

Discovery for complex DNA cases

**Carrie Wood, Appellate Attorney, Hamilton County, Public Defender's Office (Cincinnati, OH)**

Crossing the State's Expert in Complex DNA cases (mixtures, drop out and inconclusive results)

**Carrie Wood (Cincinnati, OH) and Nathan Adams (Fairborn, OH)**

## Pre-trial Litigation: Discovery, Admissibility Challenges, and Practice Tips

Carrie Wood

NACDL Pennsylvania Training Conference, February 23, 2018 – 90 Minutes

This presentation will address how a lawyer obtains and identifies the appropriate discovery for each of the areas outlined by Nathan Adams in his talk. This talk will assist attendees to use discovery to identify potential unfounded statements and/or unreliable technology, as well as flag areas where an expert for the State may oversimplify, misapply, or mischaracterize based on the discovery materials and investigation. Finally, for each identified area of concern, the talk will present pre-trial litigation strategy, law, and motions as well as practice points for cross-examination.

1. Exam/inspection for samples of interest (including serology)
2. Extraction
3. Quantification/quantitation
4. Amplification (PCR)
5. CE injection (genetic analyzer)
6. Analysis by human
7. Analysis by probabilistic genotyping software (maybe)

### Getting Started

- Discovery: One of the most critical steps in successfully challenging scientific evidence and in particular DNA evidence is obtaining the necessary discovery. It is impossible to evaluate the DNA evidence or assess the strength of it without obtaining discovery. A basic DNA report gives very little information and does not include the type of information that is necessary to effectively evaluate the DNA evidence in a particular case.
- How do I obtain discovery?
- What to include in discovery request
  - Non-cold hit discovery
  - Additional cold hit information
  - Interpretation discovery (validation studies and protocols)
  - Probabilistic Genotyping discovery
- What do I get?
  - Initial report and getting organized

- What do I do when the State won't give me what I ask for?
- Other discovery
  - Proficiency testing, log of corrective action, etc.

#### Serology issues:

- Discovery:
  - Case file, photos, protocols, etc.
- Semen and sperm
  - Discovery example
  - Pre-trial litigation:
    - Interview your lab analyst (or find a transcript)
    - Keeping out uninformative findings
      - Motion in Limine: unacceptable language
      - *Daubert/Frye* (e.g. time of deposit or “looks like it might be a sperm head but I cannot say for sure”).
        - Your own expert in a *Daubert* challenge
    - Limitations of test and conclusions
      - Motion in Limine
    - Limiting State's argument about presumptive tests or problematic findings
  - Cross-examination practice point
    - Prepare, prepare, prepare....
    - Know your expert! Know the literature!

#### Extraction issues:

- Discovery example
- Motion in Limine
  - unacceptable descriptors or attribution of differential extraction results
  - limiting state's argument about what the test can tell you
- Cross-examination
  - limitations of the test
  - theory of defense

#### Quantitation/Amplification issues:

- Quantification discovery example

- Quantification
- Stochastic effects
- Reliable amount of DNA?
  - Usually one part of larger pre-trial reliability challenge (see pre-trial challenge to reliability below)
  - Practice points: Cross-examination
- Transfer
  - Theory of transfer and cross-examination
    - Frequency of touch
    - Direct
    - Indirect
    - Secondary, Tertiary, etc.
  - Practice points: Cross-examination

Analysis/Interpretation:

- Discovery needs
  - Analytical thresholds and noise protocols
  - Stochastic threshold validation studies
  - What information did your analyst have when they marked e-gram?
- Discovery examples
  - Major-minor interpretation discovery
  - Mixture interpretation discovery
- *Daubert* challenge: Complex Mixture Interpretation
  - Subjective Interpretation Problem:
    - Number of contributors: Allele stacking and stutter (may also include reliability concerns from quantification)
  - CPI's application to complex mixture
    - What alleles count? i.e. When is CPI accurate?
      - Stochastic threshold/Drop out/stacking problem
      - Number of contributors
      - Does the analyst have the suspect's profile?
      - Practice points: box them in on contributors
  - Why does Complex Mixture Interpretation Fall Under *Daubert*?

- PCAST, SWGDAM, and Mixture Studies
    - What about likelihood ratios?
  - Practice points: Give the court a visual aid!
- Probabilistic genotyping: How do you solve a problem like complex mixture interpretation?
  - Discovery: there is a lot more information if State calls in PG
  - First litigation issue: Software quality re: probabilistic genotyping
    - Source Code Discovery, your first fight: litigation for access to source code
      - Discovery rules/Due Process/Confrontation
      - Trade secrets and protective orders
    - Get a hearing!
      - Cross examination
      - Practice points: government versus private entity? Can you prove fact in another way? (e.g. where in validation range is your case? )
  - Second litigation issue: *Daubert* and Evid.R. 702
    - *Daubert* and Evid.R. 702 (when you get access to source code):
      - Is the State's witness an appropriate/qualified expert?
      - Did lab follow their own protocols (FST)?
    - *Daubert* challenge without the source code:
      - Unspecified methods are not sufficient (cannot meet *Daubert* standard when source code is not disclosed)
      - Validation studies – are they enough? Who conducted the study?
        - PCAST
      - Still problems of subjectivity with input into software, e.g. number of contributors
      - Likelihood ratios and low template samples – still problems
        - “Likelihood ratio” statistical framework
          - Assumptions made
          - Qualitative verbal scales
        - Does analyst understand statistics? What language will they use? What do you want to prevent? (motion in limine and reliability)

- Inclusion/exclusion criteria for comparisons to suspect/victim DNA profiles
- *Daubert* problems are cross-examination for trial too!
  - Pick themes for voir dire to help jurors pick up on your cross:
    - E.g.'s likelihood not fact (based on software): anyone have problems when they relied on technology? Potential bias if you know the answer before you try to solve the puzzle.

Q&A, time permitting

Forensic DNA achieved the vaunted status of “gold standard” among the forensic disciplines based on the its objective and well-defined methods for testing and interpreting high quality, single source DNA samples, such as bloodstains, and simple, distinguishable,<sup>1</sup> two-person mixtures. These methods are backed by robust validation studies.<sup>2</sup> Many of the DNA analyses encountered in casework today are of this type, where validated protocols exist for reliable testing and interpretation. However, there has recently been a move toward processing increasingly marginal samples, and in particular, complex DNA mixtures. Complex DNA mixtures are typically samples comprised of DNA from three or more individuals in unknown proportions.<sup>3</sup> In addition, complex DNA mixtures often involve one or more low-level DNA contributors that may be in varying states of degradation and in which alleles may be stacked making interpretation difficult and often unreliable.

Validation studies and protocols for interpretation of high quality, single source DNA samples and simple mixtures are woefully inadequate when applied to complex DNA mixtures, necessarily leaving the interpretation of the DNA result to the analyst’s subjective judgment. A recent large-scale interlaboratory study conducted by the National Institute of Standards and Technology (NIST) laid bare exactly how unreliable the subjective interpretation of complex mixtures by forensic DNA analysts can be: the vast majority of participants in the study—from a cross section of local, state and federal DNA laboratories—falsely included an innocent individual in a complex DNA mixture provided by NIST.

---

<sup>1</sup> A distinguishable mixture is one in which the DNA donors contribute significantly different amounts of DNA, allowing their individual DNA profiles to be isolated, or distinguished, into a major and minor profiles. *See infra* at fn 15.

<sup>2</sup> A validation study is a test that establishes the reliability of a method (repeatability, reproducibility, and accuracy) and the parameters (e.g. amount of DNA, quality of DNA, number of DNA contributors) within which the method functions reliably.

<sup>3</sup> It is often difficult to determine the number of contributors in a mixture, and a two person mixture may also be considered complex if the components cannot be distinguished.

Less than a year ago, at the behest of President Obama, a group of the Nation’s leading scientists issued a report assessing the scientific validity of a number of forensic feature comparison methods, including complex DNA mixture interpretation and application of the associated mixture statistic known as the combined probability of inclusion (CPI). These scientists unequivocally concluded that “interpretation of complex DNA mixtures with the CPI statistic has been an inadequately specified—and thus inappropriately subjective—method. As such, the method is clearly not foundationally valid” and “has the potential to lead to erroneous results.” President’s Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts: Ensuring Validity of Feature-Comparison Methods*, at 78, 82 (Sept. 20, 2016) [hereafter PCAST Report].

Because the hallmark of legal admissibility is scientific validity and reliability, courts must exclude testimony regarding complex mixture interpretation and application of the combined probability of inclusion (CPI) statistic to such mixtures.<sup>4</sup> Exclusion is the only appropriate remedy in this case, as cross-examination cannot prevent the prejudice from the admission of a DNA inclusion result derived from a method that has not been proven valid or reliable. The significance of expert testimony at trial cannot be overstated. Scientific expert testimony carries with it the “aura of special reliability and trustworthiness,” creating a grave risk that jurors will receive it without a critical eye. *United States v. Dowling*, 753 F.2d 1224, 1236 (3d Cir. 1985); *see also United States v. Haines*, 803 F.3d 713, 730 (5th Cir. 2015) (recognizing the significance of expert testimony to juries); *People v. Kelly*, 17 Cal.3d 24, 31 (1976) (“Lay

---

<sup>4</sup> The collective experience of examiners in the field and in courts, no matter how extensive, is not a scientific basis to demonstrate validity and reliability. Claims to the contrary have been soundly rejected by the scientific community. “Nothing – not training, personal experience nor professional practices – can substitute for adequate empirical demonstration of accuracy.” PCAST Report p. 46.



jurors tend to give considerable weight to scientific evidence when presented by experts with impressive credentials.”). Perhaps because juries view forensic testimony with unflinching trust, the use of unreliable forensic science is one of the leading causes of wrongful convictions. Brandon L. Garrett, *Judging Innocence*, 108 Colum. L. Rev. 55, 83-84 (2008). This risk is exponentially magnified when it comes to DNA evidence, which is considered the “gold standard” of forensics and has an outsized impact on juror perceptions of guilt even when it is of ambiguous or low probative value.<sup>5</sup> This Court must act as a gate-keeper to protect against jurors’ uncritical reliance on DNA evidence.

For Daubert Jurisdictions: As the arbiters of evidentiary admissibility, *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 588-89 (1993); *Nations v. State*, 944 S.W.2d 795, 797 (Tex. Ct. App. 1997), trial judges must exercise caution and only admit scientific evidence if the proponent of the testimony shows it is both reliable and relevant to a contested issue. *Daubert*, 509 U.S. at 589; *Kelly v. State*, 824 S.W.2d. 568, 572 (Tex. Crim. App. 1992). “Unreliable . . . scientific evidence simply will not assist the [jury] to understand the evidence or accurately determine a fact in issue; such evidence obfuscates rather than leads to an ‘intelligent evaluation of the facts.’” *Kelly*, 824 S.W.2d at 572 (quoting Kenneth R. Kreiling, *Scientific Evidence: Toward Providing the Lay Trier with the Comprehensible and Reliable Evidence Necessary to Meet the Goals of the Rules of Evidence*, 32 Ariz. L. Rev. 915, 941-42 (1990)). Because there is no empirical evidence that complex DNA mixture interpretation and application of the CPI statistic to such mixtures is reliable, this Court should exclude it.

---

<sup>5</sup> See, e.g., L. Smith et al, Understanding Juror Perceptions of Forensic Evidence: Investigating the Impact of Case Context on Perceptions of Forensic Evidence Strength, *J Forensic Sci.*, 56(2), 409 (2011)(“the perceived strength of [DNA] evidence was significantly inflated when presented in the context of a criminal case, particularly when the evidence was of a weak or ambiguous standard”); N.J. Schweitzer and Michael J. Saks, *The Gatekeeper Effect: The Impact of Judges’ Admissibility Decisions on the Persuasiveness of Expert Testimony*, 15 Psychol., Pub. Pol’y & L. 1 (2009).

For California and Other *Frye* Jurisdictions: The *Kelly-Frye* rule governing admissibility is conservative in nature, designed to prevent the admission of evidence that is not generally accepted as reliable by the scientific community. Importantly, judges must apply particular scrutiny when the “identification technique is offered to identify the perpetrator of the crime,” as is often the case when DNA evidence is offered as evidence. *People v. Kelly*, 17 Cal.3d at 32. “When identification is chiefly founded upon an opinion, which is derived from utilization of an unproven process or technique, the court must be particularly careful to scrutinize the general acceptance of the technique.” *Id.* Complex DNA mixture interpretations and associated CPI statistics are not generally accepted by the scientific community because there is an absence of scientific evidence demonstrating its scientific validity. This Court should therefore exclude this evidence.

If this Court will not grant this motion on the papers, the defense requests the Court hold a *Daubert* [*Kelly-Frye*] hearing where the government must prove, by clear and convincing evidence (preponderance of the evidence), that the proffered testimony meets the requirements of Texas Rule of Evidence 702. *Kelly v. State*, 824 S.W.2d 568, 573 (Tex. Crim. App. 1992) [general acceptance.]. Failure to do so will violate Mr. XXXX’s evidentiary rights, his constitutional right to due process of law, and, in this most serious death penalty case, his Eighth Amendment Right against cruel and unusual punishment.

### **ARGUMENT**

In sections I and II, undersigned counsel provides the proffered testimony in the instant case, followed by background on the DNA testing process, including the difficulties and ambiguities associated with complex mixture interpretation and discussion of the only large scale interlaboratory study into the reliability of complex mixture interpretations (Section II.D).

Section III addresses the legal standard for admission of scientific evidence, and it sets forth why this Court should regard the PCAST report as an authoritative voice in the relevant scientific community. It then details the PCAST Report’s unequivocal conclusion that complex DNA mixture interpretation with the CPI statistic is not scientifically valid or reliable.

## **I. Statement of Facts**

Mr. XXXX is charged with XXXX on XXXX. The government seeks to introduce the expert testimony of XXXXX, a DNA analyst from XXXX laboratory, regarding a DNA mixture recovered from a swabbing [OR cutting, etc] of XXXX [ITEM OF EVIDENCE]. The testimony will involve [LAB ANALYST]’s comparison of the DNA profile developed from this mixture—which [LAB ANALYST] has determined to be a mixture of [NUMBER] people—to a DNA profile developed from a reference sample collected from Mr. XXXX, as well as her statistical analysis related to that comparison. According to the government, [LAB ANALYST] will testify that Mr. XXXX cannot be excluded as a contributor to this mixture, and that the combined probability that a randomly selected individual would be included in the mixture (“CPI statistic”) is 1 in XXXXXX. Exhibit A at [page], DNA Analysis Report.

## **II. Background on DNA Testing Process and Interpretation of DNA Mixtures**

### **A. Basics of the forensic DNA testing process**

Deoxyribonucleic acid, or DNA, is a double-stranded molecule that coils to form the characteristic double helix, and is found in all cells possessing a nucleus.<sup>6</sup> John Butler,

---

<sup>6</sup> Most cells, with the exception of red blood cells, possess nuclei. When it is in the nucleus, DNA is tightly packaged into two sets of 23 chromosomes; one set of 23 chromosomes is inherited from each parent. Sperm and egg cells possess only one set of 23 chromosomes each; when they unite, the resulting embryo possesses the full set of 46 chromosomes. *Fundamentals* at 23.

*Fundamentals of Forensic DNA Typing* (“*Fundamentals*”), 19 (2010).<sup>7</sup> Forensic DNA typing examines certain locations, or loci (singular: locus), on the DNA strand. The DNA typing technique at issue in this case is short tandem repeat (STR) testing. STR typing measures how many times a short piece of DNA repeats at each of the tested loci; the number of repeats is known as an allele. *Id.* at 148. An individual’s genetic type, or profile, is the compilation of his or her alleles at each locus tested. At each locus, an individual possesses two alleles: one allele inherited from each biological parent. *Id.* at 25. Thus, an individual’s DNA profile is simply a list of two numbers per locus examined.<sup>8</sup> An individual can inherit the same allele—*i.e.* same number of repeats—at a locus from both their biological parents (*e.g.* 12, 12). This means the individual is a **homozygote** at that locus. Alternatively, an individual can inherit two different alleles—two different numbers—at a locus from their parents (*e.g.* 12, 16). This means they are a **heterozygote** at that location.

The DNA testing process proceeds via a series of steps: extraction, quantitation, amplification, genetic analysis, and interpretation. The first step in generating a DNA profile from a sample is extraction, where the analyst attempts to isolate the DNA and separate it from all other cellular material and debris. *Id.* at 99. After extraction of the DNA, the sample is quantitated, *i.e.*, the total amount of DNA present in the sample is estimated. *Id.* at 114. Based on the estimated amount of DNA present, some portion of the extracted DNA is then amplified. Amplification is a process by which DNA is copied at targeted locations (*i.e.* loci) many times

---

<sup>7</sup> John Butler’s textbooks on DNA analysis – including *Fundamentals of Forensic DNA Typing* and *Advanced Topic in Forensic DNA Typing: Interpretation* – are considered authoritative in the field of forensic DNA analysis. They are both used in forensic science education, as well as cited in laboratories’ protocols.

<sup>8</sup> [INSERT NAME OF LAB] tests [NUMBER] loci plus amelogenin, an indicator for male or female sex. Thus, an individual DNA profile generated by [LAB] lists [2xNUMBER] numbers, or alleles: two for each locus (1 per chromosome) at each of the [NUMBER] loci tested, plus either X, X or X, Y at amelogenin.

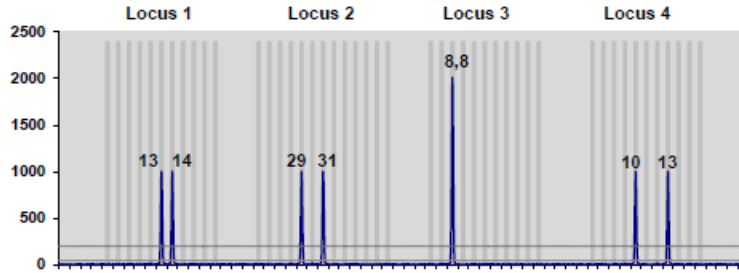
over, generating on the order of a billion copies.<sup>9</sup> *Id.* at 125-26. During the amplification process, the targeted DNA may not amplify if there is only a small amount of it present to begin with<sup>10</sup>, or if it is degraded (i.e. broken into pieces due to environmental exposure or other stressors), or if there are inhibitors (such as some fabric dyes or excess salts) present in the sample. *Id.* at 68. When targeted DNA does not amplify, that genetic information is lost in downstream steps; this loss of genetic information is known as *allelic dropout*, a concept discussed further below. *Id.* at 222.

The post-amplification sample consists of large numbers of only the copied alleles, which can then be separated on an instrument called a genetic analyzer so that each allele can be distinguished and then recorded. *Id.* at 175. The result of this process is a series of peaks on a graph, called an electropherogram. *Id.* at 194. The analyst interprets the electropherogram, generating a genetic profile for the evidence sample. Part of this interpretation process involves determining whether peaks represent “real” DNA or artifacts of the testing process. Each “real” DNA peak corresponds to an allele present in the sample and the height of each peak roughly corresponds to how much of that allele is present (*i.e.* a taller peak indicates more of a particular allele present). When testing a single source evidence sample (*i.e.* a sample originating from one individual), two peaks of roughly equivalent height should be observed at each locus where the contributing individual is a heterozygote (*i.e.* possesses 2 different alleles). At loci where the contributor is a homozygote (*i.e.* possesses two of the same allele), one, relatively high peak should be observed, because the two alleles “stack” on top of one another.

---

<sup>9</sup> Amplification is conducted via a technique called polymerase chain reaction, commonly notated as PCR.

<sup>10</sup> Quantitation gives a preliminary estimate of whether the amount of DNA in the extract falls into this low level range. However, a seemingly sufficient total amount of DNA may be comprised of low levels of DNA from multiple contributors; this is not something that can be discerned from the quantitation step, which does not distinguish between contributors but rather reports the total amount of DNA present.



**Figure 1. Electropherogram showing ideal, single-source DNA data at four hypothetical loci. Note that at Locus 3, where the DNA contributor is a homozygote (possesses two “8” alleles), his two alleles “stack” on top of one another, resulting in a single peak on the electropherogram. At each of the other three loci, where the contributor is a heterozygote (i.e. possesses two different alleles), two peaks are observed. Figure from Butler, *Advanced Topics in Forensic DNA Typing: Interpretation*, 11, Fig. 1.5 (2014).**

After the evidence sample is fully interpreted, the analyst then compares the resulting profile to the profile that the analyst developed from the reference sample(s).<sup>11</sup> [If the analyst in your case did not document their interpretation of the evidence profile before comparing to references, the material in this footnote should be elevated to the main text and made a separate section/paragraph. It can also become the basis of an as-applied challenge.] If the analyst determines that one of the reference profiles “match” or “cannot be excluded from” the evidence profile, the analyst calculates a rarity statistic to contextualize the significance of the match or

<sup>11</sup> It is well established in the scientific community that evidence samples must be fully interpreted before the analyst learns the DNA profiling results, or “genotype”, of the reference samples (i.e. samples from known contributors, such as defendant and victim). Otherwise, knowledge of the genotypes from known contributors can bias the analyst’s interpretation of DNA types from the evidence. *See, e.g.*, National Commission on Forensic Science, Ensuring that Forensic Analysis is Based Upon Task-Relevant Information, adopted Dec. 8, 2015, available at <https://www.justice.gov/ncfs/file/818196/download> (“information about the DNA profile of a suspect is irrelevant and potentially biasing when a DNA analyst is attempting to determine what genotypes are present in an evidentiary sample”). The Scientific Working Group on DNA Analysis Methods (SWGDM) has unambiguously recommended that “[t]he laboratory must establish guidelines to ensure that, to the extent possible, DNA typing results from evidentiary samples are interpreted before comparison with any known samples.” *Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories* (approved 1/4/10), at 3.6.1. The interpretation of the evidentiary samples should be contemporaneously documented. It is only after the analyst has completed the interpretation process—determined a sample to be a mixture, designated allele peaks, identified the number of contributors to the mixture, estimated the relative ratio of individuals contributing to the mixture, and considered all possible genotype combinations that could produce that mixture (including, if possible, deducing a single source “major” DNA profile from the mixture)—that the analyst should be permitted to compare the evidence profile with reference samples. J. Butler, *Fundamentals of Forensic DNA Typing* 19 (2010), at 325, Fig. 14.5.

inclusion. Statistical calculations are an essential part of the interpretation methodology, giving the trier of fact a means of assessing the possibility that the inclusion is “merely a coincidence and that the two samples did not actually come from the same person.” *United States v. Porter*, 618 A.2d 629, 632 (D.C. 1992). Indeed, without these probability statistics, evidence of a DNA inclusion is not admissible in court. *Id.* at 640. [Note: this is true for most jurisdictions. You will need to cite local law for this proposition.] This is because “it would not be scientifically justifiable to speak of a match as proof of identity in the absence of underlying data that permit some reasonable estimate of how rare the matching characteristics actually are.” National Research Council, *The Evaluation of Forensic DNA Evidence, Committee on DNA Forensic Science: An Update* 192 (1996) [NRC II]; see also National Research Council, *DNA Technology in Forensic Science* (1992) [NRC I] at 74 (“To say that two patterns match, without providing any scientifically valid estimate (or at least, an upper bound) of the frequency with which such matches might occur by chance, is meaningless”). The Scientific Working Group for DNA Analysis Methods, or SWGDAM,<sup>12</sup> to provide discipline-wide guidelines, similarly admonishes that analysts “must perform statistical analysis in support of any inclusion that is determined to be relevant in the context of a case, irrespective of the number of alleles detected and the quantitative value of the statistical analysis.” SWGDAM Interpretation Guideline 4.1 (2010).<sup>13</sup>

## **B. Difficulties associated with mixture interpretation**

Forensic DNA samples found at crime scenes often contain DNA from more than one individual. Such a DNA profile representing two or more contributors is termed a DNA mixture. *Fundamentals, supra*, at 320. An analyst knows that they are dealing with a DNA mixture,

---

<sup>12</sup> SWGDAM is an advisory group convened by the Federal Bureau of Investigation. <https://www.swgdam.org/> (accessed May 28, 2017).

<sup>13</sup> Available at [http://www.forensicdna.com/assets/swgdam\\_2010.pdf](http://www.forensicdna.com/assets/swgdam_2010.pdf).

versus a single source sample, if they observe more than two alleles at two or more loci, or if loci with only two alleles display significant peak height imbalance.<sup>14</sup> John Butler, *Advanced Topics in Forensic DNA Typing: Interpretation* (“*Interpretation*”), 129 (2014). Unlike the kind of straightforward analysis involved in interpreting a high quality single source DNA profile, mixtures are often ambiguous, and the process of interpreting them can be highly subjective. In particular, mixtures which cannot be resolved into single source components (“indistinguishable” mixtures)<sup>15</sup> involve a great deal of subjective decision making. Studies have shown that subjective interpretation of indistinguishable DNA mixtures can lead to widely divergent results from one analyst to the next, even analysts in the same laboratory applying the same set of protocols. See Dror and Hampikian, Subjectivity and bias in forensic DNA mixture interpretation, *Sci. & Justice*, 51(4), 204–208 (2011)<sup>16</sup>; NIST Interlaboratory Mixture Interpretation Study 2013 (“MIX13”)(discussed *infra*) and forerunner NIST mixture studies (e.g. MIX05). The two primary complicating factors in mixture interpretation are: (1) “the potential for allele stacking”, and (2) “potential alleles in the stutter position.” *Interpretation* at 153.

---

<sup>14</sup> Two alleles from the same contributor should be roughly the same height, within a degree of tolerance (called a “peak height ratio” (PHR). If the height of two allelic peaks observed at a given locus are not within this predetermined tolerance—i.e. they are “imbalanced”—this is a sign that they actually originate from two people rather than one.

<sup>15</sup> Mixtures may sometimes be resolved or ‘deduced’ into individual sources based on the relative amounts of DNA contributed by each source; the source contributing more DNA is the ‘major contributor’, and the source(s) contributing less DNA is the ‘minor contributor.’ Analysts use the height of allelic peaks on the electropherogram as a proxy for how much DNA is originating from each contributor. Laboratories have specific criteria regarding how much difference they have to observe between peak heights to pull out a major profile; it would be inappropriate to ‘eyeball’ a mixture to determine whether it impressionistically appears that there is ‘enough’ of a difference between contributors to deduce a major profile. Some mixtures encountered in casework do not meet these criteria and therefore the mixture must be treated in its totality rather than as individual single source profiles.

<sup>16</sup> In this study, 17 examiners from one government laboratory were provided a mixed DNA profile from a sex assault case and asked to interpret the profile and compare it to a suspect’s reference profile. The original caseworking analyst had determined that the suspect could not be excluded as a contributor to the mixture. The 17 examiners came to a variety of conclusions: 1 concluded “cannot exclude”; 12 “excluded” and 4 deemed the results “inconclusive”. Among other things, these results underscore the subjectivity of complex mixture interpretation.



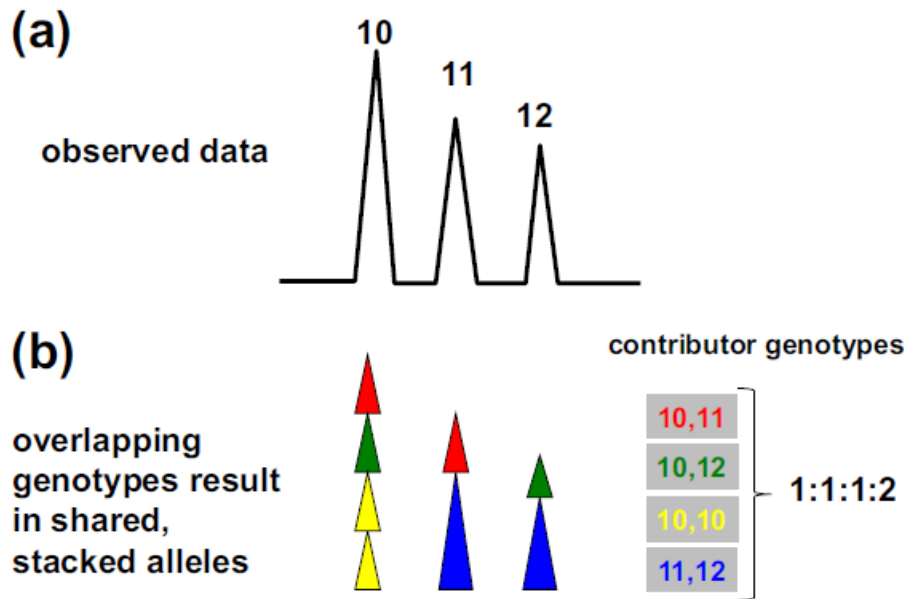
**1. Allele stacking makes it very difficult—if not impossible—to determine the number of contributors to a mixture, and to determine whether crucial DNA data might have “dropped out” of a mixture profile**

As described *supra*, when an individual has two of the same allele at a locus (*i.e.* is a homozygote), that person’s alleles “stack” on top of one another and present as a single peak on the electropherogram. Similarly, when multiple contributors to a DNA mixture possess the same allele at a locus, those alleles also “stack” and present as a single peak. *See, e.g.*, Figure 2, below. This is known as allele stacking or allele sharing.

There are two important consequences of allele stacking. One consequence is that “allele sharing makes accurately deducing the number of contributors to a mixture challenging – and the challenge only grows with each additional contributor to a DNA mixture.” *Interpretation at 169.*<sup>17</sup> If an analyst cannot accurately determine how many contributors there may be in a mixture, the analyst cannot accurately interpret the mixture. Inaccurate interpretation of the mixture impacts not only decisions to include or exclude individuals as potential contributors to the mixture, but also the associated statistical analysis. *Id.* at 335 (“some of these genotype combinations may not fit a reasonable interpretation of the data” depending on the actual number of contributors present).

---

<sup>17</sup> For example, studies have shown that, because of allelic stacking, more than 75% of known four-person mixtures would be misclassified as two- or three- person mixtures based on the maximum number of alleles detected at any given locus. Paoletti et al., Empirical Analysis of the STR Profiles Resulting from Conceptual Mixtures, *J Forensic Sci*, 1361-66 (2005).



**Figure 2. Hypothetical mixture which (a) exhibits only three alleles at a locus (and is thus suggestive of a two person mixture), (b) is actually comprised of three low level contributors plus a single higher level contributor, whose alleles stack on top of one another. Figure from *Interpretation*, at 160, Fig. 7.1.**

A second consequence of allele stacking is that it diminishes the utility of the stochastic threshold. *Interpretation* at 163 (“the potential of allelic stacking, especially with more than two contributors [], can limit the usefulness of a stochastic threshold.”). When the stochastic threshold cannot be effectively utilized, interpretation and statistical analysis are detrimentally impacted. Reliance on a stochastic threshold without considering the possibility of allelic dropout may result in a false inclusion or exclusion.

A stochastic threshold is a Y-axis value on the electropherogram (measured in relative fluorescence units, or RFUs). The value is established by the laboratory’s internal validation studies.<sup>18</sup> Data below the stochastic threshold is in the “potential ‘danger zone’ of unreliable results.” John M. Butler and Carolyn R. Hill, *Scientific Issues with Analysis of Low Amounts of*

<sup>18</sup> The stochastic threshold value(s) should be a part of the lab’s written protocols.

*DNA* (2010).<sup>19</sup> When peaks from the evidence sample fall below the stochastic threshold at a locus, there is a risk that *allelic dropout*—or loss of genetic data—is occurring at that locus. Specifically, allelic dropout occurs when only one of a DNA contributor’s two alleles at a given locus is detected by the DNA typing process.<sup>20</sup> As described *supra*, allelic dropout can happen when an individual’s DNA is present in low levels, is degraded, or is inhibited. Seeing an allele below the stochastic threshold at a given locus is a “warning indicator” that the partner allele (*i.e.* the second allele of the pair) may have dropped out. *Id.* at 163-64. Dropout of a partner allele could lead an analyst to detect a “false homozygote.” *Id.* For example, if the true contributor of an evidentiary DNA sample possesses an 8 and a 12 allele at a given locus, but due to allelic dropout only the 8 allele is detected, an individual who is homozygous for the 8 allele (*i.e.* possesses two 8 alleles) could be falsely implicated, while the true contributor could be falsely excluded. *See, e.g.*, Sec. II.D, Fig. 3 (discussion of Case 5 in MIX13 study). While the stochastic threshold serves some purpose, in that DNA data that is unambiguously in the stochastic range (*i.e.* below the stochastic threshold) is clearly at risk of being incomplete, DNA data that rises above the stochastic threshold is not necessarily safe. This is especially true with complex DNA mixtures, due to the potential for contributors to share alleles (*i.e.* allele stacking).

Allele stacking makes over-reliance on the stochastic threshold particularly dangerous with mixtures. “Just because allelic peaks at a locus are above an established stochastic threshold does not mean that no allele drop-out has occurred in a complex mixture.”

*Interpretation* at 163-64. Allele stacking can falsely elevate a peak at a particular locus above the stochastic threshold. When two or more sub-threshold alleles stack on top of each other, they

---

<sup>19</sup> <https://www.promega.com/resources/profiles-in-dna/2010/scientific-issues-with-analysis-of-low-amounts-of-dna/> (accessed May 28, 2017).

<sup>20</sup> There can also be loss of both alleles at a given location, which is called locus dropout.

may present as a peak that surpasses the stochastic threshold, which in turn may give the false impression that the DNA at that locus is “safe” from the risk of allelic dropout and can be confidently interpreted. In reality, however, each contributor to the falsely elevated peak is in the stochastic (aka dropout) zone. *See, e.g.*, Figure 2 (showing peaks from multiple low level contributors stacking upon one another and presenting as three relatively tall peaks). “The concept of a stochastic threshold *can become meaningless in complex mixtures* due to the potential for allele stacking.” *Interpretation* at 94-95 (emphasis added); *id.* at 177 (“stochastic thresholds often lose their value and meaning when allele sharing is possible with three or more contributors to a DNA mixture”).

Despite the limitations in the utility of the stochastic threshold, it is the primary means the analyst has of assessing possible allelic dropout in complex mixtures.<sup>21</sup> [You should request the validation studies (along with protocols, which will have info on the stochastic threshold) of the lab involved in the case, and if you find that they are relying on two-person mixture samples for validation purposes, you should put more emphasis on the material in the footnote.] While some level of allelic stacking will undoubtedly occur in any DNA mixture, there is no objective way to determine whether allelic stacking is occurring at any given locus in an indistinguishable DNA mixture profile because there is no way to tell whether an observed peak comes from one contributor, or actually is the combined low level (*i.e.* sub-stochastic) contributions of two or more individuals. *See, e.g., supra*, Figure 2 (the 10 allele demonstrates how an above-threshold

---

<sup>21</sup> Allelic dropout is not simply a theoretical possibility. It is “ever-present” and a “real issue faced with complex mixtures,” because “[s]ensitive DNA detection technology has the potential to outpace reliable interpretation.” *Interpretation* at 174, 177. “If a laboratory desires to develop appropriate protocols that will enable reliable interpretation of DNA from low-level DNA or mixtures involving three or more contributors, then validation studies need to be performed with known samples that mimic the amounts of DNA and complexity of profiles where stochastic effects and allele dropout are expected.” *Id.* at 164. Unfortunately, laboratories’ internal validation studies tend to be far from adequate in this respect. The latest iteration of the SWGDAM Interpretation Guidelines require internal validation studies to establish the stochastic threshold, while acknowledging that reliance on the stochastic threshold may not be appropriate in mixture samples where allele sharing is possible. SWGDAM Interpretation Guidelines (2017) 1.7, 1.7.1, and 1.7.1.3.

peak can originate from a combination of two or more individuals whose individual contributions are below the stochastic threshold). While an analyst can make an educated guess based on other information present in the DNA profile, there are no objective guidelines or protocols to guide an analyst's decision-making in this respect. Ultimately, and unavoidably, "allele drop-out and potential allele sharing from multiple contributors lead to greater uncertainty in the specific genotype combinations that can be reliably assumed." *Interpretation* at 177. And, as Dr. Butler has unambiguously warned, "[w]hen there is a high degree of interpretation uncertainty from an evidentiary sample, it makes little sense to try and draw conclusions . . . and expect those conclusions to be reliable." *Id.*

## **2. The challenge of distinguishing artifacts from real DNA is heightened in complex mixtures**

Another significant source of uncertainty in mixture interpretation is distinguishing real alleles from artifacts, particularly an extremely common artifact known as "stutter." Stutter is a by-product of the amplification (*i.e.* copying) step in the DNA testing process, and typically results in a small peak one repeat less than its parent "true allelic" peak (*e.g.* the process would produce a smaller "stutter" peak in the 7 allele position when there is a true 8 allele at that locus). "Because stutter products are the same length as actual allele PCR products, it can be challenging to determine whether a small peak is a real allele from a minor contributor<sup>22</sup> of the original sample or a stutter product of an adjacent allele created during the PCR amplification process." *Id.* at 76. When there are one or more minor contributors present whose alleles are similar in height to the stutter peaks, this task is not just "challenging," it is impossible. *Id.* at 58-59. Stutter peaks tend not to exceed a certain height relative to the associated parent "true allelic"

---

<sup>22</sup> A minor contributor is simply an individual contributing a smaller amount of DNA (which will appear on the electropherogram as smaller peaks) relative to other contributors to a DNA mixture.

peak. However, the fact that a low level peak is adjacent to a larger peak does not necessarily mean that it is stutter. *Id.* at 142 (“It is not always possible to exclude stutter since they are allelic products and differ from their associated allele by a single repeat unit”).

For complex mixtures, stutter is even more problematic. Not only does it become impossible to distinguish real DNA from stutter, but stutter peaks can stack in exactly the same way real allelic peaks do. *Id.* at 71. Thus, stutter can stack on a sub-threshold allelic peak and present as a peak that artificially surpasses the stochastic threshold. Moreover, with a mixture containing one or more low level contributors, “higher levels of stochastic variation can lead to more variability in peak height ratios of heterozygotes and more significant stutter products.” *Id.* at 160. In other words, when there are low-level DNA contributors present in a mixture, stutter peak heights can exceed expected values (i.e. the values set by validation studies) and be confused with real allelic peaks. “This variation leads to a lower confidence in appropriately allocating allele pairs into individual contributor genotypes” with complex mixtures. *Id.*

### **C. Limitations on application of combined probability of inclusion (CPI) statistic to complex mixtures.**

When a mixture cannot be resolved, it means that an analyst is unable to say—based on a scientifically valid procedure—which alleles are associated with a particular contributor/source at each locus. In other words, the analyst cannot say with any confidence which of the multiple alleles detected at each locus belong to the same pair. When this happens—such as in the instant case—the mixture is treated as a whole, rather than as individual single source contributions. The random match probability (RMP) can only be used for single source samples, or simple mixtures where individual contributors can be distinguished. RMP is not applicable to indistinguishable mixtures. Instead, analysts often apply the combined probability of inclusion (CPI) statistic to the mixture as a whole, rather than calculate statistics for individual single

source contributions. John Buckleton & James A. Curran, *A Discussion of the Merits of Random Man Not Excluded and Likelihood Ratios*, *Forensic Sci. Int'l: Genetics* 2, 343 (2008). The CPI provides an estimate of the number of people in the population that have any combination of the alleles observed in the mixture profile. *Id.* For the CPI statistic to be accurate, ***each of the alleles of every contributor to the mixture must be detected.*** In other words, the CPI statistic can only be reliably applied if there is ***no allelic dropout.*** *Interpretation* at 315; SWGDAM Interpretation Guideline 4.6.3 (2010). As the SWGDAM has pointed out, “the potential for allelic dropout raises the possibility of contributors having genotypes not encompassed by the interpreted alleles,” which are not accounted for in the CPI calculation. SWGDAM Interpretation Guidelines, 4.6.3 (2010).<sup>23</sup>

In 2010, SWGDAM recommended that laboratories implement a stochastic threshold to assess DNA profiles for potential dropout.<sup>24</sup> Although most laboratories followed the SWGDAM recommendation and put a stochastic threshold in place, “[a] number of common misconceptions have crept into the community regarding CPI.” *Interpretation* at 335. Based on his review of laboratory practices across the country, and the results of an interlaboratory study conducted by the National Institute of Standards and Technology (NIST), discussed *infra*, Dr. John Butler (NIST) compiled a list of ten “urban legends”—or scientifically erroneous assumptions—that pervade the forensic DNA analysis community about the application of CPI. Among these are:

- ***The misconception that the number of contributors to a mixture does not matter when calculating CPI.*** “The number of contributors ***always matters***

---

<sup>23</sup> Available at [http://www.forensicedna.com/assets/swgdam\\_2010.pdf](http://www.forensicedna.com/assets/swgdam_2010.pdf).

<sup>24</sup> 4.6.3. “When using CPE/CPI (with no assumptions of number of contributors) to calculate the probability that a randomly selected person would be excluded/included as a contributor to the mixture, loci with alleles below the stochastic threshold may not be used for statistical purposes to support an inclusion. In these instances, the potential for allelic dropout raises the possibility of contributors having genotypes not encompassed by the interpreted alleles.” SWGDAM Interpretation Guidelines, 4.6.3 (2010), available at [http://www.forensicedna.com/assets/swgdam\\_2010.pdf](http://www.forensicedna.com/assets/swgdam_2010.pdf).

*during the interpretation* of mixture evidence.” *Interpretation* at 335 (emphasis in original). For example, if a mixture profile contains no locus with more than four alleles, and the analyst interprets this as a two person mixture, he may believe that there is no evidence of allelic dropout, and all loci are safe to use in calculating the CPI statistic. However, if the same mixture is assumed to have three or more contributors, dropout is likely (and more likely with more contributors), and CPI cannot be reliably applied. *See, e.g.*, discussion of NIST interlaboratory study, Case 5, *infra*. [Some labs do not make the assumed number of contributors to a mixture clear in report/notes; you may need to speak with the individual analyst to determine if they made any assumptions about the number of contributors to the mixture, and if so, what that assumption was.]

- ***The misconception that low (i.e. more common) CPI statistics resulting from excluding loci where allelic dropout may be occurring<sup>25</sup> is “conservative.”*** In fact, the more loci excluded from the CPI calculation, ***the higher the chance an innocent person could erroneously be included as a potential contributor to the limited data that is left.*** As Dr. Butler points out, “there is a reduced ability to exclude innocent people when loci are dropped out from consideration in the evidence-to-known comparison due to the possibility of allele drop-out.” *Interpretation*. at 335. [If your case doesn't involve an interpretation where a number of loci were excluded from CPI calculation delete this bullet.]
- ***The misconception that “suspect-driven CPI (where the comparison of each suspect results in a different statistical result) is fine.”*** As Dr. Butler points out, “[t]he CPI statistic is calculated from the evidence profile and should not vary based on the reference profile.” *Id.* at 336. If two or more different CPI statistics are presented for the same mixture profile, this is a red flag that CPI is being misapplied. [If your case doesn't involve suspect-driven CPI – i.e. different statistical results for different known samples compared (whether suspect or victim) to the same evidence profile – then delete this bullet.]
- ***The misconception that “[i]f all peaks at a locus are above the established stochastic threshold, then the locus is safe to use” in the CPI statistical calculation.*** As Dr. Butler points out, “[a]llyle stacking is a possibility”, *id.* at 336, and as discussed *supra*, “stochastic thresholds often lose their value and meaning when allele sharing is possible with three or more contributors to a DNA mixture.” *Id.* at 177. Case 5 in the NIST interlaboratory study, discussed *infra*, is an example of this misconception borne out in practice.

---

<sup>25</sup> In order for CPI statistics to be reliable and accurate, they can only be applied to loci where there is no allelic drop-out. SWGDAM Interpretation Guideline 4.6.3 (2010); Butler *Interpretation* text; SWGDAM Interpretation Guidelines (2017).



Essentially, what Dr. Butler discovered was that laboratories were relying too heavily on the 2010 SWGDAM guidelines, which as SWGDAM itself stated “were written with single-source sample and two-person mixtures in mind.” *Interpretation*. at 162. “[T]here are nuances and limitations to the interpretation of [three or more person] mixtures, which are not fully explored in the 2010 guidelines.” *Id.* at 163. These nuances are reflected in Dr. Butler’s list of “urban legends”. On January 12, 2017, SWGDAM issued new, updated guidelines in order to address interpretation and statistical analysis of complex mixtures. SWGDAM Interpretation Guidelines (2017).<sup>26</sup> SWGDAM issued these guidelines in response to studies and case reviews conducted over the past few years that demonstrate extensive mixture interpretation errors have been made across the forensic community. These cautionary guidelines largely align with the themes found in Butler’s *Interpretation* text; for example, they emphasize that relying on the stochastic threshold as a warning indicator for allelic dropout “may not be appropriate to the interpretation of mixtures when allele sharing is possible, including the effects of sharing amongst stutter and minor allele peaks” (SWGDAM Interpretation Guidelines (2017) at Sec. 1.7).<sup>27</sup>

**D. NIST interlaboratory study indicates a lack of reproducibility and a high rate of error when interpreting complex mixtures and applying the CPI statistic**

“Exploratory interlaboratory tests are one way that the forensic community uses to demonstrate that the methods used in one’s own laboratory are reproducible in another

---

<sup>26</sup> [https://docs.wixstatic.com/ugd/4344b0\\_2a08f65be531488caa8037ed55baf23d.pdf](https://docs.wixstatic.com/ugd/4344b0_2a08f65be531488caa8037ed55baf23d.pdf) (accessed May 29, 2017).

<sup>27</sup> These guidelines put analysts on alert to the hazards of complex mixture interpretation in a way that the 2010 guidelines did not, but they do not constitute standards, protocols, or “an objective and scientifically valid method for the application of CPI.” PCAST at 78. Rather, they instruct laboratories to conduct thorough validation studies using conditions and samples representative of those encountered in casework, and establish detailed protocols for mixture interpretation with these various hazards in mind. *See, e.g.*, SWGDAM Interpretation Guidelines (2017) at 5 (core elements). Further, there have been no interlaboratory studies to demonstrate that, in the wake of the 2017 SWGDAM guidelines, analysts are now reliably and consistently interpreting complex mixtures. *See infra*, discussion of NIST’s MIX13 study.

laboratory and comparable results are generated by these laboratories. These results are essential to demonstrate consistency in results from multiple laboratories . . . .” *Fundamentals* at 303. While NIST’s Applied Genetics Group has previously conducted interlaboratory studies on mixture analysis, the 2013 study was specifically designed to “measure consistency in mixture interpretation across the U.S. after the publication of the 2010 SWGDAM guidelines.” NIST Interlaboratory Mixture Interpretation Study 2013 (“MIX13”).<sup>28</sup> In particular, they were interested in seeing if the 2010 SWGDAM guidelines’ recommendation that all laboratories implement a stochastic threshold resolved the wide variation in mixture interpretation practices within and between laboratories that had been observed in earlier NIST mixture studies (e.g. MIX05).

The NIST MIX13 study was the largest study of its kind, broadly assessing the accuracy, reproducibility, and repeatability of mixture interpretations among and across laboratories. Analysts from one hundred and eight laboratories took part, and 46 states had at least one laboratory participate; the participants were from a mix of federal, state, and local labs. As one of the study’s leading authors has noted, “[d]ue to the number of laboratories responding and the federal, state, and local coverage obtained, this MIX13 interlaboratory study can be assumed to provide *a reasonable representation of current U.S. forensic DNA lab procedures across the community.*” Dr. Michael Coble, Interpretation Errors Detected in a NIST Interlaboratory Study on DNA Mixture Interpretation in the U.S. (July 22, 2015)(“MIX13 Interpretation Errors”)(emphasis in original), attached as Exhibit \_\_\_\_, at 6.<sup>29</sup>

---

<sup>28</sup> The NIST site containing the details of study design and electronic data that was interpreted by participants is available at <http://www.cstl.nist.gov/strbase/interlab/MIX13.htm>.

<sup>29</sup> Dr. Coble’s powerpoint discussing the results of the MIX13 study is available online through NIST at [https://www.nist.gov/sites/default/files/documents/2016/11/22/interpretation\\_errors\\_detected\\_in\\_a\\_nist\\_interlab\\_study\\_on\\_dna\\_mixture\\_interpretation\\_in\\_the\\_us\\_mix13.coble\\_crim1.pdf](https://www.nist.gov/sites/default/files/documents/2016/11/22/interpretation_errors_detected_in_a_nist_interlab_study_on_dna_mixture_interpretation_in_the_us_mix13.coble_crim1.pdf).

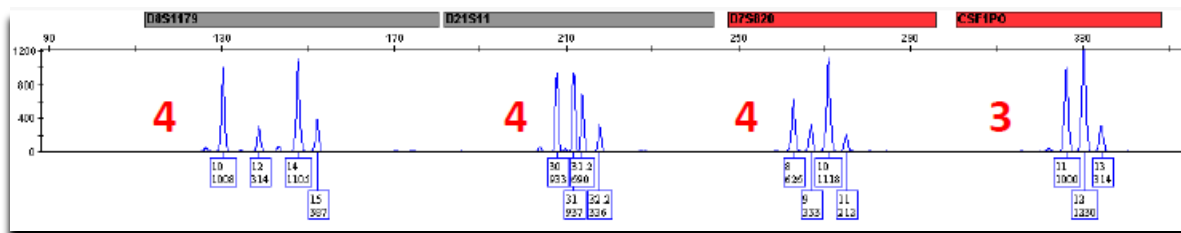
The results of the MIX13 study exposed a disturbing number of errors and showed that, following issuance of the 2010 SWGDAM guidelines, “mixture interpretation is still all over the place.” *Id.* at 37. All participants were provided with the same five mock case scenarios and the same set of five evidentiary DNA profiles to interpret, one for each case. Ground truth was known by the study’s authors for each case used in the study. As a result, the study authors were able to assess whether false exclusions or false inclusions were made. Two of the five cases (Case 1 and Case 4) involved two person mixtures, and participants were provided with reference samples for a victim and a suspect who were true contributors to the mixture; for these cases, it was not possible to make a false positive error. However, the study showed that even for two person mixtures, the calculation of statistics varied widely, with some laboratories improperly using loci with alleles below the stochastic threshold when calculating CPI. *Id.* at 26. For each of the three cases where false positives were possible (Cases 2, 3, and 5)—because non-contributors were provided among the reference samples—both false inclusions (implicating an innocent person) and false exclusions (excluding the true contributor) were made. Further, Dr. Coble noted a “major concern with labs using CPI” for their application of the statistic to a mixture where there were clear signs of allelic dropout. *Id.* at 16, 22, 23.

Case 5 involved a four person mixture which, because of significant allele stacking, could be erroneously interpreted as a two person mixture. *Id.* at 29, 30; Figure 3, *infra*. ***Sixty-nine percent of participants falsely included an innocent individual in this mixture.*** *Id.* at 34. If inconclusive opinions are removed from the total, ***92% of participants making a conclusive determination made a false positive error, implicating an innocent individual.***<sup>30</sup> Notably, all of

---

<sup>30</sup> “When reporting a false positive rate to a jury, it is scientifically important to calculate the rate based on the proportion of *conclusive* examinations, rather than just the proportion of all examinations. . . . [C]onsider an extreme case in which a method had been tested 1000 times and found to yield 990 inconclusive results, 10 false positives, and no correct results. It would be misleading to report that the false positive rate was 1 percent (10/10,000

the peaks in this mixture profile were well above stochastic threshold; unless the participants considered the possibility of allelic stacking, it would not be apparent that allelic dropout might be occurring.



**Figure 3. Portion of the mixture electropherogram from Case 5 in the MIX13 study. Note that while this mixture is actually composed of four contributors, the fact that no more than 4 alleles are detected at any locus could give the erroneous impression that the mixture is comprised of two contributors (2 alleles per contributor). The assumption of two contributors (or even three) would cause the analyst to discount the possibility that allelic stacking is bringing peaks above the stochastic threshold, and the related possibility of allelic dropout. Figure from Exh. \_\_\_ at 29.**

In his discussion of the results of the MIX13 study, along with noting that “mixture interpretation is still all over the place,” Dr. Coble opined that the high level of false positive errors in the study was actually linked to the use of the CPI statistic: “Some of this is a consequence of using a statistical approach that is inappropriate for complex mixture interpretation—*CPI is often being used as a substitute for interpretation, and has the risk of including a non-contributor.*” *Id.* at 37 (emphasis added).

### **III. THE COMPLEX MIXTURE INTERPRETATION AND ASSOCIATED CPI STATISTIC SHOULD BE EXCLUDED UNDER *DAUBERT/FRYE* BECAUSE IT LACKS SCIENTIFIC VALIDITY**

This Court should exclude the complex mixture interpretation and associated CPI statistic in this case under *Daubert/Frye* as they lack scientific validity/reliability because they are not

---

examinations). Rather, one should report that 100 percent of the conclusive results were false positives (10/10 examinations).” PCAST report at 51-52.

generally accepted in the relevant scientific community. The 2016 Report issued by the President’s Council of Advisors on Science and Technology (PCAST Report)—*Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (Sept. 2016)—explained why the use of the CPI statistic is invalid/unreliable. Under *Daubert/Frye*, this court should exclude evidence, lab reports, and testimony regarding the complex mixture interpretation and associated CPI statistic(s) calculated in this case.<sup>31</sup>

#### **A. Daubert**

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 113 S.Ct. 2786, 125 L.Ed. 2d 469 (1993), the Supreme Court held that prior to admitting evidence under Rule 702, district courts must make a preliminary assessment of “whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.” *Id.* at 593-94. The Court referred to a non-exhaustive list of factors that trial courts may consider in reviewing the reliability of proffered expert testimony, including the following: (1) whether the technique used by the expert can be, or has been, tested; (2) whether the technique has been subjected to peer review or publication; (3) the known or potential rate of error of the method used; (4) whether there are standards controlling the technique’s operation; and (5) whether the technique has been generally accepted within the relevant community. *Id.*

---

<sup>31</sup> The relevant scientific community requires that any inclusion result be accompanied by a statistic. *See, e.g., supra* (“it would not be scientifically justifiable to speak of a match as proof of identity in the absence of underlying data that permit some reasonable estimate of how rare the matching characteristics actually are.” *NRC II* at 192; *see also* SWGDAM. Further, a jury cannot give proper weight to DNA evidence without a valid statistical estimate of its rarity. The only statistic provided by the State is scientifically invalid and unreliable. Without an appropriate statistic, the DNA evidence, in its entirety, should be excluded.

The admission of fundamentally unreliable expert testimony can violate due process. *Lee v. Glunt*, 667 F.3d 397, 403 (3d Cir. 2012).<sup>32</sup> The admission of such unreliable expert testimony results in a violation of due process when it undermines “the fundamental fairness of the entire trial,” *Keller v. Larkins*, 251 F.3d 408, 413 (3d Cir. 2001), because “the probative value of [the expert] evidence, though relevant, is greatly outweighed by the prejudice to the accused from its admission.” *Bisaccia v. Attorney Gen.*, 623 F.2d 307, 313 (3d Cir. 1980) (quoting *United States ex rel. Bibbs v. Twomey*, 506 F.2d 1220, 1223 (7th Cir. 1974)).

#### **A. Frye**

The core concept at the heart of the *Frye* standard is that, in order to be admissible, a scientific technique must be accepted as reliable by the scientific relevant community. The burden of establishing the scientific reliability of the technique or procedure is on the proponent. It is also important to note that a scientific technique may be reliable for some purposes and not for others. *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923) (“Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.”).

---

<sup>32</sup> Many Supreme Court cases, from *Napue v. Illinois*, 360 U.S. 264 (1959), to *Manson v. Brathwaite*, 432 U.S. 98, 114 (1977), clearly establish that before evidence may constitutionally be used against a defendant at trial, it must satisfy a minimum standard of reliability. See also *United States v. Scheffer*, 523 U.S. 303, 309-12 (1998) (referring to “legitimate interests in the criminal trial process” which “include ensuring that only reliable evidence is introduced at trial”). Cf. *McDaniel v. Brown*, 558 U.S. --, 130 S.Ct. 665, 674-75 (2010) (per curiam) (noting existence of Due Process reliability claim based on new advances in scientific evidence, but holding claim forfeited where presented for the first time as alternate ground of affirmance in prisoner’s brief before Supreme Court).

A court may revisit previously accepted science – even under *Frye*. *State v. Hull*, 788 N.W.2d 91, 103, n. 3 & 108 (Minn. 2010)<sup>33</sup> (“We agree with the position taken by Justice Meyer in her concurrence that “lengthy use of a method by law enforcement, and even lengthy unquestioning acceptance by courts, does not [by itself] exempt expert evidence from scrutiny under the first prong of *Frye–Mack*...”); See also *State v. Cauthron*, 846 P.2d 502, 506, n. 3 (Wash. 1993) (Although courts may have previously determined a scientific theory to be generally accepted, a court must nevertheless “undertake the *Frye* analysis if one party produces new evidence which seriously questions the continued general acceptance or lack of acceptance as to that theory within the relevant scientific community.”). “Novelty” is in no way required for the admissibility of a scientific technique to be challenged.

Significantly, courts should construe the relevant scientific community broadly and include all scientists who have the background and knowledge to opine on the acceptance of a particular technique. Numerous cases recognize that the scientific community upon whose acceptance admissibility relies cannot be made up solely of those forensic scientists who practice the discipline. In a case involving the admissibility of a field sobriety test, for example, the California Supreme Court stated that “(c)onsistent with both the weight of authority and the cautious, ‘conservative’ nature of *Kelly*, we conclude that testimony by police officers regarding the mere administration of the test is insufficient to meet the general acceptance standard required by *Kelly*.” *People v. Leahy* 8 Cal.4th 587, 609 (1994). In the case of DNA mixtures, it is clear members of the National Academy of Science, PCAST and scientists at NIST are important members of the relevant scientific community.

---

<sup>33</sup> Minnesota is a modified *Frye* state. The first prong of the analysis is the basic general acceptance test of *Frye*. Once established, the court then examines whether the test is reliable.

**B. NATION’S TOP SCIENTISTS CONCLUDED THAT COMPLEX MIXTURE INTERPRETATION AND APPLICATION OF THE CPI STATISTIC IS “CLEARLY NOT FOUNDATIONALLY VALID”**

In 2016, the President’s Council of Advisors on Science and Technology (PCAST) issued a Report to the President, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (Sept. 2016) (“PCAST Report”). The authors were “an advisory group of the Nation’s leading scientists and engineers, appointed by the President to augment the science and technology advice available to him from inside the White House and from cabinet departments and other Federal agencies.” PCAST Report at 3. “PCAST is consulted about, and often makes policy recommendations concerning, the full range of issues where understandings from the domains of science, technology, and innovation bear potentially on the policy choices before the President.” The PCAST group includes the President of the Broad Institute of Harvard and MIT, experts in biology, aerospace, astrophysical sciences, natural resources and environment, string and particle theory, electrical engineering and computer science, nanotechnology, and a Medical Doctor. PCAST Report at vi.

The PCAST Report was created in response to a 2009 report by the National Research Council that was highly critical of the use, and misuse, of forensics in criminal cases—*Strengthening Forensic Science in the United States: A Path Forward* (“NRC Forensics Report”).<sup>34</sup> In 2015, President Barack Obama convened some of the country’s leading scientists to evaluate whether there were “additional steps on the scientific side,” in addition to those already taken in response to the NRC Forensics Report, “to help ensure the validity of forensic evidence used in the Nation’s legal system.” PCAST Report at x. PCAST formed a working

---

<sup>34</sup> <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf> (accessed Feb. 27, 2017).



group that included several members of the PCAST permanent advisors.<sup>35</sup> In contrast to the 2009 NRC Forensics Report, which touched on twelve separate disciplines, PCAST examined just six “forensic feature comparison” disciplines: firearms analysis; DNA analysis of single source samples, simple mixture samples, and complex-mixture samples; bitemark analysis; latent fingerprint analysis; footwear analysis; and hair analysis. The group’s goal was to determine whether those disciplines were scientifically valid and whether they had a methodology that could be reliably applied—the foundational requirements for admissibility. PCAST Report at x.

The group evaluated over 2,000 papers and studies from various sources, including papers submitted in response to PCAST’s request for information from the forensic-science stakeholder community. It consulted with forensic scientists, including those at the Federal Bureau of Investigations lab and the National Institute of Standards and Technology. *Id.* at 2.

Among the leading forensic DNA scientists and statisticians called on by PCAST were:

- John Butler, Special Assistant to the Director for Forensic Science, Special Programs Office, National Institute of Standards and Technology.<sup>36</sup>
- John Buckleton, Principal Scientist, Institute of Environment and Scientific Research, New Zealand.<sup>37</sup>
- Bruce Budowle, Professor, Executive Director of Institute of Applied Genetics University of North Texas Health Science Center.<sup>38</sup>

---

<sup>35</sup> The Group included Eric S. Lander, the President of the Broad Institute of Harvard and MIT, Michael McQuade, the Senior Vice President for Science and Technology at United Technologies Corporation, S. James Gates, Jr., the John S. Toll Professor of Physics and the Director of the Center for String and Particle Theory at the University of Maryland, College Park, William Press, the Raymer Professor in Computer Science and Integrative Biology at the University of Texas, Austin, Susan L. Graham, the Pehong Chen Distinguished Professor Emerita in Electrical Engineering and Computer Science at the University of California, Berkeley, Daniel Schrag, the Sturgis Hooper Professor of Geology and Professor Environmental Science and Engineering at Harvard University, and the Director of the Harvard University Center for Environment, two staff members: Diana E. Pankevich, AAAS Science and Technology Policy Fellow and Kristen Zarrelli, the Advisor on Public Policy & Special Projects at the Broad Institute of Harvard and MIT, and Writer Tania Simoncelli, the Senior Advisor to the Director at the Broad Institute of Harvard and MIT. PCAST Report at vii.

<sup>36</sup> <https://www.nist.gov/people/john-butler> (accessed Feb. 27, 2017).

<sup>37</sup> <https://www.nist.gov/people/john-buckleton> (accessed Feb. 27, 2017); <http://strmix.esr.cri.nz/> (accessed Feb. 27, 2017).

- Daniel MacArthur, Assistant Professor Harvard Medical School Co-Director of Medical and Population Genetics Broad Institute of Harvard and MIT.<sup>39</sup>
- Norah Rudin, President and CEO Scientific Collaboration, Innovation & Education Group.<sup>40</sup>
- Constantine A. Gatsonis, Henry Ledyard Goddard University Professor of Biostatistics Chair of Biostatistics Director of Center for Statistical Sciences Brown University.<sup>41</sup>

PCAST asked whether each forensic discipline that it assessed met two key requirements for scientific validity: (1) “foundational validity” – whether the method can, in principle, be reliably applied; and (2) “validity as applied” – whether the method has been reliably applied in practice. PCAST Report at 56; Section III.B.1 *infra*. As described below, PCAST, in consultation with the world’s leading DNA experts, found that evaluation of complex mixtures and application of CPI statistics in the manner seen in the instant case, is unreliable and without “foundational validity.” See Section IV.B.2. *infra*.

### 1. The requirements for scientific validity

Much like in 2009 NRC Forensics Report, PCAST asked whether each forensic discipline met two key requirements for scientific validity: (1) “foundational validity” – that the method can, in principle, be validly applied; and (2) “validity as applied” – that the method has been reliably applied in practice. PCAST at 56. To be “foundationally valid[,]” a field must utilize a method that has been subject to “*empirical* testing by multiple groups, under conditions appropriate to its intended use.” *Id.* at 5 (emphasis in original). Those studies must show that the method is “repeatable and reproducible.” A method is “repeatable” if, with a known

---

<sup>38</sup> <https://www.unthsc.edu/graduate-school-of-biomedical-sciences/molecular-and-medical-genetics/laboratory-faculty-and-staff/> (accessed Feb. 27, 2017); <http://www.fsc.texas.gov/profile/dr-bruce-budowle> (accessed Feb. 27, 2017).

<sup>39</sup> <https://www.broadinstitute.org/bios/daniel-macarthur-0> (accessed Feb. 27, 2017).

<sup>40</sup> <http://scieg.org/Management.html> (accessed Feb. 27, 2017).

<sup>41</sup> <http://www.stat.brown.edu/FacultyDisplay.aspx?id=1100924168> (accessed Feb. 27, 2017).

probability, an analyst can reach the same result when analyzing samples from the same sources. A method is “reproducible” if, with known probability, different examiners can obtain the same result when evaluating the same samples. PCAST at 47. Put slightly differently, a feature comparison method is *foundationally valid if studies show it has a “reproducible and consistent procedure”* for:

- (a) identifying features within evidence samples;
- (b) comparing the features in two samples; and
- (c) determining, based on the similarity between the features in two samples, whether the samples should be declared to be a proposed identification (‘matching rule’).”

PCAST Report at 48. The studies must also provide “valid estimates of the method’s accuracy,” in order to demonstrate how often an examiner is likely to draw the wrong conclusions. *Id.* “Without appropriate estimates of accuracy, an examiner’s statement that two samples are similar – or even indistinguishable – is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact.” *Id.* at 6; *see also id.* at 48 (“Without an appropriate estimate of its accuracy, a metrological method is useless – because one has no idea how to interpret its results.”).<sup>42</sup> Simply put, in order to be foundationally valid, the feature comparison method has to “show its work” through studies that document that examiners are able to do what they say they can do, and how often they get the right answer.

Once a method has been established as foundationally valid, to meet the criteria for scientific acceptance, it must also be valid “as applied.” A DNA analyst in a given case must be *capable* of reliably applying the method, and he or she must have *actually* reliably applied the

---

<sup>42</sup> For objective methods – such as the interpretation of single source DNA evidence – the field can show foundational validity by studying and “measuring the accuracy, reproducibility, and consistency of each of its individual steps” in interpretation. *Id.* at 5. For subjective feature-comparison methods such as toolmarks or fingerprint analysis, which to-date rely on an examiner’s eyeball comparison of features on known and unknown samples, the method must be evaluated “as if it were a ‘black box’ in the examiner’s head,” with many studies involving numerous examiners “render[ing] decisions about many independent tests” with corresponding error rates determined.” *Id.*

method in the case at hand. *Daubert*, 509 U.S. at 593-4/*Frye*, 293 F. at 1014. To ensure that the examiner is capable of applying the technique, the field must conduct rigorous proficiency tests evaluating how often an expert reaches the correct answer in circumstances modeling the procedures actually used in case work. *Id.* at 56. To show that the examiner has applied the method reliably in each case, the examiner must make available all procedures used, the results obtained, and any laboratory notes taken. *Id.* Finally, the examiner must make only scientifically valid assertions about the probative value of the identification. The analyst must accurately report the false positive rate for the method, and cannot overstate the significance of his conclusion by making claims that exceed the empirical evidence and the “applications of valid statistical principles to that evidence.” *Id.* at 6.

PCAST stressed that the field must prove scientific validity through rigorous empirical studies, and not through unsupported claims that an examiner can rely on his or her training or experience. “[N]either experience, nor judgment, nor good professional practices . . . can substitute for actual evidence of foundational validity and reliability. The frequency with which a particular pattern or set of features will be observed in different samples, which is an essential element in drawing conclusions, is not a matter of ‘judgment.’ It is an empirical matter for which only empirical evidence is relevant.” *Id.* at 6. Similarly, “an expert’s expression of confidence based on personal professional experience or expressions of consensus among practitioners about the accuracy of their field is no substitute for error rates estimated from relevant studies.” *Id.* at 55. “[G]ood professional practices – such as the existence of professional societies, certification programs, accreditation programs, peer-reviewed articles, standardized protocols, proficiency testing, and codes of ethics – cannot substitute for actual evidence of scientific validity and reliability.” *Id.*

## **2. PCAST on DNA Analysis**

One of the “feature comparison” methods that PCAST evaluated was DNA analysis. PCAST evaluated three separate categories of DNA analysis: (1) single-source samples, (2) simple mixture samples, and (3) complex mixture samples. PCAST at 69-83.

PCAST found that methods for evaluating single source DNA samples and simple, distinguishable mixtures were foundationally valid. *Id.* at 75. However, PCAST found that the evaluation of complex or indistinguishable DNA mixtures using Combined Probability of Inclusion (CPI)-based methods is definitively *not* foundationally valid. *Id.* at 82. This case involves the [LAB NAME]’s evaluation of complex DNA mixtures and the application of the CPI statistic to these mixture(s).

In order to understand why evaluation of complex, indistinguishable mixtures and application of the CPI statistic to these mixtures is foundationally invalid, it is important to understand why PCAST found single source and simple mixture interpretation methods foundationally valid. Understanding objective, foundationally valid DNA analysis highlights why the lab’s subjective and foundationally invalid analysis in this case is unreliable [is not generally accepted] and should be barred under *Daubert/Frye*. PCAST’s findings of foundationally valid and invalid procedures are explained in Section III.B.2.a and b below.

### **a. Foundationally valid DNA analysis**

PCAST reviewed methods for analysis of single-source DNA and simple mixtures together, since both involve interpretation of and application of statistical calculations to a distinguishable, single-source sample.

Single-source DNA analysis involves the analysis of an evidence sample from a single individual. *Id.* at 70. PCAST categorized DNA analysis of a single source sample as an

objective method because the protocols are precisely defined and “the interpretation involves little to no human judgment.” *Id.* PCAST defined a simple mixture as one where the unknown contributor can be analyzed as a single source sample—either through differential extraction or subtraction of alleles of the known individual. *Id.* PCAST also found that this is largely an objective method. *Id.*

As both single-source and simple mixture analysis are objective methods, PCAST was able to review each step in the interpretation/feature comparison process for foundational validity. PCAST found that each step of feature comparison—feature identification, comparison, and evaluation of the comparison (“matching rule”)<sup>43</sup>—for single-source and simple mixture DNA analysis is foundationally valid:

- (1) *Feature identification*: the features used in DNA analysis—the fragment lengths at the specified loci—are defined in advance. PCAST Report at 71.
- (2) *Feature measurement and comparison*: PCR amplification has been rigorously validated. Additionally, PCR kits must be externally and internally validated before use. The fragment sizes are measured by automated procedure with only a small but measureable variability. Finally, multiple validation studies for fragment sizes were done and show accuracy, precision, and reproducibility. PCAST Report at 71.
- (3) *Feature comparison*: There are clear and well-specified “matching rules” for comparing a known profile to a profile from an evidence sample. PCAST Report at 72.
- (4) *Estimation of random match probability (“RMP”) (providing statistical estimate of “match”)*: RMP—the probability of a match occurring by chance—is based on well-established principles of genetics and statistics. The frequencies of individual alleles were obtained after evaluation and study of population groups. The frequency or RMP of this pattern of alleles in an individual profile is estimated by multiplying the frequencies at the individual loci. PCAST Report at 72-3. In contrast to complex DNA mixtures which consist of indistinguishable, stacking allelic contributions, all of the alleles under consideration when applying RMP are from one individual (assuming the analyst is correct in deeming the

---

<sup>43</sup> See PCAST Report at 48; see also p. 23 *supra*.

sample single-source, or in deducing an individual contributor from a simple mixture).

The feature comparison methods of DNA analysis for single-source samples and simple mixtures are objective, foundationally valid, and reliable. *Id.* at 70, 73. However, as described *supra* and below, complex mixture analysis is subjective inconsistent, and unreliable in its application. *Id.* at 76. Subjective analysis of complex mixtures, combined with the use of the CPI statistical method, can and does lead to inaccurate and widely variable results. *Id.*; *see also supra*, Sec. II.D (MIX13 study).

**b. Use of CPI for the evaluation of complex mixtures is not foundationally valid.**

As PCAST points out, “[t]he fundamental difference between DNA analysis of complex-mixture samples and DNA analysis of single source and simple mixtures lies not in the laboratory processing, but in the interpretation of the resulting profile.” *Id.* at 75. Unlike the relatively straightforward analysis involved when an individual contributor’s allelic contributions can be distinguished, complex DNA mixtures “result in a DNA profile that superimposes multiple individual DNA profiles.” *Id.* at 75. This garbled data leads to a thorny mess that is not encountered with single source samples and simple DNA mixtures:

Interpreting a mixture profile is different for multiple reasons: each individual may contribute two, one or zero alleles at each locus; the alleles may overlap with one another; the peak heights may differ considerably, owing to differences in the amount and state of preservation of the DNA from each source; and the ‘stutter peaks’ that surround alleles (common artifacts of the DNA amplification process) can obscure alleles that are present or suggest alleles that are not present. It is often impossible to tell with certainty which alleles are present in the mixture or how many separate individuals contributed to the mixture, let alone accurately to infer the DNA profile of each individual.

*Id.* at 75-76; *see also supra* Sec. II.B (citing to Butler’s *Interpretation* text throughout).

DNA analysts largely rely on their subjective judgment to pick their way through the morass of ambiguous data posed by complex mixtures, a practice that PCAST found

“problematic, because subjective choices made by examiners, such as about which alleles to include in the [associated CPI] calculation, can dramatically alter the result and lead to inaccurate answers.” PCAST Report at 76. PCAST used examples from casework and scientific literature to illustrate the significant impact that subjectivity can have on complex mixture interpretations and associated CPI statistical analysis:

- In a 2003 double homicide in Massachusetts, DNA testing was performed on a discarded glove. At trial, the prosecution expert testified that the defendant could not be excluded from a mixture of at least three contributors. Jurors were told the chance of such a match occurring by chance was 1 in 1.1 billion. That defendant was convicted and sentenced to death. However, another expert reviewed the case in 2009 and wrote that the chance was closer to 1 in 2. In other words, 50% of the population could not be excluded from the mixture. *Id.* at 76.
- In a 2016 paper, 19 DNA experts were given a mock case involving three complex DNA mixtures, which they were to compare to reference samples from an alleged victim and an alleged suspect. Each expert was provided the same DNA profiles, but their conclusions varied wildly. One examiner excluded the suspect as a possible contributor. Another examiner declared a match and reported a random match probability of 1 in 209 million. *Id.* at 77, fn. 207.
- A 2011 study used a complex DNA mixture from an actual case in Georgia to demonstrate the impact of irrelevant case information on the subjective nature of complex mixture analysis. In the actual case, the State’s experts knew, prior to conducting their interpretation of the DNA results, that one of the suspects implicated a second man as part of a plea bargain and the suspect’s testimony could not be used unless there was corroborating DNA evidence. Both analysts involved in the case concluded that the second man could not be excluded (i.e. was included) as a contributor to the complex DNA mixture. In the 2011 study, the complex DNA mixture from the Georgia case was provided to 17 expert DNA examiners, but without the irrelevant contextual information. 12 of the examiners excluded the second man. Only 1 examiner agreed with the original experts that the second man could not be excluded. *Id.* at 76-77, fn. 206.<sup>44</sup>

PCAST also highlighted a particularly striking example of the impact of inconsistent and unreliable mixture interpretation and CPI statistical calculation coming out of Texas. In 2015, as

---

<sup>44</sup> The other 4 examiners deemed the mixture inconclusive. Whether or not the original examiners were biased in their interpretation of the complex mixture, the mixture interpretation methodology employed by both the original examiners and the 17 who performed the subsequent analysis (all of whom came from the same lab and applied the same set of mixture interpretation protocols) is unquestionably subjective, leading to widely ranging results. *See supra*, fn. 16.



a result of minor errors detected in the FBI’s population database,<sup>45</sup> the Texas Department of Public Safety issued notice that, upon request, they would conduct a recalculation of the statistics in a case. *Id.* at 77. The letter noted that the errors were expected to be relatively minor. *Id.*

However, in a number of cases, the statistics changed dramatically: “**from 1 in 1.4 billion to 1 in 36 in one case; from 1 in 4000 to inconclusive in another.**” *Id.* The Texas Forensic Science Commission (“TFSC”) determined that the remarkable change in statistics was not resulting from the corrected population frequencies, but was, in fact, the result of changes in the way the laboratories calculated the CPI statistic—specifically, how the labs dealt with “allelic dropout.” *Id.* at 78. As a result, TFSC began a state-wide review of all mixture cases that is ongoing. This review has brought to light “the extent to which DNA-mixture analysis involved subjective interpretation” and the use of guidelines that do not “clearly, adequately or correctly specify the proper use or limitations of the [CPI statistical] approach.” *Id.* at 78.

Not surprisingly, in light of the above, PCAST found that “DNA analysis of complex mixtures based on CPI-based approaches” is an “**inadequately specified, subjective method that has the potential to lead to erroneous results**” and concluded that it is “**clearly not foundationally valid.**”<sup>46,47</sup> *Id.* at 78, 82 (emphasis added). This devastating assessment joins a

---

<sup>45</sup> This is one of several databases of allelic frequencies in different population groups. The frequencies are inputted into statistical formulae (including both RMP and CPI).

<sup>46</sup> Because they concluded that complex mixture interpretation with the CPI statistic was not foundationally valid, PCAST did not reach the question of “validity as applied” for this method.

<sup>47</sup> The report noted that the experts convened by TFSC proposed specific rules for the use of the CPI statistic in a paper published just before the PCAST Report was published, but that PCAST did not have adequate time to assess whether these rules were sufficient to define an objective and scientifically valid method for the application of CPI. PCAST at 78, 82. These proposed rules largely mirror the discussion of ambiguities and difficulties inherent in complex mixture interpretation from section II.B (*supra*) and by PCAST. For example: “any of the observed peaks (true allelic or ...stutter) may overlap with a peak from the same or another donor”; “as the number of potential contributors increases, so does the uncertainty in accurately determining the true number of contributors”; “it can be extremely difficult, if not impossible to distinguish . . . even a three-person mixture from a four-person mixture” (notably, the paper expressly does not attempt to set forth “how to determine the number of contributors”); characterizes as a “misguided concept” the belief that a locus can be used in CPI calculations if all peaks observed are over the stochastic threshold. *See* Bieber, et al, BMC Genetics, 17:125 (2016). The “overriding principle” of the

chorus of criticisms of CPI-based complex mixture interpretation by varied experts in the field of forensic DNA analysis.<sup>48</sup> “When there is a high degree of interpretation uncertainty from an evidentiary sample”—an implicit feature of complex mixtures, where the number of contributors and degree of allele sharing are unknown—“it makes little sense to try and draw conclusions . . . and expect those conclusions to be reliable.” *Interpretation* at 177. “In the words of Professor Max Baur of Bonn University in Germany: ‘[CPI] is a deficient method and we should not use it!’” *Id.* at 322.<sup>49</sup>

**C. THIS COURT SHOULD EXCLUDE COMPLEX MIXTURE INTERPRETATION AND THE ASSOCIATED CPI STATISTIC AS SCIENTIFICALLY INVALID.**

As the shockingly high rate of false positive errors in the only large scale interlaboratory study conducted to date and the unambiguously critical assessment of the PCAST group demonstrate, complex DNA mixture interpretation and the associated CPI statistic lack the scientific validity that is the predicate for evidentiary admissibility. Unlike the objective, well-

---

paper mirrors Butler’s *Interpretation* text and the discussion in Sec. II.B, *supra*: “Any locus that has a reasonable probability of allele drop-out should be disqualified from use in calculation of the CPI statistic.” *Id.* at 6. The authors concede that they “cannot prescribe what is a ‘reasonable probability’” of allelic dropout but suggest that it should be very low, on the order of 1 in a 100. *Id.*

<sup>48</sup> See, e.g., Brenner, C.H., The mythical “exclusion” method for analyzing DNA mixtures – does it make any sense at all?, *Proc. of the Am. Acad. Forensic Sci.*, Vol.17, p. 79 (2011) (“Certainly no one has laid out an explicit and rigorous chain of reasoning from first principles to support the exclusion method. It is at best guesswork.”); J. Buckleton, et. al., *Forensic DNA Evidence Interpretation*, 2d Ed. (2016) 241 (“worldwide the move is away from CPI.”); Curran JM, and Buckleton, J., Inclusion probabilities and dropout, *J. Forensic Sci.* (2010)(“It is false to think that omitting a locus (during calculation of the CPI statistic) is conservative as this is only true if the locus does not have some exclusionary weight.”); M. Perlin, Inclusion probability for DNA mixtures is a subjective one-sided match statistic unrelated to identification information, *J Pathol Inform.*, 28:6:59 (2015).

<sup>49</sup> As Dr. Butler points out, “[o]ne of the most significant deficiencies of the CPI approach is that this approach does not take into account an alternative hypothesis. How can you make an informed opinion on a topic if you consider only one point of view? This is why likelihood ratios are essential in mixture interpretation and why they have been strongly recommended by the ISFG DNA Commission—because two different possibilities are compared in developing an opinion.” *Id.* at 322. There has been a definitive shift in U.S. forensic practice toward the use of probabilistic genotyping, where mixture interpretation is largely conducted by a computer program that incorporates biological models of allelic dropout and stutter. With probabilistic genotyping methods, probative value of DNA mixture evidence is communicated via likelihood ratios rather than the CPI statistic. As a result, these methods are capable of comparing the likelihood of different genotype combinations, rather than assuming dropout did or did not occur as is required under the outdated CPI framework.

defined, and validated methods for interpretation of single source DNA and simple mixtures, complex mixture interpretation methods involve the execution of a series of subjective decisions on extremely ambiguous data. *See supra* Secs. II.B-D. There is no method for complex mixture interpretation that has been sufficiently tested or generally accepted by the relevant scientific community, and there is no evidence that examiners regularly reach accurate conclusions. *See Daubert*, 509 U.S. at 593-95. In fact, there is ample evidence that DNA analysts routinely draw inaccurate conclusions when interpreting complex mixtures and calculating associated CPI statistics. *See supra*, Secs. II.D and III.B.2.b; *Daubert*, 509 U.S. at 593-95 (emphasizing the importance of a known error rate in assessing a field's reliability). This Court should thus exclude complex mixture interpretation evidence and associated statistical calculations from lay jurors' consideration.

## **COMPLEX DNA MIXTURES (WITH CPI STATISTIC): JURY INSTRUCTIONS INCORPORATING PCAST REPORT**

### **Contents**

- 1. “Expert Witness Instruction for DNA Examiner – p 1**
  - “Foundational” factors – p 3
  - “Validity as applied” factors – p 7
  - “Expert Qualifications” factors – p 8
  - “Laboratory Qualification” factors – p 9
- 2. Preliminary Instruction Addressing the CSI Effect – p 10**
- 3. Instruction Explaining “False Positive Rate” – p 10**
- 4. Instruction Designed to Avoid the “Prosecutor’s Fallacy” – p 12**

As support for the specific language choices here, we have included a version of the instructions with cites to:

- **The PCAST Report and Addendum:** Report to the President, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature Comparison Methods*, Executive Office of the President, President’s Council of Advisors on Science and Technology (Sept. 20, 2016), and *An Addendum to the PCAST Report on Forensic Science in Criminal Courts* (Jan. 6, 2017);
- **ABA Resolution 101C:** American Bar Association Resolution 101C (adopted Feb. 6, 2012).

### **1. “Expert Witness” Instruction for DNA Examiner**

The following instruction assumes that the court has admitted DNA complex mixture<sup>1</sup> typing results, and a match statistic called the “Combined Probability of Inclusion” (CPI), as *scientific* evidence, and that your efforts to exclude the testimony as lacking foundational validity have failed, but that you have made some inroads with the court about potential validity concerns such that the court has permitted cross-examination challenging the accuracy and reliability of the method.

The instruction modifies a standard “expert witness” instruction with a list of factors the jury should consider when evaluating “scientific” expert testimony.<sup>2</sup> Factors listed below are grouped into four categories: (1) “Foundational” factors, designed to ferret out whether the method has been “shown, based on empirical studies, to be repeatable, reproducible, and

---

<sup>1</sup> PCAST defines a “complex mixture” as one “with more than two contributors.” PCAST Report at 75.

<sup>2</sup> American Bar Association Resolution 101C (adopted Feb. 6, 2012) (“RESOLVED, That the American Bar Association urges judges and lawyers to consider the following factors in determining the manner in which expert testimony should be presented to a jury and in instructing the jury in its evaluation of expert scientific testimony in criminal and delinquency proceedings: . . . Whether to include in jury instructions additional specific factors that might be especially important to a jury’s ability to fairly assess the reliability of and weight to be given expert testimony on particular issues in the case”).

accurate, at levels that have been measured and are appropriate to the intended application;”<sup>3</sup>; (2) “Validity as applied” factors, designed to ferret out whether “the method has been reliably applied in practice;”<sup>4</sup> (3) “Expertise” factors, designed to help the jury assess the expert’s expertise; and (4) “Laboratory” factors, designed to help the jury assess the laboratory’s qualifications. This list of factors is in no way exclusive, nor should every factor be given in every case. Instead, the list should be targeted to the issues in your case to aid the jury without overwhelming it.

Witnesses are usually permitted to testify only about what they directly experienced, such as what they saw or what they did on a particular occasion.<sup>5</sup> Witnesses are not generally allowed to give their opinions. However, some witnesses are allowed to give their opinions because they have acquired a skill, through their training, education, or experience that few members of the general public possess.

You are about to hear the testimony of a forensic DNA examiner, who claims to have specialized knowledge and experience both in the analysis of DNA mixtures of three or more potential contributors and in the application of the so-called CPI statistic, including the [*describe method, e.g., the determination of whether a suspect’s DNA profile is consistent with having been present in a complex DNA mixture found at the scene, and if so, the determination of the statistical significance of that fact*].<sup>6</sup> [Name of expert] will testify that this opinion is not based simply on his experience in casework, but rather is based on a method that is “scientific.”

In determining the value of an expert opinion claiming to be based on *science*, you should consider a number of special factors, which I will talk about in a moment. In determining the value of any expert opinion, remember that just because a witness is allowed to give his or her opinion does not mean that you must believe his or her opinion. As with any other witness, it is up to you to decide whether you believe this testimony and wish to rely on it when making decisions about the case.

When you decide whether you believe the witness’ opinion, you may consider whether the witness has enough training and experience to give the opinion that you heard. You may also consider the method or technique or the witness used, and whether there is a strong enough scientific basis for drawing the conclusions from that method or technique that the expert has drawn.

You may completely or partially disregard the opinion if you decide that there is not a strong enough scientific basis for the opinion, or that the witness used an unreliable method to form his opinion or used a method that hasn’t yet been proven, or that

---

<sup>3</sup> PCAST Report at 4.

<sup>4</sup> PCAST Report at 5.

<sup>5</sup> This paragraph is based on the instruction approved in *United States v. Starzecpyzel*, 880 F. Supp. 1027, 1050-51 (S.D.N.Y. 1995).

<sup>6</sup> This paragraph is based on the instruction approved in *United States v. Starzecpyzel*, 880 F. Supp. 1027, 1050-51 (S.D.N.Y. 1995).

there are not enough reasons to support the opinion, or that the opinion is not based on enough training, education, or experience, or that the opinion is outweighed or contradicted by other evidence. You should consider this evidence with all the other evidence in the case and give it as much weight as you think it fairly deserves.

In deciding how much weight, if any, to give this opinion, you may consider any factors that you decide are relevant. Factors that you may want to consider include the following:

1. Factors related to the method's scientific foundation, to guide you in determining whether the method itself is scientifically valid and reliable:

*On whether the method's validity and reliability have been sufficiently proven through scientific testing:*

- a) Has complex DNA mixture analysis, accompanied by a CPI statistic, been studied by individuals or organizations that had no stake in the outcome, and who empirically tested the CPI method using an approach where the examiners did not know the right answer, and where a large number of analysts were tested?<sup>7</sup>
- b) And if so, did those studies demonstrate that the CPI method of analyzing complex DNA mixtures consistently produces accurate and reliable results?<sup>8</sup>
- c) And did those studies use evidence samples that are similar to the sample in this case? In other words, did the testing involve samples that are similar to the samples in this case, in terms of [number of potential contributors, the amount and quality of the DNA, and the mixture proportion of the person of interest [*select the particulars of your case*]]?<sup>9</sup>

---

<sup>7</sup> PCAST Report at 143 (describing requirements of scientific validity and reliability); American Bar Association Resolution 101C (adopted Feb. 6, 2012) at 11-12, 13 (“The court should consider whether additional factors such as those set forth below might be especially important to a jury’s ability to fairly assess the reliability of and the weight to be given testimony on a particular issue . . . The extent to which the forensic science technique or theory has undergone validation.”).

<sup>8</sup> PCAST Report at 48 (“The method need not be perfect, but it is clearly *essential* that its accuracy has been measured based on appropriate empirical testing and is high enough to be appropriate to the application. Without an appropriate estimate of its accuracy, a metrological method is useless – because one has no idea how to interpret its results.”); *id.* at 5 (describing “essential points of foundational validity”); American Bar Association Resolution 101C (adopted Feb. 6, 2012) at 11-12, 13 (“The court should consider whether additional factors such as those set forth below might be especially important to a jury’s ability to fairly assess the reliability of and the weight to be given testimony on a particular issue . . . The known nature of error associated with the forensic science technique or theory.”).

<sup>9</sup> PCAST Report at 66 (forensic examiners must “demonstrate that the samples used in the foundational studies are relevant to the facts of the case”); *see also* PCAST Addendum at 9 (“When considering the admissibility of testimony about complex [DNA] mixtures (or complex samples), judges should ascertain whether the published validation studies adequately address the nature of the sample being analyzed (e.g., DNA quantity and quality, number of contributors, and mixture proportion for the person of interest).”).

- d) Did the complex mixture interpretation method in this case rely on any assumptions, such as an assumption about the number of contributors, that have not been subject to empirical testing and might be incorrect or unreliable?

*On whether the method sufficiently accounts for cognitive bias:*

- a) Is complex DNA mixture analysis, accompanied by a CPI statistic, a subjective or objective method? Subjective methods involve significant human judgment based on the examiner's training and experience.<sup>10</sup> Objective methods can be performed by either an automated system or by human examiners exercising little or no judgment.<sup>11</sup>
- b) To the extent you find that complex DNA mixture analysis is at least partially subjective, did the analysis in this case take into account that subjective methods are more vulnerable to human error, bias, and variations in performance by different analysts?<sup>12</sup> Specifically, consider the following:
- i. Cognitive bias includes the natural tendency of humans to be influenced by outside information and outside pressures.<sup>13</sup> Did the DNA analyst here take steps to avoid learning any information about the facts of the case, or to speak to others about their opinions of the case, that might have affected [his/her] DNA analysis before conducting [his/her] examination and documenting the results?<sup>14</sup>

---

<sup>10</sup> PCAST Report at 47 (Subjective methods “involve significant human judgment.”). *See also id.* at 76 (describing manual interpretation of complex mixtures as a “subjective” method).

<sup>11</sup> PCAST Report at 47 (“Objective techniques or methods consist of procedures that are each defined with enough standardize and quantifiable detail that they can be performed by either an automated system for human examiners exercising little or no judgment.”).

<sup>12</sup> PCAST Report at 49 (“subjective methods . . . are especially vulnerable to human error, inconsistency across examiners, and cognitive bias”); *id.* at 76 (discussing 2011 Dror & Hampikian study showing cognitive bias in DNA mixture interpretation).

<sup>13</sup> PCAST Report at 5 (“In the forensic feature-comparison disciplines, cognitive bias includes the phenomena that, in certain settings, humans . . . may also be influenced by extraneous information and external pressures about a case.”).

<sup>14</sup> PCAST Report at 31 (“Cognitive bias refers to ways in which human perceptions and judgments can be shaped by factors other than those relevant to the decision at hand. It includes ‘contextual bias,’ where individuals are influenced by irrelevant background information; ‘confirmation bias,’ where individuals interpret information, or look for new evidence, in a way that conforms to their pre-existing beliefs or assumptions; and ‘avoidance of cognitive dissonance,’ where individuals are reluctant to accept new information that is inconsistent with their tentative conclusion.”); *id.* at 32 (Several strategies have been proposed for mitigating cognitive bias in forensic laboratories, including managing the flow of information in a crime laboratory to minimize exposure of the forensic analyst to irrelevant contextual information . . . .); SWGDAM Guidelines at 5, 3.1.1 (“The laboratory shall establish guidelines to ensure that, to the extent possible, DNA typing results from evidentiary samples are interpreted before comparison with any known samples, other than those of assumed contributors.”).

- ii. Another example of cognitive-bias in the forensic testing field is the natural tendency of human examiners to focus on similarities between profiles, and downplay differences between profiles.<sup>15</sup> Did the analyst in this case limit the effect of such bias by first documenting what [he/she] believed were the genetic markers or “alleles” found in the crime scene mixture, and where alleles might be “dropping out” from the crime scene mixture (i.e. where the data may be incomplete) *before* [he/she] compared the detected alleles to the ones in the reference or known profile?<sup>16</sup>
- iii. Did the laboratory conduct any type of “blind” verification, meaning that it asked a second analyst, who did not already know what decision the first analyst reached, to also compare the known profiles to the mixture and decide the number of potential contributors, whether the known profiles are consistent with the mixture, and, if so, what the combined probability of inclusion or “CPI” is?<sup>17</sup>

*On whether the method’s error rate is known and sufficiently low to merit trust in the analyst’s opinion:*

- a) To determine whether a complex DNA analysis method is scientifically valid, you must also consider its error rate, or likelihood of producing an inaccurate conclusion.<sup>18</sup> In complex DNA mixture analysis accompanied by a CPI statistic, an inaccurate conclusion might be not only an erroneous determination that the suspect’s profile is consistent with being present in the

---

<sup>15</sup> PCAST Report at 5 (“In the forensic feature-comparison disciplines, cognitive bias includes the phenomena that, in certain settings, humans may tend naturally to focus on similarities between samples and discount differences . . . .”); American Bar Association Resolution 101C (adopted Feb. 6, 2012) at 11-12 (“The court should consider whether additional factors such as those set forth below might be especially important to a jury’s ability to fairly assess the reliability of and the weight to be given testimony on a particular issue . . . . The extent to which the particular forensic science technique or theory relies on human interpretation that could be tainted by error; [and] The extent to which the forensic science examination in this case may have been influenced by the possibility of bias.”).

<sup>16</sup> PCAST Report at 32 (“Several strategies have been proposed for mitigating cognitive bias in forensic laboratories, including . . . ensuring that examiners work in a linear fashion, documenting their finding about evidence from crime science *before* performing comparisons with samples from a suspect.”).

<sup>17</sup> *Cf.* PCAST Report at 90 (noting that the ACE-V method of latent print comparison is problematic in laboratories that do not conduct “blind” “independent examinations,” because “the second examiner knows the first examiner reached a conclusion of proposed identification, which creates the potential for confirmation bias.”); SWGDAM Guidelines at 3 (“Upon completion of the technical aspects of DNA analysis, DNA typing results must be verified and interpreted. The verification of the DNA typing results involves a review of peak designations and other software-generated information, as well as an evaluation of quality controls.”).

<sup>18</sup> PCAST Report at 6 (“Without appropriate estimates of accuracy, an examiner’s statement that two samples are similar—or even indistinguishable—is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact.”); *id.* at 53 (“[W]ithout appropriate empirical measurement of a method’s accuracy, the fact that two samples in a particular case show similar features has *no probative value*—and, as noted above, it may have considerable prejudicial impact because juries will likely incorrectly attach meaning to the observation.”).



mixture profile, but also an erroneous estimate of the statistical significance of a suspect's profile being consistent with a mixture profile – that is, an erroneous estimate of the chance that a random person selected from the population would also be consistent with the mixture profile by coincidence.

- b) [If admitted as an objective, scientific method]: A method like DNA analysis that purports to be “objective” might produce an inaccurate conclusion for two reasons. First, the consistency between the suspect's profile and the mixture might be coincidental, meaning that the suspect is not actually a contributor to the sample but his/her profile is only coincidentally consistent with the mixture. Second, the expert's conclusion that an individual is a potential contributor, or that the combined probability of inclusion is a certain number, might be the result of human or technical failures in the process, such as an erroneous interpretation, contamination, or the like.<sup>19</sup>
- c) All DNA analysis, like any human endeavor, is subject to error and has some error rate greater than zero.<sup>20</sup> Not even highly automated tests have a zero error rate.<sup>21</sup> The error rate of a DNA analyst's method cannot be inferred from the analyst's case work alone,<sup>22</sup> nor the analyst's expression of confidence about [*his/her*] opinion or about the accuracy of the field.<sup>23</sup> Instead, a DNA analysis method's error rate can only be determined from conducting scientific experiments, sometimes called “empirical testing.”<sup>24</sup>
- d) Has the kind of “empirical testing” described in paragraph (a) been conducted with respect to [*name of expert*]’s DNA analysis method?<sup>25</sup> If so, what is the method's error rate? That is, how often does the technique or method reach an incorrect conclusion?<sup>26</sup>

---

<sup>19</sup> PCAST Report at 50 (“For objective methods, the false positive rate is composed of two distinguishable sources – coincidental matches (where samples from different sources nonetheless have *features* that fall within the tolerance of the objective matching rule) and human/technical failures (where samples have features that fall outside the matching rule, but where a proposed identification was nonetheless declared due to a human or technical failure).”).

<sup>20</sup> PCAST Report at 29 (“All laboratory tests and feature-comparison analyses have non-zero error rates . . .”).

<sup>21</sup> PCAST Report at 30 (“Even highly automated tests do not have a zero error rate.”).

<sup>22</sup> PCAST Report at 33 (“[O]ne cannot reliably estimate error rates from casework because one typically does not have independent knowledge of the ‘ground truth’ or ‘right answer.’”).

<sup>23</sup> PCAST Report at 6 (“[A]n expert's expression of *confidence* based on personal professional experience or expressions of *consensus* among practitioners about the accuracy of their field is no substitute for error rates estimated from relevant studies.”).

<sup>24</sup> PCAST Report at 50 (false positive rate must be based on “empirical measurements”).

<sup>25</sup> PCAST Report at 76-78 (noting studies showing a large number of false inclusions in DNA mixture interpretation).

<sup>26</sup> See NIST, MIX13 Study at 30, 34 (noting high (92%) false positive rate among those examiners making a conclusive determination).

- e) To the extent the field of complex DNA mixture analysis, with an accompanying CPI statistic, has not conducted the necessary empirical testing to estimate the method's error rate, you may consider the absence of such testing in deciding how much weight, if any, you wish to give the examiner's opinion in this case.

*On whether the method is governed by scientific standards:*

- a) Does complex DNA mixture analysis have standards that have been developed by reputable scientific organizations?<sup>27</sup>

2. Factors related to the validity of the method as applied, to help you determine whether the method was reliably applied by the analyst:

- a) Are there laboratory notes that show that the DNA analyst in this case properly followed each of the steps for this method?<sup>28</sup>
- b) Were those laboratory notes actually made at the time the analyst was analyzing the mixture, rather than prepared after the analysis was completed?<sup>29</sup>

---

<sup>27</sup> American Bar Association Resolution 101C (adopted Feb. 6, 2012) at 11-12 (“The court should consider whether additional factors such as those set forth below might be especially important to a jury’s ability to fairly assess the reliability of and the weight to be given testimony on a particular issue . . . The extent to which the forensic science examination in this case uses operational procedures and conforms to performance standards established by reputable and knowledgeable scientific organizations.”); *see also* SWGDAM Guidelines, § 4C (CPI), Jan. 12, 2017.

<sup>28</sup> PCAST Report at 66 (“[V]alidity as applied requires that: (a) the forensic examiner must have been shown to be *capable* of reliably applying the method, . . . and must *actually* have done so, as demonstrated by the procedures actually used in the case, the results obtained, and the laboratory notes, which should be made available for scientific review by others . . . .”); American Bar Association Resolution 101C (adopted Feb. 6, 2012) at 11-12 (“The court should consider whether additional factors such as those set forth below might be especially important to a jury’s ability to fairly assess the reliability of and the weight to be given testimony on a particular issue . . . The extent to which the forensic science examiner followed or did not follow the prescribed scientific methodology during the examination.”); *id.* at 12 (“The extent to which the forensic science examiner in this case followed the prescribed operational procedures and conformed to the prescribed performance standards in conducting the forensic science examination of the evidence.”).

<sup>29</sup> PCAST Report at 99 (noting the important of ensuring that forensic “examiners . . . document[] their findings about evidence . . . *before* performing comparisons” with known samples); *see also* National Commission on Forensic Science, Recommendation to the Attorney General National Code of Professional Responsibility for Forensic Science and Forensic Medicine Service Providers (Adopted March 22, 2016) at. (Forensic examiners “must . . . [m]ake and retain contemporaneous, clear, complete, and accurate records of all examinations, tests, measurements, and conclusions, in sufficient detail to allow meaningful review and assessment by an independent professional proficient in the discipline.”); *cf.* ABA Standards for Testing and Interpretation of DNA Evidence 16-3.2(b) Commentary (“The lack of contemporaneously prepared case notes can result in erroneous results.”); OFFICE OF INSPECTOR GENERAL, DEP’T OF JUSTICE, THE FBI LABORATORY: A REVIEW OF PROTOCOL AND PRACTICE VULNERABILITIES 107 (2004) (“If staff members are allowed to delay recording observations and test results until after they have examined all the items for a case or have completed all of their work for the day, their documentation may not be fully accurate. Also, staff members may be unduly influenced by protocol requirements when relying on memory, and document what they know should have occurred when their recollection is vague.”).

- c) Were there any times where the analyst did not follow the laboratory's protocols and procedures, and were they documented and explained to your satisfaction?
- d) If you determined that complex DNA mixture analysis has standards that have been developed by reputable scientific organizations (see above), did the complex DNA mixture analysis in this case conform to those standards?
- e) Was the expert's opinion in this case based on any assumptions, such as an assumption about the number of contributors, that was not justified based on the empirical testing that has been done of the method he/she used?

3. Factors related to whether the expert in this case is sufficiently qualified to render an accurate and helpful opinion based on the method:

*On whether the examiner's skills in applying the method have been sufficiently tested:*

- a) Has the DNA analyst in this case taken proficiency tests, that is, "test[s] that measures how often this examiner reaches the correct answer"?<sup>30</sup>
- b) If so, were those tests conducted by a third party who had no incentive to skew the performance, and did the examiner know he or she was being tested?<sup>31</sup>
- c) Did the testing used test this examiner's capacity to replicate the complexity of the task in this case, or were the samples and profiles in the test easier than the tasks in this case?<sup>32</sup>

*On the expert's individual qualifications:*

- a) Is the expert certified in complex DNA mixture analysis by a recognized organization in the field?<sup>33</sup> Does that organization evaluate examiner skill

---

<sup>30</sup> PCAST Report at 57 ("[T]he *only* way to establish scientifically that an examiner is capable of applying a foundationally valid method is through appropriate empirical testing to measure how often the examiner gets the correct answer. Such empirical testing is often referred to as 'proficiency testing.'"); *see also id.* at 76-78 (noting the dismal track record of analysts in various studies).

<sup>31</sup> PCAST Report at 57 ("To ensure integrity, proficiency testing should be overseen by a disinterested third party that has no institutional or financial incentive to skew performance.").

<sup>32</sup> PCAST Report at 57 ("Proficiency testing should be performed under conditions that are representative of casework and on samples, for which the true answer is known, that are representative of the full range of sample types and quality likely to be encountered in casework in the intended application. ").

<sup>33</sup> *See, e.g.,* American Board of Criminalistics diplomate certification program (molecular biology subject area exam; comprehensive criminalistics exam).

through testing?<sup>34</sup> Has the examiner attempted to become certified but failed? Is the testing based on reality, in the sense that it tests the ability to accurately analyze the type of samples the examiner faces in real casework?<sup>35</sup>

- b) Has the expert received training on the standards developed by a reputable scientific organization?<sup>36</sup>

#### 4. The DNA laboratory's qualifications

- a) Is the laboratory or facility in which the DNA analysis was done accredited?<sup>37</sup>
- b) Does that laboratory or facility have Standard Operating Procedures for DNA analysis?<sup>38</sup>
- b) Does that laboratory or facility have a system for recording and reporting errors or mistakes?

---

<sup>34</sup> American Bar Association Resolution 101C (adopted Feb. 6, 2012) at 11-12 (“The court should consider whether additional factors such as those set forth below might be especially important to a jury’s ability to fairly assess the reliability of and the weight to be given testimony on a particular issue . . . Whether the forensic science examiner has been certified in the relevant field by a recognized body that evaluates competency by testing.”)

<sup>35</sup> PCAST Report at 57 (“Proficiency testing should be performed under conditions that are representative of casework and on samples, for which the true answer is known, that are representative of the full range of sample types and quality likely to be encountered in casework in the intended application.”).

<sup>36</sup> American Bar Association Resolution 101C (adopted Feb. 6, 2012) at 11-12 (“The court should consider whether additional factors such as those set forth below might be especially important to a jury’s ability to fairly assess the reliability of and the weight to be given testimony on a particular issue . . . The extent to which the forensic science examination in this case uses operational procedures and conforms to performance standards established by reputable and knowledgeable scientific organizations; [and] The extent to which the forensic science examiner in this case followed the prescribed operational procedures and conformed to the prescribed performance standards in conducting the forensic science examination of the evidence.”).

<sup>37</sup> American Bar Association Resolution 101C (adopted Feb. 6, 2012) at 11-12 (“The court should consider whether additional factors such as those set forth below might be especially important to a jury’s ability to fairly assess the reliability of and the weight to be given testimony on a particular issue . . . Whether the facility is accredited by a recognized body if accreditation is appropriate for that facility.”).

<sup>38</sup> American Bar Association Resolution 101C (adopted Feb. 6, 2012) at 11-12 (“The court should consider whether additional factors such as those set forth below might be especially important to a jury’s ability to fairly assess the reliability of and the weight to be given testimony on a particular issue . . . The extent to which the forensic science examination in this case uses operational procedures and conforms to performance standards established by reputable and knowledgeable scientific organizations.”). *See also* SWGDAM Guidelines, Jan. 12, 2017, at 3 (“[L]aboratories that analyze DNA samples for forensic casework purposes are required by the Quality Assurance Standards for Forensic DNA Testing Laboratories to establish and follow documented procedures for the interpretation of DNA typing results and reporting.”).

## 2. Preliminary Instruction to Address the CSI Effect

Using a local instruction on guarding against the influences of media during a trial, you can create a proposed instruction about the portrayal of DNA analysis on television that may help dampen the impact of the misinformation jurors have received through television. For example:

Just as I have advised you to avoid any media or publicity about this case, the effort to ensure that you decide this case solely on the evidence presented in this courtroom also puts a limit on getting information from television entertainment. This applies to popular TV shows such as CSI and NCIS, which present the use of forensic methods such as DNA analysis to resolve criminal investigations. These and other similar shows will leave you with a false understanding of complex DNA mixture analysis. As far as this case is concerned, you are not prohibited from watching such shows but you may not rely on any of the information from these CSI-type programs. These programs are works of fiction. The programs often present all forensic methodologies as being scientifically valid. In this case, the scientific validity and reliability of the forensic method used – DNA analysis involving a mixture with potentially three or more contributors, accompanied by a so-called CPI match statistic – will be an issue contested by the parties. I will give you further instructions at the close of the evidence to help you evaluate the extent to which the DNA analyst’s conclusion was based on a scientifically valid and reliable method. Thus, in this case you must put aside anything you think you know about DNA evidence based on what you have seen on television. Instead you must rely on the evidence and the testimony presented in this case and follow the instructions I will provide you on the law and the assessment of evidence.

## 3. Jury Instruction Explaining False Positive Rate

For disciplines that have a reportable false positive rate, attorneys should be requesting an instruction that explains these numbers to the jury. As justification for such an instruction, you can argue that jurors are notoriously unable to make sense of numerical statements of the statistical significance of a pattern evidence match without further guidance.<sup>39</sup>

Unfortunately, no published study reports an error rate for complex DNA mixture interpretation with an accompanying CPI statistic, although ideally you have elicited during trial the fact that there have been several studies (such as Dror & Hampikian 2011) anecdotally showing erroneous conclusions by analysts.

The only large-scaled study that *could* be cited and translated into an error rate appears to be NIST’s “MIX13” Study (which is not yet published but referenced in several NIST presentations

---

<sup>39</sup> See, e.g., William C. Thompson & Eryn J. Newman, *Lay Understanding of Forensic Statistics: Evaluation of Random Match Probabilities, Likelihood Ratios, and Verbal Equivalents*, 39 L. & HUM. BEHAV. 332, 332 (2015) (“[L]ay jurors may be confused or misled by forensic scientists’ traditional characterizations of their findings.”); cf. William C. Thompson et al., *How the Probability of a False Positive Affects the Value of DNA Evidence*, 48 J. FORENSIC SCI. 1 (2003) (noting that explaining the import of a false positive rate to jurors is key to accurately assessing the probative value of a DNA profile match).

available on the Internet). The MIX13 results, to the extent they are available, suggest disturbing false positive rates in various mixture scenarios.

Nonetheless, we do not include estimates here of a false positive rate based on MIX13, because it is not yet clear, based on the information NIST has provided, what the “denominator” of the false positive rate would be (e.g. 1 in 100, 1 in 200, 1 in 300). Although 108 laboratories participated in the study (suggesting that 1 false positive would be 1 in 108), there are two additional unknown factors. First, we don’t know if there were multiple participants from each laboratory, which would make the denominator larger than 108. Second, for some of the results, we don’t know how many laboratories actually made a match call, as compared to an “inconclusive” or “excluded” call, and the PCAST report suggests that the denominator for a false positive rate should only include match calls (so, if only 79 participants made a “match” determination, and 1 was a false positive, the reported false positive rate should be 1 in 79, not 1 in 108).<sup>40</sup>

Finally, even if we had a valid estimate for a denominator, we would want to account for the possibility that the true false positive rate is actually higher than the estimate, using what is called a “confidence interval.” PCAST reported the “upper bound” of a “95% confidence interval” for the false positive rates for disciplines like firearms. Working with your expert, you should consider the possibility of translating the NIST MIX13 study results into a false positive rate estimate, using the upper bound of the confidence interval as the number you report to the jury.

Here are the MIX13 results with respect to false positives:

- Case 2, involving a 3-person mixture with a low template profile with potential allelic dropout: 1 laboratory falsely included the known profile.<sup>41</sup> While there were 108 laboratories in the study, there may have been more than one analyst from each laboratory who participated. Also, the results do not specify how many participants actually made a match determination.
- Case 3, involving a 3-person mixture with a potential relative involved: 1 laboratory falsely included profile 3B (friend). In addition, only 73% of the laboratories made a conclusive determination.<sup>42</sup>
- Case 5, involving a 4-person mixture: 69% of participants falsely included an innocent individual in the mixture; 91% (or around 9 out of 10) of participants making a conclusive determination falsely included the innocent individual.

---

<sup>40</sup> MIX13 Study at 34. *See also* PCAST Report at 51-52 (“When reporting a false positive rate to a jury, it is scientifically important to calculate the rate based on the proportion of *conclusive* examinations, rather than just the proportion of all examinations. . . . [C]onsider an extreme case in which a method had been tested 1000 times and found to yield 990 inconclusive results, 10 false positives, and no correct results. It would be misleading to report that the false positive rate was 1 percent . . . . Rather, one should report that 100 percent of the conclusive results were false positives (10/10 examinations).”).

<sup>41</sup> *See* Michael Coble, *Interpretation Errors Detected in a NIST Interlaboratory Study on DNA Mixture Interpretation in the U.S.*, July 22, 2015, at 16.

<sup>42</sup> *Id.* at 22.

If you are able to elicit from your own expert or the government's expert some sort of false positive estimate, here is what a potential instruction might look like:

As I have instructed you, you may consider a method's error rate, as established through scientific testing of examiner accuracy, in determining the extent to which that method is scientifically valid and reliable. In this trial, you have heard one estimate of something called an "upper bound" on the "false positive rate" in complex DNA mixture interpretation accompanied by a CPI statistic, based on scientific testing. That estimate was given as "1 in X."

Let me explain in more detail what those numbers would mean, if you choose to accept them. A "false positive rate" is one type of error rate. It is the chance that a given DNA mixture interpretation has resulted in an incorrect conclusion by the analyst that Mr./Ms. [Defendant] is a potential contributor to the DNA mixture found at the scene, when in fact Mr./Ms. Defendant is excluded as a potential contributor. The "upper bound" of these false positive rates means that based on the studies, the false positive rates could be as high as 1 in X.

Now, as for what these specific numbers mean. A false positive rate of 1 in X means that there is a 1 in X chance that the analyst's conclusion that the defendant is a potential contributor to the mixture is incorrect. For every 100 analyst opinions that a particular person is a potential contributor to a complex mixture, we would expect about (100 divided by X) of them to be wrong.

#### **4. Jury Instruction Designed to Avoid Prosecutor's Fallacy**

This instruction is designed to educate jurors about the fallacy of the transposed conditional, or the "prosecutor's fallacy," as it is sometimes known.<sup>43</sup>

You have heard evidence in this case of a statistic called the "combined probability of inclusion," or "CPI" for short. The CPI means the chance that a random person selected from the population will, by sheer coincidence, have a DNA profile that is consistent with the mixture profile found at the crime scene. This statistic can be easily misunderstood by laypeople. A common fallacy, or logical mistake, that laypeople make with respect to this sort of statistic is to assume that the CPI is the same as the probability that the defendant is not a contributor to the DNA mixture.<sup>44</sup> In other words, some people, if they are told that the chance of a person matching by coincidence is 1 in 10,000, will take that to mean there is only a 1 in 10,000 chance that someone other than the defendant could have contributed to the mixture. But that is wrong. Instead, it means that we would expect every 1 in 10,000 people to have a profile consistent with the

---

<sup>43</sup> See generally Andrea Roth, *Safety in Numbers?: Deciding When DNA Alone Is Enough to Convict*, 85 N.Y.U. L. Rev. 1130 (2010); *McDaniel v. Brown*, 558 U.S. 120, 128 (2010).

<sup>44</sup> See *Brown*, 558 U.S. at 128 ("The prosecutor's fallacy is the assumption that the [RMP] is the same as the probability that the defendant was not the source of the DNA sample.").

| mixture purely by coincidence. In a population of 10 million people, we would expect a full 1,000 people to be consistent with the mixture purely by coincidence.

In this case, the prosecution asserts that the CPI is 1 in \_\_\_\_\_. That means that in a population of 300 million people, we would expect a full [300 million divided by CPI] people to be consistent with the mixture purely by coincidence. We don't know anything about the other people who likely are also consistent with the mixture, such as whether they were anywhere near the scene of the crime, or are ruled out by other factors, for example, gender or age.<sup>45</sup>

Your job, as a member of the jury, is to decide whether to credit the prosecution's expert testimony that Mr./Ms. [ ] was a contributor to the mixture and, if so, whether based on that DNA evidence, along with all the other evidence both sides have presented in the case, whether the prosecution has proven beyond a reasonable doubt that Mr./Ms. [ ] is guilty of the crime charged.

---

<sup>45</sup> The last 2 paragraphs are a slightly modified version of an instruction given in the United Kingdom to stop jurors from committing a fallacy in interpreting the random match probability. See Peter Donnelly, *Appealing Statistics*, SIGNIFICANCE, Mar. 2005, at 48 (discussing *R v. Adams*, [1996] 2 Crim. App. at 468 and noting that Court of Appeal now "advocate[s] that judges should summarise cases in the" following "way": "Suppose the match probability is 1 in 20 million. If you believe that number, then on average there will be 2 or 3 people in Britain whose DNA it could be, and probably no more than 6 or 7. . . . Now your job, as a member of the jury, is to decide, on the basis of the other evidence, whether or not you are satisfied that it is the person on trial who was the assailant, rather than one of the few other possible people who match. We don't know anything about the other people who match, although they are probably spread all over the UK, may have been nowhere near the scene of the crime, and some or all may also be ruled out by other factors, for example, gender or age.").



COUNTY COURT  
COUNTY OF ST. LAWRENCE STATE OF NEW YORK

---

THE PEOPLE OF THE STATE OF NEW YORK,

-against-

**DECISION & ORDER**

DNA Analysis Admissibility  
Ind. #: 2015-15

ORAL NICHOLAS HILLARY,

Defendant.

---

---

*Mary E. Rain, District Attorney, Canton, for the People.*

*William Fitzpatrick, District Attorney, Syracuse, for the People.*

*The Legal Aid Society, New York (Jessica Goldthwaite of counsel), for the Defendant.*

*The Legal Aid Society, New York (Richard Torres of counsel), for the Defendant.*

*Dumas & Narrow, Canton (Peter Dumas of counsel), for the Defendant.*

*Emery, Celli, Brinckerhoff & Abady, New York (Earl Ward of counsel), for the Defendant.*

*Siegel, Teitelbaum & Evans, LLP, New York (Norman Siegel of counsel), for the Defendant.*

*The Legal Aid Society, New York (Clinton Hughes of counsel), for the Defendant.*

*The Bronx Defenders, New York (Adnan Omar Sultan of counsel), for the Defendant.*

---

**CATENA, J.:**

A motion was filed by defendant to preclude the prosecution from offering expert testimony as to the use of, or any results produced by, the forensic software tool STRmix alleging that the use

of this software for probabilistic genotyping is not generally accepted in the relevant scientific and legal communities as required by *Frye v. United States*, 293 F. 1013 (see, *People v. Middleton*, 54 NY2d 42 [“the test is not whether a particular procedure is unanimously indorsed by the scientific community, but whether it is generally acceptable as reliable”]). The defendant also filed a motion to preclude the use of a modified random match probability statistic (“RMP”). The defendant also challenged the admissibility of the evidence. The people filed an affirmation in opposition to the motion dated June 24, 2016, and oral argument was held on Friday July 1, 2016. By decision and order dated July 11, 2016, this court granted a limited *Frye* hearing on the issue of whether STRmix can “generate results accepted as reliable within the scientific community generally” on extreme mixture ratios where the DNA from the minor contributor is low template. Trial foundation for the admissibility of the evidence was made part of the hearing (*People v. Wesley*, 83 NY2d 417, 429). The hearing was held on Monday July 25, 2016, at the courthouse in Canton, New York. Further oral argument was held on the record on August 17, 2016.

On October 24, 2011, the Potsdam Police Department received a call from a neighbor of the victim stating that she heard moans and the word “help” coming from the victim’s apartment. An officer of the Potsdam Police Department arrived at the apartment at approximately 5:16 p.m., knocked on the door, and heard what sounded like someone walking around the apartment. Shortly thereafter, the Officer entered the apartment with the landlord and found the victim unconscious in the bedroom. No one else was in the apartment and the victim was pronounced dead at 7:18 p.m. that evening. The cause of death was determined to be strangulation.

As part of the investigation, the police collected dozens of samples of DNA from multiple areas in the apartment including the body and clothing of the victim. The police also obtained a

sample of the defendant's DNA as he had lived briefly with the victim, having an intimate relationship with the victim's mother. Because this was a New York State Police investigation, the New York State Police crime lab processed the samples. The defendant was excluded from all samples taken at the apartment where comparisons could be made except for a DNA mixture profile from fingernail scrapings taken from the victim's left hand. Due to insufficient genetic information, the defendant could neither be included nor excluded as a possible contributor to the mixture.

Beginning in 2013, the New York State Police crime lab contacted Cybergentics, Inc., to run the data obtained from the fingernail scrapings through their probabilistic genotyping software program called TrueAllele. The results were inconclusive. Nonetheless, the defendant was indicted for the victim's murder in 2014 and the New York State Police crime lab sent its data at the behest of the prosecution to the Institute of Environmental Science and Research ("ESR") which ran it through their probabilistic genotyping software program called STRmix<sup>1</sup>. The People also requested their expert, Dr. John S. Buckleton, one of the developers of STRmix, to calculate a random match probability statistic which was presented in a report dated July 1, 2016.

Dr. Buckleton conceded at the hearing that no internal validation studies were performed by the New York State Police crime lab for the use of STRmix on casework samples developed at the

---

<sup>1</sup>District Attorney Fitzpatrick first contacted Dr. Buckleton by email on November 2, 2015, stating in pertinent part "I am currently assisting a colleague in prosecuting a murder case in Saint Lawrence County . . . On 10/24/2011 a 12 year old boy [] was strangled to death in his apartment in Potsdam NY. No physical evidence was discovered at the scene except for [] fingernail scrapings . . . Those scrapings were analyzed at the New York State Police crime lab and the DNA profile generated appears to be consistent with a mixture of two individuals with the victim as the major contributor and the obligate alleles (7) being consistent with the defendant. I am hoping that you or someone you recommend might be able to provide a statistical weight to the results using a likelihood ratio or some other method. I can provide the analyst's report and the electropherograms if you decide to take a look."

lab. As a result Dr. Buckleton was forced to pick and choose data from different “reliable sources” and input parameters into the program in such a way that he believed the system would tolerate<sup>2</sup>. The reason for this was because the New York State Police crime lab was not authorized by the New York State Commission on Forensic Science to generate data from DNA samples for STRmix

---

<sup>2</sup>Dr. Buckleton testified that “[a]ll labs in the United States of America are following the SWGDAM guidelines”. The Scientific Working Group on DNA Analysis Methods (SWGDAM) provides guidelines for validation of probabilistic genotyping systems such as STRmix. As stated in the guidelines, “[v]alidation is a process by which a procedure is evaluated to determine its efficacy and reliability for forensic casework and/or database analysis.” Further, “[i]nternal validation studies is the accumulation of test data within the laboratory to demonstrate that the established parameters, software settings, formulae, algorithms and functions perform as expected . . . In particular, complex mixtures and low-level contributors should be evaluated thoroughly during internal validation, as the data from such samples generally help to define the software's limitations, as well as sample and/or data types which may potentially not be suitable for computer analysis.”

STRmix has six laboratory specific parameters to determine prior to its use. Concerning one of these parameters called stutter ratios, Dr. Buckleton testified that he received data from the New York State Police crime lab to look at “forward stutter” which “did not serve the purpose” but nonetheless was used in the April 2016 analysis. Concerning “drop-in” rates, Dr. Buckleton stated that he inputted a zero drop-in rate as the crime lab did not “have a drop-in rate because they do not do low copy number on high sensitivity methods.” A “drop-in” rate greater than zero would have benefitted the defendant. And in his affidavit dated August 18, 2016, Dr. Buckleton further stated that he used a “drop-in rate” that was based partly on “our own experience.”

Dr. Buckleton further testified that “[b]est practice is clearly validation in the lab specifically, and that is -- the optimal way is the way we recommend. It's not available to me in this case because the New York State lab has not done the relevant validation. I have two options first with that. I have the option of doing the next best practice or not doing anything. I've done next best practice. I've done this a couple times before and I'm not pertaining it's best practice, and I have candidly acknowledged in my statement that exact fact. *What I've done is attempt to take data from different reliable sources that I think applies to the circumstance.* I'm also aware of the forgiveness of the system for slight inaccuracies in certain of the parameters, so I have input parameters in such a way that I believe the forgiveness of the system will tolerate any inaccuracies I've made” (emphasis supplied).

analysis<sup>3</sup>. And although ESR had performed the necessary internal validations to be accredited by the relevant agencies for the use of STRmix, those validations were specific to data generated by ESR. Here, the only data generated was from the New York State Police crime lab<sup>4</sup>.

“The New York State Commission on Forensic Science (the “Commission”) is the governmental body tasked with developing minimum standards and accreditation programs for all forensic laboratories in New York State. In addition, the Commission approves forensic laboratories to perform specific forensic methodologies. The Commission’s objectives are to increase and maintain the effectiveness, efficiency, reliability, and accuracy of forensic laboratories, ensure that forensic analyses are performed in accordance with the highest scientific standards practicable, and set forth minimum requirements for the quality and maintenance of equipment. The DNA Subcommittee of the New York Commission on Forensic Science (“DNA Subcommittee”) is the

---

<sup>3</sup>The lab was authorized to generate data for TrueAllele analysis. The People stated at the August 17, 2016, oral argument that “each individual laboratory has to go through essentially an accreditation process to determine whether or not . . . they have the trained [personnel] and expertise to accurately employ [STRmix]”. The People further discussed the Onondaga County crime lab stating “Onondaga is going through . . . their own internal validation process to make sure that their people are properly trained to use STRmix, get accurate results with blind proficiency testing and then ultimately will present their process to ASCLD/LAB . . . for the accreditation process. Then that laboratory . . . has to go through a two year accreditation process in front of the Forensic Science Commission for that discipline.” It is clear that the New York State Police crime lab had not gone through the accreditation process and, thus, did not have the “trained [personnel] and expertise to accurately employ” STRmix. The People conceded this stating “there [were] no internal validation studies by the State Police regarding STRmix.”

<sup>4</sup>The Federal Bureau of Investigation’s quality assurance standards for Forensic DNA Testing Laboratories define “Forensic DNA analysis” as “the process of identification and evaluation of biological evidence in criminal matters using DNA technologies.” The quality assurance standards further define “internal validation” as “the accumulation of test data within the laboratory to demonstrate that established methods and procedures perform as expected in the laboratory.” Quality Assurance Standard 8.1.3 requires that internal validation be performed for forensic casework analysis.

body appointed by the Commission to perform accreditation of all DNA laboratories in New York. Further, the DNA Subcommittee is charged with assessing all DNA methodologies proposed to be used for forensic analysis. It has the sole authority to grant, deny, review, or modify a DNA forensic laboratory accreditation, which the DNA Subcommittee exercises by issuing a binding recommendation to the Commission. While the Commission can request that the DNA Subcommittee reconsider its findings, the DNA Subcommittee is the final decision maker regarding laboratory accreditation” (*U.S. v. Morgan*, 53 F.Supp.3d 732; Executive Law 995-b[1])<sup>5</sup>.

The Commission requires laboratories to comply with the standards promulgated by the American Society of Crime Laboratory Directors/ Laboratory Accreditation Board (“ASCLD/LAB”) (9 NYCRR §6190.3[b]). These standards require that DNA mixture interpretation be based on validation data (*id.*). Here, by sending its raw data to ESR, an accredited lab for STRmix analysis, the People argue that the New York State Police crime lab could bypass Commission approval for its participation in the STRmix process inasmuch as it did not run the computer program. This minimizes the importance of raw data generation in the STRmix process and emphasizes the People’s heavy reliance on the expertise of Dr. Buckleton to account for any deficiencies in the data<sup>6</sup>. Such reliance appears contrary to Commission and ASCLD/LAB requirements (*see, State v.*

---

<sup>5</sup> “‘DNA testing methodology’ means methods and procedures used to extract and analyze DNA material, as well as the methods, procedures, assumptions, and studies used to draw statistical inferences from the test results” (Executive Law §995[3]).

<sup>6</sup>As the People stated at the August 17, 2016, oral argument “[Dr. Buckleton]’s just saying that, look, I’ve got this raw data from the State Police. There are some things I don’t know, so what I’m doing, what I, John Buckleton, am doing is I am taking the most stringent, conservative approach to my input of the STRmix data.”

*Wakefield*, 57 Misc.3d 850; see, *People v. Vincent Bullard-Daniel*, Co Ct. Niagara County, March 10, 2016, Murphy, J., Ind. No. 2015-88)<sup>7</sup>.

As stated by this court in its earlier order granting the pre-trial hearing herein, “the test pursuant to *Frye v. United States* (293 F 1013) poses the . . . question of whether the accepted techniques, when properly performed, generate results accepted as reliable within the scientific community generally . . . The issues of proper foundation and of the adequacy of laboratory procedures [] are not before [the Court at a *Frye* hearing]” (*People v. Wesley, supra*). To that end, this court notes that the New York State Commission on Forensic Science issued a binding recommendation for use of STRmix in the analysis of DNA profiles upon recommendation from its DNA Subcommittee (Executive Law §995-a; *People v. Wakefield, supra* at 856 [approval by the Commission constitutes general acceptance]). And STRmix was found by a New York State court after a *Frye* hearing to be generally accepted in the relevant scientific community (*People v. Vincent Bullard-Daniel, supra*). Based upon a review of the record, this court finds that STRmix has been developmentally validated and is generally accepted as reliable within the scientific community (*id.*; *People v. Muhammad*, 14<sup>th</sup> Cir. Ct., Muskegon Co, Dec. 15, 2015). Issues concerning the manner

---

<sup>7</sup>ASCLD/LAB accreditation requirement 5.4.5.2 from ISO/IEC 17025:2005:

“Procedures for DNA profile interpretation must be based on validation data. The interpretation of a DNA profile containing a mixture of two or more individuals must be guided by a procedure that includes specific defined steps that will enable different analysts in the same laboratory to reach the same conclusion; and a competent person from outside the laboratory using the same procedure to understand how the conclusion was reached. DNA mixture interpretation procedures must be tested on mixture profiles from known contributors representing the range of mixture types (e.g., different numbers of contributors, mixture proportions, and template quantities) to which the procedure will be applied in casework. The results of this validation must be used to define the capabilities and limitations of the procedure and to verify that it produces the expected results (e.g., inclusions and exclusions).”

in which STRmix accounts for stochastic effects in its probability computations where mixture ratios are extreme and the minor contributor's DNA is low template goes to weight (*People v. Debraux*, 50 Misc.3d 247; *People v. Megnath*, 27 Misc.3d 405).

“The issue [now] shifts to a second phase, admissibility of the specific evidence--i.e., the trial foundation--and elements such as how the sample was acquired, whether the chain of custody was preserved and how the tests were made . . . Once *Frye* has been satisfied, the question is ‘whether the accepted techniques were employed by the experts in this case’. The focus moves from the general reliability concerns of *Frye* to the specific reliability of the procedures followed to generate the evidence proffered and whether they establish a foundation for the reception of the evidence at trial. The trial court determines, as a preliminary matter of law, whether an adequate foundation for the admissibility of this particular evidence has been established” (*People v. Wesley*, *supra* at 428-429). Here, the lack of internal validation by the New York State Police crime lab, as candidly admitted by Dr. Buckleton, precludes use of the STRmix results (*id.*; *see, State v. Wakefield*, *supra*; *see, People v. Vincent Bullard-Daniel*, *supra*).

Concerning RMP, while this court finds that it has been generally accepted as reliable within the scientific community under certain circumstances, the results produced in this case are unreliable based upon Dr. Buckleton's testimony that it cannot adequately account for the absence of defendant's alleles in the composite profile. As Dr. Buckleton stated, “the exact difficulty that we've come upon in this case and certain circumstances that the *Random Match Probability is not conservative and doesn't do a fair job for the defendant*, and this is one of those circumstances. The specific diagnostic is called a -- it's called a non-major allele between the profile and the accused. And we have a number of those where the accused has an allele that is not seen in the profile and



Random Match Probability is incapable of [punishing] the statistic for non-matches . . . So, in fact, it is the exact weakness of the 2p rule that has motivated me to make a probabilistic genotyping system” (emphasis supplied). Further, the defendant’s expert, Dr. Dan E. Krane, testified that “[t]he Random Match Probability is and has been considered generally accepted in many circumstances. The particulars of the evidence sample in this case do not fit within the category of those that would cause to be generally accepted. If I can put it just a different way. I’ve testified for many years that there is no generally accepted means of attaching a reliable statistical weight to a mixed sample, such as the evidence sample in this case where drop-out may have occurred, which, again, seems very likely to have occurred with the evidence sample in this case. So, on those two counts it would be quite inappropriate to rely upon a Random Match Probability approach to generate a statistical weight.”

Given Dr. Krane’s testimony, the use of RMP in this case where the People’s own expert witness testified that it “*is not conservative and doesn't do a fair job for the defendant*” must be precluded as unreliable (*People v. Wesley, supra*). In any event, even if this court were to agree with the prosecution that “[d]efendant’s concern with the methodology used by Dr. Buckleton in calculating the statistic goes to ‘the weight of the evidence, not its admissibility’”, to allow such evidence would be unduly prejudicial to the defendant (*People v. Morris*, 21 NY3d 588 [“Weighing the evidence’s probative value against its potential prejudice to the defendant is a matter of discretion for the trial court”]; *People v. Caban*, 14 NY3d 369 [“Evidence, though relevant, may be excluded where ‘it’s probative value is substantially outweighed by the danger that it will unfairly prejudice the other side or mislead the jury’”], quoting *People v. Scarola*, 71 NY2d 769, 777). In determining the motions herein, this court is reminded that “[f]orensic DNA analysis should be governed by the

highest standards of scientific rigor in analysis and interpretation” (*People v. Wesley, supra*)(Kaye, J., concurring). Neither the STRmix nor the RMP results may be used in this case (*id.*).

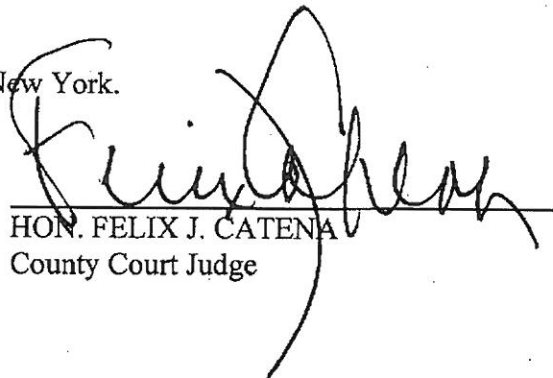
It is, therefore,

**ORDERED** that the defendant’s motion to preclude the prosecution from calling an expert witness to testify on their direct case regarding any conclusion reached by the use of STRmix is granted as the prosecution cannot lay a foundation for the introduction of evidence that had not been internally validated; and it is further

**ORDERED** that the defendant’s motion to preclude the prosecution from offering expert testimony as to any statistical results obtained by using the random match probability on the composite minor component of mixture is granted.

The above constitutes the decision and order of this Court.

Signed this 26<sup>th</sup> day of August, 2016, at Fonda, New York.



HON. FELIX J. CATENA  
County Court Judge

ENTER: