Forensic Investigative Genetic Genealogy

Lagos Forensic Symposium

Dr. Ray Wickenheiser

New York State Police Crime Laboratory System Director

Disclaimer

Opinions expressed herein are those of Ray Wickenheiser, not that of NYSP, ASCLD, FSSB OSAC, SWGDAM or any other entity.

Outline

- Current use of databases
- Direct and indirect matching
- FIGG
- EDIM and Search keys
- Case flow
- Business cases
- Recommendations

Current Practice

Current Practice

- DNA databases are very effective at producing **investigative leads**
- 206 Accredited Crime Labs in United States using CODIS (NDIS) via STRs
- Forensic Sample: Crime Scene sample from the putative perpetrator
- Direct matching: comparison of the forensic profile to known samples (individually or via a DNA database) where the entire profile is in common for the same source

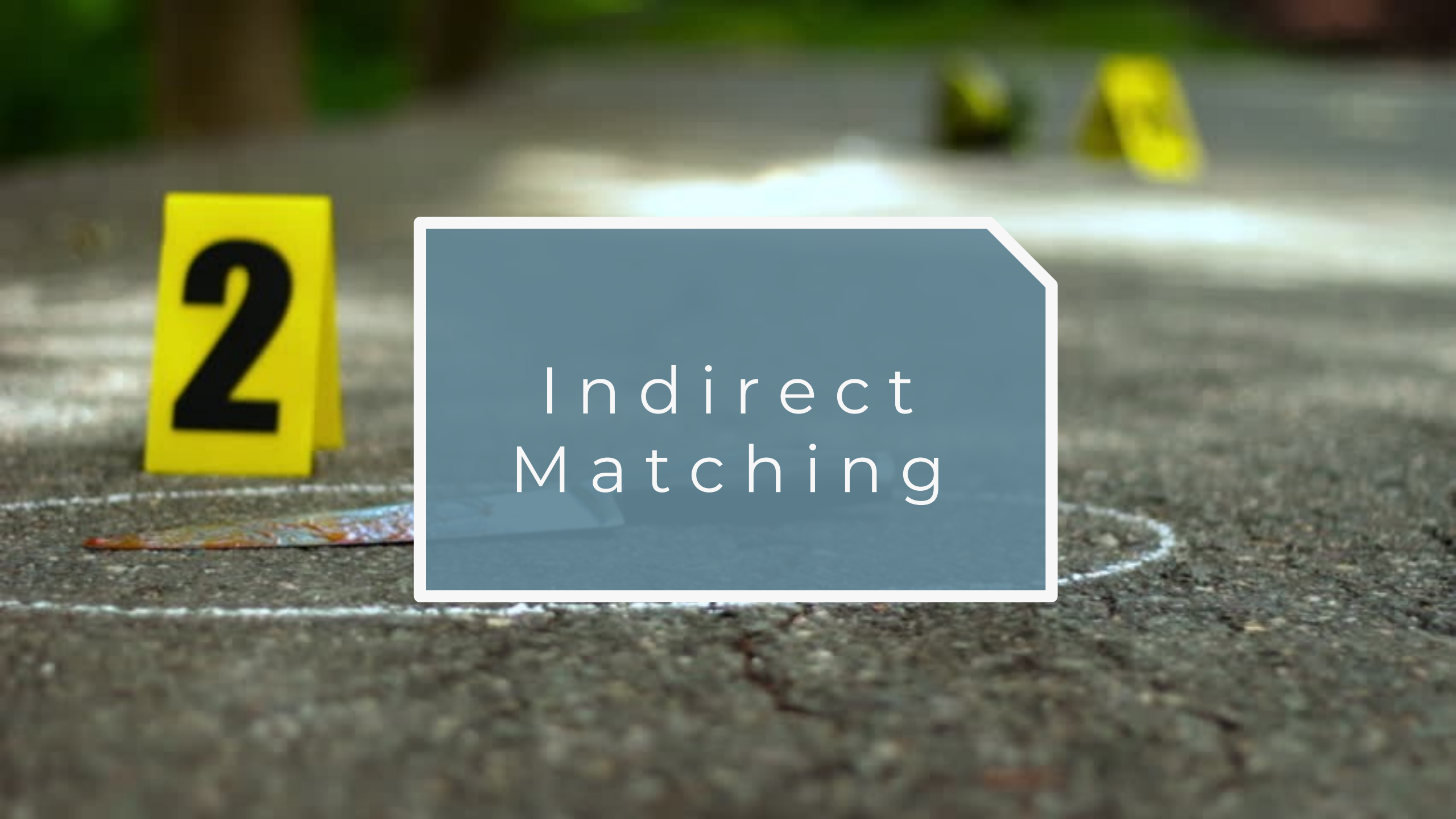
CODIS and Direct Comparison

- As of October 2021, the NDIS contained over 14,836,490 offender profiles, 4,513,955 arrestee profiles, which equals 19,350,445 known profiles available for comparison.
- There were also 1,144,255 forensic profiles in NDIS.
- Those forensic profiles recovered from biological materials deposited at crime scenes have produced 587,773 hits, which equates to roughly a 51.37% hit rate.
- This also means that roughly **556,482** forensic profiles have not been hit upon.
- The unsolved crimes represent a massive opportunity to solve crime and prevent future crimes through investigative leads.

Opportunity

- A larger database means more hits; more investigative leads
  - Directly (physically)
  - Indirectly (scientifically)

- Strategies:
  - Direct Matching
    - Arrestees
    - Lawfully owed DNA samples
    - Unidentified Human Remains (UHRs) – Direct and Indirect
  - Indirect Matching
    - Partial matching
    - Familial Searching
    - FIGG
    - EDIM



Indirect Matching

Indirect matching

- DNA profiles partially match (share common areas) due to biological relatedness
- Biological relatedness conveys known inheritance patterns
- Kinship analysis permits evaluation of indirect match to postulate the level of biological relatedness
- Infer Relationship – immediate family, extended family, paternal and maternal family lines

Indirect matching techniques

- Partial matching (discover non-identical profiles sharing significant DNA similarities) – passive search
- Familial Searching – active search of state CODIS for immediate family members
- Forensic Investigative Genetic Genealogy (FIGG) – active search of genealogy database to find potential kin, then use genealogical researching to develop investigative leads
- Indirect matching increases the size of the database by including biological relatives

Descriptions

- Partial matching and familial searching use STRs in common and the relative rarity of STRs to develop a Likelihood Ratio (LR)

- Forensic Investigative Genetic Genealogy (FIGG) uses SNPs. SNPs map the genome and are used by algorithms to compare fragments (similarities measured in centimorgans).
  - cM (centimorgans) is the measuring stick (1 cM is roughly 1 million base pairs)
  - shared fragments of DNA where those fragments have the same SNPs
  - shared number and size of fragments in common provides a measure of relatedness

- FIGG – use of SNPs to search genealogy database to find potential kin sharing portions of DNA (centimorgans), then use genealogical researching and building of family trees to develop investigative leads

FIGG

What is **FIGG**?

Generation of a SNP profile from perpetrator DNA discarded at crime scene, place in database to find relatives, use genealogy to develop leads, get DNA from suspect for direct comparison to crime scene DNA

PHASE I

Crime Scene Analysis – Crime scene DNA from perpetrator, no CODIS hit

PHASE II

Genealogy – SNP (Single Nucleotide Polymorphism) profile, database for leads, genealogy

PHASE III

Investigation – Obtain known samples for direct comparison

FIGG Phase 1

1. Crime Scene attendance
2. Submission to crime lab
3. Serology
4. DNA - STR profile
5. CODIS entry
6. No hit = FIGG candidate

FIGG Phase 2

1. Analyze SNPs: Develop SNP profile from crime scene sample (suspect)
2. Database Search: Search Direct to Consumer (DTC) genealogy database(s) for an indirect match to related individuals
3. Ascendancy Search for a Most Recent Common Ancestor (MRCA) – Ascendancy
4. Descendancy: search for descendants
5. Family: Look for crossing (triangulation) of family trees (maternal and paternal sides) of suspect family
6. Specific suspects: Search for descendants that fit the crime profile (location, sex, age, other attributes)

FIGG Phase 3

1. Known sample obtained from investitive lead
2. Submission to crime lab
3. DNA - STR profile
4. Comparison to forensic sample
5. Report and testimony if a match
6. If no match, kinship analysis to guide next steps

EDIM and Search Keys

Proposed Expanded DNA Indirect Matching (EDIM)(1)

- What is a Search Key?
- Search keys are components (aspects/types) of DNA profiles that are shared by related individuals, but are not unique to that individual, but rather tell something about their family through inheritance
- Use of a search key can effectively search samples for potential relatives
  - Y-STR – inherited paternally
  - X Chromosome – inherited from maternal family line in males, from both sides in females
  - mtDNA – inherited maternally

Forensic Y-STR as a search key

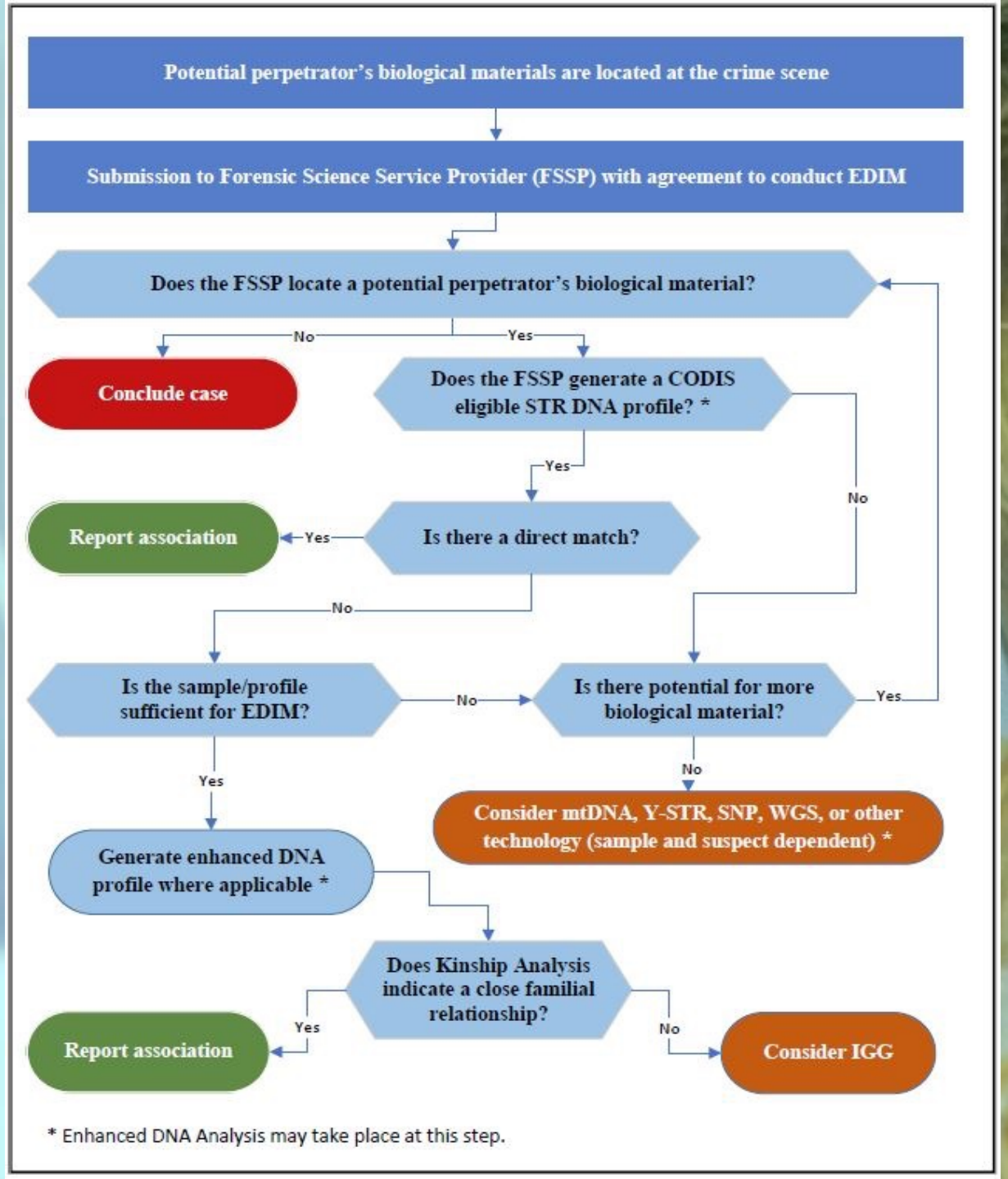
- Y-STRs are inherited paternally
  - Using the Y-STR as an example, expanding the DNA profile beyond the CODIS core loci would include a Y-STR profile
  - Y-STR used to search for candidate male biologically related individuals
  - Once a candidate is located, kinship analysis can be conducted to evaluate the level of relatedness

Proposed Expanded DNA Indirect Matching (EDIM)(1)

1. Develop an expanded DNA profile (Y-STR, X Chromosome, mtDNA, SNP, WGS)
2. Search cases against each other using search keys
3. Conduct kinship analysis between cases with the same search key
4. Report cases with close relatives
   a. Cases with previous matches or leads add value to the indirectly matched cases (biological relatives)
5. Consider SNPs for FIGG for more distant relatives

Proposed Expanded DNA Indirect Matching (EDIM) workflow (1)

- Several entry points for NGS (see *)
- 1. At original DNA analysis
- 2. After CODIS search is unsuccessful
- 3. When other typing methods are not successful
- Is this repetition the best process?
- Our goal is to maximize the evidence, therefore expanded profiles create the most matching opportunities




Bump up
the value

Unidentified Human Remains (UHRs)

- Indirect and direct matching opportunity which qualifies for CODIS

- Direct
  - High risk lifestyle (victim and perpetrator)
  - Opportunity to close open cases

- Indirect matching
  - Help ID UHRs through EDIM and FIGG
  - ID in turn feeds direct matching

- Provides closure
- Help solve case, particularly if UHR is homicide victim

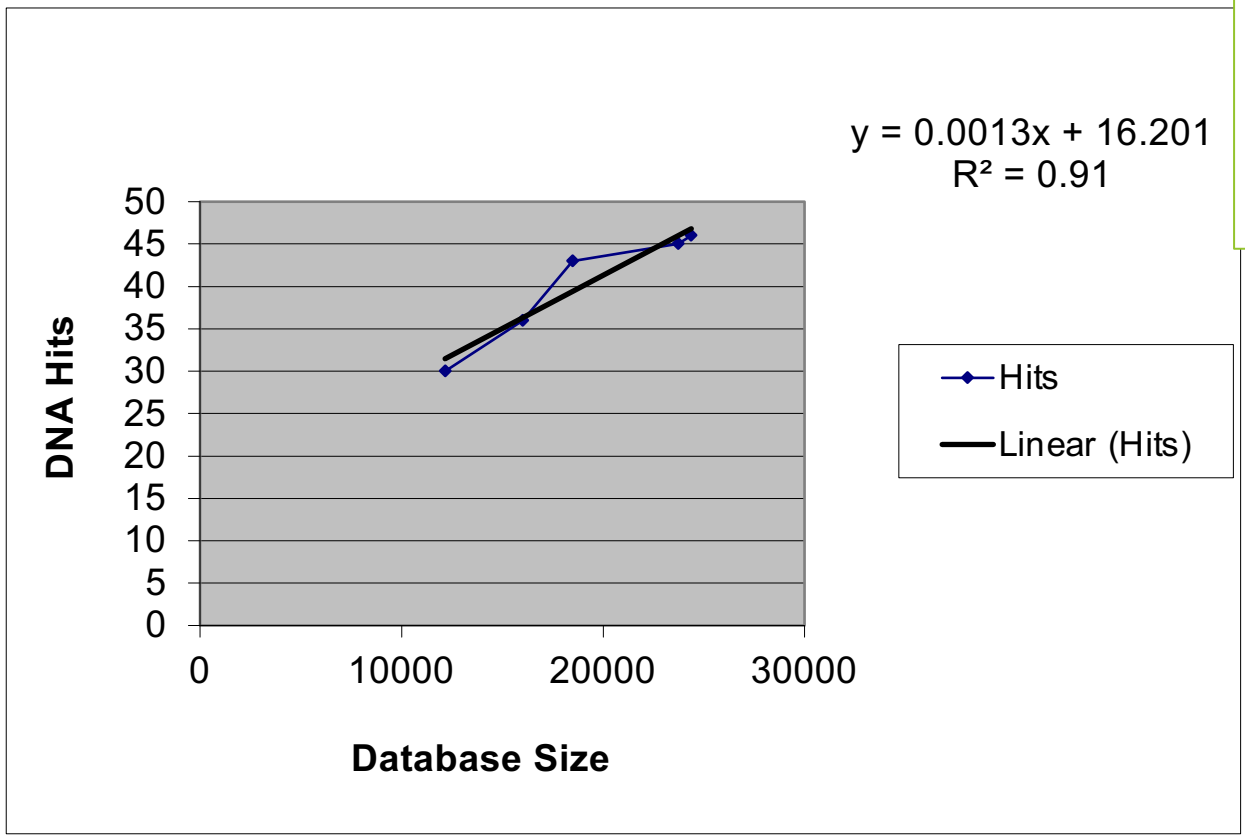
Universal DNA Database (for illustration only)

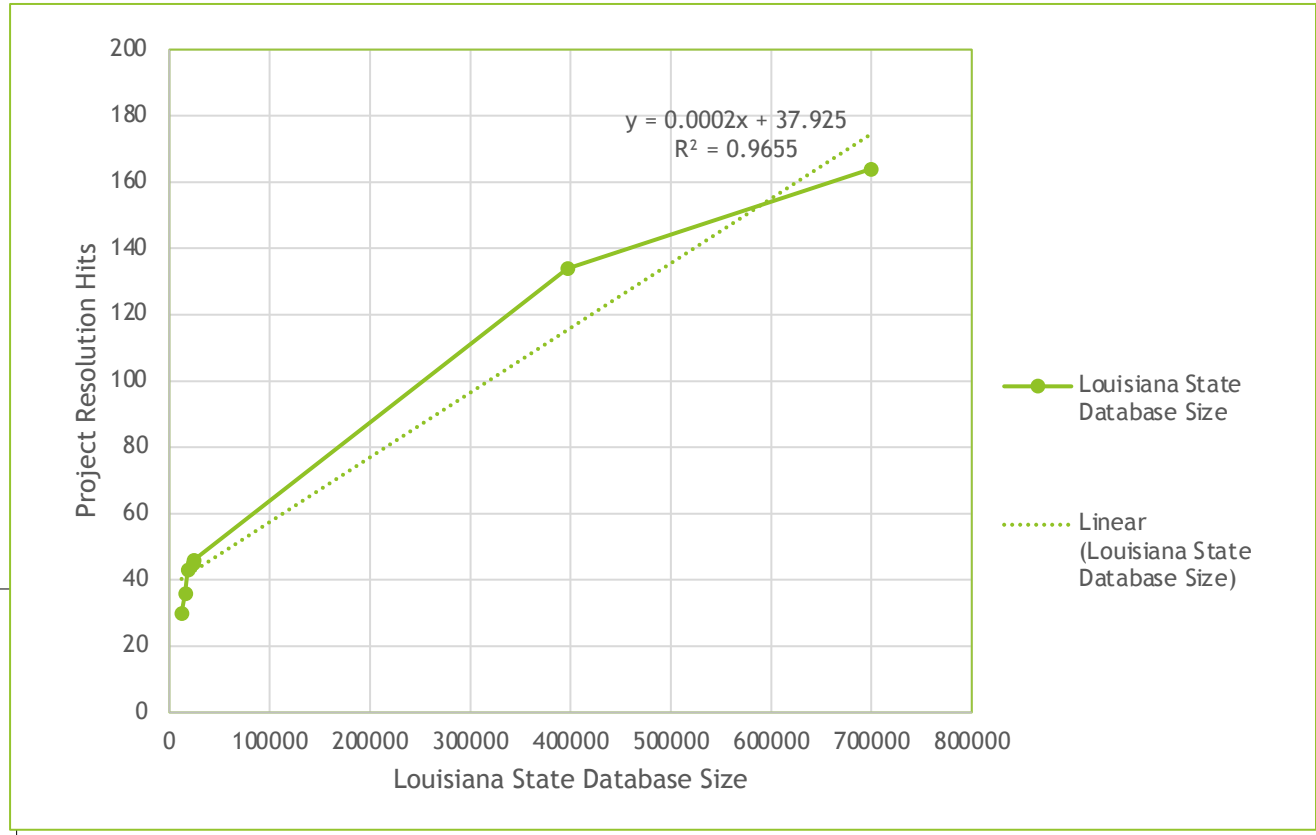
- Cost of $16.5 Billion (331.4 Million Americans at $50/sample)
- 42% plus remaining hit rate potential compared to cold case projects (Palm Beach, Acadiana and Detroit)
- 139,380 sexual assaults in US annually
- $435,419 per sexual assault
- 26.22 preventable sexual assaults per hit
- Opportunity cost is $66.8 Trillion
- **ROI is $4,050 per $1 spent**

	Cost of Sexual Assault	CODIS Hits	Project Resolution Cost	Recidivism Factor	Return on Investment
Conservative Model	$111,238	164	$286,000	7	$446.51
Aggressive Model	$435,419	164	$286,000	26.22	$6546.63

Project Resolution – Conduct DNA analysis on 605 case cuttings, 285 CODIS profiles resulted, with 164 hits

Project Resolution





- Increase in DNA database size has a direct positive relationship with increase in hits
- More known samples equals more investigative leads
- Collect samples owed
- Increase offense types
- Include arrestees
- UHRs (Unidentified Human Remains)

Cost per sexual assault	$435,419
Number of preventable sexual assaults per hit	26.22
Number of hits	164
Size of Louisiana SDIS	699,618
Louisiana SDIS offenders per hit	4,266
Cost benefit of Project Resolution	$1,872,336,534
Analysis cost per database sample	$50
Cost of Louisiana SDIS offender sample analysis	$33,980,900
Cost benefit per $1 spent	$53.52
Return on investment percentage	5,352

Cost benefit of Louisiana Database increase

Indirect matching – frequency of relatives

- Depends on biological relatives to provide leads
- Frequency of siblings in the New York State DNA database has been estimated
- In 2017 there were 536 sets of identical twins, with the database size at approximately 700,000 individuals
- Rate of identical twins is 1 in 250 live births
- 536 X 250 = 134,000 sibling pairs
- 134,000/700,000 X 100 = 19.14% (1), so 1 in 5 individuals will have a sibling (low estimate as parent-child relationships are additional)

Business Case for FIGG/EDIM (1)

- The estimated number of sexual assault hits annually is 49,964
- EDIM has potential to increase the hit rate by 19.14%
- This will solve an additional 9,563 sexual assaults annually
- The cost of a single sexual assault is $435,519 and each solved sexual assault prevents 26.22 sexual assaults (3)
- Estimated cost of crime savings is $109.18 Billion
- Potential to prevent 250,741 sexual assaults

Considerations and Recommendations

Crime Scene DNA Bioethics

- Autonomy – right to privacy and protection from unwarranted search and seizure
- Discarded material – no presumed right to privacy: autonomy is vacated
- Proportionality – weighing the benefits and risks of competing options to maximize the overall good while minimizing the overall downsides
- Recidivism - damage to society caused by a small number of individuals committing repeat and escalating crimes on new victims
- Objectivity – scientific evidence is unbiased, neutral, can be retested, challenged in court, debated (investigative lead)
- Forensic profile is developed pre-suspect and STR profile from lead is confirmed with STRs within an accredited crime lab system for court

Why Now?

- We forget that these criminals are out there now committing new crimes and we have their DNA at the crime scene in existing crimes
- FIGG/EDIM is not just for cold cases; this is for any case we do not get a CODIS hit upon
- Our mission is to maximize the value of evidence
- The huge ROI demonstrates the value of DNA investigative leads and databases

Recommendations

- Increase DNA Database size for direct matching
  - Expand qualifying offenses
  - Include arrestees
  - Maximize UHRS
  - Collect all lawfully owed DNA samples
- Increase the DNA database by utilizing indirect matching
  - NGS/EDIM is not just for cold cases; this is for any case we do not get a CODIS hit upon
  - Case for an expanded DNA profile on all forensic (crime scene) samples

References

1. Ray A. Wickenheiser, Expanding DNA Database Effectiveness, Forensic Science International: Synergy 4 (2022), 100226, 5 April 2022, https://doi.org/10.1016/j.fsisyn.2022.100226

2. Ray A. Wickenheiser, Demonstrating cost-benefit for forensic laboratory resources: Project Resolution, Forensic Science International: Synergy 3 (2021) 100158, 6 July 2021, https://doi.org/10.1016/j.fsisyn.2021.100158.

3. Wickenheiser, R.A., The Business Case for Using Forensic DNA Technology to Solve and Prevent Crime, J Biolaw and Business 2004:7 (3):34-50. http://www.dnaresource.com/documents/BusinessCaseforDNA.pdf

4. Wang, C. and Wein, L.M., Analyzing Approaches to the Backlog of Untested Sexual Assault Kits in the U.S.A., Journal of Forensic Sciences, July 2018, Vol. 63, No. 4, doi: 10.1111/1556-4029.13739

5. CODIS – NDIS Statistics; Measuring success, Federal Bureau of Investigation, Criminal Justice Information Services, Laboratory Services, Biometric Analysis https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics

Thank you!

Dr. Ray Wickenheiser

Ray.Wickenheiser@Troopers.NY.gov

(518) 457-1208