

Mining the Mixture: A DNA Analyst Explains (2)

New York State
Judicial Summer Seminars
The New York State Judicial Institute
June, 2019
Rye Brook, NY

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



Cybergenetics

Cybergenetics © 2003-2019



JUSTICE
THROUGH
SCIENCE

TrueAllele report

Item 045A (shirt collar, outside rear)

TrueAllele assumed that the evidence sample data (Item 045A) contained two contributors, and objectively inferred evidence genotypes solely from these data. The victim genotype was assumed in some calculations. Following genotype inference, the computer then compared a genotype from this evidence item to provided reference (Items 052B, 188, and 189) genotypes, relative to reference populations, to compute LR DNA match statistics. Based on these results:

A match between the outside rear of the shirt collar (Item 045A) and Brett Wentworth (Item 052B) is:
1.52 billion times more probable than a coincidental match to an unrelated Black person,
83.8 million times more probable than a coincidental match to an unrelated Caucasian person, and
249 million times more probable than a coincidental match to an unrelated Hispanic person.

A match between the outside rear of the shirt collar (Item 045A) and John Wakefield (Items 188 and 189) is:
170 quintillion times more probable than a coincidental match to an unrelated Black person,
7.54 quintillion times more probable than a coincidental match to an unrelated Caucasian person, and
119 quintillion times more probable than a coincidental match to an unrelated Hispanic person.

A match between the evidence and the defendant
is a quintillion times more probable than coincidence

The probability of the evidence given the prosecution's hypothesis
is a quintillion times more than given the defense hypothesis

Case disclosure packet

Case
Notes

Data
Table
EPG

Case Packet
People v. John Wakefield
September 17, 2014

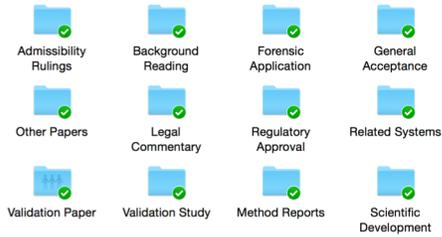
Request
Listing
Timing

Genotype
Evidence
Reference
Population

Match
Table
Locus Table
Non-contributor Analysis



Disclosure materials



Computer Interpretation of Quantitative DNA Evidence

People v. John Wakefield
October, 2014
Schenectady, NY

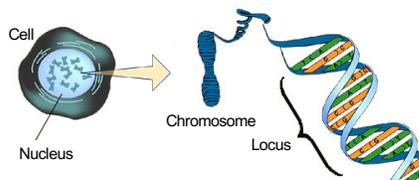
Mark W Perlin, PhD, MD, PhD
Cybergenetics, Pittsburgh, PA



Cybergenetics

Cybergenetics © 2003-2014

DNA biology



Short tandem repeat

DNA locus paragraph

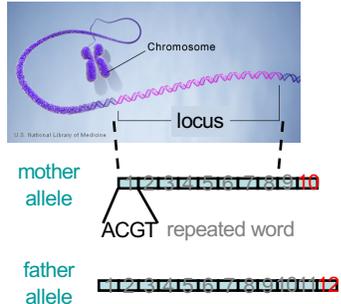


23 volumes in cell's DNA encyclopedia

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
 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

DNA genotype



Chromosome

locus

mother allele

father allele

ACGT repeated word

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

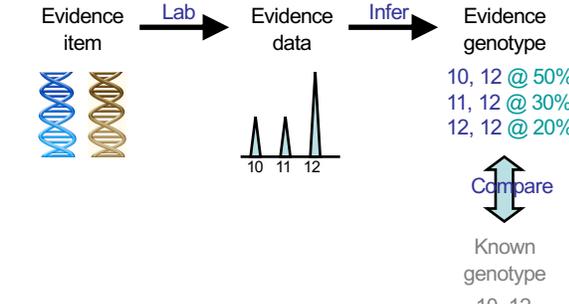
An allele is the number of repeated words.

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 evidence interpretation

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



10, 12 @ 50%

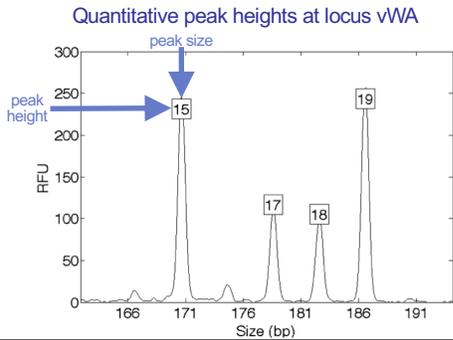
11, 12 @ 30%

12, 12 @ 20%

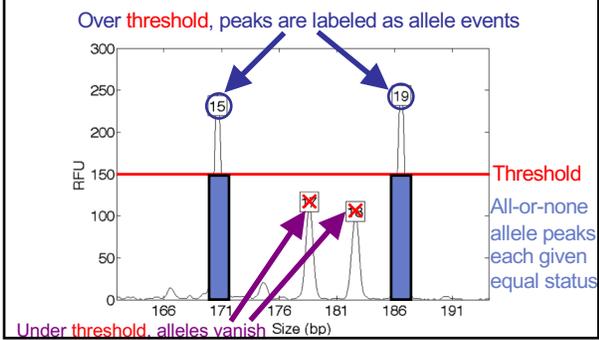
Compare

Known genotype
10, 12

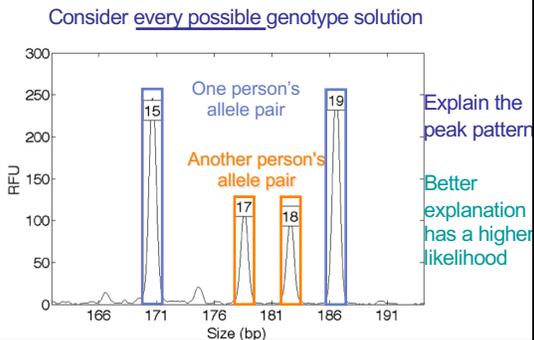
Computers can use all the data

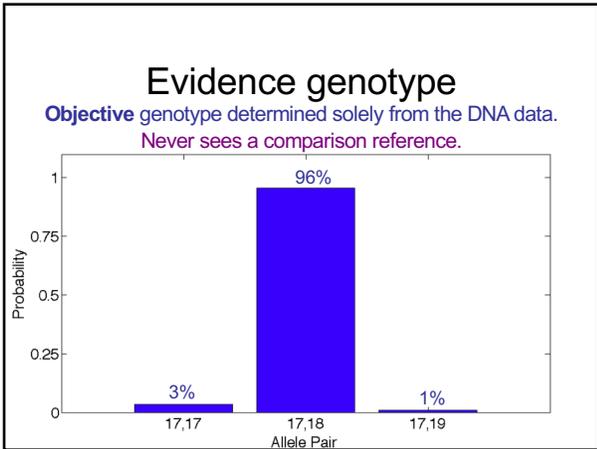


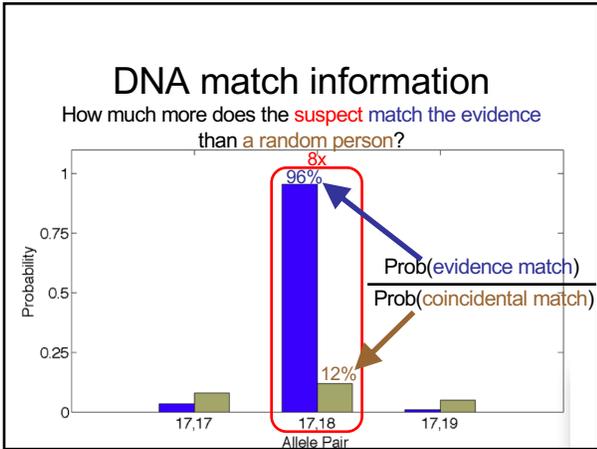
People may use less of the data

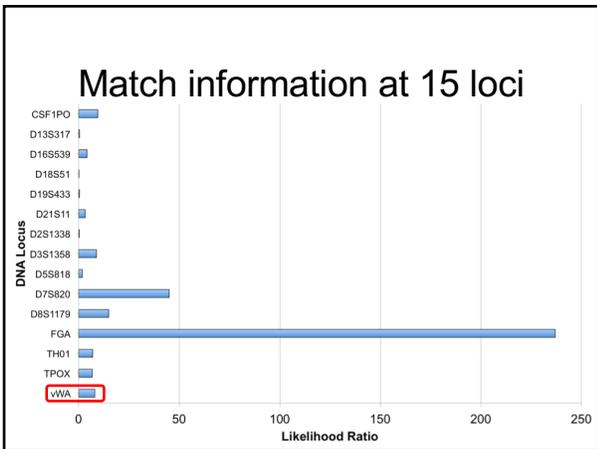


How the computer thinks









Is the suspect in the evidence?

A match between the 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

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

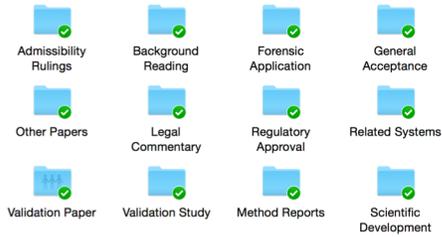
478 quadrillion times more probable than
a coincidental match to an unrelated Caucasian person

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

Match statistics

Item	Description	052B	188, 189
		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

Frye hearing



Validated genotyping method

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.

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.

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

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

Reproducibility

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

MCMC computing has sampling variation

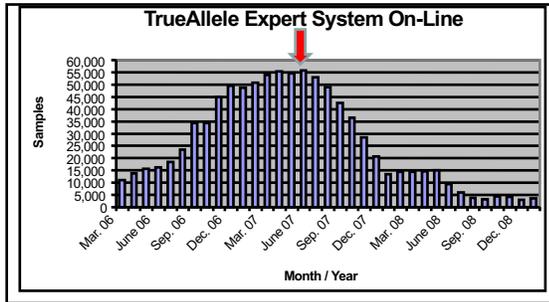
duplicate computer runs
on 101 matching genotypes
measure log(LR) variation

General acceptance

Invented math & algorithms	20 years
Developed computer systems	15 years
Support users and workflow	10 laboratories
Routinely used in casework	8 crime labs
Validate system reliability	37 studies
Educate the community	100 talks
Train & certify analysts	200 students
Go to court for admissibility	25 rulings
Testify about LR results	85 trials
Educate lawyers and laymen	1,000 people
Make the ideas understandable	700 cases, 43 states



Eliminated NYS DNA backlog





Reanalyzed WTC DNA data

18,000
victim remains

2,700
missing people



match



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: onlineibrary.wiley.com

PAPER

CRIMINALISTICS

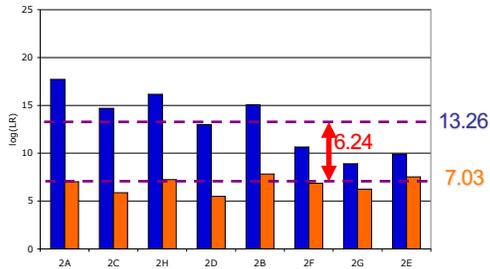
Mark W. Perlin,¹ M.D., Ph.D.; Matthew M. Legler,¹ B.S.; Cara E. Spencer,¹ M.S.; Jessica L. Smith,¹ M.S.; William P. Allan,¹ M.S.; Jamie L. Betrose,² M.S.; and Barry W. Duceman,³ Ph.D.

Validating TrueAllele[®] DNA Mixture Interpretation*[†]

ABSTRACT: DNA mixtures with two or more contributors are a prevalent form of biological evidence. Mixture interpretation is complicated by the possibility of different genotype combinations that can explain the short tandem repeat (STR) data. Current human review simplifies this interpretation by applying thresholds to qualitatively treat STR data peaks as all-or-none events and assigning allele pairs equal likelihood. Computer review, however, can work instead with all the quantitative data to preserve more identification information. The present study examined the extent to which quantitative computer interpretation could elicit more identification information than human review from the same adjudicated two-person mixture data. The base 10 logarithm of a DNA match statistic is a standard information measure that permits such a comparison. On eight mixtures having two unknown contributors, we found that quantitative computer interpretation gave an average information increase of 6.24 log units (min = 2.32, max = 10.09) over qualitative human review. On eight other mixtures with a known victim reference and one unknown contributor, quantitative interpretation averaged a 4.67 log factor increase (min = 1.01, max = 11.31) over qualitative review. This study provides a general treatment of DNA interpretation methods (including mixtures) that encompasses both quantitative and qualitative review. Validation methods are introduced that can assess the efficacy and reproducibility of any DNA interpretation method. An in-depth case example highlights 10 reasons (at 10 different loci) why quantitative probability modeling preserves more identification information than qualitative threshold methods. The results validate TrueAllele[®] DNA mixture interpretation and establish a significant information improvement over human review.



Preserves more match information



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.

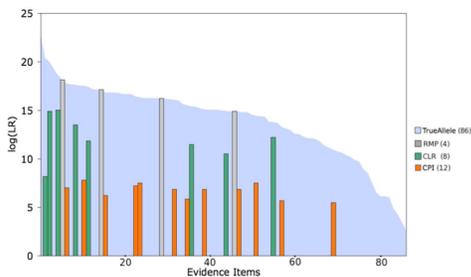
New York State TrueAllele® Casework Validation Study*

ABSTRACT: DNA evidence can pose interpretation challenges, particularly with low-level or mixed samples. It would be desirable to make full use of the quantitative data, consider every genotype possibility, and objectively produce accurate and reproducible DNA match results. Probabilistic genotype computing is designed to achieve these goals. This validation study assessed TrueAllele® probabilistic computer interpretation on 368 evidence items in 41 test cases and compared the results with human review of the same data. Whenever there was a human result, the computer's genotype was concordant. Further, the computer produced a match statistic on 81 mixture items for 87 inferred matching genotypes in the test cases, while human review reported a statistic on 25 of these items (30.9%). Using match statistics to quantify information, probabilistic genotyping was shown to be sensitive, specific, and reproducible. These results demonstrate that objective probabilistic genotyping of biological evidence can reliably preserve DNA identification information.

KEYWORDS: forensic science, DNA evidence, mixture interpretation, probabilistic genotype, likelihood ratio, developmental validation, identification information

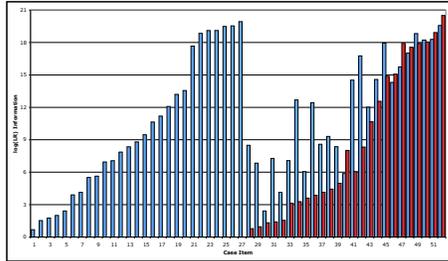


Lots more match information





That other methods discard



25
+
27



STATE OF NEW YORK DNA SUBCOMMITTEE OF THE COMMISSION ON FORENSIC SCIENCE

Approved

May 20, 2011

John Mulvaney, Esq., Chairman
 State of New York
 Department of Criminal Justice Services
 4 Lower Place, 10th Floor
 Albany, New York 12242-3764

Sean Byrne, Acting Commissioner
 Chair, Commission on Forensic Science
 Division of Criminal Justice Services
 4 Lower Place, 10th Floor
 Albany, New York 12242-3764

Dear Commissioner Byrne:
 Pursuant to Executive Law 1995-b (138) the DNA Subcommittee will assess and evaluate all DNA methodologies proposed to be used for forensic analysis and make recommendations to the Commission.

At the May 20, 2011 meeting the DNA Subcommittee reviewed and evaluated the New York State Police TrueIdent Cytingling developed by New York State Police and Cybergenetics Corp. The DNA Subcommittee offers a positive recommendation to the Commission on Forensic Science that its use by NYSP Forensic Investigation Center be approved for forensic use.

Very truly yours,

Jack Balthazart, Ph.D.
 Chair, NYSP DNA Subcommittee

cc: Gina L. Bianchi, Esq., Deputy Commissioner & Counsel, DCJS
 NYS DNA Subcommittee Members
 NYS Commission on Forensic Science Members

RECEIVED
 FEB 11 2015
 Schenectady County Court

At a Term of the Supreme Court of the State of New York held for the County of Schenectady, New York at the City of Schenectady, New York on the 30th day of January, 2015

PRESENT: HON. MICHAEL V. COCCOMA
 SUPREME COURT JUSTICE

STATE OF NEW YORK
 SUPREME COURT: COUNTY OF SCHENECTADY

THE PEOPLE OF THE STATE OF NEW YORK
 -against-

JOHN WAKEFIELD

Defendant

DECISION AND ORDER

Indictment No. A-812-29

FILED
 FEB 11 2015
 CLERK OF SUPREME COURT

STATE OF NEW YORK
COUNTY COURT : COUNTY OF CHEMUNG

THE PEOPLE OF THE STATE OF NEW YORK,
Plaintiffs,
-against-
CASEY WILSON,
Defendant.

**DECISION and ORDER
INDICTMENT NO. 2013-331**

Pursuant to an August 9, 2018 Memorandum and Order from the Appellate Division, this Court held a posttrial Frye hearing on March 26, 2019 to consider the reliability of the TrueAllele Casework system as it was when DNA testing was performed on a pair of lavender gloves in 2013.

New York TrueAllele trials

Case	Year	County	DNA Expert	Offense	Called by
New York v Casey Wilson	2014	Chemung	Dr. Mark Perlin	rape	prosecution
New York v Lewis Swift	2015	Onondaga	Dr. Mark Perlin	homicide	prosecution
New York v John Wakefield	2015	Schenectady	Dr. Mark Perlin	homicide	prosecution
New York v Adam Mogan	2015	Monroe	Dr. Mark Perlin	sexual assault	prosecution
New York v Nahktem Fields	2015	Schenectady	Dr. Mark Perlin	weapon	prosecution
New York v Sylvester Young	2015	Schenectady	Dr. Mark Perlin	homicide	prosecution
New York v Frank Thomas	2016	Onondaga	Dr. Mark Perlin	weapon	prosecution
New York v Noah Hunter	2017	Schenectady	Jennifer Hornyak	weapons possession	prosecution
New York v Akeem Williams	2017	Chemung	Jennifer Hornyak	homicide	prosecution
New York v Rene Hernandez	2018	Suffolk	Beatriz Pujols	sexual abuse	prosecution
New York v Dequan Grimes	2018	Suffolk	Beatriz Pujols	gun possession	prosecution
New York v Kimani Stephenson	2018	New York	Jennifer Hornyak	sexual assault	defense
New York v Christopher James	2018	Suffolk	Beatriz Pujols	homicide	prosecution
New York v William Hubbard	2019	Suffolk	Beatriz Pujols	homicide	prosecution

Computer Interpretation of Quantitative DNA Evidence

People of New York v. Kimani Stephenson
June, 2018
New York, NY

Jennifer M. Hornyak, MS
Cybergenetics, Pittsburgh, PA



Cybergenetics

Cybergenetics © 2003-2018

When nothing means everything

Item	Description	BC	KS
		BC	Kimani Stephenson
1, dresLB	left breast area of dress	25.8 quadrillion	one in 418 octillion
1, dress2	left front area of dress, sample 2	14.2 billion	one in 74.7 septillion
2, jackLB	left breast area of jacket	1.91 sextillion	one in 220 thousand

Computer Interpretation of Quantitative DNA Evidence

State of Georgia v. Johnny Lee Gates
May, 2018
Columbus, GA

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



Cybergenetics

Cybergenetics © 2003-2018

Match statistics

Item	Description	76C2573-004
		Johnny Lee Gates
76C2573-032	robe belt side 1 swab	one in 1.5 million
76C2573-033	robe belt side 2 swab	one in 134 thousand
76C2573-034	front of black tie swab	one in 4.33 million
76C2573-035	back of black tie swab	one in 963 million
76C2573-042	robe belt M-vac filter	one in 902 trillion
76C2573-044	black tie M-vac filter	one in 825 billion

Example testimony

DEF EXPERT: The 8,11 at the CSF locus for this particular analysis was the **fourth most probable genotype** reported.

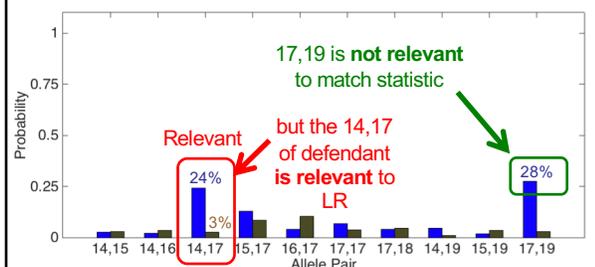
DEFENDER: Explain what you're saying to us.

DEF EXPERT: There are three genotypes other than 8,11 that have been accorded a higher probability.

DEFENDER: Okay. And D13? We're just going to go down through them.

DEF EXPERT: It was the **second highest**, this one listed in the table, is the second most probable.

Fallacy explained



How to respond

PROSECUTOR: I'm going to object to the **relevance** of this unless they can bring some sort of expert opinion to bear on it, what's the significance.

THE COURT: So I would sustain that objection. So I would disallow your ability to get into that because it's outside the scope of the expert report.

Verbal equivalents?

LR for (H ₁) and 1/LR for (H ₂)	Qualitative Equivalent
1	Uninformative
2 to <100	Limited Support
100 to <10,000	Moderate Support
10,000 to <1,000,000	Strong Support
≥1,000,000	Very Strong Support

- hides the real match strength information
- not what a DNA expert actually believes
- misleads the jury about "million" (Koehler)

Just report LR error, along with the LR,
when the match strength is under a million

Allegheny County: 75 cases, 15 trials, 8 exonerations

Crime	Evidence	Defendant	Outcome	Sentence
rape	clothing	Ralph Skundnich	guilty	75 years
murder	gun, hat	Leland Davis	guilty	23 years
rape	clothing	Akaninyene Akan	guilty	32 years
murder	shotgun shells	James Yeckel, Jr.	guilty plea	25 years
murder	fingernail	Anthony Morgan	guilty	life
weapons	gun	Thomas Doswell	guilty plea	1 year
bank robbery	clothing	Jesse Lumberger	guilty	10 years
drugs	gun	Derek McKissick	guilty plea	2 1/2 years
drugs	gun	Steve Morgan	guilty plea	2 1/2 years
murder	door, clothing	Calvin Kane	guilty plea	20 years
murder	fingernail, clothing	Allen Wade	guilty	life
murder	gun	Jaykwaan Pinckney	guilty plea	10 years
child rape	clothing	Dhaque Jones	guilty plea	6 years
shooting	gun	Anthony Jefferson	guilty plea	4 years
weapons	gun	Zachary Blair	guilty plea	15 years
weapons	gun	Delmingo Williams	guilty plea	3 years
incest rape	clothing	Terry L.	guilty	40 years
bank robbery	hat	Robert Schatzman	guilty	1 1/2 years
weapons	gun	Rashawn Walker	guilty	1.5 years
robbery	hat	Lauren Peak	guilty plea	1 year
rape	body cavity	Freddie Cole	guilty plea	2 years
murder	gun	Chaz White	guilty plea	4 years
sex crime	clothing	Christopher Stavish	guilty plea	2 1/2 years
murder	gun	Jake Knight	guilty	life

Open the past

We can overcome the past failures
of DNA mixture interpretation.

There have been 100,000's of mixtures
wrongly reported as "inconclusive"
or given inaccurate match statistics.

TrueAllele automation can open the past.
It can use all the data, without people,
to accurately reprocess old cases.

To free the innocent, and find the guilty.

More information

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



- Courses
- Newsletters
- Newsroom
- Presentations
- Publications
- Webinars

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



Cybergenetics



perlin@cybgen.com



JUSTICE
THROUGH
SCIENCE
