SUPERIOR COURT OF WASHINGTON FOR KING COUNTY

STATE OF WASHINGTON, Plaintiff, vs. EMANUEL FAIR, Defendant.	)	No. 10-1-09274-5 SEA DECLARATION OF MARK W. PERLIN

I, MARK W. PERLIN, hereby declare as follows:

1. I am over 18 years of age and I am competent to make this declaration.
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 eleven 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 twenty five jurisdictions. I am currently an adjunct faculty member at Duquesne University.
3. I reside at 5885 Marlborough Road, Pittsburgh, PA 15217.

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

4. 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.

5. The Role of TrueAllele in DNA Analysis

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

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

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

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

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

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

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

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

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

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

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.

17. Cybergenetics began developing TrueAllele 22 years ago, adding a mixture module 17 years ago.

18. The casework system underwent many rounds of testing and model refinement over 10 years before it was used in criminal casework, with the current version 25 released 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 or other individuals provides identification information that can be used as evidence.

23. TrueAllele's Widespread Acceptance

24. TrueAllele has been used in over 500 criminal cases, with expert witness testimony given in over 50 trials.

25. Courts accepting TrueAllele evidence include California, Louisiana, Maryland, New York, Ohio, Pennsylvania, South Carolina, Virginia, United States (Eastern District of Virginia), United States Marine Corps, Northern Ireland, and Australia.

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

26. Over 10 crime laboratories have purchased the TrueAllele system for their own in-house use, and 7 labs are on-line with their validated systems.

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

28. Both prosecutors and defenders use TrueAllele for determining DNA match statistics. TrueAllele is also used by innocence projects and for post-conviction relief (*Connecticut v. Ralph Birch*, *Indiana v. Darryl Pinkins*, *Maryland v. William Jamison*, *Washington v. Raymond Ben*).

29. TrueAllele's reliability has been confirmed in appellate precedent in Pennsylvania. *See Commonwealth v. Foley*, 47 A.3d 882 (Pa. Super. 2012).

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

31. 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.

32. TrueAllele is Considered to be Reliable

33. 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.

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

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

35. Over thirty validation studies have been conducted by Cybergenetics and other groups to establish the reliability of the TrueAllele method and software. Seven of these studies have been published in peer-reviewed scientific journals, for both laboratory-generated and casework DNA samples. Source code was not needed or used in any of these studies.

36. In the "peer-review" process, scientists describe their research methods, results and conclusions in a scientific paper, which they submit to a journal for publication. An editor at the journal has (at least) two independent and anonymous scientists in the field read the paper, assess its merits, and advise on the suitability of the manuscript for publication. The paper is then accepted, rejected, or sent back to the authors for revision and another round of review.

37. A "laboratory-generated" validation study uses data that has been synthesized in a DNA laboratory, and is of known genotype composition. Four published TrueAllele papers of this type are: 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;52(2):103-14; 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, and Jenkins BC. Establishing the limits of TrueAllele® Casework: a validation study. *Journal of Forensic Sciences*. 2015;60(5):1263-1276.

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

38. A "casework" validation study uses DNA data exhibiting real-world issues developed by a crime laboratory in the course of their usual casework activity. Three published TrueAllele papers of this type are: 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):1458-66; Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, and Greenspoon S, "TrueAllele® Casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. *PLoS ONE*. 2014:9(3):e92837.

39. Conducting such validations is consistent with the FBI's 2010 Scientific Working Group on DNA Analysis Methods (SWGDAM) interpretation guidelines. TrueAllele complies with the 2015 SWGDAM validation guidelines for probabilistic genotyping systems. 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.

40. TrueAllele has been admitted into evidence after opposition challenge in nine courts, located in California, Louisiana, New York, Ohio, Pennsylvania, South Carolina, Virginia, Northern Ireland and Australia.

41. Seven admissibility decisions in the United States are: People of California v. Dupree Langston, Kern County (Kelly-Frye), BF139247B, January 10, 2013; State of Louisiana v. Chattley Chesterfield and Samuel Nicolas, Parish of East Baton Rouge (Daubert), 01-13-0316 (II), November 6, 2014; People of New York v. John Wakefield, Schenectady

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

County (Frye), A-812-29, February 11, 2015; State of Ohio v. Maurice Shaw, Cuyahoga County (Daubert), CR-575691, October 10, 2014; Commonwealth of Pennsylvania v. Kevin Foley, Indiana County (Frye), 2012 PA Super 31, No. 2039 WDA 2009, Superior Court affirmed February 15, 2012; State of South Carolina v. Jaquard Aiken, Beaufort County (Jones), 20121212-683, October 27, 2015; Commonwealth of Virginia v. Matthew Brady, Colonial Heights County (Spencer-Frye), CR11000494, July 26, 2013.

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

43. Cybergenetics continually tests its software and conducts scientific validation studies to ensure TrueAllele's reliability. Source code is not used in validation studies.

44. Cybergenetics improved the speed, accuracy and generality of the user interface LR match statistic calculation in February of 2014. The previous LR estimate could understate the match statistic by around a factor of ten. Genotype computation was not affected. This change is described in Cybergenetics application note "TrueAllele® VUIer™ Likelihood Ratio Calculation."

45. Background on Software Source Code

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

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

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

49. 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).

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

50. Here is an example of MATLAB source code, simplified from a few lines of the MathWorks 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 = (q1+logpdf(y))-(q2+logpdf(x0));
    Ui = U(:,i+burnin);
    acc = Ui<= min(rho,0);
    x0(acc,:) = y(acc,:);
    accept = accept+(acc);
end
```

62. 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.

63. 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.

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

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

65. Why TrueAllele is a Trade Secret

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

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

68. 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. Patent protection is not automatic, and litigation can cost millions of dollars.

69. 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.

70. 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.

71. Cybergenetics operates in a highly competitive commercial environment.

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

73. 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.

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

74. 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.

75. 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.

76. 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.

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

78. In contrast to so-called "open source" programs, for-profit companies generally do not make their source codes available to the public. The relatively few companies that have an open source business model tend to operate in a very large market, utilize free programmer coding, conduct little innovation, and earn their main revenue by providing software services.

79. 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 tend to be relatively short programs consisting of several hundreds of lines of code that realistically can be reviewed by a human being.

80. Open source software suffers from a lack of version control and quality assurance, since any unrelated party can make code changes and release untested products. This chaotic

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

development approach is in marked contrast to the more controlled reliability and versioning requirements of forensic software that is used in criminal proceedings.

81. Irremediable Risks of Source Code Disclosure

82. Third party review of source code can divulge proprietary trade secrets wholly unrelated to reliability, but valuable to competitors. Once a review results in a release of hard-earned engineering know-how, that disclosure cannot be reversed. The source code reviewer's knowledge can be written into other software systems, shared with interested parties, or sold for profit. There are no adequate remedies for redress once this proprietary information has been released.

83. Protective orders for source code are sometimes used in expensive civil litigation for patent infringement, which is not germane to criminal proceedings. Protective orders may fail to protect valuable trade secrets, leading to unwanted disclosure of proprietary designs, methods, and know-how (*Superspeed LLC v. Google*, United States District Court for the Southern District of Texas; *Bradford Technologies, Inc. v. NCV Software.com*, United States District Court for the Northern District of California; *Apple v. Samsung*, United States District Court for the Northern District of California; *Eli Lilly & Co. v. Gottstein*, United States Court of Appeals for the Second Circuit; *Smith & Fuller, PA v. Cooper Tire & Rubber Co.*, United States Court of Appeals for the Fifth Circuit).

84. There is no real effective remedy once a protective order is violated. Courts typically merely reimburse the fees that were incurred by the party whose secrets were revealed. In a case involving source code that is a trade secret, however, once the source code has been revealed in breach of a protective order, it generally loses its status as a trade secret.

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

The genie can't be put back in the bottle, and reimbursement of legal fees does nothing to compensate for the loss of commercial value.

85. Cybergenetics uniquely provides accurate, objective, and neutral DNA identification information for criminal justice. TrueAllele DNA match results are used by both prosecution and defense for an unbiased statistical assessment of biological evidence. Crime laboratories rely on their validated TrueAllele systems for effective interpretation of complex DNA data. Jeopardizing the existence of Cybergenetics through a disclosure of its source code is unreasonable, and does not serve the interests of justice.

86. Why TrueAllele Source Code is Not Needed

87. Cybergenetics offers the TrueAllele software for license by crime labs and to other interested parties.

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

89. Individuals and companies can also submit samples to Cybergenetics for testing and analysis for a fee.

90. Cybergenetics provides opposing experts the opportunity to review the TrueAllele process, examine results, and ask questions. This review can be done in Cybergenetics's Pittsburgh office, or through an Internet Skype-like meeting. Cybergenetics regularly explains the system, and the results obtained in a case, to both prosecution and defense. This introduction to the TrueAllele method, the case data, and the application of the method to the data, is a logical first step in understanding how the system works. Source code is not necessary.

91. The TrueAllele method is inherently objective, since the computer determines evidence genotypes without any knowledge of the comparison reference genotypes. Hence there is

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

no possibility of examination bias when determining genotypes from the DNA data. Match statistics, whether inclusionary or exclusionary, are calculated only afterwards by comparing evidence genotypes with reference genotypes. Source code is not needed to understand that the TrueAllele process is objective.

92. TrueAllele's reliability was established on the evidence in this case. The report and its supporting case packet described the system's sensitivity, specificity and reproducibility on the DNA evidence. The case packet gives the data and parameter inputs used in running the program in the case. The packet also includes a case-specific mini-validation study of reported TrueAllele match statistics, measuring match specificity by comparison with non-contributor genotypes. Source code is not needed to understand or interpret these materials.

93. Additional discovery material for this case was provided on an optical disc. The DVD contains documents related to TrueAllele's reliability, such as background reading, over thirty validation studies and publications, regulatory approvals, general acceptance, and admissibility rulings. There are tutorial videos that describe TrueAllele methods and explain how the system works, as well as continuing legal education talks. The VUIer™ software for reviewing TrueAllele results is provided (with both Windows and Macintosh installers), along with instructions and user manuals. Case-specific files (data, reports, PowerPoint, case packet, VUIer input) are disclosed, enabling a thorough expert review. Source code is not needed to access these materials, read the files, use the executable VUIer software, or examine the computer results.

94. Cybergenetics offers commercial services for validating DNA mixture interpretation methods. Any party can provide DNA validation data and obtain these services to assess

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

TrueAllele reliability. 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. Source code is not needed for obtaining these services.

95. TrueAllele processing is available on-line through Cloud computing. Therefore the system's capability can be operated as an Internet service, without purchasing a product. Any party can operate TrueAllele on the Cloud, and process their own DNA case or validation data. Moreover, Cybergenetics makes this TrueAllele Cloud capability available to opposing parties at no charge so that they can conduct their own testing. Source code is not needed for assessing TrueAllele reliability, which is done by testing the executable program on actual data.

96. Although the source code for TrueAllele is a secret, the methodology it employs and implements has been disclosed. Cybergenetics has published the core mathematics of TrueAllele's underlying mathematical model for 20 years. These publications include scientific papers (1995, 2001, 2009, and 2011) and patent specifications (2000 and 2001). This information discloses TrueAllele's genotype modeling mechanism, and enables others to understand or replicate 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.

97. 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

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955

DNA data signals, the Federal Bureau of Investigation's "PopStats" for producing DNA match statistics or "CODIS" for maintaining a DNA database, and Microsoft "Excel" for conducting additional DNA data analysis. Source code is not needed to assess the reliability of these critical software programs, since they have all been tested and validated.

98. When TrueAllele source code discovery has been requested by an opposing party, no court has ever ultimately required its disclosure. The requesting parties have been unable to show why source code would be material, reasonable, and in the interest of justice. Courts have denied such discovery requests in California, Maryland, New York, Ohio, Pennsylvania, and Virginia, often providing written rulings (*California v. Martell Chubbs*, *New York v. John Wakefield*, *Ohio v. Maurice Shaw*, *Pennsylvania v. Kevin Foley*, *Pennsylvania v. Michael Robinson*). Source code was not needed in any of these cases.

Under penalty of perjury under the laws of the State of Washington, I certify that the foregoing is true and correct to the best of my knowledge and belief.

Signed and dated by me this 1st day of April, 2016, at Pittsburgh, PA.

Mark W. Perlin

Daniel T. Satterberg, Prosecuting Attorney
W554 King County Courthouse
516 Third Avenue
Seattle, Washington 98104
(206) 296-9000, FAX (206) 296-0955