# SUPERIOR COURT FOR THE DISTRICT OF COLUMBIA

**Criminal Division – Felony Branch**

**UNITED STATES OF AMERICA**

**v.**

**XXXXX**

**:**

**:**

**:**

**:**

**:**

**Criminal Case No. XXXX Honorable Judge XXXX Next Hearing Date:**

**MOTION *IN LIMINE* TO EXCLUDE EXPERT TESTIMONY ON LATENT PRINTS OR IN THE ALTERNATIVE, TO LIMIT SUCH EXPERT TESTIMONY**

Mr. XXXX XXXX, through undersigned counsel, hereby moves this Honorable Court, pursuant to [Federal Rule of Evidence 702 (Rule 702)] [*Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923), *People v. Kelly*, 17 Cal.3d 24 (Cal. 1976) and *Sargon Enterprises, Inc. v. University of Southern Cal.*, 55 Cal. 4th 747 (Cal. 2012)] [DC only:, as adopted by the District of Columbia Court of Appeals in *Motorola Inc. v. Murray*, 147 A.3d 751 (D.C. 2016) (en banc)], to exclude expert testimony related to latent fingerprint comparison in this case or, in the alternative, to limit the scope and purported level of certainty of the expert’s opinions as detailed below.

Expert testimony declaring a “match” between known and latent fingerprints based on the government’s so-called “ACE-V” method, described in more detail below, is not [sufficiently reliable, either in theory or as applied by the examiner in this case, to be admissible under *Daubert*] [generally accepted by the relevant scientific community – which extends beyond the community of forensic examiners – and is thus inadmissible under *Frye*.] To be sure, expert testimony comparing known and latent fingerprints has been admitted by various courts in the United States for decades. But this precedent does not take into account recent reports by distinguished scientific and expert groups that explain more fully why this Court should exclude the testimony in this case as scientifically invalid. Indeed, to the extent the method’s accuracy as a tool of identification has been subject to empirical testing, the estimated potential “false

positive” rates (the rates at which the method erroneously calls a “match” between a known and latent print) in the only two appropriately designed “black box” studies are as high as 1 in 18.[1](#_bookmark0) In fact, “empirically estimated false positive rates are *much higher* than the general public (and, by extension, most jurors) would likely believe based on longstanding claims about the accuracy of fingerprint analysis.” President’s Council of Advisors on Science and Technology (PCAST), *Forensic Science in Criminal Courts: Ensuring Validity of Feature-Comparison Methods* 95 (Sept. 20, 2016) [hereinafter “PCAST Report”] (emphasis in original) (footnotes omitted).[2](#_bookmark1)

Based on these recent developments, the lack of appropriate empirical testing associated with ACE-V, and the false positive rates associated with the few appropriately designed studies thus far, this Court must exclude the expert testimony as [insufficiently reliable][not generally accepted]. If this Court were nonetheless to allow such testimony, it must at the very least (1) disallow the expert from testifying to a “match” between, or common source of, the latent print and suspect print, instead limiting the expert to describing similarities and differences between prints; and (2) require the expert to acknowledge before the jury that the level of certainty of a purported “match” would be limited by the most conservative reported “false positive rate” in appropriately designed empirical studies thus far, which is 1 in 18. *See* PCAST Report at 95-96, 101-02.

1 *See* President’s Council of Advisors on Science and Technology (PCAST), *Forensic Science in Criminal Courts: Ensuring Validity of Feature-Comparison Methods* 95 (Sept. 20, 2016) [hereinafter “PCAST Report”]. The PCAST reviewed a number of studies of latent fingerprint analysis, *id.* at 91-97, and found that “[r]emarkably, there have been only two black-box studies that were intentionally and appropriately designed to assess validity and reliability.” *Id.* at 91. These are an FBI study published in 2011 and a 2014 study by the Miami-Dade Police Department Forensic Services Bureau. *Id*. at 91, 94. The FBI study found a false positive rate as high as 1 in 306 while the Miami-Dade study yielded a false positive rate of 1 in 18. *Id.* at 96. Notably, earlier studies, even though not designed to evaluate validity and reliability, also found high false positive rates. *Id.*

2 The PCAST Report was published in September 2016 at the request of then-President Obama. PCAST Report at 1. The Report reviewed several fields of forensic science, including latent fingerprint analysis, for the purpose of strengthening the various fields and clarifying the requirements for foundational validity and validity as applied. *Id.* at 4.

# BRIEF SUMMARY OF THE GOVERNMENT’S PROFFERED EXPERT TESTIMONY IN THIS CASE

Mr. XXXX XXXX is charged with XXXX on XXXX. The government seeks to introduce the expert testimony of XXXX, a latent fingerprint examiner, regarding a latent fingerprint recovered from XXXX at XXXX location. Mr. XXXX anticipates that the examiner will attempt to testify that, in his/her expert opinion, Mr. XXXX’s [right/left thumb, index,

middle, ring, pinky or palm] print is the source of the latent print left [replace with whatever the match statement will be testified to by the examiner.] on the XXXX. Exhibit A at [page], Fingerprint Report. Mr. XXXX further anticipates that the basis for this expert opinion will be the so-called “ACE-V” Method (Analysis, Comparison, Evaluation, and Verification). National Institute of Standards and Technology (NIST), Expert Working Group on Human Factors in Latent Print Analysis, *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach* 3 (2012) [hereinafter “NIST Report”].

Fingerprint examiners in the United States have used the “ACE-V” method for over fifty years, with surprisingly little change in the basic methodology. *See* National Research Council of the National Academy of Sciences, Committee on Identifying the Needs of the Forensic Science Community, *Strengthening Forensic Science in the United States: A Path Forward* 137 (2009) [hereinafter “NAS Report (2009)”]; NIST Report at 3; PCAST Report at 89. In the first step, “analysis” **(A)**, the examiner purports to study the latent (unknown) print and assess the quality and quantity of detail present. NIST Report at 3. Details might include both “class” characteristics—characteristics shared or common within certain subgroups of prints—and “individual” characteristics or “minutiae”—characteristics not known to be shared or common within subgroups. S.H. James, *Forensic Science* 4th Edition, CRC Press at 330. The examiner

then “compares” **(C)** the latent print to a “known” print of a suspect.[3](#_bookmark2) *See* PCAST Report at 89. Based on his/her subjective judgment and experience, the examiner documents what s/he believes to be similarities and differences between the latent and known print(s). NIST Report at

3. To avoid contaminating the comparison stage with cognitive bias, the examiner would have to fully analyze and document the characteristics of the latent print before comparing it to any known suspect print or computer-generated list of known prints. *Id.*; PCAST Report at 99-100. The process by which an examiner fully documents the characteristics of a latent print prior to conducting the comparison is known as “linear ACE-V.” PCAST Report at 17. Nonetheless, many examiners fail to conduct such independent documentation at the analysis stage. *See* discussion *infra* at II.C.

The examiner then “evaluates” **(E)** what s/he views as the similarities and differences between the prints and, again based on his/her subjective judgment and experience, decides his/her level of certainty about whether the prints have a common source and, if so, how to convey his/her opinion as to whether the two come from the same source. *Id.* While fingerprint examiners in past years routinely made grandiose and unsupportable claims of a “zero error rate” and individualization “to the exclusion of all other prints in the world” and such, *see* NAS Report (2009) at 43-44, 104, 142, recent scrutiny and demonstrated false positives (such as the Mayfield case, discussed *infra*) have caused a modest and laudable shift in latent print expert claims. Still, Mr. XXXX expects that the government expert in this case will testify that, in his/her expert opinion, the prints come from the same source or “match.” And lest there be any confusion, “when a latent print examiner testifies that two impressions ‘match,’ they are communicating the

3 If no suspect has been identified, the examiner may compare the latent print to a computer-generated list of potentially similar reference prints housed in a database of known fingerprints. PCAST Report at 89.

notion that the prints could not possibly have come from two different individuals.” NAS Report (2009) at 141-42.

Finally, in most latent print analysis, a second examiner will “verify” **(V)** the initial examiner’s analysis. The verification step could theoretically involve a wholly independent and “blind” comprehensive analysis of the latent print, comparison to the known prints, and evaluation by a second examiner. Unfortunately, most “verification” in common latent print examiner practice consists merely of a dependent, non-blind review, meaning that the second examiner sees the analysis and conclusions drawn by the first examiner and knows which examiner s/he is reviewing. NIST Report at 3. *See also* NAS Report (2009) at 138.

# THE EXPERT LATENT PRINT TESTIMONY IN THIS CASE MUST BE EXCLUDED UNDER [*DAUBERT*][*FRYE*]

* 1. [**The *Daubert/*Federal Rule of Evidence 702 Standard]**

Latent print expert testimony is inadmissible unless it is based on a method that is both “scientifically valid” and that “properly can be applied to the facts in issue” in a particular case. *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579, 592 (1993). Thus, *Daubert* requires, as codified in the rules of evidence, *both* that an expert’s testimony be the “product of reliable principles and methods,” *and* that the expert “reliably applied the principles and methods to the facts of the case[.]” Fed. R. Evid. 702(c), (d)*.* While the *Daubert* Court made clear that the ultimate determination of “evidentiary relevance and reliability” is a legal question to be determined by the trial judge, *Daubert*, 509 U.S. at 589, it also made clear that evidentiary reliability in the context of forensic expert testimony is determined by the underlying method’s “scientific validity.” *Id*. at 594-95.

A method’s “scientific validity” rests on a number of factors identified by the scientific community itself. The *Daubert* Court looked to literature on the scientific method to offer some non-exhaustive factors that a trial judge might consider in its 702 analysis, including: whether the scientific theory can and has been tested, whether it has been “subjected to peer review and publication,” the known or potential error rate, whether there are standards that control operation, and whether the method has been generally accepted in the relevant scientific community. *Id.* at 593-94. More recently, the recent PCAST report acknowledged *Daubert*’s charge that evidentiary reliability for Rule 702 purposes turns on a method’s scientific validity, and explained that a method is not “scientifically valid” unless it has both “*foundational validity*” and “*validity as applied*.” PCAST Report at 4 (emphasis added).

A method is not “foundationally valid” unless it has been “shown, based on empirical studies, to be *repeatable, reproducible,* and *accurate*, at levels that have been measured and are appropriate to the intended application.” *Id.*[4](#_bookmark3) To show foundational validity, the proponent of a method must establish two elements:

* + 1. 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”).
		2. Empirical measurements, from multiple independent studies, of (a) the method’s false positive rate – that is, the probability it declares a proposed identification between samples that actually come from *different* sources and (b) the method’s sensitivity – that is, probability that it declares a proposed identification between samples that actually come from the *same* source.

4 The PCAST report further clarified that “repeatable” means “that, with known probability, an examiner obtains the same result, when analyzing samples from the same sources”; “reproducible” means “that, with known probability, different examiners obtain the same result, when analyzing the same samples”; and “accurate” means “that, with known probabilities, an examiner obtains correct results both (1) for samples from the same source (true positives) and (2) for samples from different sources (true negatives).” PCAST Report at 47.

PCAST Report at 48. For a “subjective” method like ACE-V – which is based on examiner judgment and experience rather than actual population data that could be used to produce objective estimates of the likelihood that two patterns have a common source – the process “must be treated as a kind of ‘black box.’” *Id.* at 49.

The foundational validity of a subjective “black box” method like ACE-V is not inherently impossible to establish, but requires “careful scrutiny” because of its “heavy reliance on human judgment” and special “vulnerab[ility] to human error, inconsistency across examiners, and cognitive bias.” *Id.* Because the “black box in the examiner’s head” cannot be examined directly, the foundational validity of subjective methods “can be established *only* through empirical studies of examiners[’] performance to determine whether they can provide accurate answers,” i.e., through appropriately designed “black box” studies. *Id*. *See also* President’s Council of Advisors on Science and Technology, *An Addendum to PCAST Report on Forensic Science in Criminal Courts*, Jan. 6, 2017, at 1 (“[T]he *only* way to establish the scientific validity and degree of reliability of a *subjective* forensic feature-comparison method – that is, one involving significant human judgment – is to test it *empirically* by seeing how often examiners actually get the right answer.”).

In turn, a method is not valid “as applied” unless it “has been reliably applied *in practice*.” *Id.* (citing Fed. R. Evid. 702(d) as consistent with this scientific principle). To show validity as applied, the proponent of expert testimony applying a method must establish two elements:

1. **The forensic examiner must have been shown to be *capable* of reliably applying the method and must *actually* have done so**. Demonstrating that an examiner is *capable* of reliably apply the method is crucial – especially for subjective methods, in which human judgment plays a central role. From a scientific standpoint, the ability to apply a method reliably can be demonstrated only through empirical testing that measures how often the expert reaches the correct answer. (Proficiency testing[)] . . .

Determining whether an examiner has *actually* reliably applied the method requires that the procedures actually used in the case, the results obtained, and the laboratory notes be made available for scientific review by others.

# Assertions about the probability of the observed features occurring by chance must be scientifically valid.

* 1. The forensic examiner should report the overall false positive rate and sensitivity for the method . . . and should demonstrate that the samples used in the foundational studies are relevant to the facts of the case. . . .
	2. Where applicable, the examiner should report the random match probability based on the specific features observed in the case.
	3. An expert should not make claims or implications that go beyond the empirical evidence and the applications of valid statistical principles to that evidence.

PCAST Report at 56 (emphases and bold in original).

1. **[The *Frye* Standard]**

Latent print expert testimony is inadmissible unless it is based on a method that is “generally accepted” in the relevant scientific community *both* as reliable in principle for its stated purpose, *Frye*, 293 F. at 1014; *People v. Kelly,* 17 Cal.3d 24, 31 (1976), and as reliable in terms of the procedures followed in the particular case at hand. *Kelly*, 17 Cal.3d at 30 (holding that under *Frye*, the witness must be a properly qualified expert on the subject; and the person performing the test in the particular case must use correct and accepted scientific procedures). The *Kelly/Frye* rule is conservatively “austere” in general, *Daubert*, 509 U.S. at 589, and courts must be “particularly careful to scrutinize the general acceptance of the technique” where it is, like latent print analysis, an “identification technique [that] is offered to identify the perpetrator of a crime.” *Kelly*, 17 Cal.3d at 32.

The relevant scientific community for purposes of the *Kelly/Frye* rule is not the community of forensic examiners, whose very *raison d’etre* is the admission of forensic expert testimony in courts of law and who are often affiliated directly or indirectly with law enforcement. *See, e.g.*, PCAST Report at 42 n.90; *Kelly*, 17 Cal.3d at 38-39 (concluding that a

law enforcement officer without sufficient academic training and who has a vested interest in the admissibility of the discipline cannot supply evidence of general acceptance); *People v. Leahy* (1994) 8 Cal.4th 587, 609 (“Consistent 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*.”); *United States v. Porter*, 618 A.2d 629, 634-35 (D.C. 1992) (rejecting limitation of scientific community to forensic scientists for *Frye* purposes). Instead, the relevant community for assessing the reliability of feature-comparison sciences includes metrologists, statisticians, population geneticists, and other physical and life scientists from the derivative sciences and disciplines on which a specific forensic method is based.

Nor does longstanding acceptance of a method by courts establish “general acceptance” under *Kelly/Frye*. As an initial matter, the reliance on past court precedent to rubber-stamp a method as generally accepted creates a feedback loop where precedent justifies itself into the indefinite future, with no chance to scrutinize methods based on a more thoughtful or comprehensive review by future judges. *See, e.g.*, *People v. Leahy*, 8 Cal.4th 587, 606 (1994) (“To hold that a scientific technique could become immune from *Kelly* scrutiny merely by reason of long-standing and persistent use by law enforcement outside the laboratory or the courtroom, seems unjustified.”). Second, reliance on precedent to establish general acceptance denies the reality that the scientific community’s views change, often dramatically, over time based on new understandings, new research, and even new views toward the critical nature of empirical studies to the validity of a method. *See, e.g.*, *Kelly*, 17 Cal.3d at 32 (precedent must give way “new evidence . . . reflecting a change in the attitude of the scientific community.”). Indeed, trial courts have an affirmative duty to revisit the admissibility of previously accepted techniques based on

new scientific developments and research. *See, e.g.*, *Motorola Inc.*, 147 A.3d at 758 (“What the court may not do is reflexively admit expert testimony because it has become accustomed to doing so under the *Dyas/Frye* test.”). For precisely the same reasons, a technique cannot escape scrutiny under *Kelly/Frye* simply by the logic that it has been accepted by courts and is therefore no longer “novel.”

Additionally, any witness(es) called by the government in an effort to establish general acceptance must be impartial and must have “sufficient academic and professional credentials… to understand both the scientific principles involved and any difference of view concerning the method’s reliability.” *People v. Brown*, 40 Cal.3d 512, 530 (Cal. 1985). The witness should not have a personal or professional stake in the acceptance of the technique; individuals who stand to gain from admissibility “may be too closely identified with the endorsement (of the technique) to assess fairly and impartially the nature and the extent of any opposing scientific views.” *Kelly*, 17 Cal.3d at 38.

# Expert Testimony in This Case Based on ACE-V Is Inadmissible Because It Is Not Based on a “Foundationally Valid” Method

The first reason the expert testimony in this case is inadmissible under *Daubert* is that it is not “foundationally valid,” meaning that even in principle, the method has not “be[en] shown, based on empirical studies, to be *repeatable, reproducible*, and *accurate*, at levels that have been measured and are appropriate to the intended application.” PCAST Report at 4. Specifically, ACE-V fails this test both because it does not offer a “reproducible and consistent” “matching rule,” *id.* at 48, and because the false positive rate, to the extent it has been subject to two appropriately designed black box studies, is too high to justify reliance on the method as a basis for source attribution statements in a criminal trial.

# ACE-V Does Not Offer a “Reproducible and Consistent” Means of Determining Whether Two Prints Have a Common Source

To be sure, ACE-V “was long hailed as infallible, despite the lack of appropriate studies to assess its error rate.” PCAST Report at 87. The myth of ACE-V’s infallibility likely stems first from the assumption that no two people have the same friction ridge patterns on their fingers.

That assumption has not been – and cannot feasibly be – empirically tested, *see generally* Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 VAND. L. REV. 199 (2008) but in any event is wholly irrelevant to the reliability of ACE-V in identifying a latent and known ink print as having a common source. The reliability issue with ACE-V “was never about errors caused by individuals possessing duplicate fingerprint patterns.” Simon A. Cole, *More Than Zero: Accounting for Error in Latent Fingerprint Identification*, 95 J. Crim. L. & Criminology 985, 1041 (2005). Rather, the issue is that *finger patterns* are different from the *impressions* they make on surfaces, and that a latent *print* and known ink *print* may well have a lot of similarities but still be made by different fingers:

Uniqueness and persistence are necessary conditions for friction ridge identification to be feasible, but those conditions do not imply that anyone can reliably discern whether or not two friction ridge impressions were made by the same person. Uniqueness does not guarantee that prints from two different people are always sufficient different that they cannot be confused, or that two impressions made by the same finger will also be sufficiently similar to be discerned as coming from the same source.

NAS Report (2009) at 144. *See also* PCAST Report at 61 (“The issue is not whether objects or features differ; they surely do if one looks at a fine enough level. The issue is how well and under what circumstances examiners applying a given metrological method can reliably detect relevant differences in features to reliably identify whether they share a common source.”).

Indeed, that is precisely what happened in Mayfield and other publicly known cases of misattribution, discussed below. The myth also likely stems from the mere declaration by the

forensic examiner community and its governing boards, until recently, that the discipline has a zero error rate, as well as the fact that latent print misattribution is difficult to uncover and prove, even in the era of DNA testing. In a world where an ACE-V analysis is considered vindicated by the mere fact of a suspect’s conviction, and the lack of any subsequent exoneration, the myth of infallibility was to be expected.

Notwithstanding this long-existing reputation for infallibility, recent studies and current events have offered a glimpse into the high subjectivity, and fallibility, of latent print examination. Perhaps most notoriously, Brandon Mayfield was falsely identified by FBI examiners using ACE-V as the source of latent prints on a bag of detonators linked with the 2004 Madrid train bombings. U.S. Dep’t of Justice, Office of the Inspector General, *A Review of the FBI’s Handling of the Brandon Mayfield Case* 1 (2006) [hereinafter “OIG Report (2006)”].

Indeed, an FBI examiner concluded with “100 percent certainty” based on ACE-V that the latent print and Mayfield’s exemplar – which was one of several prints identified as possible matches on a computer-generated list – were in agreement at 15 “points” of comparison and were therefore definitely a match. *Id.* The alleged match was then “verified” by *two* FBI supervisors *and* the same conclusions was subsequently reached by a court-appointed independent fingerprint examiner. *See* Cole (2005) at 986; OIG Report (2006) at 1-3. Nonetheless, two weeks later, after Spanish authorities found the real perpetrator and confirmed that his prints were a better “match,” Mayfield was released from custody and cleared of all suspicion, with the FBI issuing an apology. *Id.*

The FBI initially reported that the Mayfield misidentification was the result of poor imaging and “the similarity” of the latent print to Mayfield’s print. OIG Report (2006) at 3. A comprehensive review of the case by the Department of Justice’s Office of the Inspector General

determined that the incident could also have resulted from one or more other flaws in the examiner’s technique, including:

 [You may want to leave this section out if your lab has adopted the FBI recommendations or if

these are not issues in your case.]

* The “pressure of working on a high-profile case,” *id.* at 3;
* The fact that the examiner did not conduct a “complete analysis” of the latent print before comparing it to Mayfield’s print, causing him to “disregard important differences” between the prints, *id.*;
* An “overconfidence in the power of” the database, *id.* at 3, failing to consider that the true source might not be in the database and that database searches “can find a confusingly similar candidate print,” *id.* at 7;
* The fact that the second “verifying” examiner already knew the initial examiner’s conclusions, *id.* at 4, and did not “document the features or mental processes that led to his conclusion,” making a definitive assessment of the bias’s significance difficult, *id.* at 11;
* The level of similarity between Mayfield’s print, the real perpetrator’s print, and the latent print based on 10 points of apparent convergence, which the OIG described as an “extremely unusual circumstance,” but acknowledge that it could find “no systematic study of the rarity of such an event,” *id.* at 7;
* The erroneous conclusion that certain “extremely tiny” “Level 3 details” between the prints were in agreement, *id.* at 8;
* The erroneous conclusion that certain obvious differences between the prints were the result of a “separate touch” by a different finger or person, rather than a true difference, *id.* at 9; and
* The “limited clarity” of the latent print, which caused examiners to mistake one type of characteristic (an “ending ridge”) for another (a “bifurcation”), *id.* at 9.

In response to the Mayfield affair, the FBI laboratory, in agreement with the OIG, recognized the need to “research and develop more objective standards for fingerprint identification,” to change the match criteria to minimize “implicit pressure” to “make an identification in a difficult comparison in a case involving a particularly heinous crime,” and to “assure that verifications involve complete and independent examinations and provide a more stringent safeguard against erroneous identifications.” OIG Report (2006) at 10, 11.

While the Mayfield case might be the most famous recent erroneous fingerprint attribution, there were at least 22 such false attributions in the public record as of 2005. *See* Cole

(2005) at 991. Some were discovered only as the result of a later DNA exoneration. *See, e.g.*, *id.* at 987 (discussing the 2004 Stephan Cowans false fingerprint attribution and exoneration). Logic would dictate that there are numerous other erroneous fingerprint attributions that will never come to light.

In the wake of the Mayfield case, the OIG report, and burgeoning legal scholarship questioning the assumptions underlying ACE-V’s infallibility and noting the number of known false attributions, courts and scientific advisory boards began to newly “question” the “scientific foundation of the fingerprint field,” and in particular whether “latent fingerprint identifications may not be as reliable as previously assumed.” NAS Report (2009) at 43 (citing, e.g., *Md. v.*

*Rose*, 2008 WL 2914326 (Feb. 21, 2008)) (Balt. Cty. Cir. Ct.) (concluding latent fingerprint analysis to be “a subjective, untested, unverifiable identification procedure” and refusing to allow a fingerprint examiner to testify that the defendant was the source of a latent print).

For example, in 2009, the National Academy of Sciences – one of the most distinguished groups of scholars in the United States, providing guidance to Congress and the President since 1863[5](#_bookmark4) – renewed the call for empirical studies to test the foundational validity of latent print analysis as a means of forensic identification, noting that:

* + “[T]he ACE-V method does not specify particular measurements or a standard test protocol,” making the outcomes of the method “not necessarily repeatable from examiner to examiner,” *id.* at 139;
	+ Empirical studies of actual examiners and actual cases showed that “experienced examiners do not necessarily agree with even their own past conclusions when the examination is presented in a different context some time later,” *id.* at 139 (citing Itiel E. Dror & D. Charlton, *Why Experts Make Errors*, 56 J. FORENSIC IDENTIFICATION, 600-616 (2006));
	+ Those in the latent print field cannot even say “a priori which features should be compared” because certain features of the fingerprint may not have been captured by the latent print or differences can arise from stretching between two features, *id.* at 139;

5 *See* [*http://www.nasonline.org/about-nas/mission/.*](http://www.nasonline.org/about-nas/mission/)

* + “Little research has been directed toward developing population statistics” that would give an objective basis for comparison, “although more would be feasible,” *Id.* at 139-40;
	+ “[T]he reliability of the ACE-V process could be improved if specific measurement criteria were defined” rather than relying wholly on subjective examiner judgment, *id.* at 140; and that
	+ “A simple point count is insufficient for characterizing the detail present in a latent print,” and that “more nuanced criteria are needed,” *id.* at 141.

Based on these and other concerns, the National Academy ultimately concluded that ACE-V’s foundational validity had not been established:

ACE-V provides a broadly stated framework for conducting friction ridge analyses. However, this framework is not specific enough to qualify as a validated method for this type of analysis. ACE-V does not guard against bias; is too broad to ensure repeatability and transparency; and does not guarantee that two analysts following it will obtain the same results. For these reasons, merely following the steps of ACE-V does not imply that one is proceeding in a scientific manner or producing reliable results.

NAS Report (2009) at 142.

Two years after the NAS report, the Department of Justice’s Office of the Inspector General (OIG) issued yet another report on ACE-V, lamenting that “[s]everal of the most significant recommendations” that OIG made following the Mayfield debacle “depend on the results of the FBI Laboratory’s ongoing research projects and have not been implemented.” U.S. Dep’t of Justice, Office of the Inspector General, *A Review of the FBI’s Progress in Responding to the Recommendations in the Office of the Inspector General Report on the Fingerprint Misidentification in the Brandon Mayfield Case* 3 (2011) [hereinafter “OIG Report (2011)”]. The FBI did, to its credit, implement changes in response to the 2006 OIG report, including increasing examiners’ ACE-V training from four hours to six weeks, *id.* at 17, establishing more objective measures for determining image quality, *id.* at 18, and conducting a so-called “black box” study to test the accuracy of subjective examiner decisions, *id.*, a study discussed and critiqued in Part II.B.2. A recommendation that examiners perform linear ACE-V was also

adopted. *Id.* at 27. The FBI also began requiring that verifications of *identifications* be independent and documented, *id.* at 43. But as the PCAST Report would later point out, such verifications are no longer “blind” if the second examiner knows that verification only occurs when the initial examiner’s conclusion was an identification, i.e., that the suspect was the indeed the source. PCAST Report at 90. Meanwhile, numerous state and local laboratories also only verify identifications. *Id.* at 89. [If true: one such lab is [name of lab] at which the fingerprint examination at issue here was conducted.]

Ultimately, as of the 2011 OIG report, the FBI had still failed to combat the inherent lack of repeatability and consistency in ACE-V. The OIG noted that the agency had not yet achieved its goal of “developing qualitative and quantitative standards for determining sufficiency” of points of comparison for purposes of opining about the chances that two prints have a common source. OIG Report (2011) at 17-18. The lack of variability data “based on population distributions of certain friction ridge features,” necessary to undergird any objective criteria for determining the chance that two prints have a common source, was precisely what gave the National Academy of Science the most pause about ACE-V as a valid method. NAS Report (2009) at 141. *See also id.* at 144 (“None of these variabilities – of features across a population of fingers or of repeated impressions left by the same finger – has been characterized, quantified, or compared.”).

The latent print field’s decision to rely on a wholly subjective approach is particularly problematic, given the significant levels of cognitive bias that potentially infect each stage of ACE-V. In addition to Itiel Dror’s groundbreaking 2006 study of confirmatory bias in latent print examiners, NIST conducted a human factors study in 2012 about cognitive bias in ACE-V in particular. The Working Group identified several “locations within the interpretive process of

latent print examination where subjectivity and error due to human factors might be especially salient,” NIST Report at 40, such as at the comparison stage, where “there is no systematic effort” in the latent print field “to blind forensic examiners from contextual information that is clearly not domain relevant,” such as the facts of the case or a suspect’s prior convictions. *Id.* at

44. The Group also highlighted the fact that latent print examiners essentially choose their own internal threshold for determining for the level of similarity that must be achieved before calling a match, *id.* at 7, creating a remarkable level of subjectivity and potential for cognitive bias affect results. The Group concluded that, “[f]or the most part, the empirical basis for clearly establishing the extent of these risks does not exist,” and that “[f]urther research is needed to establish whether, in what circumstances, and to what extent these theoretical issues of interpretation create an elevated risk for error.” *Id.* at 40.

Since the 2012 NIST Report, two other studies have also investigated how the conclusions of fingerprint examiners may be influenced by task-irrelevant factors. PCAST Report at 99. In a 2014 paper, researchers found that the Comparison phase and Evaluation phase were not as distinct as the ACE writers intended. *Id.* at 99-100. The results of the study indicated that examiners’ markups and documentation are influenced by their determinations. *Id.*

Researchers posit that after making a decision, examiners may go back and revise the annotations made in the Comparison phase. *Id.* at 100. This behavior may be due to a desire to provide complete support to the decision reached, given that researchers noted greater markers on a latent print when a matching exemplar print was present. *Id.* In a 2015 follow-up, research confirmed that in over 90 percent of identifications, examiners changed the markings on the latent fingerprint in response to a matching exemplar fingerprint. *Id.* The researchers argued that, as a matter of scientific validity, examiners must “complete and document their analysis of a

latent fingerprint before looking at any known fingerprint” and “must separately document any data relied upon during comparison or evaluation that differs from the information relied upon during analysis.” *Id.* While the FBI adopted these requirements following Mayfield, they have not been universally adopted. *Id.*

In short, ACE-V remains a highly subjective, black box examination process that has no demonstrably repeatable, consistent means of determining characteristics or source attribution. Distinguished scientists and academics at NAS, NIST, and PCAST have acknowledged these problems, and the laboratory in this case has not addressed them. For these reasons, the expert testimony must be excluded as lacking foundational validity.

# 2. ACE-V’s Accuracy Has Not Been Sufficiently Demonstrated Through Appropriately Designed Empirical Testing

Even if ACE-V were a repeatable and consistent method, it would still lack foundational validity because its accuracy has not been sufficiently demonstrated through empirical testing. As the PCAST Report emphasized with respect to subjective methods of pattern-evidence comparison, “nothing—not training, personal experience nor professional practices—can substitute for adequate empirical demonstration of accuracy.” PCAST Report at 46. Because ACE-V remains a subjective, black box examination process, and because it is a process known to produce false positives, the dearth of appropriately designed empirical “black box” studies demonstrating the method’s accuracy renders it scientifically invalid as a means of justifying statements of source attribution in court.

As explained in Part II.A, a “black box” study of a forensic pattern-identification discipline is an empirical test of the ability of a *subjective* method – that is, a method that relies on the judgment and experience of the examiner, which is in essence the “black box” of the examiner’s head – to render accurate answers to the questions it purports to address. PCAST

Report at 49. In short, black box studies attempt to evaluate the accuracy and reliability of a method by having many examiners analyze unknown and known samples and make conclusions. These studies are used to determine error rates in the method. *See id*. at 5-6, 48, 49. The rate at which examiners erroneously conclude that two prints have a common source when in fact they come from different sources is called the “false positive rate.”

In the latent print field, “[r]emarkably, there have been only two black-box studies that were intentionally and appropriately designed to assess validity and reliability.” PCAST Report at 91. Nonetheless, even these two studies are problematic in a number of respects and, at best, suggest that the upper bound of ACE-V’s estimated false positive rate (assuming a 95% confidence interval)[6](#_bookmark5) is too high to justify its use in criminal cases, given that “the true rate could reasonably be as high as this value.” *Id.*

The first study was conducted by the FBI in 2011 in response to the NAS Report. *Id.* at

94. The study consisted of 169 fingerprint examiners and 744 latent-known pairs, 520 “mated” pairs, known to be from the same source, and 224 “non-mated” pairs, known to be from two different sources. *Id*. Each fingerprint examiner analyzed 100 pairs of prints and classified them as an identification, an exclusion, or inconclusive. *Id*. The study found six false positive identifications among 3628 non-mated pairs, resulting in a false positive rate of “1 error in 604 cases, with the upper bound indicating that the rate could be as high as 1 error in 306 cases.” *Id*.[7](#_bookmark6)

6 *See* PCAST Report at 94. A 95% confidence interval with an upper bound of 1 in 18 means that we can be 95% certain that the true false positive rate lies somewhere between 1 in 18 at the upper end, and some smaller value at the lower end. A 99% confidence interval, arguably a more appropriately conservative statistical estimate in criminal cases, would have an upper bound even higher than 1 in 18 (meaning an even more disturbingly high false positive rate).

7 A follow up study by the same authors several months later took 75 examiners and a subset of pairs from the first study and allowed some of the examiners to reexamine the pairs that led to false positives in the first study. Not surprisingly, there were no false positives reported this time around, although the reported upper bound of a 95% confidence interval would still be a false positive rate of 1 in 160. *Id*.

The PCAST Report noted that the study, while appropriately designed, was conducted by authors with a vested interest in the admissibility of latent print examiner testimony. *Id.* at 97.

The second study was conducted by the Miami-Dade Police Department Forensic Services Bureau, with funding from the National Institute of Justice. *Id.* at 94. While the study has concluded and the results were posted on the internet, it has yet to be published in a peer- reviewed scientific journal. *Id* at 94-95. The study also did not select known-latent print pairs to be similar to each other, “which should, in principle, have made it *easier to declare exclusions* for the non-mated pairs.” *Id* at 95. (emphasis added). Even so, the study still found 42 false positives among 995 conclusive examinations, for an upper-bound false positive rate of 5.4%, or 1 in 18. *Id*.

Viewing these studies together, PCAST remarked that “[t]he empirically estimated false positive rates are *much higher* than the general public (and, by extension, most jurors) would likely believe based on longstanding claims about the accuracy of fingerprint analysis.” *Id.* (emphasis in original).

While courts have considered and rejected similar arguments about false positive rates in the past, their conclusions need to be revisited. *See U.S. v. Herrera*, 704 F.3d 480 (7th Cir.

2013); *In re O.D.*, 221 Cal.App.4th 1001 (Cal. 2013); *People v. Rivas*, 238 Cal.App.4th 967 (Cal.

2015); *U.S. v. Havvard*, 260 F.3d 597 (7th Cir. 2001); *U.S. v. George*, 363 F.3d 666 (7th Cir.

2004); *U.S. v. Crisp*, 324 F.3d 261 (4th Cir. 2003); *U.S. v. Mitchell*, 365 F.3d 215 (3d Cir. 2004).

Significantly, all relevant case law on this issue was decided without consideration of the PCAST report.

Specifically, in *Herrera*, the Seventh Circuit did not rely on the current known error rates of latent fingerprint analysis, instead relying on a review of exonerations in the United States.

704 F.3d at 487 (rejecting the defendant’s argument that latent fingerprint analysis has not been shown to be reliable enough to be admissible as evidence under Federal Rule of Evidence 702 and *Daubert*). While the court found “errors in fingerprint matching by expert examiners… to be very rare,” the Seventh Circuit was not able to take into account either the 2011 FBI study or the 2014 Miami-Dade study reviewed in the PCAST Report. *Id.* at 487; PCAST Report at 95-97. As the PCAST Report noted, the known error rate may be higher than 1 in 18. PCAST Report at 95- 96, 101-02. Additionally, the *Herrera* court focused largely on the uniqueness of fingerprints, failing to take into account other factors. *Herrera*, 704 F.3d at 487. The court noted the probability of two people having the same fingerprint as one in sixty-four billion. *Id.* While this statistic is not entirely irrelevant to the field of latent fingerprint comparison, the more important statistic is the probability of an erroneous match of a latent print to a non-matching known print, a statistic the court mistakenly assumed to be “very rare.” *Id.* Relying on an immaterial statistic, the court found fingerprint comparison evidence admissible. *See id.*

While the PCAST authors themselves ultimately described ACE-V as foundationally valid, they noted that the “false positive rate . . . is substantial.” PCAST Report at 101. The authors’ descriptions of the subjectivity of the method and “substantial” false positive rate, however, speak for themselves. Additionally, the PCAST Report authors are ultimately not the gatekeepers in a criminal trial – this Court is. And a high false positive rate, while perhaps not dispositive to a group of scientists interested in a method’s application to a number of areas, should be dispositive in a criminal trial when offered against a suspect as proof of guilt.

In sum, ACE-V lacks foundational validity not only because the method is inherently unrepeatable and inconsistent because of its high subjectivity, but because of the minimal number of empirical “black box” studies demonstrating that the accuracy of the method. Indeed,

the two appropriately-designed studies that do exist suggest that the false positive rate is disturbingly high and that the method should be further empirically tested before being used as evidence against individuals accused of crimes.

# Expert Testimony in This Case Based on ACE-V Is Inadmissible Because It Is Not Based on a Method That Is Valid “As Applied.”

Even if ACE-V were somehow a foundationally valid method in principle, the government would be unable to meet its burden under Rule 702(d) of demonstrating that ACE-V is reliable as applied by this examiner in this case – that is, that the examiner is capable of reliably applying the method, that the examiner did actually reliably apply the method, and that the examiner’s particular assertions are scientifically valid. The PCAST Report indicated that, with respect to latent print examiners, “courts should take into account” three factors in particular in determining validity as applied.

First, given the subjectivity associated with latent fingerprint analysis, the PCAST Report concluded that it is “scientifically unjustified” to find that an examiner is capable of a reliable analysis unless he has completed “regular and rigorous proficiency testing.” *Id.* at 101.[8](#_bookmark7) [If true: even if the examiner has undergone *regular* proficiency testing, the government cannot establish that any such testing was rigorous. Indeed, “testing services have stated that forensic community prefers that tests not be too challenging.” *Id.* at 57.]

Second, “[i]n any given case, it must be established that the latent print(s) are of the quality and completeness represented in the foundational validity studies.” *Id.* at 101. Third, this Court should consider the extent to which the examiner in the particular case has “taken” “measures . . . to mitigate bias during casework.” *Id.* For example, the examiner should

8 “Rigorous proficiency testing” includes, *inter alia*, double blind proficiency testing on samples that represent the full range of latent prints encountered in casework and broad disclosure of proficiency tests to the scientific community for evaluation of the tests. *See id*. at 57-58, 102, 134.

demonstrate that s/he employed “linear ACE-V” by completely documenting his/her analysis of the latent fingerprint before looking at any known fingerprint and separately documenting any additional data used during comparison and evaluation. *Id.* at 101. Similarly, safeguards should be in place to ensure the examiner was not exposed to biasing information.

[If true: There is no evidence in this case that demonstrates either that the examiner

employed linear ACE-V or that any safeguards against biasing information were in place. In fact,

discovery provided demonstrates that the examiner was exposed to irrelevant, potentially biasing

contextual information on a “Case Submission Form” provided to the examiner by the

government.]

# [if CA]: [ACE-V Is Subject to, and Fails to Pass Muster Under, *Kelly/Frye*.

No case has ever held that latent fingerprint comparison evidence satisfies the *Kelly-Frye* rule. The only hint by the courts regarding whether or not feature comparison evidence generally is subject to *Kelly* can be found in dicta in two separate lines of cases. In one line of cases, courts appear to believe feature comparison evidence is subject to *Kelly* scrutiny. See *People v. Sinaiko*, 122 Cal.App.4th 1133 (Cal. 2004) (“[t]he Kelly-Frye test applies to new scientific methods used to prove identity or the occurrence of some material fact (such as fingerprints, DNA testing, and battered woman syndrome), but is not used to test all expert testimony.”); *People v. Bledsoe*, 36 Cal.3d 236 (Cal. 1984). There is, however, a fundamental difference between rape trauma syndrome and both the battered child syndrome and the other scientific methods of proof that have in the past been evaluated against the *Frye* standard of reliability. Unlike fingerprints, blood tests, lie detector tests, voiceprints or the battered child syndrome, rape trauma syndrome was not devised to determine the "truth" or "accuracy" of a particular past event i.e., whether, in fact, a rape in the legal sense occurred but rather was developed by professional rape counselors as a

therapeutic tool, to help identify, predict and treat emotional problems experienced by the counselors' clients or patients;” *People v. McDonald*, 37 Cal.3d 351, 372-373 (Cal. 1984) “courts have invoked the *Kelly*-*Frye* rule primarily in cases involving novel devices or processes such as lie detectors, ‘truth serum,’ Nalline testing, experimental systems of blood typing, ‘voiceprints,’ identification by human bite marks, microscopic analysis of gunshot residue, and hypnosis, and, most recently, proof of guilt by ‘rape trauma syndrome’.” *Id*. at 373 (internal citations omitted). In other words, dicta from this line of cases strongly suggests latent fingerprint evidence, like human bitemark comparison and voiceprints is subject to *Kelly* admissibility requirements.

However, dicta in a second line of cases suggests otherwise based on one of two theories:

(1) feature comparison is not subject to *Kelly* because the jurors may see the evidence for themselves and presumably conduct the comparison themselves in evaluating the evidence; (2) latent fingerprint analysis is not subject to *Kelly* because it is not science. See *People v. Venegas*, 18 Cal.4th 47, 81 (Cal. 1998) (“Unlike fingerprint, shoe track, bite mark, or ballistic comparisons, which jurors essentially can see for themselves, questions concerning whether a laboratory has adopted correct, scientifically accepted procedures for generating autorads or determining a match depend almost entirely on the technical interpretations of experts”); *People*

*v. Webb*, 6 Cal.4th 494, 524 (Cal. 1993) (laser procedure for visualizing latent fingerprints not subject to *Kelly* test because the familiar image of the fingerprint makes reliability of the process readily apparent); *People v. Clark*, 5 Cal.4th 950, 1017-1019 (Cal. 1993) (blood-spatter tests not subject to *Kelly* requirements because it is common knowledge that inferences can be drawn from spatter patterns of blood expelled from the human body). Despite this dicta, it seems clear that latent fingerprint comparison evidence cannot be immune from *Kelly* admissibility

requirements particularly because jurors believe fingerprints are scientific and because the fingerprint examiner will testify that a lay person is incapable of comparing two prints and determining whether or not they “match.”

# IF NOT EXCLUDED ENTIRELY BY THIS COURT, THE EXPERT LATENT PRINT TESTIMONY MUST BE LIMITED UNDER *DAUBERT/FRYE* IN TWO KEY RESPECTS

If this Court does not exclude the expert latent print testimony in this case, it must in the alternative, at a minimum, limit the testimony in two key respects: (1) exclude testimony opining that the latent print and known print “match,” are from a common source, or that Mr. XXXX is the source of the latent print; and (2) require the expert to qualify his/her opinion by acknowledging that the level of certainty in any conclusion of source attribution would be limited by the upper bound of the highest potential false positive rate from appropriately designed empirical studies, which is 1 in 18.

Even if a method is foundationally valid, any expert testimony based on that method must also be foundationally valid. PCAST Report at 54. In particular, “*[s]tatements claiming or implying greater certainty than demonstrated by empirical evidence are scientifically invalid*.” PCAST Report at 54 (emphasis in original). Here, any statement by an expert on the stand that Mr. XXXX’s known print “matches” the latent print, or is the source of the latent print, is in essence a statement of absolute source attribution. *See, e.g.*, NAS Report (2009) at 141-42 (“[W]hen a latent print examiner testifies that two impressions ‘match,’ they are communicating the notion that the prints could not possibly have come from two different individuals.”). As detailed above, however, a statement that two prints could not possibly have come from two different individuals is a statement implying certainty that simply has no empirical foundation, even if one assumes that ACE-V is a generally reliable process.

Moreover, if the expert is allowed to testify at all, even without a statement about a “match,” his/her opinion about the similarities between the latent print and Mr. XXXX’s known prints will be an implicit signal to the jury that these similarities must have some probative value in the case (otherwise, the judge would have presumably excluded the expert’s testimony as irrelevant). The unadorned opinion of the expert about similarities in the pair of prints, without any information about the high error rate, will leave the jury to determine for itself the probative value of such similarities. Given lay jurors’ grossly inaccurate assumptions about the false positive rate of fingerprint examiners, *see* PCAST Report at 95 n. 282 (citing a study showing that mock jurors believed the false positive rate was 1 in 5.5 million), allowing the expert testimony without an explicit qualification of the level of uncertainty would be both contrary to Rule 702’s requirement of foundational validity and unduly prejudicial. *See also* PCAST Report at 56 (requiring, to establish validity “as applied,” that the expert report the false positive rate when testifying).

[This section should be tailored to the way the examiner will testify in your case.] Finally, examiners should never be allowed to testify to “scientifically indefensible claims such as: ‘zero,’ ‘vanishingly small,’ ‘essentially zero,’ ‘negligible,’ ‘minimal,’ or ‘microscopic’ error rates; ‘100 percent certainty’ or proof ‘to a reasonable degree of scientific certainty;’ identification ‘to the exclusion of all other sources;’ or a chance of error so remote as to be a ‘practical impossibility.’” PCAST Report at 19. Such statements “are not scientifically valid.” *Id. See also* NAS Report (2009) at 142-43 (“claims that… [fingerprint] analyses have zero error rates are not scientifically plausible.”); National Commission on Forensic Science, *Recommendations to the Attorney General Regarding Use of the Term “Reasonable Degree of*

*Scientific Certainty”*, (Approved March 22, 2016) https://[www.justice.gov/ncfs/file/839731/download.](http://www.justice.gov/ncfs/file/839731/download)

The stakes in allowing a latent print expert to testify to an alleged fingerprint “match” are high. Even in the age of DNA, “[n]o forensic expert witness can more convincingly place a suspect at the scene of [a] crime than the latent fingerprint examiner (LFPE).” Simon A. Cole, *Witnessing Identification: Latent Fingerprint Evidence and Expert Knowledge*, 28 SOC. STUD. SCI. 687, 688 (1998). Even when properly limited, all scientific expert testimony carries with it an “aura of special reliability and trustworthiness,” creating a risk that jurors will receive it without a critical eye and attribute greater weight to the evidence than can be established. [*Frye* jurisdictions: *See Kelly*, 17 Cal.3d at 31 (citing *Huntingdon v. Crowley*, 64 Cal.2d 647, 656 (Cal. 1966)) (holding “(l)ay jurors tend to give considerable weight to scientific evidence when presented by experts with impressive credentials.”).] [D.C. only: *Motorola Inc.*, 147 A.3d at 753 (citing *Ibn–Tamas v. United States*, 407 A.2d 626, 632 (D.C. 1979) (internal quotations and citation omitted)).] Perhaps because juries view forensic testimony with unflinching trust, the use of misleading or embellished forensic expert testimony is one of the leading causes of wrongful convictions. Brandon L. Garrett, *Judging Innocence*, 108 Colum. L. Rev. 55, 83-84 (2008).

These concerns are heightened when an expert makes grandiose claims of source attribution based on pattern evidence. As one just wrote in condemning the testimony of a toolmark examiner that markings on the bullets were “unique” to a gun recovered from a defendant’s apartment: “This is not evidence on which we can in good conscience rely, particularly in criminal cases, where we demand proof—real proof—beyond a reasonable doubt, precisely because the stakes are so high.*” Williams v. US*, 130 A.3d 343, 355 (2016) (Easterly, J.,

concurring) (disapproving of improperly admitted expert claim of certainty but affirming on plain error grounds).

# CONCLUSION

For the reasons stated above, this Court should exclude the government’s proffered latent print expert testimony or, in the alternative, limit and qualify it in the manner discussed above.

Respectfully submitted,

XXXX XXXX

**EXHIBIT A**