

# Investigative Genetic Genealogy

## Law enforcement's evolving use of genetic information

Genetics and the Law, PHG 523/LAW H 520

Sarah Nelson, MPH, PhD

March 1, 2021

 @blueyedgenes  myopenreadingframe.com  sarahcn@uw.edu

future tense

# We're Entering a New Phase in Law Enforcement's Use of Consumer Genetic Data

The favorite database for solving cold cases is now owned by a for-profit company. This could change everything.

By NILA BALA

DEC 19, 2019 • 7:30 AM

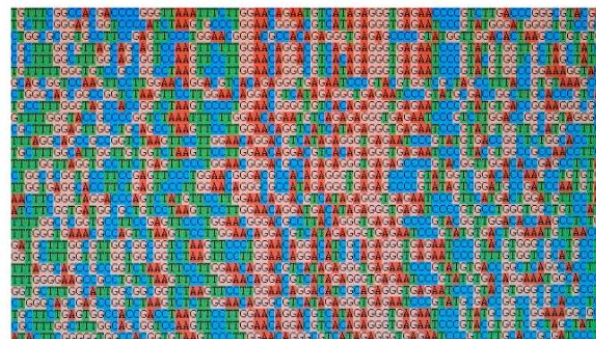


FEBRUARY 4, 2020



# About half of Americans are OK with DNA testing companies sharing user data with law enforcement

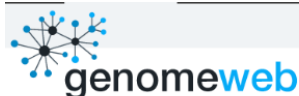
BY ANDREW PERRIN



(alanphillips via Getty Images)

# Your DNA Profile is Private? A Florida Judge Just Said Otherwise

Privacy experts say a warrant granted in Florida could set a precedent, opening up all consumer DNA sites to law enforcement agencies across the country.



Business & Policy Technology Research Diagnostics Disease Areas Applied Markets Resource

Home • Tools & Technology • Microarrays & Multiplexing • Forensic Genomics Market Advances Due to Consumer Databases, Technology |

## Forensic Genomics Market Advances Due to Consumer Databases, Technology Innovation

Jan 09, 2020 | Justin Petrone

Premium Save for later

NEW YORK – As 2020 dawns, the forensic genomics market is poised for growth as companies aim to harness the power of consumer databases coupled with advances in sequencing.

For some industry observers, the health market has often been seen as the next logical step for companies that offer consumer genomics services. 23andMe paved the way from the launch of its Personal Genome Service in 2007, and Ancestry and MyHeritage last year began offering their users health information in addition to ancestry results.



### SCIENCE

## How a Genealogy Website Led to the Alleged Golden State Killer

Powerful tools are now available to anyone who wants to look for a DNA match, which has troubling privacy implications.

SARAH ZHANG APRIL 27, 2018



### MORE STORIES

Solving a Murder Mystery With Ancestry Websites

CIARA O'ROURKE

The False Promise of DNA Testing

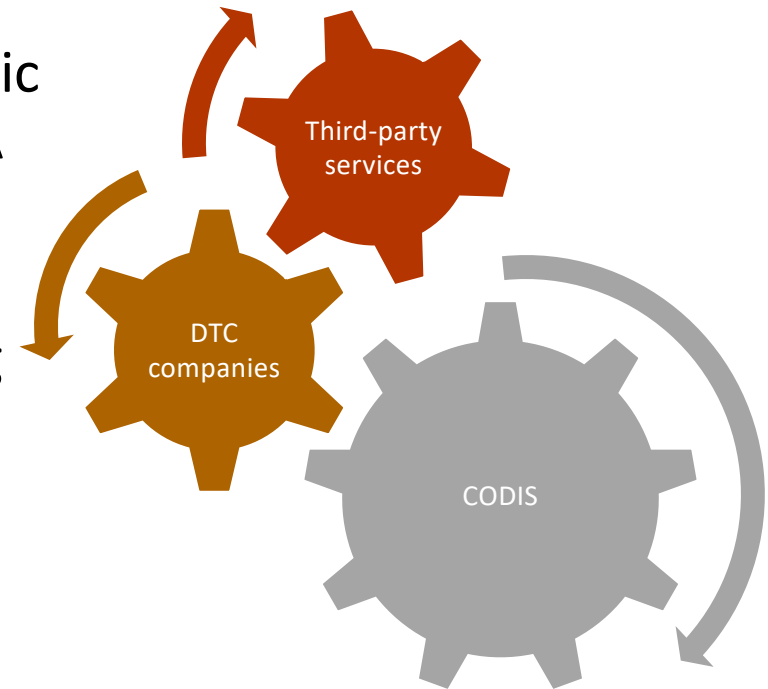
MATTHEW SHARER

23andMe Wants Its DNA Data to Be Less White

SARAH ZHANG

# Class objectives

- Understand the difference between the genetic markers used in forensic and commercial DNA testing
- Understand the differences between "traditional" and long-range familial searching
- Identify potential risks associated with long-range familial searching
- Discuss potential legal and/or policy solutions to address risks



# Outline

- I. Traditional uses of genetic information by law enforcement
- II. Familial searching (FS)
- III. Investigative genetic genealogy (IGG)



# DNA as a forensic tool

DNA can be used to

- Identify suspects
- Identify victims and missing persons
- Provide evidence to support exonerations



Image credit: Genetic Literacy Project

*Forensic* = relating to or denoting the application of scientific methods and techniques to the investigation of crime (OED)

# CODIS

- FBI Combined DNA Index System (CODIS)
  - Authorized by Congress, DNA Identification Act of 1994
- “CODIS” generically refers to FBI’s program of support for criminal justice DNA databases and the software used to run these databases
- CODIS comprises three tiers of database:

Database(s)	Maintained by
National DNA Index System (NDIS)	FBI
State DNA Index Systems (SDIS)	States
Local DNA Index Systems (LDIS)	Local police depts and sheriff’s offices

# Interactions between CODIS databases

- Federal crimes
  - DNA sample collected, sent to FBI, DNA profile entered into NDIS
- State and local crimes
  - States and local jurisdictions upload to SDIS and LDIS per their own laws and regs
  - Must meet parameters for profiles to enter NDIS
    - e.g., usable results at minimum number of CODIS core loci
    - Labs meet federally set quality assurance standards



Image - <https://www.fbi.gov/services/laboratory/biometric-analysis/codis>



# What data goes into CODIS?

- DNA profile: list of numbers indicating number of repeat units at each of the CODIS core STR loci
  - Selected Short Tandem Repeat (STR)
    - Initially 13 STR markers (in yellow)
    - 2017: Added 7 additional markers (in green)
- No names or other personal identifiers stored in NDIS
  - Kept at state or local level

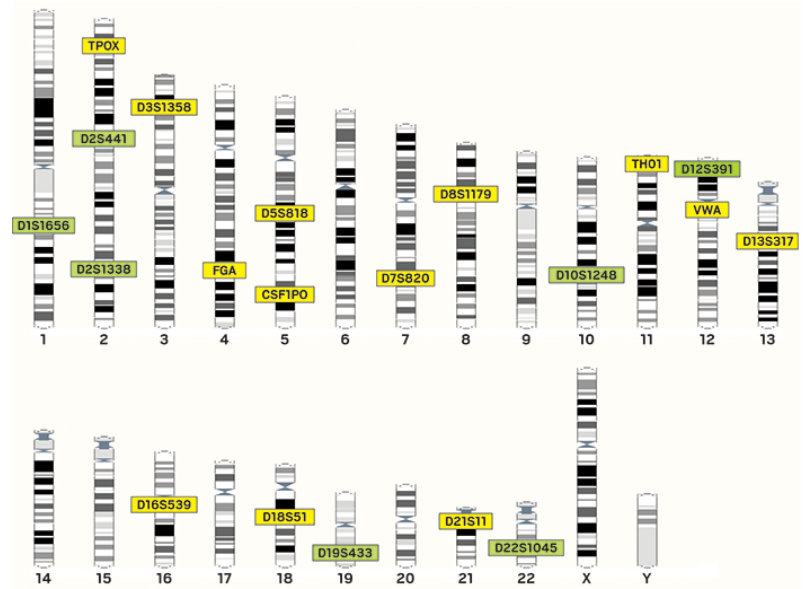


Image - <https://cen.acs.org/articles/95/i37/Thirty-years-DNA-forensics-DNA.html>

Locus	Forensic Unknown	Candidate Offender	Match Stringency
D8S1179	13	13, 14	Moderate
D21S11	28, 31.2	28, 31.2	High
D7S820	12	10, 12	Moderate
D7S820	10, 12	10	Moderate
D3S1358	15, 17	15, 17	High
.... (cont)			

<https://www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet>



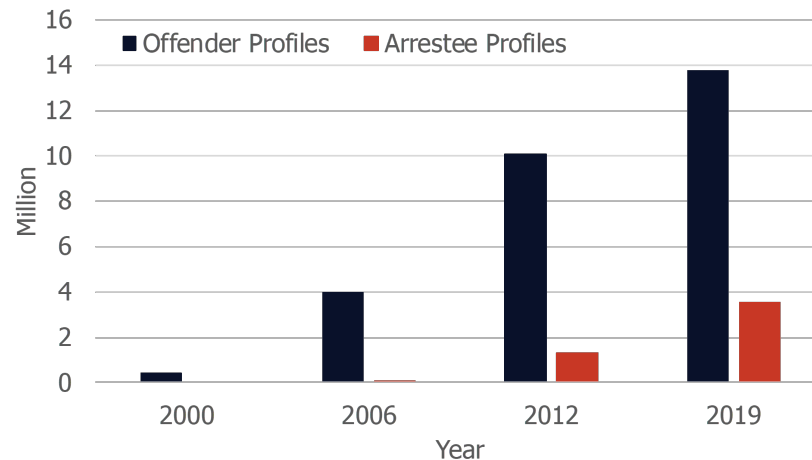
# Whose STR data goes into CODIS?

Profile type	Dec 2020 count (NDIS)
Offender*	14.4M
Arrestee**	4.21M
Forensic (i.e. from crime scene samples)	1.08M

\*Convicted offender, detainee, and legal/suspect profiles

\*\*Inclusion of arrestee samples varies by state, following 2013 US Supreme Court case *Maryland vs King*, ruling 5-4 that DNA can be taken from arrestees

Note NDIS also contains National Missing Person DNA Database (NMPDD)



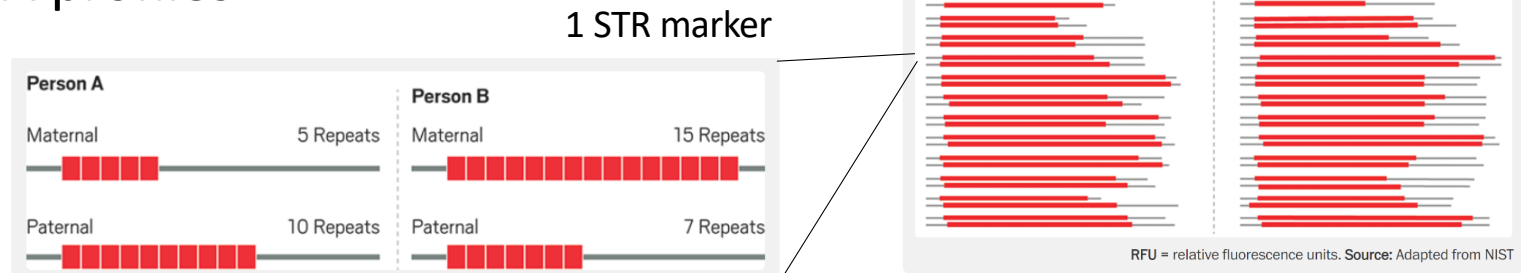
<https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics>

Personal Genetics Education Project, <https://pged.org/lesson-plans/#crime>



# What is a DNA “match”?

- Compare two DNA profiles



- **Full profile match** – two profiles are identical at all 20 loci
- **Partial match** – two profiles not identical but have many overlapping alleles. Explanations?
  - Two profiles are from two persons who are closely related
  - Two profiles are from the same person but one or both are from either DNA mixture or partially degraded DNA

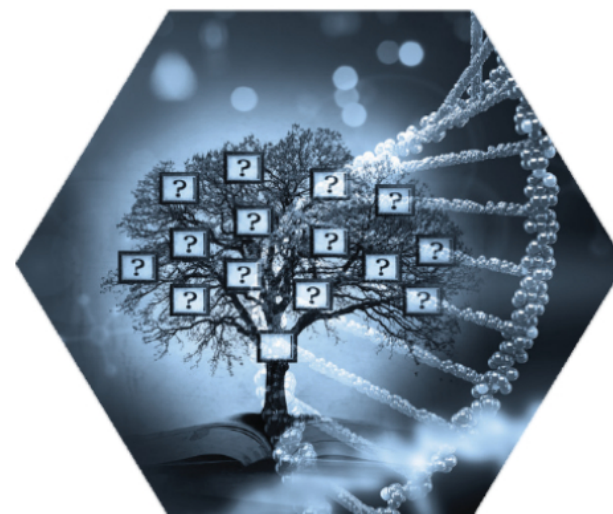
# DNA mixtures

- When a sample contains mixture of DNA from several people
- For low/trace amounts of DNA, difficult to confidently identify STR alleles
- “Double edged sword” of having more sensitive DNA detection methods
- Interpretation aided by **probabilistic genotyping**
  - Incorporates quantitative data from the genotyping or sequencing method



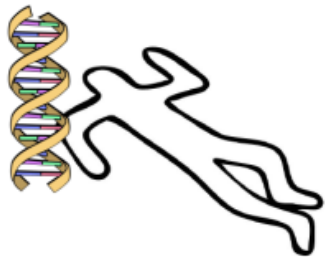
# Outline

- I. Traditional uses of genetic information by law enforcement
- II. Familial searching (FS)**
- III. Investigative genetic genealogy (IGG)

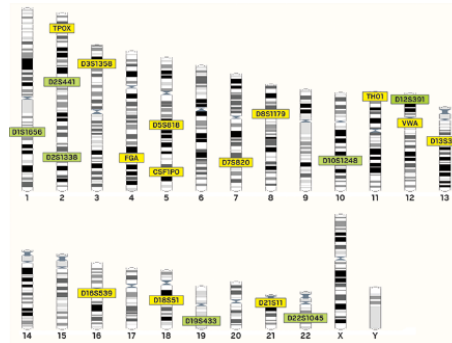


# What is familial searching (FS)?

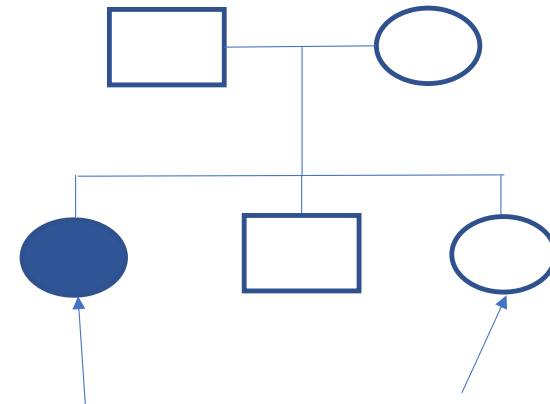
- For a crime with no leads/initial leads are exhausted
  - Including where initial CODIS searches haven't yielded full profile match
- (1) Reduce to moderate or low stringency search that returns partial match
- (2) With the intention of finding a closely related individual



1- Crime scene/forensic sample



2 – low/moderate stringency search against profiles in CODIS



Shaded =  
in CODIS

3 – Returns partial match

4 – Further investigation yields partial match's sibling as suspect

# Concerns with FS

- **Technical**

- Potential for false positives
- Requires specialized software, distinct from CODIS software

- **Ethical**

- Exacerbates existing racial biases in the databases
- Loss of privacy
- Revelation of unknown family relationships
  - Conflation of biological with social conceptions of family

- **Legal**

- Fourth Amendment concerns
  - FS presupposes that relatives of databased individuals have the same diminished expectations of privacy
  - For legal arguments *against* allowing FS, see Erin Murphy, *Relative Doubt: Familial Searches of DNA Databases*, 109 MICH. L. REV. 291 (2010)

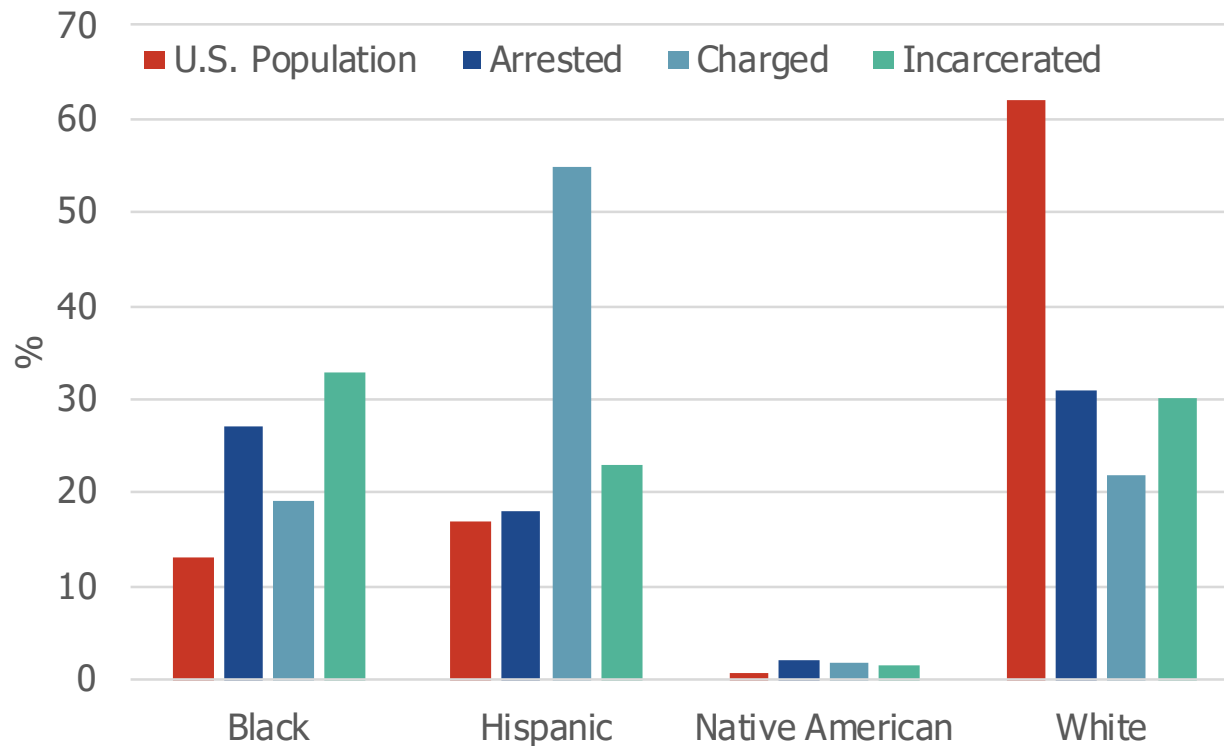


NHGRI Media Gallery



Slide courtesy of Personal Genetics Education Project, <https://pged.org/lesson-plans/#crime>

## Advances in forensic DNA technologies might disproportionately affect certain populations



Data from <https://www.bjs.gov/content/pub/pdf/p16.pdf>, <https://www.bjs.gov/content/pub/pdf/fjs1314.pdf>, and <https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/topic-pages/tables/table-21> (accessed June 12, 2019)

# Prevalence of FS

- Not conducted at federal level
- Policy and practice differs by state
  - Ten states currently perform: Arkansas, California, Colorado, Florida, Michigan, Texas, Utah, Virginia, Wisconsin, and Wyoming
  - Two jurisdictions explicitly prohibit by law: Maryland and DC

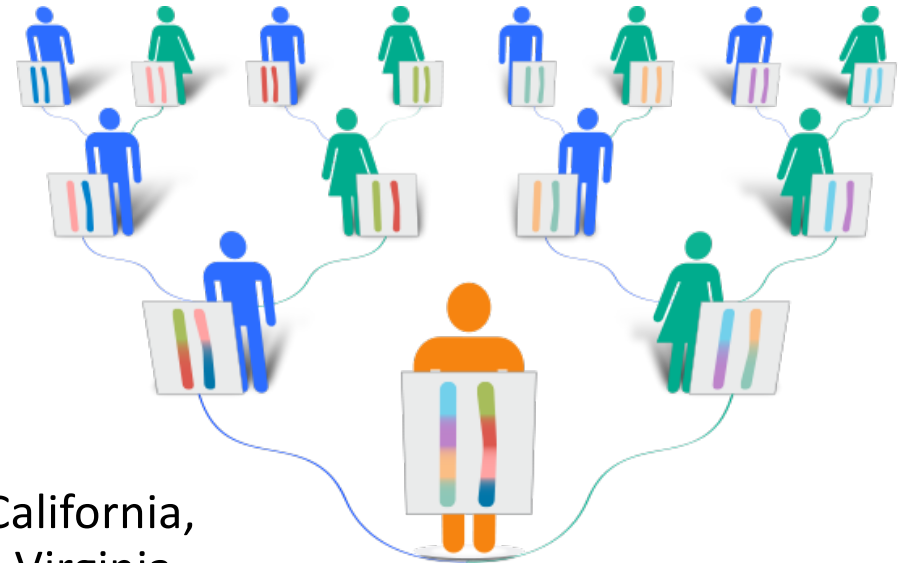
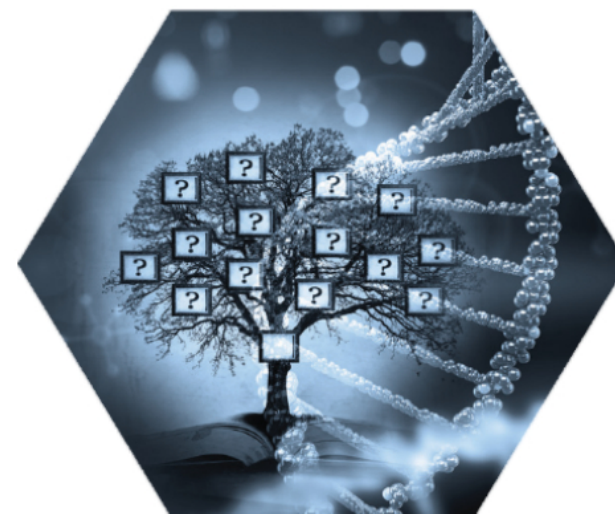


Image: FamilyTree DNA on Twitter,  
<https://twitter.com/familytreedna/status/752503030357065728>

# Outline

- I. Traditional uses of genetic information by law enforcement
- II. Familial searching (FS)
- III. Investigative genetic genealogy (IGG)**



# What is investigative genetic genealogy (IGG)?

“...the science of using genetic and genealogical methods to generate leads for law enforcement entities investigating crimes and identifying human remains”

- International Society of Genetic Genealogy Wiki,  
[https://isogg.org/wiki/Investigative\\_genetic\\_genealogy\\_FAQ](https://isogg.org/wiki/Investigative_genetic_genealogy_FAQ)

Also referred to as *forensic genetic genealogy (FGG)*

# Key features of IGG/FGG

When traditional methods have not produced a suspect

- Use of non-governmental databases containing
- ...high density genetic data that enables
  - e.g., data from genotyping array with 500K-1M SNPs
- ...long range familial searching paired with
- ...traditional genealogical methods that
- ...generates investigative leads requiring CODIS confirmation



Image: *Slate* magazine

# What is long range familial searching?

- FS with CODIS STR markers generally limited to **close relatives** (i.e., first degree rels – parent/offspring or siblings)
- Denser genetic information (e.g., array data from DTC testing) enables searching for matches with more **distant relatives** -> “long range” FS

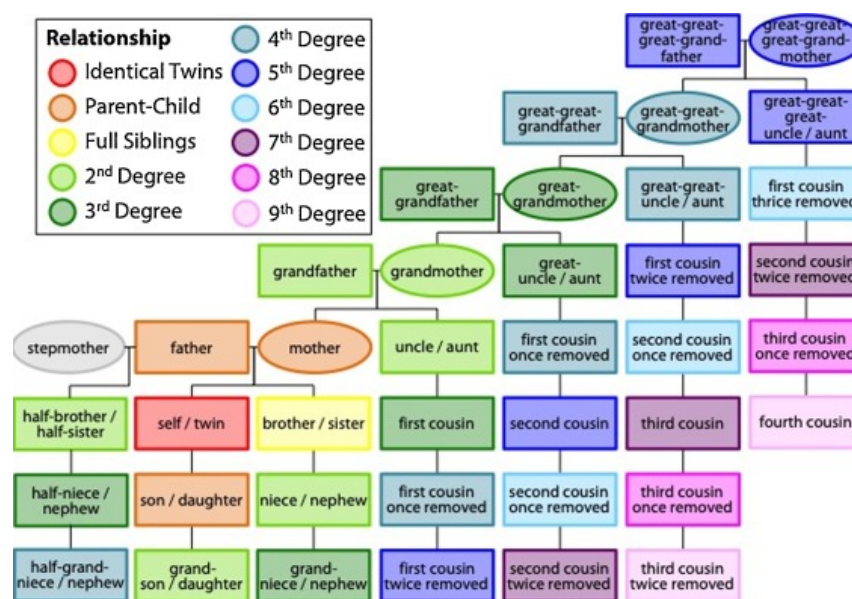


Figure 1 from Greytak, Moore, Armentrout 2019

# Implications of long range FS

- Erlich et al. 2019 - Empirical study of how likely to identify an individual
- With some simplifying assumptions, estimate that only 2% of a target population needed in a database to return 3<sup>rd</sup> cousin relative to nearly any person in the target population
- 2% of 330M in US = 6.6M
  - AncestryDNA has ~20M customers

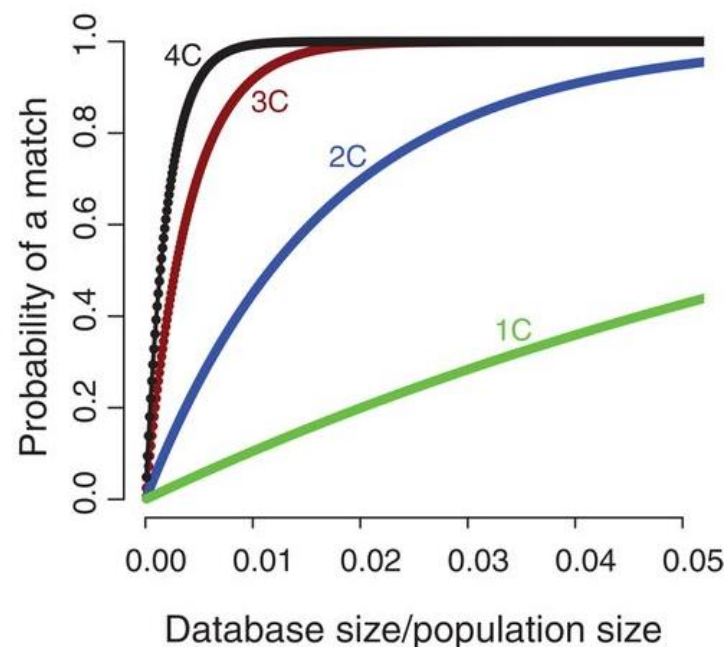


Fig 1B, Erlich et al., *Science*, 2019





## Databases that enable IGG

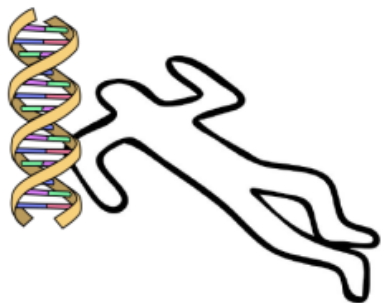
### DTC companies

- Family Tree DNA
  - Only DTC company explicitly cooperating with law enforcement (LE) requests
  - Customers automatically opted in, unless EU resident
- My Heritage requires court order or subpoena for LE usage
- 23andMe and AncestryDNA report to actively resist LE requests, including subpoenas
  - Both 23andMe and Ancestry DNA maintain transparency reports disclosing LE requests

### Third-party tools



- Users upload “raw” data files from DTC companies
- GEDmatch - publicly available genealogy site
- Apparently responsible for most of IGG
- Acquired by Verogen in Dec 2019



**Step 1: Collect DNA from Crime Scene**



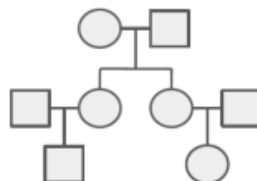
**Step 2: Convert the DNA Sample into the DTC genetic file format**



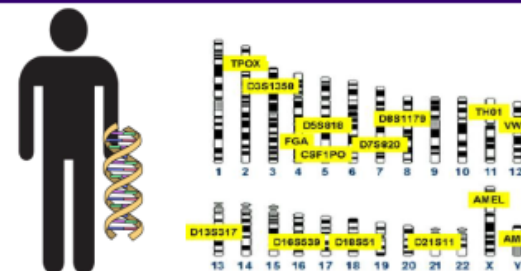
**Step 3: Upload the DTC File to GEDmatch**

Alex Smith: 2nd cousin  
Aaron Rodgers: 3rd cousin

**Step 4: Use Relative Matching to get List of People Related to the Criminal**



**Step 5: Use Public Records like Family Trees to Generate a List of Leads**



**Step 6: Confirmatory Testing with CODIS on Suspects**

# First examples of IGG

- Landmark use
  - Identifying suspect, Joseph DeAngelo, in Golden State Killer case, April 2018
- LA Times reported in Dec 2020 that My Heritage was unwittingly used for initial lead
  - Subsequent use of GEDmatch
- >200 cases since solved using IGG
  - Include murder, sexual assault, and burglary
  - Ranging from decades old “cold cases” to cases only months old



Photo of Joseph James DeAngelo via Sacramento county police department

Hill and Murphy, *New York Times*, Nov 2019  
St. John, *LA Times*, Dec 2020

# Legal basis of IGG

- Third party doctrine



“The Fourth Amendment’s protection against warrantless searches and seizures generally does not apply to material or data voluntarily shared with a third party, like a direct-to-consumer genetics testing or interpretation company or a genetic matching platform like GEDmatch”

- Ram, Guerrini, McGuire 2019

For more legal scholarship on this topic, see Abrahamson 2019, *Fordham Law Review* 87(6)

# Public Opinion on LE access of consumer dbs

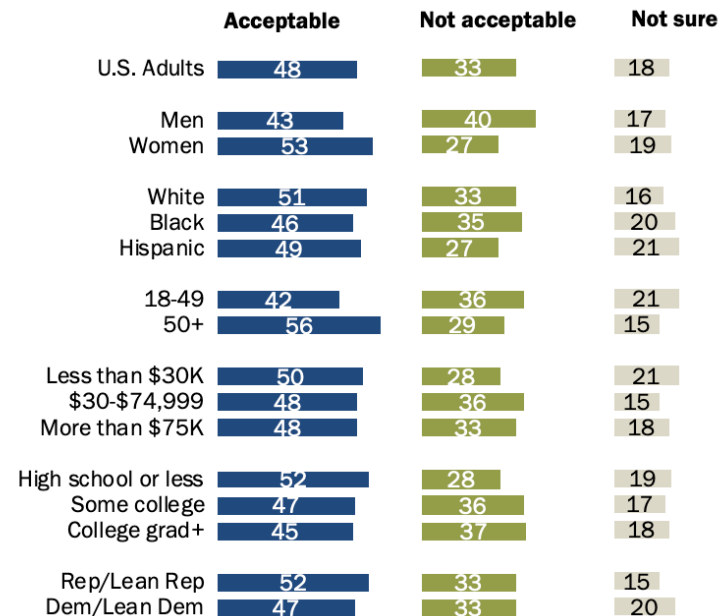
- Pew Research Center Survey conducted June 2019
- Survey also found that 16% had used a DTC genetic test
- Not addressed?
  - Type(s) of crimes
  - User awareness of LE searches
    - Visibility in Terms of Service
    - Opt-in vs opt-out

<https://www.pewresearch.org/fact-tank/2020/02/04/about-half-of-americans-are-ok-with-dna-testing-companies-sharing-user-data-with-law-enforcement/>

---

## 48% of adults find it acceptable for DNA testing companies to share users' genetic data with law enforcement

*% of U.S. adults who say DNA testing companies sharing customers' genetic data with law enforcement to help solve crimes is ...*



Note: Those who did not give an answer are not shown. Whites and blacks include only non-Hispanics. Hispanics are of any race. Rep/Lean Rep includes Republicans and independents who lean toward the Republican Party. Dem/Lean Dem includes Democrats and independents who lean toward the Democratic Party.  
Source: Survey conducted June 3-17, 2019.

PEW RESEARCH CENTER

---

# GEDmatch in flux

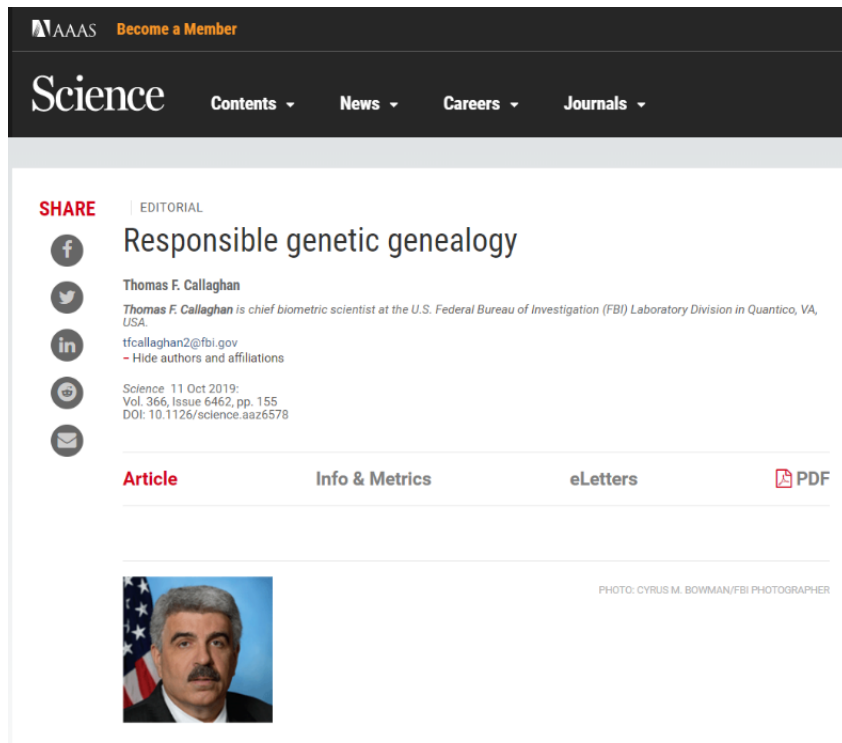
- April 2018 – GEDmatch reported as tool used to crack GSK case
- May 2018 – explicitly allow data upload by LE
  - for violent crimes only: homicide and sexual assault
- May 2019 – changed TOS to require opt-in for use by LE
  - Also expanded definition of violent crime
  - Reduced available search space from >1M to <200K
- Nov 2019 – Florida judge issues search warrant for GEDmatch
  - So does opt-in/opt-out even matter?
- Dec 2019 – GEDmatch bought by forensic genomics company Verogen

## GEDmatch in flux, cont.

- July 2020 – security breach; 3hr period where all kits were searchable
- Dec 2020 – launch of dedicated LE portal, GEDmatch Pro
- Jan 2021 – Verogen allows human remains to be searched against whole database
- Jan 2021 – Verogen announces ForenSeq Kintelligence Kit
  - Targeted assay of ~10K SNPs designed for relatedness analysis
  - Analysis of GEDmatch profiles informed SNP set



# DOJ interim policy on Forensic Genetic Genealogy



The screenshot shows the top navigation bar of the Science magazine website with the AAAS logo and a 'Become a Member' link. Below the navigation bar, the article title 'Responsible genetic genealogy' is displayed, along with the author's name 'Thomas F. Callaghan' and a brief bio: 'Thomas F. Callaghan is chief biometric scientist at the U.S. Federal Bureau of Investigation (FBI) Laboratory Division in Quantico, VA, USA.' The article is dated '11 Oct 2019' and is part of 'Vol. 366, Issue 6462, pp. 155'. The page includes social media sharing icons for Facebook, Twitter, LinkedIn, and Email. Below the article title, there are tabs for 'Article', 'Info & Metrics', 'eLetters', and a PDF icon. A small portrait photo of Thomas F. Callaghan is visible at the bottom left of the article content area.

*However, use of FGG by law enforcement has **preceded widespread development of best practices** to protect the genetic privacy of private citizens who have voluntarily submitted samples to genealogy databases...The emergence of FGG suggests that **further discussions on privacy, genomics, and the use of genealogy by law enforcement** would be beneficial. Accordingly, the FBI seeks to **engage the scientific and bioethics communities** in such a dialogue.*

....

*The scientific community and other interested parties are **encouraged to provide the FBI with comments** at [forensicgenealogy@fbi.gov](mailto:forensicgenealogy@fbi.gov)*

**Commentary** - <https://science.sciencemag.org/content/366/6462/155> - emphasis added

**Draft policy** - U.S. Department of Justice, Interim Policy on Forensic Genetic Genealogical DNA Analysis and Searching (2019); [www.justice.gov/olp/page/file/1204386/download](http://www.justice.gov/olp/page/file/1204386/download)

# Acknowledgements and References

- Received valuable feedback and input from Prof. Mastroianni, Sanne Albers, and Dr. Peter Ney
- Personal Genetics Education Project lesson on “DNA, crime, law enforcement”
- FBI CODIS website:  
<https://www.fbi.gov/services/laboratory/biometric-analysis/codis>
- Many other refs noted throughout slides....

# Readings

Kling, D., Phillips, C., Kennett, D. & Tillmar, A. Investigative genetic genealogy: Current methods, knowledge and practice. *Forensic Sci. Int. Genet.* 52, 102474 (2021)

Megan Molteni, Cops are Getting a New Tool for Family Tree Sleuthing *Wired* (12-18-20)

DOJ interim policy on Forensic Genetic Genealogy DNA Analysis and Searching (Nov 2019) -

<https://www.justice.gov/olp/page/file/1204386/download>