I. INTRODUCTION

The court's role at the intersection of science and law is to ensure that the criminal justice system operates fairly. When dealing with novel scientific evidence, a judge must ensure that the scientific evidence is sufficiently reliable to be heard by the jury. The Office of Chief Medical Examiner ("OCME") has brought on line two novel methodologies: Low Copy Number DNA Testing (LCN) and the Forensic Statistical Tool (FST). The OCME's validation studies for both of these novel methodologies, the current literature in the relevant scientific community, the testimony from scientists at this hearing, all demonstrate that these methodologies are not ready for use in court.

The OCME is pushing the boundaries of science both with their LCN methodology and the FST. In 2005, when the OCME brought LCN online, they dramatically changed the methodology employed by conventional PCR-STR DNA testing. The use of LCN is controversial because of increased stochastic effects and very serious concerns over contamination. The published scientific data strongly indicates that LCN is not ready for in court criminal use, and that the forensic DNA community is still trying to find a path forward. As recently as 2009, a heated public debate occurred among forensic DNA experts about the use of LCN. The most informative fact about the position of the relevant scientific community is that no public crime laboratory in the United States is using this methodology and even abroad there is limited support that this method is being used appropriately. While LCN is used in a few countries abroad there has been significant controversy surrounding the use of this methodology – Amanda Knox, Sean Hoey, and Bradley John Murdoch have all had LCN used in their cases abroad and all have produced unreliable results.

The OCME's LCN methodology was brought on line prematurely. Besides employing an unreliable methodology, the results generated, namely a qualitative result such as "cannot be excluded" was scientifically unacceptable. In an attempt to fix the problem that LCN could not produce a quantitative result, the OCME developed the FST without addressing all the other issues with LCN. Compounding the problems with LCN, is that the FST methodology also has a number of issues: its drop model is unique in the world, its assumptions are unsupported by scientific data, key data in the validation study went unrecorded or was discarded, and results that required further investigation were not tested further.

The OCME's pushing of scientific boundaries and lack of transparency is troubling. This is especially true given the controversy surrounding the laboratory. The New York Inspector General stated that Dr. Theresa Caragine, the Deputy Director of the Forensic Biology Department and creator of LCN and the FST, did not follow protocols, changed results that were exculpatory to inculpatory, and obstructed the discovery of this behavior by keeping a file in her office for months. Furthermore, Serrita Mitchell mishandled the evidence for an alarming number of for years before she was caught. Yet in spite of this type of behavior, by both management and criminalists, the OCME continues to not disclose data and other relevant information regarding their novel technologies that they are being used in court.

Admitting this type of evidence is contrary to the fairness and integrity that the criminal process requires. The Frye hearing before this Court demonstrated that both LCN and the FST are unreliable and not generally accepted by the relevant scientific community. Accordingly this Court should grant the defendant's motion to exclude both LCN and FST expert testimony on the ground that they do not satisfy the Frye standard.

II. FACTS

A. Witnesses Called at Hearing

Over the course of over a year this Court heard testimony from 11 experts. The People called four witnesses: two from the OCME and two from outside the OCME. The defense put on seven witnesses and called one of the People's witnesses as well.

The People called Dr. Theresa Caragine, Dr. Adele Mitchell, Dr. Hinda Haned, and Dr. Mitchell Holland. Dr. Caragine was forced to resign in between cross examination and redirect. The prosecution declined to recall her. While at the OCME she was the Deputy Director of the Forensic Biology Division, and in this position she supervised the High Sensitivity Team, the Property Crimes Team, and the Research and Validation Team. She supervised the LCN validation and played a crucial role in the FST validation. Dr. Caragine was subsequently investigated by the New York Inspector General's office after she resigned. The Inspector General found that Dr. Caragine failed to follow protocols. She was also referenced indirectly in the Sorenson report that criticized the OCME's management.

Dr. Mitchell created the FST. She was a population geneticist, and took over as leader of the Research and Validation Team after Dr. Caragine was forced to resign.

Dr. Haned is a research scientist at the Netherlands Forensic Institute and was the developer of Forensim/LRmix, a likelihood ratio program that modeled allelic drop-out and drop-in. Dr. Haned did not thoroughly review the FST validation, but rather focused on the Executive Summary and the charts and graphs in the validation papers. Dr. Haned did not review the LCN validation.

Dr. Holland is a professor from Pennsylvania State University. Dr. Holland had been a former vice-president of Bode Technologies, one of the largest private DNA forensic laboratories

in the country. Although Dr. Holland is a proponent of low copy number testing, Bode has declined to employ that methodology in its lab work. Dr. Holland only reviewed the LCN protocols and publications, and one of the FST publications.

The People indicated that they intended to call four other witnesses, but they never materialized. In early 2013, the People stated that they may call Dr. Peter Gill, and then made a motion to have Dr. Gill testify via videoconferencing. Dr. Gill is one of the pioneers of low copy number testing, an architect of an LR drop model called LoComatioN that never went online, and the principle organizer of the 2006 and 2012 DNA Commissions of the International Society of Forensic Genetics (ISFG). Shortly before the prosecution rested in September of 2013, Dr. Gill gave a talk for the ISFG annual conference in Australia, where he questioned whether a higher LR was necessarily indicative of strong evidentiary power. Dr. Gill is an advocate of open source for forensic LR software, and was quoted by the New Scientist last summer as saying open source is a requirement for in court acceptance. The prosecution indicated it would call Dr. Gill in its rebuttal case, but the Court denied its application for remote video testimony. The prosecution then declined to call Dr. Gill, citing financial reasons.

The prosecution also indicated that it would call Dr. Jack Ballantyne, a DNA forensic scientist and the chairperson of the DNA Subcommittee of the New York State Forensic Commission. Dr. Ballantyne testified on behalf of the prosecution in a Frye challenge against FST in New York County, before the suspension of Mechthild Prinz, the former director of the Department of Forensic Biology at the OCME, and before the sudden departure of Dr. Caragine. Dr. Ballantyne travels to New York City at least four times a year for the subcommittee meetings, yet in the end the People did not call him to testify.

The prosecution also indicated it would call Dr. Robin Cotton, a DNA forensic scientist at Boston University. Dr. Cotton is a former lab director for Orchid Cellmark, and has written on the relationship between general use and general acceptance of new forensic technologies. Dr. Cotton was never called.

The prosecution also indicated that it would call Dr. Peter Schneider of the University of Cologne. Dr. Schneider is a DNA forensic scientist who has published research indicating a disconnect between initial quantification values and final output for degraded DNA samples. Dr. Schneider was in New York in February, but was never called as a rebuttal witness.

The defense called Dr. Eli Shapiro, Dr. Heather Coyle, Dr. Rori Rohlfs, Dr. Angela van Daal, Dr. Noah Rosenberg, Dr. Bruce Budowle, and Dr. Ranajit Chakraborty. The defense also called Dr. Adele Mitchell in their direct case.

Dr. Shapiro is a former assistant director at the Department of Forensic Biology Division at the OCME, who was the director of training there for roughly 10 years, and leader of the Mitochondrial Team. Dr. Shapiro is a graduate of both Columbia University and Yale University. Dr. Shapiro was the manager who was part of the team that first caught Serrita Mitchell's deficiencies. Mitchell was an employee who pre-dated Dr. Shapiro at the lab, and who later became the center of scandal at the OCME because of cross-contamination. Along with Dr. Caragine, Dr. Shapiro had traveled with an OCME team in 2008 or 2009 to meet with Dr. Gill and other research scientists from the Forensic Science Service in Great Britain. Dr. Shapiro later had serious misgivings about the direction that Dr. Caragine was taking the Research and Validation Team. Dr. Shapiro did an extensive review of the OCME's FST validation.

Dr. Heather Coyle is an Associate Professor at the University of New Haven in the Department of Forensic Science. Dr. Coyle was also a former criminalist at the Connecticut State Crime Lab. Dr. Coyle reviewed parts of both the LCN and FST validations.

Dr. Angela Van Daal is a forensic DNA research scientist formerly at Bond University in Australia. Dr. Van Daal has had an extensive career in forensic DNA testing, and performed a significant amount of research. She was one of the first scientists in the world to use PCR in a forensic setting, worked developing new technologies with Applied Biosystems, the manufacturer of the kits widely used in the forensic community for DNA typing, and helped develop international accreditation standards for forensic labs. Her research included comparing replicate testing versus one amplification testing when dealing with small amounts of DNA. Dr. van Daal reviewed the LCN and parts of the FST validation.

Dr. Rori Rohlfs is a population geneticist and post-doctoral candidate at University of California at Berkeley. Dr. Rohlfs received B.S. from Carnegie Melon University and her PhD at the University of Washington while a student of Bruce Weir, a well respected expert in the forensic DNA community. Dr. Rohlfs has done research on false positive testing. Dr. Rohlfs had never testified or done any consulting. Dr. Rohlfs reviewed and analyzed the FST's false positive test.

Dr. Noah Rosenberg is a well respected population geneticist and statistician at Stanford University. He received a B.A. in mathematics from Rice University, a M.S. in mathematics from Stanford University, a PhD in biology from Stanford University, and completed his post-doctoral fellowship in a molecular and computational biology group at the University of Southern California. Dr. Rosenberg has published an extensive number of papers and is an editor for a

number of journals. Dr. Rosenberg had never testified before. Dr. Rosenberg reviewed the FST's false positive study.

Finally, the defense called Drs. Bruce Budowle and Ranajit Chakraborty of the University of North Texas Health Science Center. Dr. Budowle is one of the most famous forensic DNA scientists in the world. He is a former senior scientist and lab head with the FBI, where his career stretched over 25 years. He work was key to developing the Combined Offender DNA Index System (CODIS). Dr. Budowle reviewed both the LCN and FST validations.

Dr. Chakraborty is a preeminent forensic DNA population geneticist. Dr. Chakraborty helped developed the 13 core STR genetic markers used by the FBI and in labs throughout the world. His work was vital to the report issued by the National Research Council (NRC II). Dr. Chakraborty was a member of the DNA Subcommittee that approved LCN DNA testing and the FST. Since voting for these methodologies, Dr. Chakraborty developed serious concerns about these methodologies based on additional data and continued research in the field.

B. DNA Testing Methodology[1]

In 1985 Alec Jeffreys demonstrated that identity and generic inheritance could be determined by repeated sequences. Restriction fragment length polymorphism (RFLP) was the first method used in DNA typing, and was then followed by techniques based on polymerase chain reaction (PCR) including DQA1, PolyMarker, and amplified fragment length polymorphism (AMP-FLP). As research continued, short tandem repeats (STRs) emerged as the "genetic marker of choice." In 1997, the FBI Laboratory selected 13 core STR loci for forensic DNA testing. PCR-STR DNA testing continues to be used in forensic DNA laboratories nationwide and is the gold standard for DNA typing.

[1] The material in this section was obtained from John M. Butler's FUNDAMENTALS OF FORENSIC DNA TYPING (2011).

1. DNA Basics

DNA (deoxyribonucleic acid) is frequently referred to as our "genetic blueprint." DNA stores the information necessary for an organism to develop and function. DNA is located in every nucleated cell in the body. No two individuals, with the exception of identical twins, will have the same genetic code.

DNA is composed of a nucleobase, a sugar, and a phosphate. There are only four nucleobases: adenine, thymine, cytosine, and guanine. The order of these nucleobases determines the content of the DNA code. DNA is comprised of two strands, in which the individual nucleotides pair with their "complimentary base." The resulting pairs are adenine-thymine and guanine-cytosine. The two DNA strands form the "double helix."

DNA is made up of chromosomes, which includes 22 pairs of autosomal chromosomes and two sex-determining chromosomes. In total, a human cell will have 46 chromosomes (23 pairs of chromosomes). With regard to the sex-determining chromosomes, males have one copy of the X chromosome and one copy of the 7 chromosome; and females contain two copies of the X chromosome. With each pair of chromosome, one is inherited from a person's mother and the other from a person's father.

The DNA in chromosomes have both "coding" and "non-coding" regions, and genes are located in the coding regions. These genes are located at the same locus on each chromosome, and at these genetic loci alleles are present, which are the alternative forms of a gene. If the two alleles at a locus are different, they are known as "heterozygous," but if they are the same, then they are "homozygous."

2. PCR-STR DNA Testing

In forensic DNA testing the 13 core STR loci are analyzed as well as the sex-determining locus, with some laboratories testing a few additional loci. Forensic DNA testing involves two technologies, polymerase chain reaction (PCR) and short tandem repeats (STRs). PCR is a process by which regions of the DNA are copied, causing an exponential increase in the amount of copies of the target region to be produced.

STRs are regions where a string of nucleotides repeat, and the number of these repeats are highly variable, allowing them to be used for identification. While the number of nucleotide repeats can range from 2 to 7 base pairs, it is the 4 nucleotide repeats that are used in forensic DNA testing. An example of a four base STR is AAAC being repeated numerous times, such that the coding region would be AAAC AAAC AAAC AAAC. Alleles are determined by the number of repeats identified. Therefore, if AAAC is repeated 10 times then 10 would be the name of that allele.

In forensic DNA testing for each of the reported loci, a laboratory would report one value for each allele at a locus. For example, if the number of repeats was 15 on one chromosome and 18 on another, then a laboratory would report 15, 18 as the result at one locus.

PCR and STR are used as part of the steps of forensic DNA testing, which include examination, extraction, quantitation, amplification, electrophoresis, and analysis. During examination a lab will evaluate the evidence to determine whether there is biological material on the item. A lab may perform preliminary and confirmatory tests, which include acid phosphatase, Kastle-Meyer, and direct observation under the microscope. During extraction various steps are taken to obtain DNA depending on the source of the biological material. In quantitation, the amount of DNA present in the sample is determined.

In amplification, PCR is used to exponentially amplify the DNA sample. A PCR reaction includes the DNA, taq polymerase, and primers which target specific regions of the DNA. The PCR reaction is placed in a thermal cycler which heats and cools the mixture repeatedly. Electrophoresis is a process by which DNA is separated by size. The amount of DNA detected is based on fluorescence, a process in which a laser strikes a dye that is attached to a DNA fragment, and various filters are used to collect data on the amount of fluorescence at certain wavelengths. The amount of fluorescence is reported in relative fluorescence units (RFUs), which are visualized as the peaks seen in electropherograms.

The final step is analysis, in which an analyst interprets all the data produced by electrophoresis. An analyst will determine which peaks should be called based on the analytic threshold of the lab, and evaluate the peaks for whether they have any stochastic effects. An analyst will also have to perform a statistical analysis to determine the relevance of the data observed and peaks called.

III. LEGAL STANDARD

In New York scientific expert testimony is admissible only where the scientific principles or procedures have gained general acceptance in their particular field. People v. Wesley, 83 N.Y.2d 417, 422 (1994). The general acceptance rule was announced in Frye v. United States, 293 F. 1013 (D.C. Cir. 1923). The Frye Court stated:

> 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.

Id. at 1014. The scientific principle or procedure does not have to be 'unanimously indorsed' but must be 'generally acceptable as reliable.' Wesley, 83 N.Y.2d at 423 (citing People v. Middleton, 54 N.Y.2d 42, 49, 444 N.Y.S.2d 581, 429 N.E.2d 100).

Therefore the question for this Court is whether the reliability of the OCME's LCN and FST methodology has gained general acceptance by the relevant scientific community. See e.g., Wesley, 83 N.Y.2d at 423 (stating "the issue here concerns the acceptance by the relevant scientific community of the reliability of DNA evidence.").

A. Who is the "Relevant Scientific Community"?

In determining whether a scientific principle or procedure is generally accepted, a court must hear from the "relevant" scientific community:

> In defining the relevant scientific field, the court must seek to comply with the Frye objective of containing a consensus of the scientific community. If the field is too narrowly defined, the judgment of the scientific community will devolve into the opinion of a few experts. The field must include scientists who would be expected to be familiar with the particular use of the evidence at issue, however, whether through actual or theoretical research.

Wesley, 83 N.Y.2d at 438 (concurring opinion Chief Judge Kaye (citation omitted)).[2] The field of forensic DNA analysis draws on many other fields – molecular biology, population genetics, statistics – therefore, the relevant scientific community is not limited to forensic DNA analysts. Indeed, in the National Research Council's Report, "The Evaluation of Forensic DNA Evidence," the report had a report had a section on both the importance of population genetics and statistics. National Research Council, "The Evaluation of Forensic DNA Evidence," Chapters 4 and 5, 1996.

[2] Indeed, in Wesley the People did not rely on any experts who were from the field of forensic DNA analysis; the People's experts were in the area of molecular biology, population genetics, and a molecular biologist specializing in the study of blue-green algae. Id. at 437-38.

Indeed, some courts are skeptical in limiting the field only to practitioners in part because "practitioners tend to be materially interested in the validity of the technique. That is, they tend to stand to benefit financially if the technique is legitimated by a favorable admissibility ruling in the courts." Simon A. Cole, Out of the Daubert Fire and Into the Fryeing Pan? Self-Validation, Meta-Expertise and the Admissibility of Latent Print Evidence in Frye Jurisdictions, 9 Minn. J.L. Sci. & Tech. 453, 475. See also, Giannelli, The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later, 80 Colum. L. Rev. 1197, 1213 (stating "the supporting research is conducted by someone with a professional or commercial interest in the technique"); Wesley, 83 N.Y.2d at 441 (concurring Chief Judge Kaye) ("The opinions of two scientists, both with commercial interests in the work under consideration and both the primary developers and proponents of the technique, were insufficient to establish 'general acceptance' in the scientific field." (citing People v. Leone, 25 N.Y.2d 511, 514 (1969).).

Accordingly, in this hearing the court needed to hear from forensic DNA analysts, population geneticists, and statisticians.

B. What does it mean to be generally accepted as reliable?

As the court considers whether a scientific procedure is generally accepted as reliable, "[t]he court's job is not to decide who is right and who is wrong, but rather to decide whether or not there is sufficient scientific support for the expert's theory." Gallegos v. Elite Model Mgmt. Corp., 195 Misc.2d 223, 225 (Sup. Ct. N.Y. County 2003). Chief Judge Kaye warns in Wesley's concurring opinion:

> It is not for a court to take pioneering risks on promising new scientific techniques, because premature admission both prejudices litigants and short-circuits debate necessary to determine the accuracy of a technique. Premature acceptance of "revolutionary" forensic techniques has led to wrongful convictions. (*see,* Giannelli, *The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half–Century Later,* 80 Colum.L.Rev.

> 1197, 1224–1225 [discussing belated discovery of inaccuracy of paraffin test]; Neufeld and Colman, *When Science Takes the Witness Stand,* 262 [No. 5] Scientific Am. 46 [discussing belated discovery of inaccuracy of gunpowder detection test]). In People v. Leone, 25 N.Y.2d 511, 517–518, 307 N.Y.S.2d 430, 255 N.E.2d 696, *supra* we also warned against introduction of scientific evidence before its general reliability have been resolved in the scientific community, because " 'the value of the test * * * could easily become the question in the trial rather than that person's guilt or credibility'" (quoting People v. Davis, 343 Mich. 348, 372, 72 N.W.2d 269, 282).

Wesley, 83 N.Y.2d at 437 (n. 4, Chief Judge Kaye concurring). Therefore, a court should look to the consensus in the relevant scientific community to determine if the methodology being used is generally accepted as reliable.

New York Executive Law § 995 defines "DNA testing methodology" as "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." N.Y. Executive Law § 995 (McKinney 2014). Accordingly, when a new method is brought on line to provide a statistic for DNA evidence a court must consider all those factors when making a decision about the general acceptance of a scientific principal or procedure.

In this case, the People have failed to meet their burden to demonstrate that LCN and FST are generally accepted as reliable within the relevant scientific community. Furthermore, the defense has affirmatively proven that the relevant scientific community has not accepted the methods employed by the OCME's LCN and FST.

Mr. Collins asks this Court for an order *in limine* preventing the People from presenting expert testimony regarding any conclusion reached by the use of Low Copy Number (LCN) PCR DNA testing, and any conclusion reached by the use of the Forensic Statistical Tool (FST) on the ground that neither of these methodologies have been generally accepted as reliable by the relevant scientific community.

Similarly, Mr. Peaks asks this Court for an order *in limine* preventing the prosecution from presenting expert testimony regarding any conclusion reached by the use of the FST on the ground that the FST is not generally accepted as reliable by the relevant scientific community.

IV. THE FST IS NOT GENERALLY ACCEPTED IN THE RELEVANT SCIENTIFIC COMMUNITY AS RELIABLE BECAUSE IT IS BASED ON UNSOUND SCIENCE.

A. The FST is Based on Unsound Science.

1. This Court Should Focus on the Drop Model, Not the Likelihood Ratio

This Court should reject the argument that the only issue with FST is the use of the likelihood ratio rather than some other statistic. Arguments for the admissibility of FST have ignored the fact that in the forensic setting, models for drop-in and drop-out (drop models) are new, few, and based on different types of assumptions and calculations. The Court's inquiry should focus here. If the FST model for estimating drop-in and drop-out (or "drop model") is not generally accepted in the scientific community as reliable, no court should allow the FST statistic into evidence.

a. Shortcuts at the Expense of Sound Science

The OCME Research and Validation team developed an expedient model based upon shortcuts in lieu of sound scientific principles. The design of FST seems to have been set before any of the testing in the validation began, creating a "cart before the horse" strategy. There are strong indications that the team did not think through potential flaws in the program's design. Nor did the team show any inclination to redesign the program once these fundamental flaws came to light.

Numerous shortcuts were taken in empirically estimating the drop-out rate. First, the OCME chose to ignore the 30 percent range of error in the quant, which they had decided would be the key variable in the FST inputs. When it became apparent at the hearing that their drop-out

numbers fluctuated greatly, they resorted to the stock answer that they took steps to lower the drop-out rate, which they claimed would be conservative. However, they did not preserve any testing on this fundamental assumption, nor did they test any exceptions to their assumed rule.

Second, the team took no steps to account for allele stacking in the mixtures they created, despite the issue being highlighted in textbooks like John Butler's Fundamentals of Forensic DNA Typing. One of the prosecution witnesses, Dr. Mitchell Holland, also admitted that allele masking was not an ideal experimental model when estimating the drop-out rate. The scientific principle or assumption that drop-out rates can be counted and averaged linearly was called into question by the numbers they generated. However, they resorted to stock answers like "allele stacking happens in the real world," effectively abandoning an unknown variable in their calculations.

The third shortcut the team took was the decision to use only four basic mixture ratios in determining a set of universal drop-out rates, with no indication that they had done experiments to show that those four mixture ratios were sufficient to capture the intricacies of contributor combinations in casework.

The fourth shortcut was the decision to not preserve and publish any experiments to test the assumption that underestimating the drop-out rate would not overstate the strength of the evidence against the defendant. Nor did they apparently do any experiments to test the limits and exceptions to this assumed rule. In fact, Dr. Mitchell did not formally test this assumption until after she had been cross-examined in this hearing.

The fifth shortcut was the decision to bring the FST on-line before developing a case-by-case performance test or false positive test, even for low likelihood ratios (LRs), despite a level

of uncertainty in low LRs that possibly could change an inclusionary statistic to an exclusionary one.

The sixth shortcut was the decision to not control for key variables in their validation study, like population substructure of the contributors to their mixtures.

The seventh shortcut was the decision to not report LR results on a locus by locus basis, and to show drop-out and allele frequencies by locus.

These shortcomings will be addressed in detail below.

b. Manipulating data

OCME took a non-scientific approach towards getting the FST approach for use. Instead of quantifying and reporting the range of error for the LR generated by FST, the OCME reported a single number and lowered it arbitrarily by one standard of deviation. No comparison studies were recorded comparing their correction with the two standard deviation that is commonly employed in science to show a 95 percent confidence interval. The OCME also arbitrarily raised many of the drop-out rates from zero to .02, and capped their corrected drop-out rates well below the true drop-out rates that they recorded. The OCME tweaked and flattened drop-out rates in order to keep them from being inconsistent with its approach to making linearly –interpolated drop-out rates. Dr. Caragine and Dr. Mitchell claimed that these alterations were conservative, but Dr. Mitchell had to admit on cross-examination that their use of the term "conservative" did not necessarily comport with its universal definition within forensic science. The OCME actually did perform a study comparing higher and lower drop-out rates for the same sample, and the results indicated that the results were no better than flipping a coin for true contributors. The team buried this study in its validation studies and never reported it to anyone. The OCME also buried the data on the failed degradation model it tried to developed, failed to record the

comparison between FST and Forensim, fudged on the results of comparing high and low copy number testing of the same-sized samples, and over-exaggerated the results of its testing the independence of drop-out rate across loci.

2. Fundamental model design flaws

a. Use of the quant

i. Unique in the world

The linchpin of the FST program is the dubious belief that the quantitative PCR (or "qPCR") value, used to estimate the size of an amplified DNA sample, is directly correlated to the drop-out rates of the contributors to that DNA sample once it undergoes DNA typing. Dr. Budowle put it succinctly in his direct testimony:

> Q. In your opinion, is the FST a novel approach?
>
> A. I think it is in the way it's being used because it makes certain assumptions about DNA typing that no one else would do even in standard DNA typing. The main assumption being made is that all the rates for drop-in, drop-out are based on the amount of DNA. . .

Budowle, 12/9/2013, 793 [emphasis added].

Here, one of the top forensic scientists in the world has made clear that no other lab or program in the world would make the assumption that drop rates could be based on quantified DNA amounts. Dr. Budowle's opinion reinforced the prior testimony of Dr. Shapiro, who had pointed out the unique approach of using quant in the FST methodology:

> Q. Can you tell us the role of the quantitation value in the design of FST?
>
> A. Yes, I think the entire approach of the FST, which is a very unique one, and I think the only one in the field, is to use the quantitation values, for a sample, in order to determine the drop-out rates to apply to the mixture. . .

Shapiro, 10/15/2013, 177 .

ii. Loose correlation is not enough to be reliable.

Forensic scientists agree that as the quantity of the DNA decreases, stochastic effects like drop-out increase. However, there is a complete lack of support in the scientific literature for the OCME's use of the initial quant value as a predictive tool to accurately predict drop-out rates.

The fact that there is some relationship between the quantitative level of DNA and stochastic effects like drop-out does not mean that the OCME's reliance on quant is generally accepted in the scientific community. See generally Fraser v. 301-52 Townhouse Corp., 57 AD3d 416 (1st Dept 2008) (although there is general agreement in the scientific community that dampness and mold are associated with health problems, the observed association was not strong enough to constitute evidence of a causal relationship).

iii. The Testing of FST's Central Assumption Went Unrecorded and This Is Not Accepted in the Scientific Community

Because the OCME research team was the only group of scientists in the world claiming that it could correlate qPCR quant values to the drop-out rates of minor contributors to a mixture, it was incumbent upon them to establish with empirical evidence how changing the quant would affect the final LR. Incredibly, the team failed to perform a formal study to test this assumption, and did not record any of the data in the voluminous 47 binders:

> Q. How did you know that your drop-out rates were reasonable estimates?
>
> A. Because changing the amount of DNA in the sample does not dramatically change the likelihood ratio unless you go from say 10 picograms to 500 picograms.
>
> Q. So did you test that?
>
> A. Yes.
>
> Q. As part of the non-contributor study?

A. No. I don't believe it's part of the validation.

Q. You didn't put that in your validation?

A. No.

Q. You did a whole study about changing the quantity of DNA input to see how that could affect the likelihood ratio and didn't put that in the validation study?

A. I didn't do a formal study like that. I've altered the input value to see how it changes, but I haven't put together a formal binder of results. No.

Mitchell, 5/21/13, 28-29. Notwithstanding its lack of formality, Dr. Mitchell claimed she did a "proper validation study" to determine that the quant size of a sample tracked the drop-out rate for that sample, thereby changing the final LR, or likelihood ratio. Id., p. 29, lines 17-18. However, that "proper validation study" was effectively hidden from the scientific community. It was not reported in any of the OCME's scholarly articles, and was not reported to the DNA Subcommittee of the Forensic Commission:

Q. Did you show your work varying the input quantity of DNA to see how the likelihood ratios were affected? Did you show that work outside of OCME to anybody that –

A. No.

Q. You didn't show that to the forensic subcommittee?

A. No.

Q. You didn't write it up in any of your papers?

A. No.
…

Q. Doctor, we were talking, when we broke, about the study that you said you did comparing variation of the rates, the quantity of DNA and likelihood ratios that are produced; you recall that?

A. Yes.

Q. Okay. And did show you any of that work to the technical leader?

A. No.

Q. The technical leader is Eugene Lien; right?

A. Right.

Q. He didn't sign off on any of that work?

A. No. It was just some exploratory analysis that I was doing.

Q. Did you make any notes?

A. Not that I remember.

Q. Produce any graphs?

A. No.

Q. Do you know where they are?

A. No.

Q. Do you know how many samples you looked at?

A. No.

Q. What kind of samples you looked at?

A. Two- and three-person mixtures.

Q. But you don't know how many?

A. No.

Id. at 30-32, Dr. Mitchell testified that PDFs were created of her results, but did not know where those documents were. Id. at 32.

In analogous situations, courts have rejected novel techniques based on underlying data or testing that was unrecorded. See People v. Wernick, 89 N.Y.2d 111 (1996) (expert can only

testify on material not in evidence, when data relied upon is of the kind ordinarily accepted by experts in the field); Saulpaugh v. Krafte, 5 A.D.3d 934 (3d Dept. 2004) (broad statements of general scientific acceptance, without accompanying support, are insufficient to meet burden of establishing such acceptance); Cumberbatch v. Blanchette, 35 A.D.3d 341 (2d Dept. 2006) (concluding that the plaintiffs failed to meet their burden of proving that their expert's theory of causation was generally accepted in the medical community. The plaintiffs' expert could cite to no relevant scientific data or studies to support his causation theory.). See also People v. Seda, 139 Misc.2d 834 (Sup. Ct. New York Cty. 1988) (noting that "explanation that [the DNA research scientist] acted as his own quality control does not excuse what must, at best, be seen as a cavalier approach entirely incongruous with the empiric nature of science").

iv. Accuracy

Dr. Caragine had testified on direct examination that her team considered "measuring the amount of DNA to be a very important step." Caragine, 12/12/2012, 41. She claimed that OCME's qPCR quant process had been optimized to the point that "[t]he accuracy is within thirty percent of the expected value, which is much more accurate than commercial tests." Id. at 41 .

While the OCME's method may be more accurate than that of the wider industry, it is not accurate enough for which the application is used. The OCME mistook tightness of fit for one purpose (the role of quant in DNA typing) to tightness of fit for another purpose (figuring out the drop-in and drop-out rates for a sample).

In casework, the qPCR quant process is like figuring out how much yeast to use to bake a cake. An analyst must determine how much DNA to put into the PCR reaction.. The qPCR quant process will tell the criminalist how much to dilute the sample if there is a lot of DNA, or

how much to concentrate if there is not much DNA. Dr. Shapiro testified that the quant does not have to be that accurate for the amplification step of DNA typing. Shapiro, 10/15/2013, 183 .

See also Butler, J. Fundamentals of Forensic DNA Typing, Elsevier, 2011 edition, p. 61 ("[w]hile this degree of imprecision may seem large, recall that a factor of two corresponds to one exponential-phase PCR amplification cycle; quantitation results are usually sufficiently valid to estimate DNA template amounts that will enable optimal PCR amplification."). .

In standard DNA typing, the quant can be off by a factor of two, but still a close enough estimate as to properly amplify the DNA. The OCME process is an even a closer estimate, shaving the two-fold difference down to a 30 percent range of error.

However, that 30 percent range of error is unacceptable for use in estimating drop-out rates. The FST provides only one statistical value. If a golfer slices a putt 30 percent one way or the other, how close to the hole will the ball end up? Therefore, the use of the qPCR quant to trigger a pre-determined set of drop out rates is misleading, in that it creates an unknown error rate on the back end of the process.

v. Error Rate of the quant has unknown effect on drop-out and LR

Dr. Mitchell admitted that the error rate of the quant would affect the accuracy of the drop-out rate on the final LR:

> Q. Now, the plus or minus 30 percent is a confidence interval, right?
>
> A. It's not a calculated confidence interval. I think it was an observation, here's what we expect the quant to be when we use this method that we now use, we are within 30 percent of that value.
>
> Q. Plus or minus 30 percent?
>
> A. Yes.

Q. That means that some of the time the quant's actually going to be higher than the estimate?

A. The true quantity.

Q. Right.

A. Yes.

Q. So you're actually going to apply, if we take a 30 picogram sample or a sample that's been estimated to be 30 picograms you're going to be applying the interpolated rate for 30 picograms?

A. Right.

Q. But it's entirely possible that that sample is 35 or 40 picograms, right?

A. Yes, it's an estimate.

Q. In that situation it will be applying a higher drop-out rate than it should for that quant value, right?

A. Yes, but it's an estimate.

Mitchell, 5/1/13, 97-98 . See Matter of State of New York v Rosado, 25 Misc.3d 380 (Sup. Ct. Bronx Cty. 2009) (finding that while actuarial risk assessments have been scientifically accepted as a means of predicting recidivism, actuarial testing has not been similarly accepted to determine the existence of a mental abnormality).

vi. qPCR Method Could Produce Up to 60 percent error rate.

The questionable use of the qPCR quant to calculate the drop-out rates is exacerbated by the lab's loose standards in allowing a quant assay to pass.

When reviewing the validation studies, Dr. Shapiro discovered that the calibrator samples, the 'reality check' samples for determining how good the standard curve is, were being passed in the FST validation, even when they were more than 30 percent off of the standard

curve. In determining whether to pass a quant, protocols allow an analyst to pass a quant run with up to 60 percent error.

The architects of FST were treating the samples used in their validation studies no differently than OCME had treated samples in casework before FST was even conceived. However, the ramification of this practice is that the quants calculated in the validation could be up to 60 percent off, plus or minus from the standard curve, and still were used to calculate drop-out rates. This range of error was never carried through to the end results.

Despite four days of extensive cross-examination of Dr. Shapiro, the prosecution never touched the issue of the inaccuracy of the quant. Nor did the prosecution raise the issue with Dr. Mitchell when the defense called her right after Dr. Shapiro, even though it asked her about different issues that Dr. Shapiro had raised in his testimony. Nor was the prosecution able to muster even a single witness on rebuttal to address whether a quant that varies by plus or minus 60 percent can be a reliable source for predicting drop-out rates.

Therefore, Dr. Caragine misled the Court when she testified on direct examination that the accuracy of the quant is within 30 percent of the expected value, because OCME protocols allow it to be within plus or minus 60 percent.

vii. Quant Calculations and Degraded Samples

Dr. Peter Schneider has written that "for quantifying DNA in heavily degraded samples the length of the qPCR target size plays an important role since the overall (degraded) DNA quantity is not necessarily the amount of amplifiable DNA via STRs." Schulz, Schneider, and Rothschild, "Absolute quantification of forensic casework samples using quantitative real-time PCR (qPCR) methods," International Congress Series 1288 (2006) 765-767 (emphasis added).

Dr. Schneider's concerns were echoed by Dr. Budowle at the hearing, who testified as to the incongruity of qPCR quant value and the final DNA output on the electropherogram:

> Q. And I believe you mentioned drop-out rate and FST already, but what ramifications does this observation have for OCME's FST methodology?
> A. The concept of using total DNA is – actually makes no sense, as I explained earlier, because we use total DNA to decide how much to put in. But we see from real casework the outcome is not consistent, it depends on the quality of the DNA if it's degraded, if it's got inhibitors, these sorts of things. Using just the total quantity, assuming that the drop-out rates are the same across, doesn't address each aliquot or the total in the process. We have to look at the end result signals [peak heights].

Budowle, 12/9/2013, p. 822 (emphasis added).

In her direct testimony, Dr. Caragine cited to one of the recommendations by the 2012 ISFG DNA Commission's recommendations in order to argue that OCME was properly basing its drop-out rates on internal validation: "(2) Estimates of drop-out and drop-in probabilities should be based on validation studies that are representative of the method used."

"Evidence of experiments is properly admissible so long as the proponent establishes a 'substantial similarity between the conditions under which the experiments were conducted and the conditions at the time of the event in question'" Styles v. General Motors Corp., 20 A.D.3d 338 (1st Dept. 2005) (quoting People v Laufer, 275 AD2d 655, 655 [2000]). Here, the OCME Research and Validation team had chosen to ignore the second part of the ISFG recommendation, which was that the studies had to be "representative of the method used." FST in casework is primarily used on touch samples, which showed signs of degradation in the FST validation studies. Dr. Caragine herself testified that touch samples naturally showed signs of degradation.

Dr. Caragine had not based the drop-out rates on those touch samples, but instead had based them on pristine, purposeful and non-degraded mixtures of just a few individuals. She had ignored the critical part of the ISFG recommendation at the expense of sound scientific practices.

b. Same Low Drop-Out Rate for Prosecution and Defense

i. Unique in the World

Another reason the FST drop model is not generally accepted in the community is that one of its procedures is to hold the drop-out rate constant between the prosecutor's hypothesis and the defense hypothesis. Dr. Budowle testified that this procedure is scientifically indefensible. Dr. Budowle pointed out that it is in the prosecutor's interest to keep the drop-out rate low in the numerator, to minimize the number of potential contributors that could fortuitously be included in the mixture. However, it is in the defendant's interest to keep the drop-out rate high, to include as many potential contributors that the defendant can point her finger at as the true contributor.

ii. Gill and and others: Defense should be the defense's

Other developers of LR drop models do not agree with or follow the restrictive procedure created by the OCME. Peter Gill, Hinda Haned, and David Balding allow for varying the drop-out rates between the numerator and the denominator.

Gill and Haned make clear in their article *A new methodological framework to interpret complex DNA profiles using likelihood ratios* that the defense hypothesis should be the purview of the defense. Furthermore, they show several examples of varying the number of contributors between the numerator and denominator, and the conservative effect that it can have on the ultimate LR. Forensic Sci. Int.: Genetics 7 (2013) 251-263.

iii. Dr. Mitchell admits high drop-out rate means anyone can be in mixture

Dr. Mitchell has revealed a reason that the defense would prefer a high drop-out rate in the denominator. When asked about the apparent lack of correlation between amplicon size and drop-out in the FST studies, Dr. Mitchell responded that it was important to look at the combined drop-out rate for partial and complete heterozygous drop-out in order to see any correlation. Mitchell, 5/1/2013, 41-42.

Ironically, Dr. Mitchell Dr. Caraginenever included charts showing combined drop-out rates for partial and total heterozygous drop-out. However, from the validation studies they are fairly easy to construct. Here are examples for the locus D2, drawn from Exhibit V4 in evidence, of actual combined drop-out rates (not the rates minus one standard deviation):

Locus	(5 pg) 25 pg Minor 4:1	6.25 pg Single Source	(10 pg) 50 pg Minor 4:1	12.5 pg Single Source
D2	88 percent	92 percent	71 percent	85 percent

Leaving aside the lack of continuity in observed drop-out rates for the different levels of contributors, it is apparent that at very small levels the combined drop-out rate is very high; there is a 92 percent chance of at least one of the two alleles dropping out for single source samples at 6.25 picograms. At double that amount, 12.5 picograms, the combined drop-out rate is still a whopping 85 percent for single source samples.

This effect is not lost upon Dr. Mitchell, who testified in the Manhattan Frye hearing that if the drop-out rate is 90 percent, then anybody could fit into the mixture. Mitchell, People v. Rodriguez, N.Y. Cty. Ind. No. 5471-2009, 1/28/2013, 178. That is precisely what seems to happen to a minor contributor to a 25 pg 4 to 1 mixture (or a 5 pg contributor), who although contradictorily has a lower drop-out rate than the single source 6.25 pg, still had an actual drop-out rate of 88 percent.

iv. Budowle shows the same rate creates a prosecutorial weapon

During his testimony, Dr. Budowle demonstrated using an Excel calculator showing that maintaining the same rate in both the numerator and denominator was consistently biased towards the defendant, because higher LRs were generated. However, increasing the drop-out rate in the denominator consistently lowered the LR.

1. single source rates

Dr. Budowle suggested the possibility that OCME apply to the denominator the empirically derived drop-out rates for single source samples, where the drop-out rates were consistently higher for loci with larger amplicon sizes than for the rates in two- and three-person mixtures. There are other problems for the calculations of their drop-out rates, but Dr. Budowle's suggestion at least gave the nudge to the OCME to begin an exploratory approach to applying drop-out rates, to move towards general acceptance.

c. Pre-set Drop-out and Drop-in Rates

i. Ignores Case-by-case variables

As Drs. Budowle and Shapiro made clear in their testimony, what sets the FST methodology apart from all other drop model approaches is that FST effectively ignores critical variables that are pertinent to the case, including the level of degradation of the sample, the variability of ratios of the contributors to the mixture, the possible relatedness of the contributors to each other and to any possible suspect, the position of the allele (e.g. stutter position) and the discriminatory power of the sample itself (that is, how many non-contributors can fit into the mixture). All of these variables would affect the drop-out and drop-in rates on a case-specific basis, and thus would affect the range of reasonable LRs.

ii. Underestimating Pre-set rates is not conservative by any definition

What Drs. Caragine and Mitchell tried to convey in their testimony about FST is that the rates might be wrong, but that they are "conservative." Dr. Mitchell stated that the FST approach worked reasonably well at keeping the LRs of non-contributors below 1 and keeping the LRs of true contributors above 1.

The first reason that this approach is not generally accepted is that it is completely unsupported in the scientific literature, as Dr. Mitchell had to admit:

> Q. Have you tested this hypotheses that underestimating the drop-out rate is conservative?
>
> A. Yes.
>
> Q. Is that test part of your validation study?
>
> A. No.
>
> Q. So where is that test?
>
> A. I've done that on my own.
>
> Q. So where is the data from that test?
>
> A. I don't have it to give you.
>
> Q. I mean physically where is it?
>
> A. I've done runs and looked at the results, I don't know that I've saved everything.
>
> Q. And you didn't publish on those results?
>
> A. No, this is not published.
>
> Q. Has anybody published a paper on that hypothesis?
>
> A. I can't think of one.

Dr. Mitchell, 5/1/13.. See Cumberbatch v Blanchette, 35 A.D.3d 341 (2d Dept. 2006) (concluding that the plaintiffs failed to meet their burden of proving that their expert's theory of causation was generally accepted in the medical community; the plaintiffs' expert could cite to no relevant scientific data or studies to support his causation theory).

The second reason that this allegedly conservative approach is not generally accepted is because it seems often to be wrong. Consider this exchange between the Court and Dr. Budowle:

> THE COURT: I guess I just don't see why that wouldn't be conservative. You said you were about to come up with a sample. You are telling me if you got a higher rate it includes more false positives and gives you a lower number for the guy who actually did it. If you then lower the rate, why would you not be getting fewer false positives and a better look at the individual who actually matches, then the standard deviation would seem to drop it even further?
> THE WITNESS: Right. In a sense it seems that way but it's, again, on a case by case basis. It's related to allele frequencies, the type of mixtures you have, the type of contributors that you have. I will show you examples, even if I think it's already been entered in the Balding and Buckleton paper shows it will go up and drop down again. It depends where you are on the process. So, there is a point as you increase the drop, as you lower the drop-out rate, eventually it's going to get low but there is a certain sweet point where it will be higher. If the drop-out is higher it will be a sweet point and give a more conservative value. There is a certain point where it's very high, it will go in the opposite direction. So, it all depends on a case by case basis.

Budowle, 12/9/2013, 829-830 [emphasis added].

d. Lack of a case-specific false positive test is flawed

Another reason that the FST drop model is scientifically unsound is that the arbitrary number that it generates is provided without any evidentiary context. Dr. Shapiro testified as to the fundamental importance of case specific performance testing in the context of other drop models under development:

> THE COURT: Doctor, can you explain to me how some of the other likelihood ratio programs out there would have been better for JB when the were used, assuming that JB's sample had been compared to the crime scene sample?.....
>
> THE WITNESS: . . . When the Gill program or the Adnet [Haned] program ran their 1,000 non-contributors, they would find a tremendous amount of false positives.
>
> They would say, well, this evidence is very, very weak because we're getting 50 percent false positives or 20 percent.
>
> Balding also, they have a discrimination model, where they can change and correct the likelihood ratio represent [sic]. So that they can correct the weight of the evidence based on how many people it discriminates.
>
> And also the Balding does the same thing. He will try 100 non-contributors and one case he had 19 out of 100 false positives. <u>That tells the other programs, well, you know, this is not very discriminatory evidence or not very strong evidence for anyone</u>.

Shapiro, 10/4/13, 92-94 (emphasis added)

In September 2013, Dr. Gill traveled to Australia to give a presentation at the ISFG yearly conference, entitled "On the meaning of the likelihood ratio: is a large number always an indication of strength of evidence?" Available at http://isfg2013.org/program/, *last visited* 4/7/14. His presentation was an extension of his argument in *A new methodological framework,* where he showed an example of how a large LR is not necessarily indicative of a suspect being part of a mixture. As his PowerPoint presentation of his talk made clear, Dr. Gill believes that an exploratory non-contributor test "must apply to all models – not just LRmix."

This is why he and Dr. Haned argued that false positive performance tests were "necessary" for responsible casework with a drop model. Their position in *A new methodological framework to interpret complex DNA profiles using likelihood ratios* is consistent with Drs. Budowle's and Shapiro's position that exploration of factors going into the LR should be case-specific.

Dr. Mitchell has tried to portray here recent research with case-specific performance testing as helpful, rather than necessary. Mitchell, 10/28/2013, 789. But her position is not

consistent with those of the leading drop model pioneers in her field. Therefore, the use of pre-set drop rates, absent any reality check of a case-specific performance test, is not generally accepted in this emerging field.

3. Flawed Experimental Design

On its face, OCME's approach to doing actual biological experiments to determine the relationship between the initial calculated quantity of DNA and the drop-out observed in the electropherogram would appear to be sound science. It seems consistent with the scientific method to test the relationship, document the findings and build a forensic tool consistent with those findings. The sheer number of experiments the lab performed in the FST validation seems at first blush to be impressive as well.

However, science also includes the rigorous testing of assumptions if they are key to the performance of a method. Once various assumptions failed to pan out, OCME should have rethought the viability of the assumptions, investigated the causes of their failures, or run more tests. OCME did none of these.

a. Allele Masking

i. Unique in world

In empirically estimating drop-out rates, OCME was the only lab in the world to use mixtures without taking into account allele-masking. Allele masking occurs when two contributors to a sample share the same allele at a locus, and therefore it is unknown if one of the contributor's alleles drops out if the other contributor's allele is present. Although Tvedebrink did not devise simplistic rates like OCME, he did perform a series of lab experiments to devise a logistic regression analysis for drop models to employ, for degraded and undegraded samples,

that was described at length in the 2012 DNA Commission paper. However, Tvedebrink specifically used single source samples in order to avoid the problem of allele masking.

ii. Not accounting for allele masking is unsound

When Dr. Mitchell was cross-examined about allele masking, the best response that she could muster was that "we didn't want to [create mixtures without masking] because in real life you do have sharing." Mitchell, 5/1/2013, 70. By taking this scientifically indefensible approach, Dr. Mitchell effectively created another unknown variable in her experiments.

It is unfathomable that the Research and Validation team would not know that their decision to ignore allele masking is not generally accepted in the field of forensic DNA typing. In Fundamentals of Forensic DNA Typing, John Butler wrote, "When performing validation studies, it is best to utilize samples that are heterozygous at all tested loci if possible so that allele drop-out can be monitored." Elsevier, 2011 edition, p. 334. One of the two non-OCME prosecution witnesses, Dr. Mitchell Holland, also admitted that allele masking was not an ideal experimental model when estimating the drop-out rate. Holland, 8/16/203, 83-84.

b. Linear interpolation vs. case-specific logistical regression

The second fundamental flaw of OCME's drop model was the assumption that the drop-out rates of pre-quantified DNA amounts would be linear. They assumed that after they obtained the averages of drop-out at each locus from six different quantities (25, 50, 100, 150, 250, and 500 picograms, respectively), they would then be able to track allelic drop-out in a neat linear progression. They expected that the six measured quantities would create a downward straight line on a graph where the y-axis is the drop-out rate and the x-axis is the pre-quantified amount of DNA.

i. Unique in the world

The OCME declined to publish the empirical drop-out rates for FST, claiming that any lab wishing to use the program would need to validate its own numbers. It did not even compile the empirically-derived rates in one place, choosing instead to bury them at the end of several summaries within the FST validation. It did have a set of final drop-out rates in its Executive Summary, but those were not published either. The OCME team did not even admit in their FST paper that they used linear interpolation between the six mileposts to create their drop-out rates.

Once the empirical rates were obtained through discovery, it became apparent why they resisted providing this data to the scientific community. Dr. Mitchell admitted on cross-examination that there were many instances where drop-out rates at a certain DNA quantity, even averaged over multiple samples, did not follow a trend of increasing drop-out when DNA quantity is decreased.

At some point, Dr. Mitchell had created graphs attempting to chart drop-out rate by amplicon length, which refers to size the product of the amplification process. These graphs illustrate how the data that OCME was creating was random, even after it was averaged over several samples. Dr. Mitchell tried to explain it away by saying that perhaps with more samples the data would have revealed a more discernible trend.

In her direct testimony, Dr. Caragine attempted to utilize a quote from the 2012 "DNA commission of the ISFG article to support the notion that the commission endorsed empirical estimation of drop-out rates. The quote was "Pr(D) can be estimated by logistical analysis [13, 14] or by using an empirical approach – for example [15]." People's Exhibit 2.

The use of that quote, however, is misleading. The empirically derived rates by OCME were based on averages of a few samples, and were interpolated in a linear fashion, an approach

that the commission does not endorse. See Marso v. Novak, 89 N.Y.2d 111 (1996) (noting that the essential defense theory was an attempt to portray a pattern of behavior not generally recognized in the relevant medical context and community).

The ISFG went on to write that, "The logistic regression is fundamental to an understanding of dropout." The article noted that logistic regression was employed by Tvedebrink, who had done his own set of empirical estimates in order to obtain the data to develop a logistic regression analysis for drop-out, and included an appendix explaining the Drs. Gill's and Robert Puch-Solis'logistic model for the estimation of drop-out.

Similarly, Dr. Budowle testified at the hearing that the OCME linear calculation of drop-out rates is flawed. The linear approach is too simplistic to deal with the vagaries of the stochastic region. Dr. Budowle pointed to recent research presented by Tim Kalafut of the U.S. Army. Dr. Kalafut presented his findings at the 2013 Promega conference, showing that drop-out rates are more appropriately calculated using a logistical regression analysis. See People v. Keene, 156 Misc.2d 108 (Sup. Ct. Queens Cty. 1992) ("The People's witnesses and defendant's witnesses were in complete disagreement on whether correcting for band shift by using monomorphic probes was generally accepted in the molecular genetics community.").

Dr. Caragine also testified on direct examination while a slide from her PowerPoint displayed a large banner that read: "ISFG: One Approach to Estimate Drop-out is Empirical." The slide included a picture of an excerpt of the 2012 DNA Commission paper, with a box drawn around the sentence, "Alternatively, $Pr(D_2)$ cab be empirically estimated."

Dr. Caragine's testimony here was also somewhat misleading. $Pr(D_2)$ refers to the probability of homozygote drop-out, where a suspect has inherited the same allele from both parents, and those identical alleles are not seen in the crime scene profile, such that the

probability of drop-out must be employed. In the same article, the DNA Commission authors had discussed a correction developed by Balding for the probability of homozygous drop-out, and the authors simply stated that a lab could also empirically determine that drop-out.

This was in no way an endorsement of OCME's flawed linear interpolation of drop-out rates, because the DNA Commission article then went into detail about the use of logistic regression in experimental design.

The logistical regression analysis has been adopted for use in other drop models, particularly Balding's LikeLTD, Loehmueller's LabRetriever, and Buckleton's STRMix. The OCME team actually stood alone in the entire world in developing a drop model based on linearly-interpolated drop-out rates. See Marso v. Novak, 42 A.D.3d 377 (1st Dept. 2007) ("Plaintiff interprets Nonnon to mean that generally accepted methodology such as differential diagnosis when properly performed leads to admissible expert conclusions. This case prompts us to add 'but not when there is a generally or widely held view in the scientific community rejecting such conclusions outright.'").

Incidentally, Dr. Kalafut from the U.S. Army is a principle architect of the DNA statistical program Armed Xpert. Dr. Mitchell testified that Armed Xpert is a likelihood ratio program that models drop-out and drop-in, and that she was on the advisory board for Armed Xpert. However, Armed Xpert has not been able to model drop-out and drop-in, and from Dr. Kalafut's presentation at Promega, it is clear that he and the other Army research scientists are not adopting the OCME approach of linearly interpolating drop-out rates.

c. Serial Dilution and Splitting Samples into Replicates

The use of the quant in the experimental phase was particularly problematic because the use of serial dilution can produce inaccurate results, particularly at the low end of the DNA

amounts of tested. It is akin to the problems of splitting samples into replicates in low copy number testing.

The pool of available molecules within a solution is often simply too small at the lower quantities of DNA distribute evenly, or to create a serial dilution. Therefore, Dr. Caragine misled the Court in her direct testimony and PowerPoint presentation, when she testified that a LCN sample of 60 picograms could be split evenly into three 20 picograms aliquots, with each replicate having an equal opportunity to amplify DNA within the aliquot Caragine, 12/12/2012, 109-110.

As Dr. Budowle testified to at the hearing, a criminalist cannot be expected to pull out the correct proportionate amount of contributor DNA molecules with her micropipettor when creating samples, because of the principle of multinomial distribution:

> Q: Will the amounts of DNA in each replicate be equal?
>
> A: That's one of the fundamental problems they have. The answer is no.
>
> Q. Can you explain why it would not be equal?
>
> A. Because it's an assumption that when you take a sample and split it into thirds that you are actually moving the DNA equally into each tube. So actually you do it theoretically mathematically or you can actually do experimentation.
>
> So, for example, if I were to have 18 picograms, and if I use that, because if you look at some mixture studies they have values below that, you distribute out. It would be something like that.
>
> Ideally, you put that in the three replicates, there should be [six] picograms in each one, six picogram and one cell of DNA I will explain why I've used that as a critical point.
>
> If you plan to take a sample in this tube and pull out one-third here, and put one-third here and put another one here, <u>you have only one in five chance of actually getting six picograms in each one</u>.
>
> Now, why is that important?

> Because you are going to have sometimes maybe twelve in one, seven in another, three in another. It will vary in every single case.
>
> If the quantities of DNA are different in each tube, the assumption that OCME starts with a quantity of that DNA is an indicator of the drop-out rate. The drop-out rate in each tube is different. That's not been compensated for. . .

Budowle, 12/9/2013, 819-820.

The problems of serial dilution studies in determining drop-out rates was also addressed by Drs. Haned, Gill and three other scientists in their article, *Estimating drop-out probabilities in forensic DNA samples*, FSI: Genetics 5(2011) 525-531 [entered into evidence as People's 20].

> Experimental data sets are the easiest adapted to study drop-out. Current methodology to determine the drop-out parameter is generally based on serial dilution experiments of disrupted cells. But strictly speaking, these experiments are only valid for haploid cells (i.e. sperm), hence conclusions about the robustness of models cannot necessarily be extrapolated to all kinds of biological materials.

Id. at 526 [emphasis added].

Dr. Caragine and the other designers of the FST validation made the mistake of using diploid epithelial cells, taken from saliva samples, to create the samples used to estimate the drop-out rates.

On redirect examination, Dr. Shapiro explained the significance of Haned et al.'s critique of using serial dilutions to estimate the drop-out rates:

> A. Yes. What happens when you do the serial dilutions is that you don't always sample when you have a diploid cell with 2 versions of the alleles in heterozygotes.
>
> When you do your serial dilution and pepenae [sic, "pipette"] et cetera, you will not always sample each of the chromosomes in each of those 2 alleles in the heterozygotes equally, and this becomes more and more of a problem when you have lower amounts of DNA and when you do a greater dilution.
>
> This was the approach taken at the FST. When they tried to determine the drop-out rates, they did different types of dilutions

> of the same samples, and it's totally not recommended by Dr. Haned, who described the difficulty and problems of using that approach. And that's why they recommend the simulation approach.

Shapiro, 10/28/2013, 734-735. Because each aliquot will not have the same amount of DNA, the OCME practice of splitting samples into replicates causes FST to produce unreliable results.

d. Minimal Mixture Ratios

Scientists have also rejected the OCME drop model because only a minimal number of mixture ratios for calculating drop-out were used. They only used samples with ratios of 4:1 and 1:1 for two-person mixtures, and 5:1:1 and 1:1:1 in three-person mixtures.

i. Unique in the world

No other pioneers of drop models use this simplistic scientific principle. Every other existing drop model pioneer acknowledges that the relative proportion of the contributors of the mixture should be estimated before attempting to estimate the drop-out rate.

The OCME had indicated at the end of each of the FST volume summary that they would use different ratios to test FST in the false positive study. They ended up throwing out all of the 10:1 mixture ratio samples, without explanation, and never used them. FST has never been shown to be tested with mixture ratios more extreme than the ones used in calculating the drop-out rate.

Dr. Budowle called into question the mixture ratio of 5:1:1 for smaller amounts of DNA, because of the problems with splitting the sample into thirds, combined with stochastic effects, prevented the lab from effectively deducing out the major contributor in many of their experiments. Dr. Budowle doubted that they would be able to glean any meaningful separation among the contributors unless they took a more severe ratio. Budowle, 12/9/2013, 846-847.

Even more disturbing, as Dr. Budowle pointed out in his direct testimony, the OCME never showed what the LRs were for the mixture samples that were used to create the drop-out rate. If the world does not know what LRs FST will create for mixtures of 1:1, 4:1, 1:1:1 and 5:1:1, how can the scientific community evaluate the numbers created by mixtures with the same contributors at different mixture ratios? The OCME seemed to have purposefully left off this starting point, so that their simplistic experimental design could not be fully evaluated. See Ratner v. McNeil-PPC, Inc., 27 Misc.3d 322 (Sup. Ct. Kings Cty. 2010) (noting that while plaintiff's expert had attempted to draw a medical parallel between proper doses and greater doses of acetaminophen to conclude that acetaminophen caused cirrhosis, that theory was not accepted within the scientific community).

Until an independent agency can get their hands on the FST program itself in order to test different mixture ratios, the court and the scientific community will never know whether the FST is reliable. That is why it is important to obtain either the program or the source code, so that at least the less extreme mixtures in the FST validation can be compared with the first four basic mixture ratios.

e. Varying Injection Times and Machine Sensitivity

As Dr. Budowle testified to, the OCME performed its FST experiments on different machines, sometimes changing sensitivities and conditions in the DNA typing process. Def. Exhibit G4, slide 55). These procedures also propagated an unknown amount of error in the calculation of drop-out and drop-in rates:

> Q. Did OCME hold the injection variable constant in order to account for its estimation of drop-out rate?
>
> A. It was changed throughout and it's part of their protocol, the degree of which I can't tell you because I don't have that information.

> Q. Did that raise concerns for you, that the change in injection parameters were not held as a constant variable in the validation?
> A. It' not, this itself, about taking the data and applying it to the interpretation in the case. If you had generated the data in one way then you use a different method in another, they don't necessarily comport. They may be close, they could be far off. You have to know by the effect that occurs, you need to quantify that effect and that, I think, needs to be done.
>
> Q. Did OCME quantify that effect?
>
> A. I don't believe so. They may have data, I just don't think it's been compiled.

Budowle, 12/9/2013, 870 (emphasis added).

To make matters worse, as Dr. Shapiro has pointed out, actual FST casework is done on a wider variety of machines, that produce even more variable results.

f. Number of samples used to estimate drop-out

In OCME's FST article, they insinuated that thousands of samples had been used to calculate the drop-out rate. Dr. Caragine performed the same kind of misrepresentation in her direct testimony at the hearing when she claimed, "we counted how often we saw drop-out in thousands of samples." Caragine, 12/13/2012, 29. This is misleading at best.

OCME was able to perform thousands of amplifications associated with the FST validation. However, they did not use huge chunks of research in the validation at all. They cherry-picked the degraded samples that they created. They excluded entire studies worth of samples from their final false positive validation. They threw out all of the 10:1 ratio two person samples that they had said they would use in the false positive study. And they did perform a huge false positive study at the end of their validation, but the samples used in the false positive study had NOT been used in the earlier studies to calculate the drop-out rate. So essentially they

conflated the two sets of samples, and misled the scientific community and the courts in the number of samples that they used.

Their numbers for two-person mixtures were lower than one would expect from Caragine's testimony. As Dr. Shapiro explained with Def. Exhibit HH , there were 22 combinations of contributors for 31 cycle testing. Those 22 combinations produced a total of 468 amplification runs. There were 20 combinations of two-person mixtures run at 28 cycles, totaling 300 amplification runs. Therefore, there were a total of 768 two-person mixture amplifications.

The OCME numbers for three-person mixtures were even lower. OCME used only six combinations of contributors for all of the three-person mixtures. They produced 216 amplifications run at 31 cycles, and 128 amplifications run at 28 cycles. The total number of amplifications to calculate the drop-out rate used in FST was 1,112.

OCME's use of fewer samples for the three-person mixtures is flawed, since there can be overlap between the number of alleles observed in two- and three-person mixtures. Perez et al., Croat Med J. 2011; 52:314-26. Four-person mixtures are not recognized correctly as four-person mixtures 70 percent of the time. Buckleton et al. Forensic Science International: Genetics 1 (2007) 20–28; Paoletti et al. J Forensic Sci, Nov. 2005, Vol. 50, No. 6; Haned et al. J Forensic Sci, January 2011, Vol. 56, No. 1; Perez et al., Croat Med J. 2011; 52:314-26.

This phenomenon was borne out in the false positive study, when 31 four-person touch samples were interpreted 14 times as two-person mixtures, and 17 times as three-person mixtures. Actual three-person mixtures were interpreted as two-person mixtures 40 times. As OCME reported, the false positive rate for three-person mixtures is higher than two-person mixtures, which would have implications for their overall false positive rate.

Therefore, the need for thoroughly testing drop-out in minor contributors for complex three-person mixtures is just as crucial if not more important than calculating drop-out for two-person mixtures.

The Court might ask, why is it a problem if they used only 1,112 to calculate the drop-out rate, instead of the "thousands and thousands" that Dr. Caragine described, when the lab exceeded the number of samples recommended in the SWGDAM guidelines? There are multiple answers to this question.

First, this is only time that any lab in the world tried to determine empirically a linear association between qPCR quant and the drop-out rate for samples at different sizes. Dr. Van Daal put it succinctly in cross-examination:

> Q. How many samples are normally or usually included in a validation study?
>
> A. That's a little bit like how long is a piece of string? Because a validation study includes a whole range of different types of samples and it's a reasonable coverage of all the range of samples and depending on the purpose of the validation. So if it's a totally new technology that's being validated, that would be different than modification to a technology. So that's a little difficult to answer and there is no specific answer to that.

Van Daal, 11/22/13, 416- 417.

The second answer is that the results certainly indicated that they did not use enough amplifications to confirm their hypothesis of a linear phenomenon of drop-out rates in the stochastic region. The graphs charting drop-out rates by amplicon size indicated that no trend whatsoever in their data. Dr. Mitchell herself had to admit that numbers were all over the place.

The third answer is that they did not use enough combinations of contributors in their mixtures to calculate the drop-out rate, which is the subject of the section below.

g. Number of Contributors Used to Estimate Drop-In and Drop-Out Rates

In *Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in*, Mitchell et al. claimed that "samples of various combination of 85 contributors were amplified in duplicate or triplicate and analyzed for the purposes of drop-out and drop-in rate estimation." Forensic Sci. Int. Genet. 6 (2012) 749-61.

That statement is very misleading of the number of contributors they used to estimate their drop-out rates. As Dr. Mitchell admitted on cross-examination, the number was closer to ten. Mitchell, 5/1/2013, 60.

It matters that the OCME used only ten contributors to estimate their drop-out rates. As the DNA Commission of the ISFG pointed out in 2012, the calculation of the drop-out rate is a new enterprise. . In the addendum to its report, it suggested that a sample size of 100 profiles would be sufficient for conducting experiments to determine the probability of drop-out. Forensic Sci. Int. Genet. 6 (2012) 679-688.

It matters that OCME used only ten, rather than 85 or 100 contributors because of allele masking. As Drs. Chakraborty, Dr. Budowle and Dr. Shapiro raised in their testimonies, allele masking is exacerbated by the same combinations of contributors. The drop-out rate will therefore be distorted downward at certain loci in unpredictable random ways. The application of some correction like flattening out the drop-out rate or applying one standard deviation will not fix the inaccuracy of the data, because those corrections will not be applied in a uniform way, as Dr. Shapiro pointed out.

h. Diversity of Contributors used to Estimate Drop-Out and Drop-In Rates

In the FST paper, the OCME also claimed that the contributors used to estimate the drop-out "represented the diversity of New York City."

Since the use of 85 contributors has been shown to be false, the claim to diversity for the primary 10 contributors is likely false as well. To this date, the lab has been unwilling despite requests to provide the racial background of its contributors. In fact, the legal counsel to the Forensic Biology department had to admit to the Court that Drs. Caragine and Mitchell had tallied the races of the laboratory personnel contributors, but then thrown them away. The OCME has refused to reconstruct the tally of racial identities that Drs. Caragine and Mitchell created.

As Dr. Chakraborty testified, it is important to know and account for the racial background of contributors when calculating the drop-out rate. He noted that there could be cryptic relationships among contributors that could lead to allele masking and subsequent distortion of the drop-out rate.

Dr. Budowle graphically demonstrated the danger of ignoring the racial backgrounds of the contributors when determining the drop-out rate. In slide 79 of his presentation, he showed the significant differences in allele frequencies between African Americans and Caucasians for particular alleles and loci. The more frequent alleles in a mixture of contributors of the same race will lead to more masking, particularly when those combinations are repeated over and over, as occurred in the FST validation. To not document or take into account the racial background will be to distort the accuracy of predicting allele masking in casework.

4. Other FST Model Design Flaws

a. Only one scenario for number of contributors

When setting up its methodological approach, OCME scientists decided to take the expedient route when factoring in the number of contributors to a mixture, at the expense of good science and the accused. In the FST, the analyst estimates the minimum number of

contributors. This is a key variable and alters the drop-out rates for any alleged contributor to the mixture.

As elicited at the hearing, there has been research by Buckleton that there is a moderate risk of a non-minimal LR when the minimum number of contributors is estimated. See Perez et al., Croat Med J. 2011; 52:314-26, citing Buckleton.

There have been calls by the leading experts in the field to take an exploratory approach to analyzing mixtures, including the consideration of a number of different scenarios with differing numbers of contributors. For example, Dr. Charlotte Word in a mixture interpretation workshop and webcast sponsored by the National Institute for Standards and Technology, suggested interpreting complex mixtures as three-, four-, five- or even six-person mixtures. See Complex Mixtures (3 & 4 person, relatives), http://www.nist.gov/oles/forensics/dna-analyst-training-on-mixture-interpretation.cfm.

Furthermore, research in the field has shown a high error rate when estimating the number of contributors in complex mixtures. Most often, analysts underestimate the true number of contributors. If a non-minimal LR is created when the number of contributors is wrongly underestimated, that LR will bias the defendant.

The FST never takes into account the moderate risk that it is producing a non-minimal LR by using only the minimal number of contributors. However, the leading pioneers in the drop model field – Balding, Buckleton, Gill – have shown instances, from both simulated and real-world criminal cases, showing that a greater number of contributors will benefit the defendant.

Because estimating the number of contributors will directly affect the drop-out rate, it is a crucial aspect of the FST drop model. Therefore, the FST methodological approach of providing only one scenario for the number of contributors seems to not be generally accepted in the relevant scientific community.

Changing the number of contributors will also change the LR of a suspect, independent of the drop-out rate. Therefore, it is even more crucial that a drop model that has an unknown rate of error to follow Dr. Word's admonition to use a number of different contributor scenarios.

It is also worth noting that OCME scientists recorded zero tests in their 47 binders comparing the LRs of the same mixture under different contributor scenarios, to establish that inputting the minimal number of contributors will consistently lead to a lower LR.

The prosecutor cannot argue here that the defect in the program can somehow be cured by cross examination. Every case is fact-sensitive, and since the defendant is denied access to the program, any cross-examination on whether changing the number of contributors will be objected to as speculative.

b. Ignoring Peaks Below Threshold

The FST does not consider peaks below the 75 RFU threshold set up by the lab. Other pioneers of the drop model, like David Balding and Mark Perlin, have recognized the need to examine the data below the conventional analytical threshold in order to attempt an accurate estimate of the weight of the evidence. In the simplistic FST approach, peaks in a low copy sample that do not reach the 75 RFU threshold because of peak imbalance are completely excluded. While that approach may benefit a defendant whose alleles are present below threshold; it may prejudice a defendant who is a non-contributor. The FST drop model is biased against that defendant.

The lack of an ability for FST to take account of peaks below threshold cannot be cured at trial by cross-examination, because the defendant does not have access to the program itself to run it after interpreting the capillary electrophoresis under a lower threshold, for example 50

RFU, which most labs use. Therefore, any cross-examination on that subject would be met with an objection that it calls for speculation.

c. Ignoring the Possibility of Relatedness

When the OCME rushed the FST from the design phase to the validation phase, they purposefully ignored the opportunity to include in their drop model the capability to consider relatedness between contributors. The Research and Validation team decided to take the time-saving shortcut of not modeling relatedness, thereby treating all contributors as unrelated.

Other drop models include relatedness, including Balding's LikeLTD. Like Weir, Balding helped develop the math for modeling relatedness in complex mixtures. As Dr. Rohlfs demonstrated at the hearing, Balding's use of intra-individual theta can easily be incorporated into the defense hypothesis of FST at one locus, making it possible to incorporate into the entire program. In fact, Dr. Rohlfs showed an example of where failing to incorporate the inter-individual theta could result in bias against the defense hypothesis.

It was scientifically indefensible not to follow Balding's model. Relatedness is commonly a part of the defense hypothesis. Also, related individuals will share commonly touched items, like bicycle handlebars or door knobs. To ignore the possibility of relatedness in complex mixture is to ignore the realities of touched samples, the type of evidence sample that FST is employed with the most.

The presence of related individuals cannot be predicted. As Dr. Shapiro testified to, in *Estimating the number of contributors to two-, three-, and four-person mixtures containing DNA in high template and low template amounts*, the OCME's team randomly pulled 24 morgue profiles and used them for contributor studies, including the FST false positive study. One 61

picogram three-person mixture included so much allele sharing that the authors hypothesized that two of the contributors were brothers. Perez et al., Croat Med J. 2011; 52:314-26

Regardless of whether they were related, the drop model obviously must be prepared to encounter significant allele sharing to the extent that individuals appear related. It is not only a shortcoming of FST, it is not generally accepted.

Dr. Caragine's only response on the stand was that OCME could test a sample from the brother. This position unconstitutionally shifts the burden to the defendant, when the modeling for relatedness could have occurred long before the first test tube was filled in the FST validation.

d. Drop-in and other contamination are drastically underestimated

Drop models must not only attempt to explain away what is not there, but also what is there by contamination. Drop-in is simply contamination by another name. But after implementing its controversial low copy number testing in 2005, the OCME sought to distinguish between the occasional foreign allele that shows up in one of the three replicates, with what they defined as 'gross contamination,' that indicated another profile had somehow seeped into the sample.

When developing the FST drop model, the OCME sought to continue this artificial drop-in/gross contamination. They used pristine lab samples to calculate a drop-in rate that they could then apply to touch samples and other real world crime stain samples.

The OCME then made the unsupportable and undocumented assertion in their FST paper that underestimating the drop-in rate is likely conservative. If a defendant's profile shares many alleles with the evidence profile, a higher drop-in rate will lower the likelihood ratio. If a

defendant's profile shares very few alleles with the evidence profile, a higher drop-in rate will lower the likelihood ratio.

The OCME has failed to present data that supports this conclusion. However, Dr. Budowle demonstrated in a number of examples that increasing the drop in rate lowered the LR. Thus, OCME's assumption is unfounded.

The mock touch samples created for the FST validation displayed an enormous number of non-contributor alleles. For example, Pen B was interpreted as a three-person mixture, even though there was another profile or two reflected in the mixture. One of the three true contributors was identified as a major component, while the other two contributors shared enough alleles with the three evidence replicates that they were declared to be "cannot be excluded," precisely the qualitative definition for which FST was developed.

PenB also had a total of nine false positives, seven of which had LRs above both of the true contributors that had been defined as "cannot be excluded." One of those false positives, JB from John Butler's Caucasian American subpopulation of the NIST database, had an LR of 157, which is rated as strong evidence that the evidence is 157 more likely to be JB and two unknown, unrelated individuals than if it were three unknown, unrelated individuals.

As Dr. Budowle pointed out, the FST drop model ignores the realistic rate of contamination in touch samples, artificially lowers the drop-in rate, and prejudices every defendant. This is a reason the FST's approach of empirically deriving drop-in rates from pristine samples will never be accepted in the scientific community.

5. Manipulating and Burying Data

When OCME scientists obtained incongruent or problematic results, they did not attempt to revise the model. Rather, they ignored these results, or buried empirical data. This process did not make for a model that was generally accepted in the scientific community.

Throughout her testimony, Dr. Mitchell acknowledged that the drop-out rates were not accurate. She did claim, however, that she took efforts to lower the drop-out rate as much as reasonably possible, so that the LR in each case would be on the lower end of the range. Although she did not explicitly state it, the insinuation was clear: The LR is going to be lower than it would be if the rates were actually accurate.

a. The One Standard Deviation Correction

Dr. Mitchell admitted on cross-examination that with a normal distribution of data, two standard deviations above or below the mean would yield a 95 percent confidence interval. Mitchell, 5/1/13, 139. However, she testified that she chose to only go down one standard deviation from the 'observed' rate. Going up or down one standard deviation provides only a 66 percent confidence interval.

When questioned why she did not go down two standard deviations from the mean, Dr. Mitchell claimed that she had done preliminary testing with two standard deviations, but that it was "too low." Mitchell, 5/2/13, 15. Dr. Mitchell admitted that this testing was unrecorded, and was not reported to the DNA Subcommittee of the Forensic Commission, nor was it mentioned in any of OCME's scholarly articles. Id. at 137-138. She also testified that she had used "maybe" 20 samples in this testing. Id. at 17.

> Q. So you didn't do any kind of testing to, let's say, see what happens if you use drop-out rates that are one standard above that value?

A. I did not.

Q. But you did test to see what happens when you did two standard deviations below?

A. I did that.

Q. But that data is in the –

A. It's not.

Q. You just threw out that data?
Mr. Reeves: Objection, your Honor.

The Court: Did you throw it away?

The Witness: It was an exploratory phase, it wasn't part of the final product that we're using.

Q. So you didn't give it to the DNA subcommittee?

A. No, I don't believe so.

Q. And you didn't publish the results in a paper?

A. No.

Q. You didn't turn that over to us?

A. No. I don't have it compiled in a certain place.

Q. And you didn't even mention in your paper that you did that, right?

A. No, I don't think we did.
…
Q. You didn't look at the data and then determine one standard deviation below was the correct level to set the drop-out rates at?

A. We looked at a small set of the data. So we didn't run the entire validation with both settings. We looked at a small sample, a small set of samples, and determined that two standard deviations below was too low and we selected one standard deviation.
…

> Q. The small set that you tested or you amplified, before make the one standard deviation determination, how many samples are we talking about?
>
> A. I don't remember. 20 maybe.

That data was either unpreserved or was discarded. It is also in violation of the editorial requirements of FSI: Genetics, which requires that all data upon which a published article is based be preserved for review. The Court should take into consideration Dr. Mitchell's unwillingness to preserve this data in assessing her credibility, and give itself an adverse inference charge when weighing her testimony about whether the FST drop model is generally accepted in the relevant scientific community as reliable.

Applying one standard deviation to all of the drop-out rates was also problematic, according to Dr. Shapiro, because the deviation was highly variable. Dr. Shapiro showed an example of how the standard deviation could bias a defendant more than use of the empirical drop-out rate. This outcome is not surprising, considering all of the flaws pointed out by Dr. Budowle in the experimental design of the FST.

b. The Minimum and Maximum Drop-out Rate Corrections

Dr. Mitchell testified that minimum and maximum values of drop-out were made part of the FST program. She claimed that these minimum and maximum rates were made in consultation with Dr. Gill, along with testing that Dr. Mitchell had done herself.

Dr. Mitchell was confronted on cross-examination with the fact that some drop-out rates OCME programmed into FST were actually higher than some drop-out rate observed in their validation studies for certain sized samples. She was asked why the minimum values for those samples could not reflect reality more, for example putting .0001 percent drop-out rate for a sample instead of 2 percent.

Dr. Mitchell's response was that setting the minimum value below 2 percent was a problem because too many true contributors were excluded. She also claimed that they had not done enough samples to ascertain whether the minimum value could be lower. Mitchell 5/1/13, 127-128.

Dr. Mitchell's response here flies in the face of her earlier testimony, where she claimed that a difference in a drop-out rates of 30 percent did not change the statistic created by FST to any great degree. Id. at 97-98. How can a 30 point difference between 20 percent and 50 percent not be that significant, while the less than 2 point difference between 2 percent and .0001 percent create skewed results where true contributors are excluded?

Equally troubling is that Dr. Mitchell's research on determining the final minimum and maximum drop-out rates were unrecorded, and was not reported to the DNA Subcommittee of the Forensic Commission, nor was it mentioned in any of OCME's scholarly articles, other than to state what the minimum and maximum rates were, not how they were determined. That data was not saved, or it was thrown out.

> Q. So you couldn't have set the minimum drop-out rate to .0001?
> A. We could have but that is not in the range of reality.
> Q. Isn't actually zero what you observed at a lot of these loci for a lot of these mixture weights reality?
> A. In a sample of 100.
> …
> Q. So you would need to look at a lot more samples to determine the true drop-out rate at those values?
> A. If it's that low, yes.

Id.

c. The Fixing or Flattening Out of Data Correction

In certain instances, when the observed drop-out rates went counter to her general assumption, Dr. Mitchell lowered the observed drop-out rates to make it more consistent with her

assumption that drop-out rate decreased with the size of a DNA sample. Dr. Mitchell resisted calling this kind of action a "correction," Mitchell 5/2/13. 51, but there is no other reasonable explanation for OCME taking this action. When the data did not behave as they expected, they flattened out or fixed the aberration to fit their expectations. Dr. Mitchell could not recall whether this action was mentioned in the peer-reviewed article about FST, Id. at 53, or if she reported it to the DNA Sub-committee. See also id., at 18-21; People v. Seda, 139 Misc.2d 834 (Sup. Ct. New York Cty. 1988) ("in light of the evidence of inconsistencies and lack of memory by the witness, the court finds itself unable to make such a determination" of general acceptance).

What this hearing has brought to light, however, and what the Court should not ignore, is that nowhere in those 47 binders is this recorded. In the thousands of pages of studies, there is no evidence whatsoever of a comparison between the LRs using the empirically-derived rates, the one standard deviation rates, and the two standard deviation rates. The only arguable exception are the three pages in Volume 24 which compare mixtures using both deduced and non-deduced rates, showing that FST is as good as flipping a coin for true contributors in predicting the direction of LR values.

d. If conservative does not work, change its definition

The OCME practiced unsound science when it developed the FST software package that was biased against the defendant, but then tried to sell it as "conservative to the defendant." By doing so, the architects of FST attempted to literally change the definition of a fundamental forensic concept, in order to perpetrate such a bias.

In the Forensic DNA community, the term "conservative" is universally defined as "an assignment of the weight of evidence that is believed to favor the defense."

http://www.dna.gov/glossary; http://www.cstl.nist.gov/biotech/strbase/glossary.htm, J. Butler, Fundamental of Forensic DNA Typing, Elsevier 2010, App. 1, p. 444.

Dr. Budowle accentuated this universal definition of conservative in his testimony by calling attention to a quote in the Overview Section of the NRCII: "When in doubt, we err on the side of conservatism (that is, in favor of the defendant)."

Dr. Budowle also noted that standard DNA typing procedures from the onset have built in conservatism that attempt to not overstate the strength of the statistical analysis in forensic casework. Whenever possible, standard DNA typing procedures also seek exclusion of a suspect, rather than gauge the probability of the suspect's inclusion in a mixture. With these approaches working together, standard DNA typing procedures meet the criterion of conservatism espoused in the NRC II report, that is, assigning the weight of the evidence in a way that is believed to favor the defense.

Dr. Caragine testified that aspects of the OCME's protocols for DNA Analysis were conservative to the defendant:

> THE COURT: Why would you not want – assuming you have one at every locus or two at every locus, why would you not want to use those peaks to create a second profile?
>
> THE WITNESS: What we're doing is avoiding labeling very minor contributors to a mixture. So, these are things beyond a one to ten ratio and this in essence, favors the defense. It's a very conservative approach to looking at the results [emphasis added].

Caragine, 12/12/12, 65-66 [emphasis added].

However, after Dr. Caragine resigned in April 2013, Dr. Mitchell admitted that the term 'conservative' she and Dr. Caragine had used in their scholarly article on the FST was not the one accepted in the scientific community:

> Q: Now, is there any way for you to know that the lower rate is more accurate as opposed to the higher rate being more accurate:
>
> A: We wanted the more conservative rate.
>
> Q: To you always underestimating drop-out is more conservative?
>
> A: Yes.
>
> Q: Always?
>
> A: Not that a lower drop-out rate will always give a lower LR, that's not what I mean by conservative.
>
> Q: It depends on the sample, right?
>
> A: It depends on – We want a low drop-out rate for none [sic] contributors so by choosing a lower – we want a low LR for none contributors so by choosing a lower drop-out rate we are erring on the side of lower LR for a none contributor.

Mitchell, 5/1/13, 113-114 [emphasis added].

By switching definitions in the application of the FST, Dr. Mitchell departed from a basic approach where the defendant accused of the crime is afforded every reasonable estimation of the weight of the evidence, to one where there are fewer random matches to the mixtures, an approach that will inherently strengthen the prosecution's case, at the expense of the defense case. This arbitrary changing of the definition of 'conservative' becomes biased against the defendant.

Dr. Budowle made this very point in his testimony. He noted that a lower drop-out rate may not be conservative as the OCME had claimed. This redefinition of the term failed to address the traditional concerns that the scientific community has had for not overstating the strength of the evidence against the defendant. Finally, he advocated an implementation of a drop-out rate that is representative of each particular hypothesis in the likelihood ratio, and not to

foist an inaccurate set of drop-out rates on the defense hypothesis that is essentially not conservative.

What makes this definitional about-face so egregious is that Drs. Mitchell and Caragine failed to clearly redefine the term conservative in their 2012 paper or in their sworn testimony to the Court, until Dr. Mitchell was confronted with the fact that FST was in some cases not assigning the weight of the evidence in a way that is believed to favor the defense. Then she was forced to reveal what their true intentions were: an approach that is actually anti-conservative, in that is biased against the defendant.

e. Burying the Drop-Out Rate Comparison

One of the basic stated positions that Dr. Mitchell and Dr. Caragine espoused when applying FST was that underestimating the drop-out rate would lower the likelihood ratio. Dr. Mitchell claimed on cross-examination that she had done studies to determine that indeed underestimating the drop-out rate would lower LRs. Dr. Mitchell admitted that this testing was unrecorded, and was not reported to the DNA Subcommittee of the Forensic Commission, nor was it mentioned in any of OCME's scholarly articles. She also could not think of a single other scientist that had published studies to support this position. Dr. Mitchell, 5/1/13, 117.

Other than the preliminary runs, Dr. Mitchell denied comparing two sets of drop-out rates, anywhere in the FST validation studies:

> Q. As part of your validation of the drop-out rates, you didn't use different drop-out rates, you used different contributors?
>
> A. Yes.
>
> Q. And, so, you didn't compare two sets of different drop-out rates?
>
> A. No.

> Q. And no part of the validation was comparing one set of plausible drop-out rates to your set of drop-out rates?
>
> A. No. Early on when we were – when we made the determination that we were going to use one standard deviation below the estimate, before we made that decision, we did look at two standard deviations below the estimate. Not with very many samples, but with a few samples to see if that was a real realistic rate to use.

Id., 5/2/2013, 14-15.

That testimony does not seem to be entirely accurate. The OCME had performed a study comparing higher and lower drop-out rates applied to the exact same sample, and they performed 36 such comparisons within that study. That study was Study 3C in the larger FST false positive study, where 36 pristine three-person samples were run side by side with both higher deducible rates and lower non-deducible rates.

The results of Study 3C indicated that the higher deduced rates lead to lower LRs for true contributors almost exactly half of the time. As Dr. Chakraborty put it, "it is as good as tossing a coin." Chakraborty, 12/16/2013, 1127.

The higher deduced rates did lead to more false positives in Study 3C, as Dr. Budowle had predicted it would in his testimony; higher drop-out rates do generally lead to more fortuitous matches. But interestingly, the OCME performed no head-to-head comparisons to indicate how the drop-out rates affected individual non-contributors, to see if some actually had lower LRs with higher drop-out rates, as happened half the time with true contributors.

Most importantly, the OCME scientists never performed a side-by-side study with touch samples to show the affect of changing drop-out rates on LRs with these more complicated samples.

In fact, Dr. Mitchell decided in the middle of this Frye hearing to do yet another study, in the Summer of 2013, in an attempt to show that lowering drop-out rates will lower the likelihood ratio. OCME has resisted turning over this data, despite the fact that Dr. Mitchell brought it up and relied on it her continued testimony in the Fall of 2013. Even though she claimed that the preliminary results indicate a direct correlation, the lab has not opened up her research to the scientific community. She has not informed her peers of the details of the study, such as how many combinations of contributors that were used, or how representative her results are.

Dr. Mitchell claimed that she was performing this study only out of curiosity. Mitchell, 10/30/2013, 40. The Court should reject this response as incredible, since it involves a fundamental assumption about FST. See People v. Seda, 139 Misc.2d 834 (Sup. Ct. New York Cty. 1988) ("Dr. Shaler failed to make recordings with regard to any of the three electrophoretic setups that were performed on each of three bloodstains, saying instead that had anything of consequence occurred he would have recorded it and repeated the analysis. Dr. Shaler's explanation that he acted as his own quality control does not excuse what must, at best, be seen as a cavalier approach entirely incongruous with the empiric nature of science")

What the Court has before it, however, are the results of Study 3C, which calls that assumption into question. The Study 3C results strongly indicate that FST does not do what its makers purport it to do, which is provide an accurate or conservative statistic. It is therefore not generally accepted. In addition, Dr. Shapiro provided a one-locus example in his testimony showing that a higher drop-out rate could lead to a lower LR. Finally, there are the myriad examples in the published literature brought to light at the hearing showing that a higher drop-out rate can lead to a lower LR.

f. Burying the Comparison of FST to Forensim

OCME will not turn over the FST source code. Its refusal prevents defense experts from comparing the LRs created by FST to those of other drop model programs. The purpose of such a comparison would be to see whether the statistical values generated by FST would be reasonably similar to other programs that make similar calculations.

However, Dr. Mitchell testified on cross-examination that she did precisely the same kind of comparison between FST and LRMix, an open-source program developed by her friend and colleague Dr. Haned of the Netherlands Forensic Institute, only one of two non-OCME scientist to have testified at the Brooklyn hearing. Both Dr. Haned and Dr. Mitchell testified that the pair had done a few informal sessions where they compared results from six samples or less. This testing was unrecorded, and was not reported to the DNA Subcommittee of the Forensic Commission, nor was it mentioned in any of OCME's scholarly articles, the latest of which came out in December of 2012. Neither Dr. Haned nor Dr. Mitchell could provide specifics in their testimony about the sizes of the samples used.

g. Burying the Failed Degradation Model

In OCME's FST paper, Dr. Mitchell et al wrote, "[u]ltimately, it was determined that, in general, use of the degradation module as programmed resulted in LRs closer to 1.0 for both true contributors and non-contributors. That is, this approach did not increase the overall separation between true contributors and non-contributors (data not shown)." Def. Exhibit I.

The validation studies for the model that was used, however, includes no comparison of the rates from the degradation module and from the rates in the final version. That is, the legal and scientific communities have never seen a comparison between the current FST and a version

whose drop-out rates may be more realistic. By doing so, the FST seems to be overstating the strength of the evidence in casework.

As Dr. Budowle pointed out in his testimony, degraded DNA is common in forensic casework. It would have been useful for DNA experts to have seen the drop-out rates used in the degraded model, as well as the LRs created by that model, to compare to the online FST. However, that data was apparently removed and replaced.

The degraded module was apparently tested as part of Volume 15C, which said: "the experiments here were not used in the final validation, because they depended on a degraded module that was thrown in the garbage." See Def. Exhibit. V4.

However, of the samples in Volume 15C, that were supposed to be re-run in FST, the numbers are exactly the same as the numbers in the final study. OCME replaced them with the numbers that they ran in the final validation study.

h. FST Validation Failed to Demonstrate That the Drop-Out Rate Across Loci is Independent

Another problem that was raised in the DNA Subcommittee was whether the drop-out rate was independent across loci. As Dr. Chakraborty explained at the hearing, if drop-out among some of these loci is non-independent, it invalidates the product rule employed by FST. FST employs drop-out rates, and if the DNA fragments at different loci drop out in some inter-related way, the use of the product rule is invalid. Dr. Chakraborty further testified that after he voiced his concerns back in 2009, the OCME performed more research, providing him at the next meeting with analysis of intermediate results, that indicated that drop-out was indeed independent across loci. Chakraborty, 12/17/13, 1118. However, after he received the full validation studies including the full Volume 22, Dr. Chakraborty questioned whether indeed there was independence of the drop-out rate across loci.

With the full volume, Dr. Chakraborty could review tables which showed some correlation among drop-out in loci. He examined their conclusion that "Within each set of mixtures, drop-out at some loci was associated with drop-out (or lack of drop-out) at other loci. However, these associations were not consistent across the mixtures, indicating that there is no consistent correlation in drop-out probability among loci." See Def. Exhibit S4, p.3.

Dr. Chakraborty then examined their overall conclusion, which was that "Drop-out rates appear to be independent across loci. That is, drop-out or lack of drop-out at each locus is not consistently associated with an increased or decreased probability of drop-out at other loci." Id. p.2; hearing testimony, Chakraborty, 1114.

Dr. Chakraborty pointed out that the two statements in their conclusion were contradictory with their analyzed data. When OCME performed regression analyses on the data in these loci, they should not have had so many positive and negative correlations in their tables. The tables should have been blank, indicating no correlation at all. Instead, these tables were loaded with a "substantial number of positive and negative entries in these seven tables, right?" Id. at 1116.

Dr. Chakraborty agreed with Dr. Mitchell's conclusion that the drop-out rate was not "consistently associated" in the tables provided for particular loci. But instead of ignoring the positive and negative correlations, Dr. Chakraborty called into question the way the lab gathered their data.

In particular, Dr. Chakraborty noted that Drs. Caragine and Mitchell did not take into account allele masking when estimating the drop-out rate. That fatal design flaw gave them an inaccurate estimate of the drop-out rate. Once again, they failed to face the fundamental problems in the design of their drop model.

6. The False Positive Study

a. Relying on False Positive Study, While Minimizing the Importance of Case-Specific False Positive Testing

The OCME initially testified at the hearing that the false positive study was an incredibly important aspect of the FST validation, because it showed how reliable it was. Dr. Mitchell herself also testified on direct how it was the most important indicator to her that FST was reliable. Mitchell, 1/2/13, 45.

However, Dr. Mitchell qualified the per case false positive study as a "useful" enhancement but "not essential," despite the fact OCME is currently developing such a model. Mitchell, 10/28/13, 792. Meanwhile, Drs. Gill and Haned stated clearly in *A new methodological framework* that the performance testing of the false positive study was a necessary part of determining the weight of the evidence. His presentation at the ISFG conference in the Fall of 2013, when he traveled to Australia to talk about how a high LR was not necessarily strong evidence, demonstrated that the false positive test addressed directly the strength of the evidence for a DNA sample.

b. Discarding Racial Identity Data

The Legal Aid Society sought a chart containing the racial identifications of the individual donors used in making the 439 mixtures used in validating the FST, referred to both in *"Validation of a DNA Mixture Statistics Tool"*, supra, and in Dr. Adele Dr. Mitchell's Frye hearing testimony. The Legal Aid Society sought this information because its experts had voiced a concern that the way that OCME had developed its genetic mixtures for testing the FST might lead to racial bias in the way it is applied in criminal casework. Defense experts wanted to either prove or disprove that possibility of racial bias. However, Mimi Mairs, Special Counsel to the Department of Forensic Biology at OCME, stated that according to Dr. Mitchell, when Dr.

Mitchell and Dr. Caragine were preparing the article on FST, that they tallied up the ethnicities of the mixture donors on a "piece of scratch paper and that piece of scratch paper was not saved." Record, 6/17/13, 9.

That claim is in direct contradiction to Dr. Caragine's testimony before she resigned from OCME, that there was "a key, sort of secret, quote, unquote 'key'" that revealed "more in general, what their skin color, if someone is black or white, Caucasian." Dr. Caragine went on to testify that the key did not have the donor's ethnicity, "but we know what their skin color is." Caragine, 4/2/13, 31-32.

Drs. Shapiro, Rohlfs, Rosenberg, and Chakraborty have testified about the importance of knowing and taking into account the race, or population substructure, of the contributors to mixtures in a false positive study involving probabilities of allelic drop-out.

Dr. Mitchell tried to downplay the importance of taking racial identity into account when conducting the FST false positive study. Yet she and Dr. Caragine included the racial backgrounds specifically in the FST validation paper exactly because it was important. Because Dr. Caragine and Dr. Mitchell had thrown away the data on the races of their lab personnel contributors, and because the OCME refused to recompile this data, defense experts were denied access to the critical variable that the Research and Validation team had chosen to leave as a mystery.

c. Simulating D2 and D19 for non-contributors in the false positive study

During the hearing, Dr. Mitchell revealed that she had simulated alleles at loci D2 and D19 for non-contributor profiles from the OCME database, except for profiles in the Asian sub-population. These hybrid profiles were then used in the FST false positive study, According to

Dr. Chakraborty, that information was not revealed to the DNA Subcommittee before its deliberations on FST.

As a population geneticist, Dr. Chakraborty was concerned about the use of combinations of "partly simulated and partly raw data." The procedure for simulating a 15 locus genotype is very different from generating a single locus genotype. Had the Subcommittee known of this fact, "we would have raised that this data in no longer worth looking at." For Dr. Chakraborty, the use of the simulated D2 and D19 raised concerns about the validity of the false positive study. He indicated that simulating two loci to go with an actual person's 13-locus profile was not generally accepted in the scientific community. Chakraborty, 12/16/2013, 1121-1124.

B. The OCME'S FST Methodology Is Not Generally Accepted As Reliable By The Relevant Scientific Community

The OCME's unsound scientific methodology applied in FST has been rejected by the relevant scientific community. FST's methodology stands alone in how it determines dropout and drop-in rates; key characteristics of real crime scene samples were ignored when determining how to produce a likelihood ratio; FST's primary assumptions were not fully tested; and, alarmingly, the OCME failed to investigate results that should have been a red flag and caused them to question how FST operates.

The People had the burden of proving that the FST has gained general acceptance by the relevant scientific community. The People failed to produce witnesses to support this position besides the creator of the FST, Dr. Adele Mitchell, and Dr. Hinda Haned, a witness whose own publications and likelihood ratio software program contradict the FST's methodology.[3] In stark contrast, the defense produced witnesses with expertise in forensic DNA analysis, population

[3] The People also called Dr. Mitchell Holland, but given that he did not review the OCME validation on FST, and was not aware of how drop-out rates were determined, even though he claimed he did, his testimony is not applicable to the general acceptance of the FST. See Holland, 52 and 76-78.

genetics, and statistics, who testified to the numerous flaws and shortcomings of the FST's methodology that prevented the FST from becoming generally accepted by the relevant scientific community. Moreover, almost all the defense expert's criticisms regarding the FST's unreliable methodology were not challenged by the People during cross-examination.

Accordingly, this Court must find that the FST is not generally accepted by the relevant scientific community.

1. Defining the FST methodology

New York Executive Law § 995 defines "DNA testing methodology" as "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." N.Y. Executive Law § 995 (McKinney 2014). Accordingly, when a new method is brought on line to provide a statistic for DNA evidence a court should/must consider all those factors when making a decision about the general acceptance of a scientific principal or procedure.

In this Frye hearing the methodology being assessed is the drop model used by the FST. This Court must determine whether a drop model that uses pre-determined drop-out values based on the total amount of DNA in a sample is generally accepted by the relevant scientific community. Other assumptions made in FST, including not accounting for relatedness, not analyzing a sample based on its own characteristics, that underestimating the drop-out rate will usually lead to a lower likelihood ratio, and that underestimating the number of contributors is conservative are not supported by the scientific research and therefore, the FST must be rejected by this Court and found to be not generally accepted by the relevant scientific community.

2. The People rely on peer review publication and the DNA Subcommittee to prove that the FST is generally accepted, but this reliance is misplaced.

Peer review is critical to the scientific process. The fact of limited peer review however is insufficient to demonstrate that a methodology is generally accepted by the relevant scientific community. Peer review includes peer review publications, presenting at conferences, and presenting to oversight committees. Peer review is the beginning of scientists discussing a scientific principle, and does not indicate general acceptance by the relevant scientific community. In this Frye hearing, the People rely primarily on peer reviewed publications and the DNA Subcommittee to demonstrate general acceptance but these two factors are insufficient to prove general acceptance.

The People rely on two published papers on the FST to demonstrate that the FST is generally accepted. However, publication alone is not sufficient to demonstrate that a novel methodology is generally accepted – rather, it is frequently the beginning of the peer review process. See Coyle, 11/8/13, 33 ("No, it's [peer review] actually the beginning of the process. Submission to a journal just means that you're presenting your data, now it's open for reading by all scientists in the community and then there's further discussion further experimentation based on the content of that paper."). Furthermore, he FST has already faced criticism in a peer review publication when Drs. Ge and Budowle criticized the OCME's method for how drop-in and drop-out values are determined. See Defense Exhibit K4.

Additionally other forensic science methods that were published and generally accepted by scientists within that field that have later been discredited, and determined to have played a role in wrongful convictions. For example, microscopic hair analysis was generally accepted for years: "These [microscopic hair] comparisons have been routinely conducted in forensic laboratories and accepted both in the scientific community and in the legal community for the

past 75 years. . . . the studies cited above do show that the method is reliable and repeatable." Oien, Caryn T., Forensic Hair Comparison: Background Information for Interpretation, Forensic Sci. Communications, The FBI Laboratory Services, April 2009, Vol. 11, Num. 2 at 11. However, it has been determined that 38% of the wrongful convictions that could in part be attributed unvalidated or improper forensic science resulted from microscopic hair analysis. Garrett, B. & Neufeld, P., Invalid Forensic Science Testimony and Wrongful Convictions, 95 VA. L. Rev. 1, 47 (2009). Indeed, recently the Department of Justice has moved to review thousands of cases where hair and fiber evidence was used in criminal cases. Hsu, Spencer, Justice Dept., FBI to review use of forensic evidence in thousands of cases, Washington Post, July 10, 2012 *available at* http://www.washingtonpost.com/local/crime/justice-dept-fbi-to-review-use-of-forensic-evidence-in-thousands-of-cases/2012/07/10/gJQAT6DlbW_story.html (last visited April 10, 2014).

Consequently, this Court should not place too much faith in the peer review publications that the People use to argue that the FST is generally accepted. This is especially true since the peer review publications have errors. Specifically, the OCME claims in its paper that they used 85 contributors when in reality they only used 61. Rohlfs, 11/19/13, 144-46. Thus, a reader would be mislead by the OCME's paper and believe that they used substantially more contributors than they actually did.

The argument that approval of FST by the DNA Subcommittee demonstrates that it has gained general acceptance is also misplaced. The DNA Subcommittee is similarly a type of peer review where DNA experts review a methodology to determine whether it can go on line. However, as with many of these types of committees, it is difficult to thoroughly review a methodology given time constraints and because pertinent details do not come to light. And, as

science progresses, more information becomes available which can demonstrate that something that once seemed reliable is no longer reliable. Indeed, Dr. Chakraborty, a member of the DNA Subcommittee testified that although he originally voted to approve the FST, his vote today would be different because of new developments in the field:

> Any recommendation in the context of DNA forensics and also in other areas, by regular laboratory is not – cannot be equated to general acceptance, and the reason being is once certain recommendation is in place in that jurisdiction, it is expected that more and more people will be using that technology or method. Consequently, in the literature we will have more information about the experience of that, the methodology, published in peer-reviewed journals or discussed publicly or discussed in available public. So the general community will have an exposure of the experiences of that methodology, consequently there will be more opinions to guide us to address the question of is it generally accepted or not in the context of uncertainties of low copy number DNA labels. That's what we're finding.

Chakraborty, 12/16/13, 1098-98 (generally discussing why the DNA Subcommittee approval does not indicate general acceptance by the relevant forensic scientific community).

Therefore, the OCME's publications and review by the DNA Subcommittee are insufficient to prove that the FST is generally accepted in the relevant scientific community.

3. The use of software programs to provide a statistic in DNA testing is novel, and the OCME's lack of transparency has prevented the relevant scientific community from meaningfully analyzing the FST; therefore, the FST is not generally accepted by the relevant scientific community.

A need exists in the forensic DNA community to determine how to provide a statistic for complex DNA mixtures that cannot be deconvoluted. However, the research in this area is still in its infancy. The concept of the likelihood ratio is even difficult for most forensic analysts to comprehend. Given the complexity of this emerging field, open source is imperative so that the relevant scientific community can adequately assess the reliability of statistical software programs. In fact, Dr. Gill stated "[i]n order to gain court acceptance I argue that transparency of

software is required and therefore the way forward is open-source." Colin Barras, Software says Knox's DNA not on crime scene bra clasp, NewScientist, Aug. 20, 2013.

"Source code is the human readable form of a programming language and contains the complete set of instructions for how a computer processes input data. In the absence of source code, the inner workings of a program cannot be examined, adapted, or modified." A. Morin, et. al., SHINING LIGHT INTO BLACK BOXES, SCIENCE, Vol. 336, 13 April 2012, p. 159-160. By providing the source code for the FST, the OCME would fall in line with most of the other statistical software programs in forensic DNA testing. The OCME refuses to do so. The OCME's lack of transparency is troubling for two reasons. First, the OCME is a public laboratory that provides evidence for in court criminal cases; and secondly, and more importantly (for the Frye hearing), this lack of transparency prevents the relevant scientific community from assessing the FST's reliability.

"We do not advocate a 'black box' approach" was the position taken by the People's witness, Dr. Hinda Haned. See Gill, P., et al., DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probalistic methods, Forensic Sci. Int.: Genetics 6(2012) 679-688, at 684; see also, Haned 3/8/13, 15 (testifying that she is an advocate of open source). Dr. Haned's position is the mainstream, with almost all forensic DNA analysts advocating for open source. The OCME is in the very small minority of entities that refuse to comply with open source. Considering the OCME's novel approach to determining drop-in and drop-out, open source is crucial so that the relevant scientific community can analyze the FST and determine whether the methods used are accepted. The OCME's obstruction of this review is contrary to how science works and is impeding a real review of the FST.

Furthermore, Dr. Haned herself did not review the entire validation, but rather only read the summaries and figures. Id. at 84-85. And, Dr. Holland only read the 2012 FST paper, "Validation of a DNA mixture, statistics tool incorporating allelic drop-out and drop-in.". Not only is the OCME shielding the scientific community from using and analyzing the FST by not making it open source as the other software programs are, but also, their own witnesses who testified in support of the FST's general acceptance did not do a thorough review of the FST and its validation.

Furthermore, not only is the FST not open source, it also is unavailable to the scientific community to test. The OCME could provide a link to allow other scientists access to run their evidence through the FST to at least see how running different parameters would impact the FST's results. However, the OCME will not even permit this.

Consequently, because no scientists, except internal OCME scientists, have been allowed to review the source code for the FST it is impossible for the FST have been generally accepted by the relevant scientific community. While one U.S. based program, TrueAllele, is not open source, at least others can purchase and test how TrueAllele works. For example, the National Institute of Standards and Technology was able to conduct their own studies on TrueAllele. National Institute of Standards and Technology presents DNA mixture validation study on Cybergenetics TrueAllele Casework at international forensics conference, *available at* http://www.cybgen.com/information/newsroom/2011/sep/National-Institute-of-Standards-and-Technology-presents-DNA-mixture-validation-study-on-Cybergenetics-TrueAllele-Casework-at-international-forensics-conference.shtml (last visited 4/10/14).

4. This Court should give limited weight to the People's witnesses.

As the Court reviews the expert testimony from the Frye hearing, it should consider the weight to give to the expert testimony. In this case both the People and the Defense presented expert testimony to support their position. The weight given to the People's witnesses should be significantly less than the weight given to the Defense experts.

It has long been recognized that the expert testimony by the creators of a novel methodology should be considered with caution. "If this Court were to adopt the view that the testimony of persons who have developed and whose reputation and livelihood depends on the use of a new technique alone supports admissibility, then the views of the developer and his disciples would be substituted for the scrutiny of the marketplace of general scientific opinion and the substance of the Frye test would be eliminated." People v. Young, 391 N.W.2d 270, 276 n.4 (Mich. 1986); see also State v. Thompkins, 891 So.2d 1151, 1152 (Fla. Dist. Ct. 2005) (stating "Frye requires more than the testimony of an expert who has a personal stake in the theory or is prone to an institutional bias.").

Given both the personal and financial stake inventors have in their own developments, this Court should be careful in how much weight it attributes to Drs. Caragine and Mitchell. While Dr. Mitchell is a well respected practitioner in her field she is the primary inventor of the FST, and this Court should keep that in mind when considering her testimony.

Dr. Caragine, while not the primary creator of the FST, significantly participated in its development. Dr. Caragine's reputation has faced significant criticism since the inception of this Frye hearing. Dr. Caragine resigned from the OCME after it was discovered that she did not follow protocol. See State of New York Office of the Inspector General: Investigation into the New York City Office of Chief Medical Examiner: Department of Forensic Biology, December

2013 (hereinafter "IG Report"); see also Jacobs, S., et al., Exclusive: Pioneer in DNA testing quits Medical Examiner's Office over lab violations, Daily News, May 16, 2013; see also Beekman, D., et al., Exclusive: 'Leadership must change' at city's DNA lab: report, Daily News, May 18, 2013.

Dr. Caragine was investigated by the State of New York Office of the Inspector General ("IG"). The IG found that at least once Dr. Caragine, by changing the number of contributors input into the FST program from two to three contributors, ostensibly changed the FST's results from exculpatory to inculpatory. Specifically, a criminalist III ran an analysis on gun swabs, which included comparing the suspect's DNA to the gun's swab on the FST to produce a likelihood ratio; the results "indicated that it was slightly more likely that the suspect's DNA was not in the mixture." Defense Exhibit at 30 (emphasis added). However, Dr. Caragine changed one of the parameters of the FST, the number of contributors, from a two person mixture to a three person mixture; the FST now "resulted in a different outcome." Id. at 31 (emphasis added). Furthermore, Dr. Caragine's professional behavior speaks directly to her credibility, which should cause this Court to give her testimony little, if any, weight. These issues go directly to the issue at hand – Dr. Caragine's behavior and trustworthiness as a scientist. Furthermore, the IG report makes it clear that Dr. Caragine's testimony in this Frye hearing was not honest when she testified that there are never disputes between DNA analysts. Caragine, 3/12/13, 191-92.

Besides Drs. Caragine and Mitchell, who have a professional interest in having the FST admitted in court, the only other witness who spoke to the FST's general acceptance was Dr. Haned. Dr. Haned reviewed "description[s] of the validation study along with the figures and graphs that were also published in the validation paper, FSI Genetics" the Saturday before her testimony. Haned, 3/8/13, 84-85. Given the People's claim as to how voluminous the FST's

validation study was, it is not possible that Dr. Haned was able to thoroughly review the material in order to speak to its reliability in such a short period of time.

Furthermore, Dr. Haned's likelihood ratio software program differs significantly from the FST program, as discussed supra, and her publications advocate for a different approach than what the FST takes. Besides opposing the black box approach, Dr. Haned also advocates for the defense experts to meet with the prosecution to have input into the defense hypotheses (see generally Haned, 3/8/13, 53-54, testifying that it is better to be more inclusive with the propositions being made and allow the defense access to the defense hypothesis). The inconsistencies between Dr. Haned's practice as a forensic DNA analyst and her testimony in the current Frye hearing, as well as the limited time she spent on reviewing the FST validation, should give this Court pause in attributing any significant weight to her testimony.

The People failed to provide witnesses that could support the position that the FST was generally accepted within the relevant scientific community; and while the burden is not on the defense, the defense successfully demonstrated that the methodology employed by the FST is not generally accepted by the relevant scientific community. The People could only muster one independent witness to testify about the FST, yet the defense presented five witnesses who identified significant problems with the FST.

5. The People failed to kill the scientific message.

The People's attempt to discredit the defense witnesses by arguing that the witnesses had a financial incentive to testify must fail. Dr. Budowle, is one of the most renowned forensic DNA scientists in the world. He had a career at the FBI that spanned over 25 years and was critical to the development of forensic DNA typing technologies and assisted in developing the original DNA markers for CODIS, the national DNA databank that is used throughout the

country. Dr. Budowle's credentials cannot be impeached – and to this day, he continues to primarily testify for the prosecution; moreover, the fee he received for testifying was given to the University that employs him.

Drs. Rosenberg and Rohlfs were paid for the time they spent on the Frye hearing. Neither had ever testified in a criminal case before, and both testified in line with the same scientific principles they use in their own research and publications.

Besides cross-examining both Drs. Rosenberg and Rohlfs about their fees in their case, Dr. Rohlfs was cross-examined about her position that racism has no place in forensic DNA testing and genetics:

> Q: In fact, you wrote: "It disproportionately effects African American families due to their greater representation in the DNA database"; correct?
>
> A: That sounds right.
>
> Q: In this particular article, there's about a page-and-a-half dedicated to that concern; correct?
>
> A: It sounds right.

Rohlfs, 11/20/13, 222-23. Dr. Rohlf's continued to be questioned about her policy concerns regarding race and familial searching:

> Q: Your concern with the overrepresentation of African Americans in the databases, you share that concern [with Professor Murphy]; correct?
>
> A: Yes.
>
> Q: Do you feel that's a systemic problem in all statistical DNA analysis?
> A: Well, the overrepresentation of some groups like African Americans in databases only comes up when a database is used.

Id. at 224-25. The questioning of Dr. Rohlfs's policy concerns continued:

> Q: You wrote: "My name is Rori. Right now I live in Seattle, and I'm moving down to the Bay in midish January for a science job studying genetics in evolution and also tearing up racist assumptions in statistics used for forensic identification."

A: That sounds like something I wrote.

Id. at 228. A significant amount of Dr. Rohlfs's cross-examination focused upon her concern that with familial searching there are issues with overrepresentation of certain groups of individuals which can lead to a bias. This is a legitimate policy concern that is shared among many individuals and does not speak to any bias towards the defense. The People's attack on Dr. Rohlfs focused on these types of issues and failed to confront her with evidence that her criticism of the FST was incorrect, except for an attempt to discredit her by incorrectly characterizing her findings.[4]

With respect to Dr. Rosenberg the People were unable to contradict any of his findings as well. Dr. Rosenberg found that there were problems with the false positive test of FST's validation that indicated that the false positive rate was unreliable. Rather then confront Dr. Rosenberg with evidence contrary to his findings, the People instead pose hypotheticals that the data relied upon by Dr. Rosenberg may have been incorrect. Dr. Rosenberg did not have access to the actual races of the contributors because the OCME threw out the key to the races of their contributors. Thus, the People should not benefit from the OCME's misconduct, particularly since Dr. Rosenberg testified that the method used to reverse engineer was the best method to determine race in this situation. While he testified that he could not say there are zero errors, it is likely that most characterizations are correct. Rosenberg, 12/5/13, 82-3.

[4] Dr. Rohlfs was questioned about her conclusion that she was concerned by the number of individuals used in the false positive test. Dr. Rohlfs's notebook indicated that another statistician stated that neither of the conclusions (three Asians is enough versus three Asians is not enough) were supported. While the People tried to argue that this meant that three Asians may be enough, they missed the point that actually Dr. Slotkin was making a neutral statement, and it was Dr. Rohlfs's scientific opinion that there was reason to investigate this issue further. See Rohlfs, 11/20/13, 253-56. Further, the People tried to show a bias in what evidence Dr. Rohlfs testified about, however, there were findings that were in favor of the defense that she did not testify to because while she found a correlation in the data, she could not identify the cause. Id. at 248-49. Similar confusion occurred when the People questioned Dr. Rohlfs about Defense Exhibit PPP, where her conclusions were regarding the defense hypothesis and the need for further investigation. Id. at 262-67.

The People also tried to confuse the issue of whether likelihood ratios is generally accepted and whether how it is implemented in FST is generally accepted. Dr. Rosenberg made clear that while the likelihood ratio is a sound mathematical and statistical principle, it is the implementation that needs to be considered and whether it is correctly being used in this case. Id. at 755-56. Similar to other defense witnesses, the People failed to attack Dr. Rosenberg's findings and critiques of the serious problems that exist with the FST.

The cross examination of Dr. Shapiro consisted of days of questioning regarding his credentials and how much he was paid by the defense . The People failed to contradict many points he made, and instead made ridiculous suggestions that Dr. Shapiro was not qualified for the position he maintained for over ten years at the OCME. In reality, Dr. Shapiro's credentials are unimpeachable; after receiving his Bachelor of Science in chemical engineering and applied biology from Columbia University's Engineering School Dr. Shapiro went on to obtain a Master of Philosophy and PhD from Yale University in biology. Dr. Shapiro held a postdoctoral position at the Center for Neurobiology and Behavior at Columbia University and then continued as a research scientist and a research assistant professor there after his postdoctoral program ended. Dr. Shapiro was also a senior research associate and research associate at the Howard Hughes Medical Institute at Columbia University where his research focused on Alzheimer's disease. Additionally, Dr. Shapiro taught at numerous institutions throughout the city. See Shapiro Defense Exhibit V. Given Dr. Shapiro's educational background and his research this was truly a preposterous and unfounded accusation, and a waste of this Court's time by circling back to these issues day in and day out. See generally Shapiro Test. After all, it would be ironic for the prosecution to uphold the OCME as a leader in forensic science yet discredit the head of their mitochondrial and training units.

Dr. Shapiro was responsible for identifying the serious misconduct of the OCME's criminalist, Serrita Mitchell, that sparked New York City Council legislation and an investigation by the New York State Inspector General. Dr. Shapiro, who was not involved in Serrita Mitchell's original training, failed her during her oral exams and then was involved as a supervisor in the training team that first discovered many of the issues with her. Shapiro Test. Oct. 4, 2013 at 26-27. See also Goldstein, Joseph, & Bernstein, Nina, Ex-Technician Denies Faulty Lab Work, New York Times, January 11, 2013, *available at* http://www.nytimes.com/2013/01/12/nyregion/former-lab-technician-denies-faulty-dna-work-in-rape-cases.html (last visited 4/8/14); Goldstein, Joseph, Report Details the Extent of a Crime Lab Technician's Errors in Handling Evidence, New York Times, December 5, 2013 *available at* http://www.nytimes.com/2013/12/05/nyregion/report-details-the-extent-of-a-crime-lab-technicians-errors-in-handling-evidence.html (last visited 4/8/14); Peters, Justin, The Unsettling, Underregulated World of Crime Labs, Slate.com, January 14, 2013 *available at* http://www.slate.com/blogs/crime/2013/01/14/serrita_mitchell_dna_the_unsettling_underregulated_world_of_crime_labs.html (last visited 4/8/14). The Serrita Mitchell scandal was one of many OCME scandals that followed, which included a primary witness in this Frye hearing, Dr. Caragine. And while the People accused Dr. Shapiro of not being qualified for his position because he had no prior forensic DNA testing experience, it is curious that the People had no issue with the fact that Dr. Adele Mitchell, the primary creator of the FST also had no experience with forensic DNA testing when she arrived at the OCME in 2008 to design the FST. Dr. Mitchell did not even undergo any forensic DNA training before beginning work on the FST.

The primary focus of the People's cross examination of Dr. Shapiro was the baseless attempt to demonstrate that Dr. Shapiro was not qualified to hold his position at the OCME.

Primarily the cross examination was unwarranted attacks on Dr. Shapiro's credentials, including attacks on his wife, insinuating that his wife got him his job, and how much he was getting paid.

Dr. Chakraborty is a renowned forensic population geneticist. He has been involved with forensic DNA since its inception . His work was critical to the 1996 NRC report . He served on the New York State DNA Subcommittee for over 15 years. While Dr. Chakraborty originally voted to approve the FST, he testified that his vote today would be different, primarily because of the developments in the field. Dr. Chakraborty, like Dr. Budowle, did not keep the fee he earned for working on this Frye hearing.

The People focused on asking experts how much money they earned as a fee, rather than on contradicting the substance of their critiques about the FST. They could not challenge the substance of the scientific criticisms leveled by the experts called by the defense, so they made baseless attacks on their credibility.

6. The methodology employed by the FST is different from the other software programs in significant ways clearly demonstrating that the FST is not generally accepted by the relevant scientific community.

Forensic DNA scientists continue to search for a model to allow them to quantify mixtures that cannot be deconvoluted, and while progress has been made, still no clear answer exists. The testimony and publications make clear that the other programs are developing in the opposite direction from FST. Therefore, the FST is not generally accepted by the relevant scientific community.

a. Estimation of drop-out rates based on total DNA quantitation.

The People have argued that the FST applies the LR principle and therefore, is generally accepted. However, Courts have recognized that just because an accepted principle may be applied that does not mean that the application of the principle is correct. Furthermore, the

linchpin of the FST methodology for how the LR is determined is that drop-out rates are based on total quantitation; this method is outside the mainstream of how the forensic science community is determining drop-out values.

i. The OCME misapplied the LR principle and this explains why the FST methodology is not generally accepted by the relevant scientific community.

The LR is a sound mathematical principle. Yet, the determination of whether an LR value is correct cannot stop at this superficial analysis. The real issue is how the LR principle is applied. See Chakraborty, 12/16/13, 1091-92; see also Rosenberg, 12/6/13, 755-56. In Paoli R.R. Yard PCB Litig., 35 F.3d 717 (3d Cir. 1994) (Paoli II), the Court made clear that misapplying a methodology is critical to determining whether expert testimony is admissible. The Court held that "any step that renders the analysis unreliable under the Daubert factors renders the expert's testimony inadmissible. This is true whether the step completely changes a reliable methodology or merely misapplies that methodology." Id. at 745. While this is a Daubert decision, the principle being espoused applies to this Frye hearing. The People argue that the Court's analysis should stop once it reaches the decision that the LR is a generally accepted mathematical model, but the Paoli II court makes clear that a methodology can be misapplied and that will effect the reliability of the expert testimony. In this case, it is the information that is being used in the LR model that this Court should be concerned with; the method the OCME used to determine drop-out rates is unreliable and not generally accepted by the scientific community. Therefore, the OCME cannot shield itself from criticism about its methodology for determining drop-out rate by pointing to the fact that it uses the LR principle to produce a final result. Since the FST's final LR value depends on the drop-out values being input into the calculation, the error associated with the drop-out rate estimations directly impact the reliability and general acceptance of the FST.

ii. No other statistical program bases drop-out rates based on total quantitation values because this is an unreliable method.

The linchpin of the FST is that the drop-out rates can be estimated from the total quantitation of the DNA sample. The FST stands alone in employing this methodology to determine drop-out rates. While the FST uses a quantitation based system, other programs use an approach that takes into account peak heights and/or use a range of drop-out rates. While every approach requires more research, the FST does not even use the same approach as the others, and as such, is not generally accepted.

As discussed supra, the method employed by the OCME to determine drop-out rates is scientifically unsound. First, Dr. Mitchell claims that underestimating the drop-out rate is conservative and that this is based on a phenomenon that she witnessed in her research, but that she never actually performed any testing on this hypothesis. In In re TMI Litig., 193 F.3d 613 (3d Cir. 1999), the Court held that the expert testimony was inadmissible because a crucial part pf the analysis was a study that was never performed. Id. at 694-95. Similarly, here Dr. Mitchell has no evidence to back up the fundamental assumption of the FST. Similar to the reasoning in Paoli II, this Daubert decision demonstrates the need to for each step of a process to be reliable. Here, Dr. Mitchell lacks the evidence to demonstrate the most important assumption made by the FST is incorrect. That other forensic DNA laboratories are not employing the same method as the OCME for estimating drop-out rates is strong evidence that this method is both unreliable and not generally accepted.

As pointed out in People v. Seda, 139 Misc.2d 834, 847 (N.Y. County 1988), "[w]hile general use in crime laboratories does not necessarily connote general acceptance in the scientific community, evidence of such limited use does persuade the court that the procedure

has not been generally accepted by even the technical personnel whose standards may be less exacting than those scientists." Id. at 847 (citing Giannelli, P., The Admissibility of Novel Scientific Evidence: Frye v. United States, a Half-Century Later, 80 Colum. L. Rev. 1197, 1214-1215 (1980) (quotation omitted)). Seda determined that the results of electrophoresis analysis performed by the Office of Chief Medical Examiner could not be admitted in court because the procedure used was unreliable, and admission of this evidence would not be harmless error. Similarly, in this case, where the OCME stands alone in how it determines drop-out rates for the FST, this Court should recognize that because all other DNA laboratories are implementing a different method this is strong evidence that the FST is not generally accepted by the relevant scientific community.

The other likelihood ratio software programs include Forensim, likeLTD, True Allele, and LabRetriever. Forensim requires the user to input drop-out values (along with other parameters), and the program will then "generate a plot of [likelihood ratio] values as a function of dropout rate[] in addition to a 95% confidence interval for [drop-out]." Steele, C.D. & Balding, D.J., Statistical Evaluation of Forensic DNA Profile Evidence, Annu. Rev. Stat. Appl. 2014, 1:20.1 – 20.24, at 20.18. The FST does not generate a range of likelihood ratios, nor does it provide a confidence interval. likeLTD applies drop-out rates in order to maximize the likelihood ratio in the numerator and in the denominator – the result of this is that the drop-out rates in the prosecution and defense hypothesis may be different. People's Exhibit G, Likelihood Ratios in Forensic DNA Mixture Interpretation at 2. LabRetriever uses the average peak heights of all detected loci and inputs these values into a logistic regression equation fit to the data in order to determine the estimate of drop-out. People v. Rodriguez Ind. 0547/2009), Lohmueller 6/28/13, 160. TrueAllele uses electropherogram peak heights at every allelic position without

taking drop-out into account. Steele & Balding, supra, at 20.19. The People have failed to identify ongoing research can be identified where another forensic DNA laboratory is following in the OCME's direction.

A common theme among Forensim, Like LTD, True Allele, and LabRetriever is that these programs permit a case specific approach to determining drop-out, which is critical with real world crime scene samples. As discussed infra, crime scene samples are frequently plagued by low template DNA, degraded DNA, and multiple contributors; therefore, it is imperative to examine the case sample before determining the drop-out values. However, the FST approach is to use empirically determined drop-out values – which, in actuality are not the values used since the FST automatically uses a drop-out value that is one standard deviation lower. As discussed *supra*, this is problematic since it does not take into account the characteristics of the sample actually being analyzed; that the total quantitation originally determined in DNA testing has an average error rate of 30%; and, that combining this with the input of LCN data, the problems of each technology will be compounded because the linear drop-out model is too simplistic to address the problems of the stochastic region (see Budowle, 12/9/13, 109-11, testifying that research presented by Tim Kalafut of the U.S. Army demonstrate that drop-out rates should be calculated using a logistical regression analysis).

Dr. Budowle raised another concern about basing the drop-out values on total quant and not considering peak heights. Peak heights provide a DNA analyst with a better understanding of the quality of the DNA and what you actually have in a crime scene sample may be very different from what you originally observed when determining the total quant. Dr. Budowle testified "[s]o the FST has now removed that process [analyzing peak heights] and is making the assumption that the amount of DNA that they start with is actually a reproducible indicator of the

end result and it [the FST] can't tell you what the drop-out rate is on a case by case basis or on a sample by sample basis." Budowle, 12/9/13, 26-27. Dr. Budowle further explained that when amplifying the DNA sample, even if you start out with what you believe to be 18 picograms, when the sample is divided into triplicates during LCN testing, each amplification would have different amounts of DNA – one tube can have twelve picograms, another can have five and the third can have one. Dr. Budowle concluded: "If the quantities of DNA are different in each tube, the assumption that OCME starts with a quantity of that DNA is an indicator of the drop-out rate. The drop-out rate in each tube is different. That's not been compensated for." Id. at 50-51. Dr. Budowle further criticized the OCME's drop-out rate methodology because it does not take into account allele sharing and relatives. Dr. Budowle ultimately concluded that the OCME's methodology for determining drop-out rate is wrong given these types of factors. Id. at 94.

All of the issues that Dr. Budowle identified provide strong support that the drop-out rates the FST uses are not correct, and this helps to explain why other forensic DNA laboratories are not employing the OCME's methodology.

b. Degraded models

The OCME failed to implement a successful degraded model in its FST validation. However, Tvedebrink et al., "demonstrated how to implement the information of degradation into estimating allelic drop-out probabilities in the situation of degraded samples. This adjustment is *crucial* in order to assess the probability of drop-out of STR alleles in degraded samples since peak height imbalances caused by degradation preclude a constant level of peak intensities across the entire range of bp-values in STR profiles."[5] Tvedebrink, T., et al., Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out,

[5] This is an assumption made by Tvedebrink et al.

Forensic Sci. Int.: Genetics 6 (2012) 97-101, 100-01 (emphasis added). Indeed, likeLTD and STRmix both incorporate degradation into their likelihood ratio calculation. Steele & Balding, supra, at 20.18-20.19.

A degradation model is important because real world crime stains are frequently degraded. Yet, not only does the FST not have a degradation model, but the degradation experiments in the FST validation produced poor results, as discussed supra.

c. Range of drop-out

The majority of the research for software programs is focused on determining the drop-out and drop-in rates. While the FST uses pre-determined drop-out ranges based on quantitation of the DNA sample, many other programs use a range of drop-out values. Forensim, likeLTD, and LabRetreiver all use a range of drop-out rates. Indeed, Balding and Buckelton determined: "One important conclusion, here and in what follows, is that if drop-out is invoked to sustain the prosecution case then estimation of drop-out probabilities cannot be avoided. DNA based prosecutions that rely on drop-out and do not explicitly estimate plausible ranges for the drop-out rate parameter, are in our view, defective." Balding, D.J. & Buckelton, J., Interpreting low template DNA profiles, Forensic Science International: Genetics 4 (2009) 1-10, 5 (emphasis added). See also Haned, 3/8/13, 65-78 (testifying that splitdrop provides all plausible ranges for likelihood ranges based on varying parameters, including drop-out rate, and testifying that she prefers varying the drop-out rate between the prosecution hypothesis to make sure they are fair to each hypothesis).

In "Exploratory data analysis for the interpretation of low template DNA mixtures" authored by Haned, Slooten, and Gill, the FST's methodology for determining drop-out was criticized. While Gill et al. suggests maximizing the likelihood ratio by either exact calculations

or approximations using Monte Carlo simulations, another approach is to use drop-out rates based on "large sets of DNA mixtures obtained in different conditions. These methods derive estimates for the drop-out probabilities, which do not rely on the questioned sample, but rather on a population of samples from which the crime-sample could have originated from." Haned, H., Slooten, K., & Gill, P., Exploratory data analysis for the interpretation of low template DNA mixtures, Forensic Sci. Int.: Genetics 6 (2012) 762-774 at 768. This different approach that Gill et al., describes is similar to the approach used by the FST. Gill et al., goes on to state that they suggest a different approach:

> [W]e suggest a different approach that relies on the crime sample itself, rather than simulated or experimental samples: the crime-sample is re-simulated n times, at each iteration, a random sampling of the alleles is applied, in order to select the alleles that will drop-out from the sample. Since the true probabilities of drop-out are unknown, different drop-out probabilities, ranging from zero to one, are applied. The rationale behind this procedure is to explore the range of probabilities of drop-out that could have led to the crime-sample of interest. The simulations ultimately yield an empirical distribution for the probabilities of drop-out, and the most plausible values for these probabilities are the ones that lead to the same number of alleles that are observed in the crime-sample under investigation. The advantages of such approach is that the ranges of the drop-out probability can be evaluated separately under H_p and H_d, and that we avoid reporting values of drop-out that are supported by one hypothesis but not by its alternatives.

Id. Drs. Gill, Haned, and Slooten specifically point out that they would *not* use the approach used by the FST. The rejection of the OCME's method to devise drop-out rates, and instead advocating for a different method - use of a range of drop-out rates based on the crime scene sample itself - indicates that the relevant scientific community does not accept the FST's methodology for determining drop-out rates.

d. Access to the defense hypothesis

A likelihood ratio provides information regarding the odds of the prosecution hypothesis being true versus the defense hypothesis being true. The defense should be permitted to provide input into the defense hypothesis. As, the People's witness, Dr. Haned, testified to – LRmix, a part of Forensim, permits the defense to suggest propositions so that a defendant has a say in the type of defense he or she is presenting. Haned, 3/8/13, 53. In contrast, the FST does not allow for input from the defense. As testified to by Dr. Mitchell, there have been no conversations about permitting defense experts to come into the OCME to work with them on a defense hypothesis either. Mitchell, 10/30/13, 800-05.

One criticism that Dr. Rohlfs had was that because the FST does not employ a correction for inter-individual correlation, which is a genetic correlation between unrelated people due to distant co-ancestry, that the FST biases the defense hypothesis. Rohlfs, 11/19/13, 157. This criticism directly related to the defense hypothesis because she was able to show in Defense Exhibit PPP that without using the inter-individual correlation correction, in the worse case scenarios the defense hypothesis is being underestimated by 15 or 17 percent. Id. at 166. While the prosecution attempted to undermine Dr. Rohlfs's conclusion by indicating that this did not take the prosecution hypothesis into account this misses the point. Id. at 262-67. The defense wants a reasonable hypothesis that will maximize its likelihood ratio, and because the FST does not use a correction for inter-individual correlation the OCME is preventing the defense from maximizing its hypothesis. (See generally People's Exhibit G explaining that likeLTD selects drop-in and drop-out rates to maximize the likelihood in the numerator and in the denominator, which permits different values for each of these variables. Id. at 2.)

Although the People argue that the FST is conservative and benefits the defendant, this is just one of many examples where the FST is actually anti-conservative. Rohlfs, 11/19/13, 167. Indeed, coupling this with another principle in FST, namely that lowering the drop-out rate is conservative, Dr. Rohlfs determined that this is where the worst case scenario is observed for the defense hypothesis; her findings indicated that at minimum drop-out rates for the situation where there are two contributors and only one allele is observed at a locus, the FST underestimates the defense hypothesis by 15 to 17 percent. Id. at 166. This bias against the defense hypothesis continued into some of the middle plot drop-out scenarios, and only ended in the plots where the drop-out rates were highest.

Dr. Rohlfs performed this analysis by looking at one allele at a locus, however, if the minimum drop-out rates were used at multiple loci, then the underestimate of the defense hypothesis could build up across loci. The multiplicative effect would lead to an even more severe underestimation of the overall probability under the defense hypothesis. Id. at 172.

Dr. Rohlfs's testimony indicates why it is critical for the defense to have access to the defense hypothesis, and why many other software programs permit this type of access.

7. The population geneticists that testified in this Frye hearing have identified significant problems with the FST's False Positive Test and have concluded that the design of this test is not generally accepted by the relevant scientific community.

Drs. Rohlfs and Rosenberg testified about the significant concerns they had with the FST. Both experts' opinions were that there were numerous instances where the OCME failed to investigate areas of the validation that led to unexpected results and that the OCME did not adequately consider race in their false positive test.

Drs. Rohlfs and Rosenberg focused their analyses on the false positive study because as population geneticists and statisticians their primary concern was to determine whether the false

positive study could accurately inform the relevant scientific community on how the FST would work in the real world. As part of their analyses they needed to know the race of the individuals that were used in the false positive study. Unfortunately, the OCME did not preserve the races of the contributors. Supra at "Discarding Racial Identity". This is scientifically indefensible given how important of a role race plays in forensic DNA testing. Dr. Chakraborty testified that the racial background of contributors to mixtures is important to consider in the FST validation, and that "[i]f race and ethnicity is not accounted, the result will not be generally accepted" and the results are not reliable. See Chakraborty, 12/17/13, 1119-21.

As Drs. Rohlfs and Rosenberg did not have access to the race of the individuals used in the false positive study and the OCME refuses to turn over the FST source code, the races of those individuals used in the false positive test had to be reverse engineered. Drs. Rohlfs and Rosenberg both testified that absent the OCME providing the races to them the next option was to determine the races by taking the lowest likelihood ratio of each contributor. See Rohlfs, 11/19/13, 136 (Responding to the Court's inquiry about how to figure out the race: "[w]hen you mis-specify your allele frequencies, the likelihood ratio goes up. So the lowest likelihood ratio might be for the population group that best describes, or least mis-specifies the allele frequencies." In other words, Dr. Rohlfs is stating that a likelihood ratio is lower when you correctly identify the population group a person is from.); See Rosenberg, 12/5/13, 82-3 (explaining why the lowest likelihood ratio would be the way to determine the population the contributor is from: "But the reason for use of the lowest likelihood ratio relates to the scenario in which it's most challenging for a sample to have a higher likelihood to be in the mixture compared to not being in the mixture. When the alleles frequencies are used for the population

to which the individual belongs, that's the most challenging case for detecting presence of the individual in the sample and it produces the lowest likelihood ratio.").

a. The flawed design of the false positive study, namely that it failed to adequately take race into account, is not generally accepted by the relevant scientific community.

The importance of population genetics and specifically allele frequencies in forensic DNA testing has long been recognized. See generally NRC II, 89-124 (chapter titled "Population Genetics"). Therefore, when designing the false positive test the OCME should have taken into consideration genetic diversity. See Rohlfs, 11/19/13, 128 ("So to fully understand the false positive rate of a particular method, you need to try it out on individuals with a wide range of genomes."). However, an analysis of the false positive test has led to the conclusion that there is not sufficient genetic diversity among the individuals used and therefore, it is not generally accepted by the relevant scientific community.

In Defense Exhibit KKK, Dr. Rohlfs created a histogram demonstrating the number of times each individual was used in the validation. Id. at 129. Dr. Rohlfs determined that on average each individual from the Caucasian subpopulation was used 15 times, while individuals from the Asian, African American, and Hispanic subpopulations are used on average 23 to 27 times each. The OCME, on average, was using the same individuals from these latter groups more frequently than individuals from the Caucasian. Therefore, "since few people [in the Asian, African American, and Hispanic groups] were used many times, that fails to sample the genetic diversity of that population group." Id. at 135.

Along the same lines as Defense Exhibit KKK, Defense Exhibit LLL was used to determine the number of unique individuals who were used in the false positive study, the number of unique mixtures that were used in the false positive study, and the number of mixtures of individuals with the same population group who were used in the false positive study. Id. at

137-38. Dr. Rohlfs determined that there were 61 unique individuals used in the 480 mixtures of the false positive study; 159 unique combinations of individuals within those 480 mixtures; and, 11 out of 159 unique combinations were of individuals from the same population group. Id. at 144-46.

Dr. Rohlfs testified that only using 61 individuals seemed "fairly low." Id.at 145-46. Specifically, she determined that because typically hundreds of individuals are used to determine the allele frequencies in forensic DNA testing and that because only 30 independent trials or mixtures could be run with these 61 individuals, this was a low number of unique contributors. Id. at 146. Dr. Rosenberg supported Dr. Rohlfs's conclusions regarding the number of individuals used in the OCME's false positive study. When questioned about Dr. Rohlfs's conclusion that only using three Asian individuals was not acceptable, Dr. Rosenberg stated: "Well, these are highly – high variable microsatellites loci that have a lot of alleles and in order to capture the diversity of a population, one needs a sample that is large enough compared to the diversity of the markers and certainly three individuals would be too small to look at variability among individuals within such a large population group." Rosenberg, 12/5/13, 85 (emphasis added). Indeed, Dr. Rosenberg's own simulation analysis of the FST demonstrated that in two person mixtures, the OCME only used one mixture that was composed of only one pairing of two African Americans, one pairing of two Hispanics, and one pairing of two Asians. However, Dr. Rosenberg testified that this is insufficient and that "there's strong reason to believe that the populations of origin of the contributors would have a substantial effect on the false positive rate, and given the population of origin is likely to influence the false positive rate. It would be important to assess the false positive rate separately by populations of origin of the contributors. Here in this table the number of representatives used for any particular pairing is quite small and

would not enable a very accurate assessment of false positive rates by population of origin." Id. at 80.

Dr. Rohlfs was also concerned with the OCME only using 11 unique mixtures with individuals from the same population group. In Dr. Rohlfs's paper, Defense Exhibit NNN, titled "Familial identification: population structure and relationship distinguishability" she and her co-authors demonstrate that the false positive rate will vary depending on the population background of the individuals used. Rohlfs, 11/19/13, 146. In considering the FST, Dr. Rohlfs stated that with the FST "often you're dealing with mixtures of individuals. And when you have mixtures of individuals from – who are from the same population background, they're more likely to share alleles with a third individual of the same population background than if they're all three from different." Id. at 149. Dr. Rohlfs continued, the "highest false positive rate would be for mixtures of individuals from the same population background with other individuals who are being queried also from the same population background." Id. Yet the OCME only tested 11 of such combinations.

Dr. Rosenberg similarly supported Dr. Rohlfs's conclusion regarding the OCME's use of only 11 unique combinations of individuals from the same population group. Dr. Rosenberg testified: "I saw that as well, that most of the mixtures involved differ from the different population groups rather than pairs from the same population group. And I think that's concerning because pairs from the same population groups would be a more challenging situation to exclude false positives." Rosenberg, 12/5/13, 86.

The findings of the population geneticists and statisticians demonstrate that race, which is an important factor to consider in forensic DNA testing, was not adequately considered in the

development of the FST. Therefore, the false positive study, one of the key components of the FST's validation, is not generally accepted by the relevant scientific community.

b. The OCME failed to properly investigate numerous phenomena that occurred in the FST's false positive study which is not generally accepted by the relevant scientific community.

Dr. Rosenberg carried out a number of simulations to analyze the FST's false positive study. Through this analysis Dr. Rosenberg determined that there are systemic variables affecting the false positive study because the number of false positives per bulk run does not match what you would expect if the false positives were randomly distributed. Rosenberg, 12/5/13, 67-68. Similar to Dr. Rohlfs, Dr. Rosenberg determined that the OCME failed to investigate the role of population origin in the FST false positive study. In Dr. Rosenberg's opinion, the OCME should have analyzed the variables involved in the FST calculation to determine how each one affects the final result. He testified that the OCME should have held all the variables constant and only changed one parameter at a time, to fully understand the importance of each variable. Id. at 89. Dr. Rosenberg concluded that his analysis showed that "we don't know enough about the false positive rate that would be relevant to a real world scenario." Id. at 90.

The population geneticists and statisticians determined that the false positive study had a significant number of shortcomings. Accordingly, this type of study would be insufficient to gain general acceptance by the relevant scientific community.

G. The Only Court that Has Evaluated the FST Under the Frye Standard is People v. Rodriguez, and That Opinion is Neither Binding Nor Persuasive.

The only court in New York City to hear arguments on the general acceptance of the FST through a Frye hearing, besides this Court, was People v. Rodriguez (Ind.5471/2009). While the

Rodriguez opinion is not binding on this Court, this Court should decline to follow its logic because it misinterprets the scientific issues.

The Rodriguez court held,

> [T]hat FST rests firmly upon two pillars, Polymerase Chain [sic] Reaction-Short Tandem Repeat (PCR-STR) DNA analysis and the likelihood ratio or LR. Both have long been generally accepted by forensic scientists as reliable.

People v. Rodriguez, Ind.5471/2009, at 8. Relying on these two principles is incorrect. The point of the FST is to put a statistic on mixtures that cannot be deconvoluted after being subjected to quantitation, amplification, and electrophoresis, and analysis by a criminalist. A decision stating that the FST is generally accepted because it uses PCR-STR DNA analysis is equivalent to stating that baking a cake at 100 degrees over the recommended temperature is generally accepted because leading up to putting the cake in the oven all the right ingredients were used. Indeed, in Brim v. Florida, 695 So.2d 268 (1997), a Frye jurisdiction, the court stated:

> We reiterate that we should not treat population frequency statistics as an extension of the first step in the DNA testing process. Those statistics are a distinct step in the DNA testing process. The district court reasoned incorrectly when it found that both statistical reporting methods in this case were admissible because the chemical and biological techniques used in the first step of the DNA testing process satisfied the Frye test. It is improper to label calculations created with principles of statistics and population genetics as simply deductions from a methodology based on chemistry and molecular biology.

Id. at 273 (citation omitted). Therefore, the Rodriguez's finding that "[i]n a very real sense, an LR calculated by means of the FST software may be regarded as the final step in the PCR-STR DNA analytic process for a complex DNA mixture" is a mischaracterization of what the FST does and how it operates. Id. at 16.

Furthermore, relying on the principle that the likelihood ratio has long been generally accepted is also an incorrect. As expressed by both Drs. Chakraborty and Rosenberg, it is the

application of the likelihood ratio principle that matters. Dr. Chakraborty testified that a likelihood ratio model that accounts for allelic drop-in and drop-out in forensic DNA testing is not generally accepted yet. The likelihood ratio is a methodology that has been in existence for more than 200 years. The question is not whether likelihood ratios are accepted, but how that the likelihood ratio is applied and what the modeling assumptions are that are employed. Chakraborty, 12/16/13, 1091-92.

Similarly, Dr. Rosenberg testified that the likelihood ratio is a sound mathematical and statistical principle. However, it is the implementation of the likelihood ratio and how it is being used that needs to be considered when determining whether it is generally accepted in this context. Rosenberg, 12/6/13, 755-56. Accordingly, the logic applied with regard to likelihood ratios in Rodriguez, as well as the position taken by the People in this case would be that a likelihood ratio that accounts for drop out and uses a drop out rate which has been calculated from the phases of the moon would be generally accepted because likelihood ratios have been around for over hundreds of years.

While the People may point to the Rodriguez decision as persuasive authority, that opinion should carry no weight with this Court given that it is primarily based on two incorrect findings. The primary question this Court must answer is whether the *application* of likelihood ratios using the OCME's FST drop model is generally accepted by the relevant scientific community. The answer to that questions is no, as no one in the relevant scientific community is employing or even trying to employ the methods used by the FST. Indeed collectively, the other programs are using methods that have similarities with one another, while the FST stands alone with its methodology.

8. Conclusion

The FST's employs a methodology that has been rejected by the relevant scientific community. The approach that the OCME uses to provide a statistic for mixtures that cannot be deconvoluted is unique and no other laboratory is performing research to follow in the OCME's footsteps. Accordingly, this Court must find that the FST is not generally accepted as reliable by the relevant scientific community.

V. THE OCME'S LCN METHODOLOGY IS NOT GENERALLY ACCEPTED IN THE RELEVANT SCIENTIFIC COMMUNITY BECAUSE IT IS BASED ON UNSOUND SCIENCE

A. OCME's LCN methodology is based on unsound science.

1. Introduction

Fourteen years have passed since Peter Gill first published on the low copy number technique; eight since OCME brought their LCN methodology online for use in forensic criminal casework; five years since OCME published their LCN validation paper. And nearly four years have passed since the court in Megnath held that OCME's LCN methodology is generally accepted. This court should find it alarming that after all of this time and after all of these developments, not one other public forensic laboratory in the United States has followed suit. Strikingly few labs in the world perform low copy number testing for presentation in court.

It is clear why LCN DNA testing is so unpopular in the scientific community. LCN DNA testing for use in court radically diverges from the methods used in gold standard conventional DNA testing. Conventional DNA typing methods produce results that are reliable, reproducible, and robust. LCN is neither reproducible nor reliable. The pervasiveness of stochastic, or random, effects strip LCN of its reliability because any results produced by this method suffer from a high degree of uncertainty. With stochasticism, pieces of DNA that should be present go missing (allele drop out); pieces of DNA that shouldn't show up in the evidence do

(contamination: allele drop-in and transference); testing artifacts masquerade as true alleles (stutter), and proven ways of determining that two pieces of DNA belong to the same contributor no longer work (peak height imbalance).

Because of these effects, LCN testing results are not reproducible. A reproducible method "means that the same or very similar results are obtained each time a sample is tested." John Butler, ADVANCED TOPICS IN FORENSIC DNA TYPING: METHODOLOGY, p. 174, Elsevier (2011). If an OCME analyst runs the same LCN test on the same sample three times, they will not get the same results three times. OCME's LCN testing does away with the notion of reproducibility, the cornerstone principle of good science.

Statistics are supposed to convey the strength of the testing results. Any statistical method used to place a weight on LCN results must account for this uncertainty and lack of reproducibility. The forensic scientific community has not yet developed a statistical method that adequately accounts for the level of uncertainty that exists in LCN profiles.

SWGDAM and other scientific bodies require statistics for any positive association of a comparison profile to a mixture. The use of the combined probability of inclusion (CPI) without accounting for drop in and drop out was severely criticized. For years, OCME presented results in LCN cases in court with no accompanying statistic or with a flawed application of CPI. to needing a statistic for mixture cases. Dr. Budowle testified about this in Megnath and he and others published about it. The pressure on OCME to figure out a way to properly weight the uncertainty of LCN profiles. OCME developed the FST to account for these stochastic phenomenon in mixtures that are typed with LCN testing or contain low template components. As discussed supra, the FST employs fatally flawed assumptions because the uncertainty in LCN

profiles were not adequately accounted for, and methods used over a decade ago to bring LCN online originally were never rethought as science progressed.

The People essentially argue that LCN is simply an extension of the gold standard PCR STR DNA testing used in labs around the country and the world. As Dr. Holland testified, "it's simply turning up the dial". Whatever departure the LCN method makes from conventional testing Dr. Caragine and Dr. Mitchell justified as being conservative for the defendant. Yet the claims of conservativeness are conclusory. Indeed, neither of the two witnesses the People called that were not from the OCME even reviewed the OCME LCN validation. Yet the People ask this court to grandfather LCN in under the principles of conventional DNA typing and find the OCME's LCN methodology generally accepted.

Far from being a simple extension of existing gold standard PCR STR testing, LCN testing turns the basic philosophy of forensic DNA testing on its head:

> In a typical forensic DNA analysis, attempts are made to exclude the suspect (or victim), and only after failing to exclude, inferences are made about the rarity of the DNA profile. Because of the extreme sensitivity of detection, background level DNA and DNA from casual contact may and will be detected. Thus, LCN typing cannot be used for exculpatory purposes. Thus, the paradigm for the use of LCN typing differs fundamentally from general DNA typing regarding the capacity to exculpate individuals.

Def. Exhibit H4. Bruce Budowle, et al., Low Copy Number: Consideration and Caution, 2001.

LCN DNA testing has not left the experimental stage. Cutting edge science has no place in the courtroom until it withstands the rigorous scrutiny of the scientific community. The courtroom is not a laboratory for experimenting on people.

The People's main witness on the general acceptance of OCME's methodology was Dr. Theresa Caragine. This court should find that Dr. Caragine is not credible and her testimony should be weighted accordingly.

This was demonstrated at the hearing. When asked on cross examination whether there were disagreements over allele calls, she answered that OCME analysts just follow the protocols and there is nothing more to the matter. Caragine, 3/12/13, 187-. 188. The following cross examination ensued:

> Q: Is there ever disagreement about say, the number of contributors to a DNA mixture sample?
>
> A: We have written guidelines to determine the number of contributors that refer analysts to the guidelines, to the protocols, and you follow the protocols.
>
> Q: My question was, though, is there ever disagreement between a criminalist and a supervisor, say about the number of contributors to a sample?
>
> A: I would say, no, because we have written guidelines for how to determine the number of contributors, so you use those to guide your decisions.
>
> Q: So, the disagreement is not possible; is that correct?
> [Objection; objection overruled].
>
> A: Not when you have protocols because someone can't just not follow the protocol. The job of the reviewer is to insure that we are following our protocols and that there are no transcriptional errors or no omissions. That's what we do when we review a case, so I wouldn't call that disagreement. It's just a reminder, "Here is the guideline, and according to the guidelines, this is the number of contributors that we can say are in this mixture.

Caragine, 3/12/13, 191-192.

Yet, in the middle of this Frye hearing, after those questions, Dr. Caragine was forced to resign, in part because of an incident stemming from a disagreement about the *number of*

contributors to a mixture in an FST case. The New York State Inspector General investigated this in its December 2013 report, Investigation into the New York City Office of Chief Medical Examiner, Department of Forensic Biology, Defense Exhibit.

That report revealed Dr. Caragine did not follow a protocol designed to ensure proper supervisory review of results. OCME protocols dictate that when there is disagreement among analysts, the dispute is to be resolved by the Technical Leader. In direct contradiction to her testimony, Dr. Caragine disagreed with another analyst over the number of contributors to a mixture that was analyzed with FST. Dr. Caragine believed there to be three contributors; the analyst two. Dr. Caragine overruled the analyst's interpretation and changed the results. She never consulted with the technical leader.

Dr. Caragine's testimony about lack of disagreement among scientists is misleading at best. When the Court considers this testimony in light of the New York State Inspector General report, this Court should be highly suspicious of her testimony. A court's inquiry under Frye is focused on the state of scientific acceptance, including disagreements among scientists. Since Dr. Caragine was misleading in her testimony about disagreement in her lab, this Court should accord her testimony about general acceptance in the larger scientific community no weight.

2. What is OCME'S LCN DNA methodology?

Definitions of LCN DNA testing abound. See United States v. McCluskey, 954 F.Supp.2d 1224, 1277 (U.S.Dist. NM. 2013). Dr. Budowle defines LCN testing as "the analysis of any results below the stochastic threshold for normal interpretation." Def. Exhibit H4, supra. The stochastic threshold separates an area where you have confidence in the good quality of the profiles from an area below the threshold where the confidence in reliability drops. By

definition, all LCN testing OCME performs is in this stochastic zone. See Budowle, 12/9/13, 792("By nature of calling it low copy, everything under that is stochastic threshold by design.").

LCN testing is sometimes referred to as "low template" DNA testing. The labels do not always refer to the same methodologies. Sometimes low template DNA testing refers only the quantity of DNA tested, usually under 100-200 pg.

OCME calls their LCN testing "high sensitivity" DNA testing. They are the only lab in the world to use this moniker, which is not surprising because they are the only lab which does this type of testing in the United States for use in court in criminal cases. OCME coined this term because of the controversy associated with the technique. Dr. Caragine testified, "I didn't come up with the name. If I recall correctly, it was our former director. To me, it is a way of just describing what we do." Caragine, 3/12/13, 167. However, in Megnath, Dr. Caragine admitted that, "…New York City named it High Sensitivity to kind of put or [sic] more positive spin on it…" Id. at 168-169. The use of the term "high sensitivity" attempts to obfuscate the OCME's radical departure from the gold standard PCR STR testing methodology in place in forensic labs around the country.

This public relations campaign for a controversial testing technique is inimical to responsible science. See Bruce Budowle, Arthur Eisenberg, Angela van Daal "Validity of Low Copy Typing and Applications to Forensic Science," attached as Exhibit A. ("Publicizing the potential of the application of LCN typing without describing its limitations is not a responsible role for the forensic scientist to take"); National Academies of Science: Strengthening Forensic Science in the United States: A Path Forward, p.S-16, National Academies Press 2009 ("Forensic reports, and any courtroom testimony stemming from them, must include clear

characterizations of the limitations of the analyses, including measures of uncertainty in reported results and associated estimated probabilities where possible").

As noted above, N.Y. Exec. Law 49-b, §995(3) defines "'DNA testing methodology' as "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." Dr. Chakraborty agreed with definition and in fact testified that that was essentially the definition used in DNA Advisory Board and SWGDAM deliberations. Chakraborty, 12/16/13, 1084.

OCME's LCN methodology attempts to test very small amounts of DNA with a suite of novel techniques. These techniques are designed to detect low level DNA. Chakraborty, 12/16/13, 1086; van Daal, 11/21/13, 306. Using the definition in the New York Executive Law as a guide, these techniques are:

1) replicate testing, consisting of the splitting a sample of 300 pg or less into three smaller samples intended to be equally sized

2) amplifying samples of 100 pg or less;

3) increasing the number of amplification cycles from 28 to 31, against the guidelines of the manufacturer of the amplification kit;

4) testing in an ultra-clean environment;

5) altering protocols for interpretations;

6) changes in exclusion criteria; and

7) applying a statistical weight to the results: a) use of the random match probability for major donors or single source profiles; and b) use of FST for statistical weight computation for mixed LCN profiles that cannot be separated out into individual profiles, discussed supra.

These techniques are used to increase the sensitivity of the testing to allow an analyst to see DNA fragments which you wouldn't be able to see otherwise and to attempt to compensate for some of the stochastic effects which are part and parcel of LCN DNA testing. Both the amount of DNA tested and the way in which it is done affect the end result. Dir. Chakraborty, 12/16/13, 1086.

3. OCME's testing flouts manufacturer's guidelines

The quantity of DNA used in gold standard testing is approximately 0.5 nanograms (ng) to 1.25ng (which equals 500 pg to 1,250 pg). Applied Biosystems, the manufacturer of the kit which OCME and many other labs use, performed developmental validation before selling its kit.See http://www3.appliedbiosystems.com/cms/groups/applied_markets_support/documents/generaldocuments/cms_041201.pdf (hereinafter ABI manual) *last visited* 4/7/2014. When forensic labs purchase the kit, they perform internal validation. The reliability of the use of the kit under the manufacturer's guidelines are established because of both validations and worldwide use of the kit for years.

The kit was manufactured and validated to function with a specific amount of DNA at a certain number of cycles. The manufacturer states, "[a]t 28 cycles, 1.0ng of AmpFℓSTR Control DNA 9947A amplifies reliably and specifically following the conditions outline in this guide." ABI Manual. at 68. ABI did test a range of cycles and determined this was the recommended parameter. Id. In fact at 31 cycles, ABI found that the test produced off scale data. ABI specifically states that the input range of DNA should be .5ng to 1.25ng. Id.

OCME uses this particular kit in all of its STR testing, and despite the manufacturer's recommendations, OCME uses 100 pg or less for each amplification in its LCN testing at 31

cycles. As discussed above, a sufficient starting sample of DNA to go into the PCR reaction is necessary to ensure there is enough DNA to be copied reliably.

OCME increases the amplification cycles used during PCR from 28 to 31. During each cycle of PCR, pieces of DNA are copied exponentially, e.g., from 2 to 4 to 8, and so on. There are approximately 67 million copies at 28 cycles. After 31 cycles, there are 536 million. Butler, FUNDAMENTALS OF FORENSIC DNA TYPING, 128-129, Elsevier (2011). This is a far cry from "turning up the dial" as prosecution witness Mitchell Holland termed it, "making it a little more sensitive," as Dr. Caragine phrased it, or the "slight modifications" which OCME publicizes on their website, *available at* http://www.nyc.gov/html/ocme/html/hss/hss_faskedq.shtml#9, last visited 4/11/14.

The point of increasing the amplification cycles is to increase the sensitivity of detection and see DNA at levels that you would not be able to see with the standard methodology. Budowle, 12/9/13, 789. However, because of the great increase in the sensitivity, the test is more likely to detect contaminant alleles, low level contributors, and DNA transferred onto a touched item, as discussed infra. Increased amplification cycles on low level DNA also exacerbate peak height imbalance and the height of stutter peaks which can confound interpretation, especially in a mixture.

Although the prosecution witnesses would have you believe that this is akin to the "extra dry" cycle on the your clothes dryer, it isn't. It is akin to jerry-rigging more cycles on your dryer that was manufactured for a certain number by crossing some wires, drastically increasing the chance of a fire hazard.

4. OCME's own comparison of the effect of running 28 v. 31 cycles on 100pg demonstrates that any benefit to the increase of sensitivity may be exaggerated

OCME's own FST validation study demonstrates that the drawbacks to using increased amplification cycles may not be outweighed by benefits of less drop-out. OCME overstated the efficacy of the 31-cycle amplification versus the 28 cycle amplification. In the summary of Volume 16 of the FST validation, OCME concluded that "Drop-out rates were generally higher for these samples [100pg samples amplified at 28 cycles] than for 100pg samples amplified with 31 cycles, although in some cases it was lower." It would be expected that a more sensitive assay would produce less drop out when template amount was held constant.

Volume 16 summarized OCME's study that looked at the estimation of drop-out and drop-in rates in 100pg two-and three-person mixtures amplified at 28 cycles. Dr. Shapiro compared the partial and total heterozygous drop out rate for 1:1 and 4:1 mixtures amplified at both the 28 cycles and 31 cycles. He found that the 100 pg samples amplified with 31 cycles exhibited higher drop out rates than the 28 cycles nearly half the time: seven times it was higher, while eight times it was lower. With homozygous drop out rates for 1:1 and 4:1 minor contributors, the drop-out rates the results were even more surprising: eight times the 31 cycled samples dropped out at a greater rate, versus one time lower. Thus, OCME's stated conclusion in Vol. 16 summary is wrong. Clearly there was more drop out at the 31 cycles with the 100 pg samples. Shapiro, 10/16/13, 265 et seq. In fact, Dr. Shapiro also performed an R squared test, a type of statistical test that is designed to measure the how good a fit the function is to the data, and concluded that the 28 cycle assay was a better fit. Id. at 270- 271. This is yet another example of OCME obtaining results that buck the expected trend, failing to investigate the cause, and misleadingly characterizing the results.

5. The dangers of testing in the stochastic zone

LCN testing is defined by stochasticism. 'Stochastic' means random. Stochastic effects in LCN testing arise from two things: 1) the small quantity of DNA being tested and 2) the different techniques that are used in LCN testing to increase sensitivity, e.g., increased amplification cycles. Stochastic effects are random effects that occur during the amplification stage because of random sampling of the pieces of DNA which are amplified.

Drs. Budowle, Chakraborty, Coyle, and van Daal described stochastic effects. See, e.g., Chakraborty, 12/16/13, 1086 (describing six interrelated characteristics of low copy number crime scene samples that affect the ability to interpret them: 1) how much DNA is in the sample; 2) the number of contributors to the sample, which causes "PCR competition"; 3) locus and allele size; 4) allelic drop-out; 5) allelic drop-in; and 6) increased stutter.).

Dr. Caragine, referring to Dr. Budowle's 2001 paper "Consideration and Caution," testified that "many of the concerns are indeed shared by other scientists in the field and they have worked to address these concerns." Caragine, 12/12/12, 90. Dr. Caragine's position is that OCME's interpretation protocols and statistical methods account for stochastic effects. She testified that, "We maintain that these are not unforeseen effects, that we have very well characterized these stochastic effects." Id. at 80. However, the only thing that can be predicted is that the random *will* happen, but not what the random will be. The line in the old children's song Teddy Bear's Picnic is the hallmark--and the paradox--of LCN DNA testing: "you're sure of a big surprise".[6]

[6] "You're Sure of a Big Surprise," Ed., The Guardian, *available at* http://www.theguardian.com/science/political-science/2013/jul/10/science-policy1. Last visited 4/11/14.

a. Peak height imbalance and drop-out: imbalance and absence

As described above, when an analyst interprets a DNA profile, they look at the information provided in an electropherogram, a graphical depiction of the fragments of DNA measured by the capillary electrophoresis machine. DNA fragments are represented by peaks on the electropherogram. With standard DNA testing, peak heights from the same contributor at a heterozygotic locus will be fairly well balanced—in that they will be approximately the same height. Coyle, 11/8/13, 38. See also ABI Manual at, (stating "…alleles amplified with the Identifiler Kit have similar peak height values for a heterozygous genotype within a locus. This balance can be used as an aid in detecting an interpreting mixtures").

Peak heights aid in determining how much (relative) DNA was amplified in a particular aliquot, because they represent the strength of the signal the DNA fragments emit as they pass the laser in the capillary electrophoresis machine. They are the only measure of how much DNA different individuals contributed to a mixed sample. Peak heights are also an indicator of whether a sample is degraded: A degraded sample will show a "ski slope" effect on an electropherogram and the peak heights will decrease from the smaller-sized loci on the left of the electropherogram to the larger-sized loci on the right.

In LCN testing, peak height balance is radically skewed and an analyst may not be able to conclude that the two peaks are from the same individual. See Budowle, 12/9/13, 791("…It's a variation in the amplification of any one target site so one allele may be amplified more than another. And those kinds of effects will be more exaggerated in the low copy area just because you have less DNA to start with.").

Drop out is a form of peak height imbalance in which one of the two peaks of a heterozyotic locus is completely missing. Coyle, 11/8/13, 38. As discussed supra, the danger

with allelic drop-out is that a person who is heterozygous at a loci may appear homozygous. This would be erroneous. Id. at 39. Additionally, locus drop-out occurs when both alleles may drop-out at a heterozygotic locus, which would lead to complete loss of information at that location.

OCME developed protocols to attempt to address peak height imbalance and drop-out in assigning a major donor in a mixture and in assigning homozygotes. However, OCME rules sometimes get the answer wrong. Dr. Caragine testified about the protocols for allelic assignment in LCN testing are conservative. One of the concerns with allelic drop out is that an analyst may falsely call a locus homozygotic (same allele from both parents, i.e. a 12, 12 which is sometimes simply written as a 12) when it is in fact heterozygotic (different allele from each parent, i.e. 12, 14) because the analyst does not see the other allele. This would lead to an incorrect interpretation. The difference of just one allele between the suspect profile and the comparison profile (provided both are full profiles) means that they are different contributors. OCME uses a "Z" designation when the analyst is unsure whether an allele dropped out or whether the locus is homozygotic.

While the Z designation is used in both conventional and LCN testing at OCME, its use is extensive in LCN because of the amount of drop-out. In fact, OCME's protocols state that a Z is to be used in all homozygotic loci in samples amplified with less than 20 pg. See Exhibit V4, LCN Validation, Binder 8A addendum. Dr. Caragine testified that less than 25 % of the samples are amplified with 20 or fewer picograms, although stated she needed to check that number. Caragine, 4/2/13, 75. In the LCN touched study, 12.84% of the items fell into the category or 10-20 pg amplified. Furthermore, a Z designation is to be used for any potential homozygote at any of the less efficient loci, which constitute seven of the fifteen somatic loci.

Essentially this is an admission that OCME simply does not know what should be at the location. In order to deal with this problem, the OCME created a blanket rule to incorporate a Z in the allele calls. This blanket rule was wrong 38% of the time in the OCME's LCN Validation study of touched items. See Def. Exhibit V4, Binder 8A, LCN Validation. This demonstrates the level of uncertainty inherent in the LCN profile.

Peak heights assist an analyst in determining whether two peaks come from the same contributor in a mixed sample. However, because of peak height imbalance and drop-out it becomes very difficult to deduce contributors. See Coyle, 11/8/13, 38.

Dr. Budowle demonstrated the extreme peak height fluctuation in LCN samples. Dr. Budowle, in over 20 slides showing separate examples in his powerpoint demonstrated stochasticism present in 25 pg three-person mixtures in various ratios. Budowle, 12/9/13, 835 et. seq.

b. Drop-In: Contamination

Additional amplification cycles drastically increase the sensitivity of the DNA test. LCN testing is so sensitive that it detects single alleles unconnected with the sample being tested. This is known as drop-in. Drop-in is contamination. The proponents of LCN testing distinguish this contamination from gross contamination, which occurs when, for example, an analyst sneezes into a tube prior to testing and some or all of the analyst's alleles are amplified. This could happen in conventional DNA testing or LCN testing.

Contaminated results are unreliable. Forensic labs employ strict protocols to avoid contamination, including gloving, special cleaning procedures, and rules about evidence examination. But contamination is unavoidable in LCN testing. As such, the proponents of LCN DNA, such as OCME and Dr. Gill have sought to rename contamination and call it "drop-

in" in hopes of convincing the relevant scientific community that contamination is acceptable in LCN DNA testing. See generally Coyle 11/8/13, 41 (explaining that OCME distinguishes between gross contamination and drop-in).

Drop-in is simply not an important factor with conventional, 28 cycle testing given that the elevated cycling causes it. LCN is so sensitive that it picks up sporadic contamination that conventional testing does not.

i. Negative Controls

In forensic DNA testing, controls are used to help ensure that DNA testing is producing reliable results. See Coyle, 11/8/13, 51. Because of their essential role, FBI Quality Assurance Standards (QAS) require them. See Standard 9.5 et. seq., Quality Assurance Standards For Forensic DNA Testing Laboratories, FBI, p. 19, *available at* http://www.fbi.gov/about-us/lab/biometric-analysis/codis/qas_testlabs. ("The laboratory shall monitor the analytical procedures using the following controls and standards. 9.5.2: Positive and negative amplification controls associated with samples being typed shall be amplified concurrently with the samples at all loci and with the same primers as the forensic samples. All samples typed shall also have the corresponding amplification controls typed.")

A positive control is known DNA that is run and analyzed to make sure that the testing produces the correct genotype. Negative controls test for contamination. Negative controls are run with the reagents used in DNA testing along with water and are completely free of DNA. Negative controls are run in the same batch with the samples being tested. If in the negative control the presence of DNA is indicated, then it is likely that the reagents or some aspect of the run is contaminated. Contamination in the negative control indicates that the results from the

samples being tested should not be trusted as the reliability and accuracy of the results have been affected. LCN methodology disposes of this critical safeguard.

In OCME's standard DNA testing, a run will fail if there is contamination in the negative controls. In OCME's LCN DNA testing, an analyst may pass a run even if contaminant alleles are still seen. OCME's protocols state that in LCN testing a run can still pass if its negative controls have spurious peaks unless (1) the allele occurs in two of the two or three amplifications…; 2) if more than two repeating peaks are present in a negative control, or 3) if there are more than 9 drop-in peaks distributed over at least two of the three amplification aliquots for three amplifications. (emphasis added) OCME Protocols for Forensic STR Analysis, *available at* http://www.nyc.gov/html/ocme/downloads/pdf/Fbio/Protocols%20for%20Forensic%20STR%20Analysis.pdf.

Indeed, OCME keeps stretching the limits of what is acceptable. Dr. Caragine testified that the number of contaminant alleles allowed in the negative control increased from 2007 from up to five drop-in peaks to up to nine drop in peaks. Caragine, 12/13/12, 107. However, Dr. Coyle testified that finding contamination in your negative controls is was not generally accepted: "Certainly, in 28 cycle PCR testing, that's not acceptable. In most forensic DNA tests, having contamination in your negative control is considered unscientific." Coyle, 11/8/13, 52. In answering a question from the Court, Dr. Coyle also explained that with LCN testing it is possible to "have a completely clean negative control and still have contamination in your sample and vice versa." Id. at 53.

This is a matter of admissibility, not of weight. When courts have found that challenges based on possible contamination went to weight rather than admissibility they did so because the

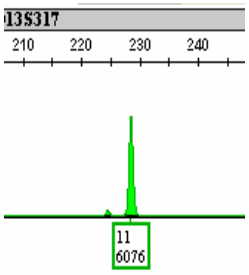
possibility of contamination was adequately mitigated by proper procedures. See, e.g., United States v. Beasley, 102 F.3d 1440, 1446-1447 (8th Cir. Minn 1996). In this case, Dr. Caragine established numerous times in her testimony that a strict quality control program was an essential component of the LCN methodology. See, e.g., Caragine, 12/12/13, 109; Caragine, 3/13/13, 278, 282. That is because the increased sensitivity of the LCN method makes contamination inevitable, not just a risk derived from sloppy human behavior.

As such, the replicate testing methodology was developed to account for the unavoidable contamination caused by the increased sensitivity of the method. The triplicate method was designed to weed out drop in alleles. If an allele was seen in one of the three runs it was considered less reliable, i.e. drop-in, and discounted. Because the methodology itself was developed around contamination, this is an admissibility and not weight issue.

c. Stutter: Pseudo Allele

Stutter is an artifact of DNA testing. Stutter appears on an electropherogram as a small peak to the immediate left (or sometimes immediate right) of the true allele. It usually occurs at one repeat less than the true allele and is known as "-4 stutter". When a stutter peak appears to the right of a true allele it is known as +4 stutter.

Thus stutter will appear in a profile like:



The small blip to the left of the larger peak is stutter in a conventionally typed profile.

Stutter is easily identified in a single source conventional profile. Extensive testing has demonstrated that a stutter peak occurs during the amplification stage and manufacturers and

labs test the percentage height of the stutter product. In contrast, with LCN testing, the heights of the stutter peaks increases. "When analyzing LCN DNA, the percent stutter increases and can be greater than the true allele. The degree of stutter is not predictable." Def. Exhibit H4, LCN: Consideration and Caution.; see also Gill et. al., "An Investigation of the Rigor…" 112 Forensic Sci. Int'l, p.27 (2000) ("For LCN, the relative size of the peak (area) is not as informative since a stutter peak may actually exceed the size of the associated allelic peak…").

Dr. van Daal echoed this concern. She testified, " In the case of a low copy number profile, it's almost impossible to tell what is and what is not stutter, given the exaggerated effects of these—that stutter peaks can be seen at. So it impacts quite dramatically on the interpretation of the profile." Van Daal, 11/21/13, 307-308; Chakraborty, 12/16/13,1088.

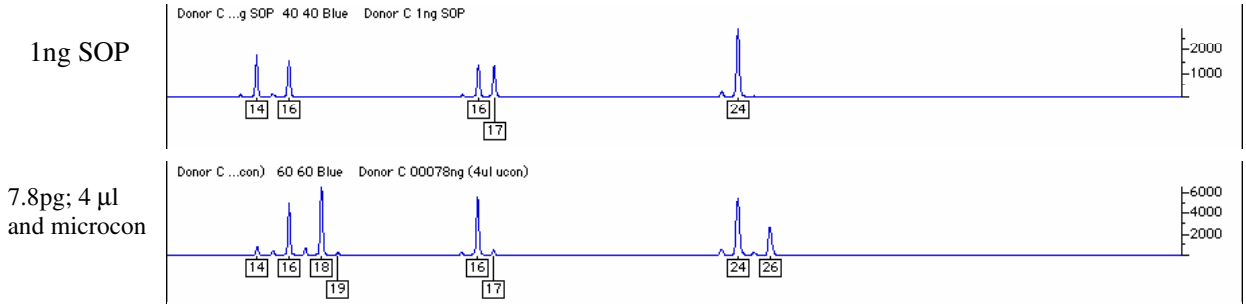
Exaggerated stutter complicates interpretation in different ways. What appears to be stutter can mask the presence of a real allele. Similarly, stutter can appear to be a real allele in a complex mixture with peaks present at a range of heights.

Yet, the OCME claims they can account for the exaggerated stutter caused by LCN typing through its replicate profiling and stutter filters, which are set by software. Dr, Caragine testified however that stutter filters missed stutter in LCN profiles in 3.1% of the time in 100pg samples, and 5.4% in samples under 50pg, compared to missing stutter 0% in the 28 cycle testing. See People's Exhibit 2. Again, OCME relies on replicate testing to compensate for the stochastic effect of exaggerated stutter arguing that stutter is reproducible; however, exaggerated height of stutter is not. Yet, exaggerated stutter can still be called as an allelic peak and input into FST.

Furthermore, stutter is not edited out in mixtures because "in a mixture, particularly, for the minor contributors, you may not known for sure if that peak is truly from a minor contributor

or is it a stutter peak." Caragine, 4/2/13, 110. In OCME's FST validation, a true two person mixture was miscalled as a 3-person mixture because of elevated stutter. Id. at 107. Stutter artifacts were miscalled as alleles. That increase in number of contributors may be anti-conservative. Although Dr. Caragine claimed that analysts don't disagree about stutter, that following protocols resolves all scientific disagreement—stutter can and has lead to incorrect results.

Dr. Budowle's powerpoint depicted examples of some of the stochastic effects pervasive in LCN testing:



Def. Exhibit G4, slide 2.

The top run ("1ng SOP") represents the amplification of 1000 pg or 1 ng which is in the target range recommended by ABI, the manufacturer of the amplification kit used by OCME The 14 and 16 peaks, as well as the 16 and 17 peaks are approximately the same size. There is one peak at the next locus, a 24, because the DNA contributor was a homozygote and it is taller than the heterozygote peaks because the strength of the signal as measured as the height of the peak is from two of the same alleles. The second run, "7.8 pg" is a LCN sample (although typed with different methodology than OCME, but still using enhanced sensitivity techniques). In the LCN run, the 14 and 16 are radically imbalanced, as are the 16 and 17. Additionally, there are several peaks in this second run that are not present in the first run—the 18 and 19 and the 26--all drop-in or contaminant alleles. Budowle, 12/9/13, 802.

6. LCN applied to the real world: mixtures, touch, degraded samples

A method must be tailored to its use. If the forensic method cannot handle the challenging realities of the crime scene samples on which it is used then it is not fit for purpose. LCN samples are almost always mixtures, touch samples, and frequently degraded. Yet, OCME's LCN methodology—already problematic with single source profiles—is radically under-equipped to produce reliable results under these conditions. Although OCME conducted a LCN validation, the validation was insufficient with regard to the types of samples on which the methodology would actually be used: mixtures and touch samples.

Dr. Van Daal observed that the mixture data she reviewed in the LCN validation data "were very typical of, and demonstrated typically, the sorts of stochastic effects that I and others have expressed in the literature as being of concern So they showed numerous and repeated examples of the allele drop-outs, locus drop-outs, exaggerated peak height imbalances, and it was difficult to tell what the ratio, for example of mixtures were give that change in the ratios of mixtures form one replicate to another in particular." Van Daal, 11/21/13, 311.

a. LCN Was created to deal with touch samples

LCN was borne from the desire in typing touched items. Although touched items can contain amounts of DNA sufficient for conventional analysis, many touched evidentiary items are in the low template range. The first paper in which OCME described their LCN protocol (a 2 ½ page article) was entitled, "Maximization of STR DNA Typing Success for Touched Objects." Attached as Exihibit B. Caragine testified that"[w]hat is being tested in the low template area are typically the touched items where we expect less DNA." Caragine, 12/12/12, 92. She claimed that OCME had typed 5,940 touched items with their LCN technique. Caragine, 4/2/13, 54.

Caragine added that "the majority of the samples that we see with touched items are mixtures." Caragine, 12/12/12, 129. She also admitted that complex mixtures are harder to interpret than single source profiles. Caragine, 4/2/13, 59.

Yet, the touch study in the LCN validation was insufficient because OCME only used one known contributor with the touch items; the method of handling the items and any cleaning procedures were not clearly explained, and the source of the noncontributory alleles was not adequately investigated.

b. Touch mixtures are degraded

Complicating the interpretation of DNA touch mixture samples is that they are naturally degraded. Caragine admitted that touched and handled items demonstrate this degradation. Caragine, 4/2/13, 54. In typing mixtures for the FST validation studies, her team discovered that "touched samples did show degradation just naturally." Caragine, 3/13/13, 329.

This phenomenon is not unknown to the many DNA laboratories that have rejected LCN methodology as practiced by OCME. It is a commonplace that cells sloughed off in touch DNA are in an apoptotic, or dying, state. Because the cell is dying, the DNA is degrading.

Degradation complicates the interpretation of touched mixtures. Alleles drop-out, and peak heights decrease from the smaller to the larger loci, creating a "ski slope" effect, as described above.

In fact, so many samples are degraded that the OCME attempted to develop a degradation module for the FST program during the validation, a venture that ultimately failed. When real world drop-out rates, which better reflected the degraded nature of the samples which were used, FST had trouble distinguishing between true- and non-contributors.

Therefore, in evaluating the LCN methodology and any statistical mixture interpretation of LCN--such as FST--it is critical that the Court analyze the response by the relevant scientific community for what LCN was designed for, and is most used in: complex degraded touch mixtures. See Del Maestro v. Grecco, 791 N.Y.S.2d 139, 141 (2d Dept. 2005) (upholding trial court's finding that plaintiff had failed to show general acceptance of expert theory; theory was based on literature dealing with premature babies although case at bar dealt with full-term infant).

c. Uncertainty and non-reproducibility in mixtures

OCME's use of increased amplification cycles in the stochastic region has not overcome the uncertainty associated with stochastic effects. As Dr. Van Daal put it, "with a mixture you don't know what—when you only see one or two peaks, whether that's a true reflection of the alleles in that mixture or if, in fact you lost, one, two, three or four alleles." Van Daal, 11/21/13, 308. See also Def. Exhibit H4, Budowle, et. al, LCN Consideration and Caution, supra ("mixture analyses and confirmation of a mixture are not reliable with LCN typing because of imbalance of heterozygote alleles, increased production of stutter products and allele drop in can occur.").

A review of Defense Exhibit O4, particularly the three-person touched mixtures in FST validation Study 3E that were interpreted by OCME criminalists as single source samples, bears out Dr. Van Daal's observation. The 14 four-person touch mixtures that were interpreted as two-person mixtures also reflects that drop-out in a low copy sample simply cannot be reliably predicted. However, the LCN validation studies did not involve touch mixture samples.

Dr. Caragine claimed that OCME could account for the uncertainty. In 2008 at the Megnath hearing, she claimed that "[w]e showed that our protocols are 100 percent reliable

because when we look at our known profiles, 100 percent of the trial [sic] we got correct results." In 2012, she did not extend this level of confidence, but still maintained that LCN testing as practiced by OCME was reliable.

After reviewing the LCN validation studies, Dr. Van Daal challenged Dr. Caragine's 2008 statement that 100 percent of the results were correct, "[b]ecause the validation studies show that that's not the case." Van Daal, 11/21/13, 312. To provide an illustration of this, Dr. van Daal prepared a chart, Defense Exhibit Triple R from mixture data she reviewed in the LCN validation.

As discussed supra, OCME uses Z to indicate the possible presence of another allele at a locus. Dr. van Daal pointed out that inconclusive results are not the same as correct: "so, again, it's not correct to say 14, Z is the correct genome type." Id. at 318. She further dryly noted "if you make no conclusion, you can't get it wrong." Id. at 319. The effect is having more than incomplete data, more than just having a partial profile: "the major concern is not so, per se, having a partial profile but in the ability to interpret the profile and draw conclusions from it. If the data is not interpretable, which a lot of this isn't, then it's not good data." Id. at 319-320.

The Court also inquired whether the data even as a partial profile still had value. Dr. van Daal replied "…my concern with interpretation of data like this is whether its in fact, valid in the first place. Certainly for a profile with standard testing where the cycle number is not raised and those exaggerated stochastic effects don't exist, they you're right to have eight good loci it still give you some value as evidence. But the concern here is whether in fact, the interpretation is valid in the first place." Van Daal, 11/21/13, 321.

What gave Dr. Van Daal concern about the validity of the interpretation was the lack of reproducibility:

> if you did a standard DNA profile, if you ran it three times, you would see the same result three times, in effect. You would see slight variations in the peak height balance say from 60 to 80 percent but not from [zero] to 80 percent. And you are almost certainly not going to see drop-in and allele drop-out. So you would get the same result if you repeated that sample. Clearly, you don't have that happening here. When you take a low copy number, increase cycle sample, you don't see the same result. It's not reproducible.

Van Daal, 11/21/13, 321- 322.

d. Contamination

Lab contamination was a possible reason for the enormous number of foreign alleles in the touch mixture studies. Because of the boosted signal of increased cycles, background contamination is also a major concern for LCN casework.

Dr. Coyle testified that as a result there is "a lot more information in the mixture. The mixtures are going to be more complex. And by having additional contamination, there is a concern that the mixture doesn't represent the sample from the scene." Coyle, 11/8/13, 87.

Both OCME's LCN validation and FST validation show that LCN methods when applied to touch samples produce enormous amounts of non-contributor alleles, even for items that had been cleaned prior to being touched. Drs. Van Daal, Coyle, and Budowle testified about these noncontributor alleles and possible causes, such as contamination, inadequate cleaning protocols, and transference.

Dr .van Daal also testified about extraordinarily high amounts of foreign alleles in both the LCN and FST validation studies with the touched samples. See Def. Exhibit SSS. For example, in the volume 8 of the LCN validation, touched items exhibited up to 11 drop-in alleles, that is, alleles from an unknown source. Van Daal, 11/21/2013, 329. The 'fingerprint' sample had 11 non-contributor alleles in one run; the 'watch' had five; the 'id cover' had nine, and the 'calculator' had six drop-ins in one run.

Dr. Caragine made clear in her testimony that part of the LCN methodology involved extensive cleaning measures and prevention of contamination. Dr. Caragine testified that if the OCME's protocols and quality control program are followed then LCN is acceptable for casework. Caragine, 3/12/13, 278. She also testified that the cleaning procedures are an important part of testing, id. at 282, and that the quality control protocols were developed to address the issues with low copy number DNA testing, including stochastic effects. Caragine, 12/12/12, 109.

Dr. Budowle pointed out that these protocols and quality assurance standards set in place for LCN testing were unable to prevent "'very high levels of…noncontributory alleles" in the FST touch studies. Budowle, 12/10/13, 947.

Furthermore, since these touch samples were created in a controlled setting, it is likely that there is a "higher prevalence of drop-in in real world type samples than more of a controlled lab interpretation." Id. at p. 918[emphasis added]. This also has implications for the drop-in rate employed by FST. Dr. Budowle demonstrated in his testimony that under-estimating the drop-in rate could bias the defendant, counter to the claim made by Dr. Mitchell and Dr. Caragine, as discussed supra.

Tellingly, the prosecution and OCME offered no explanation for the high level of foreign alleles. The lab utterly failed to demonstrate that it had taken steps to investigate the cause of the possible contamination. The only line of questioning the prosecution pursued with Dr. van Daal was that those noncontributor alleles could be stutter. The People's suggestion would of course mean that elevated stutter—a product of LCN testing—was being extensively miscalled as real alleles—a real problem for casework. See van Daal, 11/22/13, 436 (by prosecution: "Would you agree that most of the drop in exhibited in this exhibit are in the minus 4 stutter position?").

The purpose of validation experiments, especially those involving touched items, is to simulate case work. Id. at 332. Thus, this court has good cause to conclude this high level of possible contamination is also occurring in case work as a result of the LCN methodology.

e. Transference

DNA can be transferred from a person to an object. This is called primary transfer. It can also involve cells from saliva. Def. Exhibit N4, Warshauer et al, An Evaluation of the Transfer of Saliva Derived DNA (2012).

With secondary transfer, DNA from one source is found on another through an intermediary vector. For instance, if you put a pen in your mouth, the DNA on the pen from your mouth is primary transfer. If your friend should then pick up the pen and get your DNA from the epithelial cells in your saliva on his hand, then this is secondary transfer. Tertiary transfer occurs when your friend would grab another object and deposit your DNA on it. Id.

Understanding transference has become critical with the growth of testing touch mixture samples. An innocent individual unconnected with the crime event could have his DNA on a piece of evidence because it was transferred there by someone else.

One of the concerns with LCN testing with its increased cycles is its ability to pick up transferred DNA—that is, DNA that is unconnected to the crime scene evidence, but was rather transferred to crime scene evidence by another individual who did come into contact with the crime scene evidence.

As noted above, Dr. Budowle has written that "because of enhanced sensitivity, casual contact must be considered. If the victim and the accused had any previous contact, the result may have no relation to the case at hand." Def. Exhibit H4, LCN: Consideration and Caution. In 2009, Dr. Budowle reiterated that "[d]ue to the enhanced sensitivity, secondary transfer cannot

be ruled out as a possible explanation for LCN typing results." Budowle, Eisenberg, van Daal "Validity of Low Copy Number. . .," Exhibit A. Dr. Budowle's prescient warning came to light after the FST touch studies became available.

In Defense Triple T, some of the alleles repeat in all three replicates. Some samples had repeating alleles at numerous loci. Dr. van Daal hypothesized one possible source of this extensive contamination could have been contamination in a reagent, but because that level of contamination was not seen in most of the samples used with that reagent, that explanation was not supported by the data. Van Daal, 11/21/13, 349.

The more likely explanation was transference: "That explanation is possibly what's happening here, that DNA that hasn't actually truly contributed to the sample has ended up on the sample as result of what's referred to as secondary transfer." Id. at 349-350.

OCME's inability to control or account for transference is critical to whether LCN is viewed by the relevant scientific community as being suitable for use in-court criminal casework. See id. at 350 ("With standard DNA testing there is usually, potentially, sufficient DNA or more likely to be sufficient DNA not to see that small amount of my DNA. However, with touch items or items, or these kinds of items, very low level items, that DNA that's unrelated to the actual event that forms part of the crime, eventually I guess, is unrelated to it. So it is much more likely that DNA unrelated to a crime will be seen on a sample as a result of these very low levels of FNA that are being analyzed.").

OCME scientists conducted a transference study, that indicated transference could be an issue in DNA testing. "When DNA is obtained from epithelial cells on handled items, rather than from body fluids, lower quantities of DNA are expected and mixtures are often encountered. In addition, particularly with the use of measures to increase the sensitivity of genotyping

protocols, secondary or tertiary transfer of FNA may be an issue." Jeannie Tamariz, Andrew Wong, Jaheida Perez, Adele Mitchell, Mechthild Princz, Theresa Caragine, "Investigation of the Detection of DNA through Secondary and Teritary Transfer," Attached as Exhibit C.

Yet the OCME scientists attempted to diminish the results of its study, when it concluded that "[t]hese findings demonstrate that, although secondary and tertiary transfer of DNA may occur it is unlikely to be a factor in a criminal investigation, as transfer was not detected except in contrived circumstances, such as intimate contact between Person A and Person B." Id.

Their conclusion ignores the fact in 11 percent of their trials, a mixture from the person actually handling an item and the transferred person was obtained. Additionally, OCME detected alleles from tertiary transfer when contact followed immediately or when the two individuals were intimate partners. Moreover, the FST touch studies strongly indicate that LCN testing of touch mixture samples creates a significant danger that foreign alleles will either contaminate the sample, or be transferred into the mixture. Id.

A stark example of the problems of applying LCN testing to complex and degraded touch mixtures is the 60 pg three-person sample PenB in Defense Exhibit X and O4. As Dr. Shapiro pointed out, two of the three contributors to PenB were ruled as "cannot be excluded" (CBE) by the analyst examining the mixture. FST included one of those CBE contributors, and excluded the other. However, because of the virtual soup of alleles created by the true contributors and 37 extra foreign alleles that were called by the criminalist, a total of 9 non-contributors were included by FST in the mixture, on of whom, "JB," was classified as a contributor by strong evidence.

This level of uncertainty has been rejected by the relevant scientific community as unreliable forensic science. The data in both the FST and LCN touch studies do not support its

use for criminal case work. There is "simply too great an analytical gap between the data and the opinions offered." McCluskey citing Paoli II: Mitchell v. Gencorp Inc., 165 F.3d 778, 782 (10th Cir. 1999).

6. Splitting the sample into three: ensuring reproducibility or making problems worse?

The OCME's LCN protocols call for the splitting of the DNA sample into three smaller samples, called aliquots, intended to be equally sized. The three aliquots are then amplified and run through the capillary electrophoresis separately. The three aliquots are then combined and run again. Originally the theory was that only alleles which appear in two out the three amplifications are included in what is known as a consensus profile. Only alleles which were present in at least 2 out 3 amplifications were determined to be reliable and therefore, included into the consensus profile. This is an effort to avoid the drop in phenomenon pervasive with LCN testing. Of course, while usually successful in avoiding drop-in, the method would exclude an allele associated with the crime scene sample that dropped out of the amplifications two of the three times.

However, there is no consensus in the scientific community on whether splitting an already small sample into even smaller samples adequately offsets the stochasticism inherent in subjecting small samples increased amplification cycles, or simply exacerbates the problems: "[c]ritics of the biological model suggest that splitting an already low level sample into multiple aliquots would increase the stochastic effects seen in LTDNA profiles because fewer template molecules are subject to the PCR process in each reaction." Def. Exhibit U3, Grisedale and van Daal, Comparison of STR Profiles from Low Template DNA Extracts With and Without the Consensus Profiling Method, Investigative Genetics, 2013.

Dr. Caragine testified that with replicate testing (and pooling) controlled for various stochastic affects, such as drop-in, increased peak height imbalance, and increased stutter, both problems which plague low copy number DNA testing. See, e.g., Caragine, 12/12/12, 79, 117. She testified that they split the sample into thirds "to ensure that we don't make a mistake." Id. at 82.

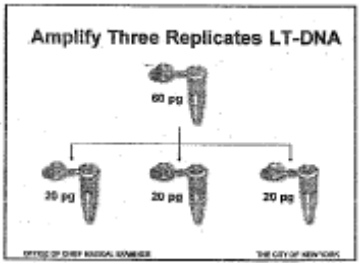
However, far from solving problems, it exacerbates them. Dr. van Daal designed a peer reviewed study that tested whether this consensus method produced more accurate profiles than concentrating the sample and testing it once. The study showed that the "loss of information that occurs when a sample is split for amplification indicates that consensus profiling may not be producing the most informative DNA profile for samples where the template amount is limited." Def. Exhibit U3, Comparison of STR Profiles, at 1.

Furthermore, Dr. van Daal and her colleague found that "[f]or both starting template amounts [100 and 25 pg], the data show that splitting the sample into three aliquots and constructing a consensus profile did not result in the most informative profile compared with a profile where the DNA extract was amplified in one reaction. While the consensus profile approach did eliminate allele drop in, all other measures of profile quality were improved when the sample was not split….Overall, this study has demonstrated that performing standard cycling STR typing on non-split DNA extracts will result in profiles with a higher percentage of total loci compared with the consensus profiling technique." Id. at 8.

In addition to publication, the paper was presented at Promega as a poster and feedback was that it made sense not to split the sample. Van Daal, 11/8/13, 363. Dr, van Daal testified that the "overall conclusion was that the-there was more information in the profile obtained when the sample wasn't split, and the profile was of a better quality. The only advantage to splitting

the sample appeared to be that any random lo-level drop in allele was removed from the consensus profile sequence." Id. at 360.

Dr. Caragine represented in her PowerPoint that the triplicate method performs like this:



People's Exhibit 2.

This is misleading. Dr. Caragine never revealed to the court that under basic scientific principles the amount of DNA in each replicate will not be the same. Dr. Budowle demonstrated that the amount of DNA in each replicate is not equal. Thus, further exacerbating the issues with dividing the sample into three smaller samples is that the replicates do not contain the same amount of DNA. Dr. Budowle calls this a "fundamental problem." Budowle, 12/9/13, 819.

Dr. Budowle confirmed this finding in a study published in 2014. That study compared the consensus and the single "complete" methodologies and evaluated their success on whether they obtained the true genotype and were capable of detecting a drop out or drop in allele. The study examined one locus in a single source sample using logistic model and empirical data. Def. Exhibit K4, Jianye Ge and Bruce Budowle, Modeling One Complete Versus Triplicate Analyses in Low Template DNA Typing, Int'l J. Legal Medicine (2014).

Dr. Budowle observed: "If everybody is moving in the direction of accommodating drop-in and drop-out, as I said already, to me it is—you have to ask the question does it really make technological sense to split something into third to try to get some robustness out of it because you don't have procedures to analyze it, were you say if I see two out of three times or three out

of three times, I have confidence its there. That no longer applies. In fact, it's actually detrimental to the long term capabilities of efficiently analyzing and interpreting low level evidence. It makes more sense to go with a greater amount of DNA you can put in one amplification." Budowle, 12/9/13, 874-875.

Thus, the evidence presented at the hearing demonstrated that splitting a LCN sample only exacerbates the problems intrinsic to LCN profiles.

7. Statistical interpretation of LCN profiles

a. LCN builds in bias

One of the earliest concerns with LCN testing was how to appropriately weight the results. Any stochastic phenomenon would have to be considered in the computation so that the statistical significance would not be accorded undue weight. Dr. Budowle pointed out in 2001 that "statistical interpretations should be modified to better represent the uncertainty associated with LCN typing." As presently practiced by the OCME, LCN testing builds bias against the defendant right into forensic DNA analysis. In the 2009 LCN validation paper, Dr. Caragine and colleagues stated that "[s]ince allelic drop out caused by stochastic effects is a common occurrence for LT-DNA samples, the absence of a comparison sample's alleles from a mixture is not necessarily indicative of an exclusion." The authors concluded that "[a]n analyst should consider whether the alleles detected are what one would expect to see had an individual contributed to the mixture."

During the hearing, Dr. Caragine was asked whether OCME had criteria to exclude a possible contributor. She responded that it "depends upon the sample. If many alleles of the suspect sample are not seen in the mixture, that would be an exclusion. So probably, if it is six or seven alleles that would be an exclusion. . .Low template for example we would not say they

cannot be excluded." Caragine, 1/2/13, 41-42. See also §7(b)(i) of OCME mixture protocols, OCME Protocols for Forensic STR Analysis, *available at* http://www.nyc.gov/html/ocme/downloads/pdf/Fbio/Protocols%20for%20Forensic%20STR%20Analysis.pdf, last visited 4/11/14 (suspect would be excluded if "1)Three or more alleles seen in the DNA profile of the comparison sample are absent at the efficient loci; 2) Many alleles seen in the DNA profile of the comparison sample are absent at any locus.").

Therefore, bias infects the process by permitting a mismatch between the suspect's profile and the alleles in the mixture can be explained by resorting to allelic drop-out and drop-in.

The OCME's practices stand in stark contrast to the prevailing concept of conservatism as described in NRC II. NRC II defines "conservative" as "[w]hen in doubt we err on the side of conservatism, that is in favor of the defendant." Budowle, 12/9/13, 812.

Dr. Budowle further explained that this means "when possible and you can exclude somebody, you should, and that's just as important as including somebody…it's just as important when erring on conservatism not falsely including somebody and also not stating the strength of the evidence." Id. at 813, 814.

The OCME's LCN methodology does not follow this norm. Id. at 815. Dr. Budowle illustrated how the OCME's LCN methodology was a sharp departure from NRC II using an example the OCME itself provided at a recent conference in Asia. Sheila Dennis, the head research scientist for the OCME's Forensic Biology Department, gave a presentation where she explained that under her lab's consensus method, a locus would not be counted if no alleles were called in the second and third replicate runs.

Dr. Budowle explained that that practice was not conservative. If a suspect's profile did not match the alleles in the first run, that that locus should be accorded exclusionary weight, not be deemed inconclusive: "I think that's something that has to be addressed to ensure that we don't falsely include somebody." Id. Therefore, the OCME's protocols ignore this exculpatory data is inconsistent with NRC II.

The Court asked whether other labs would have similar protocols as the OCME for exclusion. Dr. Caragine responded that other labs use the 'no conclusions can be drawn' category. When the Court pressed whether those labs base it on the same criteria as the OCME, Dr. Caragine answered, "[p]robably similar criterion, not other labs in the US are doing low template [sic]; they would not have those criteria, that's our criteria." 12/13/12, 42.

When the Court attempted to further clarify whether other labs would have those standards with respect to high template mixtures, Dr. Caragine stated, "They would have protocols in their lab. I know labs say no conclusions can be drawn. They don't have FST like we do so now we can put a number on that." 12/13/12, 42-43.

The Court's question regarding whether other labs were using similar protocols, went directly to the general acceptance standard, but Dr. Caragine could not answer it. Even if other American labs actually did LCN testing, it would be impossible to compare the amorphous standard of "many." The only guidance an analyst is given is "many." This ill-defined protocol may lead to biased interpretation because it leaves discretion to the analyst. Furthermore, there is no way to evaluate a standard like that, as what one lab may define as many may be very different from what another lab would define as many.

b. OCME's past lack of recognition of the need to account for stochastic phenomena in a statistic

In 2001, Dr. Budowle warned that "statistical interpretations should be modified to better represent the uncertainty associated with LCN typing." For over five years, OCME ignored concerns from Dr. Budowle and other concerned scientists, and used two indefensible statistical methods in its LCN casework. First, the OCME was reporting a qualitative conclusion, "cannot be excluded," for mixtures where some of the defendant's alleles were missing. This practice was in contravention to Section 4.1 of the SWGDAM guidelines, which clearly states that , "The laboratory 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." Secondly, OCME was using a CPI that did not account for allelic drop-out and drop-in: As Dr. Budowle explained at the hearing, "I've been an open critic about the CPI for low copy all along. In fact most people are." Budowle, 12/10/2013. These criticisms were leveled in Megnath and in the literature and created pressure on OCME as the only public lab performing LCN typing for in court use to develop a statistic which could account for stochastic phenomenon.

c. Mixture statistic does not fix fundamental problem

In developing FST, the OCME never bothered to examine the original assumptions on which LCN testing was built to determine whether they were consistent with the application of a new statistical model. As Dr. Budowle testified, "OCME created a low copy number procedure based on this, the consensus approach, three replicates." Budowle, 12/10/13, 925.

After five years, the lab attempted to resolve its statistical problem without dealing with the underlying issue of replicate testing and the consensus approach. As Dr. Budowle explained, "you're trying the sort of put-the finger-in the-dike-to-stop-the-flood fix. This is a consequence

to what you do if it doesn't [fix it]. Id. Dr. Budowle leveled his criticism against the attempted statistical fix by OCME, for the fundamental problem that seems to have kept the rest of the American forensic community from following the lab down the LCN rabbit hole: "If you're really doing drop-in and drop-out, it makes absolutely no sense to divide that into thirds." Id. Dr. Caragine was anchored in replicate testing and consensus profiling, where they claimed they could use peak heights to discern individual profiles in the stochastic region. Without questioning that first approach, they ironically ended up with a statistical method that abandoned peak heights, ironically not comporting with their consensus profiling analysis.

This irony caused Dr. Budowle to conclude, "they either need to rethink the front end or they need to calculate something that describes this dependency that they base reliability on when they present the profile." Budowle, 12/10/13, 925.

d. Failure to Rethink RMP: Uncertainty of LCN Typing Not Reflected in RMP

OCME protocols presently allow criminalists to deduce out major donors from LCN mixtures. However, using an RMP for deduced-out major donor in a LCN mixture is flawed, because the deduction does not adequately account for the uncertainty caused by stochastic effects, such as drop-out and drop-in. See Budowle, 12/10/13, 927-928. As discussed supra, any statistical weight accorded to a DNA testing result must properly account for the weight of the evidence. The high level of uncertainty in a LCN profile must be accounted for even if the OCME analyst deduces out a major component. In 2001 Budowle warned that "statistical interpretations should be modified to better represent the uncertainty associated with LCN typing." Def. Exhibit H4. Yet, this has yet to be done with deduced major profiles.

Paradoxically, the OCME does not consider peak heights reliable enough to look at to determine drop out rates in its FST program. However, the lab allows its analysts to deduce out a major contributor by relying on those same unreliable peak heights.

B. THE OCME'S LCN METHODOLOGY IS NOT GENERALLY ACCEPTED AS RELIABLE BY THE RELEVANT SCIENTIFIC COMMUNITY

1. Courts are split in ruling on admissibility of LCN DNA

Case law concerning the admissibility of LCN evidence is sparse. This is unsurprising given the lack of use of LCN DNA testing. Nevertheless, some courts have considered LCN challenges. However, no appellate court has ruled on the admissibility of LCN DNA testing.

Courts have excluded what they call the results of LCN DNA testing in three cases after holding a hearing that defendant is aware: Espino, Lapointe, and McCluskey and admitted it in Megnath. The Southern District's ruling in Morgan appears to be tabled until the DNA Subcommittee engages in a review of the lower limit of OCME's LCN DNA testing.
Some courts have heard challenges to DNA evidence based on the unreliability of LCN DNA testing but ruled that the DNA testing in the cases before them did not qualify as LCN DNA testing. In United States v. Grinnage, 486 Fed. Appx. 325, 329 (3d Cir. 2012), the defendant claimed that the DNA testing in his case involved "an unusually low amount" of DNA. The Third Circuit found however, that "this is not one of those [LCN] cases," as the template in at least one of the samples tested was .65 ng (650 pg). However, the Court did note that "…there maybe cases where a Daubert hearing will be necessary to discern a minimum acceptable mass threshold, below which the PCR STR methodology is unreliable."

In Lapoint v. Commissioner of Correction, the Court held that the methodology used by Orchid Cellmark was unreliable. Orchid Cellmark was increasing sensitivity by employing post

amplification clean up. Orchid Cellmark, a private lab, did "not consider its procedure to be a LCN procedure." 2011WL1759129 (Conn. Super. 2011); *rev'd on other grounds*. The court ruled that the methodology was unreliable. Id.

In United States v. McCluskey, the federal district court in New Mexico ruled that the results of LCN DNA testing were inadmissible after conducting a Daubert hearing. 954 F.Supp.2d 1224 (U.S.Dist. NM. 2013). The methodology employed by the New Mexico lab is not the same as that used by OCME. First, the challenged sample in the case was 215 pg. Id. at 1276. Secondly, the New Mexico lab stated that it used no special procedures or methods of interpretation for LCN testing. Id. at 1277, 1284. However, the New Mexico lab tested evidence with increased injection parameters. Increased injection parameters—which OCME also does—increases the sensitivity of the test. Because the sample was injected with increased time parameters, it was not eligible for upload to NDIS. Id. at 1283.

The court found critical that the sample was not able to be uploaded to CODIS. The Court found that the government had not met their burden in proving that the New Mexico lab was able to retain reliable results under 250 pg, the lab's stochastic threshold. The Court further noted that "peer review" and publications had raised serious questions about LCN testing. Finally the court found that "the reliability of LCN testing is not 'generally accepted in the relevant scientific community.'" Id. at 1288.

Thus, even under the more liberal standard of admissibility imposed by Daubert, the McCluskey court ruled that the results of the LCN testing were not reliable enough to go before a jury.

In People v. Espino, (CA. Sup. Ct., L.A. Co., 2009, NA076620), YSTR testing underwent a montage process to increase sensitivity. The Espino court found that there was no

general acceptance as to the methods used in the LCN testing; nor was there general acceptance in the scientific community as to interpretation of those results, including any statistical measures that could contextualize the results. See Oral Decision, Attached as Exhibit D.

During this Frye hearing, both parties referred to the LCN Daubert hearing in the Southern District of New York, United States v. Johnny Morgan (12Cr. 22 VM). The People represented that the Court had ruled that the LCN results in that case admissible. However, subsequent to that ruling from the bench, the defense moved to reopen based on information she had learned. The defense learned that OCME represented to the DNA Subcommittee and the Forensic Commission that OCME would not test DNA samples in forensic casework under 20 pg. See DNA Subcommittee proceedings, People's 34, 35.

Nevertheless, in Morgan, the OCME amplified a 14 pg touch sample. The OCME's protocols now allow for the amplification of DNA amounts under 20 pg (total DNA quantity of 60pg) with a supervisor's approval. See OCME Protocols for Forensic STR Analysis, *available at* http://www.nyc.gov/html/ocme/downloads/pdf/Fbio/Protocols%20for%20Forensic%20STR%20Analysis.pdf. Thus the final status of the ruling remains undecided at this time.

On March 26, 2014, the New York State Forensic Commission met and Commission Member Barry Scheck voiced the concern that the OCME had made the representation to the Commission that samples under 20 pg would not be tested. The Commission voted 9-2 to send two questions to the DNA subcommittee to investigate: "to determine if there is a quantity of DNA that is required before the method becomes valid, and for the DNA Subcommittee to review any changes to the standard operating procedure for LCN DNA testing at OCME due to advances in science, and to determine if the changes have been validated as well." Meeting

Minutes, New York State Commission On Forensic Science, attached as Exhibit E. This issue is on the agenda for the next DNA subcommittee meeting on May 30.

The only other New York court to hold a Frye hearing on OCME's LCN DNA testing is Megnath. Other New York courts have denied Frye hearings on LCN DNA testing based on the Megnath decision. See, e.g., People v. Garcia, 39 Misc.3d 482 (Sup.Ct. Bx 2013). Megnath's ruling is not binding on this court; nor should this Court follow Megnath's logic because it is flawed.

Megnath found that "LCN DNA testing is simply a more sensitive form of HCN DNA testing which uses the gold standard Polymerase Chain Reaction Short Tandem Repeat (PCRSTR) technique to identify a person's DNA profile." People v. Megnath, 27 Misc.3d 405, 411 (Qns. County Court 2010). The court reaches this conclusion by finding that the "LCN DNA testing process uses the same procedures as HCN DNA testing;" "the principal difference merely being the number of amplification cycles and the manner in which the scientific data is interpreted when lesser amounts of DNA templates are tested using the LCN DNA method." (emphasis added) Id. at 410-411.

The "mere" difference in number of amplification cycles and the interpretation of scientific data are, obviously, critical for the reasons discussed supra. Nowhere does the court mention the splitting of the sample into three smaller samples for LCN testing, which is highly controversial and the subject of research that was presented at this hearing. The court misunderstands the difference between LCN and HCN testing, stating that "…many of the same scientific issues that arise in HCN DNA testing, such as stutter, allelic or locus drop-out, and allelic drop-in also occur in LCN DNA testing," although the court does admit they occur more frequently with LCN testing. Id. at 410. As discussed supra, it is extremely rare to see these

effects in conventional DNA testing. Furthermore, one of the concerns with stutter in LCN profiles is its exaggerated height which makes it more likely to be mistaken for an allele.

Megnath relied on the fact that LCN was used in some other countries and was approved the by the New York Forensic Commission. The court also relied on the fact that LCN has been used in missing persons investigations and increased cycling during PGD. Id. at 407-408. However, as is discussed infra, this reliance is misplaced.

In the end of the decision, the court ruled, somewhat gratuitously given the year long hearing he held, that OCME's LCN methodology is not novel. Furthermore, his reasons for concluding that LCN is novel is highly criticized by the legal community. See, e.g., FAIGMAN, DAVID, JEREMY BLUMENTHAL, EDWARD CHENG, JENNIFER MNOOKIN, ET.AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY (Thomson 2013) ("[s]uch analyses are not just simple, they are also simplistic—it is akin to assuming that if a recipe works by cooking the dish for 30 minutes at 300 degrees, it will work equally well cooking it for 15 minutes at 600").

The federal court in McCluskey also criticized the Megnath court for making "questionable findings that the Frye test only applies to novel scientific evidence and that LCN testing was 'not a novel scientific technique' (despite changes in procedure and interpretive methods); and "not account[ing] for the increase in stochastic effects in LCN testing, merely observing that stochastic effects may occur in any testing." McCluskey, supra, 954 F.Supp. at 1280.

2. The scientific community has rejected LCN for use in court in criminal cases

"Where controversy rages, a court may conclude that no consensus has been reached." Wesley, supra, at 438. As one text put it, "it is fair to say that LCN typing is the subject of great

dispute among some of the leading lights of the forensic community." FAIGMAN, D. ET AL, MODERN SCIENTIFIC EVIDENCE, supra; Natasha Gilbert, "Science in Court: DNA's Identity Crisis," Nature, Vol.464, p.347-348 (2010) (discussing the "highly charged debate in the scientific and law-enforcement communities about low-copy number analysis"); JOHN BUTLER, ADVANCED TOPICS IN FORENSIC DNA TYPING: METHODOLOGY, p. 318 Elsevier (2011) ("The debate of how to handle LT-DNA samples reached a boiling point at two large international scientific conference in the fall of 2009.").

3. OCME is the only lab in the USA using LCN for in court use in criminal cases

The OCME is the only lab in the United States which uses increased amplification, replicate LCN DNA testing for use in court in criminal cases. Caragine 12/12/12, 67. Dr. Budowle testified that other labs in the United States are not using it because "it's in area where it's –there lots of risk and uncertainty, they don't feel comfortable at this time going to that level of pushing the sensitivity to interpret DNA profiles." Budowle, 12/10/13, 961. When the Court enquired what risk and uncertainty exists, he answered, "[w]ell, it's about getting an answer that you feel comfortable is representative of the DNA in the sample, making associations to people who may be considered potential source of the evidence." Id.

While the People elicited testimony from Dr. Caragine about international labs which engage in LCN typing for criminal case work, no details were provided. Dr. Haned testified that Santiago, Chile; the Netherlands, Norway, and perhaps France and Spain were using LCN DNA typing. However, the exact method employed is unknown. In order to show that use of LCN by other labs constitutes general acceptance it must be first be shown that the methodology is the same or fundamentally similar. If this were not the case, then use by a lab of a method different

from the one challenged in this case would have no bearing on the issue. See, e.g., McCluskey, supra, 954 F.Supp.2d at 1279.

The People claimed Australia was using LCN testing for criminal casework, but Dr. van Daal an Australian scientist, testified that to her knowledge that this is not true. Van Daal, 11/21/13, 324. Dr. van Daal testified about one case in Australia in which the prosecution attempted to introduce results produced by LCN DNA testing, but that testing was performed by the British Forensic Science Service, not an Australian lab. Ultimately, the DNA results were not used in court because the sample was contaminated with the lab director's DNA. Id. at 370.

Moreover, Dr. van Daal testified that while the New Zealand lab uses LCN testing, the methodology is unclear. In a New Zealand case, the laboratory attempted to both include and exclude the defendant through the use of a single locus because no results were produced at the other loci. This is an example of the type of use of LCN DNA testing being used abroad and the reason why the majority of laboratories in the world do not use LCN testing. Id. at 377. See generally People v. Ferguson, 526 N.E.2d 525, 531 (Ill. App.3d 1988) (overturning conviction where shoeprint evidence admitted even though witness had testified that individuals in other countries were doing same work that she did).

4. Unlike gold standard method, LCN is not in general use

Forensic laboratories will not use something for evidence in court unless they believe it is generally accepted as reliable. There is a need in the forensic DNA community to test low template DNA. Despite the demand, and the fact the methodology has been available for well over a decade, no laboratory in the United States besides the OCME uses it. This proves that the OCME's LCN DNA testing methodology is not generally accepted in the relevant scientific community.

The only explanation Dr. Caragine was able to provide as to why no other DNA labs in the United States use LCN DNA testing is that it is costly and requires a dedicated lab space. Dr. Caragine, responding to the Court's inquiry whether labs are not using LCN DNA testing because it is unreliable stated, "[m]y feeling from speaking with people at meetings, it is generally accepted and in the scientific community." [sic] Test. Direct Caragine, 12/13/12, 68. However, Drs. Van Daal and Coyle both testified that other labs are not using LCN testing because the results are cannot be interpreted reliably with the existing methodologies. See Van Daal, 11/21/13, 389 ("because of the associated interpretation issues with the profile"); Coyle, 11/8/13, 98 ("because increased artifacts and contamination "confounds interpretation significantly").

Forensic DNA labs are constantly being asked to provide DNA evidence in serious felony cases where touched items have low template DNA. This type of evidence includes clothing from burglaries and rapes, recovered weapons, and any items that are believed to have been handled by a perpetrator. Similarly, other crime scene samples that are not touched can contain low template levels of DNA. Therefore, the need is there to bring LCN testing online. If LCN testing was generally accepted it would have started to be implemented in other laboratories in the same way that conventional PCR STR forensic DNA testing came online. See People v. Seda, 139 Misc.2d 834, 847 (Sup.Ct. NY Co. 1988) (holding OCME's 4 in 1 system of blood testing not generally accepted and stating, "[w]hile general use in crime laboratories does not necessarily connote general acceptance in the scientific community, evidence of such limited use does persuade the court that the procedure has not been generally accepted by even the technical personnel whose standards may be less exacting than those of scientists").

The People attempt to defeat this simple logic with the claim that other labs do not have the resources that OCME has and thus cannot implement this type of testing. First, large public labs—the FBI, state laboratories—would have the resources to undertake the validation and implement this methodology if they believed the data was there to prove that it produces reliable results.

Second, even privates do not use OCME's LCN methodology. As Dr. Holland admitted on cross examination, Bode Technologies, the largest private DNA lab in the United States, does not use LCN DNA testing. That Bode does not do LCN testing is even more revealing because Dr. Holland—who came to court to support the admission of OCME's LCN methodology-- was a vice president there. A private lab makes money by catering to the needs of its customers and if private DNA labs believed in the reliability of the technology they would invest in it.

The People provided no credible evidence that other labs are not implementing LCN testing because of a lack of resources. Therefore, this lack of general use by other DNA labs is strong evidence that LCN testing is not generally accepted by the relevant scientific community. See e.g., Advanced Health Care, Inc., v. Guscott, 173 Wash.App. 857, 872 (Ct. of App. 2013); State v. Kunze, 97 Wash.App. 832, 853 (1999) (discussing general use as indicator of general acceptance).

6. LCN results are not uploadable into the federal DNA databank, NDIS

The Combined DNA Index System (CODIS) is the software and program of government DNA databases. The National DNA Index System (NDIS) is a part of the overarching CODIS system, and is the federal/national DNA databank. NDIS contains DNA profiles contributed by federal, state, and local participating forensic laboratories developed from both convicted offenders and crime scene samples.

There are rules about what kinds of crime scene samples can be uploaded to NDIS. NDIS rules are contained in a document called NDIS Operational Procedures Manual. Low copy number profiles cannot be uploaded to NDIS. The Manual states: "Low Template or Low Copy DNA analyses shall not be submitted to NDIS." §4.2.1.10 NDIS Operational Procedures Manual, January 2013, available at http://static.fbi.gov/docs/NDIS-Procedures-Manual-Final-1-31-2013-1.pdf. Relevant section Attached as Exhibit WHAT.

Low Template or Low copy DNA analyses are defined as "[b]ased upon a laboratory's internal validation, any DNA typing results generated from limited quantity and/or quality DNA template using conditions that have demonstrated increased stochastic effects…" §4.2.1.10.1, NDIS Manual, *supra*. As part of that definition, the NDIS manual lists the stochastic effects seen with those types of analyses as allelic drop in or drop out, increased stutter and increased intra locus peak height imbalance."

NDIS includes in their listing of the technologies additional amplification cycles, post amplification purification, reduced reaction volume, injection enhancement by increased voltage or time, and nested PCR. OCME employs three of these technologies as part of its low copy number DNA testing: additional cycles, reduced volume, and injection enhancement. Thus, OCME may not upload its low copy number DNA samples to NDIS. These federal guidelines preventing LCN DNA profiles from being uploaded demonstrates that the FBI does not believe that the results are reliable. Courts have found that acceptance by CODIS to be a relevant factor in determining general acceptance. See United States v. McCluskey, 954 F.Supp.2d 1224, 1260 (U.S.Dist. NM. 2013) (finding whether sample could be uploaded to NDIS critical to the issue of general acceptance).

Furthermore, SWGDAM has an Enhanced Detection Methods and Interpretation (EDMI) Committee which was tasked with "prepar[ing] guidelines for the validation and interpretation of enhanced detection methods," and review standards and the NDIS manual and make recommendations. See http://www.swgdam.org/committees.html. Last visited 4/11/2014. The chair of the EDMI committee is Eugene Lien, the Technical Leader of OCME. The committee was formed in 2009. The most recent NDIS manual was issued in January 2013. The committee has been in place for years, yet no guidelines have been written, and no changes to NDIS were made. Although Dr. Caragine testified about EDMI as a sign of approval of use by the forensic community, SWGDAM also has a committee on next generation sequencing which is not employed in casework because it is still in the Research and Validation stage. Id.

7. The People's reliance of use of increased sensitivity and replicate testing in other fields does not constitute evidence of general acceptance because the applications differ in substantial ways**.**

The use of a methodology in other contexts other than the one being challenged in court may be relied upon to establish general acceptance. However, the uses must be comparable. Here, the People presented testimony that the use of LCN-like methods in other contexts prove its reliability for use in court in criminal cases. Those claims ignored both scientific and policy differences between the usages. While usage in other contexts may be relevant in a Frye inquiry, because of scientific and policy differences between the contexts, the reliance here is misplaced. Whether a methodology is reliable relates to what it is and *how* it is used.

In Wilson v. State, 370 Md. 191 (Md. 2002), the Maryland Court of Appeals reversed a conviction for murder. The trial court admitted expert testimony in which the product rule was used to compute the statistical probability that two of defendant's children died of Sudden Infant Death Syndrome (SIDS). Clearly the product rule is a well-accepted statistical formula, but the

question in that case was the context in which it was used. The Court of Appeals found that "'the proper choice of statistical techniques is dependent on an underlying scientific phenomenon or principle,'" id. at 203, citing Armstead v. State, 342 Md. 38, 80 n.33. The appellate court reversed and held,

> …the trial court erred in admitting expert testimony based on the product rule because a condition necessary to the proper application of the product rule was lacking: there was inadequate proof of the independence of Brandi and Garrett's deaths. As evidenced by the authorities above cited, there is not general agreement in the scientific community as to the relationship between SIDS deaths within a single family. Stated another way, there is not general agreement in the medical community that multiple SIDS deaths in a single family are genetically unrelated. The literature continues to reflect a lively debate concerning the role of genetics in SIDS. Moreover, the recent study in the Journal of the American Medical Association suggests that there may well be a genetic component to SIDS. *[Citation omitted]* If there is any consensus in the field, it is that more research into the question is necessary before general acceptance is reached.

Dr. Caragine testified that LCN methods are used in preimplantation genetic diagnosis (PGD), missing persons work, and next generation sequencing in order to support a finding of general acceptance of usage for court use in criminal cases. The People's reliance on their claims about the use of low copy number DNA testing in these contexts is misplaced. The concerns which strip LCN testing of reliability in forensic DNA usage are not present in these other contexts.

Key aspects of these types of testing distinguish them from the LCN context in forensics. While PGD and cancer screening and single cell amplification are used with known sources of DNA, forensic testing by its nature is not. Additionally, while PGD, cancer screening and single cell amplification deal with single sources of DNA, forensic testing often does not, and it is never known a priori whether it will. Forensic LCN uses STR testing; PGD, cancer screening and single cell amplification do not. In fact, Dr. Caragine admitted that STR testing was not employed in any of the articles she cited to support the general acceptance of LCN. The Court

further questioned Dr. Caragine on this matter, and in response to the court's questioning she admitted that no profile was developed in these techniques as is done with STR testing. Caragine, 12/12/12, 26.

Dr. van Daal explained that PGD does not support the conclusion that LCN is generally accepted for criminal casework. PGD is the testing of a single cell of an embryo for genetic diseases. The region of DNA where a mutant gene is amplified. Van Daal 11/21/2013, 382-383. The technique used in PGD employs increased amplification cycles to boost the sensitivity. Yet, the use of the increased cycling in this context is completely different than in the LCN context. Furthermore, evidence profiles are degraded and cells in PGD are not. Additionally stutter does not occur in PGD or mutation screening, which as explained supra is a serious complication in interpreting LCN testing results.

Dr. Caragine's slide also referenced next generation sequencing technology, or "next gen" to support replicate testing in the LCN context. Dr. Caragine was misleading in testifying that, "replication is the cornerstone of science. Before I came into forensics, you always replicated your results to make sure it was accurate." Caragine, 12/12/12, 113. Replication in science is performing the same experiment multiple times; it is not what the OCME does, which is using a limited sample, that is not sufficient for one experience and dividing that sample into even smaller quantities. LCN by its nature is not reproducible. Furthermore, Dr. van Daal testified that the research that is being done on Next generation sequencing or massively parallel sequencing is not evidence that low copy number DNA testing is generally accepted because the connection between the methodologies is non existent. Van Daal, 11/21/13, 380-381.

Dr. Caragine testified that UNTHSC—the lab Dr. Budowle and Dr. Chakraborty are affiliated with-- uses the LCN methodology in the missing persons context, and that this is

evidence it is generally accepted. Several defense witnesses explained to the court that there are scientific and policy reasons why using LCN in the missing person context is different than for in court use. See Chakraborty, 12/16/13, 1088-1089.

First, with missing persons work involving a bone or a hair, an analyst can remove surface contaminants from the item before testing it. This is impossible with a crime scene sample since cleaning an item of evidence would remove the DNA. Second, because surface contaminants are cleaned off, a bone or hair will be expected to be a single source and thus the concerns with LCN and mixtures are absent. Chakraborty, 12/16/13, 1089.

Additionally, with mass disasters, you would likely be looking at what is known as a closed population, as opposed to an open population. A closed population is one for which you know the identity of the members, i.e. the passenger manifest for an airplane crash. This significantly limits the source of the DNA and relatives of the victims can provide DNA samples for comparison purposes.

Dr. Budowle testified that LCN could be used to generate investigative leads in some circumstances. Dr. Budowle explained that LCN can be used as an investigatory lead but not for evidence in court because, "I worry again on that because every time you do a low copy analysis there are certain amount of uncertainties, every single allele, that is a critical part to understand. You have to place a lot of uncertainty on it, but if I was using it as an investigative lead and I said that is all I am using it for, I would not want to use it in that situation." Budowle, 12/9/13, 810.

Similarly, Dr. Van Daal explained that "there clearly isn't the same requirement for reliability because an investigative leads is a piece of information that can be used with the limitations to narrow a suspect pooled[sic] but, obviously, its not an identification of somebody.

It's simply a way of potentially narrowing the suspect but it should [not] be the only thing used to do that." Van Daal, 11/21/13, 385-386. In fact, Dr. van Daal compared the use of LCN as an investigative lead to the use of polygraphs. Id., at 391.

The United States Supreme Court has recognized that a methodology may be fit for use in some out-of-court contexts and that limited use does not render it admissible in a trial, where stricter standards are meant to exclude unreliable evidence. In United States v. Scheffer, 523 US 303, 312 (1998) the Supreme Court held that a per se rule barring polygraphs did not violate the defendant's Fifth or Sixth Amendment rights. The Court stated:

> "Respondent argues that because the Government-and in particular the Department of Defense-routinely uses polygraph testing, the Government must consider polygraphs reliable. Governmental use of polygraph tests, however, is primarily in the field of personnel screening, and to a lesser extent as a tool in criminal and intelligence investigations, but not as evidence at trials. See Brief for United States 34, n. 17; Barland, The Polygraph Test in the USA and Elsewhere, in The Polygraph Test 76 (A. Gale ed.1988). Such limited, out of court uses of polygraph techniques obviously differ in character from, and carry less severe consequences than, the use of polygraphs as evidence in a criminal trial. They do not establish the reliability of polygraphs as trial evidence, and they do not invalidate reliability as a valid concern supporting Rule 707's categorical ban."

Id. at 312 (emphasis added). The use of LCN DNA testing for investigative purposes is similar to the use of the polygraph, which is not generally accepted nor admissible in court.

8. Previous admission in other courts does not establish general acceptance absent a showing of a thorough vetting of the issues

Dr. Caragine testified that at the time of her testimony the results of OCME's LCN testing had been admitted at trial 143 times. Caragine 12/13/12, 75. Although the Court overruled defense counsel's objection, the Court noted that these may not have been cases in which the methodology was challenged. Unless the methodology was specifically challenged and the court made a determination of admissibility based on general acceptance and reliability,

the mere fact of admission is meaningless: "'[u]nless the question of general acceptance has been thoroughly and thoughtfully litigated in the previously case, …reliance on judicial practice is a hollow ritual." [citations omitted]. People v. Canulli, 341 Il..App.3d 361, 369-70 (Ill. App. 2003). When produced from a novel methodology is being offered a judge must consider whether that methodology is generally accepted, as here. See Test. Dir. Chakraborty, p.1096, l. 1-3 (noting that the New York DNA Subcommittee reviews methodologies because they are new).

9. OCME's claims about the sufficiency of peer review are misleading

The OCME's claim that LCN has been peer reviewed and as such, is generally accepted is misleading at best. The claim that OCME publication of a peer reviewed article on the LCN methodology establishes general acceptance ignores the nature of peer review and the function of publication of scientific articles.

Dr. Caragine testified that when an article is published in a peer reviewed journal, "[i]t means your results have been accepted by …your peers, by the scientific community." Caragine, 12/12/12, 30, the publication of OCME's LCN paper in a peer reviewed journal—the Croatian Medical Journal—established general acceptance. When asked whether the acceptance of a paper for publication in a peer reviewed journal was the end of peer review, Dr. Coyle explained "…it's actually the beginning of the process. Submission to a journal just means that you're presenting your data, now it's open for reading by all scientists in the community and then there's further discussion, further experimentation based on the content of that paper." Coyle, 11/8/1333. As such, publication in a peer review journal is necessary for general acceptance but not sufficient to demonstrate general acceptance.

Furthermore data critical to a more comprehensive review of OCME's LCN validation was missing from OCME's articles on their validations. Dr. van Daal testified that OCME never investigated the extraordinary level of noncontributor alleles seen in their LCN and FST validations and never included this data in their publications. Van Daal, 11/21/13, 345.

This position is supported by the fact that other methodologies have enjoyed publication in a peer reviewed journal but were subsequently determined to be unreliable. For example, comparative bullet lead analysis (CBLA) as example of methodology that was published in peer reviewed publications. It was later determined that the methodology was fundamentally flawed and convictions based on the evidence have been overturned. Coyle, 11/8/13, 33. Indeed, courts relied on publication to determine that this evidence was admissible. See e.g., Ragland v. Commonwealth, No. 2002-SC-0388-MR, 2004 WL 2623926 at *12 (Ky. Nov. 18, 2004) (admitting CBLA evidence and finding it scientifically reliable, in part because the expert "testified that the analysis has been subjected to peer review in a number of scientific journals.) This court later reversed this decision and found the evidence inadmissible after the National Research Council published an extensive report criticizing CBLA. See National Research Council, National Academy of Science, Forensic Analysis: Weighing Bullet Lead Evidence (2004).

Furthermore, although developmental validation are supposed to be published in a peer reviewed publication, OCME did not publish their validation study paper until 2009—at least three years after the method was brought online for case work. The paper was published during the ongoing Megnath Frye hearing. As discussed supra, this publication did not include key emplirical data—specifically, the levels of contamination that occurred. As Professor Paul Giannelli explained "[s]everal lessons can be gleaned from the CBLA experience. First, the

failure to publish the empirical data that supports scientific conclusions is unacceptable. Scientists 'are generally expected to exchange research data as well as unique research materials are essential to the replication or extension of reported findings.'" Giannelli, Paul C., Comparative Bullet Lead Analysis: A Retrospective, 47 Crim. L. Bull. 306, 316 (2011) (citing National Research Council, National Academies of Sciences, Responsible Science 11 (1992). See also National Research Council, National Academies of Sciences, Sharing Publication-Related Data and Materials: Responsibilities of Authorship in Life Sciences 4 (2003) (advocating a "uniform principle for sharing integral data and materials expeditiously" or UPSIDE).

10. Passing an audit does not demonstrate general acceptance.

The People further sought to rely on claims by Dr. Caragine that passing audits showed that LCN was generally accepted in the scientific community. This claim is not true. Both Drs. Coyle and van Daal testified that an auditor in passing a lab for accreditation purposes, is not making a statement that the methodology is generally accepted in the forensic community—or even that it is reliable and robust. An auditor is only checking that a lab has the appropriate documentation and procedures in place. Quite literally, an auditor works off a checklist. Coyle, 11/8/13, 36; van Daal, 11/21/13, 378.

Dr. van Daal, who helped develop forensic DNA accreditation standards with the National Association of Testing Authorities (NATA), the oldest forensic DNA accreditation body in the world, and a precursor to ASCLAD-LAB. Dr. van Daal is a qualified inspector for both NATA and ASCLAD-LAB, and has audited the FBI lab as well as Australian labs. Test. Van Daal, 11/21/13, 298. While the OCME relies on having passed audits to demonstrate that their LCN methodology is generally accepted, Dr. van Daal testified that,

> "Both auditing and accreditation are process-oriented events, whereby, the laboratories are checked for a number of things. For example, a check will be made that on the physical nature of the laboratory, is it secure? Does it have the right kind of separate facilities for certain operations? Are things securely stored? Is evidence securely stored? Are processes in place for all sorts of things that there need to be? So are there processes for the receipt of evidence for analysis of evidence, examination of evidence? Have validation studies been done to support the testing that's been done? So it's a range of things but its about checking the processes and the quality system of the organization."

Van Daal, 11/21/2013, 378. Dr. van Daal also explained that when an auditor looks at a validation study, the auditor is "looking to see that the validation has been done. So if I use an example, currently an auditor would look to see against the guidelines for the validation, so in DNA, for example, that's mixture studies and sensitivity and all the things that are required to have been done." Id. at 379. The testimony of Drs. Van Daal and Coyle indicate that auditing is about the processes not about the reliability of the science.

Furthermore, the ASCLAD-LAB's Accreditation and auditing practices, which Dr. Caragine relied on in her testimony and powerpoint presentation, have come under fire. See, e.g., Mandy Locke and Joseph Neff, "Inspectors Missed all SBI Faults," The News and Observer (Raleigh), 8/26/10, *available at* http://www.newsobserver.com/2010/08/26/648075/inspectors-missed-all-sbi-faults.html, last visited 4/12/2014 ("Because audits are conducted by peers in the forensic community some fear that there's an expectation to be gentle or pay for it when your lab is examined.").

11. The New York DNA Subcommittee's approval does not demonstrate general acceptance.

The People placed particular emphasis on approval by the New York DNA Subcommittee as evidence of general acceptance. While this is a factor this Court may consider, evidence at this hearing demonstrates that the review was insufficient and that the Subcommittee

did not possess relevant and material information that would have influenced the DNA Subcommittee vote.

The defense called a member of the DNA Subcommittee, Dr. Chakraborty. Dr. Chakraborty was invited to, and served on the NY State DNA Subcommittee for over 15 years. Chakraborty, 12/16/13, 1094. The DNA Subcommittee is a body which reviews lab methodologies and issues recommendations to the New York Forensic Commission pursuant to New York Executive Law 49(b) §995. The DNA Subcommittee voted to approve the OCME's LCN DNA testing. Their binding recommendation was then submitted to the New York Forensic Commission, which voted to approve LCN in December 2005.

While a sitting member of the DNA Subcommittee, Dr. Chakraborty voted to approve the use of LCN for forensic casework. However, he testified in the Frye hearing that had he known at the time of the vote what he later learned about the data, he would not have voted to approve it:

> Q: Doctor, in sum, is OCME's low copy number DNA testing methodology generally accepted in the scientific community as reliable for use in court in criminal cases?
> A: My testimony now would be no.
> Q: If you voted as a member of the DNA subcommittee whether to approve OCME's LCN methodology, what would your vote be today?
> [objection overruled]
> A: With the data we have in hand I would have voted no.

Chakraborty, 12/16/13, 1150.

Dr. Chakraborty described for the court the workings of the DNA subcommittee. As the only non-OCME witness called to testify at this hearing, he described the workings of the subcommittee. He testified that during his time on the DNA subcommittee there was not one

instance that he could recall in which the subcommittee did not issue a binding recommendation to the Forensic Commission to approve a new methodology. Chakraborty, 12/16/13, 1096-1097.

Dr. Chakraborty testified that a large amount of data was submitted and there was not enough time to review it. This complaint was of course was heard in the DNA Subcommittee because Dr. Anne Walsh raised this concern during the initial meeting of the Subcommittee at which OCME presented on LCN. She refused to vote because she didn't have enough time to review the material. See People's 33, 34. Moreover, discussion would inevitably gravitate to certain topics, and other topics would be neglected over the short time allotted to discuss novel methodologies: "Our reviews were not exactly as total as would have been needed if that was a particular research portion of my interest and I am part of that so, with that limitation I would say that the review of such issues are not necessarily completely exhaustive." Chakraborty, 12/16/13, 1102.

As a scientist who was invited and sat on the DNA subcommittee for over 15 years, Dr. Chakraborty stated that a vote of approval by the DNA subcommittee does not mean that the methodology is generally accepted in the larger scientific community as reliable. See generally, Chakraborty 12/16/2013, 1097-98 (generally discussing why the DNA Subcommittee approval does not indicate general acceptance by the relevant forensic scientific community).

Furthermore, Dr. Chakraborty learned information after he voted to approve LCN testing by OCME which caused him to be concerned about the OCME's LCN methodology. The OCME's LCN validation article was published several years after the presentation to the DNA Subcommittee and the bringing on line of the methodology. Similarly, numerous other scientific articles have been published that explore new ways to approach low template testing as the field is in the exploratory stage. Dr. Chakraborty read the OCME's LCN validation and FST articles

and the continuing research published on those topics, and concluded LCN testing is not yet reliable and robust. Chakraborty, 12/16/2013, 1108. He was concerned that OCME's validation studies did not adequately address the uncertainty factors and their interrelatedness. Id. p. 1103.

12. Conclusion.

The experts called by the defense testified that the great uncertainty in LCN profiles, particularly mixtures, has not been adequately accounted for in the testing and the statistical weighting of the results. Both Drs. Budowle and van Daal conveyed optimism that techniques will be developed in the future for robust typing of low template profiling, which includes the appropriate statistical weight to be put on inclusionary results. As Dr. Chakraborty stated however, low copy number DNA testing is not reliable for use in court in criminal cases "[b]ecause we have not dealt with solutions of all of these uncertainties in a way that makes the, ultimately, protocol reliable, robust and generally accepted." Chakraborty, 12/16/2013, 1090.

The forensic science community is developing methods to deal with the problems of low template analysis reliably. The controversy in the field and the lack of labs using the methodology OCME points to the lack of a solution. The time has not come yet for LCN typing to be admitted into court as evidence in a criminal case.

VI. CONCLUSION

The People have failed to establish that both OCME's LCN and FST methodologies are generally accepted as reliable in the relevant scientific community under Frye v. United States, 293 F. 1013 (D.C. Cir. 1923). Accordingly, this Court must grant the defendants' motion to preclude the LCN and FST evidence.

Dated: April 14, 2014
New York, New York

Respectfully submitted,

_____

Jessica Goldthwaite, Esq.
The Legal Aid Society, DNA Unit
199 Water Street
New York, NY 10038
(212) 577-3444
Susan Friedman, Esq.
Clinton Hughes, Esq.