



Overview of Investigative Genetic Genealogy

This overview describes the technique known as investigative genetic genealogy¹ and distinguishes it from the investigative use of law enforcement DNA databases, such as those in the Combined DNA Index System (CODIS). The overview also identifies issues for agencies considering the use of investigative genetic genealogy in their investigations.

Investigative genetic genealogy refers to the DNA analysis of a crime scene sample or unidentified human remains (UHR) sample (both of which are referred to as a sample from a “person of interest”), using a high density single nucleotide polymorphism (SNP) array² to generate genetic data that is used to search a third-party³ SNP database for genetic relatives of the person of interest. The potential genetic relatives identified by that search may then be researched by genealogists and law enforcement using genealogical and other available records to assess possible genealogical relationships to the person of interest. The goal of the SNP database search and genealogical research is to provide leads to identify the person of interest. There are several points during this process at which law enforcement intersects with vendor laboratories and third-party SNP databases. The steps of this process⁴, some of which are described in more detail below, typically include the following:

- Collection of a probative biological sample at a crime scene or collection of a UHR sample;
- Development of a short tandem repeat (STR) DNA profile (consisting of the CODIS Core Loci) from the collected sample;
- Pursuit of all viable investigative leads, including no matches resulting from a search of CODIS using the STR DNA profile;

¹ For the purposes of this document, the SWGDAM Committee of Correspondence on Forensic Genealogy is using the term investigative genetic genealogy although the following additional terms are being used to describe this technique: forensic genealogy, forensic genetic genealogy, forensic genetic genealogical DNA analysis and searching, genetic genealogy and investigative genealogy, see, for example, U.S. Department of Justice Interim Policy Forensic Genetic Genealogical DNA Analysis and Searching, Effective 11/01/2019, available at <https://www.justice.gov/olp/page/file/1204386/download> (last access date February 3, 2020).

² In addition to SNP arrays, other forms of high-density genotyping, including next-generation whole exome and/or whole genome sequencing, can generate SNP profiles of equivalent information content to permit genealogical analysis.

³ Third-party refers to non-governmental, non-forensic SNP databases such as GEDmatch and FamilyTreeDNA.

⁴ See, generally, Wickenheiser, R.A., *Forensic genealogy, bioethics, and the Golden State Killer case*, *For. Sci. Int'l: Synergy* 1:114-125 (2019); available at <https://www.sciencedirect.com/science/article/pii/S2589871X19301342> (last access date February 3, 2020).

- Development of genomewide SNP data from the collected sample;
- Search of one or more third-party SNP databases using the SNP data to identify potential genetic relatives in the database;
- Assessment of possible genealogical relationships between the potential genetic relatives and the person of interest;
- Investigation of leads generated by the research to identify the person of interest; and
- Obtaining DNA from the person of interest for development of an STR DNA profile in order to perform a one-to-one comparison to the crime scene STR DNA profile for exclusionary/inclusionary purposes⁵; or collection of DNA from potential relatives for development of an STR, YSTR or SNP profile, or mitochondrial DNA sequence, in order to perform a kinship analysis comparison with the UHR.

Most third-party SNP databases are maintained by direct-to-consumer (DTC) genetic testing providers. Currently, the providers that have the largest databases are Ancestry, FamilyTreeDNA, MyHeritage, and 23andMe.⁶ These providers generate genetic data from customers' DNA samples using SNP microarrays that produce between 600,000 to 700,000 SNPs.⁷ They then analyze certain markers in these SNPs for a range of purposes. Depending upon the provider, these purposes might include reporting to customers their biogeographical ancestry information and identifying potential genetic relatives for customers who elect to participate in relative matching services. They might also include reporting to customers their health, wellness, trait conditions or predispositions.⁸

Many DTC genetic testing providers maintain their customers' SNP data in a database. However, the providers with the largest databases also allow the removal of a customer's data, at any time, upon request. In addition, many DTC genetic testing providers permit customers to retrieve their data to personally maintain, control and share their SNP file. Individuals can share their SNP file with researchers and third-party services, such as GEDmatch, that offer to interpret their SNP data.⁹ GEDmatch does not provide genetic testing services but instead provides a central location for users to upload and share their SNP file and provides tools to users to help

⁵ The Marshall Project, *In an Apparent First, Genetic Genealogy Aids a Wrongful Conviction Case*, July 16, 2019; available at <https://www.themarshallproject.org/2019/07/16/in-an-apparent-first-genetic-genealogy-aids-a-wrongful-conviction-case> (last access date February 3, 2020).

⁶ Larkin L., Database sizes—September 2018 update. The DNA Geek. <https://thednageek.com/database-sizes-september-2018-update/> (last access date February 3, 2020).

⁷ International Society of Genetic Genealogy Wiki, Autosomal DNA Testing Comparison Chart available at https://isogg.org/wiki/Autosomal_DNA_testing_comparison_chart (last access date February 3, 2020).

⁸ For more information on the use of phenotyping for law enforcement identification purposes, see Samuel, G., Prainsack, B. (2018) *The regulatory landscape of forensic DNA phenotyping in Europe*, VISAGE; available at http://www.visage-h2020.eu/Report_regulatory_landscape_FDP_in_Europe2.pdf (last access date February 3, 2020).

⁹ Nelson, S.C., Fullerton, S.M. "Bridge to the literature"? Third-party genetic interpretation tools and the views of tool developers. *J Genet Counsel.* 27:770-781 (2018).

them identify possible genetic relatives among other users in the database.¹⁰ GEDmatch allows participation in its database and use of many of its tools for free.

Relative matching services offered by DTC genetic testing providers and GEDmatch function by ranking and reporting a user's relative matches according to how many centimorgans (cM) are shared between the user's SNP data and each potential genetic relative's SNP data. A cM is a measure of shared DNA segments or genetic linkage. The higher the number of shared cM, the closer is the potential familial relationship.¹¹ Upon receiving this information, the user (or one or more genetic genealogists working on behalf of the user) can evaluate the genetic results and review/construct family trees in an effort to determine genealogical relationships and identify other relatives. In some instances, relatives identified in a family tree might volunteer DNA samples for analysis to support or eliminate genetic relationships.

Participation in genetic databases maintained by DTC genetic testing and third-party services such as GEDmatch are typically governed by some combination of terms-of-service policies, privacy policies, and informed consent agreements. These policies vary by provider,¹² but they generally inform users how their genetic information will and potentially might be used, shared, and searched. Providers often post their policies online, and users are required to accept (or are deemed to have accepted) the policies upon purchase of the services, registration with the provider, or use of services. Providers expect that users will comply with their policies, and users expect that other users will comply with the policies as well.¹³

Currently, two providers – FamilyTreeDNA and GEDmatch – allow law enforcement to search their genetic databases for investigative purposes under the same terms as other users^{14,15} with the important caveat that their policies are dynamic and subject to change. At this time both providers permit law enforcement to use their databases to assist in identifying UHRs. Additionally, FamilyTreeDNA allows law enforcement to use its database to assist in identifying perpetrators of homicide and sexual assault, whereas GEDmatch permits law enforcement to use its database to assist in identifying perpetrators of violent crime, which GEDmatch defines as murder, non-negligent manslaughter, aggravated rape, robbery, and aggravated assault.

Both providers have addressed individual genetic privacy by allowing users of their genetic databases to choose to exclude their SNP data from law enforcement searches. FamilyTreeDNA allows users to opt out of law enforcement searches, whereas GEDmatch users

¹⁰ GEDmatch.com Terms of Service and Privacy Policy at <https://www.gedmatch.com/tos.htm> (last access date February 3, 2020).

¹¹ See, generally, Bettinger, B., *The Genetic Genealogist*, August 2017 Update to the Shared cM Project, available at <https://thegeneticgenealogist.com/2017/08/26/august-2017-update-to-the-shared-cm-project/> (last access date February 3, 2020).

¹² See generally Hazel J, Slobogin C. *Who knows what, and when?: A survey of the privacy policies proffered by U.S. direct-to-consumer genetic testing companies*. Cornell J Law Pub Policy. 28:35-66 (2018). See also <https://www.familytreedna.com/legal/privacy-statement>; <https://www.23andme.com/about/privacy/>; <https://www.gedmatch.com/tos.htm> (last access date February 3, 2020).

¹³ See, for example, Kennett, D., *Using genetic genealogy databases in missing person cases and to develop suspect leads in violent crimes*, For. Sci. Int'l. 301:107-117 (2019).

¹⁴ See <https://www.familytreedna.com/legal/privacy-statement> at "5. E. Law Enforcement Matching".

¹⁵ See <https://www.gedmatch.com/tos.htm> at "Raw DNA Data Provided to GEDmatch" and "DNA Data".

are automatically opted out of law enforcement searches but may choose to opt in to such searches. However, individual privacy rights may be implicated in other ways, such as when law enforcement approaches a person to obtain that person's DNA for testing, knowing that the person is not the suspected perpetrator but appears to be a relative of a person of interest based on a search of a SNP database.¹⁶ The SNP analysis of possible relatives of the person of interest is sometimes referred to as "targeted collection" or "targeted analysis." Other issues relating to privacy may include: the terms of consent for knowingly providing a DNA sample for further testing if identified on a family tree relating to a person of interest; or, so as not to disclose the interest of law enforcement in an individual, the surreptitious collection of an abandoned DNA sample(s) from a relative of a person of interest.

Generally, law enforcement agencies that currently employ investigative genetic genealogy are using this technique on cold cases or active, unsolved violent crimes.¹⁷ When operating in collaboration with a forensic laboratory, the law enforcement agency will first send the sample from the person of interest to the forensic laboratory for analysis and searching of any resulting STR DNA profile in their state and the national DNA databases. If the search of the STR DNA profile within the law enforcement DNA databases does not produce any investigative leads, the law enforcement agency or forensic laboratory may decide to employ investigative genetic genealogy in accordance with applicable agency protocols. This may include a review to ensure that there is sufficient crime scene sample and/or DNA extract to proceed with this technique as well as consultation with the prosecutor's office to obtain approval (and to ensure that any reasonable leads will be pursued). Once approval to proceed is obtained, the law enforcement agency and/or laboratory will transmit the crime scene sample and/or DNA extract to a commercial genetic laboratory that will analyze the sample/extract using SNP technology.

The resulting SNP data are either provided to the law enforcement agency for searching in third-party SNP databases that permit it, or a law enforcement agency may contract out the genealogy services and searching of the SNP data to a genetic genealogist or commercial laboratory. The search of the third-party SNP database returns a list of the potential genetic relative matches ranked by the amount of shared DNA. Once the genetic search results are obtained, the genetic genealogist will assess the genetic distances or relationships between matches and, based upon the amount of DNA shared among the matches, construct and review family trees. Working in conjunction with the law enforcement agency, other non-genetic investigative databases and research can be used to eliminate individuals based on sex, age, location, and other information to generate the best investigative leads. As an additional step, when a likely person of interest is identified, the law enforcement agency will obtain a DNA sample from the person identified with the crime scene for forensic STR analysis for exclusionary/inclusionary purposes. In UHR investigations, the law enforcement agency will

¹⁶ Kennett, D., *Using genetic genealogy databases in missing person cases and to develop suspect leads in violent crimes*, *For. Sci. Int'l* 301:107-117 (2019).

¹⁷ Greytak, E., Moore, C., Armentrout, S., *Genetic Genealogy for cold case and active investigations*, *For. Sci. Int'l* 299:103-113 (2019).

obtain a DNA sample from the relative(s) of the person of interest for forensic DNA analysis for exclusionary/inclusionary purposes.

In contrast to these genealogy services and databases, law enforcement DNA databases are only accessible to criminal justice agencies through their forensic DNA laboratories. The DNA samples are analyzed for designated law enforcement identification markers (20 STR loci known as the CODIS Core Loci) that have been specifically selected because they are not predictive of medical conditions or disease status.¹⁸ DNA records in the law enforcement DNA databases have been collected, analyzed and databased in accordance with applicable state and federal laws that specify offenders and, in some cases, arrestees that are required to provide DNA samples. For UHR investigations, DNA samples from family members of missing persons can be voluntarily submitted and the profiles entered into a separate index of the law enforcement databases.¹⁹

Using a law enforcement DNA database, the crime scene or UHR DNA profiles are searched against the offender/arrestee DNA profiles to generate candidate matches which are reviewed and as appropriate, confirmed. The confirmation process requires the verification of the offender's biographical information and, in many laboratories, a reanalysis of the offender's/arrestee's DNA sample and direct comparison to the crime scene or UHR profile as an additional quality assurance measure. It is only following confirmation that the forensic DNA laboratory releases the name of the person of interest to the law enforcement agency that submitted the crime scene evidence or UHR sample.²⁰ At present, most forensic DNA laboratories do not have the capability to develop the SNP data needed for genealogy database searching purposes. Accordingly, law enforcement agencies will typically outsource the analysis of the crime scene or UHR sample to a laboratory to obtain SNP genetic data to search in the third-party SNP database.

Although some commentators have compared investigative genetic genealogy to forensic familial DNA searching,²¹ familial searching is an additional search of a law enforcement DNA database which is conducted after a routine search of the crime scene evidence DNA profile has been completed with no resulting matches. Whereas both approaches involve working with non-genetic investigative databases, investigative genetic genealogy begins with a search of a third-party SNP database. A forensic familial search is performed using the STR DNA profile, and with a different search algorithm, for the purpose of identifying close biological relatives of the unknown DNA profile developed from the crime scene sample. Familial searching can create privacy concerns that are similar to those associated with investigative genetic genealogy. In consideration of these privacy concerns, jurisdictions employing familial searching have

¹⁸ See <https://www.fbi.gov/services/laboratory/biometric-analysis/codis> (last access date February 3, 2020) at "Planned Process and Timeline for Implementation of Additional CODIS Core Loci"; Hares, D.R., *Selection and implementation of expanded CODIS core loci in the United States*. For. Sci. Int'l Genetics 17: 33-34 (2015).

¹⁹ These family reference profiles are only searched against UHR profiles and are prohibited from being searched against unknown crime scene profiles.

²⁰ *Ibid.* at "Frequently Asked Questions on CODIS and NDIS"

²¹ Murphy, E., *Law and policy oversight of familial searches in recreational genealogy databases*, For. Sci. Int'l 292: e5-e9 (2018); available at <https://doi.org/10.1016/j.forsciint.2018.08.027> (last access date February 3, 2020).

generally established the following threshold criteria for the use of this searching tool: (1) the search is used for single source crime scene DNA profiles; (2) the search is limited to serious violent felony offense cases such as homicide, sexual assault, and kidnapping; (3) all other investigative leads must have been exhausted; and (4) the prosecutor must agree to pursue investigative lead information in the case.²²

Law enforcement agencies exploring the use of investigative genetic genealogy should include the relevant legal authority representative in their discussions and consider the following:

- A search of CODIS should be performed first to ensure that there are no matches in that state's or the national law enforcement DNA database.
- Policies/procedures should be established which consider applicable privacy policies and the database provider's terms of service, a level of transparency of techniques employed, and maintenance of the public trust.
- Keeping in mind proportionality²³, only serious violent felonies (such as homicide, sexual assault and aggravated assault) and identification of human remains should be considered for the use of investigative genetic genealogy.
- Investigative genetic genealogy should be used only after viable investigative leads have been pursued or there is a significant public safety threat warranting use of investigative genetic genealogy while viable leads are being pursued.
- Law enforcement should process investigative genetic genealogy cases subject to the approval of, and in consultation with, applicable legal representatives, such as the agency legal counsel or prosecutor.
- DNA samples selected for investigative genetic genealogy must be from a single source or deduced²⁴ single source and attributable to the person of interest.
- Once a person of interest is identified following the use of investigative genetic genealogy, law enforcement should obtain a reference sample for STR DNA analysis for exclusionary/inclusionary purposes.
- Once a person of interest is included by STR DNA analysis, surreptitious or abandoned collections of targeted reference samples should be expunged from the third-party SNP database(s).

²² See <https://www.fbi.gov/services/laboratory/biometric-analysis/codis> at "Familial Searching"; see also Field, M.B, et al., *Study of Familial DNA Searching Policies and Practices: Case Study Brief Series*, August 2017 available at <https://www.ncjrs.gov/pdffiles1/nij/grants/251081.pdf> (last access date February 3, 2020); BFS DNA Frequently Asked Questions, California's Familial Search Policy available at <https://oag.ca.gov/bfs/prop69/faqs#familial> (last access date February 3, 2020). See also, Recommendations from the SWGDAM Ad Hoc Working Group on Familial Searching at http://media.wix.com/ugd/4344b0_46b5263cab994f16aeedb01419f964f6.pdf (last access date February 3, 2020).

²³ See, generally, Guerrini, C. J., Robinson, J. O., Peterson, D., McGuire, A.L., *Should police have access to genetic databases? Capturing the Golden State Killer and other criminals using a controversial new forensic technique*, PLoS Biol. 16(10) (2018); available at <https://doi.org/10.1371/journal.pbio.2006906> (last access date February 3, 2020); Wickenheiser, R.A., *Forensic genealogy, bioethics and the Golden State Killer case*, For. Sci. Int'l: Synergy 1:114-125 (2019) available at <https://doi.org/10.1016/j.fsisyn.2019.07.003>.

²⁴ Deduced single source profiles are obtained by mixture deconvolution for which the perpetrator can be readily determined from the sample (e.g., rape kit for which there are victim and/or elimination standards available).

Approved by the Scientific Working Group on DNA Analysis Methods --February 18, 2020

- Once a person of interest is included by STR DNA analysis, the crime scene or UHR sample should be expunged from the third-party SNP database(s).
- In jurisdictions where authorized, forensic familial searching should also be considered in accordance with the jurisdiction's policies/procedures.
- Prioritization of resources should be evaluated for best allocation of funding/personnel.

*Approved by the Scientific Working Group on DNA Analysis Methods
February 18, 2020*