



Forensic Investigative Genetic Genealogy and the Privacy Questions Facing Crime Laboratories

Sona Mahlangu^{1*}, Thabo Subramaniam², Agnieszka Gumede³

¹Department of Science, Pannonian Institute for Advanced Science, Greece

²Department of Science, Rhone Research Institute of Science, South Africa

³Department of Science, Limpopo Graduate Institute of Science, Uzbekistan

Article Info

Volume 1, Issue 1, 2026

[doi: 10.62715/ijfs.2026.v1.i1.pp8-13](https://doi.org/10.62715/ijfs.2026.v1.i1.pp8-13)

Received: 23 November 2025

Accepted: 16 January 2026

Published: 12 March 2026

Abstract

Forensic investigative genetic genealogy combines dense single nucleotide polymorphism data from a crime scene sample with the family trees of people who have uploaded their own data to genealogy databases. Since 2018 it has produced investigative leads in cases that conventional database searching had left unsolved, and it has drawn sustained criticism on privacy grounds. Most of that debate has been conducted in terms of police powers and the rights of database users. This Perspective argues that the crime laboratory is a neglected party to it. The laboratory decides whether a sample is suitable, generates or commissions the genotype file, chooses the vendor, and holds the residual extract and data afterwards. Each of those steps raises a privacy question that the laboratory, and not the investigator or the genealogist, is best placed to answer. The article sets out those questions, considers the arguments that privacy concerns are overstated or belong to others, and proposes that laboratories treat genealogy work as a distinct service with its own acceptance criteria, data rules and audit trail. It reports no new data and describes no individual case in detail.

Keywords: *Investigative genetic genealogy, Genetic privacy, Crime laboratory policy, SNP genotyping, Familial searching, Forensic ethics*

© 2026 Sona Mahlangu et al. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1. Introduction

In April 2018 investigators in California announced an arrest in the Golden State Killer case after comparing a crime scene profile with data in a public genealogy database and building family trees from the partial matches. The method was not secret, but it was new to most of the forensic community (Phillips, 2018). Within a few years it had been used in many investigations in the United States and had been taken up in Europe. The technique is now usually called forensic investigative genetic genealogy, and its methods have been reviewed in detail (Greytak et al., 2019; Kling et al., 2021).

Commentary followed quickly. Legal scholars asked whether police should have access to databases built for family history, and on what terms (Ram et al., 2018; Murphy, 2018). Survey work suggested that public support was broad for serious violent crime and weaker for lesser offences (Guerrini et al., 2018).

* Corresponding author: Sona Mahlangu, Department of Science, Pannonian Institute for Advanced Science, Greece. E-mail: sona.mahlangu@pannonian.edu

© 2026 Sona Mahlangu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The debate has largely been framed as a contest between the investigator who wants a lead and the database user who did not expect to become a genetic informant.

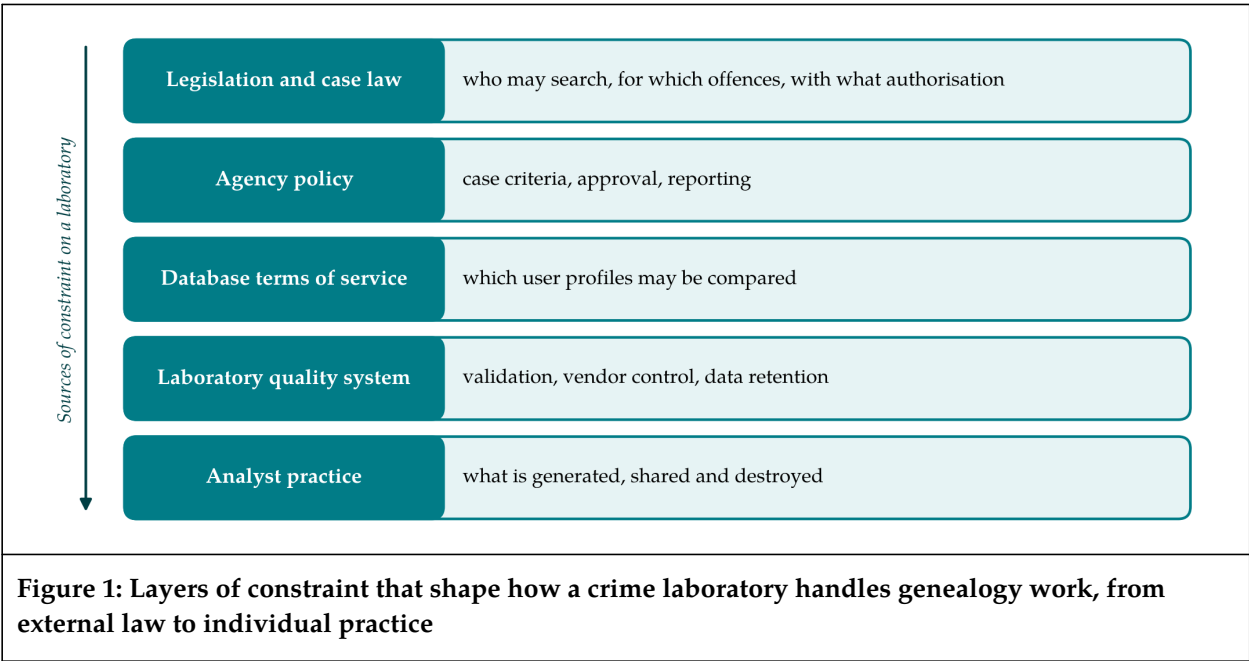
This Perspective takes a different vantage point. It asks what the privacy debate means for the crime laboratory, the institution that holds the evidence, produces or commissions the genetic data and answers to an accreditation body for what is done with them. The argument is that laboratories face privacy questions of their own, distinct from those facing police, and that they are better served by answering them openly than by treating genealogy as somebody else's business.

2. How the Technique Departs From Routine Casework

Conventional forensic DNA typing examines a small set of short tandem repeat loci chosen for identification and compares the result with a government database of profiles from convicted offenders, arrestees and crime scenes. The loci were selected to reveal little beyond identity, the database is created by statute, and its use is restricted. When the United States Supreme Court upheld the taking of DNA from arrestees it relied in part on those features ([Supreme Court of the United States, 2013](#)). Familial searching of such databases, which looks for close relatives of an unknown offender, was already controversial before genealogy arrived ([Bieber et al., 2006](#); [Murphy, 2010](#)).

Genealogy changes each element. The data are hundreds of thousands of single nucleotide polymorphisms, generated by microarray or by whole genome sequencing. The comparison is made against profiles that private individuals have uploaded to commercial or volunteer-run services. The matches of interest are distant ones, third cousins and beyond, and the route from match to suspect runs through public records, obituaries and social media. [Erlich et al. \(2018\)](#) showed that a database covering only a small percentage of a population is enough to give most members of that population a detectable relative in it. Reach therefore extends far beyond the people who chose to take part.

For the laboratory the departure is practical as well as conceptual. Few public laboratories run dense SNP assays in-house, so the extract or the remaining evidence is commonly sent to a private vendor. The genotype file that comes back is a new kind of record, richer than any profile the laboratory has held before, and the genealogical research that follows is done by people outside the quality system. Figure 1 sets out the layers of constraint that bear on a laboratory in this position, from legislation down to the conduct of the individual analyst.



3. Whose Privacy Is Engaged

Four groups have interests at stake, and the laboratory touches each of them. The first is the database user whose profile is compared. Whether that person agreed to the comparison depends on the terms of service in force and on the settings chosen, both of which have changed over time. The second and much larger group is the user's relatives, who appear in the resulting family tree without having been asked. Consent by one member of a family cannot sensibly be treated as consent by all, and the arithmetic of [Erlich et al. \(2018\)](#) means that the unconsulted greatly outnumber the consenting.

The third group is the people investigated and cleared. A genealogical lead typically narrows the field to a branch of a family, and several individuals may be looked at, and sometimes sampled, before one is confirmed or all are excluded. The fourth is the donor of the crime scene sample. In most cases that is the offender, whose claim to privacy in abandoned biological material is weak. In others it is a victim, an unidentified decedent or a person with an innocent reason for leaving DNA at the scene, and the SNP file generated from the sample can disclose ancestry and health-related variants as well as kinship.

Security research has added a further concern. Databases that let users upload raw files and see matching segments can be probed with crafted uploads to recover the genotypes of other users ([Edge and Coop, 2020](#)). A laboratory that uploads a casework file to such a service is placing evidential data in an environment whose safeguards it does not control.

4. The Questions That Fall to the Laboratory

Table 1 lists the decisions in a genealogy case that sit with the laboratory and the privacy question each one carries. None of them is answered by a policy addressed to investigators alone.

Table 1: Laboratory decision points in investigative genetic genealogy and the privacy question attached to each		
Decision point	Privacy question	Possible laboratory control
Case acceptance	Is the intrusion proportionate to the offence and the state of the inquiry?	Written criteria; prior STR typing and database search
Sample selection	Is the sample single-source and attributable to the offender?	Documented assessment before any extract is released
Choice of vendor	Who else will hold the extract and the genotype file?	Contract terms on security, reuse, return and destruction
Markers generated	Does the assay reveal more than kinship requires?	Restrict analysis and reporting to kinship markers where possible
Database upload	Have the users being compared agreed to law enforcement matching?	Use only services whose terms permit it; identify the upload as such
Confirmation	Are relatives or cleared persons sampled, and how?	Direct STR comparison with a lawfully obtained reference
Retention	How long are the SNP file and match lists kept, and by whom?	Retention schedule; removal of the profile when the case closes

The first decision is whether to take the case at all. The interim policy issued by the [US Department of Justice \(2019\)](#) confines the technique, for agencies within its reach, to violent crimes and unidentified human remains, requires that a conventional profile has been searched against the national database without success, and requires that investigators identify themselves to the genealogy service. A laboratory can adopt comparable acceptance criteria whether or not the policy binds it. Doing so gives the analyst a principled basis for declining a request, which is harder to find once a detective is standing at the counter with a compelling story.

The second cluster of decisions concerns the data. A microarray or sequencing run produces far more

information than kinship estimation needs. The laboratory can specify what the vendor may analyse, what it may keep and for how long, and whether the file may be used to develop the vendor's own products. It can decide whether phenotype and ancestry inference are run as a matter of course or only on request. Guidance from the Scientific Working Group on DNA Analysis Methods treats genealogy as an investigative tool whose result must be confirmed by conventional typing ([SWGDM, 2020](#)). That confirmation step is the laboratory's, and it is where the quality system reasserts itself.

The third cluster concerns what is left behind. When a case closes, the SNP file, the list of matches, the family trees and any reference samples taken from relatives still exist somewhere. In 2021 Maryland enacted the first comprehensive American statute on the field, with requirements that include judicial authorisation, consent before relatives are sampled and destruction of data at the end of proceedings ([Ram et al., 2021](#)). Elsewhere the laboratory's own retention schedule is frequently the only rule there is.

5. Objections Considered

Three objections can be raised against the view taken here. The first is that the privacy interest is slight. People upload their data voluntarily, relatives have always been able to incriminate one another, and the result is only a lead that must be confirmed by a direct comparison. There is force in this. A genealogical match is not evidence of guilt and is not presented as such. But the objection underrates the difference between a relative who chooses to speak to the police and a database in which a whole extended family can be read from a handful of volunteers. Since [Supreme Court of the United States \(2018\)](#), American law has been less willing to assume that information exposed to a third party carries no expectation of privacy, and commentators have asked whether genetic data are a natural candidate for the same treatment ([Ram et al., 2018](#)).

The second objection is that these are questions for legislatures, courts and police, and that a laboratory should confine itself to science. The difficulty is that legislation is sparse and uneven, and in its absence the practical limits are set by whoever handles the material. A laboratory that releases an extract without conditions has made a policy choice, whether or not it recognises it as one. [Wickenheiser \(2019\)](#) reaches a similar conclusion from a bioethical analysis: the technique can be justified, but only within a framework that someone has to write and maintain.

The third objection is distributive. Government offender databases over-represent the groups that are most heavily policed, whereas recreational genealogy databases over-represent people of European ancestry. Some authors have argued that a universal database with strict controls would be fairer and more protective than the present patchwork ([Hazel et al., 2018](#)). Whatever one makes of that proposal, it shows that the status quo is not a neutral baseline. A laboratory that adopts genealogy should be able to explain why its reach, which falls unevenly, is nonetheless acceptable.

6. Implications for Laboratory Practice

If the argument is accepted, several consequences follow. Genealogy should be defined in the laboratory's quality documents as a distinct service, with written acceptance criteria, a named approver and a record of each decision. Sample suitability should be assessed before material is consumed or sent out, since a mixed or degraded sample may produce a genotype file that implicates the wrong family. Contracts with vendors should cover security, secondary use, the return of extracts and the certified destruction of data, and should be reviewed by someone who understands what a genotype file can reveal.

Uploads should be made only to services whose terms allow law enforcement matching and only in the manner those terms prescribe. Reviews of the field describe how the larger services have differed, and changed, on this point ([Katsanis, 2020](#); [Kling et al., 2021](#)). The laboratory should know which regime applies on the day of the upload and should keep a copy of the terms with the case record. When the case ends, the profile should be withdrawn from the service and the fact recorded.

Finally, the laboratory should be prepared to disclose. Defence access to the genealogical stage of an investigation has been contested, on the ground that the lead was independent of the confirming profile.

From the standpoint of scientific accountability the better course is a file that can be produced: what was sent, to whom, what came back, what was uploaded and when it was removed. A method that depends on public trust in voluntary databases is poorly served by opacity.

7. Conclusion

Investigative genetic genealogy has given investigators a way into cases that had none, and it will remain in use. Its privacy costs fall mainly on people who never chose to be involved: relatives of database users, individuals investigated and cleared, and sometimes the donor of the sample. Crime laboratories control more of the points at which those costs are incurred than the public debate has recognised. Acceptance criteria, limits on what is generated, vendor contracts, retention rules and a disclosable record are all within their ordinary competence. Treating them as part of the laboratory's own responsibility, and not as matters to be settled elsewhere, is the more defensible position and the one most likely to preserve the conditions on which the technique depends.

Conflict of interest

To be declared by the author(s) before publication.

Funding

To be declared by the author(s) before publication.

References

- Bieber, F.R., Brenner, C.H. and Lazer, D. (2006). [Finding criminals through DNA of their relatives](#). *Science*, 312(5778): 1315-1316. <https://doi.org/10.1126/science.1122655>
- Edge, M.D. and Coop, G. (2020). [Attacks on genetic privacy via uploads to genealogical databases](#). *eLife*, 9: e51810. <https://doi.org/10.7554/eLife.51810>
- Erich, Y., Shor, T., Pe'er, I. and Carmi, S. (2018). [Identity inference of genomic data using long-range familial searches](#). *Science*, 362(6415): 690-694. <https://doi.org/10.1126/science.aau4832>
- Greytak, E.M., Moore, C. and Armentrout, S.L. (2019). [Genetic genealogy for cold case and active investigations](#). *Forensic Science International*, 299: 103-113. <https://doi.org/10.1016/j.forsciint.2019.03.039>
- Guerrini, C.J., Robinson, J.O., Petersen, D. and McGuire, A.L. (2018). [Should police have access to genetic genealogy databases? Capturing the Golden State Killer and other criminals using a controversial new forensic technique](#). *PLoS Biology*, 16(10): e2006906. <https://doi.org/10.1371/journal.pbio.2006906>
- Hazel, J.W., Clayton, E.W., Malin, B.A. and Slobogin, C. (2018). [Is it time for a universal genetic forensic database?](#) *Science*, 362(6417): 898-900. <https://doi.org/10.1126/science.aav5475>
- Katsanis, S.H. (2020). [Pedigrees and perpetrators: Uses of DNA and genealogy in forensic investigations](#). *Annual Review of Genomics and Human Genetics*, 21: 535-564. <https://doi.org/10.1146/annurev-genom-111819-084213>
- Kling, D., Phillips, C., Kennett, D. and Tillmar, A. (2021). [Investigative genetic genealogy: Current methods, knowledge and practice](#). *Forensic Science International: Genetics*, 52: 102474. <https://doi.org/10.1016/j.fsigen.2021.102474>
- Murphy, E. (2010). [Relative doubt: Familial searches of DNA databases](#). *Michigan Law Review*, 109(3): 291-348.
- Murphy, E. (2018). [Law and policy oversight of familial searches in recreational genealogy databases](#). *Forensic Science International*, 292: e5-e9.
- Phillips, C. (2018). [The Golden State Killer investigation and the nascent field of forensic genealogy](#). *Forensic Science International: Genetics*, 36: 186-188. <https://doi.org/10.1016/j.fsigen.2018.07.010>
- Ram, N., Guerrini, C.J. and McGuire, A.L. (2018). [Genealogy databases and the future of criminal](#)

- investigation. *Science*, 360(6393): 1078-1079. <https://doi.org/10.1126/science.aau1083>
- Ram, N., Murphy, E.E. and Suter, S.M. (2021). *Regulating forensic genetic genealogy*. *Science*, 373(6562): 1444-1446. <https://doi.org/10.1126/science.abj5724>
- Supreme Court of the United States (2013). *Maryland v. King*. 569 U.S. 435.
- Supreme Court of the United States (2018). *Carpenter v. United States*. 585 U.S. 296.
- SWGDM (2020). *Overview of Investigative Genetic Genealogy*. Scientific Working Group on DNA Analysis Methods.
- US Department of Justice (2019). *Interim Policy: Forensic Genetic Genealogical DNA Analysis and Searching*. US Department of Justice, Washington, DC.
- Wickenheiser, R.A. (2019). *Forensic genealogy, bioethics and the Golden State Killer case*. *Forensic Science International: Synergy*, 1: 114-125. <https://doi.org/10.1016/j.fsisyn.2019.07.003>