Innovation and Transparency for Reliable Forensic Software

The New Abnormal:
Data Protection in a More Virtual World

Santa Clara High Tech Law Journal
Santa Clara University School of Law
March, 2022
Santa Clara, CA

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

Cybergenetics

JUSTICE THROUGH SCIENCE

California v. Martell Chubbs

In December 1977, a 17-year-old Long Beach mother was found raped and murdered in her home, toddler at her feet.

In June 2011, a DNA lab tested the vaginal swabs, and found a three-person mixture – victim, elimination & unknown.

In September 2012, the major sperm-fraction profile was found to match Chubbs with a CPI statistic of 1 in 10,000.

In September 2013, Pittsburgh-based Cybergenetics found a TrueAllele® LR statistic to Chubbs of 1.62 quintillion.

Chubbs approach to DNA

Hard to attack TrueAllele reliability. Extensively validated probabilistic genotyping software. Kelly-Frye admitted.

How to suppress the incriminatory DNA evidence?

Demand irrelevant trade secrets that a company can't release.

The **"source code" gambit.** Hire an expert. Falsely claim that source code is needed to assess software reliability.

What is source code?

Algorithm
A procedure for solving a problem

Patent & publish
(*Alice v. CLS*)

Programmer

Source code

Trade secret

```
computer instructions written in a
human-readable computer language
```

Compiler

Input → Software application → *Output*

Release

What users have and test

Scientific reliability

Science is based on **empirical testing** of hypotheses.

There are **national standards** and guidelines for testing probabilistic genotyping software (PGS).

Computer source code is **entirely irrelevant** to scientific testing, national standards, or admissibility standards.

Cybergenetics makes its TrueAllele software application program **available for** (free) **testing** by defense experts.

Transparency

42 validations, 8 peer-reviewed

6

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):e92837.

Test other PGS on Chubbs data

Software	log(LR)	LR
mCPI	4.00	Thousand
CPI	9.68	Billion
LRmix	9.73	Billion
LabRetriever	12.21	Trillion
LikeLTD (v4)	14.86	Trillion
STRmix	16.39	Quadrillion
EuroForMix	16.98	Quadrillion
LikeLTD (v6)	18.08	Quintillion
TrueAllele	18.21	Quintillion

Broad TrueAllele community

Post-conviction. 10 exonerations
Defendants. 200 cases; acquittals
Crime labs. 10,000 criminal cases
WTC disaster. 18,000 victim remains
DNA scientists measuring information
Police & prosecutors
Civil & parentage disputes
Automated DNA databases
Exact error rate determination
Removing human bias from forensics

Chubbs appellate opinion

Trade Secret Privilege
Defense requires a prima facie, particularized showing that the source code is relevant and necessary.

Necessity of Source Code
Chubbs has not demonstrated how TrueAllele's source code is necessary to his ability to test the reliability of its results.

Confrontation Clause
The Sixth Amendment right to confrontation does not confer a right to discover privileged information before trial.

2016: Chubbs pleaded "no contest", for 7 years 8 months

Other source code decisions

Hearing with expert testimony
- Pennsylvania v. Kevin Foley
- Virginia v. Matthew Brady
- Ohio v. Maurice Shaw
- New York v. John Wakefield
- Pennsylvania v. Michael Robinson
- Washington v. Emanuel Fair
- Missouri v. Reginald Clemons
- Nebraska v. Charles Simmer
- Tennessee v. Demontez Watkins
- Georgia v. Thaddus Nundra
- Georgia v. Monte Baugh and Thaddeus Howell
- Georgia v. Alexander Battle
- Georgia v. Adedoja Bah
- Pennsylvania v. Anthony Spudis

No hearing or expert testimony
- Virginia v. Darwin Bowman
- Maryland v. Adan Espinoza Canela
- California v. Martell Chubbs
- Pennsylvania v. Allen Wade
- Pennsylvania v. Jake Knight
- Pennsylvania v. Chelsea Arganda and Chester White
- California v. Billy Ray Johnson
- New Jersey v. Corey Pickett
- United States v. Lafon Ellis

Under suitable protective orders, Cybergenetics has provided its TrueAllele source code to defendants.

Gambit feeds on false facts

California v. Chubbs "access to this code is the only satisfactory and professionally recommended way to fully consider the validity of the TrueAllele analysis" – Expert

New Jersey v. Pickett "[source code access] would allow that expert to independently test whether the evidentiary software operates as intended" – Court

Harvard Data Science Review "In a criminal trial in New York, the judge ultimately excluded results from two competing DNA mixture algorithms, STRMix, which was developed by the FBI, and **TrueAllele**, created by a commercial developer, which reached opposite results." – Law Professor

Escalating demands

Read the source code. (170,000 lines)

Get **electronic** source code. (If they lose it?)

Run source code in development environment.

Obtain the full software **build** environment.

Rebuild the entire software system themselves.

Never mention **testing** software application on data, which is how real scientists assess **reliability**.

Goals & consequences

Goal 1: suppress reliable scientific evidence

Goal 2: eliminate forensic science innovation

Impact: blocking scientific truth leads to injustice

Goal 3: eliminate trade secret protection for all software (little patent protection: *Alice Corp. v. CLS Bank*)

Impact: reducing Silicon Valley competitiveness

Conclusions

In science and law, reliability is based on **empirical testing**.

Source code doesn't provide empirical testing. Demands are **irrelevant to reliability**, and **destructive to innovation**.

Transparency requires software **testing access**.
Trade secrets are **not material** and can ruin companies.

Cybergenetics

JUSTICE THROUGH SCIENCE