1

Finding truth in evidence: Computational intelligence

National Courts and Sciences Institute
Kansas City, Missouri
October 2024

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

Cybergenetics

2

New York v John Wakefield

In April, 2010, Brett Wentworth (41) was found dead in his apartment, strangled with an electric guitar cord.

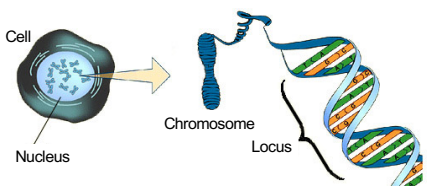
The police collected biological evidence was from amp cord sections, plus his shirt collar and forearm.

The New York State Police lab examined the DNA mixtures.

Comparing the evidence with suspect Wakefield's profile found very little DNA match information.

3

DNA biology



4

Short tandem repeat

23 volumes in cell's DNA encyclopedia

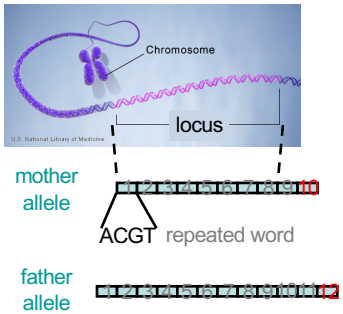
DNA locus paragraph

Take me out to the ball game
take me out with the crowd
buy me some peanuts and Cracker Jack
I don't care if I never get back
let me
root root root root root root root root root root root
for the home team,
if they don't win, it's a shame for it's one, two, three strikes, you're out
at the old ball game

"root" repeated 10 times, so allele length is 10 repeats

5

DNA genotype



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.

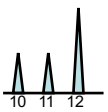
10, 12

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

6

DNA evidence interpretation

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



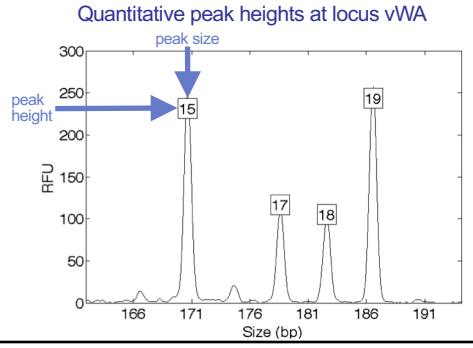
10, 12 @ 50%
11, 12 @ 30%
12, 12 @ 20%

Compare

Known genotype
10, 12

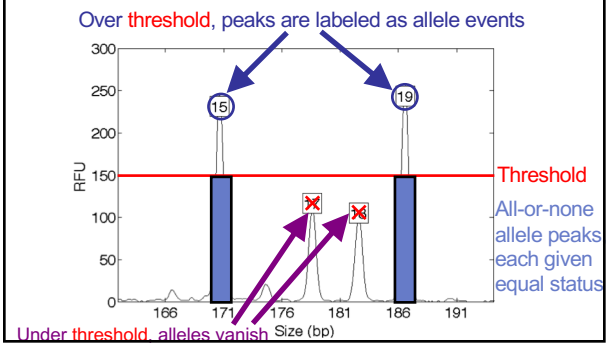
7

Computers can use all the data



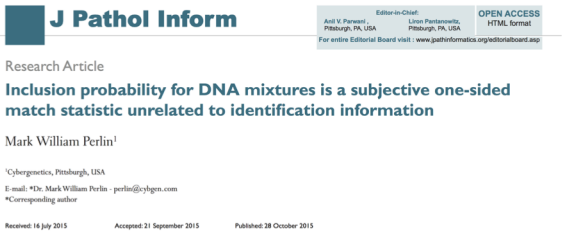
8

People may use less of the data



9

Human review CPI match statistic: subjective, biased & uninformative



J Pathol Inform

Research Article

Inclusion probability for DNA mixtures is a subjective one-sided match statistic unrelated to identification information

Mark William Perlin[1]

[1]Cybergenetics, Pittsburgh, USA

E-mail: *Dr. Mark William Perlin - perlin@cybgen.com
*Corresponding author

Received: 16 July 2015 Accepted: 21 September 2015 Published: 28 October 2015

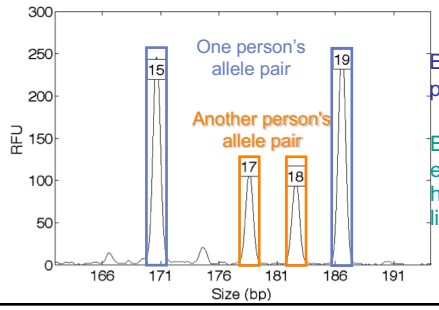
10

Finding truth through science

11

How the computer thinks

Consider every possible genotype solution



Explain the peak pattern

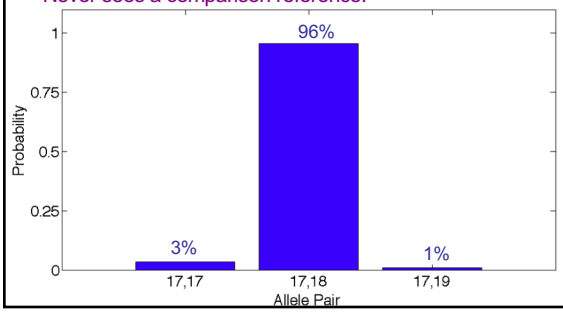
Better explanation has a higher likelihood

12

Evidence genotype

Objective genotype determined solely from the DNA data.

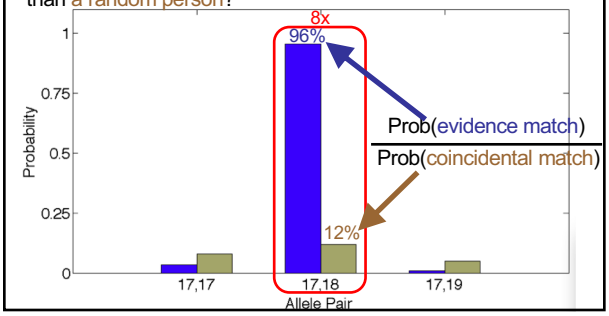
Never sees a comparison reference.



13

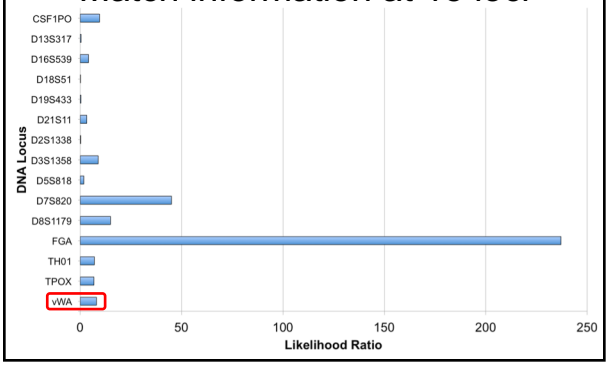
DNA match information

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



14

Match information at 15 loci



15

Is the suspect in the evidence?

A match between the amp cord
and John Wakefield is:

5.88 billion times more probable than
a coincidental match to an unrelated Black person

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

2.25 billion times more probable than
a coincidental match to an unrelated Hispanic person

16

Is the victim in the evidence?

A match between the amp cord
and Brett Wentworth is:

221 quintillion times more probable than
a coincidental match to an unrelated Black person

478 quadrillion times more probable than
a coincidental match to an unrelated Caucasian person

906 quadrillion times more probable than
a coincidental match to an unrelated Hispanic person

17

Match statistics

		052B	188, 189
Item	Description	Brett Wentworth	John Wakefield
004A-C	Swabs 0-3 ft of amp cord	18.81	-0.10
004D-F	Swabs 3-6 ft of amp cord	18.81	0.15
004G-I	Swabs 6-9 ft of amp cord	18.81	2.90
004J-L	Swabs 9-12 ft of amp cord	18.81	-16.69
004M-O	Swabs 12-15 ft of amp cord	17.68	8.48
004P-R	Swabs 15-18 ft of amp cord	18.70	-1.49
004S-T	Swabs 18-20 ft of amp cord	18.81	-1.09
045A	Shirt collar, outside rear	7.92	18.88
045C	Shirt collar, outside front	18.81	10.07
052F1-2	Victim forearm swabs	18.81	6.36

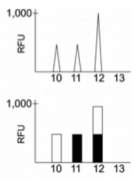
18

Explaining DNA mixtures

JOURNAL OF FORENSIC SCIENCES

Mark W. Perlin,[1] Ph.D., M.D., Ph.D. and Beata Szabady,[1] Ph.D. 2001

Linear Mixture Analysis: A Mathematical
Approach to Resolving Mixed DNA Samples



DNA peak height data

Sum of genotypes

19

TrueAllele® computer solution

- Accurate. 43 validation studies, 8 published
- Objective. Workflow removes human bias
- Accepted. Reported in 46 states, used by 10 labs
- Transparent. Give math, software (4GB DVD)
- Neutral. Can statistically include or exclude

Peer-reviewed validation studies[20]

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

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):103-114.

Perlin MW, Hornyak J, Sugimoto G, Miller K. TrueAllele® genotype identification on DNA mixtures containing up to five unknown contributors. *Journal of Forensic Sciences*. 2015;60(4):857-868.

Greenspoon SA, Schiermeier-Wood L, Jenkins BC. Establishing the limits of TrueAllele® Casework: a validation study. *Journal of Forensic Sciences*. 2015;60(5):1263-1276.

Bauer DW, Butt N, Hornyak JM, Perlin MW. Validating TrueAllele® interpretation of DNA mixtures containing up to ten unknown contributors. *Journal of Forensic Sciences*. 2020; 65(2):380-398.

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):1430-1447.

Perlin MW, Belrose JL, Duceman BW. New York State TrueAllele® Casework validation study. *Journal of Forensic Sciences*. 2013;58(6):1458-1466.

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.

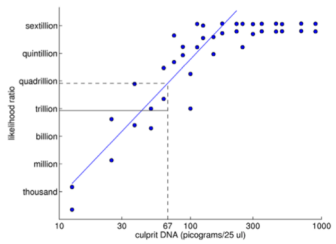
21

TrueAllele predictability

An Information Gap in DNA Evidence Interpretation

Mark W. Perlin[1]*, Alexander Sinelnikov[2]

PLoS one 2009



22

TrueAllele reliability

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

Mark W. Perlin[1]*, Kiersten Dormer[1], Jennifer Hornyak[1], Lisa Schiermeier-Wood[2], Susan Greenspoon[2]

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

PLoS one 2014

Validation axes
- sensitive
- specific
- reproducible

23

Sensitivity

The extent to which interpretation identifies the correct person

True DNA mixture inclusions

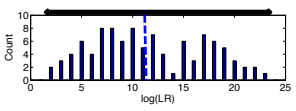
101 reported genotype matches
82 with DNA statistic over a million

24

TrueAllele sensitivity

log(LR) match distribution

11.05 *(5.42)*
113 billion

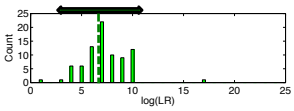


TrueAllele

25

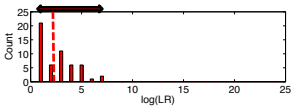
Comparison with human review

6.83 *(2.22)*
6.68 million



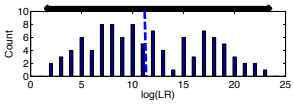
CPI

2.15 *(1.68)*
140



mCPI

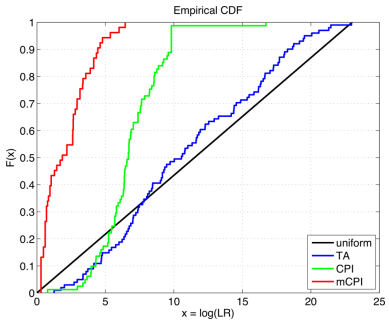
11.05 *(5.42)*
113 billion



TrueAllele

26

TrueAllele accuracy



27

Specificity

The extent to which interpretation does
not misidentify the wrong person

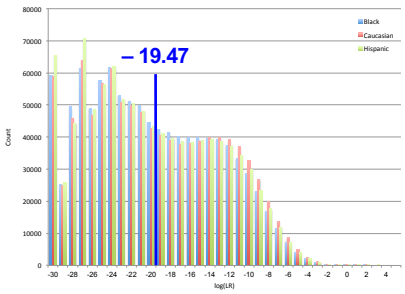
True exclusions, without false inclusions

101 matching genotypes x 10,000 random references
x 3 ethnic populations,
for over 1,000,000 nonmatching comparisons

28

TrueAllele specificity

log(LR) nonmatch distribution



29

False positives

in over 1,000,000 comparisons per group

Tail distribution	Black	Caucasian	Hispanic
0	39	32	29
1	8	11	9
2	2	1	1
3	0	0	1
log(LR) > 0	49	44	40

false positive rate is under 1 in 20,000 (0.005%)
for LR > 100, rate is 1 in 1,000,000 (0.0001)%

30

Higher human error rate

TrueAllele specificity (million samples)
From noncontributor distribution, for LR > 100:
Error rate = 1 in 1,000,000 (**0.0001**)%

CPI – analytical threshold
5 false positives in 81 comparisons
Error rate = 5 in 81 (**6%**)

mCPI – stochastic threshold
17 inconclusive results
1 false positive in 53 comparisons
Error rate = 1 in 53 (**2%**)

31

Reproducibility

The extent to which interpretation gives the same answer to the same question

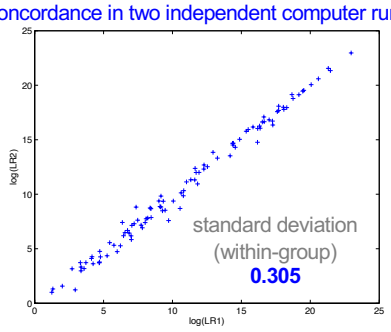
MCMC computing has sampling variation

duplicate computer runs
on 101 matching genotypes
measure log(LR) variation

32

TrueAllele reproducibility

Concordance in two independent computer runs



33

JOURNAL OF FORENSIC SCIENCES

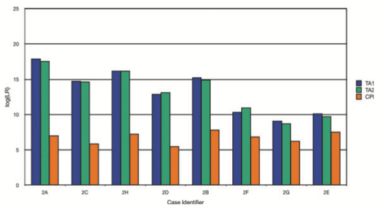
J Forensic Sci, November 2011, Vol. 56, No. 6
doi: 10.1111/j.1556-4029.2011.01859.x
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Mark W. Perlin,[1] M.D., Ph.D.; Matthew M. Legler,[1] B.S.; Cara E. Spencer,[1] M.S.; Jessica L. Smith,[1] M.S.; William P. Allan,[1] M.S.; Jamie L. Belrose,[2] M.S.; and Barry W. Duceman,[3] Ph.D.

Validating TrueAllele® DNA Mixture Interpretation*,†



34

JOURNAL OF FORENSIC SCIENCES

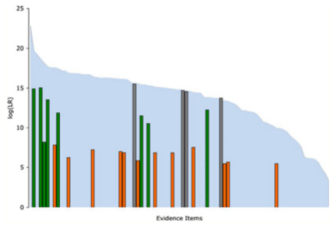
J Forensic Sci, November 2013, Vol. 58, No. 6
doi: 10.1111/1556-4029.12223
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Mark W. Perlin,[1] M.D., Ph.D.; Jamie L. Belrose,[2] M.S.; and Barry W. Duceman,[3] Ph.D.

New York State TrueAllele® Casework Validation Study*



35

JOURNAL OF FORENSIC SCIENCES

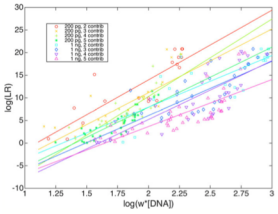
J Forensic Sci, July 2015, Vol. 60, No. 4
doi: 10.1111/1556-4029.12788
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

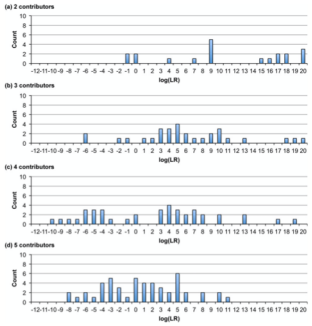
Mark W. Perlin,[1] Ph.D., M.D.; Jennifer M. Hornyak,[1] M.S.; Garett Sugimoto,[2] M.S.; and Kevin W.P. Miller,[2] Ph.D.

TrueAllele® Genotype Identification on DNA Mixtures Containing up to Five Unknown Contributors*



36

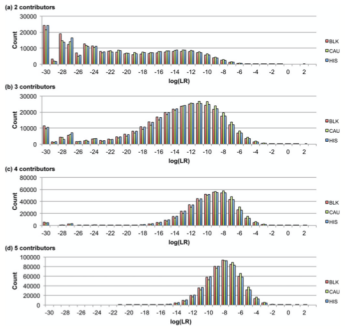
TrueAllele sensitivity



200 pg

37

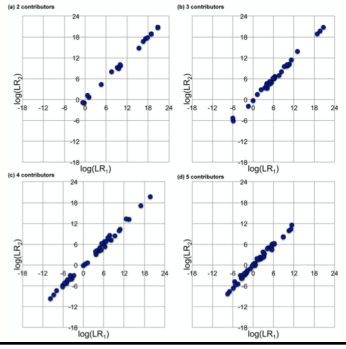
TrueAllele specificity



200 pg

38

TrueAllele reproducibility



200 pg

39

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, 2019
doi: 10.1111/1556-4029.14204
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

David W. Bauer,[1] Ph.D.; Nasir Butt,[2] Ph.D.; Jennifer M. Hornyak,[1] M.S.; and Mark W. Perlin,[1] Ph.D., M.D., Ph.D.

Validating TrueAllele® Interpretation of DNA Mixtures Containing up to Ten Unknown Contributors*

TABLE 4—*Independent analysis.*

	Operator Site	
	Cybergenetics	CCRFSL
Genotypes	78	78
Minimum	−5.16	−9.14
Mean	8.36	8.48
Median	5.98	5.61
Maximum	29.03	29.12
SD	8.37	8.54
σ_m	0.70	

TABLE 7—*Peeling sensitivity.*

	Peeling Round						
Mixture weight (%)	0	1	2	3	4	5	6
13	7	K	K	K	K	K	K
22	6	7	K	K	K	K	K
12	5	4	5	K	K	K	K
16	4	4	5	6	K	K	K
13	4	3	2	1	6	K	K
15	3	3	4	1	6	8	K
2	1	1	1	1	3	3	4
2	0	2	2	3	2	3	4
4	0	1	1	1	2	0	2
1	0	0	0	0	−1	−1	0

40

Heliyon

Received:
9 July 2018
Revised:
18 September 2018
Accepted:
24 September 2018

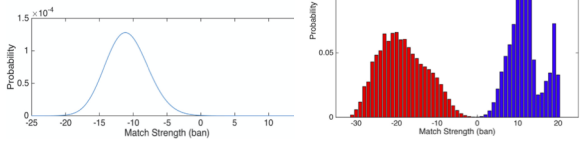
Cite as:
Mark W. Perlin. Efficient construction of match strength distributions for uncertain multi-locus genotypes. Heliyon 4 (2018) e00824. doi: 10.1016/j.heliyon.2018.e00824

Efficient construction of match strength distributions for uncertain multi-locus genotypes

Mark W. Perlin*

Cybergenetics, Pittsburgh, PA, USA

*Corresponding author.
E-mail address: perlin@cybgen.com (M.W. Perlin).



41

44 US admissibility rulings

Commonwealth of Pennsylvania v Kevin Foley (admitted, 2009; appellate precedent, 2012)
People of California v Dupree Langston (admitted, 2013)
Commonwealth of Virginia v Matthew Brady (admitted, 2013)
State of Ohio v Maurice Shaw (admitted, 2014)
State of Louisiana v Chatilley Chesterfield & Samuel Nicolas (admitted, 2014)
People of New York v John Wakefield (admitted, 2015; appellate precedent, 2019; high court precedent, 2022)
State of South Carolina v Jaquard Aiken (admitted, 2015)
Commonwealth of Massachusetts v Heidi Bartlett (admitted, 2016)
State of Indiana v Dugniqo Forest (admitted, 2016)
State of Indiana v Malcolm Wade (admitted, 2016)
State of Washington v Emanuel Fair (admitted, 2017)
State of Louisiana v Harold Houston (admitted, 2017)
State of Indiana v Randal Coalter (admitted, 2017)
State of Nebraska v Charles Simmer (admitted, 2018; appellate precedent, 2019)
State of Indiana v Vaylen Glazebrook (admitted, 2018)
State of Ohio v David Mathis (admitted, 2018)
State of Florida v Lajayvian Daniels (admitted, 2018; appellate precedent, 2021)
State of Tennessee v Demontez Watkins (admitted, 2018; appellate precedent, 2021)
State of Georgia v Thaddus Nundra (admitted, 2019; appellate precedent, 2023)
State of Georgia v Monte Baugh & Thaddeus Howell (admitted, 2019)
State of Louisiana v Kyle Russ (admitted, 2019)
People of New York v Casey Wilson (admitted, 2019)
State of Georgia v Alexander Battle (admitted, 2019)
United States v Lenard Gibbs (admitted, 2019)
State of Georgia v Guy Sewell (admitted, 2019)
State of Georgia v Adedoja Bah (admitted, 2019)
State of Georgia v Nathaniel Day (admitted, 2019)
State of Tennessee v Abdullah Powell (admitted, 2021)
State of Georgia v Zarren Garner (admitted, 2021)
United States v Curtis Johnson, Jr. (admitted, 2021)
State of Georgia v Rahul Joseph Das (admitted, 2021)
State of Maryland v Tyrone Harvin (admitted, 2021)
State of Maryland v Gregory Jones (not used, Daubert not applied, 2021)
State of Georgia v Lashumbia Session (admitted, 2021)
State of Georgia v Bryan Byers (admitted, 2022)
State of Louisiana v Dermell Lewis, Corey Major, & Gerald Parker (admitted, 2022)
State of Louisiana v James Tabb (admitted, 2022)
State of Louisiana v Shawn Briscoe and Lance McIntyre (not used due to timeliness, 2022)
United States v Hunter Anderson (admitted, 2023)
State of Louisiana v Corlious Dyson (admitted, 2023)
United States v Ravel Mills (admitted, 2023)
United States v Damond Lockett (admitted, 2023)
State of Georgia v Erin Stephon Arms (admitted, 2023)

42

TrueAllele today

Invented math & algorithms	30 years
Developed computer systems	25 years
Support users and workflow	10 laboratories
Routinely used in casework	450 agencies
Validate system reliability	43 studies
Educate the community	175 talks
Train or certify analysts	400 students
Admissibility challenges	42 rulings, 15 states and federal
Testify about LR results	145 trials
Educate lawyers and public	1,000 people
Make the ideas understandable	1,250 cases, 46 states

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.

43

Conclusions

TrueAllele Casework DNA mixture interpretation is:

A reliable method
- objective
- sensitive
- specific
- reproducible
- accurate

TrueAllele computer genotyping is more effective than human review