

Data-driven DNA forensic science: TrueAllele® computation

Computational Health Informatics
University of Cape Town, South Africa
August 2024

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



Cybergenetics © 2003-2024

STR stutter artifact

Am. J. Hum. Genet. 57:1199–1210, 1995

Toward Fully Automated Genotyping: Genotyping Microsatellite Markers by Deconvolution

Mark W. Perlin,¹ Giuseppe Lancia,² and See-Kiong Ng¹

¹ Computer Science Department and ² Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh

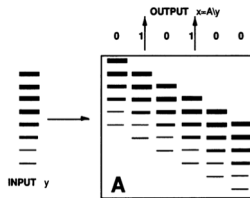


Figure 4 One deconvolution procedure. The stutter pattern matrix A is divided by the observed (input) data vector y to compute a "best" (output) genotype vector z that fits the data.

DNA database backlog

Year 2000. Britain. The FSS.

The first national DNA database.

Human review of DNA data fails.

Backlog of 350,000 swabs.



Enter **computer automation**.

Cybergenetics **TrueAllele** clears the backlog.

Every year, does another 350,000 samples.

Accurate: eliminates human error (1 in 2,000).

Fast: "swab to DB" turnaround time in 5 days.

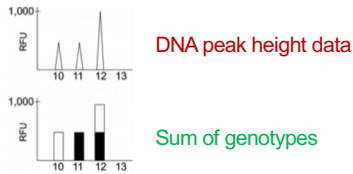
Labor-saving: from 100 people down to 6.

Inexpensive: 2-3 computers, not 100 people.

Solving DNA mixtures

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

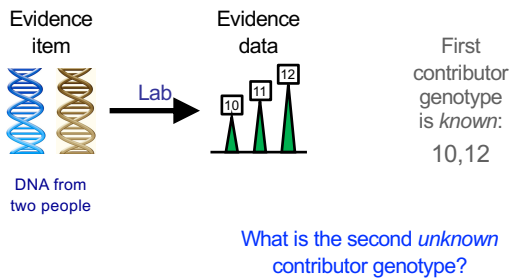
Linear Mixture Analysis: A Mathematical Approach to Resolving Mixed DNA Samples



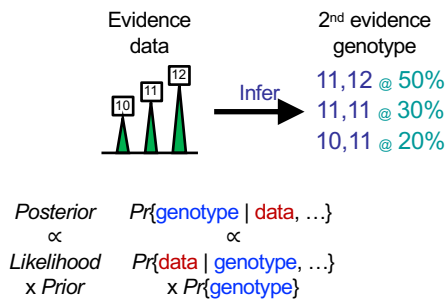
DNA peak height data

Sum of genotypes

DNA mixture data

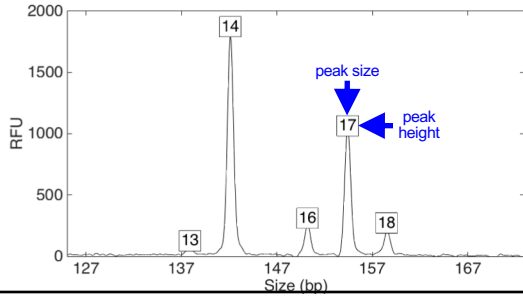


Bayesian modeling & computation



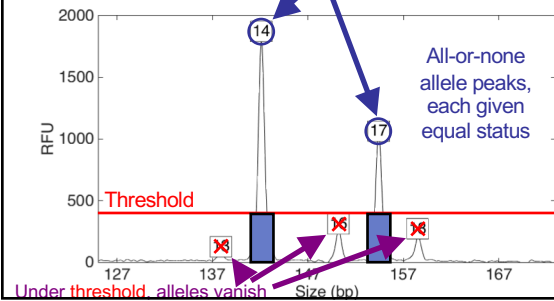
Computers can use all the data

Quantitative peak heights at locus vWA



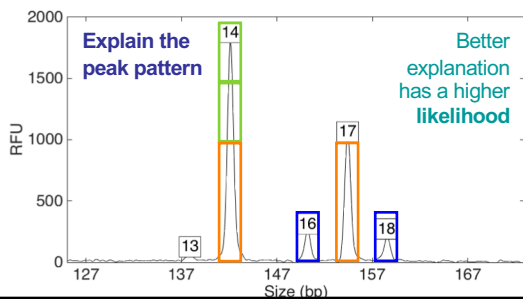
People may use less of the data

Over threshold, peaks are labeled as allele events



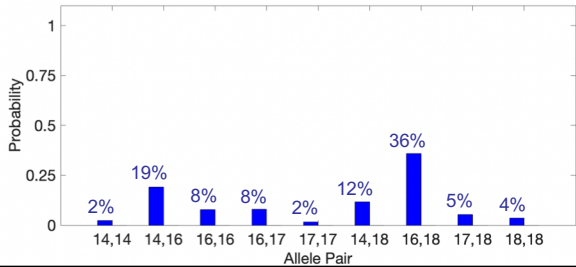
How the computer thinks

Consider every possible genotype solution



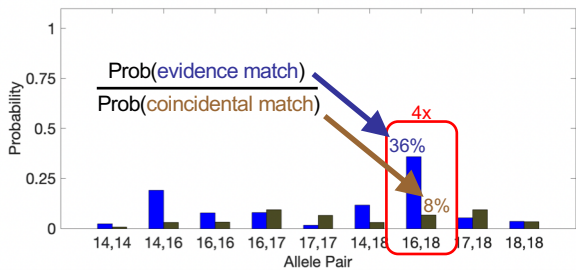
Unmixed contributor genotype

Objective genotype determined solely from the DNA data.
Evidence solution doesn't know comparison reference.

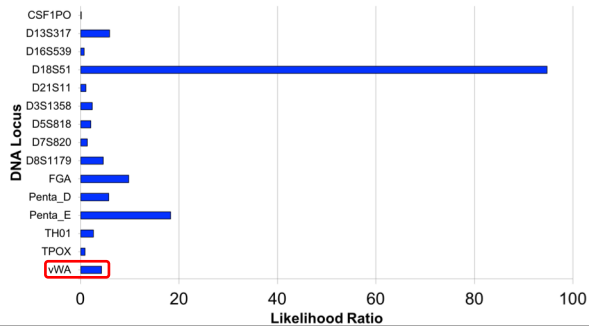


DNA match information

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



Match information at 15 loci



Is the suspect in the evidence?

A match between the **item**
and the **defendant** is:

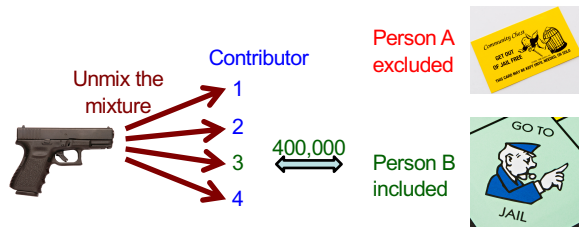
971 million times more probable than
a coincidental match to an unrelated **African-American** person

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

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

Informative genotyping

(Inconclusive uninterpretable DNA mixture)
Cybergentics TrueAllele analysis
Match statistics provide information



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

World's first modern DNA database case – TrueAllele

World Trade Center

18,000
victim remains

2,700
missing people



match



September 11, 2001 – New York City

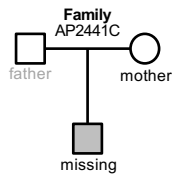
Komatipoort Bus Crash

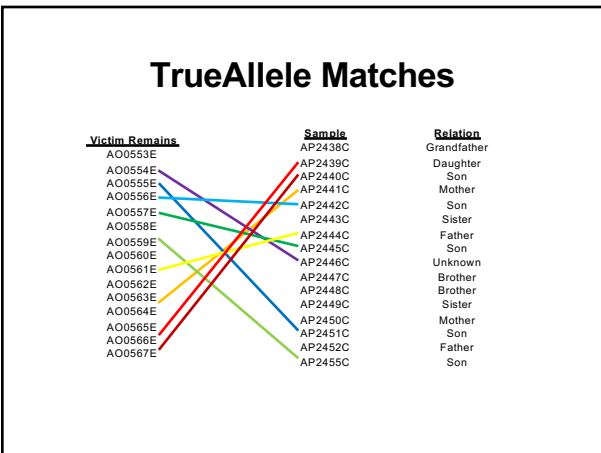


- 2008 bus crash in Komatipoort, South Africa
- Police recover burned victim remains
- Relatives submit DNA to help identify remains
- SAPS lab unable to identify victim remains

Genotype Matching

Victim Remains
AO0563E





Student Exercise

15 victim genotypes x 16 kinship genotypes
= 240 genotype match comparisons

240 comparisons x 20 students
= 4,800 total match comparisons

before lunch:	students upload genotypes
during lunch:	TrueAllele solves all problems
after lunch:	students review identification results


TrueAllele automation: fast workflow

World's first modern DNA analysis case – TrueAllele

Pennsylvania v. Kevin Foley

Apr 2006: Blairsville Dentist John Yelenic murdered

Nov 2007: Trooper Kevin Foley charged with crime





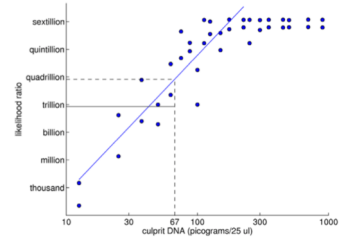
February 2008: Defense questions 13,000 DNA match score

March 2009: Jury hears 189,000,000,000 TrueAllele statistic

TrueAllele predictability

An Information Gap in DNA Evidence Interpretation

Mark W. Perlin^{1*}, Alexander Sinehnikov²  PLoS one 2009



Peer-reviewed validation studies

Perlin MW, Sinehnikov 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.

TrueAllele reliability

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

 PLoS one 2014

Validation axes

- objective
- sensitive
- specific
- reproducible
- accurate

Sensitivity

The extent to which interpretation identifies the correct person

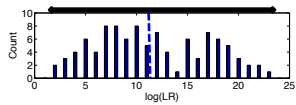
True DNA mixture inclusions

101 reported genotype matches
82 with DNA statistic over a million

TrueAllele sensitivity

log(LR) match distribution

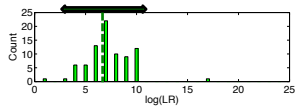
11.05 (5.42)
113 billion



TrueAllele

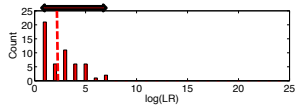
Comparison with human review

6.83 (2.22)
6.68 million



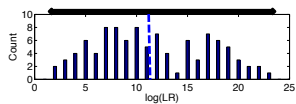
CPI

2.15 (1.68)
140



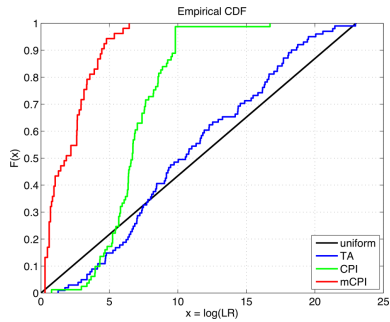
mCPI

11.05 (5.42)
113 billion



TrueAllele

TrueAllele accuracy



Specificity

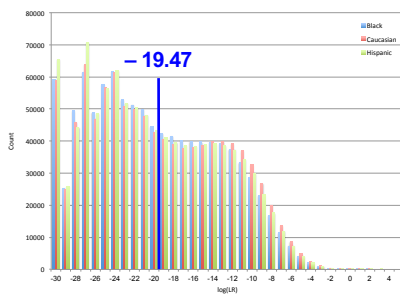
The extent to which interpretation does not misidentify the wrong person

True exclusions, without false inclusions

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

TrueAllele specificity

log(LR) nonmatch distribution



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%**)

Reproducibility

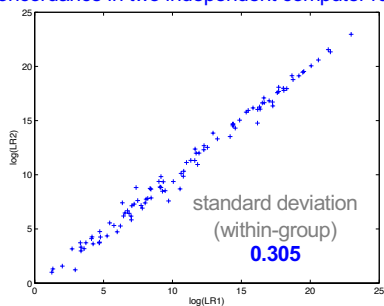
The extent to which interpretation gives
the same answer to the same question

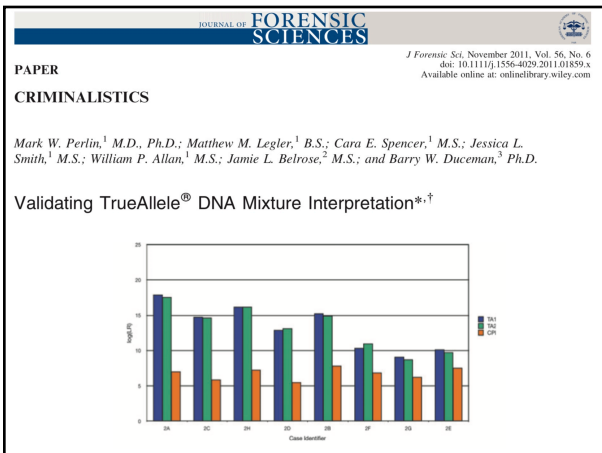
MCMC computing has sampling variation

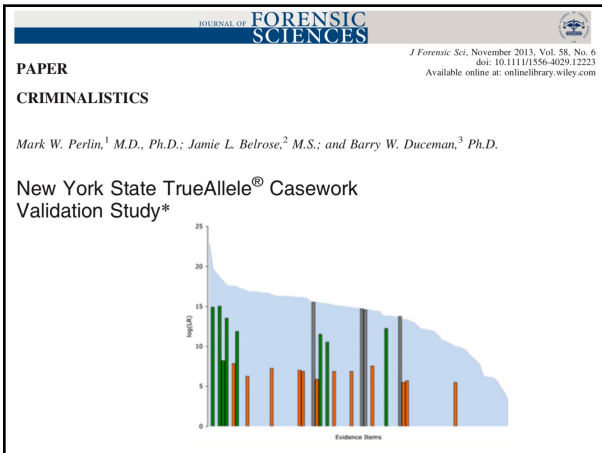
duplicate computer runs
on 101 matching genotypes
measure log(LR) variation

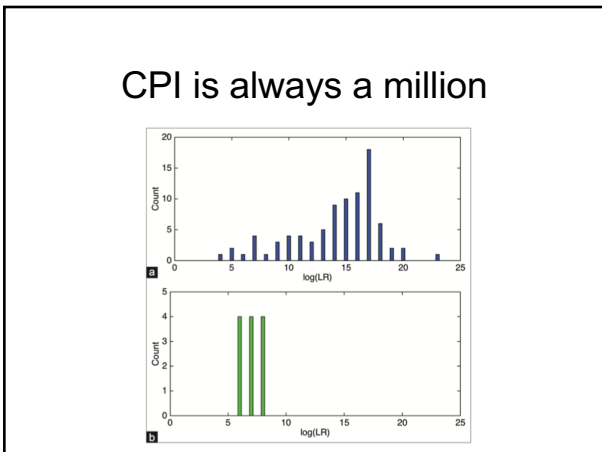
TrueAllele reproducibility

Concordance in two independent computer runs









Statistics lack scientific basis

Misled courts for 15 years on countless DNA mixtures

J Pathol Inform

Editor in Chief: **André A. Pomeroy**, Pittsburgh, PA, USA
Editor: **Liisa Parkkinen**, Pittsburgh, PA, USA
For entire Editorial Board visit: www.jpathinformatics.org/fortheboard.asp

OPEN ACCESS
HTML format

Research Article

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

Mark William Perlin¹

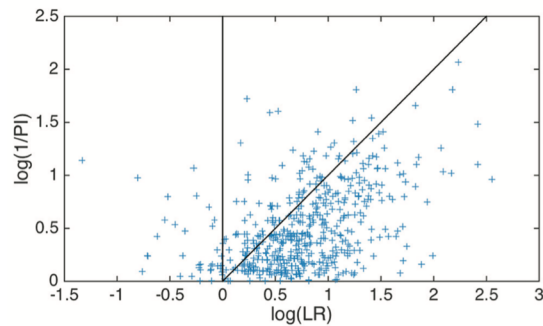
¹Cybergenetics, Pittsburgh, USA

E-mail: Mark William Perlin - perlin@cygn.com

*Corresponding author

Received 14 July 2013 Accepted 21 September 2013 Published 28 October 2013

Uncorrelated with information



Inclusion just counts tests

- 1 One-sided Match Statistic
- 2 Truncated Normal Distribution
- 3 Positive Tail Centered at Zero
- 4 Uncorrelated with Identification Information
- 5 Inclusion Distribution Has a Positive Mean
- 6 Law of Large Numbers
- 7 Why CPI is Always a Million

A subjective one-sided match statistic
unrelated to identification information

Wolfe sisters homicide



On February 6, 2014, Susan Wolfe (44) and her younger sister Sarah (38, left) were killed in their East Liberty home in Pittsburgh.

Pennsylvania v. Allen Wade

Thresholds failed to interpret most DNA mixtures

Hat	No conclusions
Cup	Insufficient data
Fingernails	Contamination, insufficient data
Gear shift	Insufficient data
Seat lever	Cannot be excluded
Knit hat	Insufficient data
Sock	Too complex, no conclusions

Pennsylvania v. Allen Wade

The crime lab reported 5 DNA mixture matches
TrueAllele found 17 matches on the same data

Hat	65.3 thousand	Allen Wade
Cup	20.5 thousand	Susan Wolfe
Fingernails	6.06 trillion	Allen Wade
Gear shift	9.37 million	Sarah Wolfe
Seat lever	385 billion	Sarah Wolfe
Knit hat	25.7 thousand	Allen Wade
Sock	300	Sarah Wolfe

Allen Wade Found Guilty On All Counts In East Liberty Sisters' Slaying

CBS News, May 23, 2016

PITTSBURGH (KDKA/AP)

- A man accused of killing two sisters who lived next door to him in East Liberty has been found guilty on all counts.
- Allen Wade was accused of shooting Sarah and Susan Wolfe after they returned from work on Feb. 6, 2014, apparently to steal a bank card.
- On Monday morning, a jury found Wade guilty of first-degree murder, robbery, burglary and theft by unlawful taking.

Pennsylvania v. Allen Wade

Thresholds failed to interpret DNA mixture
TrueAllele succeeded on the same data

A hat left from a burglary of the Wolfe sister's home
six weeks before the murder matched
Allen Wade with a 65.3 thousand statistic



Preventable Crime

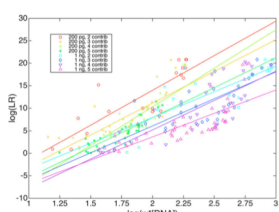
JOURNAL OF FORENSIC SCIENCES

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.; Gareth Sugimoto,² M.S.; and Kevin W.P. Miller,² Ph.D.

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



JOURNAL OF FORENSIC SCIENCES

J Forensic Sci. 2019
doi: 10.1111/1556-4029.14294
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 7—Peeling sensitivity.

TABLE 4—Independent analysis.			Mixture weight (%)	Peeling Round							
				0	1	2	3	4	5	6	
	Operator Site		13	7	K	K	K	K	K	K	K
	Cybergentics	CCRFSL	12	6	7	K	K	K	K	K	K
Genotypes	78	78	16	5	4	5	K	K	K	K	K
Minimum	-5.16	-9.14	13	4	3	2	6	K	K	K	K
Mean	8.36	8.48	15	3	3	4	1	6	8	K	K
Median	5.98	5.61	2	1	1	1	3	3	4		
Maximum	29.03	29.12	2	0	2	2	3	2	3	4	
SD	8.37	8.54	4	0	1	1	1	2	0	2	
σ_c		0.70	1	0	0	0	0	-1	-1	0	

Indiana v. Darryl Pinkins



Wrongfully convicted
Imprisoned for 25 years
Released from prison
April 25, 2016

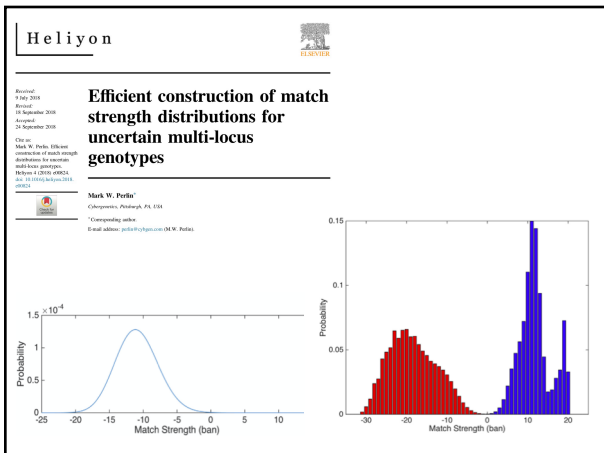
CBS News 48 Hours
"Guilty Until Proven Innocent"

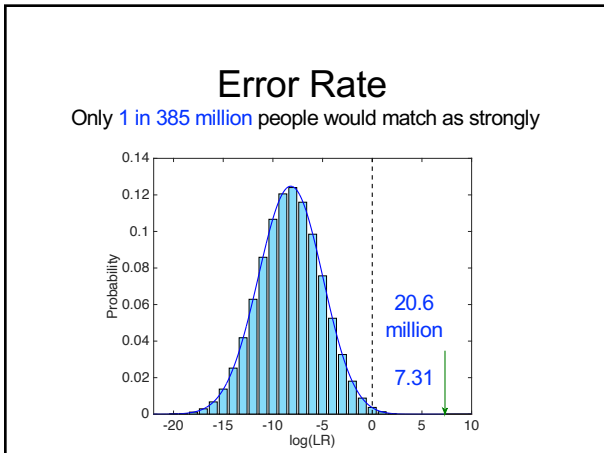
TrueAllele Pinkins findings

1. compared *evidence with evidence*
2. calculated *exclusionary match statistics*
3. revealed *5% minor mixture contributor*
4. *jointly analyzed DNA mixture data*
5. showed three perpetrators were *brothers*

found 5 unidentified genotypes,
defendants not linked to the crime

Search CODIS?





New York v. Nick Hillary

Garrett Phillips (12)
Died from strangulation
October 24, 2011

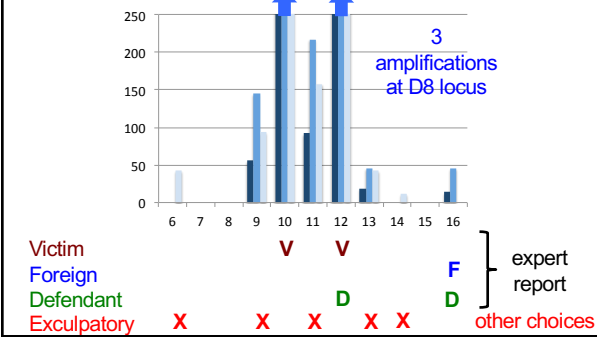
Oral "Nick" Hillary
Arrested for murder
May 15, 2013

TrueAllele findings in Hillary

2013. 26 Identifier tests on left fingernails
 Mixture of 95% victim + 5% other
 No statistical connection to Hillary
 Advised Minifiler for degraded DNA

2014. More lab data on left fingernails
 9 tests using new kit & machine
 NYSP requested TrueAllele analysis
 Again, no connection to Hillary

Other PG software forces users to choose their data



Thresholds change answers

RFU	Data choices	All stutters
80	0	9
70	30	51
60	250	1,660
50	15,500,000	69,200 include
40	0	0 exclude
30	0	0

Different choices, different answers
 Software does not agree with itself

Judge's admissibility ruling

The Expert conceded at the hearing that **no internal validation studies were performed** by the State crime lab for the use of the Software on casework samples developed at the lab.

As a result the Expert was forced to **pick and choose data** from different "reliable sources" and **input parameters** into the program in such a way that he believed the system would tolerate.

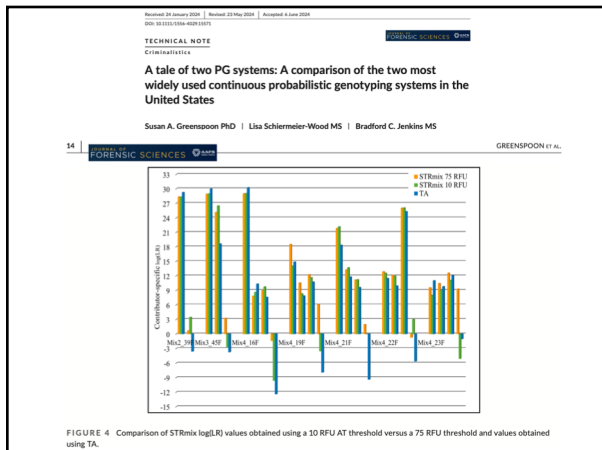
ORDERED that the **defendant's motion to preclude** the prosecution from calling an expert witness to testify on their direct case regarding any conclusion reached by the use of the Software is **granted** as the prosecution cannot lay a foundation for the introduction of evidence that had not been internally validated.

Nick Hillary acquitted

The New York Times

Oral Nicholas Hillary Acquitted in Potsdam Boy's Killing





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 Chantley 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 Barlett (admitted, 2016)
State of Indiana v Dagniqo 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 Cozler (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 Lajayvan Daniels (admitted, 2018; appellate precedent, 2021)
State of Tennessee v Demontez Watkins (admitted, 2018; appellate precedent, 2021)
State of Georgia v Thaddeus Nundie (admitted, 2019; appellate precedent, 2023)
State of Georgia v Monte Baugh & Thaddeus Howell (admitted, 2019)
State of Louisiana v Kyle Ross (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 Adedoga Bah (admitted, 2019)
State of Georgia v Nathaniel Day (admitted, 2019)
State of Tennessee v Abdulrah 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 Dernel 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 Corious Dyson (admitted, 2023)
United States v Travel Mills (admitted, 2023)
United States v Diamond Lockett (admitted, 2023)
State of Georgia v Erin Stephon Arms (admitted, 2023)

TrueAllele today

Invented math & algorithms	30 years
Developed computer systems	25 years
Support users and workflow	10 laboratories; 100,000 items
Routinely used in casework	450 agencies
Validate system reliability	43 studies
Educate the community	175 talks
Train or certify analysts	400 students
Admissibility challenges	44 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 (of 50) states

TrueAllele can help Africa



- Informative**
- Accurate
 - Unbiased
- Automated**
- Few analysts
 - Large population
- Effective**
- Solve crime
 - Prevent rape