Law Enforcement Use of Probabilistic Genotyping

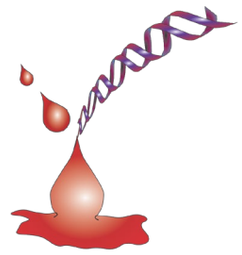
Norah Rudin, Ph.D.

NAS Workshop

March 13, 2024

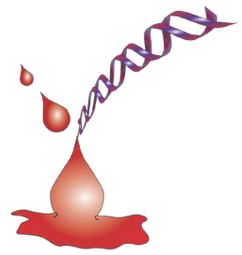
Probabilistic genotyping programs

- All technologies are tools
- Probabilistic genotyping is a technology
  - Implemented via various software implementations
- Two commercial programs enjoy name recognition in the U.S.
  - Currently the U.S. market is dominated by one of these programs



Probabilistic genotyping programs

- Both programs are expensive and proprietary
  - Their use is restricted to:
    - Those who can afford them
    - Those who the purveyors of the programs deem worthy
    - Those who the purveyors of the programs deem non-competitors



All tools have capabilities and limitations

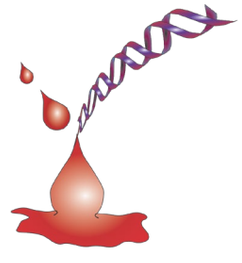
- Capabilities and limitations determined by validation studies
- Capabilities of PG
  - Uses more profile information
    - How much depends on flavor
- More accurate weighting of profile
- Allows relevant questions to be asked

All tools have capabilities and limitations

- Limitations of PG
  - Requires more understanding, additional education and training
  - Difficult to communicate results to laypersons
  - Does not overcome sample, profile limitations
  - No true value upon which to calibrate systems
  - Risk of patina of science without understanding sources of uncertainty and variability

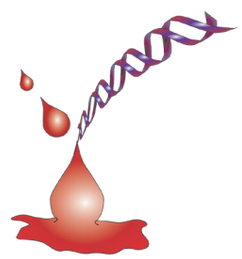
Greatest sources of variation extrinsic (originate outside the software)

- Technological sources of variation
    - Different PCR amplification tubes of the same sample (replicates) can produce wildly different profiles, resulting in massive differences in PG results
        - Possibly combination of stochastic effects from both pre-PCR sampling and PCR process itself
- Differences between sample replicates can exceed differences due to the algorithms and implementation of different software programs

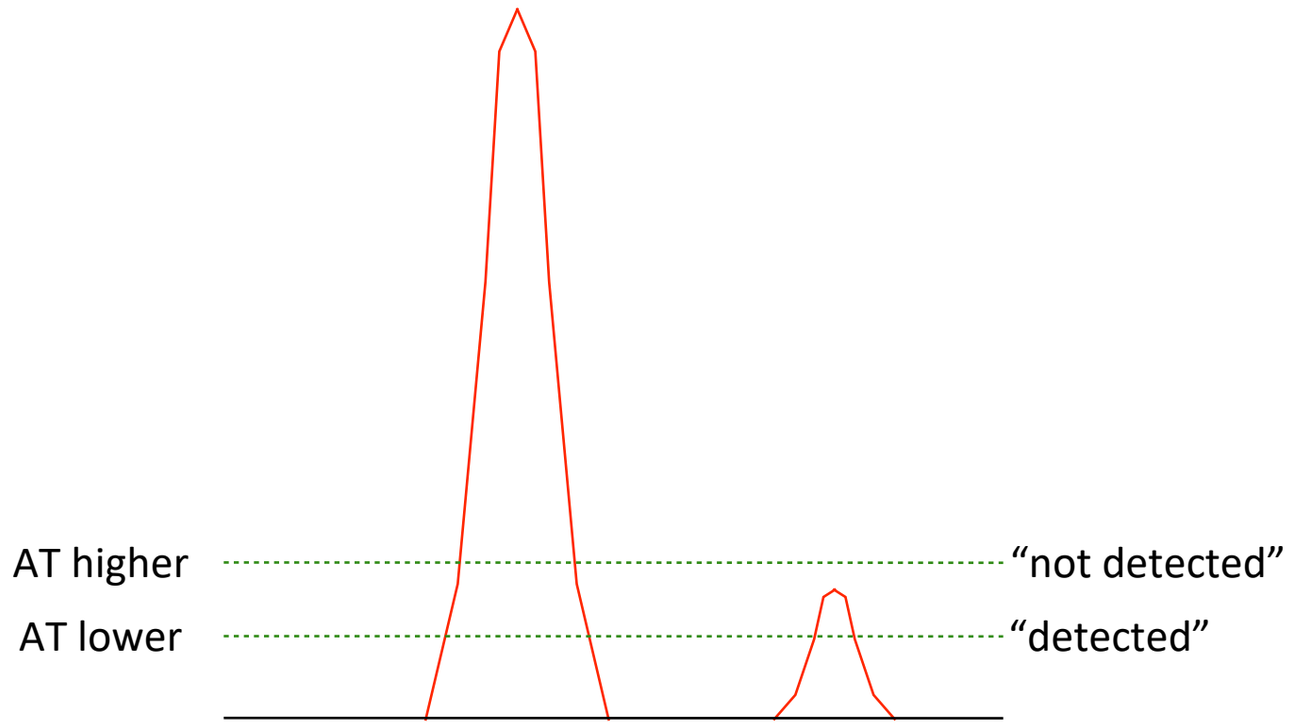


Greatest sources of variation extrinsic (originate outside the software)

- Human factors
  - Analytical threshold (how much data is detected)
    - Not modeled by any software
    - Can affect apparent number of contributors
    - Can affect apparent allelic drop-out (missing information)
- Can affect final result
  - Statistic
  - Potential contributors



Impact of analytical threshold



What is the Question?

If you don’t ask the right question,
you won’t get the right answer,
regardless of the
sophistication and power of your analysis

Greatest sources of variation extrinsic (originate outside the software)

- Human factors
  - Choosing hypotheses to be compared
    - Generally chosen by primary laboratory
    - Some factors to be considered
      - Number of contributors (NoC)
      - Assumed contributors
      - Concurrent contributors
      - Relatives

Fair and equal access

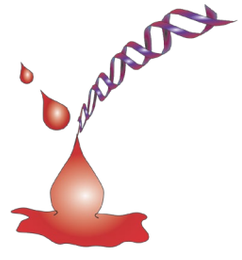
- Science requires independent review to determine (or confirm) the capabilities and limitations of a software program

- Justice requires fair and equal access to technologies used in investigating and prosecuting crimes

  - Because individuals of low socio-economic status are over-represented in the criminal justice system, they are, by definition, disproportionately impacted by any issues with the technologies

- Because of restrictions in access to the commercial proprietary programs, true independent testing and comparison have never been possible

  - Only owners of the commercial programs can perform comparison testing with other programs



Fair and equal access

- A plethora of PG software programs now exist
  - The commercial programs used by most U.S. labs are extraordinarily expensive
    - At least one won’t sell to anyone they deem a competitor
      - Trainings also closed to perceived competitors
    - Research licenses highly curated or flatly denied
- Free of charge, open-source programs available
  - Use different algorithms

Fair and equal access

- Who is independent?
  - Manufacturer should not be the arbiter
- Who should have access to the software?
  - Make available for research use at reasonable cost
    - Manufacturer should not be the arbiter of who can use the software for research
    - Manufacturer should not be the arbiter of how the software is used for research
  - Make available for case review at reasonable cost
    - Manufacturer should not be the arbiter of who can use the software for case review

Summary

- All PG software tools have capabilities and limitations that encompass both technological and human factors

- Commercial programs that are expensive and proprietary result in unequal access that can hamper the ability of independent parties to explore these capabilities and limitations, perform comparison studies, and also to review casework

- Options to provide more equitable access for both research and casework should be explored

Thank you for your attention

Appendices

Likelihood Ratio

- The results of PG are reported as a likelihood ratio
- Example of correct phraseology

The probability of seeing this mixture is 1 billion times more likely IF it originates from Mr. S than IF it originates from a random unrelated individual.

Likelihood Ratio

$$\frac{\text{Hypothesis 1}}{\text{Hypothesis 2}} \rightarrow \frac{\text{H1}}{\text{H2}}$$

H1: Hypothesized individual is the true contributor

H2: Unknown individual is the true contributor

$$LR = \frac{\text{Probability of the Evidence IF H1}}{\text{Probability of the Evidence IF H2}}$$

$$LR = \frac{P(E|H1)}{P(E|H2)}$$