

DNA-led investigation through computer interpretation of evidence

Pennsylvania State Police
Training Seminar
Hershey, PA
April, 2014

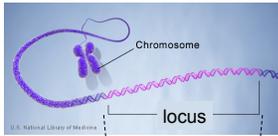
Mark W Perlin, PhD, MD, PhD
Cybergenetics, Pittsburgh, PA



Cybergenetics

Cybergenetics © 2003-2014

DNA genotype



mother allele



father allele



A genetic locus has two DNA sentences, one from each parent.

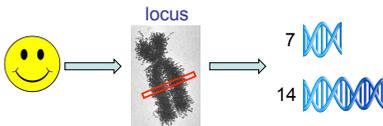
An **allele** is the number of repeated words.

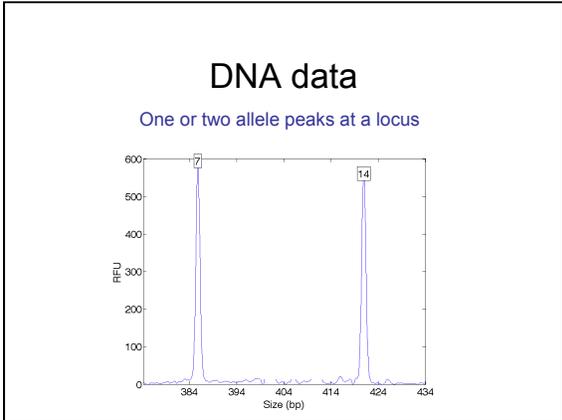
A **genotype** at a locus is a pair of alleles.

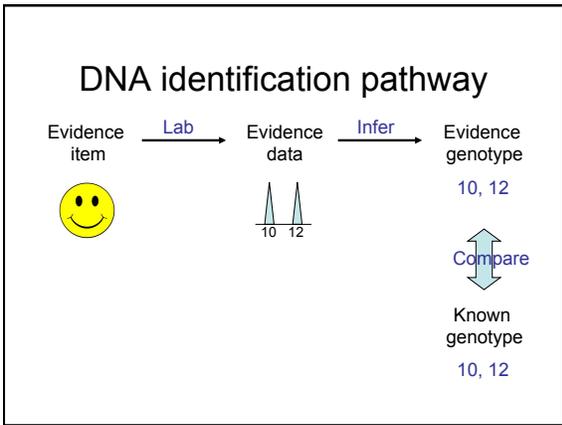
10, 12

Many alleles allow for many many allele pairs. A person's genotype is relatively unique.

One person, one genotype





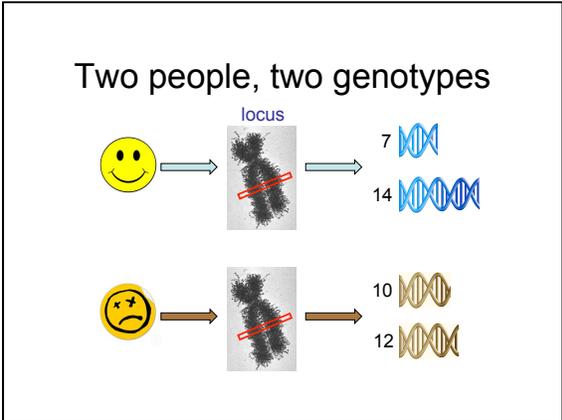


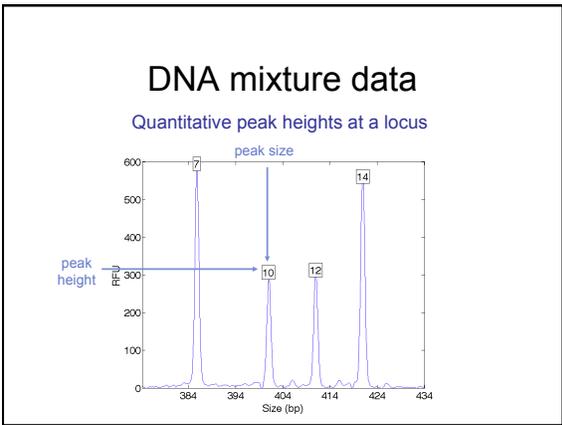
Match information

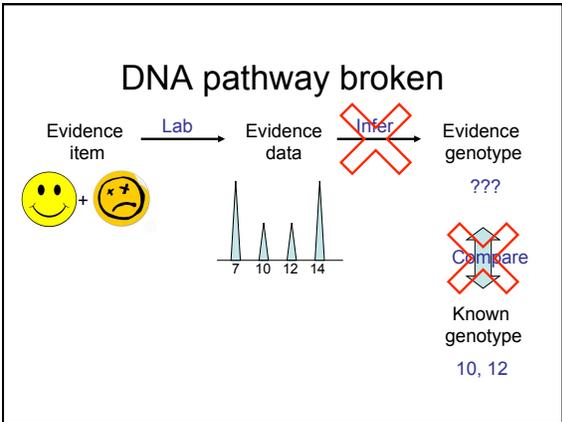
At the suspect's genotype,
identification vs. coincidence?

after (evidence) \uparrow data \downarrow before (population)

$$\frac{\text{Prob}(\text{evidence matches suspect})}{\text{Prob}(\text{coincidental match})} = \frac{100\%}{5\%} = 20$$







Human interpretation issues

Evidence

- call good data inconclusive
- peaks are too low for them
- too many contributors to handle
- potential examination bias

Database

- hit by association, not by match
- comparison: make false hits
- restrict upload: lose true hits

TrueAllele® Casework

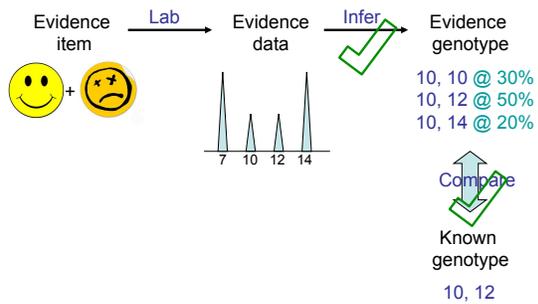
Evidence

- preserve data information
- use all peaks, high or low
- any number of contributors
- entirely objective, no bias

Database

- hit based on LR match statistic
- sensitive: find true hits
- specific: only true hits

DNA pathway restored

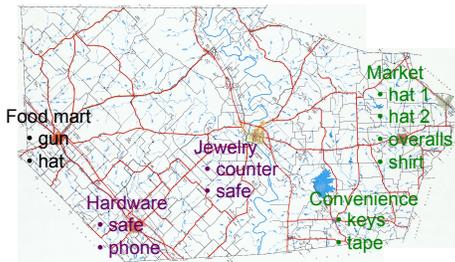


Match information preserved

At the suspect's genotype,
identification vs. coincidence?

$$\begin{array}{l}
 \text{after} \\
 \text{(evidence)} \\
 \uparrow \text{data} \\
 \text{before} \\
 \text{(population)}
 \end{array}
 \frac{\text{Prob}(\text{evidence matches suspect})}{\text{Prob}(\text{coincidental match})} = \frac{50\%}{5\%} = 10$$

Gang DNA from 5 crime scenes



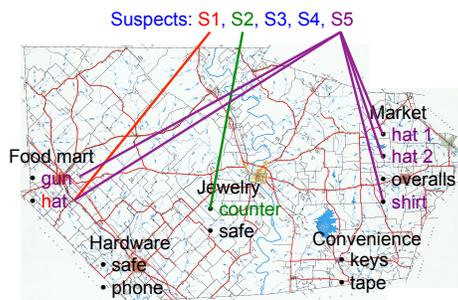
Laboratory DNA processing

- | | |
|-------------------|--------------------|
| 12 evidence items | 10 reference items |
| Scene 1 • gun | 5 victims |
| • hat | • V1 |
| Scene 2 • safe | • V2 |
| • phone | • V3 |
| Scene 3 • counter | • V4 |
| • safe | • V5 |
| Scene 4 • keys | 5 suspects |
| • tape | • S1 |
| Scene 5 • hat 1 | • S2 |
| • hat 2 | • S3 |
| • overalls | • S4 |
| • shirt | • S5 |

Cybergentics TrueAllele® timeline

Day	Activity
1	Received evidence data from lab
2	Started computer processing
4	Replicated evidence results
9	Received known references
10	Calculated DNA match statistics
12	Reported match results to lab

TrueAllele computer matches



DNA match statistic:
553 million

People of California v. Charles Lewis Lawton
and Dupree Donyell Langston
November, 2012
Bakersfield, CA

Admissibility hearing
and trial testimony



Peer-reviewed validations

Perlin MW, Sineelnikov A. An information gap in DNA evidence interpretation. *PLoS ONE*. 2009;4(12):e8327.

Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW. Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences*. 2011;56(6):1430-47.

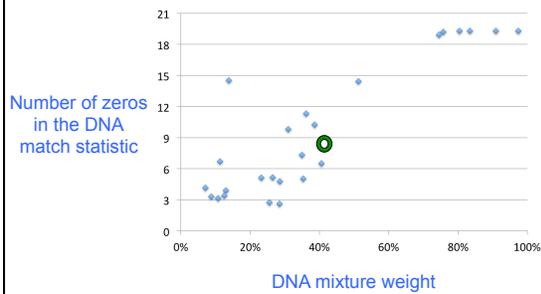
Ballantyne J, Hanson EK, Perlin MW. DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations: Combining quantitative data for greater identification information. *Science & Justice*. 2013;53(2):103-14.

Perlin MW, Belrose JL, Duceman BW. New York State TrueAllele® Casework validation study. *Journal of Forensic Sciences*. 2013;58(6):1458-66.

Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S. TrueAllele® Casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. *PLOS ONE*. 2014;9(3):e92837.

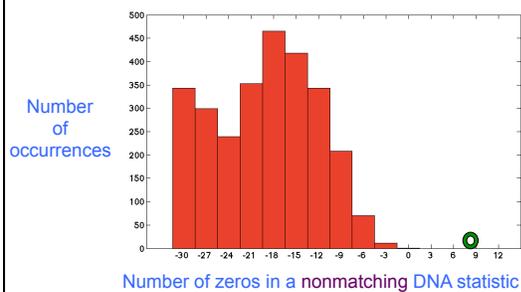


Expected match statistic



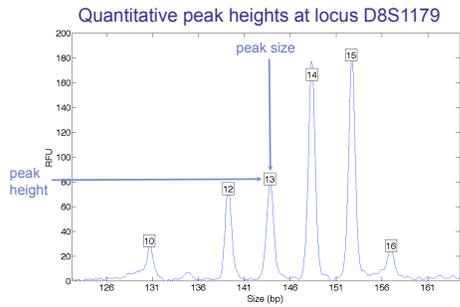


Specific match statistic



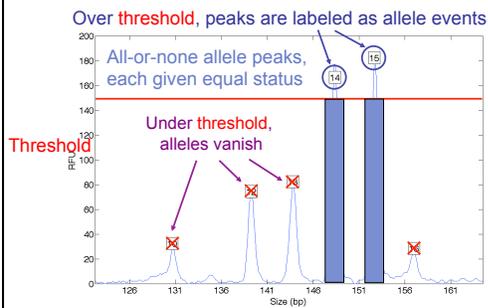


Computers can use all the data



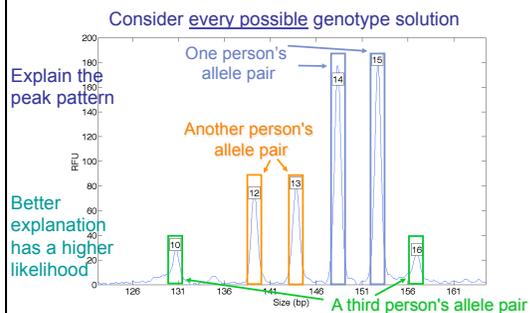


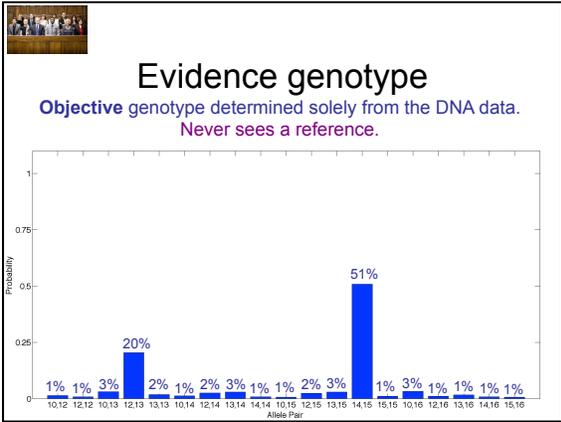
People may use less of the data

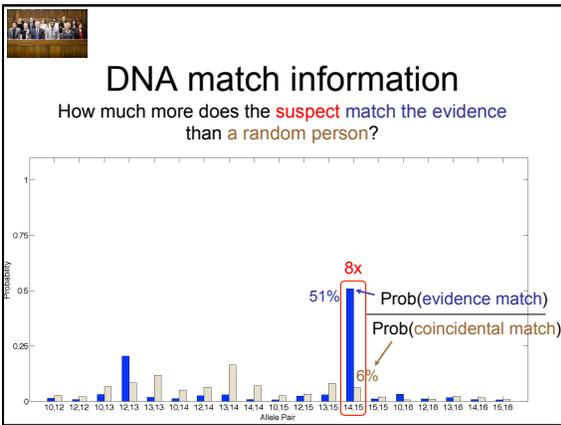


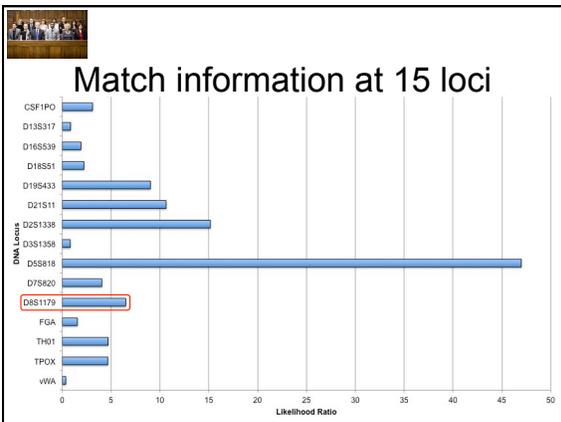


How the computer thinks











Is the suspect in the evidence?

A match between the front counter
and Dupree Langston is:

553 million times more probable than
a coincidental match to an unrelated Black person

731 million times more probable than
a coincidental match to an unrelated Caucasian person

208 million times more probable than
a coincidental match to an unrelated Hispanic person



TrueAllele reinterpretation

The Washington Post

Virginia reevaluates DNA evidence in 375 cases

July 16, 2011

“Mixture cases are their own little nightmare,” says
William Vosburgh, director of the D.C. police’s crime
lab. “It gets really tricky in a hurry.”

“If you show 10 colleagues a mixture,
you will probably end up with 10 different answers”
Dr. Peter Gill, Human Identification E-Symposium, 2005

Virginia mixture study

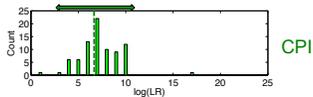
- 72 criminal cases
- 92 evidence items
- 111 genotype comparisons

Criminal offense

- 18 homicide
- 12 robbery
- 6 sexual assault
- 20 weapon

Old manual interpretation

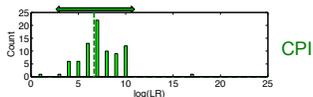
6.83 (2.22)
6.68 million



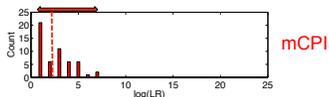
Combined Probability of Inclusion (CPI)
analytical threshold

New manual interpretation

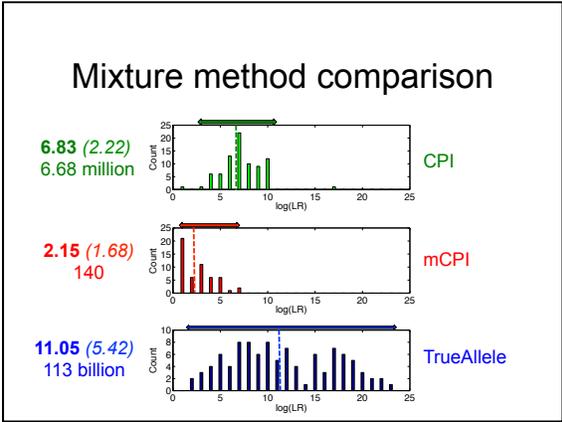
6.83 (2.22)
6.68 million



2.15 (1.68)
140



modified Combined Probability of Inclusion (mCPI)
stochastic threshold
analytical threshold



OPEN ACCESS Freely available online **PLOS ONE**

TrueAllele Casework on Virginia DNA Mixture Evidence: Computer and Manual Interpretation in 72 Reported Criminal Cases

Mark W. Perlin^{1*}, Kiersten Dormer¹, Jennifer Hornyak¹, Lisa Schiermeier-Wood², Susan Greenspoon²

1 Cybergenetics, Pittsburgh, Pennsylvania, United States of America, 2 Department of Forensic Science, Richmond, Virginia, United States of America

Abstract

Mixtures are a commonly encountered form of biological evidence that contain DNA from two or more contributors. Laboratory analysis of mixtures produces data signals that usually cannot be separated into distinct contributor genotypes. Computer modeling can resolve the genotypes up to probability, reflecting the uncertainty inherent in the data. Human analysts address the problem by simplifying the quantitative data in a threshold process that discards considerable identification information. Elevated stochastic threshold levels potentially discard more information. This study examines three different mixture interpretation methods. In 72 criminal cases, 111 genotype comparisons were made between 92 mixture items and relevant reference samples. TrueAllele computer modeling was done on all the evidence samples, and documented in DNA match reports that were provided as evidence for each case. Threshold-based Combined Probability of Inclusion (CPI) and stochastically modified CPI (mCPI) analyses were performed as well. TrueAllele's identification information in 101 positive matches was used to assess the reliability of its modeling approach. Comparison was made with 81 CPI and 53 mCPI DNA match statistics that were manually derived from the same data. There were statistically significant differences between the DNA interpretation methods. TrueAllele gave an average match statistic of 113 billion, CPI averaged 6.68 million, and mCPI averaged 140. The computer was highly specific, with a false positive rate under 0.000%. The modeling approach was precise, having a factor of two within-group standard deviation. TrueAllele accuracy was indicated by having uniformly distributed match statistics over the data set. The computer could make genotype comparisons that were impossible or impractical using manual methods. TrueAllele computer interpretation of DNA mixture evidence is sensitive, specific, precise, accurate and more informative than manual interpretation alternatives. It can determine DNA match statistics when threshold-based methods cannot. Improved forensic science computation can affect criminal cases by providing reliable scientific evidence.

Citation: Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S (2014) TrueAllele Casework on Virginia DNA Mixture Evidence: Computer and Manual Interpretation in 72 Reported Criminal Cases. PLoS ONE 9(3): e92837. doi:10.1371/journal.pone.0092837

TrueAllele Virginia outcomes

144 cases analyzed
72 case reports – 10 trials

City	Court	Charge	Sentence
Richmond	Federal	Weapon	50 years
Alexandria	Federal	Bank robbery	90 years
Quantico	Military	Rape	3 years
Chesapeake	State	Robbery	26 years
Arlington	State	Molestation	22 years
Richmond	State	Homicide	35 years
Fairfax	State	Abduction	33 years
Norfolk	State	Homicide	8 years
Charlottesville	State	Homicide	15 years
Hampton	State	Home invasion	5 years

TrueAllele in criminal trials

Over 150 case reports filed on DNA evidence

Court testimony:

- state
- federal
- military
- foreign

Crimes:

- armed robbery
- child abduction
- child molestation
- murder
- rape
- terrorism
- weapons

TrueAllele usage in the US



Casework system
Interpretation services
Admissibility hearing

All the DNA, all the time

Objective, reliable truth-seeking tool

- solves the DNA mixture problem
- handles low-copy and degraded DNA
- provides accurate DNA match statistics
- automates DNA evidence interpretation

Currently used to:

- eliminate DNA backlogs
- reduce forensic costs
- solve crimes
- find criminals
- convict the guilty
- free the innocent
- create a safer society

TrueAllele today

Invented math & algorithms	20 years
Developed computer systems	15 years
Support users and workflow	10 laboratories
Used routinely in casework	3 labs
Validate system reliability	20 studies
Educate the community	50 talks
Train & certify analysts	200 students
Go to court for admissibility	5 hearings
Testify about LR results	20 trials
Educate lawyers and laymen	1,000 people
Make the ideas understandable	150 reports

More TrueAllele information

<http://www.cybgen.com/information>



- Courses
- Newsletters
- Newsroom
- Presentations
- Publications

<http://www.youtube.com/user/TrueAllele>
TrueAllele YouTube channel