Four misconceptions about investigative genetic genealogy

Christi J. Guerrini1,*, Ray A. Wickenheiser2,†, Blaine Bettinger3,**, Amy L. McGuire1,‡‡ and Stephanie M. Fullerton4,¶

1 Baylor College of Medicine, Center for Medical Ethics and Health Policy, Houston, TX 77030, USA
2 New York State Police Crime Laboratory System, Albany, NY 12226, USA
3 Bond, Schoeneck & King, PLLC, Syracuse, NY 13202, USA
4 University of Washington School of Medicine, Department of Bioethics and Humanities, Seattle, WA 98195, USA
*Corresponding author. E-mail: guerrini@bcm.edu

ABSTRACT

Investigative genetic genealogy (IGG) is a new technique for identifying criminal suspects that has sparked controversy. The technique involves uploading a crime scene DNA profile to one or more genetic genealogy databases with the intention of identifying a criminal offender’s genetic relatives and, eventually, locating the offender within the family tree. IGG was used to identify the Golden State Killer in 2018 and it is now being used in connection with hundreds of cases in the USA. Yet, as more law enforcement agencies conduct IGG, the privacy implications of the technique have come under scrutiny. While these issues deserve careful attention, we are concerned that their discussion is, at times, based on misunderstandings related to how IGG is used in criminal investigations and how IGG departs from...
Four misconceptions about investigative genetic genealogy

traditional investigative techniques. Here, we aim to clarify and sharpen the public debate by addressing four misconceptions about IGG. We begin with a detailed description of IGG as it is currently practiced: what it is and—just as important—what it is not. We then examine misunderstood or not widely known aspects of IGG that are potentially confusing efforts to have constructive discussions about its future. We conclude with recommendations intended to support the productivity of those discussions.

KEYWORDS: Fourth Amendment, forensic, genetic databases, genetic genealogy, genetic privacy

I. BACKGROUND

Investigative genetic genealogy (IGG) is a new technique for identifying criminal suspects that has sparked controversy. The process of IGG involves uploading a crime scene DNA profile to one or more genetic genealogy databases with the intention of partially matching it to a criminal offender’s genetic relatives and, eventually, locating the offender within their family tree. In April 2018, investigators announced the successful use of IGG to identify Joseph James DeAngelo as the Golden State Killer (GSK) responsible for at least 13 murders and 45 rapes throughout California in the 1970s and 1980s.1 According to one expert, by September 2020, IGG had led to the successful identification of over 150 suspects.2

Most of the major firms that maintain genetic genealogy databases—including Ancestry, 23andMe, and MyHeritage—have adopted policies that forbid law enforcement from participating in their databases for investigative purposes, either through requirements that users provide only their own DNA for analysis or explicit bans on the conduct of IGG in their databases.3 However, in December 2018, FamilyTreeDNA (FTDNA), which maintains a genetic genealogy database consisting of 1.15 million autosomal DNA profiles,4 adopted a policy permitting law enforcement to participate

1 Matthias Gafni & Lisa M. Krieger, Here’s the ‘Open-Source’ Genealogy DNA Website that Helped Crack the Golden State Killer Case, EAST BAY TIMES, Apr. 26, 2018, https://www.eastbaytimes.com/2018/04/26/ancestry-23andme-deny-assisting-law-enforcement-in-east-area-rapist-case/.
2 CeCe Moore & Paul Fronczak, The Foundling—Resolving a Case of Unknown Identity Through the Use of Genetic Genealogy, International Symposium on Human Identification (Sept. 14, 2020). Details of some cases are provided in Sara. H. Katsanis, Pedigrees and Perpetrators: Uses of DNA and Genealogy in Forensic Investigations, 21 ANN. REV. GENOMICS HUMAN GENET. 12.1, 12.10–12.14 (2020).
3 Terms of Service, 23ANDME.COM, https://www.23andme.com/about/tos/ (updated Sept. 30, 2019); 23andMe Guide for Law Enforcement, 23ANDME.COM, https://www.23andme.com/law-enforcement-guide/ (last visited Apr. 20, 2020); Ancestry Terms and Conditions, ANCESTRY.COM, https://www.ancestry.com/cs/legal/termsandconditions&Eligibility (updated July 25, 2019); Ancestry Guide for Law Enforcement, ANCESTRY.COM, https://www.ancestry.com/cs/legal/lawenforcement (last visited Apr. 20, 2020); MyHeritage—Terms and Conditions, MYHERITAGE.COM, https://www.myheritage.com/terms-and-conditions (updated Apr. 18, 2020).
4 International Society of Genetic Genealogy, Autosomal DNA Testing Comparison Chart, ISOGG Wiki, https://isogg.org/wiki/Autosomal_DNA_testing_comparison_chart (last visited Apr. 25, 2020) [hereinafter ISOGG, Comparison Chart].
in its database to identify violent criminals and human remains.\textsuperscript{5} FTDNA database participants can choose whether to make their information available for law enforcement searches in a process known as ‘law enforcement matching.’\textsuperscript{6} Specifically, registered participants located in the USA are automatically opted in to law enforcement matching, but they can choose to opt out of law enforcement matching at any time by selecting this option in their user profile.\textsuperscript{7} By contrast, on May 18, 2019, all registered participants of the genetic genealogy database known as GEDmatch were automatically opted out of law enforcement matching.\textsuperscript{8} At that time, the GEDmatch database consisted of approximately 750,000 unique single nucleotide polymorphism (SNP) profiles designated as public (in April 2020 that number was approximately 900,000).\textsuperscript{9} However, GEDmatch participants can choose to opt in to law enforcement matching at any time by selecting this option in their user profile.\textsuperscript{10} Individuals joining GEDmatch after May 18, 2019 are required to decide at the time of registration whether they will opt in to or out of law enforcement matching, where the opt-in choice is now selected by default.\textsuperscript{11} In December 2019, GEDmatch was acquired by Verogen, a forensic genomics company, and in January 2021, FT DNA's parent company announced its merger with myDNA, an Australian personalized genomics company.\textsuperscript{12}

As more law enforcement agencies use IGG in investigations, the privacy implications of the technique have come under scrutiny.\textsuperscript{13} Privacy concerns associated with IGG stem from the sources and types of genetic information maintained in genetic genealogy databases, which differ in important respects from the composition of law enforcement databases, such as the National DNA Index System (NDIS) in the USA. US law enforcement databases—which are generally referred to by the acronym

\textsuperscript{5} Salvador Hernandez, One of the Biggest At-Home DNA Testing Companies is Working with the FBI, BuzzfeedNews, Jan. 31, 2019, https://www.buzzfeednews.com/article/salvadorhernandez/familytree-dna-fbi-investigative-genealogy-privacy.

\textsuperscript{6} Law Enforcement Matching—Frequently Asked Questions, FamilyTreeDNA.com, https://learn.familytreedna.com/ftdna/law-enforcement-faq/ (last visited Apr. 25, 2020).

\textsuperscript{7} Id.

\textsuperscript{8} Jon Schuppe, Police Were Cracking Cold Cases with a DNA Website. Then the Fine Print Changed, NBC News, Oct. 23, 2019, https://www.nbcnews.com/news/us-news/police-were-cracking-cold-cases-dna-website-then-fine-print-n1070901.

\textsuperscript{9} Personal communication with C. Rogers & B. Williams, Verogen, with authors (C.J.G. & A.L.M.) (April 29, 2020). In September 2020, Verogen reported 1.1 million registered GEDmatch participants. Swathi Kumar & Kirk Campbell, GEDmatch: A Platform for Data Driven Forensic Intelligence, International Symposium on Human Identification (Sept. 14, 2020).

\textsuperscript{10} GEDmatch.com Terms of Service and Privacy Policy, GEDMATCH.COM, https://www.gedmatch.com/tos.htm (revised Dec. 9, 2019).

\textsuperscript{11} Id.

\textsuperscript{12} Nila Bala, We’re Entering a New Phase in Law Enforcement’s Use of Consumer Genetic Data, Slate (Dec. 19, 2019), https://slate.com/technology/2019/12/gedmatch-verogen-genetic-genealogy-law-enforcement.html; Pharmacogenetic and Genealogy Pioneers Merge for Historic Partnership, PR Newswire, (Jan. 7, 2021), https://www.prnewswire.com/news-releases/pharmacogenetic-and-genealogy-pioneers-merge-for-historic-partnership-301202798.html.

\textsuperscript{13} More recently, Othram, a private forensic laboratory, launched a genetic database known as DNA Solves, which exists primarily and explicitly to facilitate IGG. DNA SOLVES, https://dnasolves.com/ (last visited Oct. 22, 2020). Members of the public are invited to upload their DNA profiles to the database for use by law enforcement and are provided ancestry information based on their uploaded DNA, but they are not allowed to participate in the database. For these reasons, IGG conducted in DNA Solves does not seem to raise the same privacy and consent issues as IGG conducted in genetic genealogy databases.
Combined DNA Index System (CODIS) for the software that supports them—are comprised of DNA profiles of persons who have been convicted of, and in some cases arrested for, crimes. The profiles consist of 20 short tandem repeats (STRs) generated by accredited forensic laboratories that must comply with a host of quality assurance standards and requirements.

By contrast, genetic genealogy databases are populated voluntarily by individuals interested in exploring their ancestry and family lineage. The genetic data that they contribute are autosomal DNA profiles consisting of 600,000–700,000 SNPs generated by commercial test providers. Unlike STRs, SNPs are more evenly (and densely) distributed throughout a person’s genome and hence can carry information about a person’s medical history and appearance. If analyzed with regard to patterns of linked variation along sections of chromosomes, SNPs can also be used to identify more distant genetic relatives than STRs.

For these reasons, IGG represents an expansion over standard CODIS searching in terms of the population of persons whose genetic information might be searched in an investigation, even if the search objectives are different, and the kinds of information that are the basis for identification. Although this is presumably known and accepted by genetic genealogy database participants who opt in to law enforcement matching, the same cannot be said of all of their non-participant relatives whose names might become part of an investigation by virtue of the fact that they are members of a suspect’s family tree. Because most genetic genealogy database participants are persons of

14 34 U.S.C. § 12592(a)(1) (authorizing the establishment of NDIS); Natalie Ram, Genetic Privacy after Carpenter, 105 Va. L. Rev. 1357, 1376–77 (2019) (summarizing jurisdictional inclusion criteria for contribution of DNA profiles to CODIS).
15 Ray A. Wickenheiser, Forensic Genealogy, Bioethics and the Golden State Killer Case, 1 FORENSIC SCI. INT’L SYNERGY 114, 114, 121–22 (2019). These include standards to attain and maintain accreditation, which is a requirement to participate in CODIS. Federal Bureau of Investigation, QUALITY ASSURANCE STANDARDS FOR DNA DATABASING LABORATORIES (2011), https://www.fbi.gov/file-repository/quality-assurance-standards-for-dna-databasing-laboratories.pdf/view; Federal Bureau of Investigation. QUALITY ASSURANCE STANDARDS FOR FORENSIC DNA TESTING LABORATORIES (2011), https://www.fbi.gov/file-repository/quality-assurance-standards-for-forensic-dna-testing-laboratories.pdf/view. They also include over 400 requirements set forth in ISO/IEC 17025:2017 GENERAL REQUIREMENTS FOR THE COMPETENCE OF TESTING AND CALIBRATION LABORATORIES, https://www.iso.org/obp/ui/?iso:std:iso-iec:17025:ed-3:v1:en.
16 In addition, some individuals participate in direct-to-consumer genetic testing for health and wellness insights. See Erin Murphy, Law and Policy Oversight of Familial Searches in Recreational Genealogy Databases, 292 FORENSIC SCI. INT’L e5, e5 (2018). Some of these testing providers, such as 23andMe, maintain genetic genealogy databases and offer related genealogical services.
17 ISOGG, Comparison Chart, supra note 4.
18 Murphy, supra note 16, at e5.
19 Id. at e6.
20 Id.; see also Natalie Ram, Incidental Informants: Police Can Use Genealogy Databases to Help Identify Criminal Relatives—But Should They?, MARYLAND BAR J., Sept./Aug. 2018, at 9, 10.
21 Ram, supra note 20, at 10; Debbie Kennett, Using Genetic Genealogy Databases in Missing Persons Cases and to Develop Suspect Leads in Violent Crimes, 301 FORENSIC SCI. INT’L 107, 114 (2019); Christi J. Guerrini et al., Should Police Have Access to Genetic Genealogy Databases? Capturing the Golden State Killer and Other Criminals Using a Controversial New Forensic Technique, 16 PLOS BIOLOGY e2006906, at 7 (2018); Stephanie Fullerton & Rori Rohlfs, Should Police Detectives Have Total Access to Public Genetic Databases?, LEAPSMAG, July 23, 2018, https://leapsmag.com/should-police-detectives-have-total-access-to-public-geneticdatabases/.
European ancestry, the privacy of their relatives is especially at risk.\textsuperscript{22} Moreover, IGG has reportedly been conducted in at least one database without the company’s—or their participants’—knowledge and consent.\textsuperscript{23} That practice would violate the database’s current terms of service, which explicitly prohibit IGG.\textsuperscript{24}

While these issues deserve careful attention, we are concerned that their discussion is, at times, based on misunderstandings related to how IGG is conducted and used in criminal investigations and how IGG departs from traditional investigative techniques. Here, we aim to clarify and sharpen the public debate by addressing four misconceptions about IGG. We begin with a detailed description of IGG as it is currently practiced: what it is and—just as important—what it is not. We then examine misunderstood or not widely known aspects of IGG that are potentially confusing efforts to have constructive discussions about its risks and benefits. Along the way, we identify persistent concerns and controversies related to each misconception that might benefit from policy intervention. We conclude with broad recommendations intended to support the productivity of discussions about the future of IGG.

\section*{II. IGG Step by Step}

Put simply, IGG describes the process of using information about genetic similarities and known family relationships to generate investigative leads. The basic information used in IGG falls into two categories: genetic relative information, which is generated by a genetic genealogy database based on its internal comparison of SNP profiles in the database; and genealogical and other often publicly accessible information, such as information from census records and obituaries, that describes family relationships. Law enforcement integrates these two categories of information to develop family trees and then identifies and investigates high-likelihood suspects within those trees.

Thus, IGG comprises two steps that are book-ended by standard police work (Fig. 1).\textsuperscript{25} According to best practices,\textsuperscript{26} when an offender leaves a biological sample (e.g., blood or semen) at a crime scene, an accredited forensic laboratory first generates an STR profile from that sample, which is called the ‘forensic sample.’ If there are no suspects, the STR profile is then uploaded to CODIS to identify a possible match.

\begin{itemize}
\item \textsuperscript{22} In 2018, researchers projected that genetic genealogy database searches of 60\% of Americans of European descent will result in at least a third-cousin match. Yaniv Erlich et al., \textit{Identity Inference of Genomic Data Using Long-Range Familial Searches}, 362 SCIENCE 690 (2018). As explained in a published letter responding to this report, however, probability of a relative match to a source profile does not equate to probability of identification of the source profile. Ellen McRae Greytak et al., \textit{eLetter} (Oct. 28, 2018), https://science.sciencemag.org/content/362/6415/690/tab-e-letters. Identifying individuals on the basis of their matches requires extensive genealogical work, which is made more difficult when matches are to distant relatives. \textit{Id.}; see also infra notes 70–71 and accompanying text.
\item \textsuperscript{23} Specifically, IGG was reportedly conducted in the MyHeritage database to identify the GSK. See Paige St. John, \textit{The Untold Story of How the Golden State Killer Was Found: A Covert Operation and Private DNA}, L.A. TIMES, Dec. 8, 2020, https://www.latimes.com/california/story/2020-12-08/man-in-the-window.
\item \textsuperscript{24} See MyHeritage—\textit{Terms and Conditions}, MYHERITAGE.COM, supra note 3.
\item \textsuperscript{25} See generally Wickenheiser, supra note 15, at 122.
\item \textsuperscript{26} Conducting a CODIS search before initiating IGG is required by IGG-specific guidelines developed by the US Department of Justice, the Sacramento County District Attorney's Office, the Scientific Working Group on DNA Analysis Methods, and the American Society of Crime Laboratory Directors. See infra note 80 and accompanying text.
\end{itemize}
with any of the 18 million-plus profiles in the database (Step 1). If the forensic STR profile matches a CODIS profile, following manual confirmation of the match to ensure no administrative errors occurred in connection with analysis of the CODIS sample, the name of the matching offender in CODIS is released to law enforcement as an investigative lead.

If the CODIS search does not result in a confirmed match, investigators usually attempt to identify and pursue leads through the normal channels. Among other things, they conduct interviews, collect and analyze evidence, and surveil suspects and their associates. This work might also include the collection of biological samples from one or more high-likelihood suspects, the generation of STR profiles from such samples, and direct comparison of those profiles with the forensic STR profile (Step 1).

But if these normal investigative channels are also unsuccessful, an agency with the resources to do so might turn to IGG to develop additional leads. If it does, first, a SNP profile is generated from the forensic sample, typically by a private lab on behalf

---

27 CODIS-NDIS Statistics, FBL.GOV, [https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics](https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics) (last visited Oct. 27, 2020). Although it is possible that investigators who are not bound by relevant guidelines might elect to skip this step and proceed directly to IGG, it usually will not make logistical or financial sense to do so. That is because conducting a CODIS search is significantly less expensive, time-consuming, and resource-intensive than conducting IGG. Given budgetary constraints of investigative agencies, the best use of limited funds and resources in most cases will be to first search for a match in CODIS, which is a relatively quick and straightforward process.

28 National DNA Index System (NDIS) Operational Procedures Manual v. 8 § 6.0 (2019), [https://www.fbi.gov/file-repository/ndis-operational-procedures-manual.pdf/view](https://www.fbi.gov/file-repository/ndis-operational-procedures-manual.pdf/view).
Four misconceptions about investigative genetic genealogy

II. OF THE INVESTIGATIVE AGENCY (Step 2). (This is because no US forensic lab currently has the technology to generate SNP profiles—a consequence of the fact that resource acquisition, instrument and method accreditation, and implementation is a painstaking and costly process.\(^{29}\)). The forensic SNP profile is then uploaded to one or more genetic genealogy databases, each of which generates a list of genetic relatives and the basis for their identification—specifically, the amount and length of shared DNA blocks.\(^{30}\) At this point, investigators or genealogists working on behalf of the agency search for genealogical and other information about the genetic relatives identified by the databases to develop a family tree that includes the source of the forensic sample (Step 3). In some cases, DNA samples might be collected from family members for SNP testing and comparison to narrow the search to specific branches of the larger tree in a process called ‘targeted testing.’

Investigators then identify individuals within the tree who are high-likelihood suspects because, among other things, they align with the known demographics or activities of the offender, and those suspects are investigated using standard techniques (Step 4). This investigation might unearth new information about family relationships and result in expanding or modifying the family tree or eliminating specific branches from investigative consideration.

Again, according to investigative best practices,\(^{31}\) the investigation eventually proceeds to the collection of a biological sample from a high-likelihood suspect, the generation of an STR profile from that sample, and direct comparison of that profile with the forensic STR profile. If the profiles match, the suspect is subjected to further investigation and may be arrested if the totality of the evidence supports the arrest. If the profiles do not match, the person is cleared as a suspect and the investigation continues.\(^{32}\)

III. FOUR ASPECTS OF IGG THAT ARE MISUNDERSTOOD OR NOT WELL KNOWN

This description of IGG provides a foundation from which to dispel certain misconceptions about IGG that we have observed in the media and online forums and heard in private conversations. The problem with allowing them to persist is that they potentially can confuse the public debate around IGG and lead to the adoption of policies intended to address concerns that do not reflect the current practice of IGG while misdirecting attention away from concerns that do.

\(^{29}\) Wickenheiser, supra note 15, at 118.

\(^{30}\) Ellen M. Greytak, CeCe Moore & Steven L. Armentrout, Genetic Genealogy for Cold Case and Active Investigations, 299 FORENSIC SCI. INT’L 103, 106–07 (2019).

\(^{31}\) Direct STR comparison matching before arrest is required by IGG-specific guidelines developed by the US Department of Justice, the Sacramento County District Attorney’s Office, and the Scientific Working Group on DNA Analysis Methods. See infra notes 43–45 and accompanying text.

\(^{32}\) Although it is possible that some investigators who are not required to engage in this step might choose to skip it, there are strong incentives to conduct direct STR comparison testing before making an arrest in an investigation involving DNA evidence. Perhaps most importantly, the evidence it generates is considered less vulnerable to legal challenge, especially where the collection of the suspect’s sample used in testing is supported by a warrant.
III.A. Misconception #1: When law enforcement conducts IGG in a genetic genealogy database, they are given special access to participants’ SNP profiles

Some criticisms of IGG rest on the assumption that police conducting IGG have access to the 600,000-plus SNPs comprising each profile of genetic genealogy database participants. Relatedly, and more broadly, it is sometimes suggested that IGG gives law enforcement special access to information in genetic genealogy databases that is off limits to other database users. Both statements are inaccurate.

When investigators conduct IGG, they participate in genetic genealogy databases on at least the same—and in some respects more restrictive—conditions as other database participants. FTDNA and GEDmatch are the only genetic genealogy databases that currently permit law enforcement participation, and both databases permit such participation only to solve violent crimes and identify human remains. Further, FTDNA requires law enforcement to apply to participate in its database on a case-by-case basis supported by specific documentation.

When investigators participate in any genetic genealogy database, after uploading the forensic SNP profile, they are provided the same kinds of information as other database participants. Thus, for each uploaded profile, the database provides law enforcement a list of genetic relatives of the offender (identified among participants who have opted in to such sharing in the database) that is generated by the database’s internal algorithm, and for each match, the length and location of shared DNA blocks, as well as contact and other information that the person chose to share with their matches. (Notably, that person can choose to remain de-identified—for example, by participating using a pseudonym or by allowing another person to manage their account.) The databases do not share with any participant, including law enforcement, the SNP profile of any other participant.

Although it is possible for law enforcement to participate in genetic genealogy databases against their terms of service or via a warrant, it does not appear that either route has resulted thus far in access to participants’ SNP profiles. If law enforcement participates against terms of service—that is, without self-identifying as law enforcement or for purposes other than allowed—they still participate on the same basis as other participants and, again, no genetic genealogy database gives its participants access to other participants’ SNP profiles. The only known warrants that have thus far been submitted to genetic genealogy databases requested either the identity of a relative match or permission to participate in the entire database and obtain the name, pseudonym, or contact information associated with a match’s account (effectively overriding the consent preferences of participants who opted out of law enforcement.

---

33 GEDmatch.com Terms of Service and Privacy Policy, supra note 10; FamilyTreeDNA Law Enforcement Guide, FamilyTreeDNA.com, https://www.familytreedna.com/legal/law-enforcement-guide (last visited May 1, 2020).
34 FamilyTreeDNA Law Enforcement Guide, supra note 33.
35 Greytak et al., supra note 30, at 106–07.
36 Id.
37 Peter Aldhous, Genetic Genealogy Helped Finally Crack the 1996 Murder of 18-Year-Old Angie Dodge, BUZZFEEDNEWS, May 16, 2019, https://www.buzzfeednews.com/article/peteraldhous/angie-dodge-cold-case-murder-genetic-genealogy-parabon.
Four misconceptions about investigative genetic genealogy

While it is unclear how law enforcement conducting IGG might use warrants in the future, it is unlikely that such requests will encompass SNP profiles. That is because the information that is useful to law enforcement conducting IGG is the estimated genetic relatedness of participants to the offender and contact information for close matches, not any participant’s SNP profile.

Nevertheless, any participant, including law enforcement, can potentially infer the sequence of the DNA segments of participants that match the uploaded SNP profile via a chromosome browser tool provided by some databases. Chromosome browsers provide a graphical view of DNA segments shared between matches on individual chromosomes. When browsers provide the location of the beginning and ending base pairs of minimum-length matching segments, the relative’s base pairs in those segments can be inferred given that the base pairs of the forensic SNP profile are known. It bears emphasizing, however, that chromosome browsers are available to all database participants and privacy concerns about them are not specific to IGG but are inherent to the phenomenon of genetic ‘identity by descent’ that the tool is based on.

Recently, bioinformaticians described how segment matching features can be exploited to uncover a substantial amount of a person’s genotype via techniques that involve uploading multiple real or artificial data sets to genetic genealogy databases. Law enforcement could theoretically seek to reconstruct a person’s SNP profile using these techniques, but the results would not be more informative in suggesting genetic relationships between participants and an offender than the location and size of shared DNA blocks. There are also straightforward solutions to these concerns: according to experts, the hack would not be possible if testing services cryptographically signed the SNP profiles they provide to participants or exchanged data directly at participants’ request.

---

38 One request involved service of a warrant on GEDmatch, which reportedly complied. Kashmir Hill & Heather Murphy, Your DNA Profile is Private? A Florida Judge Just Said Otherwise, NY TIMES, Nov. 5, 2019, https://www.nytimes.com/2019/11/05/business/dna-database-search-warrant.html. GEDmatch was subsequently purchased by Verogen, whose CEO stated that the company will fight such a warrant if one is served on Verogen in the future. Jon Schuppe, New Owner of Consumer DNA Database GEDmatch Vows to Fight Police Search Warrants, NBC NEWS, Dec. 10, 2019, https://www.nbcnews.com/news/us-news/new-owner-consumer-dna-database-gedmatch-vows-fight-police-search-n1099091. Ancestry.com has received several requests from law enforcement to participate in its database, but thus far it has not complied with any of them. Ancestry Transparency Report: July 2020, ANCESTRY.COM (July 10, 2020), https://www.ancestry.com/cs/transparency; Ancestry Transparency Report: 2019, ANCESTRY.COM, https://www.ancestry.com/cs/transparency-2019 (last visited Dec. 19, 2020); see also 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.

39 Michael D. Edge & Graham Coop, Attacks on Genetic Privacy Via Uploads to Genealogical Databases, 9 eLIFE e51810, at 2 (2020). In rare circumstances, it might also be possible to infer a genetic relative’s medical condition or predisposition. See Leah Larkin, Cystic Fibrosis: A Case Study in Genetic Privacy, THE DNA GEEK, Mar. 20, 2017, https://thednageek.com/cystic-fibrosis-a-case-study-in-genetic-privacy/.

40 Edge & Coop, supra note 39.

41 See id. at 14, 24 app. 1; Erlich et al., supra note 22, at 692–93 (2018); Peter M. Ney et al., Computer Security Risks of Distant Relative Matching in Consumer Genetic Databases, at 12, CoRR ARXIV (2018), https://arxiv.org/pdf/1810.02895.pdf.
In addition, it might be appropriate from a public trust perspective to prohibit law enforcement from using any information obtained from genetic genealogy databases to reconstruct some or all of a person’s genome. This prohibition would serve to complement the US Department of Justice’s (DOJ) existing ban (set forth in its Interim Policy on Forensic Genetic Genealogy) on law enforcement’s use of samples and information obtained during IGG to determine a person’s genetic predisposition for disease or any other medical condition or psychological trait.42

III.B. Misconception #2: Law enforcement will arrest a genetic genealogy database participant’s relatives based on the genetic information the participant provided to the database

Another concern that has been raised is that participating in genetic genealogy databases is the equivalent of ratting out one’s criminal relatives. By referring to database participants as ‘genetic witnesses’ or ‘genetic informants’, some might be led to believe that genetic relatedness is itself evidence of a relative’s criminal activity.

As described above, however, IGG is used only to generate leads. When investigators receive an anonymous phone call that a murderer works in auto maintenance, they follow up on that tip by focusing investigative efforts on individuals locally employed in that field. Similarly, when investigators learn through IGG that a criminal suspect is genetically related to certain genetic genealogy database participants, they follow up on that scientific tip by focusing investigative efforts on individuals who lie within the family tree of those participants. In both cases, the tip serves to narrow the universe of individuals who might warrant a closer look in the investigation. At that point, corroborating evidence connecting the person to the crime, such as access to the crime scene, possession of the weapon, and eyewitness accounts, must provide the justification for their arrest and conviction.

At a minimum, according to the DOJ Interim Policy and other IGG best practices, corroborating evidence should include a confirmed match resulting from a direct comparison of the STR profile of the suspect, generated from DNA collected directly from them, and the forensic STR profile.43 Indeed, the DOJ Interim Policy, which applies to every investigative agency in the USA receiving DOJ funds to conduct IGG and every criminal investigation involving the DOJ in the genealogy step of IGG, specifically prohibits the arrest of any person based solely on a genetic association generated by a genetic genealogy database.44 Even if the perpetrator is a database participant (a prospect we view as highly unlikely) and the genetic genealogy database indicates a perfect SNP match, direct STR comparison testing will be performed to confirm the

---

42 United States Department of Justice, Interim Policy: Forensic Genetic Genealogical DNA Analysis and Searching 6–7 (2019), https://www.justice.gov/olp/page/file/1204386/download [hereinafter DOJ Interim Policy].
43 Id. at 4; SACRAMENTO COUNTY DISTRICT ATTORNEY’S OFFICE, MEMORANDUM OF UNDERSTANDING: INVESTIGATIVE GENETIC GENEALOGY SEARCHING 2, https://chia187.wildapricot.org/resources/Documents/Sacramento%20County%20District%20Attorney's%20Office%20-%20IGG%20MOU%20Example.pdf (last visited May 1, 2020) [hereinafter SACRAMENTO COUNTY MOU]; SCIENTIFIC WORKING GROUP ON DNA ANALYSIS METHODS (SWGDAM), OVERVIEW OF INVESTIGATIVE GENETIC GENEALOGY 6 (2020), available at https://1ecb9588-ea6f-4feb-971a-73265dbf079c.filesusr.com/ugd/4344b0_6cc9e7c82cc4fc0b5d10217af6e31b.pdf [hereinafter SWGDAM OVERVIEW].
44 DOJ INTERIM POLICY, supra note 42, at 2, 4, 7.
match.\textsuperscript{45} Given the reliability of direct STR comparison testing,\textsuperscript{46} when it is conducted and results in a confirmed match, it is unlikely that it will result in a ‘mistaken identity’ arrest. The rare exceptions involve cases where the offender is an identical twin or stem cell or bone marrow transplant donor.\textsuperscript{47} However, the possibility of mistaken identity exists in either case regardless of whether IGG is part of the investigation, and IGG should minimize the possibility of the arrest of an innocent identical twin because the genealogical work is likely to uncover this relationship.

That said, an innocent person might experience significant stress in the time between when their DNA is collected for direct STR comparison testing and the return of a negative result. This was the case with Michael Usry, a New Orleans filmmaker whose father’s participation in a now-defunct genetic genealogy database consisting of Y-chromosome (not autosomal) DNA led law enforcement to consider him in 2014 in connection with a murder.\textsuperscript{48} Based on evidence that included Mr. Usry’s possible access to the crime scene and production of slasher films, including one titled ‘Murderabilia’, investigators homed in on him as a high-likelihood suspect and obtained a warrant to collect and test his DNA.\textsuperscript{49} Mr. Usry was never arrested because the results of direct STR comparison testing were negative, but it took nearly a month for those results to be returned, which was understandably distressing to him.\textsuperscript{50}

Since Mr. Usry’s ordeal, testing technologies have evolved so that now direct STR comparison testing can usually be performed in less than a week’s time, or if recently approved ‘Rapid DNA’ analysis systems are used, within just a few hours.\textsuperscript{51} In addition, when pursuing a lead (including but not limited to leads generated by IGG, as explained in more detail below), investigators will sometimes collect a DNA sample from a high-likelihood suspect through surreptitious methods, such as from used and discarded soda cans or cigarettes. While we appreciate the privacy concerns that have been raised about surreptitious testing, if that person is innocent, it can mean that they will not experience the stress or embarrassment of knowing that they were ever under suspicion. On the other hand, if that person is guilty, surreptitious testing avoids the possibility that they might flee or destroy evidence upon becoming aware that they are a suspect.

\textsuperscript{45} Id., at 4; see also SACRAMENTO COUNTY MOU, supra note 43, at 2; SWGDAM OVERVIEW, supra note 43, at 6.

\textsuperscript{46} See Yun S. Song et al., Average Probability That a ‘Cold Hit’ in a DNA Database Search Results in an Erroneous Attribution, 54 J. Forensic Sci. 22, 24 (2009). We assume, of course, that its reliability is not one day disproven.

\textsuperscript{47} See Heather Murphy, When a DNA Test Says You’re a Younger Man, Who Lives 5,000 Miles Away, NY. TIMES, Dec. 7, 2019, https://www.nytimes.com/2019/12/07/us/dna-bone-marrow-transplant-crime-lab.html.

\textsuperscript{48} Jim Mustian, New Orleans Filmmaker Cleared in Cold-Case Murder; False Positive Highlights Limitations of Familial DNA Searching, NEW ORLEANS ADVOC, Mar. 12, 2015, https://www.nola.com/article_d58a3d17-c89b-543f-8365-a26197196f0.html.

\textsuperscript{49} Id.

\textsuperscript{50} Id.

\textsuperscript{51} Rapid DNA, FBI.GOV, https://www.fbi.gov/services/laboratory/biometric-analysis/codis/rapid-dna (last visited July 10, 2020).
III.C. Misconception #3: IGG necessarily involves collecting and testing DNA samples from a larger number of innocent persons than would be the case if IGG were not used in the investigation

A common criticism of IGG draws from the fact that it can involve collecting and analyzing DNA from multiple persons, all but one of whom might be guilty of the crime. The argument is that such testing constitutes an invasion of privacy that undermines the ethical acceptability or even lawfulness of IGG. However, when a crime involves DNA evidence, it is not uncommon for investigators to collect and test DNA from individuals who are eventually cleared of suspicion, whether or not those investigations involve IGG. Moreover, one of the benefits of IGG is that, especially in high-profile cases where the perpetrator is unknown to the victim, it potentially can reduce the overall number of such tests that are conducted.

As an initial matter, it is important to understand that DNA collections generally occur in three investigative contexts: (i) voluntarily, with the knowledge and consent of the donor; (ii) involuntarily, with perhaps the knowledge but not the consent of the donor, pursuant to a warrant or as authorized by law; and (iii) involuntarily, without the knowledge or consent of the donor, based on the surreptitious collection of the donor’s abandoned DNA. In the USA, DNA collections in all three contexts are lawful. With respect to surreptitious collections in particular, although regularly challenged, courts have consistently upheld law enforcement’s collection and analysis of ‘shed’ or abandoned DNA under the Fourth Amendment.52 Even states that have adopted statutes prohibiting DNA testing without consent make exceptions for law enforcement purposes.53

DNA collections in all three contexts are routinely performed for the purposes of investigating suspects identified from all kinds of tips, not just leads generated by IGG. Moreover, DNA testing of suspects following surreptitious collection is especially common where the offender is a stranger to the victim and DNA is the primary evidence supporting the association of any suspect to the crime.

What is unique to IGG, however, is the targeted DNA testing of non-suspect family members that sometimes occurs to narrow searches to specific branches of family trees. Specifically, law enforcement might know that certain individuals in the family tree are innocent—for example, they are female and the perpetrator is known to be male—but ask them to participate in SNP testing for the purpose of including or eliminating individuals on specific branches of the family tree. This practice of targeted testing was apparently used with an Oregonian man residing in a nursing home, who voluntarily provided a DNA sample to law enforcement investigating the GSK case in 2017.54 Although it was reported that the man had been misidentified as a suspect,55

---

52 See Ram, supra note 20, at 12; Ellen W. Clayton et al., The Law of Genetic Privacy: Applications, Implications, and Limitations, 6 J. L. BIOSCI. 1, 33–34 (2019); Benjamin E. Berkman et al., Is it Ethical to Use Genealogy Data to Solve Crimes?, 169 ANNAALS INTERNAL MED. 333, 333 (2018); Albert E. Scherr, Genetic Privacy and the Fourth Amendment: Unregulated Surreptitious DNA Harvesting, 47 GA. L. REV. 445, 447–48, 454 (2013).

53 See Katsanis, supra note 2, at 12.16; Clayton et al., supra note 52, at 34.

54 Joe Douglass, ‘Golden State Killer’ Investigator Say Oregon City Man Was Never a Suspect in DNA Probe, KATU News, Apr. 30, 2018, https://katu.com/news/local/golden-state-killer-investigator-says-oregon-city-man-was-never-a-suspect-in-dna-probe.

55 See, e.g., Michael Balsamo, Jonathan J. Cooper & Gillian Flaccus, Police Using Genetic Sites Misidentified Oregon Man as Golden State Serial Kill Suspect in 2017, CHI. TRIB., Apr. 28, 2018, https://www.chicagotribune.com/
Four misconceptions about investigative genetic genealogy

investigators told one news outlet that they never viewed him as a person of interest but sought his DNA to understand if they were focused on the right branch of the family tree.\textsuperscript{56}

While it might seem that targeted testing would increase the total number of DNA collections associated with IGG, we are informed by individuals involved in DNA testing of suspects identified both with and without IGG that, in high-profile cases in particular, this is often not the case.\textsuperscript{57} To understand this point requires an appreciation of the potential scale of the use of DNA testing in a criminal investigation. Especially when a case involves a serial offender who eludes detection for many months or years, the investigation might encompass DNA collections from tens, hundreds, or even thousands of persons of interest who, following direct STR comparison testing, are eliminated as suspects. For example, in the investigation of the South Louisiana Serial Killer during the early days of CODIS, direct STR comparison testing ruled out approximately 2000 suspects before the offender was identified.\textsuperscript{58} DNA ‘dragons’ have also been initiated by law enforcement hoping to identify an offender among discrete populations of individuals who live or work near a crime scene but otherwise have not done anything to arouse suspicion.\textsuperscript{59} The first DNA dragon reportedly occurred in the 1980s when police in the UK collected DNA samples from over 5000 males connected to a village where two girls were raped and murdered.\textsuperscript{60}

IGG can decrease the number of DNA samples that are tested in connection with an investigation by shrinking the pool of individuals from whom such samples are collected. This is true even after accounting for targeted DNA testing because the strategic testing of one individual can eliminate entire branches of the family tree from further consideration—and therefore potentially reduce the overall number of persons tested. For example, in the investigation of the 1979 murder of Michelle Martinko in Iowa, DNA from more than 125 individuals was collected and tested before investigators turned to IGG.\textsuperscript{61} They were able to close the case after conducting targeted testing

\textsuperscript{56} Douglass, \textit{supra} note 54.

\textsuperscript{57} We are not aware of a public dataset that would facilitate a direct comparison of number of DNA samples collected and tested in investigations conducted with and without IGG. Accordingly, this discussion is based on confidential conversations with individuals who conduct or otherwise facilitate IGG as well as publicly available data from specific cases.

\textsuperscript{58} \textit{Derrick Todd Lee-Baton Rouge Serial Killer}, \textit{JUST SCIENCE}, 2017, \url{https://forensiccoe.org/category/season-2/} (season 2, episode 9).

\textsuperscript{59} Aaron B. Chapin, Note, \textit{Arresting DNA: Privacy Expectations of Free Citizens Versus Post-Convicted Persons and the Unconstitutionality of DNA Dragnets}, 89 MINN. L. REV. 1842, 1845–46 (2005).

\textsuperscript{60} Ian Cobain, \textit{Killer Breakthrough—The Day DNA Evidence First Nailed a Murderer}, \textit{THE GUARDIAN}, June 7, 2016, \url{https://www.theguardian.com/uk-news/2016/jun/07/killer-dna-evidence-genetic-profiling-criminal-investigation}.

\textsuperscript{61} Luke Nozicka, \textit{Police Collected DNA From Dozens While Investigating 1979 Killing of Michelle Martinko}, \textit{DES MOINES REG.}, Apr. 15, 2019, \url{https://www.desmoinesregister.com/story/news/crime-and-courts/2019/04/15/michelle-martinko-jerry-burns-suspect-murder-homicide-cedar-rapids-iowa-dna-para bon-labs-cold-case/3476459002/}.
with a few relatives, who led investigators to three brothers, one of whom was arrested after his STR profile made the match. 62

Fewer tests, in turn, can reduce the total cost and length of an investigation, as well as the number of victims when the perpetrator is a serial offender. The GSK, for example, eluded identification by law enforcement using non-IGG techniques for 42 years. During that time, investigators had developed a list of at least 8000 suspects and collected and tested DNA from over 300 of them. 63 Upon uploading the perpetrator’s SNP profile to GEDmatch, it took investigators only 63 days to make an arrest. 64

Still, targeted testing solely for the purpose of including or eliminating branches of the family tree can be unwelcome, and some might view such testing, whether conducted voluntarily, surreptitiously, or with a warrant, as a violation of privacy. Some individuals approached by law enforcement are apparently quite willing to help investigations by providing information about their families or participating in SNP testing. 65 But where offenders are close relatives or law enforcement is not truthful about the purpose of DNA sample collections, reactions to targeted testing have been negative. 66

Asking known innocents for information, personal items, or even biological samples that might be relevant to a crime or otherwise help law enforcement identify the offender—and even concealing the true purpose of these collections when case circumstances suggest it is necessary to do so—is, again, routine police work and not unique to IGG. But we believe that there are opportunities to improve the nature of those interactions to reduce psychosocial harms. Among other things, we support training for investigators who facilitate targeted testing to ensure that communications with donors are respectful and that testing is performed with knowledge and consent when case circumstances do not suggest that the donor’s understanding of the testing purpose would impede the investigation.

Also unique to IGG are the privacy issues associated specifically with surreptitious DNA collections from relatives. The DOJ Interim Policy limits this activity to instances in which investigators believe that requesting a DNA sample ‘would compromise the integrity of the investigation,’ 67 and as a general matter, surreptitious collections are not favored given that they can require more time and resources than voluntary and warrant sample collections. Even if such collections are rare, however, the practice

62 Search Warrant Shows How Relative’s DNA Led Police to Manchester Man in Michelle Martinko’s Death, The Gazette, Mar. 12, 2019, https://www.thegazette.com/subject/news/public-safety/michelle-martinko-dna-search-warrant-straw-jerry-lynn-burns-family-tree-cold-case-murder-20190312.
63 Anh Do & Matt Hamilton, Thousands of Suspects, Many Dead Ends; The Quest for the Golden State Killer, L.A. Times, Apr. 26, 2018, https://www.latimes.com/local/lanow/la-me-ln-golden-state-killer-centerpiece-20180426-story.html; Steve Kramer, FBI Case Studies and Its Application in Solving Cold Cases, International Symposium on Human Identification (Sept. 16, 2020).
64 David Corriveau, Genealogist Who Helped Solve Allenstown Four Cold Case Speaks at Dartmouth, Valley News, May 2, 2019, https://www.vnews.com/Former-patent-lawyer-turned-DNA-crime-solver-discusses-method-at-DHMC-25276977.
65 Jon Schuppe, ‘They Lied to Us’: Mom Says Police Deceived Her to Get Her DNA and Charge Her Son with Murder, NBC News, Feb. 22, 2020, https://www.nbcnews.com/news/us-news/they-lied-us-mom-says-police-deceived-her-get-her-n1140696.
66 Id.
67 DOJ INTERIM POLICY, supra note 42, at 6. In addition, both the investigative agency and prosecutor must agree with the decision to proceed with surreptitious collection before it can be undertaken. Id.
warrants ethical and legal scrutiny. To provide empirical insight into these and related concerns, we support the collection and public reporting of information relevant to DNA sampling in criminal investigations conducted with and without IGG.

Finally, concerns have long been raised about the retention of DNA in informal genetic databases not subject to regulatory oversight. While these concerns are not unique to investigations involving IGG, they are especially salient in such investigations given that the conduct of IGG might encompass DNA collections from relatives who are known innocents. Although the DOJ Interim Policy provides for the destruction of all reference samples, it might be necessary for policymakers to intervene on this specific issue to ensure uniformity of practices for the protection of these third parties.

III.D. Misconception #4: IGG is or soon will be ubiquitous because there are no barriers to IGG that limit the cases in which it can be conducted

Finally, some have described IGG as a dystopian tool for state surveillance. The suggestion is that genetic genealogy databases will soon be, if not already are, consulted by law enforcement agencies in the investigation of every crime, from serial murder to petty theft. Although we share concerns about ‘scope creep,’ there are currently a number of technical and practical barriers to IGG, as well as emerging policy barriers, that serve to temper them.

First, the genealogy step of IGG—that is, the process of building out family trees based on genetic relationships—serves as a significant practical barrier to the widespread conduct of IGG because it is expensive and requires specialized support. Specifically, the genealogy step must be conducted by individuals skilled in genetic genealogy and can take months to complete, depending on the complexity of the relevant family tree, sometimes without resulting in a useful lead. Factors that increase the complexity of this work include the number and genetic distance of identified relatives, immigration events, misattributed parentage, unrecorded adoption, surname changes, endogamy, and pedigree collapse resulting from, for example, incest.

The number and genetic distance of identified relatives, in particular, are dependent on the total number, as well as the genetic ancestries, of database participants who opt in to law enforcement matching. To date, more than 96% of US-based SNP profiles are opted in to law enforcement matching in the FTDNA database, which adopted an opt-in default for participants in 2019. Conversely, approximately 30% of public SNP profiles are currently opted in to law enforcement matching in GEDmatch, which opted out all participants in 2019. Although 83% of new GEDmatch users are

---

68 See, e.g., Erin Murphy, Forensic DNA Typing, 1 Ann. Rev. Criminol. 497 (2018).
69 DOJ INTERIM POLICY, supra note 42, at 8. In cases that result in criminal prosecution, the policy provides for destruction of reference samples upon entry of a court order. Id. Otherwise, samples must be destroyed “after their investigative use is complete.” Id.
70 Greytak et al., supra note 30, at 107–08.
71 Id.
72 Personal communication with C. Bormans, Gene by Gene, with author (C.J.G.) (Apr. 28, 2020).
73 Personal communication with C. Rogers & B. Williams, supra note 9; Kumar & Campbell, supra note 9.
74 See Schuppe, supra note 8 and accompanying text. As noted earlier, however, new GEDmatch participants are required to decide at the time of registration whether or not they will opt in to or out of law enforcement matching.
opting in to law enforcement matching,\textsuperscript{75} it is well-known that the overall reduction in relative matches available to law enforcement on GEDmatch has impeded the ability to successfully conduct IGG. When the offender is believed to be of largely non-European ancestry, agencies might be reluctant to devote scarce resources on IGG given that most of the databases’ participants are presumed to be of European ancestry.\textsuperscript{76} Further, if the offender is believed to have recently immigrated to the USA, agencies might decline to conduct IGG given the difficulty of reconstructing family trees using foreign records in addition to the reduced likelihood of finding a close relative match for demographic reasons.

But even if an agency has the resources to conduct IGG in a particular case and the chances of finding a genetic relative of the offender are high given their presumed genetic ancestry, it still might not be possible to conduct IGG because the quantity or quality of the forensic sample does not meet minimum standards to generate a database-compatible, high-density SNP profile. At this point, the instruments that generate SNP profiles generally require at least 20 ng of DNA to produce a profile, although laboratories have produced profiles based on 1 ng of DNA or less.\textsuperscript{77} Where the quantity of DNA is sufficient, success might still be impeded by other factors, including the extent of degradation of the DNA; the source of the DNA, where SNP extraction is generally more successful when performed on semen than blood or bones; and where the sample is a mixture (i.e., it contains the DNA of more than one person), the proportions of DNA in the mixture and whether reference samples are available for non-suspect contributors.\textsuperscript{78} Thus, it might be possible to generate an IGG-eligible SNP profile from 5 ng of DNA extracted from fresh, single-source semen, but not from a 5-year-old blood mixture, where the offender’s blood accounts for 30% of the mixture. While some samples that are currently not candidates for SNP extraction might later become candidates as the technology advances, for at least the near future, these technical issues are an important hurdle to IGG.

Finally, the DOJ Interim Policy, as well as best practices endorsed by the California District Attorney who prosecuted the GSK case and the Scientific Working Group on DNA Analysis Methods, which advises the Director of the Federal Bureau of Investigation, limit the conduct of IGG to the identification of human remains and investigation of violent crimes, where reasonable investigative leads have already been pursued.\textsuperscript{79} In addition, as discussed above, investigators cannot conduct IGG until they have first conducted a CODIS search and that search failed to produce a probative and

\textsuperscript{75} Kumar & Campbell, supra note 9.

\textsuperscript{76} Greytak et al., supra note 30, at 107. In 2019, Parabon NanoLabs, which provides IGG services to law enforcement and has consulted on hundreds of cases, reported that it had evaluated 40% of cases involving suspects of non-European ancestry as not solvable. Id. at 107–108.

\textsuperscript{77} Id. at 103–04. One nanogram is one billionth of a gram.

\textsuperscript{78} Id.

\textsuperscript{79} DOJ Interim Policy, supra note 42, at 4–5; SACRAMENTO COUNTY MOU, supra note 43, at 1; SWGDAM OVERVIEW, supra note 43, at 6. The SWGDAM statement allows for the conduct of IGG while viable leads are being pursued where there is a significant public safety threat. SWGDAM OVERVIEW, supra note 43, at 6. Guidelines by the American Society of Crime Laboratory Directors limit the use of IGG to sexual assault and homicide cases. AMERICAN SOCIETY OF CRIME LABORATORY DIRECTORS, INVESTIGATIVE GENETIC GENEALOGY (Oct. 18, 2019), https://www.ascld.org/wp-content/uploads/2019/10/ASCLD-IGG-10.18.19.pdf [hereinafter ASCLD GUIDELINES].
confirmed match. 80 These policies serve to limit the cases in which IGG is conducted to those that are considered public safety and criminal justice priorities and where other investigative tactics have been attempted but failed.

Of course, these barriers are not static. While the genealogy work associated with IGG is arduous and will likely continue to require specialized expertise for some time, advances in DNA extraction and testing should improve the ability to successfully generate SNP profiles from forensic samples. In addition, the policy landscape will likely change over time. As IGG evolves to encompass whole-genome sequencing data, 81 and as databases modify their tools and services, 82 some existing barriers will disappear as new ones emerge. Regardless of these changes, genetic genealogy database participants are themselves the ultimate check on the indiscriminate conduct of IGG. If they disagree with changes to how law enforcement conducts IGG, including its participation in databases that have adopted terms prohibiting IGG, they are free to withdraw their data. The mass exodus, in turn, will render the databases of limited value to law enforcement given that the success of IGG depends on robust participation. For this reason, the conduct of IGG in ways that participants perceive to be unprincipled or unbridled is not in law enforcement’s interests.

IV. LOOKING AHEAD

By clarifying the steps of IGG, including how it does and does not depart from traditional investigations, we hope to help distinguish concerns about IGG that are based on misinformation from those that are grounded in how it is currently conducted and its results are used in practice. To be clear, we have concerns associated with certain practices associated with IGG. In addition to those already described, we view as ethically troubling the practices of conducting IGG in genetic genealogy databases against their terms of service and identifying genetic relatives by name in search warrants. We support public conversations about these and other concerns and believe that they are necessary for the development of best practices and policies for IGG.

However, to prevent misunderstanding, we believe those conversations would benefit from more thoughtful use of language and understanding of the relevant technologies. For example, we favor descriptions of IGG that state the information provided to law enforcement is a list of genetic relatives of an offender among genetic genealogy database participants, and their estimated relatedness to the offender, not participants’ DNA. Although the latter makes for shorter (and more attention-grabbing) headlines, it is incorrect and promotes confusion. We also favor referring to those identified by IGG as genetic relatives and not as genetic informants or witnesses, which might wrongly suggest that genetic relatedness is itself evidence of criminal wrongdoing. Third, targeted testing of known innocents to help zero in on specific

80 DOJ INTERIM POLICY, supra note 42, at 5; see also SACRAMENTO COUNTY MOU, supra note 43, at 1–2; SWGDAM OVERVIEW, supra note 43, at 6; ASCLD GUIDELINES, supra note 79.

81 Andreas Tillmar et al., Whole-Genome Sequencing of Human Remains to Enable Genealogy DNA Database Searches—A Case Report, 46 FORENSIC SCI. INT’L GENETICS 102233 (2020).

82 For example, Verogen has announced that it will launch a new portal called ‘GEDmatch Pro’ for law enforcement to upload forensic SNP profiles. Brett Williams & Nicola Oldroyd Clark, Reimagining Human Identification, VEROGEN (2020), https://info.verogen.com/reimagining-human-identification-watch?submissionGuid=ee61d2cb-78c5-42ff-8b80-d0cb66a5dd28.
family tree branches should not be confused with direct STR comparison testing of high-likelihood suspects who are eliminated from investigation when a confirmed match is not made to a forensic sample. These are different tests conducted for different purposes, but neither is a case of mistaken identity. Fourth, identifying genetic relatives of the source of a SNP profile among genetic genealogy database participants is not the same thing as identifying the source. Relative matching is an automated function performed by genetic genealogy databases for participants opting in to this service. Identifying an unknown source with the help of relative match information is an investigative activity performed by persons skilled in genetic genealogy, can continue for months, and is not always successful. Because IGG necessarily encompasses both activities, conducting IGG should not be analogized to conducting a CODIS search.

Finally, we believe that going forward, it will be useful to distinguish concerns about IGG that are unique to its practice from those that are not. As one example, criticisms of IGG often include privacy concerns related to surreptitious DNA collection and testing. But that practice began many years before IGG, and as explained above, it is routinely used today in unsolved cases that do not involve IGG. This is not meant to suggest that privacy concerns related to surreptitious DNA collection and testing are unfounded. Indeed, these concerns as they relate to targeted testing in particular are compelling. Rather, the point is that they would be more effectively addressed at a policy level far broader than IGG given that even a complete ban on IGG would not eliminate this practice.

Looking ahead, we hope that any policies that are adopted relevant to IGG take into account at least these points. Policies that reflect up-to-date information are needed to navigate the important individual and public interests at stake.

ACKNOWLEDGEMENTS
Research conducted by C.J.G., A.L.M., and S.M.F. was supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number R01HG011268. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health, the authors’ employers, or any institutions with which they are or have been affiliated. The authors wish to thank Whitney Bash Brooks for research support.

CONFLICT OF INTEREST
C.J.G., R.A.W., B.B., and S.M.F. are members of the Investigative Genetic Genealogy Working Group of the Scientific Working Group on DNA Analysis Methods. A.L.M. is a former member of the FamilyTreeDNA Citizen’s Panel.