found at the crime scene. If it matches, police will then seek a warrant to obtain a sample directly from the suspected individual. If that sample confirms the earlier match from the discarded item, police will make an arrest.

So far, FGG has mostly been reserved for egregious cold cases involving homicide, serious sexual assault, and multiple victims. However, this will likely change as more police departments learn of the technique, more vendors offer these services, and costs start to decrease. For example, in 2019, police used FGG to implicate a teenager in a sexual assault.[27] Without meaningful regulations on the use of FGG (discussed further below), it is possible in the near future that police will use FGG at earlier stages of investigations to generate leads, as well as for lower level crimes or even in questionable cases, such as to identify political protestors[28] or abortion seekers.[29]

III. Regulation is Minimal

Unlike CODIS and established state-run DNA databases, there are still few rules governing how and when FGG may be used. At the federal level, the Department of Justice maintains "interim" policy guidelines that describe conditions under which federal officers may conduct FGG searches.[30] For example, the policy states that agencies

[27] *DNA, Genealogy Research Lead to Teen's Arrest in Sex Assault*, Associate Press (Oct. 21, 2019) https://www.usnews.com/news/best-states/pennsylvania/articles/2019-10-21/dna-genealogy-research-lead-to-teens-arrest-in-sex-assault.

[28] ***DNA from cigarette leads to Dakota Access arrest 3 years on*, Associated Press (Sept. 6, 2019) https://apnews.com/abb444c2e6f14ca49a675e82d4b0d520 (DNA from a cigarette butt was used to identify and charge a man who protested the Dakota Access pipeline.).**

[29] FGG has already been used in several cases to prosecute mothers of abandoned newborns. *See, e.g.,* Jennifer Lynch, *Police May Not Need a Warrant to Rummage Through Your Trash, But Warrantless Collection of DNA Is Unconstitutional*, EFF.org (March 11, 2020) https://www.eff.org/deeplinks/2020/03/police-may-not-need-warrant-rummage-through-your-trash-warrantless-collection-0; *How DNA Helped Meriden Police Solve a 32-Year-Old Cold Case*. NBC Connecticut (Jan. 14, 2020) https://www.nbcconnecticut.com/news/local/meriden-police-to-provide-update-on-32-year-old-investigation-into-death-of-infant/2209179/.

[30] *Department of Justice Announces Interim Policy on Emerging Method to Generate Leads for Unsolved Violent Crimes*, DOJ (Sept. 24, 2019) https://www.justice.gov/opa/pr/department-justice-announces-interim-policy-emerging-method-generate-leads-unsolved-violent; *United States Department of Justice Interim Policy Forensic Genetic Genealogical DNA Analysis and Searching* (hereinafter "Interim Policy"), DOJ (Sept. 2, 2019)

may consider using FGG when investigating "unsolved violent crime" and that it is not to be used as the sole basis for arrest.[31] However, these guidelines have yet to be formalized, lack the force of law, and can be changed at any time.[32] Further, while they describe some layers of bureaucracy that federal officers must go through before the DOJ will authorize an FGG search, they do not require any legal process at all, much less a warrant or independent judicial oversight.[33]

Only two states—Maryland and Montana—have explicit statutory limitations on the use of FGG and familial searches.[34] Maryland's law,[35] passed in 2021, is the broader of the two. It requires judicial authorization for FGG and places strict limits on when and under what conditions law enforcement officers may conduct FGG searches. For example, it limits the use of FGG to cases of rape, murder, felony sexual offenses, and criminal acts that present "a substantial and ongoing threat to public safety or national security."[36] Before officers can pursue FGG, they must certify to the court that they have already tried searching existing, government-run criminal DNA databases like CODIS, that they have pursued other reasonable investigative leads, and that those searches have failed to identify anyone. And FGG may only be used with consumer databases that have provided explicit notice to users about law enforcement searches and sought consent from those users.

Maryland's law regulates other aspects of genetic investigations as well. For example, it places strict limits on and requires judicial oversight for the covert collection of DNA samples from both potential suspects and their genetic relatives. This is an important protection because, as noted above, officers almost always secretly collect and search DNA from free people after a FGG search.[37] Maryland's law also mandates informed consent in writing before officers can collect DNA samples from third parties (non-suspects like family members) and precludes covert collection from a third party who has refused to provide a sample. It requires destruction of DNA samples and data when an investigation ends. It also requires licensing for labs that conduct DNA sequencing used for FGG and for individuals who perform genetic genealogy. It creates criminal penalties for violating the statute and a private right of action with liquidated

---

https://www.justice.gov/olp/page/file/1204386/download (pdf). The Scientific Working Group on DNA Analysis Methods (SWGDAM), an organization chaired by an FBI appointee and made up of representatives from federal, state and local forensic DNA laboratories around the country, has also developed its own recommended policy guidelines for FGG searches. *See Overview of Investigative Genetic Genealogy*, 6, SWGDAM (Feb. 18, 2020) https://www.swgdam.org/_files/ugd/4344b0_6cc9e7c82ccc4fc0b5d10217af64e31b.pdf (.pdf).

[31] Interim Policy at 4.

[32] In fact, the Interim Policy itself states that it "does not impose any legal limitations on otherwise lawful investigative or prosecutorial activities or techniques utilized by the Department of Justice, or limit the prerogatives, choices, or decisions available to, or made by, the Department in its discretion." Interim Policy at 1, n.1.

[33] *Forensic Genetic Genealogical DNA Analysis and Search (FGGS)*, Governor's Office of Crime Prevention, Youth, and Victim Services, https://goccp.maryland.gov/crime-statistics/law-enforcement-reports/fggs/.

[34] For more on these two laws, *see* Jennifer Lynch, *Maryland and Montana Pass the Nation's First Laws Restricting Law Enforcement Access to Genetic Genealogy Databases*, EFF.org (June 7, 2021) https://www.eff.org/am/deeplinks/2021/06/maryland-and-montana-pass-nations-first-laws-restricting-law-enforcement-access.

[35] S.B. 187, 2021 Leg., 442nd Sess. (Md. 2021), https://mgaleg.maryland.gov/2021RS/bills/sb/sb0187T.pdf.

[36] S.B. 187, 2021 Leg., 442nd Sess. (Md. 2021), https://mgaleg.maryland.gov/2021RS/bills/sb/sb0187T.pdf.

[37] Courts have yet to hold the Fourth Amendment prohibits these searches. *See, e.g., Raynor v. State*, 99 A.3d 753 (2014) (testing of DNA that defendant inadvertently left on chair at a police station was not a "search" for Fourth Amendment purposes); *State v. Burns*, 988 N.W.2d 352 (Iowa 2023) (similar).

damages so that people can enforce the law through the courts. It requires the governor's office to report annually and publicly on law enforcement use of FGG and covert collection. Finally, it states explicitly that criminal defendants may use the technique as well to support their defense (but places similar restrictions on a defendant's use).

Montana's statute[38] is much more limited than Maryland's, but it does require a warrant before government entities can use familial DNA or partial match search techniques on either consumer DNA databases or the state's criminal DNA identification index. The statute defines a "familial DNA search" broadly as a search that uses "specialized software to detect and statistically rank a list of potential candidates in the DNA database who may be a close biological relative to the unknown individual contributing the evidence DNA profile."

For defendants and consumers outside Maryland and Montana, there are few, if any, limitations on FGG searching. The companies themselves have only minimal contractual restrictions and no technical measures to protect data from unauthorized or unwanted access. GEDmatch states that it only allows law enforcement to access genetic data from users who have expressly opted in to police searches,[39] while FamilyTreeDNA allows police to search all users' data unless those users have explicitly opted out.[40] In contrast, MyHeritage "prohibits law enforcement use of its DNA Services" and will not provide information to law enforcement "unless required by a valid court order or subpoena for genetic information."[41] But, like the Department of Justice's Interim Policy, genetic genealogy companies' policies are not set in stone and could be changed at any time.

IV. FGG Can Both Misidentify Suspects and Exonerate Wrongfully Convicted People

FGG searches have already implicated innocent people who happen to have DNA markers similar to the forensic sample, and they will do so again. Although police used FGG to identify the "Golden State Killer," an earlier search in the same case identified a different person as the likely perpetrator.[42] In 2014, a similar search in an Idaho cold

[38]https://www.leg.mt.gov/bills/mca/title_0440/chapter_0060/part_0010/section_0040/0440-0060-0010-0040.html.
[39]**https://www.buzzfeednews.com/article/peteraldhous/this-genealogy-database-helped-solve-dozens-of-crimes-but**
[40] FamilyTreeDNA, FamilyTreeDNA - Privacy Statement, Family Tree DNA https://www.familytreedna.com/legal/privacy-statement.
[41] **MyHeritage Privacy Policy (March 15, 2023) https://www.myheritage.com/privacy-policy. Similarly, Ancestry and 23andMe have required warrants for specific user information (as opposed to access to their entire databases) since at least 2015. *See* https://www.23andme.com/transparency-report/; https://www.ancestry.com/cs/transparency-2015.**
[42]https://www.independent.co.uk/news/world/americas/golden-state-killer-2018-case-solved-identity-genetic-genealogy-websites-dna-joseph-james-deangelo-a8326946.html; https://www.eff.org/deeplinks/2018/05/distant-relatives-arent-only-ones-looking-your-dna-geneology-sites-law-enforcement

case led police to suspect an innocent man.[43] Even without FGG, DNA matches have led officers to suspect—and jail—the wrong person. In 2012, a California man spent five months in jail after a database search linked his DNA to DNA found on the fingernails of a murder victim—although he was in the hospital when the murder occurred.[44] Prosecutors believe paramedics may have transferred his DNA to the murder victim when they responded to the crime scene hours after dropping him off at the hospital.[45]

DNA transfers and the sensitivity of modern DNA collection tools and analysis are likely to lead to more misidentifications. We shed DNA constantly[46] and, as the California example above shows, our DNA can easily be transferred from item to item or from one location to another, even if we never touched the item ourselves or were never at the scene of the crime. One study showed that after two people shook hands and then each handled a separate knife, "[i]n 85% of cases, the DNA of the other person was transferred to the knife and profiled. In one-fifth of the samples, the DNA analysis identified this other person as the main or only contributor of DNA to the 'weapon.'"[47] Another study discovered that sperm could be transferred from a man's underwear to the underwear of another person when their clothes were washed together.[48] With the increasing sensitivity and sophistication of DNA collection tools and analysis techniques, labs can in some cases, identify a source, even if the forensic sample is extremely small, contains genetic material from multiple people, or is seriously degraded. These advances, combined with the challenges in accurately mapping genetic lineages, increase the risk that an FGG search could lead police to suspect the wrong person.

On the other hand, FGG has cleared some people who were wrongfully prosecuted and led to the exoneration of two wrongfully-convicted individuals. In the Idaho case mentioned above, a man was prosecuted and spent 20 years in prison after coercive interview tactics may have elicited a false confession—despite the fact that his DNA did not match that found at the crime scene. A later FGG search finally led police to the true perpetrator.[49] A Canadian researcher who has collected information on over 500 cases involving FGG has found FGG was involved in helping to clear 18 wrongful prosecutions and 7 wrongful convictions.[50] Maryland's FGG law explicitly allows defendants to

[43]Jennifer Lynch, *How Private DNA Data Led Idaho Cops on a Wild Goose Chase and Linked an Innocent Man to a 20-year-old Murder Case*, EFF.org (May 1, 2015) https://www.eff.org/deeplinks/2015/05/how-private-dna-data-led-idaho-cops-wild-goose-chase-and-linked-innocent-man-20.
[44] Henry Lee, *How Innocent Man's DNA Was Found at Killing Scene*, SF Gate (June 26, 2013), http://www.sfgate.com/crime/article/How-innocent-man-s-DNA-was-found-at-killing-scene-4624971.php.
[45]https://www.themarshallproject.org/2018/04/19/framed-for-murder-by-his-own-dna
[46] Humans shed roughly 500 million skin cells per day. Leonard M. Milstone, *Epidermal Desquamation, Journal of Dermatological Science* 36, no. 3 131–40 (December 1, 2004)).
[47] Cynthia M. Cale, *Forensic DNA evidence is not infallible*, Nature (Oct. 28, 2015), https://www.nature.com/articles/526611a.
[48] Erin Murphy, *Inside the Cell: The Dark Side of Forensic DNA* 40–41 (2015) (citing research).
[49] Neil Vigdor, *An Innocent Man Went to Prison for Murder. Now, It's the Real Killer's Turn*, N.Y. Times (June 8, 2021) https://www.nytimes.com/2021/06/08/us/angie-dodge-murder.html.
[50] Tracey Leigh Dowdeswell, *Forensic Genetic Genealogy Coding Book & Annotated Bibliography 2022* (Feb. 21, 2023) https://data.mendeley.com/datasets/jcycgvhm96/1/files/77cfadae-dc7b-47a8-a31d-9bc3ce287188.

conduct their own FGG searches, pursuant to similar restrictions on law enforcement searches.[51]

V. Discovery and Building a Record in FGG Cases

Before defense attorneys can challenge an FGG search, they will need to build a complete record with facts about how FGG searches operate in general, how FGG was used in their case, how accurate these searches are, and their limitations.

First, defense attorneys need to identify whether FGG has been used in their case. In most cases that have involved FGG, officers have readily disclosed this fact. However, there may be future cases where this is not disclosed. In that case, it is important to question how the police found and identified your client. As part of that query, attorneys should also ask whether traditional criminal investigative methods—including a CODIS search—were exhausted before officers turned to FGG.[52]

Second, defense attorneys should ask for lab records and reports that describe exactly how the company extracted a SNP profile from the forensic sample. These profiles are more difficult to obtain from forensic samples because, in many cases samples are degraded and quantities of DNA are low. This means labs must increase processing and amplification efforts to obtain genotypes of sufficient quality to perform a search. Attorneys should also ask if the sample contained a mixture of DNA sources, and, if so, how many individuals are believed to be represented in the sample and whether it was possible to isolate the DNA alleged to be from your client. Trying to use low copy number DNA and multi-source samples for FGG searches could lead to serious errors or misidentifications.[53] Finally, if other individuals' DNA was present in the forensic sample or otherwise present at the crime scene, did the law enforcement agency run FGG searches on each individual's DNA or just your client's?

Next, defense attorneys should ask for specific information about how the FGG search was conducted and the genetic genealogy site or sites that were used. Are there any validation studies for the specific genetic genealogy technology used? How accurate is it?[54] What are its limitations? Did officers or genetic genealogists disclose to the site that they were conducting a criminal search or did they upload the forensic sample secretly using a pseudonym?[55] How many profiles were searched? Did existing site users know

[51] https://legiscan.com/MD/text/SB187/id/2416137.
[52] Both the DOJ's Interim Policy and Maryland's FGG statute require this.
[53] Parabon states that it is able to work with two-person mixtures, but only if the suspect's DNA represents at least 40% of the mixture and "a single-source reference sample from the second contributor is available." Ellen Greytac et al, *Genetic genealogy for cold case and active investigations*.
[54] A discussion of how to bring a Daubert or Frye challenge is beyond the scope of this article, but the answers to the questions in this section may help attorneys determine whether to make such a challenge.
[55] DOJ's interim policy on FGG searches requires officers to exhaust traditional criminal investigative tools first and to identify themselves to the site owner. https://www.justice.gov/olp/page/file/1204386/download

about or consent to law enforcement searches?[56] Did officers get a warrant? If so, what was the basis for probable cause?

It is also important to learn more about who conducted the FGG search and how it was conducted. Did the police investigator work with a private company or an individual? What are the credentials and training of the individual who performed the FGG search and built out the family tree?[57] Did the company or individual produce any reports documenting their work? If the investigator worked with Parabon Nanolabs, the company produces at least two reports for each case that could be helpful to the defense. The first is a "Snapshot Phenotype Report" that describes how it performed its extraction and search as well as its assessment of how likely the case is to be solved using FGG.[58] Parabon also produces a "Snapshot Genetic Genealogy Report," which describes how the company performed its additional genealogical research, how it determined potential relatives, how it built its family tree, and what its recommendations for next steps were.[59]

Once defense attorneys have these reports, they should look for information about how many centimorgans the forensic DNA sample shared with the user(s) on the site who were identified as potential relatives.[60] If that number is low, there may be an equal or better chance that the DNA similarities happened by chance rather than heredity. Sometimes the report will also state how confident the company was in the match(es) it found. For example, in one case, Parabon wasn't able to find a close relative but did identify two individuals who could have been between sixth to eighth degree relations. Based on this result, Parabon stated it thought there was a "low probability" the case would be solved using FGG, but that it was "likely to generate actionable info."[61]

It is also important to know how many site users were identified as related to the forensic sample, how the algorithm determined those matches, how law enforcement narrowed down its search, whether law enforcement followed up on any other potential matches, and, if some potential matches were rejected, why they were rejected. These facts may be useful in determining if confirmation bias was at play—did law enforcement, after learning of potential matches from different genetic lines, decide to

[56] Whether individual users of genetic genealogy databases have consented to law enforcement searches is an important fact to know, but is somewhat beside the point. In all cases that we know of so far, site users who have uploaded their own DNA are not the suspects. Instead, law enforcement is using their DNA to try to find someone else who is not on the genetic genealogy site. The site users cannot meaningfully consent to a search on behalf of someone who has not personally uploaded their own genetic data but who may nevertheless have genetic data on the site, solely by virtue of sharing their DNA with a relative.
[57] Maryland now requires genetic genealogists to obtain specific training and requires both the labs that process genetic samples and the genetic genealogists that interpret FGG results to obtain licenses from the state Office of Health Care Quality. https://mgaleg.maryland.gov/2021RS/bills/sb/sb0187T.pdf.
[58] See, e.g., https://www.eff.org/document/state-v-bentaas-defendants-exhibit-2-support-motion-suppress. This report also provides Parabon's predictions of other features of the person behind the forensic sample, including hair color, skin pigmentation, ancestry (and by extension ethnicity), eye color, and even whether the person has freckles.
[59] See, e.g., https://www.eff.org/document/state-v-bentaas-defendants-exhibit-3-support-motion-suppress.
[60] See, e.g., https://www.eff.org/document/state-v-bentaas-defendants-exhibit-2-support-motion-suppress.
[61] See https://www.eff.org/document/state-v-bentaas-defendants-exhibit-2-support-motion-suppress.

follow up and build a family tree for only one of those genetic lines because it pointed toward an individual who they already suspected?

VI. Fourth Amendment Challenges in Cases Involving FGG

A. Challenging the Extraction and Search of DNA From Discarded Items

There are several ways to challenge police searches that involve FGG. The most common approach so far has been to focus not on the FGG search itself but on challenging the DNA collected from the discarded item used to directly identify their client. Relying on cases like *California v. Greenwood*, 486 U.S. 35 (1988), prosecutors have argued the Fourth Amendment does not require a warrant to search DNA found on items, like a cigarette butt or a straw, that a suspect has discarded because the suspect abandoned their privacy interests in their DNA when they left it behind on those items.[62]

In response, defense may argue that the Fourth Amendment creates a high bar against collecting DNA from people outside the criminal justice system, even if it is found on items the person has voluntarily discarded. In *Greenwood*, the Supreme Court held that the Fourth Amendment does not protect the contents of people's trash left for pickup because they "abandoned" an expectation of privacy in the trash. But unlike a straw or a cigarette butt, our DNA contains our entire genetic makeup—intensely personal and private information wholly different in scope and kind from a piece of trash.[63] And despite the uniquely revealing nature of DNA, we cannot avoid leaving behind the whole of our genetic code wherever we go. For these reasons, we can never "abandon" a privacy interest in our DNA. Even if police do not need a warrant to rummage through trash (and several states disagree on this point),[64] police should be required to get a warrant to rummage through DNA.

Prosecutors have also argued that *Maryland v. King*, 69 U.S. 435 (2013), exempts them from a warrant requirement in this context. In *King*, the Supreme Court held that warrantless collection of DNA from arrestees does not violate the Fourth Amendment, in part because an arrestee's privacy interest in the small portion of their DNA entered into CODIS is minimal when compared against the government interest in identification. Similarly, in cases involving FGG, prosecutors have argued that collecting DNA from a discarded item is at most a de minimus privacy invasion. However, defendants can counter that the government seized and has access to *all* of the defendant's genetic

[62] See https://www.eff.org/document/state-v-bentaas-state-brief-opposition-motion-suppress
[63] See *Riley v. California*, 573 U.S. 373, 393 (2014) (finding "Modern cell phones, as a category, implicate privacy concerns far beyond those implicated by the search of a cigarette pack.")
[64] See, e.g., *State v. Krivda*, 5 Cal.3d 357 (1971); *State v. Hempele*, 120 N.J. 182 (1990); *State v. Lien*, 364 Or. 750 (2019).

information, not just the 20 loci used for a CODIS profile. As the U.S. Supreme Court has made clear in several cases decided after *King*, the Fourth Amendment is concerned with the entirety of the private information revealed to police through a search—not just the pieces of information the government ultimately considers useful. A search that turns up nothing useful is still a search. *See, e.g., Birchfield v. North Dakota*, 136 S. Ct. 2160, 2178 (2016). Further, *King* involved the suspicionless and mass collection of DNA from arrestees, individuals with arguably diminished privacy interests. Unlike arrestees, DNA collected from a free person should be entitled to full Fourth Amendment protection—and in fact, in most cases police have obtained a warrant to collect a DNA sample directly from a suspect before their arrest.

B. Challenging the FGG Search Directly as a General Warrant

Defense attorneys may also decide to challenge the FGG search directly as an unconstitutional "general warrant." FGG searches are fishing expeditions through millions of innocent people's DNA. They leave it up to an officer's discretion to determine who to search further and who to arrest. As such they are similar to the general warrants and writs of assistance (broad warrants that allowed officials to search anywhere for smuggled goods) that this country's founders sought to preclude in drafting the Fourth Amendment. See *Stanford v. Texas*, 379 U.S. 476, 481-82 (1965).

Defendants who decide to challenge an FGG search directly may face standing challenges or an argument that they lack a privacy interest in the genetic data searched because the defendant did not upload their own genetic information to the genetic genealogy database. Instead, investigators searched data uploaded by the defendant's biological relatives to generate leads in the investigation. If the individual who uploaded the data is not the person who will ultimately stand trial (or, put the other way, if the defendant on trial did not upload their own genetic data), courts may resist finding the defendant had a protectable privacy interest in the data. The Supreme Court has held that Fourth Amendment rights are personal; defendants cannot assert them for someone else, even if the defendant is the target of the search. *See Rakas v. Illinois*, 439 U.S. 128, 133–34, 36 (1978).

However, it is possible to argue the defendant *does* have a personal privacy interest in the data because some of the defendant's own genetic code is in the database; the genetic data that the defendant shares with their biological relatives is what led the police to the defendant in the first place, even though that genetic data was uploaded by someone else. The Supreme Court recognized in *Carpenter v. United States* that we can still have a privacy interest in data shared with another party, even if that data is outside our own control. 138 S. Ct. 2206, 2220 (2018). So, like the cell site location information at issue in *Carpenter*, theoretically the Fourth Amendment should protect genetic genealogy data, even when it is shared involuntarily with a relative who has chosen to upload it to a consumer website accessible to the police.

Assuming defendants survive a standing challenge, they can argue that FGG searches fail the Fourth Amendment's requirements that a search be particularized, supported by probable cause, and generally be conducted pursuant to a warrant. FGG searches lack particularity because they do not identify any specific person or genetic profile to be searched. *See Stanford v. Texas*, 379 U.S. at 485-86[65] They are overbroad because they encompass intensely private data from potentially millions of people. *Id.* at 727. And these searches cannot be supported by probable cause because there is no evidence that any individual site user is in any way personally connected to the crime. The mere possibility that some users may share some genetic data with the forensic sample should not be sufficient to support probable cause to search through all users' data. *See Ybarra v. Illinois*, 444 U.S. 85, 92 (1979) (a person's "mere propinquity" to criminal activity is insufficient to establish probable cause to search that person).

FGG searches are dragnets, similar to officers searching every house in a town with a population of millions on the off chance that some of those houses could contain evidence useful to finding the perpetrator of a crime. Like the original general warrants, FGG searches "specif[y] only an offense . . . and [leave] to the discretion of the executing officials the decision as to which persons should be arrested and which places should be searched." *Steagald v. United States*, 451 U.S. 204, 220 (1981). Almost by definition, FGG searches give police access to private information, regardless of whether they use it. Therefore, an FGG search is precisely the sort of "general, exploratory rummaging" the Fourth Amendment was intended to preclude. *Coolidge v. New Hampshire* (1971) 403 U.S. 443, 467; *Andresen v. Maryland,* (1976) 427 U.S. 463, 479-480. With or without a warrant, the Fourth Amendment prohibits searches like this in the physical world. It should prohibit genetic dragnets as well.

VII. Conclusion

The ability to research family history and disease risk shouldn't carry the threat that our data will be accessible to police and used in ways we never could have foreseen. Research shows that individuals in the United States believe that data stored with genetic genealogy companies is intensely private. They rank FGG searches on par with searches of bedrooms, text messages, or emails.[66] That's because these searches invade our privacy in unique ways—they allow law enforcement to access information about us that we may not even know ourselves, that we have no ability to hide or change, and that could reveal more about us in the future than scientists know now. For all these reasons, legislators should follow Maryland and Montana's lead and enact sufficient restrictions or bans on FGG searches. Until that time, defense attorneys should continue to challenge them in their cases.

[65] FGG searches are, by design, not targeted at finding specific users or based on individualized suspicion—a fact the police must admit because they don't know who their suspect is.
[66] Christopher Slobogin & James Hazel, "'A World of Difference?': Law Enforcement, Genetic Data and the Fourth Amendment," Duke L.J., Vol. 70 at 31, 2020, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3585241.