SUPERIOR COURT OF WASHINGTON FOR KING COUNTY

STATE OF WASHINGTON, Plaintiff, vs. EMANUEL FAIR, Defendant.	No. 10-1-09274-5 SEA SECOND DECLARATION OF MARK W. PERLIN in RESPONSE TO DEFENSE MOTION TO COMPEL

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 have read and reviewed the defense Motion to Compel in *State v. Emanuel Fair* and recognize the arguments and claims by Dr. Chakraborty and Dr. Krane as they have been made repeatedly in courts around the nation and have been rejected repeatedly by courts around the nation. My responses to those declarations and the other arguments by defense are stated below.
3. **Casework in State v. Fair**
4. Cybergenetics provided a discovery case packet (DCP) that discloses how the reported DNA match statistics were calculated. The DCP provides case notes; gives DNA data

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

tables and figures; lists the settings used in each computer run; shows the inferred evidence genotypes, reference profiles, and population databases; and tabulates all the DNA match statistics, with table entries giving the numerical strength of association between every inferred evidence genotype (row) and reference profile (column). This DCP was supplemented by a discovery DVD, as detailed in my Declaration. A retained expert may incorrectly read the DCP, and then make unfounded claims based on their incorrect reading, thereby fomenting unnecessary confusion.

5. No source code was used in processing the *Washington v. Fair* case.

6. Testing. TrueAllele testing was conducted and reported properly.

7. Variation. TrueAllele correctly captured the data variation in the case items, reflected as more variation in the match statistics.

8. Competency. The opposition experts incorrectly read the match tables in the DCP. They looked at different contributors to an evidence item, not understanding that those match statistics to a reference profile are expected to be different, not similar. They did not read the instructions, or contact Cybergenetics for assistance.

9. Bias. The paid experts have an interest in not understanding the report or additional discovery materials. TrueAllele accurately and reproducibly resolved the DNA mixtures in this case. Their points of confusion have been previously addressed in other cases where they were involved, but without advancing their knowledge.

10. **Declarations of Defense**

11. Defense counsel makes a number of unfounded assertions. Most of these are based on incorrect statements from his experts, which are refuted in detail later on in this

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

document. However, a few of these mistakes stand out and warrant immediate correction here.

12. Declaration of Counsel #5 - TrueAllele is not "novel". The first mixture interpretation version was developed in 1999. It is not "evolving". The currently used TrueAllele version 25 was completed in 2009, which is the system that has undergone extensive validation testing.

13. Declaration of Counsel #13 - The concept that "without the software source codes of the system, it is impossible" to conduct normative science to assess reliability is just plain wrong. Scientists test their software on data; they do not read source code. There is not one example in a standard forensic DNA laboratory of software being validated with "source code"; the programmer text is irrelevant to testing an executable program.

14. Declaration of Counsel #15 - "Without the source code … any expert will be unable to verify". This is utter nonsense. Source code is never used for validating forensic software (and was not used in conducting over 30 TrueAllele validation studies). There is no way to actually use source code in a validation study, which tests the reliability of an executable computer program. Source code is not available for any of the commercial (GeneMapper®, Excel) or FBI (PopStats, CODIS) software that crime labs rely on; source code is irrelevant for testing the executable program on real data.

15. Declaration of Counsel #17 - The "hundreds of variables" referred to are hierarchical groupings of just a few variables. Hierarchical modeling is explained in published TrueAllele articles and is standard practice in statistical modeling. The objection makes no sense.

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

16. Declaration of Counsel #18, 19 - Validation is the testing of a system on data to assess its performance. The FBI's SWGDAM validation standards (e.g., 2010, 2015) make no reference to source code. Source code is used to develop software; it is not used to test software.

17. **Defense Claims of Error in Testing in *State v. Fair***

18. Defense Motion page 30, line 21 - *The Defense motion discusses quantitative peak data from STR experiments on biological evidence, and TrueAllele's assessment of the electronic information obtained by WSPCL. Examining item Robe-6 at locus D3, they note that the TrueAllele peak table lists 31 potential "alleles", while the crime lab report lists only 5 "alleles" for WSPCL.* The Defense confuses data *input* "alleles" with modeled *output* "alleles". TrueAllele's peak table does not show biological alleles, just all the data peaks that are above 10 RFU for both the Profiler Plus® and COfiler® STR experiments. TrueAllele's genotype modeling assigns probability to allele pairs, not to alleles. TrueAllele statistically explored all possible allele pairs during the genotype separation process. The Defense needs to read the provided documentation, not source code, in order to understand these tables.

19. Defense Motion page 32, line 4 - *The Defense has concerns that they saw different match statistic results, with large variation, which might lead to different conclusions about strength of evidence.* The problem here is that Defense experts are not reading the table correctly. A detailed description of match table rows and columns was provided on page 125 of the case packet. Here are some of the Defense errors in reading the match table.

- Comparing to the wrong evidence genotype contributor.
- Not examining reproducibility within the same contributor assumption.
- Not accounting for natural variation in a low-level 5% contributor.

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

• Ignoring the consistent statistical association of Fair with the DNA evidence.

20. Defense Motion page 32, line 24 - *TrueAllele identified a genotype that appears to be entirely absent from the data.* - TrueAllele uses statistical modeling to reapportion probability based on data. All possible genotypes have some probability before seeing the evidence. The evidence changes these genotype probabilities, increasing it for some genotype values and decreasing it at others. The match statistic is a ratio of the "after" to the "before" probabilities, so no probability can ever be zero. TrueAllele correctly gave a very low probability of under 1% to a genotype possibility 16,23 that had little support in the data.

21. Defense counsel are alarmed at the prospect of standard probability modeling working properly. They insist that probabilities of zero be assigned, which makes no mathematical sense. They need to read a lot more about probability modeling. Reading the TrueAllele source code will not help them in this educational process.

22. **Studies – New South Wales (NSW)**

23. The New South Wales (NSW) crime laboratory located outside Sydney, Australia uses TrueAllele® to resolve complex DNA mixtures that their routine software cannot interpret. The NSW lab has recommended the use of TrueAllele, and has testified alongside Cybergenetics in a quintuple homicide case involving a mixture containing DNA from (up to) five relatives mixed together in a single sample. Cybergenetics provided the Defense with this NSW study as one of 31 TrueAllele validation studies. This Australian study was introduced as evidence of software reliability in a successful Sydney admissibility hearing two years ago.

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

24. Testing. The study design was oriented toward data for older human review methods, where relative amounts of DNA in a mixture have little impact on the match statistic.

25. Variation. With low-level mixtures, TrueAllele correctly ascertained that there was more data variation, which introduces more genotype possibilities, lowering match statistic values and increasing their variability.

26. Competency. The NSW scientists properly operated the TrueAllele software. However, they had a limited concept of statistical variation, and did not fully appreciate how data variation naturally leads to genotype uncertainty.

27. Bias. There was experimenter bias in this study. Statistical software infers answers solely from the data, but the scientists knew how they had created the mixtures. Therefore, they assessed the computer's results based on what they put into the data, instead of what an accurate interpretation would get out of the data.

28. No source code was used in conducting the NSW validation study.

29. **Defense Arguments from NSW Study**

30. Defense Motion page 26, line 10 *[summary] "TrueAllele will assess all peaks as potential contributors to the DNA genotypes and will not disregard or give less weight to apparent artifacts…. TrueAllele found most probable genotype at FGA to be 20.3, 24, even though NSW team knew 20.3 peak was an artifact and true genotype was 24, 24."* The TrueAllele analyst could eliminate obvious artifact peaks from the data, but chose to keep the 20.3 peak in their analysis. TrueAllele's answers are based on the provided data, and are given as probability distributions. The "most probable" genotype is not meaningful, since all possibilities are considered and assigned probability. What a person "knows to be correct" is irrelevant to what can be statistically inferred from data.

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

TrueAllele gave the 24,24 genotype a 0.26 probability (second highest). The NSW team described proper functioning of the TrueAllele system in their validation study, which did not use or require any source code.

31. Defense Motion page 26, line 17 *[summary] "If the artifact overlapped a smaller actual peak, than the effect could have been to falsely include a suspect's genotype when it was actually not present."* Yes, data peaks (real or artifactual) can overlap one another. That is why TrueAllele statistically separates genotypes out of the data, and works with genotypes (not unseparated data) in making comparisons to calculate DNA match statistics. TrueAllele assigns a probability to every possible genotype outcome, and does not jump to incorrect all-or-none conclusions. Source code is not needed to know this.

32. Defense Motion page 27, line 3 *[summary] "Minor components of a mixture at same height as peak stutters, genotype probabilities are not consistent with what would be expected given a reasonable consideration of stutter." TA placed 97.7% on 15 17, even though 15 peak was stutter (true minor was 17 18).* The NSW team confused "what they knew" with "what the data showed". TrueAllele works from the data, and is impervious to outside knowledge of what an answer "should be". That is why the system is objective, accurate and reliable. It gives neutral answers, used by both prosecution and defense, that can either include or exclude a person from DNA evidence. Source code is unnecessary for this understanding.

33. In science, uncertainty is handled by assigning probabilities to possible outcomes. More data variation translates into more diffuse or varied probabilities. TrueAllele accounts for possibilities that human analysts do not consider. The human analysts on the NSW team

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

were striving for definite answers, whereas TrueAllele gives more accurate and appropriate probabilities, based on the observed data.

34. The NSW team inappropriately used the Explain window teaching tool to incorrectly conclude that there was "no attribution of stutter to the 15 peak position". In fact, TrueAllele did assign probability to the different genotype possibilities. Cybergenetics' response to the NSW question is documented in the NSW study report (pages 50-59):

> "TrueAllele does not assign a "probability" to the event that a particular peak is stutter. Rather, the entire data pattern is examined relative a proposed peak pattern, with all relevant variables considered (genotypes, mixture weight, stutter, relative amplification, peak variance, etc.). TrueAllele's stutter modeling found some stutter appearing at this locus, which is why we see genotypes in the posterior distribution that can explain the 15 data peak as (at least in part) a stutter that shadows the large 16 allele peak. Given that the 18 peak's height of 40 rfu is below most human detection thresholds, it is gratifying to see that TrueAllele gave positive probability to allele pairs that included an 18 allele."
>
> "Most solutions here that are "reasonable" to human threshold review ignore the quantitative peak height pattern of the observed data. These qualitative solutions are not all that "reasonable" to quantitative modeling, since the data shows excellent balance between the two major contributor peak heights, and the two "stutter" peaks."
>
> "The Explain window is good for teaching, but not as good for examining the data in great detail. This interface only employs a limited number of variables (genotype, mixture weight), and does not provide a full probability model. Similarly, probability should be used for understanding our inferences about belief, rather the mathematical and unintuitive likelihood construct."

35. **Studies – California Department of Justice (CalDOJ)**

36. The California Department of Justice (CalDOJ) conducted a comparison study between TrueAllele and another software program, STRmix from ESR in Auckland, New Zealand. CalDOJ's goal, as indicated in their sole source procurement document, was to

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

purchase STRmix as an improvement upon their previous PopStats DNA mixture software from the FBI, without changing their process. The CalDOJ mixture interpretation process entails (a) subjective human selection of DNA data, and (b) subjective human determination that a defendant's DNA is in a mixture, before ever (c) calculating a DNA match statistic. Both PopStats and STRmix adhere to this subjective human-centered workflow, whereas the objective TrueAllele process instead calculates a statistic without human data choices. The comparison study was apparently conducted to justify CalDOJ's noncompetitive purchase of STRmix.

37. Testing. The Defense selected only one axis of the study to discuss, low-level three-person mixtures.

38. Variation. Low-level mixtures have more data variation. This data variation induces more variation in genotypes and match statistics. TrueAllele's match numbers correctly reflected this variation.

39. Competency. The CalDOJ scientist crippled the TrueAllele software by changing key parameters. The scientist removed a scientifically determined statistical cutoff constant for the match statistic in order to make a "fair" comparison with the limited STRmix software. He imposed an unwarranted reproducibility tolerance, and did not understand when to reset computer runs. CalDOJ scientists did not complete Cybergenetics TrueAllele operator course, and never learned how to properly solve the more complex DNA mixtures of this study axis. They did not understand what they were doing, nor did they contact anyone knowledgeable for assistance. Their results are invalid.

40. Bias. The key bias in this comparison study was favoring STRmix in order to avoid a fair and open procurement process. The lab altered TrueAllele in nonstandard ways that

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

rendered it inoperable. By comparing STRmix with a damaged TrueAllele, the lab avoided open bidding and could continue with their subjective DNA analysis.

41. No source code was used in conducting the CalDOJ comparison study.

42. **Defense Argument from CalDOJ Comparison Study**

43. Defense Motion page 28, line 19 - *"In its comparison of TrueAllele to STRmix, the California DOJ found that in some situations running pairs of MCMC runs gave identical results."* The TrueAllele system runs parallel computer processes, solving multiple DNA problems simultaneously. Occasionally, one process will overwrite another, giving the appearance of an identical result. The proper procedure is to reset, and run the process again, which resolves the issue. CalDOJ elected to not follow the proper procedure. They did not contact the manufacturer Cybergenetics for assistance. Instead, they blundered on with their invalid study, making unwarranted assumptions and collating meaningless results. Having source code does not eliminate user mistakes.

44. Defense Motion page 29, line 1 - *"DOJ found that on some occasions different MCMC runs resulted in values that were 6.5 quadrillion times different than one another."* This outcome has not been seen in valid operation of TrueAllele in any of the other 30 validation studies. From the provided information, it is impossible to know what the Cal DOJ team did wrong here. The observed variation seems to be a defect in the unskilled operators, not in the technology. Source code does not eliminate user error.

45. Defense Motion page 29, line 2 - *"Many deviations were observed that could lead to different conclusions about the strength of the evidence (e.g., LR << 1 in one interpretation becoming LR >> 1 in another).* With low template DNA data, variation in data is expected to translate into variation in statistical measures, such as the match

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

statistic (LR). And, indeed, the expected large differences like this were observed for the lowest template weights in the study. These answers are correct, with the comparison STRmix software similarly finding uncertainty on these samples, having the LR vary around 1.

46. Unfortunately, the CalDOJ team altered the TrueAllele software in irreparable ways, rendering their validation study invalid. One change they made was to remove the 0.01 lower bound on LR values, implemented to ensure correct Bayesian probability ratios. The Cal DOJ results therefore arbitrarily changed the match statistics to incorrect values, particularly on the low-level mixtures they studied. Source code does not help users follow instructions.

47. Defense Motion page 29, line 13 - *"Heights and allele designations were seen to occasionally change from the values listed prior to upload to the server and those listed after processing in the server"* The Analyze module conducts the bulk of the signal processing, transforming the raw data signals into quantitated allele peak events. The interpretation process provides a Bayesian completion to the signal processing, rectifying possible errors in allele sizing through probability. Asking the software manufacturer would have answered this question, not reading source code.

48. **Opinion – Dr. Ranajit Chakraborty**

49. Professor Chakraborty is a frequent paid opposition expert to TrueAllele evidence. Exhibit A of the Ranajit Chakraborty declaration (RCD) enumerates eleven TrueAllele cases he has reviewed (pages 9-11). Of note, no court has ever agreed with his sworn contrarian views on TrueAllele, which conflict with mainstream science. Specifically:

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

Admissibility. TrueAllele was admitted after defense challenge in (1) *Regina v. Duffy & Shivers*, (2) *Virginia v. Brady*, (5) *New York v. Wakefield*, and (6) *Ohio v. Shaw*.

Source code. Discovery of TrueAllele source code was denied in (3) *Virginia v. Bowman*, (4) *Maryland v. Canela*, (5) *New York v. Wakefield*, (6) *Ohio v. Shaw*, (8) *California v. Chubbs*, and (9) *Pennsylvania v. Robinson*.

Outcome. The defendant was convicted in (1) *Regina v. Shivers*, (2) *Virginia v. Brady*, (3) *Virginia v. Bowman*, (4) *Maryland v. Canela*, (5) *New York v. Wakefield*, (6) *Ohio v. Shaw*, (8) *California v. Chubbs,* (10) *Regina v. Toland,* and (11) *New York v. Frank Thomas*.

50. Chakraborty says that source code is needed to validate TrueAllele. But source code is not used in validation studies. He has not provided the source code for his own commercial software MPkin. His published MPkin validation study has only three example problems, evidently sufficient for his software but not for anyone else. He stated under oath that he published the source code for MPkin (*New York v. Collins*), but then said under oath that he didn't (*Pennsylvania v. Robinson*), and was forced to redefine source code incorrectly in order to avoid further self-contradiction. The *Robinson* court simply ignored his full day of testimony and misstatements.

51. **Response to Chakraborty Declaration (Defense Appendix A)**

52. Page 3, Paragraph 9 - *On March 4th, 2016 I received documents that suggest that a license of the "read-only VUler software" (a component of the TrueAllele system) can be made available with an expiry date of 96 days for which the details of computer infrastructure needed to view the software is not detailed, nor what can be accomplished from it is described. Likewise, the invitation to join the TrueAllele cloud platform to process DNA data using the TrueAllele without having to purchase a system is also equally vague and unspecified with regard to its scope of analyses.* Cybergenetics

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

provided documentation along with the discovery DVD. This details the installation and use of the read-only version of the VUIer software. Cybergenetics extended an invitation to the defense experts in this case to let them to run TrueAllele on their own. They need only respond to the invitation.

53. Page 3, Paragraph 10 - *In absence of the availability of the software source code, several claims made in the court reports on DNA data interpretation using the TrueAllele system as well as those published in the validation studies of the TrueAllele software are impossible to verify. In contrast, some claims are clearly at best misrepresentation of the tasks actually performed by the True Allele system.* Source code is never used in forensic validation. Perhaps reading and understanding the scientific papers would assist this expert in understanding the science. Watching the six hours of lectures (discovery DVD) on what TrueAllele is and how it works might help him as well.

54. Page 3 Paragraph 10 - *For example, it is obvious that in contrary to the claim that the system processes each evidentiary sample objectively (see e.g., see the METHODS section of the Cybergenetics report submitted in this case. dated December 17, 2015), it does not analyze any evidence sample but rather, it re-analyzes the DNA data on the evidentiary items generated by other laboratories (the Washington State Police Laboratory in this case).* Since TrueAllele is software that analyzes DNA data, clearly the sample in this context is the data file generated by a laboratory from biological evidence. There is no possibility of confusion.

55. Page 3, Paragraph 11 - *While this generic description of the TrueAllele system may be correct, the mathematical details of the system, published in the validation studies are very generic and does not give details of several critical features of complex DNA*

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

mixtures such as the ones analyzed in this case. In other words, without the software source codes of the system, it is impossible to verify whether the underlying mathematical models of the system are accurately translated in the source code instructions, or implemented accurately in computations. Source code is used for developing software, not for these fanciful purposes. To learn more about the statistical methods of the TrueAllele system, first read the published scientific papers. Then read the descriptions and equations provided in the TrueAllele Methods document. For additional mathematical details, please read the references that are cited in the Methods document. Scientists assess the accuracy of a system by conducting validations on actual data. They learn about how it works by reading articles. They do not read source code.

56. Page 3, Paragraph 12 - *Not generally accepted practices (e.g., his system uses data part of which may be generated from the artifacts of the DNA amplification process of laboratory analysis of samples, which are generally filtered out by other laboratories invoking the concept of 'thresholds", not used in the TrueAllele system).* TrueAllele has withstood nine admissibility challenges where "general acceptance" was a prong to be established. Its methods are generally accepted in the scientific community. Using all data is what modern statistical computing does, accounting for artifacts through mathematical modeling. Old-fashioned "thresholds" from the late 20th century discard potentially informative data, and have been discredited by the federal government (e.g., the National Institute of Standards and Technology) and many other forensic scientists.

57. Page 4, Paragraph 14 - *Peaks below such threshold heights and the data, which they represent are deemed unreliable and are excluded from the report in which interpretations are made and conclusions reached by the forensic laboratories.* Data is

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

not reliable or unreliable. It is the interpretation of data that can fail, particularly when using older methods that discard important data. Rigorous statistical modeling uses all the data. TrueAllele models laboratory artifacts and baseline noise, determining these variation parameters from the evidence data. In 2010, the FBI's SWGDAM guidelines, paragraph 3.2.2, acknowledged that a lab could forgo using thresholds if they instead used a validated probabilistic genotyping method.

58. Page 4, Paragraph 15 - *In so doing, as input variables, TrueAllele is given all allele peak heights including the ones that fall below the thresholds of the standard operating procedures of the forensic laboratories, but apparently its initial step of analysis (VUIer software) filters out some of the data without any explicit explanation of how is done (that must be present in the software not made available as yet).* By default, TrueAllele uses all the quantitated peaks above 10 relative fluorescent units (RFU). As documented, this baseline convenience cutoff can be lowered, if necessary, to admit more data. The analysis step produces quality-checked quantitated peaks above 10 RFU.

59. Page 4, Paragraph 17 - *Consequently, the use of this data by TrueAllele is a novel and experimental innovation in forensic DNA analysis, which has not gained general acceptance within the scientific community.* TrueAllele data analysis has been around for over twenty years, using most all of the peak height data. Since 1995, there have been 8 published peer-reviewed validation papers establishing the scientific acceptance of its analysis module. The system has withstood 9 admissibility challenges, establishing general acceptance. Moreover, in Commonwealth of Pennsylvania v. Kevin Foley, the Superior Court determined in a published opinion that the TrueAllele method was not novel.

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

60. Page 5, Paragraph 18 - *In this case, the TrueAllele system has been applied to DNA mixtures in compromised evidence samples, whose profiles clearly exhibit lack of clear presence of one or more DNA components of the possible contributors of DNA in these mixtures. This creates another level of complexity of DNA mixture that was not adequately presented to the New York State DNA Subcommission at the time of seeking approval for the TrueAllele system from the Subcommittee. In addition, neither the allele degradation model in the TrueAllele software, nor the incorporation of allele drop-in is explicitly explained in any of the publications or operating procedures of the TrueAllele system.* At a 2011 New York DNA subcommittee meeting, this defense expert seconded a recommendation for using TrueAllele in forensic casework (without restriction), approved unanimously based on extensive validation studies. These studies included complex DNA mixture samples, some of which were low-level or contained three people, and have been provided here in Discovery. Please refer to the TrueAllele Methods document for a more detailed description of DNA degradation and drop-in modeling.

61. Page 5, Paragraph 19 - *There have been suggestions that TA is 'black box'. Currently while the mathematics for key variables such as mixture weight amplification variance, and baseline variance have been disclosed in publications, the handling of other parameters such as stutter, relative amplification of alleles at a locus, and DNA degradation are not disclosed. This makes it difficult to determine how TA handles these issues.* Please refer to the TrueAllele Methods document for a mathematical description of stutter, relative amplification, and DNA degradation.

62. Page 5, Paragraph 20 - *For example, the equation (5) of this document clearly illustrates that the TA system assumes independence of allele peak heights at a locus to model*

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

distributions of peak height variance and baseline variance. However, the lack of exposition of modeling of variables such as the PCR stutter, relative amplification, DNA degradation, and dye separation and their mathematical implementation, noted in the NSW report, illustrates the "Black Box" nature of the TA analysis, which could have been deciphered through the analysis of the specific instructions in the source code of the software. Please refer to the TrueAllele Methods document for a mathematical description of stutter, relative amplification, and DNA degradation. (Dye separation is done in the laboratory by genetic analyzer equipment and software, prior to TrueAllele processing.) The source code is irrelevant to these mathematical descriptions.

63. Page 5, Paragraph 21 - *With ample of specific examples, this report illustrates that the comfort region of unambiguous genotype inference of contributors by the TA analysis is rather narrow, and inapplicable for complex DNA mixtures, in terms of the parameter space of describing a complex DNA mixture not recognized in any of the publications or reports on the TA system produced by Cybergenetics.* Perhaps scientific probability modeling is beyond the "comfort region" of this defense expert, who yearns for the "unambiguous genotype inference" of his youth that is unattainable in nature and science. Modern scientists model uncertainty to get answers, including the calculation of accurate DNA match statistics for complex DNA data. The NSW laboratory does not share his antiquated concerns, since they currently use TrueAllele to resolve their complex DNA mixtures.

64. Page 6, Paragraph 22 - *I may note that sensitivity as well as specificity of TA inference in the cases of complex DNA mixtures in the experiments done by CAL-DOJ were of far less acceptable quality than the ones reported in the Cybergenetics publications. In*

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

particular, both sensitivity and specificity substantially diminished with increase in number of contributors (even from 2-person mixture to 3-person mixtures), with skewed mixture weights of contributors, and with DNA degradation. In the CAL-DOJ study, the scientists altered the TrueAllele software by changing the usual settings to unworkable values. Their results from altered software are meaningless. There have been 31 TrueAllele validation studies done by trained operators using correct parameters. The CalDOJ study is not one of them.

65. Page 6, Paragraph 23 - *These are examples of evidence that I would use now to assert that the TrueAllele system has failed to gain general acceptance in the scientific community and it has not been adequately validated for the type of caseworks it is now being applied.* This defense expert can only cite a failed California study that was done improperly by untrained operators, and his slanted misreading of an Australian study that established TrueAllele reliability for both the Sydney DNA laboratory and court system. He ignores the usual indicia of general acceptance – over 30 TrueAllele validation studies (7 of them published), and withstanding 9 admissibility challenges. Moreover, mainstream scientists ignore Chakraborty's paid opinions on interpreting complex DNA evidence. On this topic, Chakraborty has never written a scientific paper, conducted a validation study, done original research, programmed a computer, understood the statistical computing machinery, developed a scientific result, or presented any of his wild assertions at a scientific meeting. His main DNA mixture interpretation legacy is the failed combined probability of inclusion (CPI) mixture statistic method, which he helped develop 20 years ago and is now discredited.

66. Page 6, Paragraph 24 - *In addition, as of present, a great majority of these laboratories*

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

are yet to produce DNA evidence in criminal trials on their own, based on TrueAllele-based analyses of DNA mixture data. Currently, seven crime laboratories are using TrueAllele in their own independent casework. Forensic DNA analysts in California, Maryland, South Carolina and Virginia prepare TrueAllele case reports and testify in court.

67. Page 7, Paragraph 26 - *Further, Dr. Perlin has refused to reveal the full details concerning the input and output data of the applications of his software.* These details are transparently provided in the discovery case packet, available for any competent lawyer or scientist to read on their own. Hundreds of trained users can operate the TrueAllele system, inputting their data and outputting accurate DNA match results. Chakraborty is welcome to use the TrueAllele Cloud system on his own data, and join the ranks of these informed scientists.

68. Page 7, Paragraph 27 - *TrueAllele represents complicated technology providing novel scientific evidence whose general acceptance remains questionable.* Yes, this 21st century technology may be too complicated for Chakraborty. However, TrueAllele has withstood 9 admissibility challenges and has proven to be generally accepted in the relevant scientific community.

69. Page 7, Paragraph 29 - *There is no documentation of any validation study for evolution of revisions of the system by scientists working independently of the Cybergenetics Corporation.* The current version 25 was released early in 2009. There have been 7 published validation studies done on this final version. Moreover, independent groups have completed 10 validation studies on the current version of TrueAllele. This information is provided in the discovery material in possession of the Defense that gives

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

PDF-readable reports on 31 validation studies.

70. Page 7, Paragraph 30 - *Until such time as TrueAllele reveals its source code, the flow charts for the use of its system of equations and its input /output data for its software, it can only be considered "a work in progress".* By this absurd logic, virtually all software used by society is a "work in progress." The source code, flow charts, etc. are not available for Microsoft Excel, the Google search engine, the iPhone, automobile computer systems, aircraft navigation systems, and the other 99% of software we (and crime laboratories) rely on in our everyday lives.

71. Page 7, Paragraph 31 - *It [the report] makes ambiguous and misleading statements (such as 'The TrueAllele® system processed each evidence item ...") and improperly uses the term "match" when comparing a mixture DNA profile to the DNA profile of a single individual.* A probability of match is not a "match." A ratio of probabilities of match is not a "match." Match probabilities have been used in science and law for over 20 years. Chakraborty can learn more about the mathematics of match probability by reading the provided Likelihood Ratio Application Note.

72. Page 7, Paragraph 32 - *TrueAllele cannot be meaningfully validated [without source code, flow charts, etc.]. Consequently, without any such meaningful validation, it remains novel and experimental and it has not been generally accepted within the scientific community.* Merely repeating an incorrect statement many times does not make it true. Validation is done by testing software on data. Source code is not used. Chakraborty is mistaken, and seems to lack relevant expertise in this area.

73. Page 8, Paragraph 35 - *Neither a review of these validation studies, most of which were performed by Cybergenetics itself, nor a "walk through" of the program is an adequate*

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

substitute for the revelation of the source code itself, as a way of validating TrueAllele. Again, source code is not used for validating software. The source code is human readable text. The executable program is a machine-readable program that inputs data and outputs results. Testing is done by inputting data into the executable program, and observing the output. This has nothing to do with source code text that a computer cannot execute, does not input data, and does not output results. If Chakraborty, or any competent scientist, wanted to test TrueAllele, they would enter their data into the TrueAllele executable program and observe the results. Cybergenetics has offered this testing option to Chakraborty in the past, and again now. He continues to decline this normative approach to testing reliability.

74. Page 8, Paragraph 36 - *Graphic results of the VUler software have many alleles missing that are apparently included in data submitted to the computer for runs of this sample for TrueAllele analysis. Clearly, this is discordant with the claim that TA used all data in interpreting the DNA mixture of the Robe 6 evidentiary item.* The VUIer data images provided in the Case Packet dated December 30, 2015 show the data in a display image where peaks are labeled at a peak height cutoff label indicated inside the image. This peak height cutoff is for display purposes only, and does not affect the TrueAllele interpretation. Moreover, this display cutoff is fully described in the VUIer manuals that have been provided.

75. Page 8, Paragraph 37 - *Need further discovery data … request #5 of the discovery demand made on February 3, 2016, and request #3 of the discovery demand of February 11, 2016).* Point #5 of the 2/3/16 discovery request refers to the Analyze rules. These rules are described in the Analyze manual, found on the DVD in the file: 5-VUIer > 2-

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

Manuals > 03-Analyze.pdf. Point #3 of the 2/11/16 discovery request is addressed by the provided TrueAllele Methods document.

76. **Opinion – Dr. Daniel Krane**

77. Professor Krane is a regular paid opposition expert to TrueAllele evidence. He testified for the defense in the *Ohio v. Maurice Shaw* Daubert hearing, where TrueAllele was admitted into evidence, source code discovery was denied, and the defendant ultimately pleaded guilty to the homicide.

78. Krane's Bioforensics company has long marketed its "Genophiler" software for forensic use, and (like most companies) protects its proprietary source code as an undisclosed trade secret.

79. In his declaration, Krane appears to reject conventional testing of software on real data to validate its reliability. Yet in his own academic and commercial endeavors, Krane conducts exactly that sort of generally accepted testing, running software on data to reach conclusions. Nowhere in his published scientific studies does he discuss using source code. Rather, like normal scientists, he runs software on data to calculate results.

80. Krane seems greatly alarmed in his declaration at variation appearing in data analysis. Yet in his published scientific studies, he accepts such variation as normative. His papers abound with statistical measures that mathematically describe the variation he observes when running software on data. He uses standard deviations, p-values, percentiles, bootstrap methods, z-scores and correlations – all measures of variation – in his scientific papers. Those papers are the basis of his academic qualifications. Yet he decries the very variation he so assiduously studies in his scientific day job when opining on forensic

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

statistical analysis. If Krane is comfortable with variation in science, he should be equally comfortable with variation in forensics.

81. **Response to Krane Declaration (Defense Appendix B)**

82. Page 2, Paragraph 3 - *Dr. Perlin has not provided any validation studies or published, third-party reviews of the hundreds of variables or their sub-models and their associated uncertainties, boundaries and interrelationships that constitute the underlying probability model of TrueAllele. . . It is important to know how the value of each variable and uncertainty is determined, how those values affect TA analysis, and if alternative implementation of this method would deliver a different, reasonable explanation of the data being evaluated.* As described in published TrueAllele papers, and summarized in the TrueAllele Methods document, modern statistical software uses hierarchical modeling. Therefore, one variable (e.g., mixture weight) can manifest itself as 16 variables, 1 for the DNA template, plus 1 for each of the 15 STR locus experiments. To wonder at "hundreds of variables" when it is clear that there are just a few main variables (e.g., mixture stutter, stutter) with many groupings is disingenuous. To know how variables and their uncertainty are determined, Krane is referred the published TrueAllele papers and the TrueAllele Methods document.

83. Page 3, Paragraph 4 - *Best way to evaluate TA's probability model is through a review of its source code. . .Source code is the precise, yet human-readable description of the sequence, branches, and loops of computer instructions that constitute a computer program. . . Source code directly analogous to laboratory SOPs.. . Difficult to find DNA profiling expert who felt that a list of the approaches used to generate a DNA profile could possibly take the place of a review of the labs SOPs that describe implementation*

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

of that method….It would be difficult to find a computer scientist that accepted that a computer program's reliability could be assessed without having access to the source code. This is specious reasoning, based on a false analogy between manual human procedures and automated computer processes. And so by this logic, Krane leads us to a false conclusion, that software can only be assessed with access to source code. But, in fact, that doesn't happen in the real world, because commercial software (i.e., no source code provided) represents virtually all of the software that society relies on. Krane would have us believe that Consumer Reports does not assess the reliability of automobiles, since without access to the manufacturer's software source code (undisclosed trade secrets), their cars cannot be assessed.

84. Page 4, Paragraph 5 - *TA produces LRs and cannot be validated with only black box testing because the correct answer cannot be known (and therefore cannot be compared to results generated by program)….From Steele and Balding paper: Measuring a quantity can be validated by showing the measured value consistently lies within an acceptable range or error relative to true value. Such a validation is infeasible for LR computing software because LR has no underlying true value – expresses uncertainty about an unknown event and depends on modeling assumptions that cannot be precisely verified in context of crime scene data.* Not true. TrueAllele accuracy was demonstrated on an ensemble of a hundred DNA match statistics (see the Virginia TrueAllele validation paper published in *PLoS ONE* in 2014). Krane's logic leads us to the wrong idea that match statistics are unknowable, and can never be validated. This flies in the face of over 30 TrueAllele validation studies, and the 2015 SWGDAM guidelines that describe validation requirements.

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

85. Page 5, Paragraph 6 - TrueAllele has been validated on samples of known composition. The genotypes in these data are known and can be compared to the TrueAllele separated genotypes to see if TrueAllele is giving accurate answers without using the LR. Also, running TrueAllele on the same data multiple times demonstrates reproducibility (genotypes, match statistics, etc.).

86. Page 5, Paragraph 7 - *The article Case for Open Computer programs summarizes need for review of source code: With some exceptions, anything less than release of actual source code is indefensible approach for any scientific results that depend on computation – because not releasing the code raises roadblocks to reproducibility.* An Open University professor who promotes open access software is certainly entitled to his opinion. His academic view is wrong here, because testing software on real data can demonstrate reproducibility. In the real world, outside the university, most software we regularly use is extensively tested and its source code is not disclosed.

87. Page 6, Paragraph 8 - *Steele and Balding – Some progress can be made in evaluating the validity and performance of software – courts need evaluations to have confidence in results – open source highly desirable in court because openness to scrutiny by any party is invaluable source.* Academics like Krane, Steele and Balding can be passionate believers in open source software, based on their university experience. They write short academic computer programs, often not more than a few hundred lines long, of limited functionality and little commercial value. Their colleagues help ferret out software bugs. Academic software is less prevalent out in the real world. The commercial development model usually engenders funding for extensive advance testing before software is released, while academic programs lacking such funding are not as thoroughly tested.

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

Academic professors are entitled to their opinions, but (as in this situation) those views may have no applicability in the real world, or to the forensic crime laboratory.

88. Page 6, Paragraph 9 - *ISFG – no black box approach – open source strongly encouraged since it offers unrestricted peer review and best assurance that methods are fit for purpose.* The TrueAllele method has been independently peer-reviewed in several publications. Some authors of this ISFG paper are open-source advocates, whose academic business model (government grant funding, limited software prototypes) may depend on open source as a requirement of their university funding. Their business model is unrelated to the reliability needs of forensic scientists.

89. Page 7, Paragraph 10 - *Others have noted significant concerns about TA without access to source code. CALDOJ – precision – identical results – large difference between runs. Seems to apply to TA results in this case, since some reported LRs are very different from others for the very same test results of individual samples.* Krane can only cite a failed validation study that wrecked the TrueAllele software before testing it. He ignores 31 other studies conducted by competent scientists whose goal was to test software for its reliability, not to provide a rationale for a closed bid.

90. Page 8, Paragraph 11 - *NSW – lack of consistency in stutter modeling.* This has been addressed at length in the "NSW" section above.

91. Page 9, Paragraph 12 - *Review of source code would allow the following:*

    1. *determination of what computations were performed*
    2. *determination of scientific accuracy of computations by*
        a. *evaluating whether computations performed and conclusions are consistent with the published claims of Dr. Perlin*
        b. *evaluating whether the computations and conclusions are consistent with generally accepted principles that are routinely employed by human experts*
    3. *determination if these methods were properly translated from concept to source code and no mistakes were made during writing of source code*

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

4. *determination of whether alternative explanations of observed data could have produced similar results to those produced by TA*

Professor Krane makes our point. As he says, review of source code would enable the reverse engineering of the TrueAllele technology, allowing others to learn the trade secrets that keep Cybergenetics solvent. If Cybergenetics wants to stay in business, and continue providing objective and accurate DNA identification to all parties in criminal justice, it cannot disclose its source code. Also note that TrueAllele's accuracy has been established in peer-reviewed and other studies. Moreover, these points can be evaluated by testing the system on DNA data of known composition. The core mathematics that underlies TrueAllele's internal calculations is disclosed in scientific papers and the TrueAllele Methods document.

92. Page 10, Paragraph 13 - *Important to know precisely why and how TA arrived at the results in this case. Careful evaluation of computational steps taken would allow it to be determined if the program:*

1. *reflects what is described by Dr. Perlin*
2. *consistent with practices of forensic DNA profiling community*
3. *free from bugs and errors*
4. *if TA can and does provide sufficient explanations for observed data in this case*

The many validation studies reflect how well the system performs relative to its design. There has been no suggestion that TrueAllele has errors. Rather, a "careful evaluation of computations steps" would not be possible for 170,000 lines of code within any reasonable time frame, but it would enable reverse engineering of protected trade secrets.

93. Page 11, Paragraph 14 - *Important to evaluate the data generated by TA during course of analysis and to evaluate how it progressed through the program. Preserved data serves as snapshot of state of TA at each point during analysis. Intermediate data especially*

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

important in the evaluation of algorithms such as Metropolis-Hastings MCMC – important to know the intermediate results and how they vary and the methods that are used to hone in on final result to have confidence in conclusions. Variance of results produced can be affected by starting value used for MC, acceptance ratio, # cycles performed, # cycles whose data were discarded, any fixed values that serve as starting points and other factors that are not described in Dr. Perlin's publications. There are no intermediate data or files generated by TrueAllele during its interpretation process. The results are seen in the output in the VUIer software. The prior probabilities are available in peer-reviewed publications, as well as in the TrueAllele Methods document.

94. Page 12, Paragraph 15 - *Do not know if self-checks TA did that resulted in TA performing corrections on its own data or avoiding potential computational pathways. Human experts expected to explain how they arrive at a conclusion using alternative approaches when preferred analysis fails. Same expectation can and should apply to a computer program.* Human TrueAllele experts explain at how they arrive at their conclusions using TrueAllele. Unlike experts in older methods where only a few alternatives are entertained, TrueAllele can consider 100,000's of possible alternatives, and summarize them using probability. The TrueAllele expert can explain the input, process, and results to a court, on direct and cross examination.

95. Page 12, Paragraph 16 - *Review of source code would help in understanding TA report by allowing a flow chart that outlines what operations were performed and in what order as TA evaluated input. Flow chart at present time is 'TA given input data humans felt too complicated for conventional interpretation. TA evaluated data. TA arrived at conclusion regarding data'. Human expert using that approach to explain arriving at conclusion*

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

would not be considered credible. Not possible to assess or confront TA's conclusions without particularized understanding of analysis it performs. There is no flow chart. Statistical sampling just repeatedly considers the different random variables to propose alternatives and evaluate their relative probabilities. A human TrueAllele expert can understand and explain the objective computer process. Materials needed for understanding TrueAllele analysis have been disclosed to the defense. It is up to their experts to expend the requisite time and effort to learn the material.

96. Page 12, Paragraph 17 - *Scientific principles dictate that new manners of applying methodologies should be made available for outside review and confirmation before they are relied upon. Essential that TA model and specific implementation of model are carefully evaluated before relied upon in this case.* TrueAllele's methodology has been described in many peer-reviewed publications. Cybergenetics and other groups have conducted over 30 TrueAllele validation studies. Defense experts can test the TrueAllele system on their own data, independently of the company and at no charge.

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 3rd day of April, 2016, at New York.

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