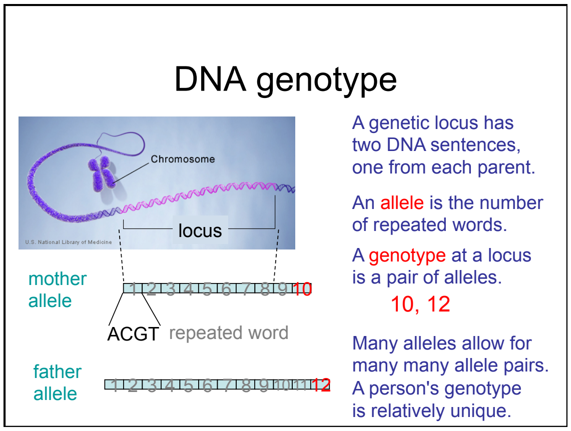
TrueAllele® Computing: All the DNA, all the time

Continuing Professional Development
Sydney, Australia
March, 2014

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

Cybergenetics



DNA genotype

Chromosome
U.S. National Library of Medicine
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

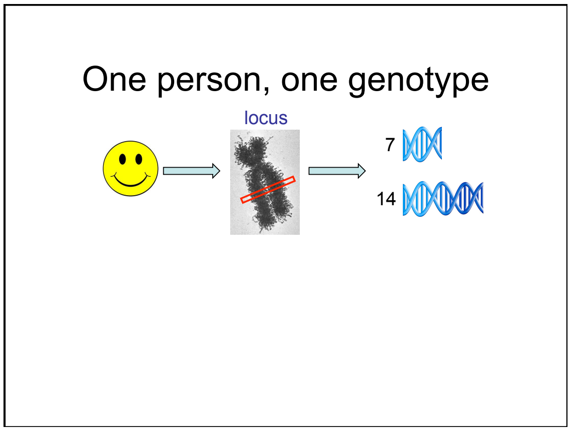
A genetic locus has two DNA sentences, one from each parent.

An allele is the number of repeated words.

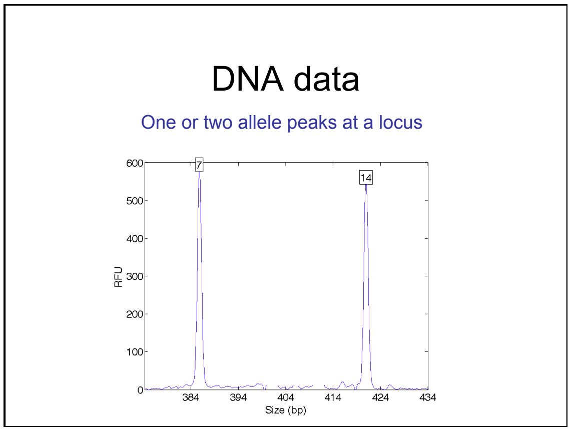
A genotype at a locus is a pair of alleles.

10, 12

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

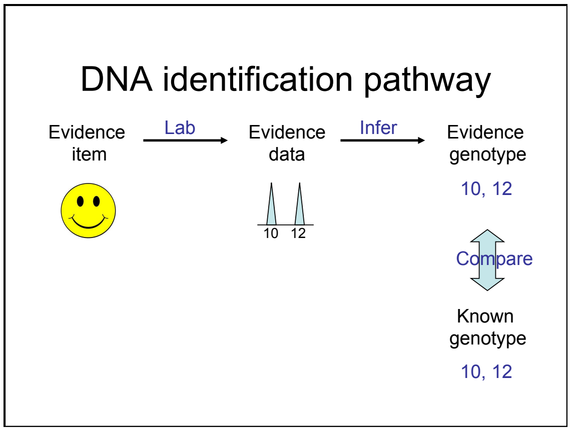


One person, one genotype



DNA data

One or two allele peaks at a locus



DNA identification pathway

Evidence item $\xrightarrow{\text{Lab}}$ Evidence data $\xrightarrow{\text{Infer}}$ Evidence genotype

10, 12

Compare

Known genotype

10, 12

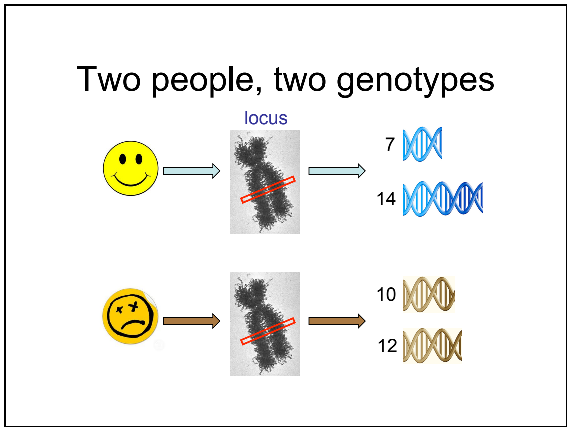
Match information

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

after
(evidence)
↑ data
before
(population)

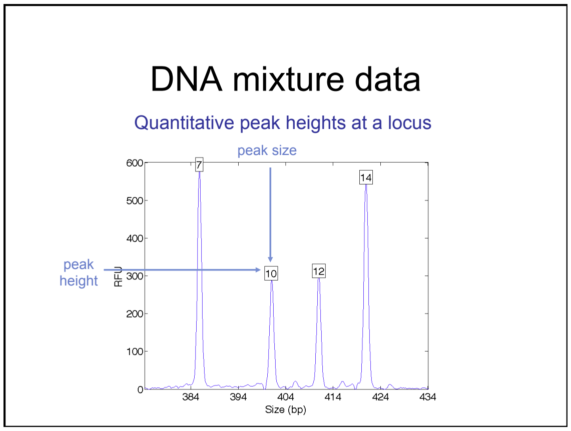
$$\frac{\text{Prob(evidence matches suspect)}}{\text{Prob(coincidental match)}} = \frac{100\%}{5\%}$$

$$= 20$$



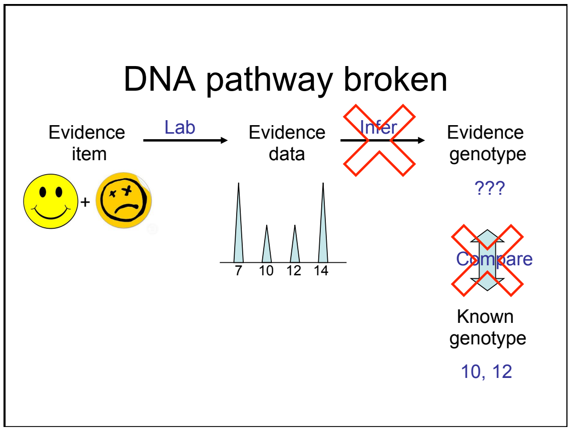
Two people, two genotypes

locus
7
14
10
12



DNA mixture data

Quantitative peak heights at a locus



DNA pathway broken

Evidence item
Lab
Evidence data
Infer
Evidence genotype
???
+
7
10
12
14
Compare
Known genotype
10, 12

Human interpretation issues

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

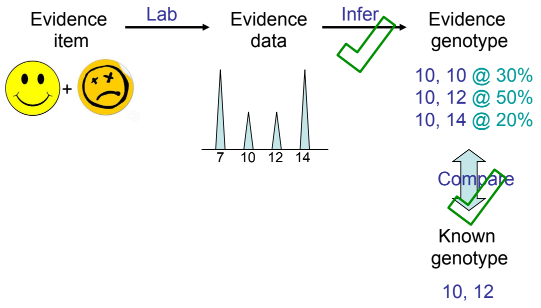
Database
- hit by association, not by match
- comparison: make false hits
- restrict upload: lose true hits

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
7
10
12
14
Compare
Known genotype

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

10, 12

Match information preserved

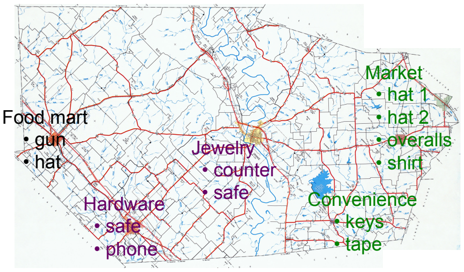
At the suspect's genotype,
identification vs. coincidence?

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


Food mart
• gun
• hat
Hardware
• safe
• phone
Jewelry
• counter
• safe
Market
• hat 1
• hat 2
• overalls
• shirt
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

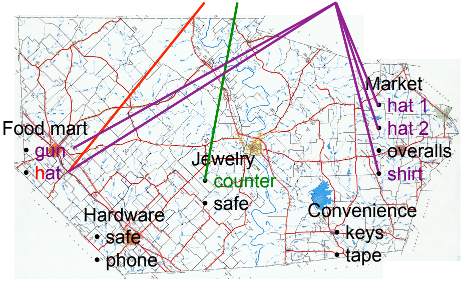
5 suspects
- S1
- S2
- S3
- S4
- S5

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.

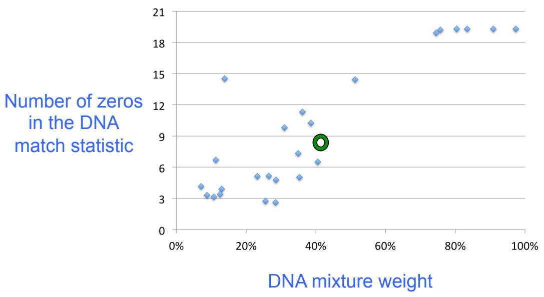
Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW. Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences*. 2011;56(6):1430-47.

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.

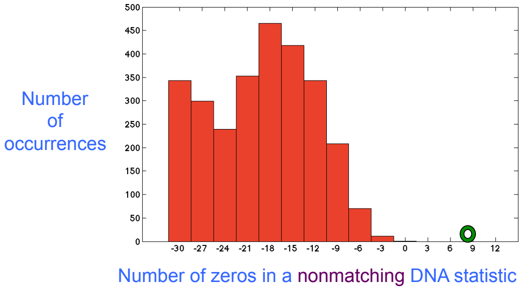
Perlin MW, Belrose JL, Duceman BW. New York State TrueAllele® Casework validation study. *Journal of Forensic Sciences*. 2013;58(6):1458-66.

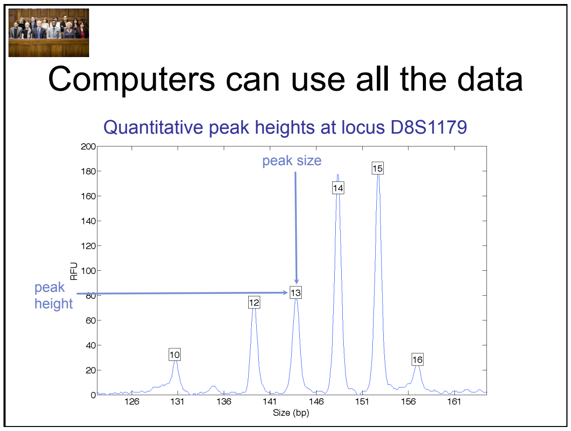
Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S. TrueAllele® Casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. *PLOS ONE*. 2014:*in press*.

Expected match statistic



Specific match statistic



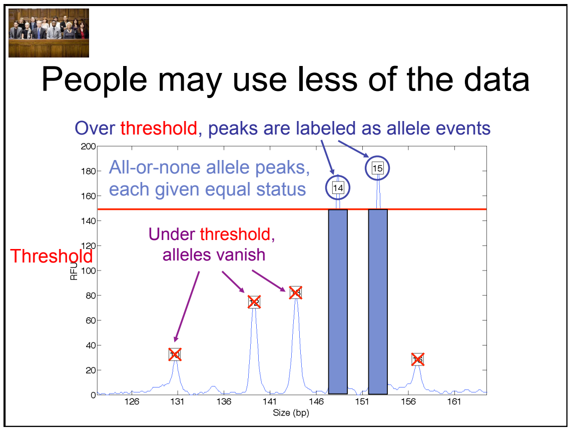


Computers can use all the data

Quantitative peak heights at locus D8S1179

peak size

peak height



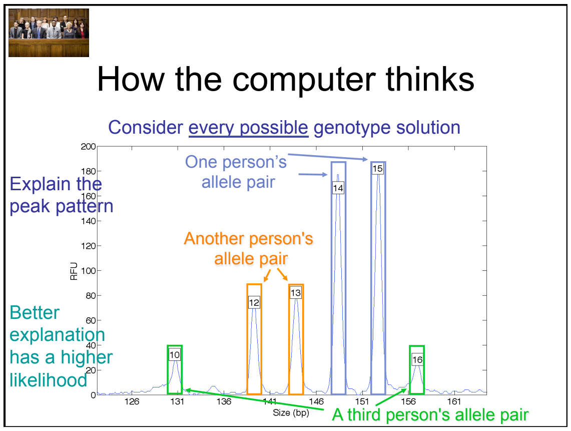
People may use less of the data

Over threshold, peaks are labeled as allele events

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

Threshold

Under threshold, alleles vanish



How the computer thinks

Consider every possible genotype solution

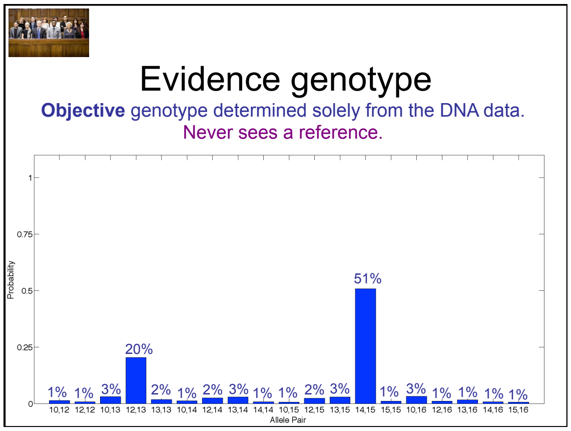
Explain the peak pattern

One person's allele pair

Another person's allele pair

Better explanation has a higher likelihood

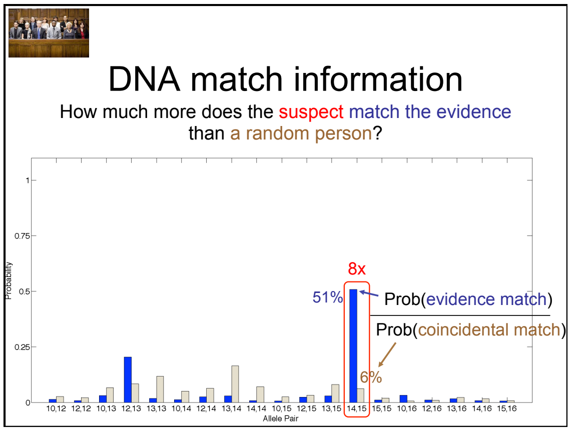
A third person's allele pair



Evidence genotype

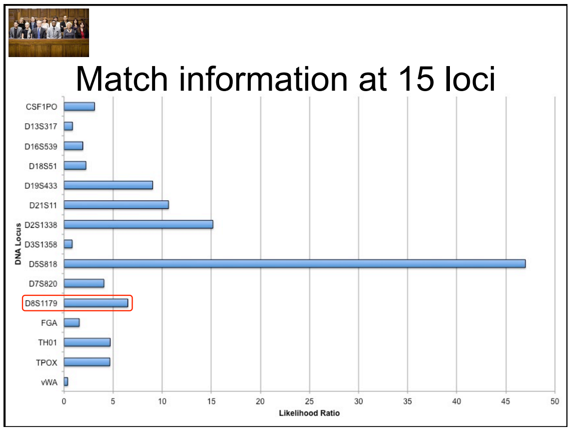
Objective genotype determined solely from the DNA data.

Never sees a reference.

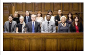


DNA match information

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



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

NEWS
17
IN HIGH DEFINITION
IN HD
WORKING IN THE SPIRIT OF THE GOLDEN EMPIRE

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."

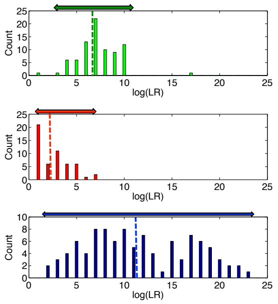
"If you show 10 colleagues a mixture,
you will probably end up with 10 different answers"
Dr. Peter Gill, Human Identification E-Symposium, 2005

Virginia validation study

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

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

Mixture method comparison

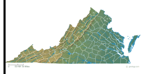


6.83 *(2.22)*
6.68 million
CPI

2.15 *(1.68)*
140
mCPI

11.05 *(5.42)*
113 billion
TrueAllele

DNA mixture crisis

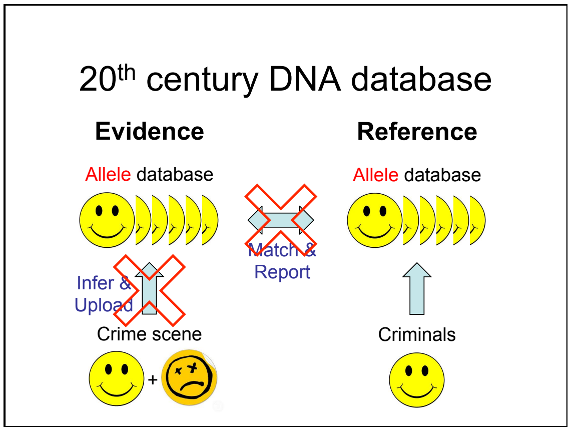


375 cases/year x 4 years = 1,500 cases
320 M in US / 8 M in VA = 40 factor
1,500 cases x 40 factor = 60,000 inconclusive

1,000 cases/year x 4 years = 4,000 cases
320 M in US / 8 M in NY = 40 factor
4,000 cases x 40 factor = 160,000 inconclusive

+ underreporting of DNA match statistics

DNA evidence data in 100,000 cases
Collected, analyzed & paid for – but unused



20th century DNA database

Evidence

Allele database

Infer & Upload

Crime scene

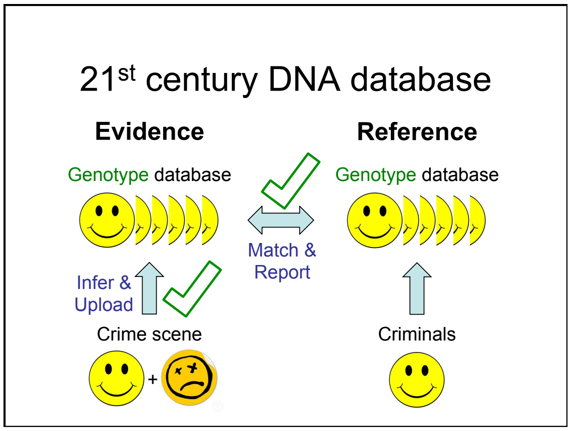
+

Match & Report

Reference

Allele database

Criminals



21st century DNA database

Evidence

Genotype database

Infer & Upload

Crime scene

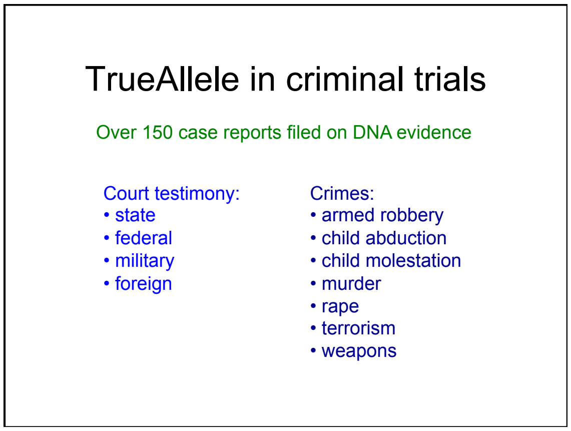
+

Match & Report

Reference

Genotype database

Criminals



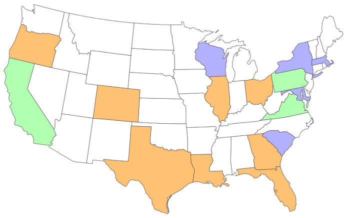
TrueAllele in criminal trials

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

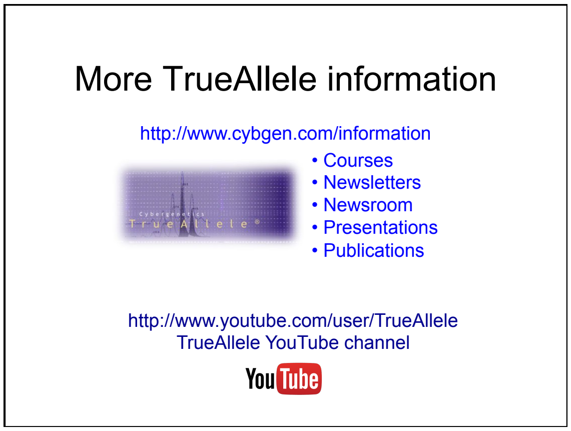
All the DNA, all the time

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