1 LISA S. GREEN, District Attorney
   County of Kern
2 **Cynthia J. Zimmer, SBN 116401**
   Deputy District Attorney
3 Kern County Justice Building
   1215 Truxtun Avenue
4 Bakersfield, California 93301
   (661) 868-2340
5

6

                    SUPERIOR COURT OF CALIFORNIA
7
            COUNTY OF KERN - METROPOLITAN DIVISION
8

9

10 THE PEOPLE OF THE STATE OF CALIFORNIA,)   Case No.: **BF 151825 A**
)
11   Plaintiff,   )
)
12   )   **PEOPLE'S OPPOSITION TO**
              vs   )   **DEFENDANT'S MOTION TO**
)   **REVEAL TRUEALLELE SOURCE**
13   )   **CODES**
    BILLY RAY JOHNSON,   )
14   )
)   Date: January 5, 2015
15   )   Time: 8:30 a.m.
              Defendant,   )   Dept.:   One
16 _____)

17

18   Comes now the plaintiff, the People of the State of California, by and through their attorneys,

19 **LISA GREEN, District Attorney**, Cynthia J. Zimmer, Deputy District Attorney, who respectfully

20 submit the following Opposition to Defendant's Motion to Reveal TrueAllele Source Codes.

21   The People assert that discovery of the source codes should not be ordered on the grounds that

22 (i) the materials sought by Defendant are privileged pursuant to Evidence Code Section 1060; and (ii)

23   -1-

    People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
24   BF151271 A

the materials sought by Defendant are not material or necessary.

I.

FACTS

A.   Procedural Background.

The defendant is now seeking discovery of the TrueAllele source codes. The defense did not comply with discovery statutes set out in Penal Code Section 1054 et seq. Therefore, the request is untimely.

B.   Background on TrueAllele Casework and DNA

TrueAllele is a probabilistic genotyping computer system that interprets DNA evidence using a statistical model. (Perlin Declaration, ¶ 8). The program and its source codes are owned by Cybergenetics, a Pennsylvania corporation that Dr. Perlin founded and currently is employed by. (Perlin Declaration, ¶ 6). TrueAllele is used to analyze DNA evidence, particularly in cases where human review might be less reliable or not possible. (Perlin Declaration, ¶ 9).

1.   The Role of TrueAllele in the Analysis of Uncertain DNA Evidence

A definite genotype can be readily determined when abundant DNA from one person produces unambiguous genetic data. (Perlin declaration, ¶ 10). However, when data signals are less definitive, or when two or more people contribute to the evidence, uncertainty arises. (Perlin Declaration, ¶ 11). This uncertainty is expressed in the derived contributor genotype, which may describe different genetic

-2-

identity possibilities. (Perlin Declaration, ¶ 12). Such genotype uncertainty *may* translate into reduced identification information when a comparison is made with a suspect. (Perlin Declaration, ¶ 13).

The DNA identification task can thus be understood as a two-step process:

1. objectively *inferring genotypes* from evidence data, accounting for allele pair uncertainty using probability, and
2. subsequently *matching genotypes,* comparing evidence with a suspect relative to a population, to express the strength of association using probability. (Perlin Declaration, ¶ 14).

The match strength is reported as a single number, the likelihood ration (LR), which quantifies the change in identification information produced by having examined the DNA evidence. (Perlin Declaration, ¶ 15).

The TrueAllele® Casework system is Cybergenetics' computer implementation of this two-step DNA identification inference approach. (Perlin Declaration, ¶ 16). Cybergenetics began developing TrueAllele 20 years ago, adding a mixture module 15 years ago. (Perlin Declaration, ¶ 17). The Casework system underwent many rounds of testing and model refinement over 10 years, with the current version 25 released in 2009. (Perlin Declaration, ¶ 18).

The TrueAllele computer objectively infers genotypes from DNA data through statistical modeling, without reference to a known comparison genotype. (Perlin Declaration, ¶ 19). To preserve the identification information present in the data, the system represents genotype uncertainty using probability. (Perlin Declaration, ¶ 20). These probabilistic genotypes are stored on a relational database. (Perlin Declaration, ¶ 21). Subsequent comparison with suspects provides evidentiary identification information. (Perlin Declaration, ¶ 22).

-3-

People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
BF151271 A

1

2. TrueAllele Enjoys Widespread Acceptance

2

3

TrueAllele has been used in about 200 criminal cases, with expert witness testimony given in

over 20 trials. (Perlin Declaration, ¶ 23). Courts accepting TrueAllele evidence include California,

4

5

Pennsylvania, Virginia, Federal, United States Marine Corp, Northern Ireland and Australia. (Perlin

Declaration, ¶ 24). Over 10 crime laboratories have purchased the TrueAllele System for their own in-

6

house use, and 3 labs are on-line with their validated systems. (Perlin Declaration, ¶ 25).

7

8

TrueAllele was used to identify human remains in the World Trade Center disaster, comparing

18,000 victim remains with 2,700 missing people. (Perlin Declaration, ¶ 26). Both prosecutors and

9

defenders use TrueAllele for determining DNA match statistics. (Perlin Declaration, ¶ 27). TrueAllele

10

reliability has been confirmed in appellate precedent in Pennsylvania.

11

3.   TrueAllele is Reliable

12

13

Over twenty TrueAllele validation studies have been conducted to establish the reliability of

14

the method and software.[1] Seven of these studies have been published in peer-reviewed scientific

15

journals, on both synthetic[2] and casework[3] data. Conducting such validations is consistent with the

16

17

18

[1] Perlin MW, Szabady B. Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *J Forensic Sci*. 2001;46(6):1372-7.

19

[2] 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

20

information. Sci 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. *J Forensic Sci*. 2015;*in press*; Greenspoon SA, Schiermeier-Wood L, Jenkins BA. Pushing the limits of TrueAllele® Casework: a validation study. *J Forensic Sci*. 2015:*in*

21

press.

[3] Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW. Validating TrueAllele® DNA

22

mixture interpretation. *J Forensic Sci*, 2011;56(6):1430-1447; Perlin MW, Belrose JL, Duceman BW. New York State TrueAllele® Casework validation study. *J Forensic Sci*. 2013;58(6):1458-1466; Perlin MW, Dormer K, Hornyak J,

23

-4-

People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
BF151271 A

24

FBI's 2010 Scientific Working Group on DNA Analysis Methods (SWGDAM) interpretation guidelines.[4]

TrueAllele has been admitted into evidence after opposition challenge in seven courts, located in California, Louisiana, Ohio, Pennsylvania, Virginia, Northern Ireland and Australia. (Perlin Declaration, ¶ 35).

Cybergenetics has a strong financial incentive to ensure the reliability of its widely used TrueAllele system. (Perlin Declaration, ¶ 36). Cybergenetics continually tests its software and conducts scientific validation studies to ensure TrueAllele's reliability. (Perlin Declaration, ¶37).

4. Background on Software Source Code

People write a computer program in a programming language using "source code". (Perlin Declaration, ¶ 38). This source code is later translated into computer-readable "executable" software. (Perlin Declaration, ¶ 39). The source code details step-by-step human-readable instructions that describe to the computer and programmers how the program operates. (Perlin Declaration, ¶ 40).

TrueAllele is written in MATLAB (for *MATrix LABoratory),* a high level mathematical language for programming and visualizing numerical algorithms made by the MathWorks (Natick, MA. (Perlin Declaration, ¶ 41). Here is an example of MATLAB source code, simplified from a few lines of the built-in "Mhsample" function that performs Metropolis-Hastings statistical sampling:

```
U = log(rand(nchain,nsamples+burnin));
```

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.
[4] SWGDAM. Interpretation guidelines for autosomal STR typing by forensic DNA testing Laboratories. 2010; http://www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines at paragraph 3.2.2.)

People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
BF151271 A

```
      for i = 1-burnin:nsamples;

      y = proprnd(x0);

      q1 = logproppdf(x0,y);

      q2 = logproppdf(y,x0);

      rho = (ql+logpdf(y))-(q2+logpdf(x0));

      Ui = U(:,i+burnin);

      acc =  Ui<=min(rho,0);

      x0(acc,:) = y(acc,:);

      accept = accept+(acc);

      end
```

Thus, source code is written in language that humans are capable of understanding, but only if they are fluent in reading, writing and interpreting the particular language that the program is written in. (Perlin Declaration, ¶¶ 42-43).

    5.   TrueAllele is a Trade Secret

    People can easily copy a computer program if they have its source code. (Perlin Declaration, ¶ 46). Source code contains the software design, engineering know-how, and algorithmic implementation of the entire computer program. (Perlin Declaration, ¶47). Cybergenetics has invested millions of dollars over two decades to develop its TrueAllele system, the company's flagship product.

People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
BF151271 A

(Perlin Declaration, ¶ 48). Cybergenetics does not disclose the source code to anyone outside the company. In fact, the source code has never been disclosed. (Perlin Declaration ¶ 49).

Cybergenetics operates in a commercially competitive environment. (Perlin Declaration, ¶ 51). In recent years, at least five other groups have developed similar software. (Perlin Declaration, ¶ 52). There is keen interest from competitors to find out how to replicate TrueAllele. The TrueAllele software represents a technological breakthrough that has not been successfully replicated by any other company as of this date. (Perlin Declaration, ¶ 53) Disclosure of the TrueAllele source code trade secret would cause irreparable harm to the company, enabling competitors to easily copy the company's proprietary products and services. (Perlin Declaration, ¶ 54) Ownership of the TrueAllele program and source code provides Cybergenetics with an advantage over its competitors who do not know the proprietary code and could not legally duplicate it. (Perlin Declaration, ¶ 55)

Cybergenetics takes reasonable measures to protect the secrecy of the source code. (Perlin Declaration, ¶ 56). All information relating to the source code is housed on secure computers. (Perlin Declaration, ¶ 56). TrueAllele derives value from remaining secret. (Perlin Declaration, ¶ 57).

In contrast to so-called "open source" programs, for-profit companies do not make their source codes available to the public. (Perlin Declaration, ¶ 58). Open source programs typically are not validated prior to release, because the process of perfecting software is costly. (Perlin Declaration, ¶ 59). Cybergenetics offers the TrueAllele software for license by crime labs and to other interested parties. (Perlin Declaration, ¶ 60). The company currently charges a base license fee of $60,000. (Perlin Declaration, ¶ 61).

-7-

1   Individuals and companies can also submit samples to Cybergenetics for testing and analysis

2 for a fee. (Perlin Declaration, ¶ 62). Cybergenetics provides opposing experts the opportunity to

3 review the TrueAllele process, examine results, and ask questions. (Perlin Declaration, ¶ 63). This

4 review can be done in Cybergenetics' Pittsburgh office, or through an Internet Skype-like meeting.

5 (Perlin Declaration, ¶ 63).

6   II.

7   LEGAL ARGUMENT

8
   A.   The TrueAllele Source Code is Privileged Under Cal. Evid. Code § 1060
9
10   Evidence Code 1060 provides that "the owner of a trade secret has a privilege to refuse to

11 disclose the secret, and to prevent another from disclosing it, if the allowance of the privilege will not

12 tent to conceal fraud or otherwise work injustice." (*Bridgestone/Firestone, Inc. v. Superior Court*

13 (1992) 7 Cal. App. 4th 1383). The *Bridgestone/Firestone* Court held that, because Evidence Code

14 section 1060 creates a privilege for trade secrets, use of a heightened standard is necessary before

15 disclosure of documents containing trade secrets will be compelled. (*Id*. at p. 1393.)

16
           Under this heightened standard, the party seeking discovery must make a prima facie,
17
   particularized showing that the information sought is relevant and necessary to the proof of a claim or
18
   defense, and that it is "reasonable to conclude that the information sought is essential to a fair
19
   resolution" of the proceeding. (*Bridgestone//Firestone, Inc. v. Superior Court, supra*, 7 Cal.App.4th at
20
   p. 1393 (emphasis added).) If the party seeking disclosure makes such a particularized showing, the
21
   burden shifts to the privilege holder "to demonstrate any claimed disadvantages of a protective order."
22

23   -8-

Id. In addition, the *Bridgestone* Court held that "either party may propose or oppose less intrusive alternatives to disclosure." *Id*.

In this case, there is no need to demonstrate the disadvantages of a protective order, because (i) the TrueAllele source code is indisputably a trade secret, and (ii) the defendant has failed to make a particularized showing that the information sought is relevant and necessary to his defense.

      1.   The Source Code is a Trade Secret

The California Uniform Trade Secrets Act defines a "trade secret" as including "information, including a formula, pattern, compilation, program, device, method, technique, or process, that: (1) Derives independent economic value, actual or potential, from not being generally known to the public or to other persons who can obtain economic value from its disclosure or use; and (2) is the subject of efforts that are reasonable under the cirumstances to maintain its secrecy." Cal. Civ. Code § 3426.1. the trade secret privilege incorporates this definition by reference. Cal. Evid Code § 1061(a). Software source code, if commercially valuable and maintained byits owner as confidential, generally qualifies as a trade secret. *See, e.g., Wellogix, Inc. v. Accenture, LLP,* 716 F.3d867 (5th Cir. 2013)(affirming conclusion that proprietary source code was a trade secret); MAI *Sys. Corp. v. Peak Computer, Inc.,* 991 f.3D 511, 522 (9TH Cir. 1993)(applying California version of the UTSA); *B&B Microscopes v. Armogida,* 532 F. Supp. 2d 744 (W.D. Pa. 2007)(finding that source code for image analysis software developed for crime lab was a trade secret); *Aries Inf. Sys., Inc. v. Pacific Management Sys. Corp.,* 366 N W.2d 366. 368 (Minn. Ct. App. 1985) (applying Minnesota version of UTSA); *TDS Healthcare sys. Corp. v. Humana Hops. Illinois, Inc*., 880 f. Supp. 1572, 1582 (N.D. Ga. 1995)(applying Georgia version of UTSA). *See also United States v. Agrawal,* 726 F.3d 235 (2d Cir.

-9-

2013)(affirming conviction for theft of source code qualifying as trade secret under the Economic Espionage Act).

As is established by the Declaration of Dr. Perlin submitted in support of this motion, the TrueAllele source code qualifies as a trade secret that is entitled to legal protection under California law. The TrueAllele source code has never been disclosed to the public. (Perlin Declaration, ¶ 57). Cybergenetics operates in a highly competitive market. (Perlin Declaration, ¶ 51). The TrueAllele software represents a technological breakthrough that has not been successfully replicated by any other company. (Perlin Declaration, ¶ 53). Although the technology is patented, the source code itself is not disclosed by any patent and cannot be derived from any publicly disclosed source. (Perlin Declaration, ¶ 48). The source code is the product of 15 years' worth of programming effort and an investment of millions of dollars. (Perlin Declaration, ¶ 48). Cybergenetics takes extensive measures to protect the source code and maintain its proprietary status. (Perlin Declaration, ¶¶ 49, 56).

The source code is not known by Cybergenetics' competitors. (Perlin Declaration, ¶¶ 49, 50, 57). The fact that the source code is kept secret provides Cybergenetics with a significant advantage over others who do not have access to the source code and do not have the programming know-how or are not willing to make the investment necessary to develop comparable software. (Perlin Declaration, ¶ 50). These facts have not and cannot be rebutted by the defendant.

2.   The Defense Has Not Made the Required "Particularized" Showing

A party seeking access to source code that qualifies as a trade secret is required to meet a heightened standard. (*See Bridgestone/Firestone, Inc. v. Superior Court, supra,* 7 Cal.App.4th at p. 1393). In the *Bridgestone* case, the plaintiffs sought to compel the trade secret formulas for a tire in a

-10-

People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
BF151271 A

1   products liability action. In support of the motion, the Plaintiffs submitted a declaration of their expert

2   that asserted that the formulas were needed to reach conclusions about the cause of the tire failure.

3   The expert in *Bridgestone* "gave specific examples of the manner in which the formulas were helpful

4   in evaluating the reasons why tire components fail," and explained "from his own experience and that

5   of others in the field" that the information contained in the formulas "was important in an analysis and

6   proof of why a tire failed." *Id.* at 1396.

7   The *Bridgestone/Firestone* Court found this was insufficient because the expert failed to

8   "describe with any precision how or why the formulations were predicate to his ability to reach

9   conclusions in this case." *Id.* at 1397. The court also noted that, although the expert's declaration

10 indicated that the formulas "would be helpful to the analysis of the case and to [the expert's] ability to

11 reach conclusions and render opinions, "simply being helpful is not enough.

12

13   The defense has made a much weaker showing in this instance. The defense fails to explain

14 with any precision or particularity how a review of the TrueAllele source code would enable the

15 defense to determine what assumptions were made, or how reviewing the highly technical code would

16 help defense counsel cross examine Dr. Perlin at Kelly-Frye hearing. Moreover, the defense has failed

17 to explain why the documentation and reports that have already been provided are insufficient, and

18 why the source code also is necessary. As discussed in more detail below, other additional methods of

19 validating and testing the software are available. The source code therefore is not necessary, or even

20 relevant.

21   Because the defense has failed to make a prima facie showing meeting the threshold standard,

22 consideration of claimed disadvantages of a protective order is not necessary.

23   -11-

24

B.   The TrueAllele Source Code is Not Material or Necessary to the Defense

The defendant fails to provide any explanation of why a review of source code would be useful to their proposed inquiry, or how the source code will provide answers to any of the questions posed by the defense. Neither the defense nor any expert it has identified has argued that the software is defective or has errors that could be detected by reviewing 170,000 lines of source code.

For those reasons, defendant has failed to make any showing that the TrueAllele source code is material. Materiality of evidence looks to the relationship between the proposition for which the evidence is offered and the issues in the case. McCormick on evidence § 541 (Edward W. Cleary, Lawyer's Ed. 1984). The mere fact that testing equipment or computer-based analysis is used in a criminal case does not demonstrate that the defense is entitled to examine the testing method or obtain the equipment's source code. For that reason, courts faced with source code requests have required a particularized showing demonstrating that "observed discrepancies" in the testing require access to the source code. *See, e.g., state v. Bastos,* 985 So. 3d 37 (Fla. Dist. Ct. App.3d Dist. 2008); *Young v. State,* 749 s. e. 2D 423 (Ga. Ct. App. 2013) (holding that defendant must show that there is a specific logical connection between the source code to be examined and some consequential fact).

In order to demonstrate materiality of the source code in this case, Defendant must come forward with some admissible evidence – and not simply hearsay statements or proffers by counsel – capable of supporting a finding that examination of the source code will provide evidence that is both relevant to and favorable to the defense. That has not happened here. Moreover, production of the source code is simply not necessary for the following reasons.

-12-

1 **1.** **The TrueAllele program is Validated.** The underlying source code does not need to

2 be made available because software reliability can be assessed through validation studies. Ruling on

3 an appeal regarding the admissibility of TrueAllele evidence in a 2009 homicide trail Pennsylvania

4 Superior Court Judge Panella wrote:

> "Foley's third reason for exclusion is misleading because scientists can validate the reliability
>
> of a computerized process even if the "source code" underlying that process is not available to
>
> the public. TrueAllele is proprietary software; it would not be possible to market TrueAllele if
>
> it were available for free. Nevertheless, TrueAllele has been tested and validated in peer-
>
> reviewed studies. One study used laboratory-generated DNA samples and found that
>
> quantitative analysis performed by TrueAllele was much more sensitive than qualitative
>
> analysis such as that performed by the FBI."

Commonwealth v. Foley, 47 A.3d 882, 889 (Pa. Super. 2012). The defendant has not provided any

basis for reaching a different conclusion regarding the necessity of a source code review.

  2. **TrueAllele is Reliable.** TrueAllele is an extensively validated system. Validation

studies assess system accuracy and reliability, measuring sensitivity (how well contributors are

included), specificity (how well non-contributors are excluded) and reproducibility (how well

independent solutions agree). Seven studies have been published in peer-reviewed scientific journals,

studies have been conducted. (Perlin Declaration, ¶ 33). Regulatory bodies in New York and Virginia

have had independent scientists review validation studies before they granted approval for their state

crime laboratories to use TrueAllele for casework. (Perlin Declaration, ¶ 34). Cybergenetics

People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
BF151271 A

thoroughly tests the software before it is released. (Perlin Declaration, ¶ 32). Because TrueAllele is a validated reliable system, there is no need for disclosing source code.

3. **Review of the source Code is Unrealistic.** TrueAllele has about 170,000 lines of computer source code, written by multiple programmers over two decades. (Perlin Declaration, ¶ 44). The computer code is dense mathematical text. (Perlin Declaration, ¶ 44). It can take hours for a person to read through even a few dozen lines of MATLAB to decipher what it does. (Perlin Declaration, ¶ 44). It is wholly unrealistic to expect that reading through TrueAllele source code would yield meaningful information. (Perlin Declaration, ¶ 45). Since having the source code does not serve any useful purpose, there is no need to disclose it.

4. **Ample and Sufficient Evidence Has Already Been Provided.** TrueAllele's reliability was established on the evidence in this case. (Perlin Declaration, ¶ 64). The report and its supporting case packet described the system's sensitivity, specificity and reproducibility on the DNA evidence. (Perlin Declaration, ¶ 64). Since this reliability was already established on the evidence in this case, there is no need for source code. (Perlin Declaration, ¶ 64).

5. **There Are Alternatives Means to Inspect the Software.** Cybergenetics offers experts and attorneys the opportunity to review the TrueAllele process, examine results, and ask questions. (Perlin Declaration, ¶ 63). This review can be done in Cybergenetics's Pittsburgh office, or through an Internet Skype-like meeting. (Perlin Declaration, ¶ 63). Cybergenetics regularly explains the system, and the results obtained in a case, to both prosecution and defense. (Perlin Declaration, ¶ 63). This introduction to the TrueAllele method, the case data, and the application of the method to the data, is a

-14-

logical first step in understanding how the system works. (Perlin Declaration, 20 63). Source code is not necessary.

6. **There Are Alternatives Means of Validating the Program.** Cybergenetics offers commercial services for validating DNA mixture interpretation methods. (Perlin Declaration, ¶ 65). Any party can provide DNA validation data and obtain these services to assess True-Allele reliability. (Perlin Declaration, ¶ 65). Since TrueAllele is an objective process, and produces unbiased DNA identification results that do not "know" comparison genotypes during analysis, it is easy for Cybergenetics to perform these studies. (Perlin Declaration, ¶ 65). Source code is not needed for obtaining these services. (Perlin Declaration, ¶ 65).

7. **Source Code Production is Rare, if not Unprecedented.** Source code is not made available for the commercial software that is regularly used and relied upon in the area of forensic DNA identification. (Perlin Declaration, ¶ 67). Such software includes Life Technology's "Genemapper ID" for generating and analyzing DNA match statistics, and Microsoft "Excel" for conducting additional DNA analysis. (Perlin Declaration, ¶ 67). Source code is not needed to assess the reliability of these critical software programs, since they have all been tested and validated. (Perlin Declaration, ¶ 67).

III.

CONCLUSION

The information sought is protected by the Trade Secret Privilege. The Defense has failed to make a prima facie showing of the particularized need for TrueAllele's source codes.

-15-

People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
BF151271 A

1   Therefore, the Defendant's Motion to Reveal Source Codes should be denied.

2

3

4

5   Dated: January 16, 2015   LISA GREEN
                                                            DISTRICT ATTORNEY
6

7   _____
                                                            Cynthia J. Zimmer
8   Supervising Deputy District Attorney

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23   -16-
     People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
24   BF151271 A

DECLARATION OF DR. MARK W. PERLIN

I, Mark W. Perlin, declare I have personal knowledge of the following, and if called upon to do so, could and would testify competently to the matters contained herein:

1.   I have been retained to testify as an expert witness on behalf of the prosecution in People v. Johnson, BF151825, pending before this Court (Johnson).

2.   I hold the following academic degrees: a B.A. in Chemistry from SUNY/Binghamton, a Ph.D. in Mathematics from CUNY/Graduate School, an M.D. from the University of Chicago Pritzker School of Medicine, and a Ph.D. in Computer Science from Carnegie Mellon University. I hold ten patents. Prior to founding my own technology company, I was a senior research faculty member of Carnegie Mellon University's School of Computer Science. I have been qualified to testify as an expert in fifteen jurisdictions. I am currently an adjunct faculty member at Duquesne University.

3.   I reside in Pittsburgh, PA.

4.   As part of my engagement in the Johnson case, Cybergenetics, the company that I founded and currently work for, analyzed evidence samples and issued a report using Cybergenetics' proprietary TrueAllele® software. Copies of the report and supporting materials have been provided to defense counsel as part of pretrial discovery practice. I understand that the defense is seeking access to the source codes and pseudosource codes for the TrueAllele software.

5.   I have provided consent to the people to assert the Trade Secret privilege pursuant to Cal. Evid. § § 1060 and 1061 on behalf of myself and Cybergenetics in relation to the source code and pseudo source codes sought by the defense.

6.   Cybergenetics is a Pennsylvania corporation located at 160 North Craig Street, Suite 210, Pittsburgh, PA 15213. Cybergenetics is the owner of the TrueAllele software, as well as its proprietary source code.

-17-

People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
BF151271 A

7. Cybergenetics does not maintain an office in the state of California. Although I have previously testified on behalf of parties in the state of California, neither I nor Cybergenetics regularly conduct business in that state.

The Role of TrueAllele in DNA Analysis

8. TrueAllele is a probabilistic genotyping computer system that interprets DNA evidence using a statistical model.

9. TrueAllele is used to analyze DNA evidence, particularly in cases where human review might be less reliable or not possible.

10. A definite genotype can be readily determined when abundant DNA from one person produces unambiguous genetic data.

11. However, when data signals are less definitive, or when two or more people contribute to the evidence, uncertainty arises.

12. This uncertainty is expressed in the derived contributor genotype, which may describe different genetic identity possibilities.

13. Such genotype uncertainty may translate into reduced identification information when a comparison is made with a suspect.

14. The DNA identification task can thus be understood as a two-step process:

    (1.)   objectively inferring genotypes from evidence data, accounting for allele pair uncertainty using probability, and

    (2.)   subsequently matching genotypes, comparing evidence with a suspect relative to a population, to express the strength of association using probability.

15. The match strength is reported as a single number, the likelihood ratio (LR), which quantifies the change in identification information produced by having examined the DNA evidence.

16. The TrueAllele® Casework system is Cybergenetics' computer implementation of this two-step DNA identification inference approach.

-18-

17. Cybergenetics began developing TrueAllele 20 years ago, adding a mixture module 15 years ago.

18. The Casework system underwent many rounds of testing and model refinement over 10 years, with the current version 25 release in 2009.

19. The TrueAllele computer objectively infers genotypes from DNA data through statistical modeling, without reference to a known comparison genotype.

20. To preserve the identification information present in the data, the system represents genotype uncertainty using probability.

21. These probabilistic genotypes are stored on a relational database.

22. Subsequent comparison with suspects provides evidentiary identification information.

TrueAllele's Widespread Acceptance

23. TrueAllele has been used in over 200 criminal cases, with expert witness testimony given in over 20 trials.

24. Courts accepting TrueAllele evidence include California, Louisiana, Maryland, New York, Ohio, Pennsylvania, Virginia, United States Marine Corps, Northern Ireland and Australia.

25. Over 10 crime laboratories have purchased the TrueAllele system for their own in-house use, and 3 labs are on-line with their validated systems.

26. TrueAllele was used to identify human remains in the World Trade Center disaster, comparing 18,000 victim remains with 2,700 missing people.

27. Both prosecutors and defenders use TrueAllele for determining DNA match statistics.

28. TrueAllele's reliability has been confirmed in appellate precedent in Pennsylvania. *See Commonwealth v. Foley,* 47 a.3D 882 (Pa. super. 2012)

29. The TrueAllele calculation is entirely objective: when it determines the genotypes for the contributors to the mixture evidence, the computer has no knowledge of the comparison genotypes. Genotype comparison and match statistic determination are only done *after* genotypes

-19-

have been computed. In this way, TrueAllele computing avoids human examination bias, and provides a fair match statistic.

30.   I agree with the conclusions that were reached in the *Foley* case, which found that (i) scientists can validate the reliability of a computerized process even if the source code is not available to the public; (ii) it would not be possible to market TrueAllele if it were available for free; (iii) TrueAllele has been tested and validated.

31.   There is no genuine controversy as to the validity and reliability of the TrueAllele method. To the contrary, computer analysis of uncertain data using probability modeling is the scientific norm. Forensic science researchers see this as the best approach.

32.   Cybergenetics thoroughly tests its software before it is released.

33.   Over twenty TrueAllele validation studies have been conducted to establish the reliability of the method and software. Seven studies have been published in peer-reviewed scientific journals, for both laboratory-generated and casework DNA samples.

34.   Conducting such validations is consistent with the FBI's 2010 Scientific Working Group on DNA Analysis Methods (SWGDAM) interpretation guidelines. Regulatory bodies in New York and Virginia have had independent scientists review validation studies before they granted approval for their state crime laboratories to use TrueAllele for casework.

35.   TrueAllele has been admitted into evidence after opposition challenge in seven courts, located in California, Louisiana, Ohio, Pennsylvania, Virginia, Northern Ireland and Australia.

36.   Cybergenetics has a strong financial incentive to ensure the reliability of its widely used TrueAllele system.

37.   Cybergenetics continually tests its software and conduct's scientific validation studies to ensure TrueAllele's reliability.

Background on Software Source Code

38.   People write a computer program in a programming language using "source code".

People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
BF151271 A

39.   This source code is later translated into computer-readable "executable" software.

40.   The source code details step-by-step human-readable instructions that describe to the computer and programmers how the program operates.

41.   TrueAllele is written in MATLAB (for MATrix LABoratory), a high level mathematical language for programming and visualizing numerical algorithms made by the MathWorks (Natick, MA).

42.   Here is an example of MATLAB source code, simplified from a few lines of the built-in "mhsample" function that performs Metropolis-Hastings statistical sampling:

```
U = log(rand(nchain,nsamples+burnin));

for i = 1-burnin:nsamples;

y = proprnd(x0);

q1 = logproppdf(x0,y);

q2 = logproppdf(y,x0);

rho = (ql+logpdf(y))-(q2+logpdf(x0));

Ui = U(:,i+burnin);

acc =  Ui<=min(rho,0);

x0(acc,:) = y(acc,:);

accept = accept+(acc);

end
```

//

People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
BF151271 A

43. Thus, source code is written in a language that humans are capable of understanding, but only if they are fluent in reading, writing and interpreting the particular language that the program is written in.

44. TrueAllele has about 170,000 lines of computer source code, written by multiple programmers over two decades. The computer code is dense mathematical text. It can take hours for a person to read through even a few dozen lines of MATLAB to decipher what it does.

45. In my opinion, it is wholly unrealistic to expect that reading through TrueAllele source code would yield meaningful information.

Why TrueAllele is a Trade Secret

46. People can easily copy a computer program if they have its source code.

47. Source code contains the software design, engineering know-how, and algorithmic implications of the entire computer program.

48. Cybergenetics has invested millions of dollars over two decades to develop its TrueAllele system, the company's flagship product. Although the technology is patented, the source code itself is not disclosed by any patent and cannot be derived from any publicly disclosed source.

49. Cybergenetics considers the TrueAllele source code to be a trade secret. Cybergenetics does not disclose the source code to anyone outside the company. In fact, the source code has never been disclosed. The source code is not distributed to employees of Cybergenetics, and copies are not provided to individuals, businesses or government agencies that use or license the software.

50. The fact that the source code is kept secret provides Cybergenetics with a significant advantage over others who do not have access to the source code and do not have the programming know-how or are not willing to make the investment necessary to develop comparable software.

-22-

51.   Cybergenetics operates in a highly competitive commercial environment.

52.   In recent years, at least five other groups have developed similar software.

53.   There is keen interest from competitors to find out how to replicate TrueAllele. The TrueAllele software represents a technological breakthrough that has not been successfully replicated by any other company as of this date.

54.   Disclosure of the TrueAllele source code trade secret would cause irreparable harm to the company, enabling competitors to easily copy the company's proprietary products and services.

55.   Ownership of the TrueAllele program and source code provides Cybergenetics with an advantage over its competitors who do not know the proprietary code and could not legally duplicate it.

56.   Cybergenetics takes reasonable measures to protect the secrecy of the source code. For example, all information relating to the source code is housed on secure computers.

57.   TrueAllele's source code derives value from remaining secret, and has never been disclosed to the public.

58.   In contrast to so-called "open source" programs, for-profit companies do not make their source codes available to the public.

59.   Commercial software programs are extensively validated while in development and before release and commercialization. By their nature, open source programs typically are not validated prior to release, because the process of perfecting software is costly. Open source forensic DNA analysis software programs tent to be relatively short programs consisting of several hundreds of lines that realistically can be reviewed by a human being.

-23-

1   60.   Cybergenetics offers the TrueAllele software for license by crime labs and to other

2   interested parties.

3   61.   The company currently charges a base license fee of $60,000.

4   62.   Individuals and companies can also submit samples to Cybergenetics for testing and

5   analysis for a fee.

6   63.   Cybergenetics provides opposing experts the opportunity to review the TrueAllele

7   process, examine results, and ask questions. This review can be done in Cybergenetics Pittsburgh

8   office, or through an Internet Skype-like meeting. Cybergenetics regularly explains the system, and

9   the results obtained in a case, to both prosecution and defense. This introduction to the TrueAllele

10   method, the case data, and the application of the method to the data, is a logical first step in understand

11   how the system works. Source code is not necessary.

12   64.   TrueAllele reliability was established on the evidence in this case. The report and its

13   supporting case packet described the system's sensitivity, specificity and reproducibility on the DNA

14   evidence. Source code is not needed to understand or interpret these materials.

15   65.   Cybergenetics offers commercial services for validating DNA mixture interpretation

16   methods. Any party can provide DNA validation data and obtain these services to access TrueAllele

17   reliability. Since TrueAllele is an objective process, and produces unbiased DNA identification results

18   that do not "know" comparison genotypes during analysis, it is easy for Cybergenetics to perform these

19   studies. Source code is not needed for obtaining these services.

20   66.   Although the source code for TrueAllele is a secret, the methodology it employs and

21   implements has been disclosed. Cybergenetics has published the core mathematics of TrueAllele's

22   underlying mathematical model for over 10 years (in 2001, 2009 and 2011). This information

23   -24-

24   People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
                                        BF151271 A

discloses TrueAllele's genotype modeling mechanism, and enables others to understand the basic

method. Indeed, at least five other groups have independently developed software that uses

TrueAllele's linear mixture analysis approach. The source code is not necessary or helpful to

understand or test the methodology or reliability of the analysis.

67.   To my knowledge, source code is not made available for other commercial software that

is regularly used and relied upon in the area of forensic DNA identification. Such software includes

Life Technology's "Genemapper ID" for generating and analyzing DNA data signals, the Federal

Bureau of Investigation's "Popstats" for producing DNA match statistics, and Microsoft "Excel" for

conducting additional DNA data analysis. Source code is not needed to access the reliability of these

critical software programs, since they have all been tested and validated.

I declare the above is true and correct under penalty of perjury under the law of the state of

California, executed this 16th day of January, 2015, in Pittsburgh, Pennsylvania.

By: _____
                      Dr. Mark Perlin

People's Opposition to Defendant's Motion to Reveal TrueAllele Source Codes
BF151271 A