DNA Knowledge, DNA Power:
How Computers Interpret Evidence
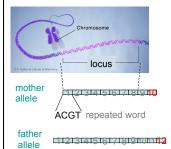
Cybergenetics Webinar
August, 2013

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

Cybergenetics

Cybergenetics © 2003-2013

DNA genotype

Chromosome

locus

mother allele

1 2 3 4 5 6 7 8 9 10

ACGT repeated word

father allele

1 2 3 4 5 6 7 8 9 10 11 12

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.
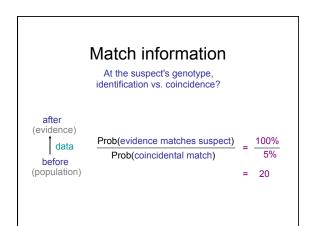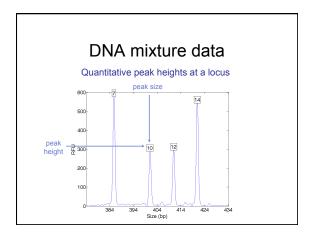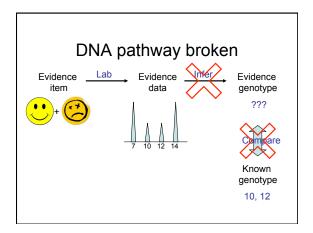

DNA identification pathway

Evidence item

Lab →

Evidence data

10 12

Infer →

Evidence genotype

10, 12

Compare

Known genotype

10, 12

Match information

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

after
(evidence)

↑ data

before
(population)

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

$$= 20$$

DNA mixture data

Quantitative peak heights at a locus

peak size

peak height

RFU

Size (bp)

7   10   12   14

DNA pathway broken

Evidence item →(Lab)→ Evidence data →(Infer ✗)→ Evidence genotype ???

✗ Compare

Known genotype
10, 12

7 10 12 14

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
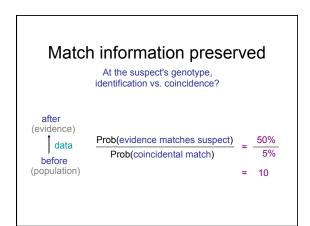- restrict upload: lose true hits
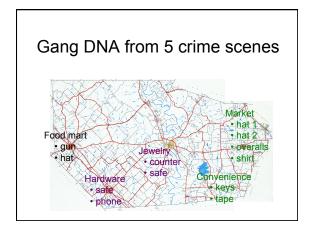
TrueAllele® Casework

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

Evidence item → Lab → Evidence data → Infer → Evidence genotype

10, 10 @ 30%
10, 12 @ 50%
10, 14 @ 20%

Compare

Known genotype

10, 12

Match information preserved

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

after
(evidence)

data

before
(population)

$$\frac{\text{Prob(evidence matches suspect)}}{\text{Prob(coincidental match)}} = \frac{50\%}{5\%}$$

$$= 10$$

Gang DNA from 5 crime scenes

Market
• hat 1
• hat 2
• overalls
• shirt

Food mart
• gun
• hat

Jewelry
• counter
• safe

Hardware
• safe
• phone

Convenience
• keys
• tape

Laboratory DNA processing

12 evidence items

Scene 1 • gun
         • hat
Scene 2 • safe
         • phone
Scene 3 • counter
         • safe
Scene 4 • keys
         • tape
Scene 5 • hat 1
         • hat 2
         • overalls
         • shirt

10 reference items

5 victims
• V1
• V2
• V3
• V4
• V5

5 suspects
• S1
• S2
• S3
• S4
• S5

Cybergenetics 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

Suspects: S1, S2, S3, S4, S5

Food mart
• gun
• hat

Jewelry
counter
• safe

Hardware
• safe
• phone

Market
• hat 1
• hat 2
• overalls
• shirt

Convenience
• keys
• tape

DNA match statistic:
553 million

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

**Admissibility hearing
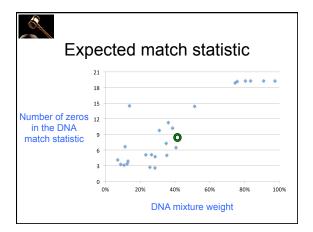and trial testimony**
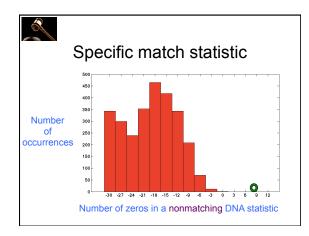
Peer-reviewed validations

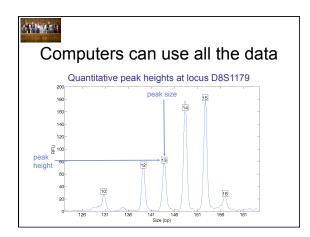Perlin MW, Sinelnikov A. An information gap in DNA evidence interpretation. *PLoS ONE*. 2009;4(12):e8327.

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.
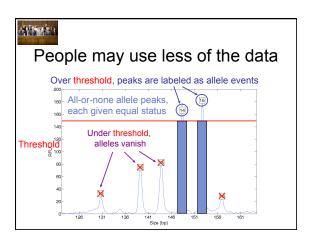
Ballantyne J, Hanson EK, Perlin MW. DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations: Combining quantitative data for greater identification information. *Science & Justice*. 2013;53(2):103-14.
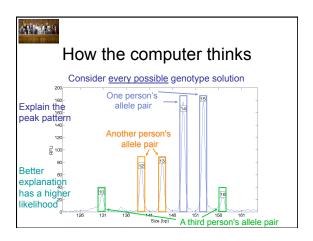
Perlin MW, Belrose JL, Duceman BW. New York State TrueAllele® Casework validation study. *Journal of Forensic Sciences*. 2013;58(6):in press.

Expected match statistic

Number of zeros in the DNA match statistic

DNA mixture weight

Specific match statistic

Number of occurrences

Number of zeros in a nonmatching DNA statistic

Computers can use all the data

Quantitative peak heights at locus D8S1179

peak size

peak height

RFU

10 12 13 14 15 16

126 131 136 141 146 151 156 161
Size (bp)

People may use less of the data

Over threshold, peaks are labeled as allele events

All-or-none allele peaks, each given equal status

14 15

Under threshold, alleles vanish

Threshold

RFU

126 131 136 141 146 151 156 161
Size (bp)

How the computer thinks

Consider every possible genotype solution

One person's allele pair

Explain the peak pattern

Another person's allele pair

14 15

12 13

Better explanation has a higher likelihood

10 16

RFU

126 131 136 141 146 151 156 161
Size (bp)

A third person's allele pair

Evidence genotype
Objective genotype determined solely from the DNA data.
Never sees a reference.

DNA match information
How much more does the suspect match the evidence
than a random person?

8x
51% Prob(evidence match)
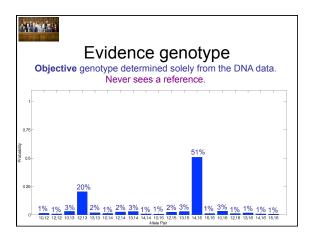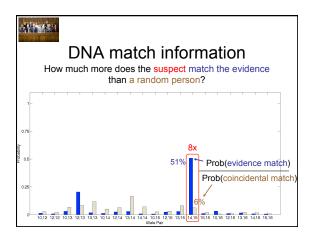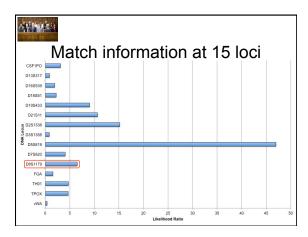Prob(coincidental match)
6%

Match information at 15 loci

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
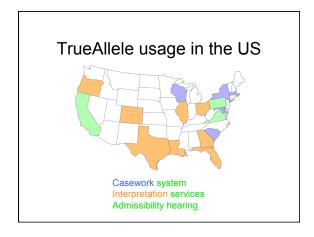

TrueAllele in criminal trials

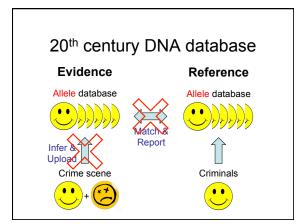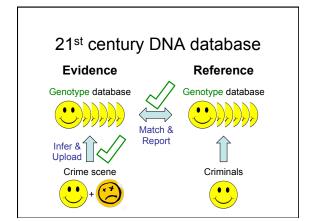Over 100 case reports filed on DNA evidence

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

20th century DNA database

Evidence — Reference

Allele database — Allele database

Match & Report

Infer & Upload

Crime scene — Criminals

21st century DNA database

Evidence — Reference

Genotype database — Genotype database

Match & Report

Infer & Upload

Crime scene — Criminals

TrueAllele computer age

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

More information

http://www.cybgen.com/information
• Courses
• Newsletters
• Newsroom
• Presentations
• Publications

http://www.cybgen.com/support/sending_data.shtml

Cybergenetics   perlin@cybgen.com