The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title:	**Low-Template DNA Mixture Interpretation: Determining the Number of Contributors**
Author(s):	**Catherine Grgicak, Ph.D.**
Document No.:	**249157**
Date Received:	**September 2015**
Award Number:	**2011-DN-BX-K558**

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this federally funded grant report available electronically.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Boston University School of Medicine
Biomedical Forensic Sciences

72 E Concord St. Rm R806B
Boston, Massachusetts 02118
T 617-638-1950 F 617-638-1960
www.bumc.bu.edu/gms/biomedforensic

Low-Template DNA Mixture Interpretation: Determining the Number of Contributors

FINAL TECHNICAL REPORT

August 25, 2014

Department of Justice, National Institute of Justice
Award Number: 2011-DN-BX-K558
(January 1, 2011 – August 31, 2014)

Principal Investigator:
Catherine Grgicak, PhD
Assistant Professor
Program in Biomedical Forensic Sciences
Boston University School of Medicine
Boston, MA 02118
Phone: (617) 638-1968
Fax: (617) 638-1960
e-mail: cgrgicak@bu.edu

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

ABSTRACT

A biological sample obtained from a crime scene might be a mixture containing DNA from two or more individuals. An assumption regarding the number of contributors to a sample is needed to compare the crime scene profile with that of a known if the Likelihood Ratio is the statistic utilized to convey the 'weight of the evidence'. Usually, the number of contributors to a question sample is unknown and is specified by the analyst based on the electropherogram obtained. This can be challenging in the case of complex, low template samples that exhibit allele sharing and often contain artifacts like dropout and stutter.

NOCIt – a computational tool that calculates the probability distribution for the number of contributors to a DNA sample is presented. Unlike existing methods, which operate on the number of peaks in the signal and/or the rarity of the alleles, NOCIt uses both the rarity of the alleles and the quantitative data in the signal, i.e. the heights of the peaks. NOCIt was calibrated using single source samples amplified from various low level DNA amounts (0.007 – 0.25 ng) and three different times of injection (5, 10 and 20s). The peak heights (peak height ratios in the case of stutter peaks) were modeled using the Gaussian distribution and appropriate parameters (mean and variance) were obtained from the calibration data. Dropout rates were also computed at each DNA mass.

To test the performance of NOCIt, 1-, 2-, 3-, 4- and 5- person samples were created and interpreted with NOCIt. The number of contributors with the highest probability was taken as the answer supported by NOCIt. In addition, the performance of NOCIt was compared with the Maximum Allele Count (MAC) and the Maximum Likelihood Estimator (MLE) [1] methods. MAC uses the number of peaks at a locus to determine the minimum number of contributors while MLE uses the number of peaks as well as the allele frequencies. The performance of NOCIt was consistently better than MAC and MLE across all DNA amounts and all times of injection for the 1-, 2- and 3- person samples. Both NOCIt and MLE (20 sec injection) showed improved performance over the allele counting method for the 4- and 5- person samples. Since the MAC and MLE methods utilize analytical and stutter thresholds, the accuracy of MLE and MAC were found to be dependent on injection time. When samples contained at least one contributor with < 2 cells worth of DNA (i.e., 12.6 pg), the accuracy of all methods decreased due to allele dropout and allele sharing. Further, the difference between the calculated and actual NOC was larger for these samples. For example, the percentage of samples where N_{Calc}-N_{actual} $\leq$ -2 was 55%, 55% and 44% for the MAC, MLE and NOCIt methods, respectively. However, in cases where NOCIt failed to pick the correct number of contributors, it was able identify the region in which the NOC is most likely to lie. NOCIt has been implemented using the Java programming language and is currently available on www.bu.edu/dnamixtures.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Contents

Executive summary....4
Main Body of the Final Technical Report....7
Introduction....7
Statement of the Problem....7
Literature Review....8
Rationale for the Research....9
Methods....12
Results....18
Conclusion....26
References....27
Dissemination of Research Findings....29
Tables....30
Figures....38

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Executive summary

Interpreting an STR profile obtained from an evidentiary sample can be complicated because the number of contributors to the sample is unknown in most cases. While it is straightforward to identify the number of contributors in the case of high template samples containing 1 or 2 contributors, the interpretation becomes challenging when complex, low template mixtures are evaluated. The signal from low template samples is difficult to interpret because of allele dropout. Moreover, the stutter ratio becomes elevated in low template samples, making it hard to distinguish between stutter peaks and allelic peaks. In response to the aforementioned issues, a probabilistic procedure and corresponding software tool to infer the number of contributors in a forensic DNA sample was developed. The procedure and software tool is called NOCIt.

Forensic DNA mixture interpretation of short tandem repeats (STRs) typically uses the following procedure: 1) an analytical threshold is applied to distinguish peak from noise, 2) the number of contributors is assessed, 3) the genotypes from the item of evidence are de-convoluted, 4) the inferred genotype obtained from the evidentiary profile is compared to the genotype from a known, standard or suspect, and lastly 5) a conclusion and 'match-statistic' is reported.

The number of contributors (NOC) is an important assumption in the DNA interpretation process since it has a direct impact on the way in which the evidentiary profile is de-convoluted. Further, it may be necessary to determine the most likely number of contributors for investigative purposes. Two approaches to determine the number of contributors are currently employed. The common method is the Maximum Allele Count (MAC) technique and the second method is the Maximum Likelihood Estimator (MLE) technique. Both of these methods rely upon the application of an analytical threshold and stutter filter to the signal. All peaks below the AT are considered indistinguishable from noise while all peaks above the AT are considered allelic peaks and included in the interpretation. Application of an AT to a low template profile is not optimal because of the increased risk of labeling a noise peak as an allelic peak or an allelic peak as a noise peak. Furthermore, an allelic peak could be masked by stutter. Thus, any peak in the stutter position could be a stutter peak, an allelic peak or a combination of the two. A stutter-ratio threshold is used by both MAC and MLE. Since the height of a stutter peak is dependent upon the height of the parent peak, a stutter-ratio is calculated and any peak in the stutter position with a ratio less than the stutter-ratio threshold is removed.

MAC is a widely used method and gives the minimum number of contributors that explain the signal. MAC works by counting the number of alleles that are visible above the AT at all loci. The maximum value is divided by two and rounded up to give the minimum number of contributors that explain the profile.

MLE was more recently developed and utilizes more information than MAC[1] and is currently available on [2]. In addition to using the number of peaks observed, MLE uses the background frequencies of the alleles above the AT to calculate the likelihood a locus contains alleles from a certain number of contributors. Based on the profile obtained, all possible genotypes for *n* individuals are considered. The likelihoods at all the loci are multiplied with each other to obtain the overall likelihood for the sample given *n* contributors. The number of individuals that results in the highest likelihood is taken to be the most likely number of contributors in the sample. Though MLE utilizes the qualitative data obtained in the signal (i.e., the allele frequencies), the quantitative data (i.e., the heights of the peaks) is not considered.

The peak heights in the signal are directly proportional to the amount of DNA originally present in the sample. They are also a good indicator of the mixture ratio that gave rise to the sample. The

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

background noise may be dependent upon the amount of DNA originally present in the sample. Average stutter ratios and allele dropout are also related to the amount of sample. This information could be used while calculating the number of contributors.

NOCIt works upon the entire signal obtained – the background frequencies of the alleles observed as well as the peak heights - while calculating the number of contributors. To calibrate the software, 1555 single source samples from 68 donors with known genotypes were created. These samples were generated using the AmpF*l*str® Identifiler® Plus kit (Life Technologies, Foster City, California). During creation of the sample profiles, 3 times of injection (5, 10 and 20s) were used during electrophoresis. The injection time is typically increased in the case of low template samples to increase the signal-noise ratio and hence 3 different sets of data, each at a different injection time, were used to study the change in signal with injection time. At each injection time, samples were amplified from 7 template DNA amounts (0.008 – 0.25ng) to analyze how the DNA amount impacted the signal obtained. In the profiles obtained, the peaks were separated into 1 of 3 categories: True peaks (all peaks representing the alleles in the sample), Stutter peaks (all peaks in the stutter position of True peaks) and Noise peaks (all other peaks in the signal). Only reverse stutter was considered. The heights of the peaks were modeled using the Gaussian distribution. Calibration parameters (namely the mean and the variance) for the 3 categories of peaks were computed for each DNA amount at the 3 injection times for every locus. Dropout rates were also computed at each DNA amount at the 3 injection times for every locus.

Models to estimate the rise in baseline noise and allele heights were developed. Both the baseline noise and allele peak heights increased linearly with target. Although the linear model was chosen for baseline noise in this study, further valuation into the model is required since it was observed that the linear growth occurred after a certain mass of DNA was added to the PCR, i.e., > 0.125 ng. The exponential function was found to be a good approximation of 1) the frequency of dropout and 2) the average stutter ratios at a locus. Gaussian distributions were assumed, and the approximation was tested for baseline noise and peak height. Results suggest the Gaussian distribution describes the variability it the peak height well. Baseline noise is better described using a log-normal distribution, and studies that assess the impact of this finding on determining the NOC are warranted. Stutter ratios were calculated for the entire locus and a Gaussian distribution was assumed.

A Monte Carlo approach is used by NOCIt to compute the likelihood for the number of contributors. In every iteration of the Monte Carlo process, genotypes for the *n* contributors are chosen based on the frequencies of the alleles in the frequency table. Only alleles present in the frequency table are sampled. A mixture ratio is chosen at random - all mixture ratios are assumed to occur with equal probability. To obtain parameters corresponding to DNA amounts that are not among those used for calibration, modeling of the dropout frequencies, means and variances of true peak heights, means and variances of the noise heights, and the means and variances of stutter ratios is carried out at each locus using the following: exponentially decreasing curve ($Pr(D) = ae^{-bx}$) was used to model allele dropout, an exponentially decreasing curve ($\overline{SR} = ae^{-bx}+c$) was used to model the average stutter ratio (SR), a straight line with a positive slope ($\overline{H_{allele}} = mx$) was used to model the peak height and a straight line with positive slope ($\overline{H_{noise}} = mx+b$) was used to model noise. For every allele in the genotype of the contributors, dropout of the allele is simulated by a Bernoulli trial. Two assumptions are made with regard to dropout: a) Dropout of one allele of a contributor is independent of dropout of the contributor's other allele and b) Dropout of an allele from a contributor is independent of dropout of the same allele from another contributor. Based on the evidence observed, the likelihood of observing the heights of the peaks given the genotypes

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

of the contributors, the mixture ratio, the amount of DNA amplified and the time of injection is computed using the calibration data. This is repeated a number of times. The average of the values computed is the likelihood of observing the evidence at a locus, given *n* contributors. Since the loci are assumed to be independent of each other, the likelihood values at all the loci are multiplied with each other to give the overall likelihood for *n*. The *n* that results in the highest likelihood is taken to be the number of contributors most supported by the evidence as calculated by NOCIt.

The performance of NOCIt was tested on 1-, 2-, 3-, 4- and 5- person mixtures. Like the calibration data, these samples were also generated using 3 injection times. Samples were amplified using 7 DNA amounts. The performance of NOCIt was compared with the MAC and the MLE methods by running those methods on the same samples. As previously described, MAC and MLE both need the setting of an AT to calculate the number of contributors. For comparison purposes, 2 types of thresholds for MAC and MLE were evaluated. The first threshold was a constant threshold of 50 RFU at all the loci, which is a commonly used threshold in forensic laboratories. The second threshold is a variable threshold. At each injection time, this threshold varies with DNA amount and dye color. This threshold was set by picking the height of the highest noise peak observed in the calibration data corresponding to a DNA amount, dye color and time of injection and setting that height as the threshold for that DNA amount, dye color and time of injection. NOCIt does not depend upon the setting of a threshold and works on the entire electropherogram obtained. Application of MAC and MLE also uses a stutter threshold to filter out the peaks in the stutter position of allelic peaks. The stutter filter specified by the AmpF*l*str® Identifiler® Plus manual was used to filter the stutter peaks at each locus. Allele frequencies from the Caucasian population specified in the AmpF*l*str® Identifiler® Plus manual were used for NOCIt and MLE.

The performance of MAC and MLE, using both thresholds, increased with an increase in injection time. NOCIt outperformed both methods at all 3 times of injection for the 1-, 2- and 3-person samples. Changing the injection time did not have a significant effect on the performance of NOCIt. Both NOCIt and MLE resulted in similar accuracy rates for the 10 and 20 second injections of the 4- and 5- person mixture samples. NOCIt outperformed MAC and MLE for the 5 second injections of the 4- and 5- person mixture samples. MLE was affected by injection time, while NOCIt was not. Overall, the accuracy of all the 3 methods increased with an increase in DNA mass.

NOCIt can compute the likelihood a forensic sample has up to 5 contributors in approximately 9 hours on a regular PC with an Intel quad core processor. Fewer numbers of contributors can be evaluated – as specified by the user – if desired. The output of NOCIt is presented as the probability distribution over 0 to 5 contributors thereby giving the user information regarding, not only the most likely number of contributors, but the uncertainty associated with the measurement.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Main Body of the Final Technical Report

Harish Swaminathan[1], Sarah Norsworthy[2], Ullrich Mönich[3], Genevieve Wellner[2], Kayleigh Rowan[2], Desmond S. Lun[1], Muriel Médard[3] and Catherine Grgicak[2]*

[1]Rutgers University, Center for Computational and Integrative Biology, Camden, NJ, 08102;
[2]Boston University School of Medicine, Program in Biomedical Forensic Sciences, 72 East Concord Street, Boston, MA 02118;
[3]Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 77 Massachusetts Avenue, Cambridge, MA 02139;
**Corresponding Author*

Introduction

Statement of the Problem

Many biological samples deposited at and collected from crime scenes contain mixtures from two or more individuals. The elucidation of individual donors in mixed biological samples has traditionally been a problem with serological testing and remains one today. Even with experienced analysts, complex mixtures are not interpreted in many laboratories.

There are two general approaches to interpreting DNA profiles. One is the CPE/CPI (Combined Probability of Exclusion/Inclusion) method, which is a binary scheme where alleles are considered either present or absent and the resultant genotype match information is derived from all possible allele pairs. This approach requires a stochastic threshold (ST) be defined to avoid interpretation of products which are too low and prone to stochastic effects. The apparent benefit of this method is that it makes no assumptions with regard to the number of contributors. However, this method cannot always be accurately applied to complex mixtures. Although the number of contributors theoretically has no bearing on whether CPI could be used as the statistic, it does have a significant effect on the assumption regarding the fact that all alleles have been detected. That is, as the number of contributors' increases, the ability to confidently state all alleles were detected is lost. The inability to confidently conclude that all alleles are observed is mainly due to allele sharing and the difficulty associated with mixture de-convolution when greater than 2 contributors are present. This is exacerbated when amplifying low amounts of DNA. An example of such a case is shown in Figure 1, which displays the electropherogram of a three person mixture amplified using the AmpFlstr® Identifiler® Plus Amplification Kit and 125 pg of DNA. The AT is set to 30 RFU and the ST is 150 RFU. In this example, there are 5 peaks at D18S51 and because the peak heights for alleles 16, 18 and 20 are below the ST of 150, the CPI would not be calculated for this locus. However, at TPOX all alleles are above ST and it would therefore appear that TPOX could be used to calculate the CPI. However, if the minimum number of contributors is 3 then there is a substantial amount of 'allele stacking' at TPOX and it no longer becomes possible to ascertain whether all alleles from every contributor is detected. As a result, uncertainty at that locus precludes the use of CPI at TPOX. Since CPI is to be used in instances where all alleles are detected, it is evident that CPI can only be applied when there is little or no ambiguity in the profile. As a result, CPI has limited applicability to complex, low-template samples where there is a significant level of uncertainty associated with the profile.

Another approach to obtaining match information is the utilization of the likelihood ratio (LR) method which can readily incorporate information which is not utilized in CPI. The LR test is conducted by computing the ratio

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

$$LR = \frac{\Pr(E|H_P)}{\Pr(E|H_D)} \qquad \text{(Equation 1)}$$

Where H_P represents the prosecution's hypothesis, H_D the hypothesis of the defense and E represents the evidentiary DNA profile.

There are a number of state-of-the-art methods, algorithms and software packages that report the strength of a match in terms of the LR.[3-10] Though the methods differ in the algorithmic processes and the ways in which they model stutter[10, 11], allele dropout[12, 13] and/or baseline noise[14], they all require an assumption regarding the number of contributors. Since the use of any formula for mixture interpretation should be applied in cases where the assumptions are reasonable, assigning the correct NOC to an evidentiary profile may arguably be a very important step in DNA mixture interpretation pipeline.

Literature Review

Currently, the common approach to determine the NOC is to utilize the MAC (Maximum Allele Count) method. This is a 'discrete' method which counts the number of alleles above the analytical threshold (AT). Typically, the number of obligate alleles above the AT at a locus is divided by two and rounded up. The maximum value across the loci is taken to be the minimum number of contributors that gave rise to the evidentiary DNA profile. This minimum NOC is then used as the NOC to determine the LR.

A number of issues with this approach exist. First, this method does not work well with complex mixtures because of allele sharing between contributors. That is, it does not take into account the frequency of the alleles and the propensity for multiple contributors to possess common alleles. Therefore, as the number of actual contributors increases the probability that the actual NOC equals the minimum NOC decreases. For example, in simulation studies using the SGM+™ multiplex loci, it was shown that ~66% of four-person mixtures would present six or fewer alleles at all loci. As a result, these complex mixture samples would likely be incorrectly interpreted as a mixture from three or fewer people.[15] Similarly, Paoletti et al. also showed that ~76-77% of the simulated four-person mixtures tested would not have been recognized as four-person mixtures based on the maximum allele count as they had only five or six alleles (or fewer) at all 13 CODIS loci.[16] Based on the aforementioned studies, it is unlikely that the minimum number of contributors for the majority of mixtures containing DNA from three or more people will accurately reflect the actual or true number of contributors. In fact, the minimum calculation could easily be off by one or two contributors.

Another allele counting method, similar to the MAC method, was published by Perez et al.[4] In this work the total number of alleles above the AT were counted and guidelines for estimating the number of contributors for high template and low template samples were established. Like MAC, this method is prone to misclassification due to the fact that it does not take into account allele sharing, stutter and dropout.

Methods that do not solely rely upon the number of alleles to determine NOC exist. For example, an approach developed by Biedermann et al. [17] employed a Bayesian network to infer the number of contributors to forensic samples. This method was shown to work better than MAC with degraded DNA and with higher numbers of contributors. Haned et al. [1] extended the work of Egeland et al. [18] to develop a Maximum Likelihood Estimator (MLE) for the number of contributors, taking into account population substructure. This method was also shown to give more accurate results than MAC with higher number of contributors and degraded DNA. A

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Probabilistic Mixture Model was used by Paoletti et al. [16] to infer the number of contributors to a sample based on the frequencies of the alleles observed. This method, like MLE, accounts for correction due to allele overlap.

These methods, though they use the qualitative data available, i.e. the frequencies of the alleles observed, do not use the quantitative data obtained, i.e. the heights of the peaks in the signal. When peak heights are not considered a plethora of interfering signal that impacts the ability to deduce the actual NOC is also not examined. This renders both the counting and frequency based methods ineffectual for low-template, complex forensic mixture interpretation.

Rationale for the Research

The first impediment to accurately inferring the number of contributors to complex mixtures originates from the chance that an allele may not have been detected during testing. To illustrate the effect of allele dropout on the ability to infer the NOC, consider the allelic peaks obtained from a typical DNA electropherogram presented in Table 1. The probabilities that one random individual gave rise to the observed peaks in positions 13 and 16 at the D8S1179 locus can be computed by examining the frequencies of the alleles in question and the probability of dropout (Pr(D)). If one person gave rise to the stain, then there is only one way that a single person could have resulted in a D8S1179 signal at allele positions 13 and 16: That person must have genotype G_1 = 13,16, and no dropout could have occurred. Therefore, (population substructure is not considered here for simplicity of exposition) the likelihood that the evidence at D8S1179 originated from one contributor is

$$L(E|n=1) = 2f_{13}f_{16}\left(1 - Pr(D)\right)^2. \qquad \text{(Equation 2)}$$

Here, n is taken to be the NOC, f_{13} is the frequency of observing the allele 13 within the population, f_{16} is the frequency of observing allele 16 and Pr(D) is the probability of allele dropout. To calculate the possibility that two unrelated individuals gave rise to the signal at positions 13 and 16, three scenarios are considered. First, the profile could have originated from two random individuals, where neither of the individuals' alleles dropped out (Rows 1-3 Table 2). The second scenario is that two random individuals' DNA gave rise to the peaks, but one allele dropped out (Rows 4-8, Table 2). The last scenario is that two random individuals gave rise to the signal, but two alleles dropped out (Rows 9-11, Table 2). If dropout is considered - and the frequency of the other, not-observed, allele is taken to be 1 less the frequency of the observed alleles - then the chance that two random individuals gave rise to the alleles at D8S1179 is,

$$L(E|n=2) = [(4f_{13}f_{16}^3+6f_{13}^2f_{16}^2+4f_{13}^3f_{16})\cdot(1-Pr(D))^4] + [4Pr(D)\cdot(1-f_{13}-f_{16})\,(3f_{13}f_{16}^2+3f_{13}^2f_{16})\cdot(1-Pr(D))^3] + [6Pr(D)^2\cdot(1-f_{13}-f_{16})^2\,(2f_{13}f_{16})\cdot(1-Pr(D))^2] \qquad \text{(Equation 3)}$$

A similar approach is used to calculate the $L(E|n=3)$, $L(E|n=4)$, etc. Therefore, Equations 2 and 3 show that when Pr(D) is considered it can have a significant effect on the ability to accurately assess the NOC. Figure 2 shows the APP (*a posteriori* probability) that 1 versus 2 versus 3 versus 4 contributors gave rise to the stain when the information from all loci is combined. Therefore, for this profile, when $Pr(D) = 0.4$, there is a probability of 0.04 that 2 random individuals gave rise to a profile that seems to have originated from only one contributor. This is a direct result of the loss of information associated with a high dropout rate and includes the possibility that the entire locus of one person may not be detected. This example demonstrates that counting the number of

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

peaks observed may not the optimal technique to determine the actual number of contributors to a low level stain.

One way to minimize the loss of allele information is by decreasing the AT. However, ATs which are too low may lead to false detections, while ATs which are too high lead to high levels of false non-detects [19]. For example, work performed by Bregu et al. [20] showed that baseline noise increases with target for some kits, suggesting that noise may need to be modeled based on target mass or other similar parameter. This would be the optimal method for examining allele signal in the presence of noise and ensures all data in the signal is analyzed. Further, by examining evidentiary profiles in this way, the Pr(D) is kept to a minimum for a given laboratory process. If the Pr(D) is minimized, then the soundest assessment of the NOC can be performed.

Determining the NOC may also be complicated by the fact that the ratio of contributors may be indeterminable and, like the LR ratio, contributor ratio assessments are dependent on the assumption on the NOC. Table 3 shows a comparison of the likelihoods that the NOC was 1 versus 2 people when different ratio combinations are considered for a sample amplified with a total of 1, 0.6, 0.2 and 0.1 ng of DNA. Table 3 shows that, not only does the likelihood that 2 persons contributed to the stain increase as template levels decrease, but the L(E|n=2) for a 1:9 2-person mixture increases at a faster rate. This exemplifies the need to evaluate numerous mixture ratios such that all scenarios are taken in to account during interpretation.

Signal interference from stutter also inhibits accurate interpretation of complex, low-level mixtures. Current practice dictates that when analyzing single source samples, if the %stutter is below the stutter threshold it is assumed that the peak is derived from stutter. If the stutter percent value is above the %stutter cutoff value, the peak is considered an allele. This determination becomes error-prone when a mixture of DNAs is present. Using a binary method to determine whether a peak is stutter versus an allele can have significant effects on the interpretation of samples with major/minor components. For example, Figure 3 shows that in a mixture where the amount of DNA from one contributor is significantly below the amount of DNA from the second (major) contributor, it is possible to have a minor allele and a stutter peak in the same position. In this instance, the amount of signal from the minor 15 allele and the amount of signal from stutter cannot be deciphered. Therefore, if the %stutter threshold is 8%, the 15 allele would be incorrectly ignored. This is evidence that a more robust model of stutter is needed to correctly interpret stutter/allele likelihoods when attempting DNA interpretation.

Due to the complex nature of DNA signal and the known signal interferences associated with artifacts such as stutter and dropout, samples containing low-template mixtures with > 2 contributors cannot properly be interpreted using methods currently available. Current methods do not utilize signal strength information to determine the NOC. Further, all currently available methods rely on utilizing an AT resulting in higher than necessary Pr(D), which has an effect on the ability to determine the NOC. If a sample is amplified with optimal DNA targets and the DNA from each component is relatively large then stutter can be easily characterized.[11] However, in cases where it is expected that the mass of DNA added to the PCR reaction was sub-optimal (i.e. < 0.25ng), elevated stutter may be expected.[21] Additionally, at optimal amplification inputs, peak height ratios at a heterozygous locus are close to 1 and the peak heights are large, leading to a probability of dropout of ~ 0. Therefore, for simple samples amplified using optimal DNA concentrations, the minimum number of contributors is usually equivalent to the actual number of contributors. However, the PHR variance has been shown to significantly increase with decreasing input levels.[22, 23] For situations where there are more than 3 contributors and suboptimal DNA inputs, the ability to infer either the NOC or the genotypes becomes impossible to accomplish

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

manually. As previously stated, the LR relies on the ability of the analyst to make an assessment regarding the number of contributors. The question then becomes; how can and does a DNA analyst determine the actual number of contributors for low-level, complex mixture profiles? And, are there methods to ascertain the actual number of contributors such that the analyst does not have to rely on the assumption that the minimum number of contributors is the actual number of contributors? The fact that there is potential for allelic dropout in a sample automatically precludes the analyst from confidently assessing the actual number of contributors by utilizing the number of alleles. Therefore, despite the recent advancements in complex mathematical systems to infer evidentiary genotypes - one question still remains; if a suspect is included as a potential contributor to an evidence stain and the hypothesis of the prosecution is based on a certain number of contributors - which may have been derived from a qualitative assessment of the number of alleles - is the defense required to agree with that assumption? For example, the prosecution may hypothesize that the suspect (S) and one unknown (U) were the contributors, while the defense may hypothesize that three unknowns (U1 and U2 and U3) contributed to the DNA profile. The number of contributors under H_P and H_D do not necessarily need to be equivalent, nor has it been determined in the literature whether the hypotheses need to be exhaustive. It has been suggested that they should be exhaustive or at least care must be taken to ensure no relevant hypotheses are omitted in the denominator.[15, 24] The SWGDAM Guidelines[25] and Budowle *et al.*[26] suggest numerous LR's be calculated if need be and all assumptions clearly stated in the report. This could potentially result in 2 or more individual LRs with no clear indication of which LR is to be regarded as the best estimate. This highlights the importance of the assumption regarding the number of contributors and brings to the forefront the concept that there is a need to develop systems, algorithms or interpretation processes which allow the laboratory to assess 1) the NOC that gave rise to the evidence and 2) the uncertainty associated with that number.

This research represents a detailed study into the effects interfering factors have on the ability to accurately infer the number of contributors to mixed stains, and is subdivided into two phases. The purpose of the first phase was to create single-source and mixture profiles from multiple contributors amplified at low target masses. Further, 4 factors that were expected to impact complex mixture interpretation were evaluated. Specifically, changes in baseline noise, stutter ratios, allele dropout and allele peak heights were evaluated at various targets, and multiple models describing each factor were compared. Once the optimal model to describe each factor was established the second phase of the project ensued. In the second phase, a method to determine the NOC to complex, low-template DNA mixtures was developed. The method was one that utilizes an APP (*a posteriori* probability), where the $\mathrm{APP}(i) := \Pr(n = i \mid E)$, which can be interpreted as the probability that a certain number (n) of contributors gave rise to the evidence. The work resulting from Phase I and II culminated into the development of a computational software tool that calculates the probability distribution for the number of contributors to a DNA sample, and is referred to as NOCIt. Tests which examine the results from the MAC, MLE and NOCIt were performed. Specifically, 1- to 5- person mixtures, amplified using 0.25 to 0.008 ng of total DNA were generated. The NOCs were determined via the MAC, MLE and NOCIt methods. The NOCs derived from each method was then compared to the real number of contributors that gave rise to the profile and accuracy rates were used as a means to compare methods.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Methods

The procedures are in accordance with the ethical standards of the Institutional Review Board. All reagents were purchased from Sigma Aldrich (Sigma Alrich, St. Louis MO) unless stated otherwise.

High molecular weight DNA was extracted from 68 single source samples using standard organic extraction procedures. The samples were whole blood, dried blood stains or saliva. The blood stains were either on Whatman® paper or cloth swatches. Saliva samples were either whole saliva or dried buccal swabs on cotton. Briefly, the organic extraction consisted of incubating the sample in 300 µg/mL of Proteinase K and 2% v/v SDS (sodium dodecyl sulfate) solution at 37 ^{o}C for 2 hours to overnight. Purification was accomplished with phenol/chloroform and alcohol precipitation. The DNA was dissolved in 50 µl of TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) at 56°C for 1 hour. Absolute DNA quantification was performed using real-time PCR and the Quantifiler® Duo™ Quantification kit according to the manufacturer's recommended protocol and one external calibration curve.[27, 28] A 7500 Sequence Detection System (Life Technologies, Inc.) was used for C_t (cycle threshold) detection. The extracted DNA was amplified using the manufacturer's recommended protocol (29 cycles) for AmpFℓSTR® Identifiler® Plus Amplification Kit (Life Technologies, Inc).[29] Single source samples were amplified using 0.5, 0.25, 0.125, 0.063, 0.047, 0.031, 0.016 and 0.008 ng of DNA. In addition, 2-, 3-, 4- and 5- person mixture samples were created by mixing the appropriate volumes of DNA extracts to attain the various ratios in Table 4. Once mixed, these samples were re-quantified and amplified using the same target masses used for the single-source samples. The PCR reaction consisted of 15 µL of master mix, the calculated volume of template DNA based on target mass required, and enough Tris-EDTA (TE) buffer (10 mM at pH 8.0) to bring the total reaction volume to 25 µL. Amplification was performed on Applied Biosystems' GeneAmp® PCR System 9700 using 9600 emulation mode. Positive and negative amplification controls were also run and showed expected results (data not shown). Fragment separation was accomplished using a 3130 Genetic Analyzer (Life Technologies) and a mixture containing appropriate amounts of HiDi (highly-deionized) formamide (8.7 μl/sample) (Life Technologies. Inc) and GeneScan™-600 LIZ™ Size Standard (0.3 µL/sample) (Life Technologies, Inc). A volume of 9 µL of that mixture and 1 µL of sample, negative or ladder was added to the appropriate wells. The samples were incubated at 95°C for 3 minutes and snap-cooled at -20°C for 3 minutes. Five, ten, and twenty second injections at 3 kV were performed for each of the samples and run according the manufacturers recommended protocol.[29] Fragment analysis was performed using GeneMapper IDX v1.1.1 (Life Technologies, Inc) using Local Southern sizing and an RFU threshold of 1. Known artifacts such as pull-up, spikes, -A, and artifacts due to dye dissociation were manually removed. A peak was considered pull-up if it was the same size (+/- 0.3bp) as a larger peak in another color and below 5% of the height of the more intense peak. Peaks were determined to be 'spikes' if they were in greater than 2 colors and in the same position. Peaks were considered as originating for incomplete adenylation (i.e. –A) if they were one base pair smaller than an allele, and peaks determined to originate from dye dissociation had to be in the same position, in the same color channel and be observed in multiple samples. The Genotypes Table, which included the File Name, Marker, Dye, Allele, Size and Height, was then exported for further analysis.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Phase I.

To examine baseline noise and the effects of target on establishing an optimal AT, the data obtained from the single source samples were exported from GeneMapper IDX v1.1.1 and filtered in order to remove stutter and allele peaks. Two methods to determine ATs were contrasted and compared. First is the method based on the approach proposed by Kaiser.[30] In this method the mean and standard deviation of the baseline noise are determined and the AT is calculated by adding some factor 'k' times the standard deviation to the mean noise signal as per,

$$AT_1 = \bar{x}_{noise} + ks_{noise} \quad \text{(Equation 4)}$$

Where AT_1 is the analytical threshold obtained via Equation 4, $\bar{x}_{noise}$ is the mean of the noise peak heights, k is a factor which is usually 3, and s_{noise} is the standard deviation of the noise heights. This value is also known as the MDS (minimum distinguishable/discernable signal) and the L_c (critical level) and is differentiated from the stochastic threshold or limit of quantification.[31] Another method to determine the AT is by examining the allele peak heights from DNAs amplified with known target masses and performing a least squares regression to obtain the *y*-intercept and its standard deviation.[20, 31, 32] In this work the AT derived by this method was determined by plotting the average peak height, corrected for diploidy, at each locus against target mass and performing a WLS (weighted least squares) regression. Once the *y*-intercept and its standard deviation were obtained, then the following equation was utilized to estimate the AT.

$$AT_2 = b + ks_b \quad \text{(Equation 5)}$$

Here AT_2 is the AT determined via Equation 5, b is the *y*-intercept obtained through a WLS regression, k is 3 and s_b is the standard deviation of the *y*-intercept.

To explore the impact that different target ranges have on the resultant ATs determined via Equations 4 and 5, 3 experimental designs utilizing different target ranges were tested.[33] The experiment, named *Range 1*, consisted of examining the baseline noise and *y*-intercepts obtained from 0.25, 0.125, 0.063 and 0.047 ng samples. *Range 2* examined the noise and intercept obtained when examining the lowest masses (0.047, 0.031, 0.016 and 0.008 ng), while *Range 3* examined the results from 4 points evenly spread throughout the entire range of masses (0.25, 0.062, 0.031, and 0.008 ng).

The distribution of the noise peak heights was examined and a signal model for noise was established. We call an index i isolated if neither at n - 4 nor at n + 4 there is a single or double peak. In other words an index i is isolated if it is neither in the n + 4 (forward stutter) nor in the n - 4 (reverse stutter) position of an allele. We consider the set

$$N = \{i : x_i = 0 \text{ and } i \text{ is isolated}\} \quad \text{(Equation 6)}$$

which is the set of indexes i where we have an isolated noise peak, i.e. a peak that is not an allele peak ($x_i = 0$), and which is not in n - 4 nor at n + 4 stutter position. Further, by $N_L = \{i \in N : i \text{ is on locus } L\}$ we denote the set of noise peaks indexes on locus L. Since, we do the analysis separately for all loci, for the sake of brevity, the index L is dropped in the following.

The peak heights of the measurements that are available for the analysis are quantized. In the statistical literature this is also known as grouped data. More precisely, we do not have the true

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

values y_i but only the quantized values Qy_i, which we abbreviate with $\tilde{y}_i$. Q denotes the quantization operator that is defined as

$$Qx = \left\lfloor x + \frac{1}{2} \right\rfloor \quad \text{(Equation 7)}$$

Where $\lfloor x \rfloor$ denotes the largest integer smaller than or equal than x. For large signal values the quantization can be neglected, because the introduced relative error is small, however for the noise analysis it is problematic, because the additive noise peak heights are small. We call the peaks $\{\tilde{y}_i : i \in N\}$ quantized noise peaks or noise peak measurements. Note that it is possible that $\tilde{y}_i = 0$ for some $i \in N$. In Figure 7 the histogram of the noise peak measurements is plotted for the different loci. Since most of the probability mass is concentrated in zero, i.e., most of the quantized noise peaks have zero height, further analysis only considers the non-zero noise peak measurements. We denote by $N^+ = \{\text{i} \in N : \tilde{y}_i > 0\}$ the set of indexes with a non-zero noise peak measurement.

Two distributions to describe the non-zero noise measurements were assessed, i.e. log-normal and Gaussian distribution. The Kolmogorov-Smirnov test was utilized in order to check whether the distribution of the non-zero noise peak measurements follows a quantized log-normal distribution. The log-normal cumulative distribution function (CDF) is given by

$$F_{m,s}(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\ln(x) - m}{s\sqrt{2}}\right)\right] \quad \text{(Equation 8)}$$

where erf is the complementary error function, and m and s are parameters. In a first step the parameters *m* and *s* are estimated. The maximum-likelihood estimates are given by

$$\hat{m} = \frac{1}{|N^+|} \sum_{i \in N^+} \log(\tilde{y}_i) \quad \text{(Equation 9)}$$

and

$$\hat{s} = \sqrt{\frac{1}{|N^+|} \sum_{i \in N^+} \log(\tilde{y}_i - \hat{m})^2} \quad \text{(Equation 10)}$$

Both estimators are for unquantized data, however, our simulations show that the estimates obtained are good enough. Kolmogorov-Smirnov tests were also conducted to test whether the normal distribution would also be an acceptable assumption.

Since the quantization of a continuous distribution induces a discrete distribution, the X^2-test was also utilized. The intervals [0:5; 3:5),[3:5; 4:5), [4:5; 5:5), [5:5; 6:5), [6:5; 7:5), [7:5; 8:5), [8:5; 9:5), and [9:5;1) were used to bin the data.

To examine any changes in baseline signal with respect to target, plots of the mean and standard deviation of noise against target mass were created for each locus. An ordinary least squares (OLS) regression was performed and the correlation coefficient (R^2) used as a means to examine whether the fit was satisfactory.

Similarly, two distributions to describe the allele peak heights were compared. Specifically, histograms of the peak heights for a given target, for each locus –using only heterozygous results - were plotted and Gaussian and log-normal fitting ensued. Once the distribution functions were obtained, the Kolmogorov-Smirnov test, which quantifies a distance between the empirical distribution of the samples and the Gaussian or log-normal cumulative distribution, was applied.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

A p-value of 0.05 was used to accept or reject the null hypothesis that the peak height values did not come from a normal (or log-normal) population. The mean of the peak heights and their standard deviations were then plotted against target and fit to a line with intercept of 0. The correlation coefficient was used to assess the goodness of fit.

Further, the frequency of allele dropout with respect to target was evaluated. To calculate the frequency of dropout, only heterozygous results were considered during this analysis. The frequency of dropout was calculated by dividing the number of non-detected alleles by the number of expected alleles. This was performed for each target and for each locus. Then four methods to characterize dropout were evaluated – two 'indirect' methods based on observed peak heights and two 'direct' methods using observed dropout frequencies. The first of the 'indirect' methods, named *Method 1*, included an assessment of the peak heights of the detected alleles to determine the predicted level of dropout. Specifically, a histogram of allele peak height was created using Igor Pro v6.12 and the auto-set bin function. A Gaussian curve was fitted to the data using

$$f(y) = 0.5\left[1 + \frac{\mathrm{erf}(y-\mu)}{\sqrt{2\sigma^2}}\right] \quad \text{(Equation 11)}$$

where μ is the mean, σ is the standard deviation and $\mathrm{erf(z)} = \frac{2}{\pi}\int_0^z e^{-t^2}\,dt$ and y is the peak height. Therefore, to approximate the probability of dropout, *f(y=1)* was determined and was taken to be the probability that an allele is not detected at an RFU threshold of 1. These probabilities were then compared to the frequency of dropout. The second model to describe dropout, *Method 2*, was similar to *Method 1* in that is utilized data obtained from detected alleles and is the second of the 'indirect' methods. However, in this instance the mean of the peak heights were calculated via

$$\mu = \overline{y} = \frac{\sum y}{n} \quad \text{(Equation 12)}$$

and

$$\sigma = s = \sqrt{\frac{\sum(y-\overline{y})^2}{n-1}} \quad \text{(Equation 13)}$$

where μ is the mean, σ is the standard deviation n is the number of peaks and $\overline{y}$ is the arithmetic mean of the peak height observed for a locus at a specific target mass. The values of $\overline{y}$ and s were then used as the μ and σ in Equation 11 and *f(y=1)* was determined and taken to represent the probability of non-detection (i.e dropout).

Two 'direct' methods to determine the Pr(D) at a given target mass were also evaluated. This was accomplished by plotting the frequencies of dropout against the target amount of DNA (ng). Logistic[34] (*Method 3*) and exponential (*Method 4*) curves were fit to the data using Igor Pro v6.12. The logistic curve takes the form

$$f(x) = \frac{1}{1+e^{-(a+bx)}} \quad \text{(Equation 14)}$$

While the exponential curve is

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

$$f(x) = ae^{-bx} \quad \text{(Equation 15)}$$

Therefore the probability of dropout , i.e. *f(x)*, could be approximated for any target mass *x*.

The differences between the observed frequencies of dropout versus the probability of dropout, estimated via each of the methods, was used to assess which model to use during algorithm development. Additionally, 45 single source samples, amplified using 0.25 to 0.008 ng of DNA, and 12 two-person mixture samples (1:1 or 1:2 mixture ratios), amplified using a target of 0.25 to 0.016 ng, were used to further test the validity of the chosen model. The likelihoods (calculation described in Literature Review using Equation 2 and 3) for n =1, 2, 3 and 4 were calculated using the frequency of dropout and the calculated probability of dropout for each contributors target mass. The probability of dropout was estimated using *Method 4* and APP differences < 0.05 were considered indications of the validity of *Model 4*.

Lastly, stutter models which describe the change in stutter ratio per locus for every target were established. In this instance, the average stutter ratios were calculated as per,

$$\left[\overline{\frac{H_{1-a}}{H_a}}\right] \quad \text{(Equation 16)}$$

where H_{1-a} is the height of the peak in stutter position and H_a is the height of allele a. The average stutter ratio was then plotted against target mass for each locus. The standard deviations of $\left[\frac{H_{1-a}}{H_a}\right]$ were also plotted against target mass. A decreasing exponential curve was fitted to the points and the goodness of fit was assessed by examining the residuals.

Once the four models which describe the 1) baseline noise, 2) allele peak height, 3) allele dropout and 4) stutter ratios with respect to template mass were chosen, Phase II of the project – which focused on the development and testing of the algorithm to determine the NOC – ensued. The software and its algorithm will henceforth be referred to as NOCIt.

Phase II.

NOCIt calculates the *a posteriori* probability (APP) on the number of contributors N given a particular evidence sample (electrophoresis profile) E. That is, it calculates $\Pr(N = n \mid E)$ for $n = 1,2,3, \ldots$. We assume that *a priori* N is uniformly distributed between 1 and $n_{\max}$, the maximum possible number of contributors. Thus, by Bayes' rule, we obtain

$$\Pr(N = n \mid E) \propto \Pr(E|N = n), \quad \text{(Equation 17)}$$

for $n = 1, \ldots, n_{\max}$. Let G_i, Θ_i be the genotype of and fraction of total DNA mass, respectively, contributed by $i \in \{1, \ldots, n_{\max}\}$, and let $\boldsymbol{G}$ and $\boldsymbol{\Theta}$ be the $n_{\max}$-component vectors of the G_i and Θ_i, respectively. We have

$$\Pr(E|N = n) = \int_{\boldsymbol{\theta} \in \Delta^{n-1}} \sum_{\boldsymbol{g} \in \mathscr{g}^n} \Pr(E|\boldsymbol{G} = \boldsymbol{g}, \boldsymbol{\Theta} = \boldsymbol{\theta}, N = n) \Pr(\boldsymbol{G} = \boldsymbol{g}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \quad \text{(Equation 18)}$$

where $\Delta^{n-1} := \{(x_1, \ldots, x_n) \in \mathbb{R}^n | \sum_{i=1}^{n} x_i = 1, x_i \geq 0 \ \forall \ i\}$ is the unit $n-1$ simplex, $\mathscr{g}$ is the space of possible genotypes (for both alleles of a contributor) in the population, and $f_{\boldsymbol{\Theta}}$ is the probability density function of $\boldsymbol{\Theta}$, which we assume to be uniform over Δ^{n-1}. The distribution

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

$\Pr(E|\boldsymbol{G}=\boldsymbol{g},\boldsymbol{\Theta}=\boldsymbol{\theta},N=n)$ is known because it is derived from calibration samples with known genotype.

We implement NOCIt using a Monte-Carlo sampling algorithm. We generate random samples of $\boldsymbol{g}$ and $\boldsymbol{\theta}$ using the background population allele frequencies and $f_{\boldsymbol{\Theta}}$ and, for each sample, we compute $\Pr(E|\boldsymbol{G}=\boldsymbol{g},\boldsymbol{\Theta}=\boldsymbol{\theta},N=n)$. After a large number of samples, we average all the computed values of $\Pr(E|\boldsymbol{G}=\boldsymbol{g},\boldsymbol{\Theta}=\boldsymbol{\theta},N=n)$ to obtain an approximation of Equation 18. We then calculate the APP according to

$$\Pr(N=n|E)=\frac{\Pr(E|N=n)}{\sum_{n=1}^{n_{\max}}\Pr(E|N=n)}. \qquad \text{(Equation 19)}$$

To determine the probability of the evidence given a specified genotype $\boldsymbol{g}$ and $\boldsymbol{\theta}$, the baseline noise, reverse stutter proportions, dropout rates and allele heights/areas are all considered and modeled as a function of target amount (i.e. mass). In every iteration of the Monte Carlo process, a genotype $\boldsymbol{g}$ is randomly chosen based on the frequencies of the alleles provided in any allele frequency table. In the current iteration of NOCIt, modeling of the dropout frequencies, means and variances of true peak heights, means and variances of the baseline noise heights, and the means and variances of stutter ratios is carried out using the following: An exponentially decreasing curve ($\Pr(D) = ae^{-bx}$) was used to model allele dropout, an exponentially decreasing curve ($\overline{SR} = ae^{-bx}+c$) was used to model the average stutter ratio (SR), a straight line with a positive slope ($\overline{H_{allele}} = mx$) was used to model the peak height, and a straight line with positive slope ($\overline{H_{noise}} = mx+b$) was used to model height of the baseline noise. For every allele in the genotype of the contributors, dropout of the allele is simulated by a Bernoulli trial. Two assumptions are made with regard to dropout: a) Dropout of one allele of a contributor is independent of dropout of the contributor's other allele and b) Dropout of an allele from a contributor is independent of dropout of the same allele from another contributor. Based on the evidence observed, the likelihood of observing the heights of the peaks given the genotypes of the contributors, the mixture ratio, the amount of DNA amplified and the time of injection is computed using calibration data. The STR loci are assumed to be independent of each other and the distributions for peak height, noise height and stutter ratios are assumed to be Gaussian.

To test the algorithm and software system, 1555 single source samples from 58 donors with known genotypes were used to calibrate NOCIt. During creation of the sample profiles, 3 times of injection (5, 10 and 20s) were used. The injection time is typically increased in the case of low template samples to increase the signal-noise ratio and hence 3 different sets of data each at a different injection time were used to study the change in signal with injection time. To test the performance of the software, NOCIt was run on 1-, 2-, 3-, 4- and 5- person mixtures. The performance of NOCIt was compared to the MAC and MLE methods. MAC uses the number of peaks observed in the signal to determine the minimum number of contributors while MLE uses the number of peaks as well as the allele frequencies. Both methods depend upon the establishment of a threshold to determine the set of true peaks. The threshold is typically chosen by a laboratory based on internal validation data. Two different types of thresholds were used for MAC and MLE for comparison purposes. The first threshold was a constant threshold of 50 RFU at all the loci, which is a commonly used threshold in forensic laboratories. The second threshold is a variable threshold. At each injection time, this threshold varies with DNA amount and dye color. Thus, this threshold was set by determining the height of the highest noise peak observed in the calibration data corresponding to a DNA amount, dye color and time of injection and setting that

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

height as the threshold for that DNA amount, dye color and time of injection. Application of MAC and MLE also used a stutter threshold to filter out the peaks in the reverse stutter position. Thus, any peak in the stutter position with a ratio less than the stutter filter was removed. The stutter filter specified by Applied Biosystems in the AmpF*l*str® Identifiler® Plus manual was used.[29] Allele frequencies from the Caucasian population specified in the AmpF*l*str® Identifiler® Plus manual were used for the NOCIt and MLE methods.[29]

Results

Phase I.

If an AT is to be utilized during evidence interpretation, consideration as to the which template ranges to study are of importance[35]; this is particularly true when examining low-level, complex mixtures since it is necessary to keep the AT to a minimum if allele dropout is to be minimized.[19] To explore the impact different target ranges have on determining the ATs, three ranges of target masses were evaluated. *Range 1* consisted of evaluating the baseline noise heights and *y*-intercepts obtained when 0.25, 0.125, 0.063 and 0.047 ng samples were utilized. *Range 2* examined the noise and intercept obtained when examining the lowest mass range, 0.047, 0.031, 0.016, 0.008, and, *Range 3* examined the results from 4 targets evenly spread throughout the entire range of masses tested (0.25, 0.063, 0.031, 0.008 ng). Figure 4. shows either the *y*-intercept and 3 times the standard deviation or the average height of the noise and 3 standards deviations obtained when data from *Ranges 1*, *2* and *3* were utilized. It is observed that, as discussed and predicted by Currie[33], the range of templates utilized to estimate the AT using Equation 5 can have a significant impact on the final result. For example, when *Range 1* and Equation 5 are utilized to determine the AT, the result is substantially higher than the AT derived using *Ranges 2* and *3*. Further, the ATs derived using this method yields AT which are up to 2 orders of magnitude larger than the ATs derived when the baseline noise is directly examined, and is similar to previously published results.[20] The ATs derived by directly examining the heights of the baseline noise using data from each of the Ranges is provided in Table 5. The higher range of masses, *Range 1*, results in higher ATs than the other two ranges tested, suggesting that when determining ATs for forensic purposes, mass ranges used during validation studies must take into account the template masses tested and expected in casework. Therefore if low-template DNA analysis is sought, a representative data set which includes the full-range of template masses observed in casework is recommended.

Although utilization of an AT determined via careful consideration of the template masses may decrease the Type II error rates (rates of false non-detection of true alleles) associated with allele detection, Figure 5 shows that the average and standard deviations of the baseline noise for each color channel change with template mass, suggesting that rather than utilizing an AT, a model which describes the baseline noise may prove useful for determining the NOC for low-template samples. As a result, the average and standard deviation of the baseline noise for each locus was plotted against target and showed the same trend. Representative plots are provided in Figure 6. It is observed that the averages and standard deviations of the noise remains relatively stable between 0.008 and 0.125 ng, but then increase with target. When an OLS regression is performed for similar data at every locus, at all three injection times, the trends are similar. Although the linear fit tends to overestimate the noise at targets of 0.125 and 0.25 ng, the residuals for the average differ did not exceed 2 RFU, suggesting the linear model may be useful for purposes of describing the increase in baseline noise with target masses. Further the R^2 values for the mean noise heights were 0.82, 0.93, 0.94 and 0.90 for the CSF1P0, D3S1358,

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

D18S51 and D5S818 loci respectively. The R^2 values for the standard deviations for the same loci were 0.90, 0.97, 0.94 and 0.56. All loci showed similar results.

To further examine which distribution to utilize in order to describe the noise, two distributions were tested. In Figure 7a the histogram of the noise peak measurements (all targets) is plotted for the different loci. Most of the probability mass is concentrated in zero, i.e., most of the quantized noise peaks have zero height. Therefore, for further analysis, we consider only the non-zero noise peak measurements.

Figure 7b shows the empirical CDF of the quantized noise peaks for samples amplified with 0.25 ng of DNA

$$F^E(x) = \frac{1}{|N^+|}\sum_{i\in N^+} 1_{\{\tilde{y}_i \le x\}} \quad \text{(Equation 20)}$$

and $F^Q_{\hat{m},\hat{s}}(x)$, i.e., the CDF of a quantized log-normal distributed random variable with parameters $m = \hat{m}$ and $s = \hat{s}$ for each locus. In the title of each sub-figure the Kolmogorov-Smirnov test statistic

$$\text{KS-stat} = \sup_x \left|F^E(x) - F^Q_{\hat{m},\hat{s}}(x)\right| \quad \text{(Equation 21)}$$

and the corresponding p-value are given. In Figure 7c the histogram of the quantized noise peaks and a pseudo PMF of a quantized log-normal distributed random variable with parameters $m = \hat{m}$ and $s = \hat{s}$ is visualized. The high p-values show that the null hypothesis that the quantized noise peaks follow a quantized log-normal distribution cannot be rejected. We are aware that our approach has two conceptual weaknesses. First, we estimate the parameters *s* and *m* from the same data that we use for the Kolmogorov-Smirnov test. Second, we use the Kolmogorov-Smirnov test for quantized data, and it is known to be too conservative, i.e., giving p-values which are too large. Therefore, the χ^2-test as also applied on the data and utilized to assess whether the log-normal or normal distributions are reasonable assumptions for the noise distribution.

Figure 7d shows a comparison of the log-normal and the normal distribution. The parameters of both distributions were obtained from the data by maximum likelihood estimators. It can be seen that the p-values of the Kolmogorov-Smirnov test for normal distribution are in the range from 0 to 0.651. In particular, loci D2S1338, D3S1358, D8S1179, D18S51,D19S433, D21S11, FGA, and TH01 have a p-value smaller than 0.05. Thus, given a significance level of 0.05, the null-hypothesis that the quantized noise peaks follow a quantized normal distribution for these loci would be rejected. In Figure 7e we see the p-values of the χ^2-test for the log-normal and the normal distribution. Except for locus D2S1338 the p-values for the log-normal distribution are all larger than 0.05. This confirms the previous findings that the log-normal distribution is a reasonable assumption for the noise distribution. In contrast, except for locus D13S317, all p-values for the normal distribution are smaller than 0.05, which is significant for the rejection of the null-hypothesis.

To evaluate the change in allele peak height, a similar analysis was performed. In this instance the average peak heights (only heterozygous results are used) for each locus was plotted against the target mass and an OLS regression ensued. Figure 8 shows the results for samples injected for 10 sec. In all cases, the linear model fits well to the data, whereby the R^2 values for all loci were ≥ 0.97. Plots of the standard deviation of the peak heights versus target resulted in

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

$R^2 \geq 0.92$ for all loci. Although endpoint PCR has traditionally been considered a sub-optimal way to quantify peak heights, low-template samples are not expected to exhibit obvious plateauing effects at cycle numbers as low as 29. That is, if the amplification efficiency of the PCR is 100% at the end-point cycle number of 29, and it is assumed the portion of the product and the RFU is directly proportional to the concentration of amplified product at 29 cycles, the following linear relationship is obtained,

$$RFU = \phi C_{j,0} 2^{29} + A \qquad \text{(Equation 22)}$$

where the A is the y-intercept (and may be expected to be 0 if the proportionality is unbiased) and the slope is 2^{29} multiplied by a proportionality constant (ϕ). Therefore, if optimal conditions are met, the samples contain accurate DNA concentrations, and the approximation that PCR efficiency does not change between concentrations is valid, a plot of RFU versus $C_{j,0}$ results in a straight line. However, the target mass at which plateauing effects will begin to be significant is dependent on the number of cycles and the template mass. It is to be emphasized that the current work focused on low-template ($\leq$ 0.25 ng) masses and therefore the dynamic ranges for this approach would need to be evaluated for each kit or amplification process. For example, if the dynamic range for an amplification kit which utilizes 29 cycles is 0.008 to 0.5 ng, it may not be assumed that the dynamic range is equivalent for a process that uses 32 or 35 cycles. However, previous studies have shown that the dynamic range can be quite large and were up to 1 ng for the AmpF*l*str® Identifiler® kit (28 cycles).[20] The distributions of the allele peak heights were also examined and the results for representative locus D21S11 are shown in Figure 9. Qualitatively, both the log-normal and Gaussian distributions seem reasonable. To confirm this quantitatively, Kolmogorov-Smirnov tests were applied and the results for the 0.25 ng samples are depicted in Figure 10 and 11 for the Gaussian and log-normal distributions, respectively. The KS statistic and the p-values are provided along with the cumulative distribution functions. The smallest p-value, when comparing the Gaussian to the empirical distributions, was 0.108 at D2S1338 and the smallest p-value for the log-normal comparison was 0.143 at D5S818. Since all loci resulted in $p > 0.05$, both the Gaussian and log-normal distributions were considered appropriate for algorithmic development purposes.

The probability of dropout and the modeling thereof has been extensively studied in the literature. Typically, the logistic model is applied.[10, 12, 13] However, dropout models are typically dependent upon the ability to measure the dropout rate as a function of total peak height, or surviving sister allele. Therefore, it was of interest to examine whether the same – or similar – models can be applied when the independent variable is the mass. Since the peak height can be a good indication of the input mass, the logistic model is hypothesized to be of value in this instance. Three other methods to estimate dropout probability were also assessed. *Methods 1* and *2* utilize peak height data to extrapolate that information in order to estimate the probability of non-detection of an allele. Results on the cumulative distribution of the empirical data compared to the Gaussian curve generated using *Method 1* and *2* for a representative locus, D16S519, is shown in Figure 12 for single source samples amplified using 0.25 and 0.008 ng. The associated means and standard deviations are shown in Table 6. Although similar, *Method 1* results in smaller residuals and a better fit to the data. However, both methods overestimate the dropout rates at high targets and underestimate the rates at low targets. Figure 13 shows the frequency of dropout for the same locus, D16S519, plotted against target. Here the data is fit to either a logistic (*Method 3*) or exponential (*Method 4*) function and the residuals are also indicated on the figure. The coefficients

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

(Equations 14 and 15) were 0.25 and 0.72 for coefficient *a*, and -131 and -105 for coefficient *b* for the logistic and exponential fits, respectively. Similar trends were obtained for all loci tested. A summary of the estimated probabilities of dropout for the D16S519 locus, calculated via *Methods 1-4* at 7 target amounts are summarized in Table 7. This shows that the 'indirect' methods for characterizing dropout overestimate dropout at higher targets and underestimate it at lower targets, while both 'direct' methods are appropriate ways of characterizing allele dropout. The underestimations associated with *Methods 1* and *2* suggest other factors beyond detection and PCR variation contribute to allele non-detection, as hypothesized by Gill et al.[36] Further, across all loci, the frequency of dropout increased as target amount decreased and increased with increasing molecular weight. Since *Method 4* was deemed an appropriate model to determine Pr(D), the validity of *Method 4* to estimate the NOC was tested. In this experiment, 45 single source samples, amplified using 0.25 to 0.008 ng of DNA were utilized. The likelihoods that n=1,2,3 or 4 contributors gave rise to the profile were determined using the either the frequency of dropout or the Pr(D) calculated via *Method 4*. The APP was then plotted and the results are shown in Figure 14a. The same examination was performed on 12, two-person mixtures and the results are presented in Figure 14b. Figure 14 shows that the APPs obtained using the model versus the actual frequencies of dropout are very close, whereby the difference never exceeded $1x10^{-6}$. This is an indication that *Method 4* is a viable method for determining the Pr(D) for purposes of determining the NOC to low-template DNA mixtures.

The last parameter to be assessed and characterized was stutter. To accomplish this, the average stutter ratios were plotted against target mass and a decreasing exponential curve was fit to the data. A representative plot for the D8S1179 locus is shown in Figure 15. Qualitatively it is observed that the mean and standard deviation of the stutter ratio increases as target amounts decrease. This phenomenon was observed for all loci tested. Therefore, using the fit information provided in Figure 15, one can estimate the average stutter expected for the D8S1179 locus amplified with the AmpF*l*str® Identifiler® Plus chemistry/protocol. Therefore at 0.008 ng, an average stutter ratio is expected to be

$$average\%stutter = 0.2207e^{-161.74x} + 0.0651, \text{ where } x\text{=0.008 ng} \quad \text{(Equation 23)}$$
$$average\%stutter = 12.5\%$$

The standard deviation of stutter is modeled in a similar fashion where the y0, *a* and *b* coefficients for the D8S1179 locus were determined to be 0.5336, -191.54 and 0.0237, respectively. Therefore the standard deviation of %stutter can be calculated in a similar fashion

$$Stdev\%stutter = 0.5336e^{-196.54x} + 0.0237, \text{ therefore at } x\text{=0.008 ng,} \quad \text{(Equation 24)}$$
$$Stdev\%stutter = 13.4\%$$

The average %stutter at D8S1179 for a 0.008 ng sample of 12.5 ± 13.4% is significantly different from the %stutter calculated for x = 0.25 ng which is 6.5 ± 2.4%. The %stutter obtained from the 0.25 sample are similar to the range of values provided in the manufacturer's manual for the D8S1179 locus, i.e. 2 – 11% for all alleles.[29] It should be noted that although the stutter ratios could also be modeled on a per allele basis, the model shown here takes the average stutter ratio for all alleles within a locus. It is expected that stutter models which utilize the length of the longest uninterrupted sequence would be of value, but are beyond the scope of this work. Each locus' average and standard deviation of stutter was plotted and fit in this manner and the

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

aforementioned model was used to determine the NOC. Stutter distributions were assumed to be Gaussian.

Phase II.

To determine the NOC to a DNA sample, all parameters were modeled as a function of template and all distributions were assumed to be Gaussian. Specifically, the baseline noise was modeled as a line with a non-zero *y*-intercept, the allele peak heights were modeled as a straight line with a *y*-intercept of 0, the average stutter ratios were modeled as an exponential. The standard deviations were modeled using the same functions as the means. The probability of allele dropout and non-occurrence of stutter was modeled as an exponential. Each of the models was examined, either through analysis of the residuals or via more extensive means, to confirm their validity. The normal distribution was shown to be reasonable for describing the variability in the peak heights at all loci at various targets. Gaussian distributions were assumed for stutter ratio and noise calculations. Future implementations of the algorithm are expected to incorporate improvements in the noise and stutter models as per findings in Phase I of this study.

The performance of NOCIt was compared to the Maximum Allele Count (MAC) and the Maximum Likelihood Estimator (MLE) methods. Since the MAC and MLE methods rely on the use of ATs, two different ATs were tested. The first was the common AT of 50 RFU and the second AT was the maximum noise peak observed for a given color channel. Figure 16 shows the performance of MAC, MLE and NOCIt for 30, 1-person samples amplified using 0.25 – 0.16 ng of DNA (i.e. 5 samples amplified at 0.25, 0.125, 0.063, 0.047, 0.031, 0.16 ng). Each sample was injected three times; once utilizing a 5 second injection, once with a 10 second injection and once using 20 second injection. The %accuracy is calculated by determining the most likely NOC divided by the actual NOC, multiplied by 100%. Figure 16 shows the summary of results. For single-source samples, the performance of the MLE and MAC methods are dependent upon injection time. Specifically, MACs accuracy decreased from 93 to 70% when an AT of 50 RFU was utilized. In contrast, when an injection specific AT was used the %accuracy of MAC increased from 70 to 90%, indicating the importance of ATs which are specific to the laboratory process. The %accuracy of the MLE method decreased with injection time when a constant AT was used, while %accuracy increased when an injection specific AT was applied. The accuracy of NOCIt also decreased with an increase in injection time from 100% to 93%. However, for every sample at every injection, NOCIt resulted in the highest accuracy rates and was not drastically impacted by time of injection. Figure 17 shows the %accuracy of 2-person mixture samples which consisted of 30 samples amplified at various targets (0.25, 0.125, 0.063, 0.047 and 0.031 ng) using various ratios (1:1, 1:2, 1:4, 1:9, 1:19). It should be noted that samples containing less than 2 cells worth of DNA (< 0.012 ng) from any one contributor were not utilized to assess accuracy.

Changing the injection time resulted in minor decreases in performance for all methods. Figures 18, 19 and 20 are the results from the 3-, 4- and 5- person mixtures respectively. The 3-person accuracy study was conducted by amplifying 13 samples using 0.047, 0.063, 0.0125 and 0.25 ng of DNA at ratios of 1:1:1, 1:2:1, 1:4:1, 1:9:1, 1:2:2, 1:4:4, 1:9:9. The 13, 4-person mixtures contained 0.25, 0.125 and 0.063 ng of DNA mixed in 1:1:1:1, 1:1:2:1, 1:1:4:1, 1:1:9:1, 1:2:2:1, 1:4:4:1 and 1:9:9:1 ratios. Lastly, 14, 5-person mixtures consisted of samples mixed in 1:1:1:1:1, 1:1:2:1:1, 1:1:4:1:1, 1:1:9:1:1, 1:1:2:2:1, 1:1:4:4:1, 1:2:2:2:1, and 1:4:4:4:1 ratios, amplified using 0.063, 0.125 and 0.25 ng total mass of template DNA. Similar to the 1-person samples, a minimum of 2 cells was required for accuracy testing.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

For the 2-person and 3-person samples, at all 3 injection times, NOCIt underestimated the number of contributors at the lower DNA amounts (2-person samples were called as 1-person samples, while most of the 3-person samples were called as 2-person samples and a few as 1-person samples). At the higher DNA amounts (> 0.047 ng), NOCIt had 100% accuracy with the 2-person and the 3-person samples. There were no overestimates from NOCIt for the 2-person and 3-person samples. The 4- and 5- person samples show that the overall %accuracy of all methods decreases and falls below 80%, regardless of injection time and/or AT utilized. However, there is a marked increase in the accuracy rates between the MAC and NOCIt methods for these complex mixtures, where the %accuracy of NOCIt was 2-fold greater than MAC. Further, the MLE method was superior to the MAC method for high-level mixtures and also resulted in a two-fold increase in the %accuracy over MAC. However, MLE was dependent on both injection time and AT, where the highest %accuracy was obtained with a 20 second injection and a constant AT of 50 RFU. Interestingly, MAC nearly always underestimated the 5-contributor mixtures (only one sample was correctly identified as a 5-person mixture). MLE resulted in higher accuracy rates with the highest %accuracy (i.e. 71%) originating from the dataset analyzed using a 20 second injection time and a constant AT of 50 RFU. NOCIt resulted in the highest %accuracy of 64% when the 20 sec injection was utilized. The MLEs accuracy is highly dependent on the injection time for the 4- and 5-person samples, where the %accuracy was 15%, 46% and 69% for the 4-person mixtures injected for 5, 10 and 20 seconds, respectively. The %accuracy of MLE for the 5-person mixtures was 14%, 57% and 71% for the 5, 10 and 20 second injections respectively. This is hypothesized to be the effect of increasing the rates of allele detection by increasing the amount of product in the capillary. In contrast, NOCIt results were not as affected by injection time for these samples and were 69%, 61% and 69% (4-person) and 50%, 43% and 64% (5-person) for the 5, 10 and 20 second injections. A summary of results for each time of injection for all samples/mixtures is shown in Figure 21. In summary, for all samples and mixtures tested, NOCIt resulted in higher accuracy rates for all times of injection. Further, it was unaffected by injection time, suggesting an approach which utilizes information procured from the laboratory and takes into account the changing stutter, baseline noise, dropout rates and peak heights is essential for accurate interpretation of low-template mixtures. This information further highlights that the minimum number of contributors is not equivalent to the actual NOC for low-template or highly mixed samples. Although the MLE method resulted in higher rates of accuracy than MAC, especially for samples containing 4- and 5- contributors, it is highly dependent on the rate of allele detection. If MLE is to be utilized for low-template samples, enhancing the amount of product injected into the capillary is one way to improve its accuracy.

The %accuracy rates of < 80% for the 3-, 4- and 5- person, low-template mixtures highlights the issues associated with complex mixture interpretation. For example, Figure 22 is the electropherogram and corresponding output from NOCIt for a sample amplified using a low-template (0.016 ng) mixture of 2 contributors injected for 10sec using a 3 kV injection voltage. The peaks which can confidently be discerned from baseline are highlighted with a red arrow. Within the arrow is the allele designation (i.e. STR allele). The loci and true genotypes of contributor 1 and 2 (i.e. G_1 and G_2) are also depicted. As expected, the electropherogram shows there is a substantial amount of allele dropout and there are multiple loci with complete dropout. Note that the alleles 9,12 at the D7S820 are a composite of two heterozygous contributors where one allele from each contributor dropped out. Therefore, unless the laboratory has a standard operating procedure which explicitly prohibits the interpretation of such low-template samples using this sample and others like it are at high risk for mis-classification using MAC and MLE as

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

there is no quantitative indication that an NOC of 2 is a possibility. Thus, extremely corrupted samples, such as the one depicted in Figure 22 were also tested in an attempt to evaluate whether it is possible to deduce the true NOC without significant probability of error. To evaluate this, 53 samples which contained very low quantities (i.e., < 2 cells) from any contributor were evaluated and the results are shown in Table 8. Table 8 depicts the difference between the calculated NOC and the true NOC obtained via the MAC, MLE and NOCIt methods. Overall, the accuracy of these samples is lower than samples which do not exhibit excessive allele drop-out from any one contributor and resulted in accuracy rates of 26%, 26% and 30% for the MAC, MLE and NOCIt methods respectively. It is also observed that the extent to which the samples are underestimated is substantial in that they may be underestimated by more than one contributor. MAC and MLE grossly (i.e. NOC_{Calc}-$NOC_{True} \leq -2$) underestimated the NOC's 55% of the time. In contrast, for samples containing a minor contribution containing < 2cells, NOCIt grossly underestimated the NOC 45% of the time.

Since there is a non-trivial probability of error associated with evaluating NOC, NOCIt provides an estimate on the NOC. Figure 22 shows the output obtained when the sample was interpreted using NOCIt - using an n_{max} =2 - and shows that the output does not provide a single number (i.e. the estimated NOC). Rather than provide the most likely NOC, NOCIt presents the end-user with the probability that the stain originated from 0 versus 1 versus 2 separate, independent biological sources. By providing the probability estimates, NOCIt supplies the measure of uncertainty associated with the assessment. This allows the ability to assess the validity of the assumption on the NOC reported and can also be an indication of the complexity of a sample. The APPs of 1E-429, 0.91 and 0.09 for n= 0,1 and 2, respectively, shows that although NOCIt suggests the most likely NOC is 1, it also suggests there is a reasonable probability that the NOC may also be 2. Figure 23 shows the comparison between the %accuracy of NOCIt (as depicted in Figures 16 to 20) versus the percentage of time NOCIt resulted in a probability of at least 1% for the number of contributors in the sample. For example, for the data set containing 2-contributors, the percentage of time NOCIt returned an APP $\geq$ 0.01 for n=2 was determined. An APP of 0.01 was utilized as a cutoff indicating that NOCIt showed a 'reasonable' probability that the sample may have originated from the correct NOC. As indicated in Figure 23, the 'accuracy' increases for every mixture set. This again is an indication that complex DNA mixtures may need to be evaluated under multiple assumptions using probabilistic analysis methods – particularly when the number of contributors exceeds 2 and the total template mass is $\leq$ 0.25 ng. Since the output is in the form of an APP distribution, an assessment of the complexity of the profile prior to comparison is possible. Figure 24 shows the APP distribution for a low-template 2-person mixture when only 10 loci are considered, and shows the output can be used to inform the analyst of the potential complexity of a sample. For example, if the probabilities are not strongly peaked, then this information could be utilized during the comparison to a known. That is, multiple LR's may need to be calculated and reported, or the LR may incorporate multiple assumptions on NOC, or the laboratory may report that since the NOC is in question, the sample is too complex to render a meaningful conclusion with respect to a known.

A prototype of NOCIt has been developed in Java and the interface is shown in Figure 25. The NOCIt application, technical manual and tutorial may be downloaded from www.bu.edu/dnamixtures. Modeling of 1) peak height with respect to target, 2) reverse stutter ratios with respect to target, 3) baseline with respect to target and 4) allele dropout with respect to target have been integrated into the system. Output files that contain the APPs on the NOCs, the

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

input information (i.e. calibration file used, frequency information used and sample file) and the analysis time is available for the end-user to utilize with reports or further studies.

NOCIt can be used within the forensic DNA laboratory process and does not require interruption to the current laboratory scheme. That is, data analyzed using the current laboratory software (i.e. GeneMapper) is exported in tabular form and used by NOCIt. The following is a description of the standard operating procedure used with NOCIt and details the user/laboratory interaction;

1. NOCIt has been written in Java and is designed to work on Windows systems running JRE 1.7.
2. Download NOCIt. The software is present as part of a zip file. Extract the contents of the zip file. 'NOCIt.jar' is the file used to run the software. The folder 'lib' contains the Apache Commons Math files used for curve fitting. 'NOCIt.jar' and the folder 'lib' should always be kept in the same directory.
3. Double click on 'NOCIt.jar' to launch the software.
4. NOCIt requires the users to provide 6 inputs, the formats of which are described below:
   a. Calibration file: This file contains the single source samples used for calibration of the software and should be in the 'csv' format. For example, it is the GeneMapper output file, analyzed using a 1RFU peak threshold. The first line is a header line. The first column is 'General information' and contains details about the calibration samples. This column is not used by the software during the calculation. The second column contains the DNA input (in ng) used to amplify the sample. The third and the fourth columns contain the 2 known alleles present in the sample at the locus. Homozygous alleles are listed twice. (i.e. 10,10). The fifth and sixth columns are 'Marker' and 'Dye'. From the seventh column onwards, the description of the peaks are present, with each peak having 3 characteristics: 'Allele', 'Size' and 'Height', in that order.
   b. Allele frequency file: This is the file containing the frequencies of the alleles in the population the user is testing. The file should be in the 'csv' format. The first line is a header line. The first column is 'locus', the second column is 'allele' and the third column is 'frequency'. At every locus, only the set of alleles present in the frequency table are considered potential alleles while performing the calculation.
   c. Sample file: This is the 'evidence' file with the unknown number of contributors that the user is interested in analyzing. This file should be in the 'csv' format. This is exported from the laboratories' data analysis software (i.e. GeneMapper) using an RFU threshold of 1. The first line is a header line. The first column is a 'General information' column that contains details about the sample and is not used by the software. The second and third columns are 'Marker' and 'Dye'. From the fourth column onwards, the description about the peaks are present, with each peak having 3 characteristics: 'Allele', 'Size' and 'Height', in that order.
   d. Output file: This file contains the results of the calculation. The output file is in the 'txt' format. The user can browse and choose a txt file as the output file. Alternatively, if no output file is provided, NOCIt creates an output file in the same directory as the sample file and with the default name " 'Sample file name'_NOCIt_output.txt ". An example of the output file is provided in Figure 26.

      The beginning of the output file contains the inputs specified by the user - the calibration file, the frequency file, the sample file and the sample DNA input.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

After that, the results of the calculation are shown. For each number of contributors, NOCIt displays:

- The time taken (in minutes)
- The probability of observing the evidence at each locus, given the number of contributors
- The set of alleles at every locus that had the highest probability while doing the sampling
- The probability of observing the entire evidence, given the number of contributors. This is computed as the product of the probabilities at all the loci. The value is displayed as 'Likelihood'.
- The probability of the sample coming from the number of individuals, given the evidence. This is computed by normalizing all the likelihood values to add to 1. The value is displayed as 'Probability'.

e. Maximum number of contributors: This is the maximum number of individuals for which the user is interested in computing the likelihood. Valid values: Any integer between 0 and 5.

f. DNA input: This represents the amount of DNA amplified to obtain the sample profile. Valid values: Any real number greater than 0.

5. Once all the inputs have been specified, click on 'Start' to initiate the calculation.
6. A progress bar starts moving and indicates that the software is running.
7. Click on 'Stop' at any moment to terminate the calculation. The values that have been computed up to this point would be written to the output file.
8. If the software continues running uninterrupted, at the end of the calculation a pop up box informs the user that the software has finished running.
9. Open the output file to view the results.

Conclusion

Allele dropout was shown to have a detrimental effect on the ability to infer the actual number of contributors. As allele dropout increased, so did the probability of underestimating the NOC.

Models that estimate the rise in baseline noise and allele heights were developed. Both the baseline noise and allele peak heights increased linearly with target. Although the linear model was chosen for baseline noise in this study, further valuation into the model is required since it was observed that the linear growth occurred when target mass > 0.125 ng. Further, it is unclear if this increase in baseline noise at targets > 0.125 ng is consistent between kit chemistries, laboratories, instruments, etc. Therefore, future research would aim at elucidating an approach that utilizes information provided in 1) blank/negative samples and 2) samples containing DNA. The exponential function was found to be a good approximation of 1) the frequency of dropout and 2) the average stutter ratios at a locus. Since the stutter ratio is dependent on the number of uninterrupted repeats, examination into the implementation of sequence specific stutter models are warranted. Gaussian distributions were assumed, and the approximation was tested for baseline noise and peak height. Preliminary results suggest the Gaussian distribution describes the variability it the peak height well. Baseline noise is better described using a log-normal distribution, and studies that assess the impact of this finding on determining the NOC are warranted. Stutter ratios were calculated for the entire locus and a Gaussian distribution was assumed. Though the stutter ratio model and distribution used for NOCIt resulted in high levels of accuracy, improvements to the models/distributions describing stutter are of interest.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

As the signal-noise ratio increased with an increase in injection time, so did the accuracy of MLE. MAC was shown to be an insufficient method for determining the actual NOC for samples containing more than 2 contributors. NOCIt was minimally affected by a change in the injection time, as the software is designed to use parameters from calibration data corresponding to a specific laboratory process. Overall, the accuracy of all 3 methods increased with an increase in DNA amount. Samples that contain at least one contributor with fewer than 2 cells are prone to gross underestimation, where the actual and calculated true NOCs differ by at least 2. Thus, in addition to providing estimates on the most likely NOC, NOCIt provides the APP distribution across 0 to 5 contributors such that the uncertainty associated with likely NOC is available to the analyst and trier-of-fact. Information on NOCIt training and the use of NOCIt can be found on www.bu.edu/dnamixtures.

References

1. Haned, H.; Pene, L.; Lobry, J. R.; Dufour, A.; Pontier, D., Estimating the number of contributors to forensic DNA Mixtures: Does Maximum Likelihood Perform Better than Maximum Allele Count? *J. Forensic Sci.* **2011,** 56, (1), 23-28.
2. http://forensim.r-forge.r-project.org/ (July 24, 2014).
3. Gill, P.; Buckleton, J., A Universal Strategy to Interpret DNA profiles that does not require a definition of low-copy-number. *Forensic Sci. Int. - Genetics* **2010,** 4, 221-227.
4. Gill, P.; Kirkham, A.; Curran, J., LoComatioN: A software tool for the analysis of low copy number DNA profiles. *Forensic Sci. Int.* **2007,** 166, 128-138.
5. Gill, P.; L. Gusmao; Haned, H.; Mayr, W. R.; Morling, N.; Parson, W.; Prieto, L.; Prinz, M.; Schneider, H.; Schneider, P. M.; Weir, B. S., DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluationn of STR typing results that may include drop-out and/or drop-in using probabilistic methods. *Forensic Sci. Int. Genetics* **2012,** 6, 679-688.
6. Perlin, M. W.; Szabady, B., Linear Mixture Analysis: A Mathematical Approach to Resolving Mixed DNA Samples. *J. Forersic Sci.* **2001,** 46, (6), 1372-1378.
7. Wang, T.; Xue, N.; Birdwell, J. D., Least-Square Deconvolution: A Framework for Interpreting Short Tandem Repeat Mixtures. *J Forensic Sci* **2006,** 51, (6), 1284-1297.
8. Curran, J. M.; Gill, P.; Bill, M. R., Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure. *Forensic Sci. Int.* **2005,** 148, 47-53.
9. Curran, J. M., A MCMC method for resolving two person mixtures. **2008,** 48, 168-177.
10. Puch-Solis, R.; Rodgers, L.; Mazumbder, A.; Pope, S.; Evett, I.; Curran, J.; Balding, D., Evaluating Forensic DNA Profiles using Peak Heights, Allowing for Multiple Donors, Allelic Dropout and Stutters. *Forensic Sci. Int. Genetics* **2013,** 7, 555-563.
11. Bright, J.-A.; Taylor, D.; Curran, J. M.; Buckleton, J. S., Developing Allelic and Stutter Peak height Models for a Continuous Method of DNA Interpretation. *Forensic Sci. Int. Genetics* **2013,** 7, 296-304.
12. Haned, H.; Egeland, T.; Pontier, D.; Pene, L.; Gill, P., Estimating drop-out probabilities in forensic DNA samples: A simulation approach to evaluate different models. *Forensic Sci. Int. Genetics* **2011,** 5, 525-531.
13. Tvedebrink, T.; Eriksen, P. S.; Mogensen, H. S.; Morling, N., Estimating the Probability of Allelic Drop-out of STR alleles in forensic genetics. *Forensic Sci. Int. - Genetics* **2009,** 3, 222-226.
14. Perlin, M. W.; Sinelnikov, A., An Information Gap in DNA Evidence Interpretation. *Plos One* **2009,** 4, (12), e8327.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

15. Buckleton, J. S.; Curran, J. M.; Gill, P., Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *Forensic Sci. Int. - Genetics* **2007,** 1, 20-28.
16. Paoletti, D. R.; Krane, D. E.; Raymer, D. L.; Doom, T. E., Inferring the Number of Contributors to Mixed DNA Profiles. *IEEE/ACM Trans. on Computational Bio. and Bioinformatics* **2012,** 9, (1).
17. Biedermann, A.; Bozza, S.; Konis, K.; Taroni, F., Inference about the number of contributors to a DNA mixture: Comparative analyses of a Bayesian network approach and the maximum allele count. *Forensic Sci. Int. Genetics* **2012,** 6, 689-696.
18. Egeland, T.; Dalen, I.; Mostad, P. F., Estimating the Number of Contributors to a DNA Profile. *Int J Legal Med* **2003,** 117, 271-275.
19. Rakay, C. A.; Bregu, J.; Grgicak, C. M., Maximizing allele detection: Effects of analytical threshold and DNA levels on rates of allele and locus drop-out. *Forensic Sci. Int. Genetics* **2012,** 6, (6), 723-728.
20. Bregu, J.; Conklin, D.; Coronado, E.; Terrill, M.; Cotton, R. W.; Grgicak, C. M., Analytical thresholds and Sensitivity: Establishing RFU thresholds for forensic DNA analysis. *J. Forensic Sci.* **2012,** 58, 120-129.
21. Hill, B., Low Copy Number (LCN) DNA Analysis. In *Mid-Atlantic Association of Forensic Sciences Meeting*, Hunt Valley, Maryland, 2009.
22. Bright, J.-A.; Turkington, J.; Buckleton, J., Examination of the variability in mixed DNA profile parameters for the Identifiler multiplex. *Forensic Sci. Int. - Genetics* **2010,** 4, 111-114.
23. Ruiz, E.; Grgicak, C. M.; Cotton, R. W., Amplification Reproducibility of profiles generated using Identifiler and MiniFiler PCR Amplification Kits: Effects on Mixture Interpretation. In *Northeastern Association of Forensic Sciences*, White Plains, New York, 2008.
24. Brenner, C. H.; Fimmers, R.; Baur, M. P., Likelihood ratios for mixed stains when the number of donors cannot be agreed. **1996,** 109, (4), 218-219.
25. SWGDAM SWGDAM Interpretation Guidelines for Autosomal STR Typing. http://www.fbi.gov/about-us/lab/codis/swgdam.pdf (11/22/2010),
26. Budowle, B.; Onorato, A. J.; Callaghan, T. F.; Manna, A. D.; Gross, A. M.; Guerrieri, R. A.; Luttman, J. C.; McClure, D. L., Mixture Interpretation: Defining the Relevant Features for Guidelines for the assessment of Mixed DNA Profiles in Forensic Casework. *J. Forensic Sci.* **2009,** 54, (4), 810-821.
27. Grgicak, C. M.; Urban, Z. M.; Cotton, R. W., Investigation of Reproducibility and Error Associated with qPCR Methods using Quantifiler Duo DNA Quantification Kit. **2010,** 55, (5), 1331-1339.
28. Cicero, M. C.; Grgicak, C. M., Examination into the Applicability and Stability of a Single External Calibrator for Forensic DNA Quantification. In *NorthEastern Association of Forensic Scientists*, Cromwell, CT, 2013.
29. AppliedBiosystems, *AmpFlSTR(R) Identifiler(R) Plus PCR Amplification Kit User's Manual.* 1st ed.; Applied Biosystems: 2006.
30. Kaiser, H., Part II: Quantitation in Elemental Analysis. *Anal. Chem.* **1970,** 42, (4), 26A-59A.
31. Winefordner, J. D.; Long, G. L., Limits of Detection: A Closer Look at the IUPAC Definition. *Anal. Chem.* **1983,** 55, (7), 712A-724A.
32. Mocak, J.; Bond, A. M.; Mitchell, S.; Scollary, G., A Statistical Overview of Standard (IUPAC and ACS) and New Procedures for Determining the Limits of Detection and

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Quantification: Application to Voltammetric and Stripping Techniques. *Pure & Appl. Chem.* **1997,** 69, (2), 297-328.
33. Currie, L. A., Detection: International Update, and Some Emerging Di-lemmas Involving Calibration, the Blank and Multiple Detection Decisions. *Chemometrics and Intelligent Laboratory Systems* **1997,** 37, (1997), 151-181.
34. Puch-Solis, R.; Kirkham, A. J.; Gill, P.; Read, J.; Watson, S.; Drew, D., Practical Determination of the Low Template DNA Threshold. *Forensic Sci. Int: Genetics* **2011,** 5, 422-427.
35. Currie, L. A., Detection: International update, and some emerging di-lemmas involving calibration, the blank and multiple detection decisions. *Chemometrics and Intelligent Laboratory Systems* **1997,** 37, 151-181.
36. Gill, P.; Curran, J.; Elliot, K., A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. *Nucl. Acids Res.* **2005,** 33, (2), 632-643.

Dissemination of Research Findings

1 H. Swaminathan, C.M. Grgicak, M. Medard and D. S. Lun. NOCIt: *A Computational Method to Infer the Number of Contributors to DNA Samples Analyzed by STR Genotyping*. Forensic Science International: Genetics, 16, 172-180 (2015).

2 66th Annual AAFS Scientific Meeting – Sarah Norsworthy, Desmond S. Lun, Harish Swaminathan, Muriel Medard and Catherine M. Grgicak. *Characterizing Rates of Allelic Dropout and the Impact on Estimating the Number of Contributors.* (February 2014)

3 25th International Symposium on Human Identification and 39th Northeastern Association of Forensic Scientists– Kayleigh Rowan, Genevieve Wellner, Desmond S. Lun, Muriel Medard and Catherine M. Grgicak. *Characterization of the Sources of Peak Height Uncertainty Resulting from Ordinary Alterations During Forensic DNA Processing: Examining Validation Schemes for the Calibration of NOCIt.* (October 2013)

4 25th International Symposium on Human Identification – Harish Swaminathan, Catherine M. Grgicak, Muriel Medard and Desmond S. Lun. NOCIt: *A High-Accuracy Computational Method for Determining the Number of Contributors in an STR DNA Profile*. (September 2013).

5 39th Northeastern Association of Forensic Scientists – Harish Swaminathan, Catherine M. Grgicak, Muriel Medard and Desmond S. Lun. NOCIt: *A Computational Tool to Infer the Number of Contributors to a Forensic DNA Samples.* (September 2013).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Tables

Table 1. Allelic peaks observed in a typical DNA electropherogram

Locus	Peak 1	Peak 2
D8S1179	13	16
D21S11	29	32.2
D7S820	8	11
CSF1PO	11	12
D3S1358	15	16
TH01	6	9
D13S317	11	11
D16S539	11	12
D2S1338	19	24
D19S433	15	15
vWA	18	19
TP0X	8	11
D18S51	13	14
D5S818	10	12
FGA	20	20

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 2. Possible genotype combinations that may explain how two random individuals could have given rise to the genotype in Table 1.

Row	Possible Genotypes Person 1	Possible Genotypes Person 2
1	13,13	13,16 or 16,16
2	13,16	13,16 or 13,13 or 16,16
3	16,16	13,13 or 13,16
4	13,13	16,O
5	13,16	13,O or 16,O
6	16,16	13,O
7	13,O	13,16 or 16,16
8	16,O	13,13 or 13,16
9	13,16	O,O
10	13,O	16,O
11	16,O	13,O
12	O,O	13,16

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 3. A comparison of the likelihoods of n=1 and n=2 for a simulated profile with ≤ 2 alleles at every locus for different targets.

0.6 ng			**0.2 ng**		
L(n=1)	**L(n=2)**		**L(n=1)**	**L(n=2)**	
9.08e-20	1:1	2.67e-30	9.08e-20	1:1	3.51e-30
	1:2	2.67e-30		1:2	5.58e-30
	1:4	2.67e-30		1:4	3.13e-30
	1:9	2.67e-30		1:9	7.19e-30

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 4. The DNA ratios of multiple contributors when generating the 2-, 3-, 4- and 5- person mixtures.

No. of people in the mixture	**2**	**3**	**4**	**5**
	1:1	1:1:1	1:1:1:1	1:1:1:1:1
	1:2	1:2:1	1:1:2:1	1:1:2:1:1
	1:4	1:4:1	1:1:4:1	1:1:4:1:1
	1:9	1:9:1	1:1:9:1	1:1:1:9:1
	1:19	1:2:2	1:2:2:1	1:1:2:2:1
		1:4:4	1:4:4:1	1:1:4:4:1
		1:9:9	1:9:9:1	1:1:9:9:1
				1:2:2:2:1
				1:4:4:4:1
				1:9:9:9:1

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 5. The average noise and 3 standard deviations for the noise peak heights obtained from samples amplified using *Range 1* (0.25 – 0.047 ng), *Range 2* (0.047 – 0.008 ng) and *Range 3* (0.25 – 0.008 ng) of target DNA mass. The AT (avg + 3stdev) is also shown.

	Range 1			***Range 2***			***Range 3***		
Locus	**Avg Noise Height (RFU)**	**3StdDev Noise Height**	**AT**	**Avg Noise Height (RFU)**	**3StdDev Noise Height**	**AT**	**Avg Noise Height (RFU)**	**3StdDev Noise Height**	**AT**
D8S1179	3	6	10	3	5	9	3	7	10
D21S11	3	6	9	3	4	7	3	5	8
D7S820	3	4	7	3	4	7	3	5	8
CSF1PO	3	6	9	3	4	7	3	4	7
D3S1358	5	8	13	5	7	12	5	8	13
TH01	4	8	12	4	6	10	4	6	10
D13S317	4	9	13	4	5	9	4	7	12
D16S539	4	6	10	4	5	9	4	6	10
D2S1338	4	7	11	4	6	10	4	6	11
D19S433	8	10	18	7	9	16	8	9	17
vWA	7	9	16	7	8	15	7	8	15
TP0X	8	12	20	7	10	17	7	10	18
D18S51	7	8	16	7	8	15	7	8	15
D5S818	10	19	28	8	16	23	9	16	25
FGA	8	9	16	7	8	15	8	9	16
D8S1179	7	8	15	7	8	15	7	8	15

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 6. The mean and standard deviation for *Method 1* (fitted cumulative Gaussian) and *Method 2* (non-fitted cumulative Gaussian) for the representative D16S539 locus obtained with 68 single-source samples amplified using 0.25 and 0.008 ng of template DNA.

.

	0.25ng		**0.008 ng**	
Method	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
1	944	446	37	48
2	1076	440	58	43

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 7. Allele dropout and estimated probabilities of dropout calculated by *Methods 1- 4* at 7 target amounts for representative locus D16S539, showing that *Methods 1* and *2* overestimated dropout at higher target amounts but underestimate dropout at lower target amounts whereas *Methods 3* and *4* are both appropriate characterizations of dropout. Highlighted cells signify most the accurate approximation.

Target (ng)	**Observed**	***Method 1***	***Method 2***	***Method 3***	***Method 4***
0.25	0	0.017	0.007	7.25E-15	2.72E-12
0.125	0	0.005	0.016	9.67E-8	1.40E-6
0.0625	0	0.104	0.050	3.53E-4	1.01E-3
0.047	0.008	0.074	0.027	2.69E-3	5.14E-3
0.0313	0.032	0.158	0.081	0.021	0.027
0.0156	0.136	0.209	0.085	0.143	0.140
0.0078	0.319	0.226	0.091	0.316	0.314

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 8. The $NOC_{Calc} - NOC_{True}$ obtained when examining samples containing minor components with < 2 cells worth of DNA with the MAC, MLE and NOCIt methods.

	NOC_{Calc} - NOC_{True} (N=53, Minor < 12.6 pg)					
	-4	**-3**	**-2**	**-1**	**0**	**1**
MAC	3	11	15	9	14	1
MLE	3	11	15	9	14	1
NOCIt	1	12	11	13	16	0

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Figures

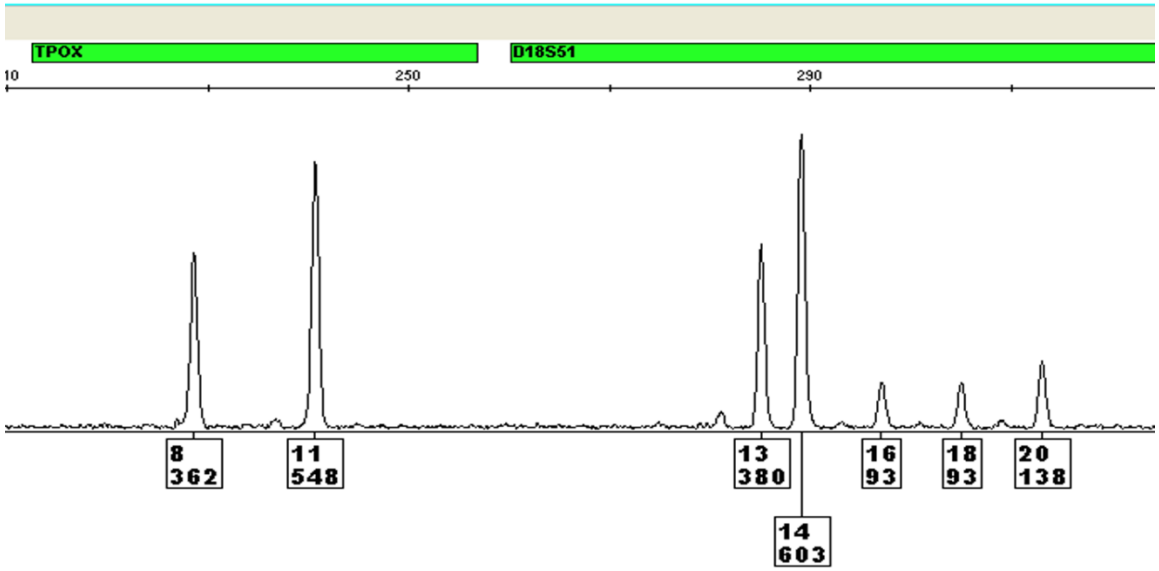


Figure 1. Complex low-level DNA profiled amplified using 125 pg of DNA using the AmpF*l*str® Identifiler® Plus Kit, injected for 10s on a 3130 Genetic analyzer. This data represents three contributors in a 1:1:2 ratio.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

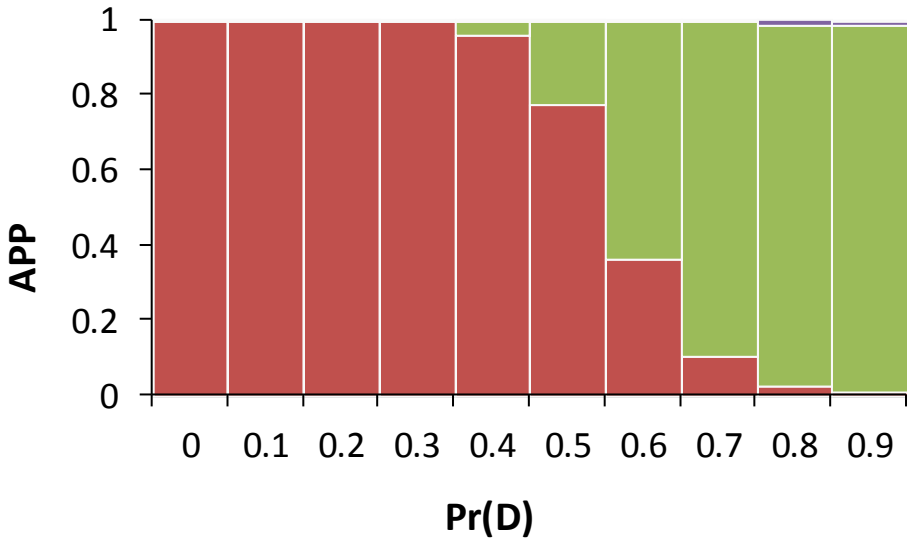


Figure 2. The probability that the profile from Table 1 resulted from (■) 1 versus (■) 2 versus (■) 3 versus (■) 4 contributors at various levels of dropout.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

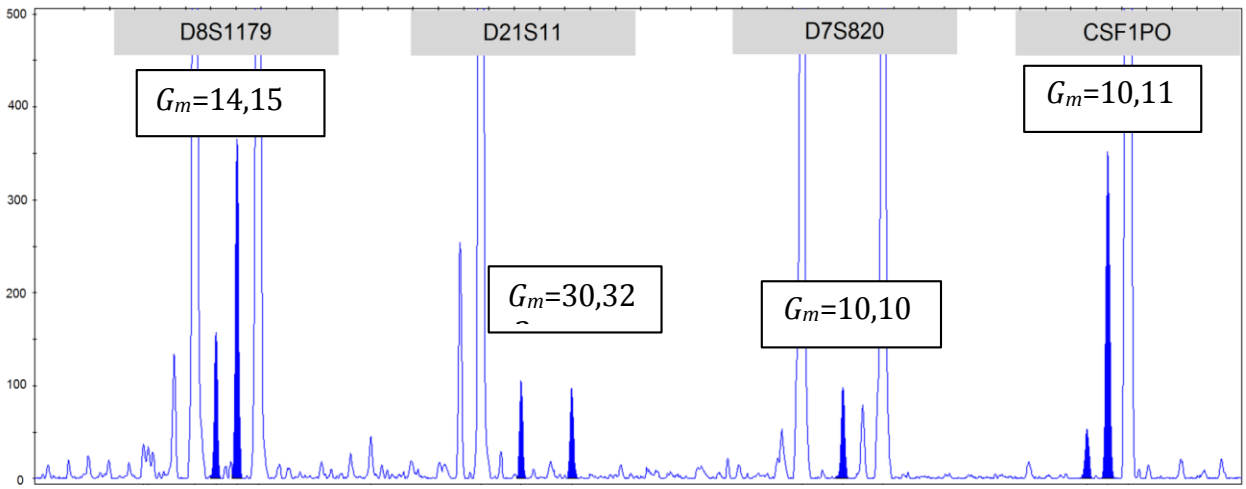


Figure 3. The blue channel of a 1:9, 2-person mixture, where the minor components' known genotype (G_m) is provided in the box above the peaks, and the minor peaks are highlighted. The 15 allele of minor contributor in D8S1178 is in stutter position and the %stutter of 7.5% is below the stutter cutoff value of 8%.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

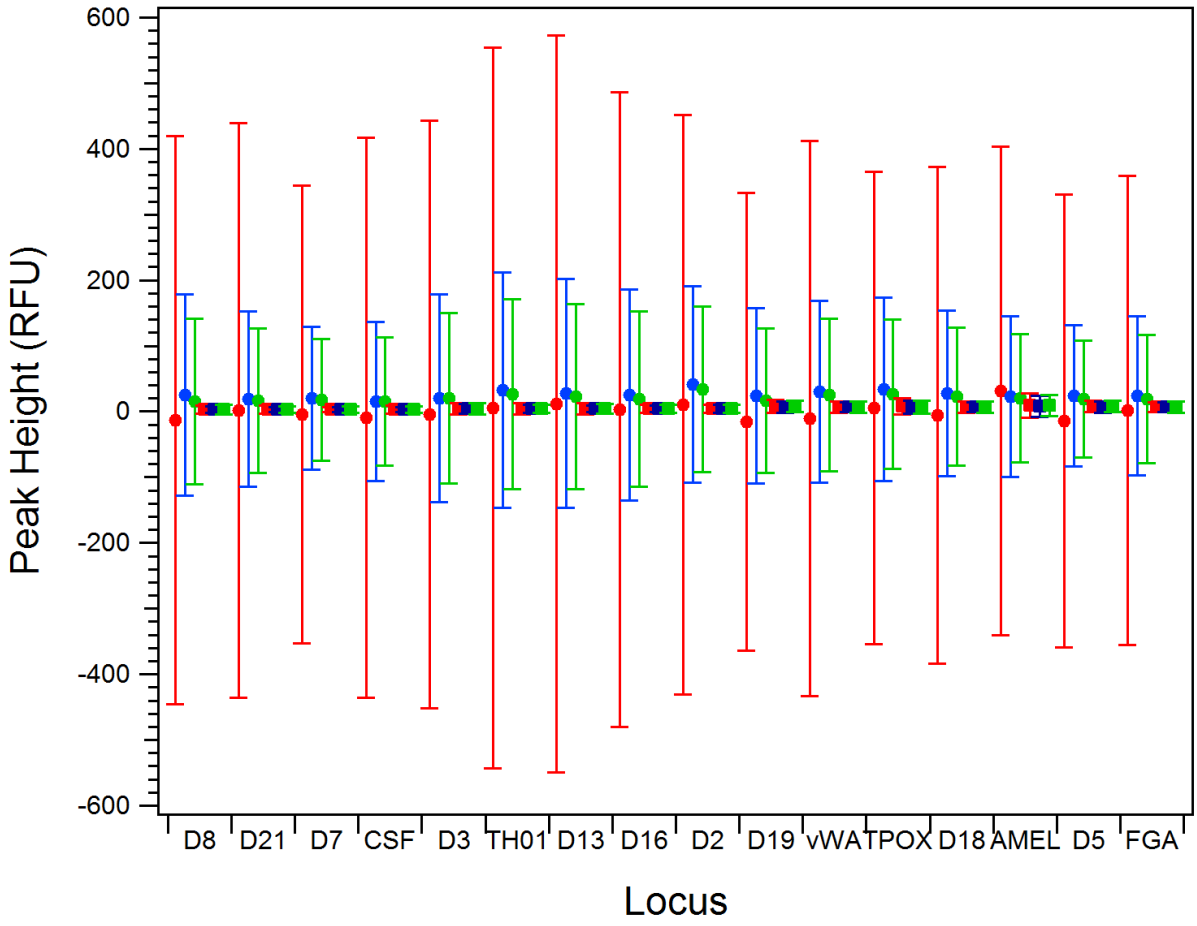


Figure 4. The *y*-intercept and 3 standard deviations obtained from a WLS linear regression obtained from (●) *Range 1*, (●) *Range 2* and (●) *Range 3*. The average and 3 standard deviations of the baseline peak heights from samples amplified with targets in (■) *Range 1*, (■) *Range 2* and (■) *Range 3*.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

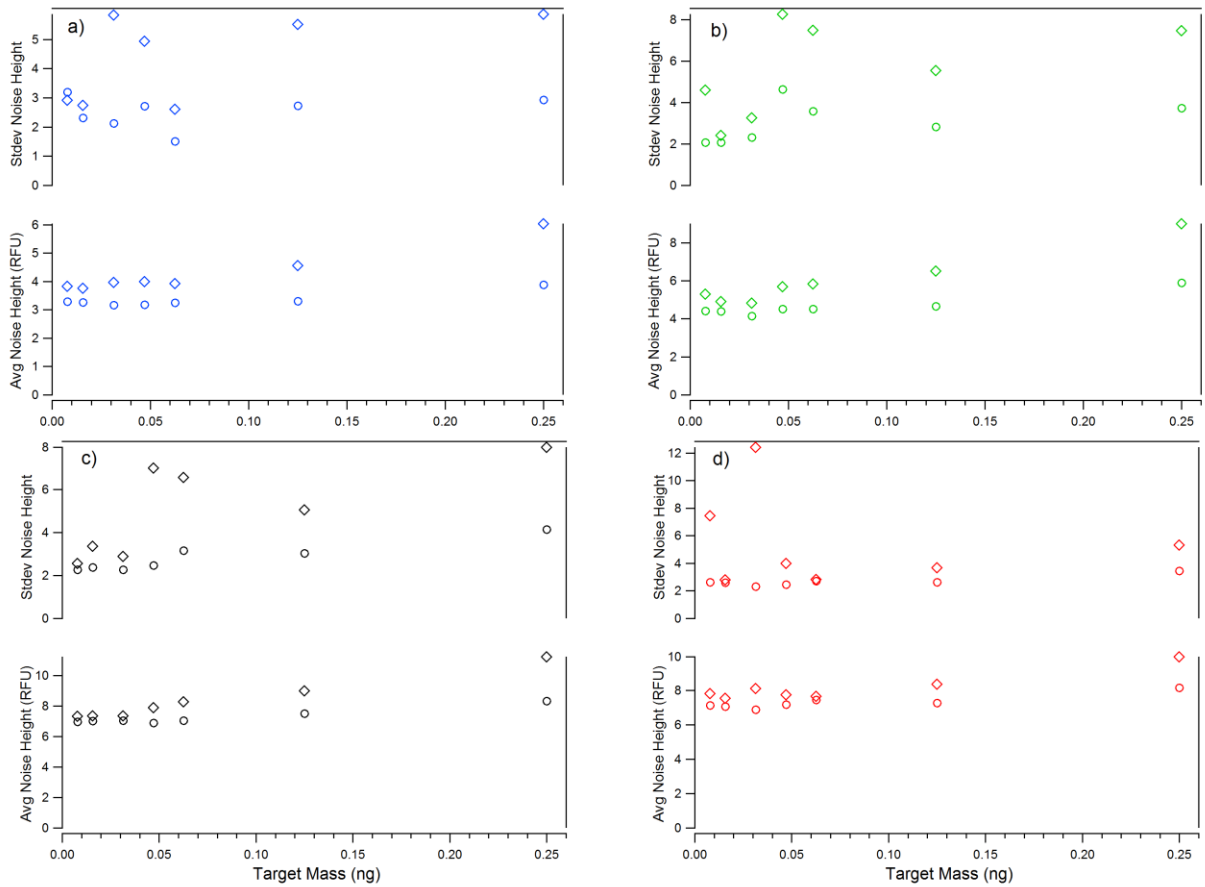


Figure 5. The average baseline noise and standard deviation of baseline noise plotted against target for the (a) blue, (b) green, (c) yellow, and (d) red color channels. Data for (○) 10 and (◊) 20 second injections are shown.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

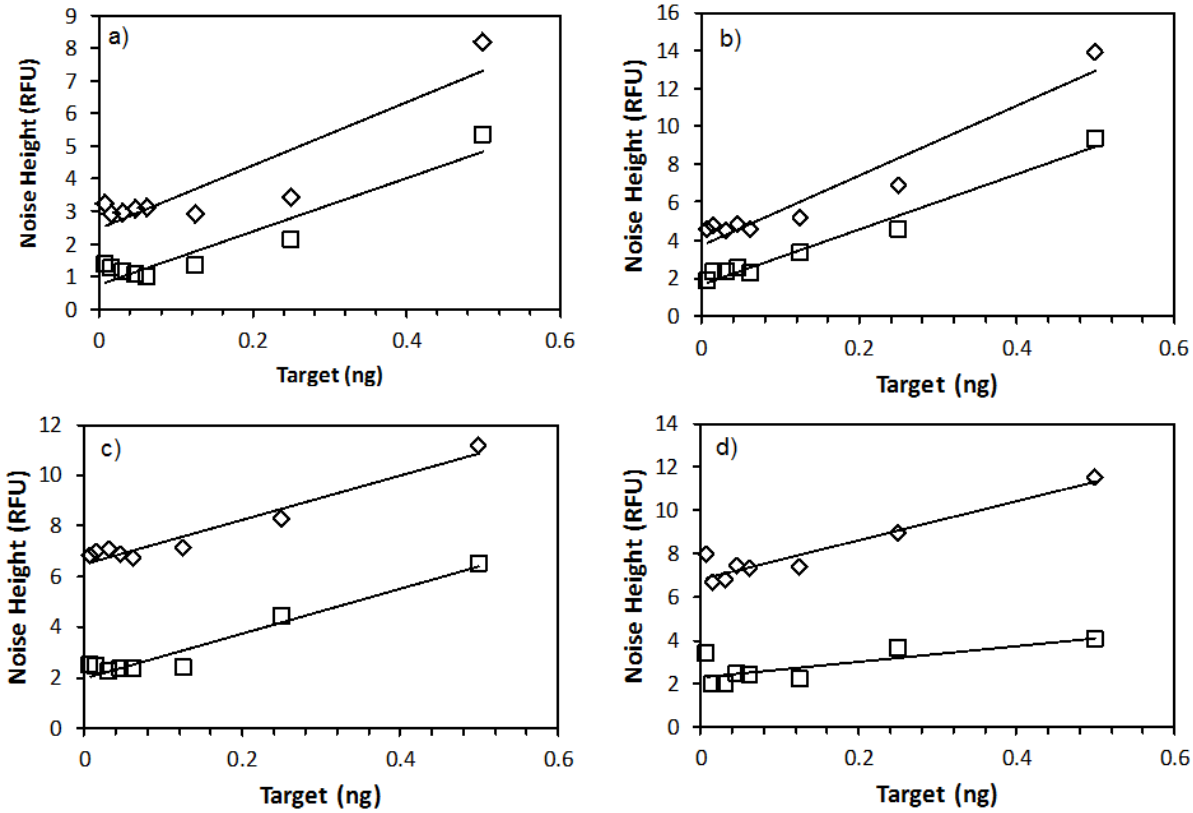


Figure 6. The (◊) average noise peak heights and the (□) standard deviation of the noise heights plotted against target for samples amplified with 0.007 to 0.5 ng of DNA for the four representative loci a) CSF1P0, b) D3S1358, c) D18S51 and d) D5S818. The injection time was 10 seconds. Also presented is the resultant trendline with and R^2 of > 0.82 and > 0.56 for the average and standard deviations, respectively.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

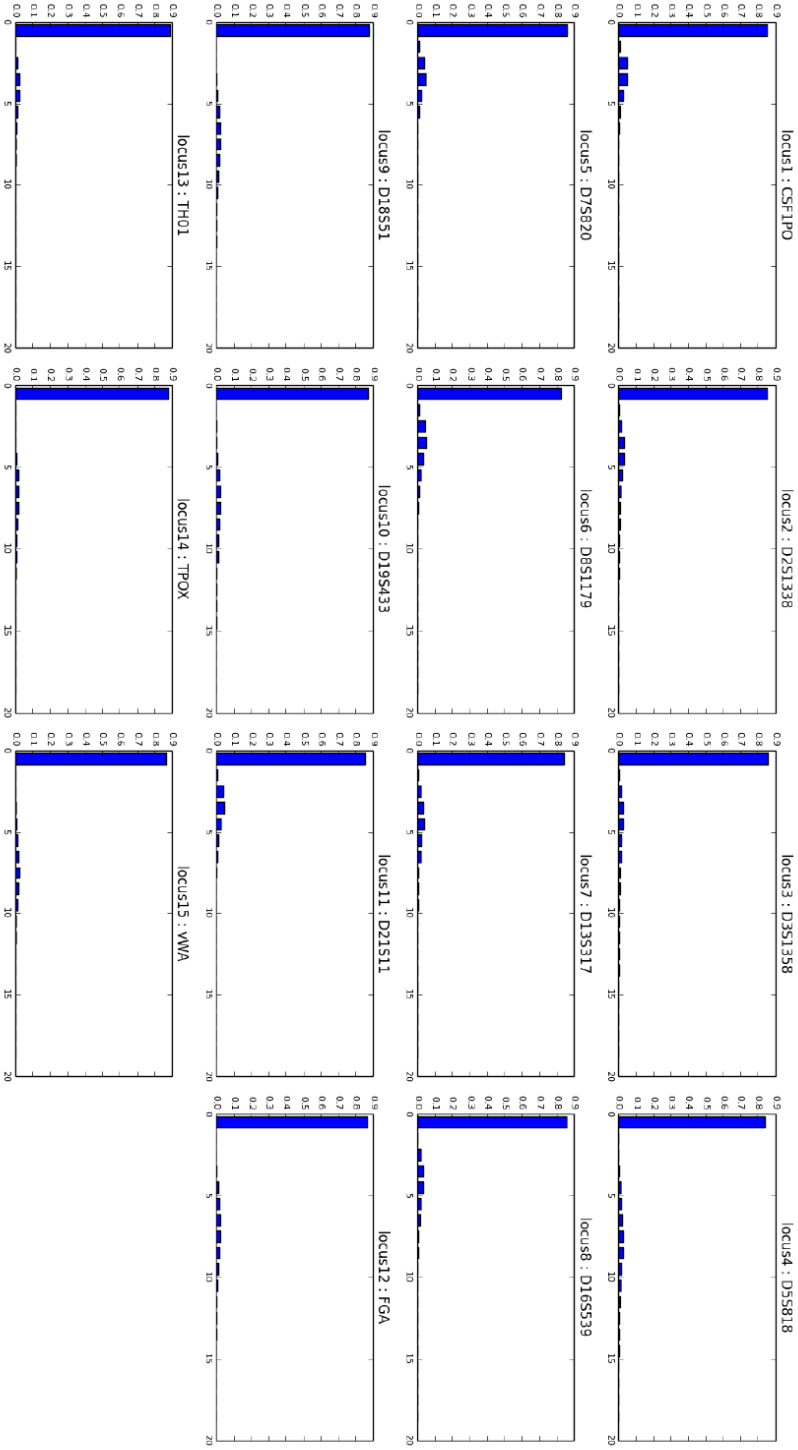


Figure 7a. Histogram (▮) of the noise peak heights (all targets) in an allele position at each locus.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

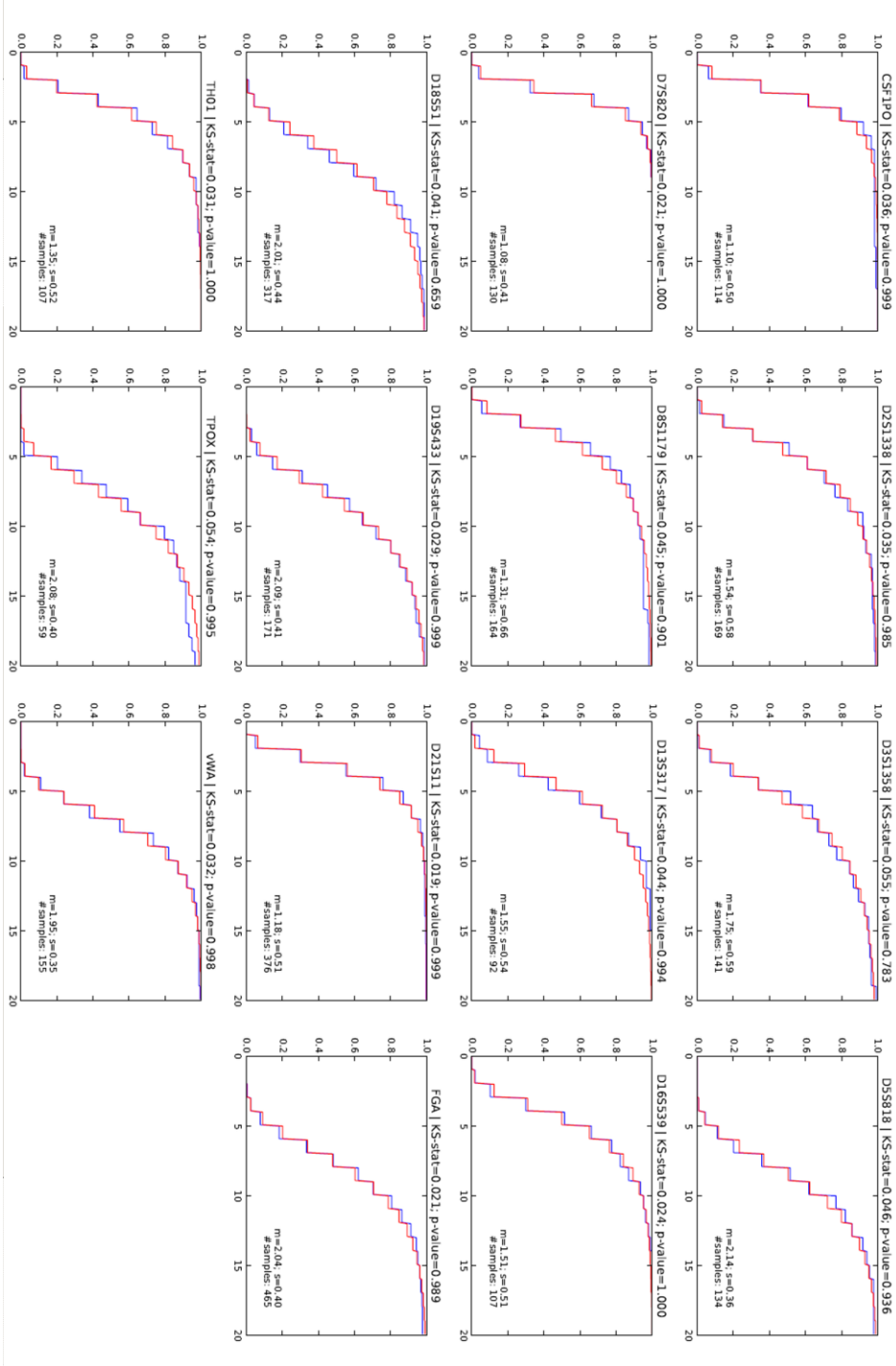


Figure 7b. Empirical CDF of the noise peaks (blue) and CDF of a quantized log-normal distributed random variable with (red).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

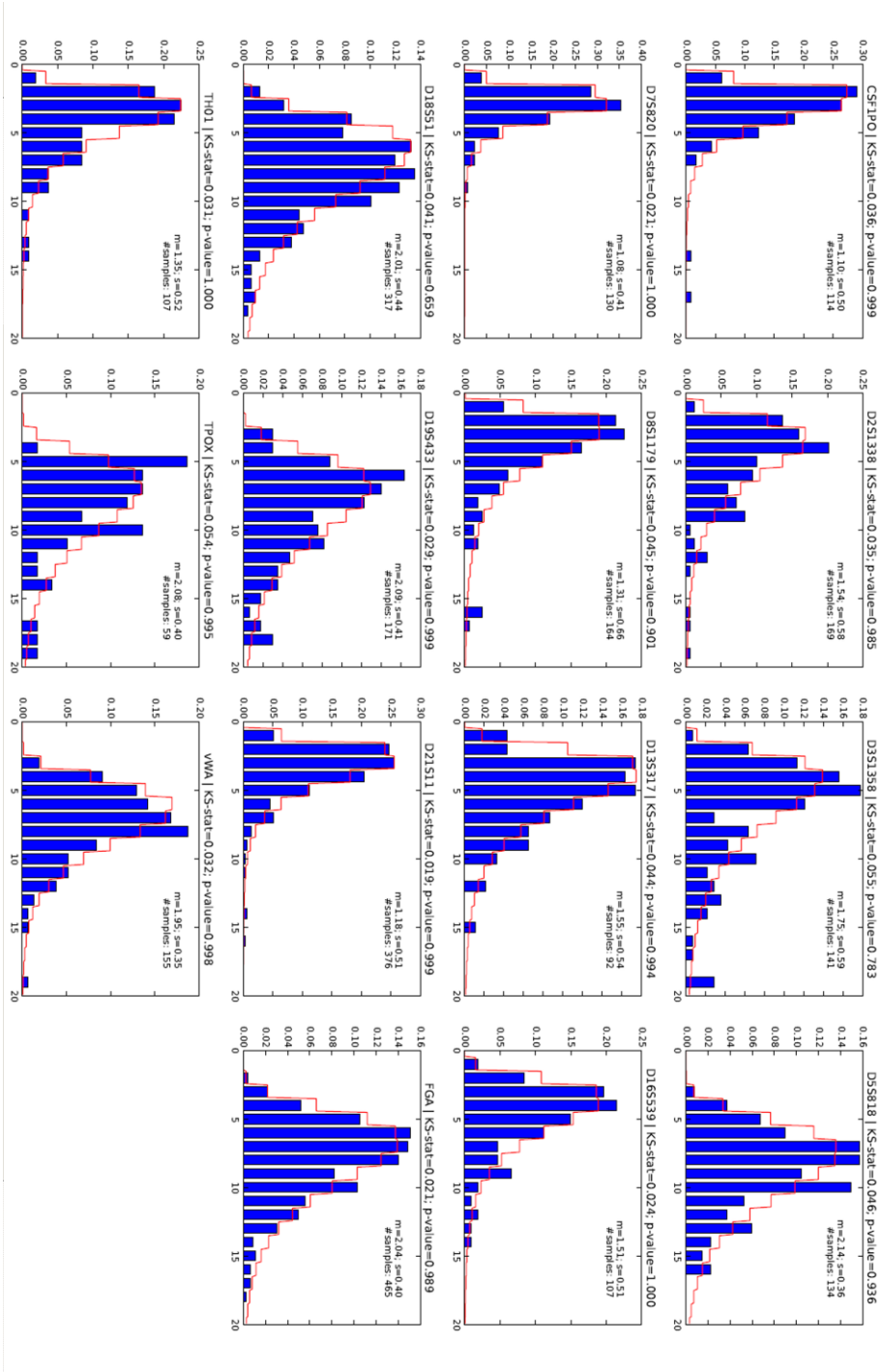


Figure 7c. Histogram of the noise peaks (blue) and pseudo PDF of a quantized log-normal distributed random variable (red).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

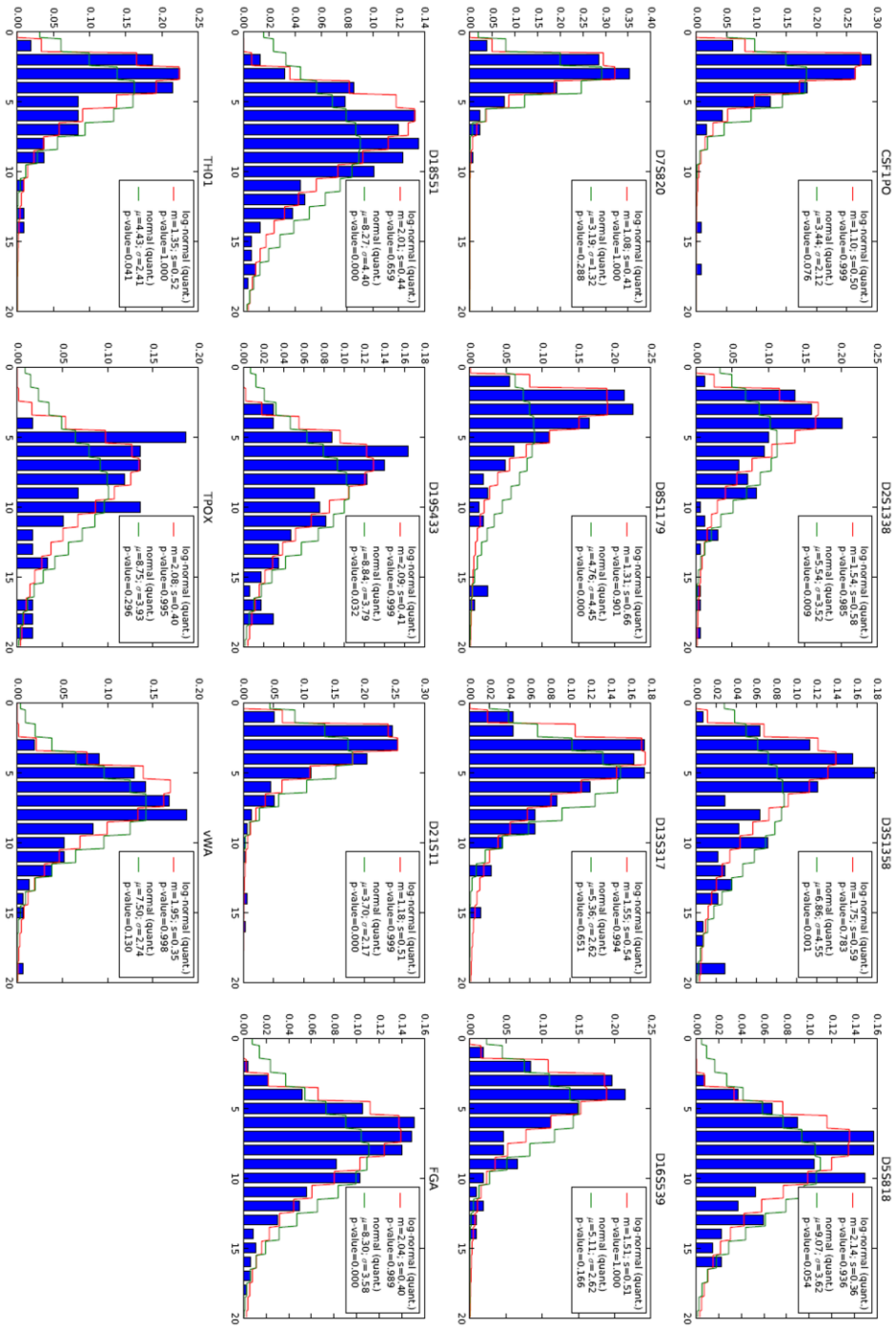


Figure 7d. Histogram of the noise peaks (blue) and pseudo PDF of a quantized log-normal distributed random variable (red) and quantized normal distributed random variable (green).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

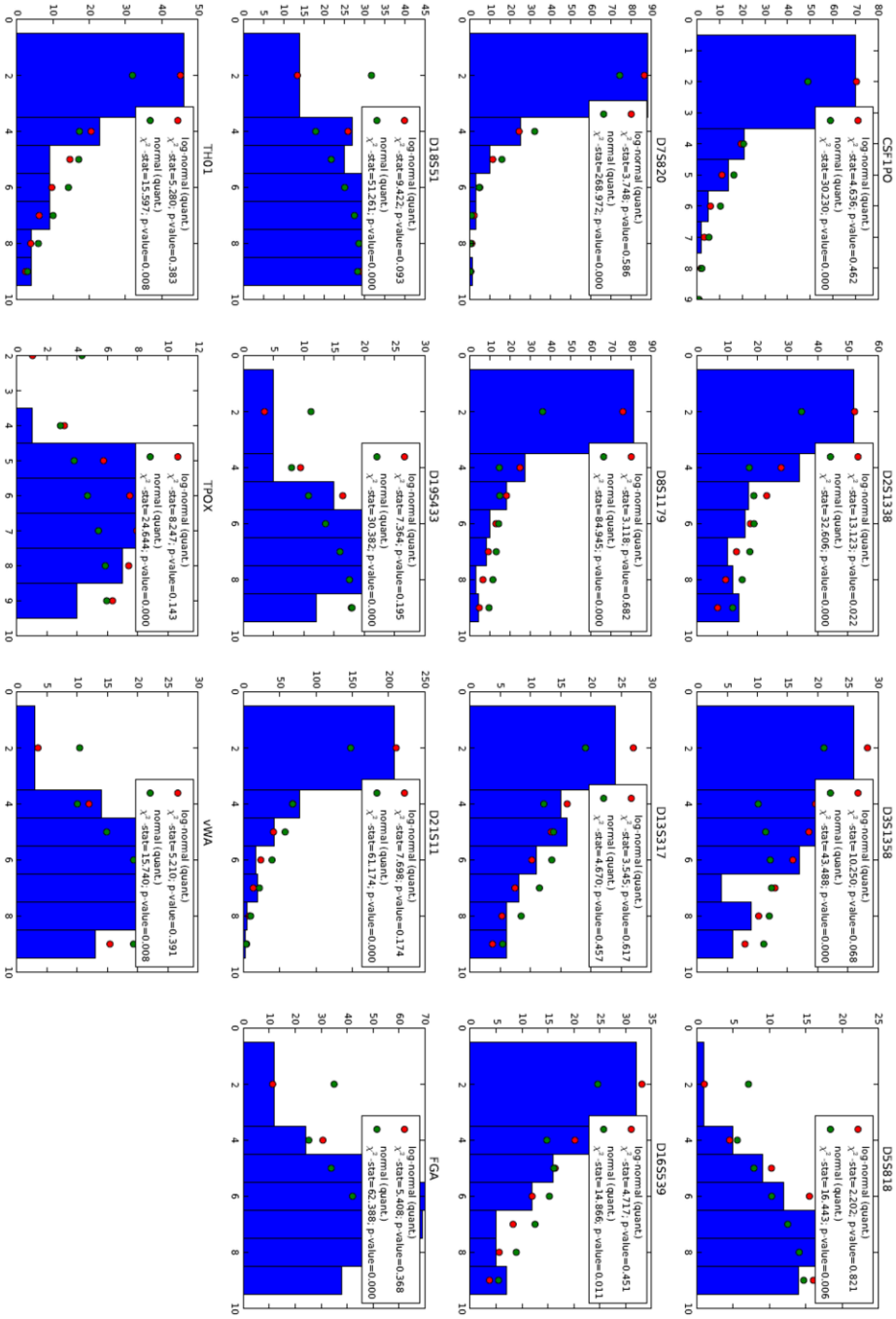


Figure 7e. Histogram (specially binned) of the noise peaks (blue), PMF of a quantized log-normal distributed random variable (red), and PMF of a quantized normal distributed random variable (green).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

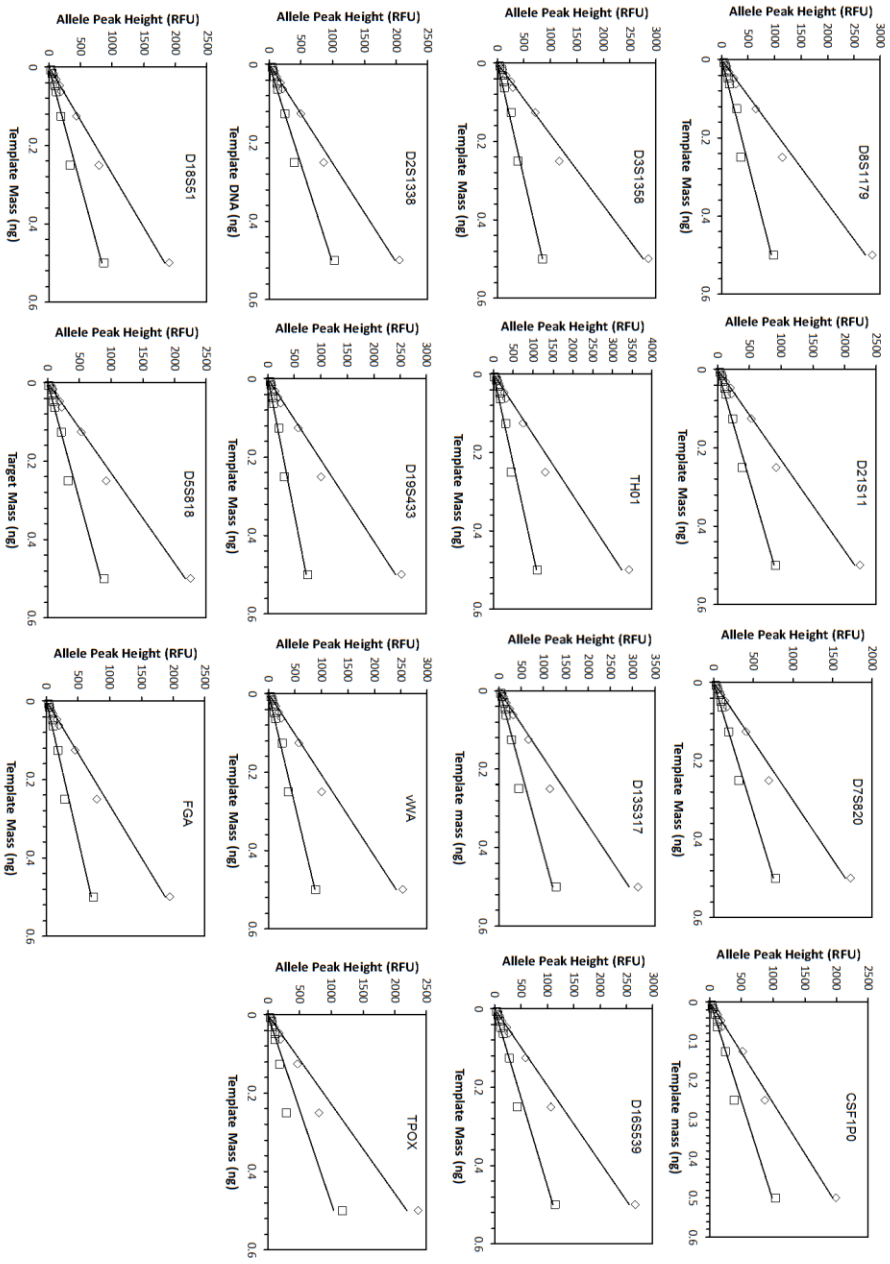


Figure 8. The (◊) average and (□) standard deviations of allele peak heights (heterozygous peaks only) plotted against target mass for all loci. The linear trendline is also shown.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

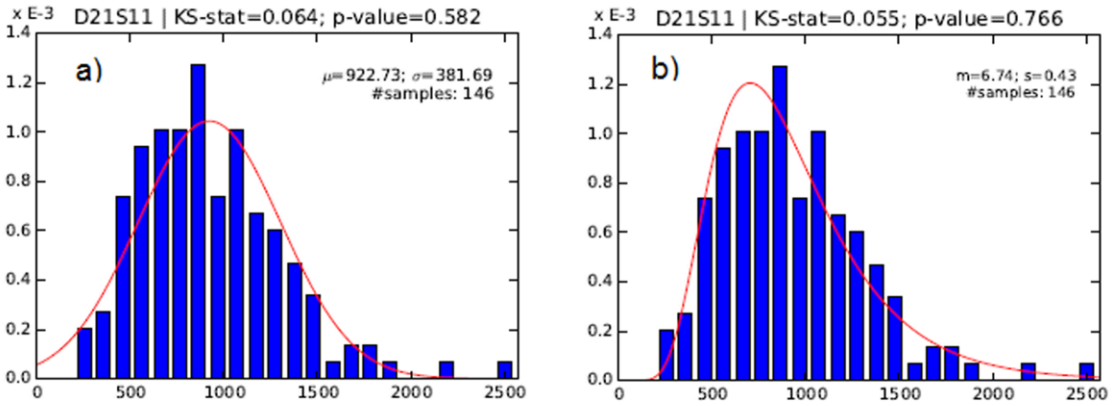


Figure 9. Representative histograms of the allele peak heights of heterozygous alleles obtained from 68 single source files amplified using 0.25 ng and the resultant (a) Gaussian and (b) log-normal fits. Representative locus D21S11 is shown.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

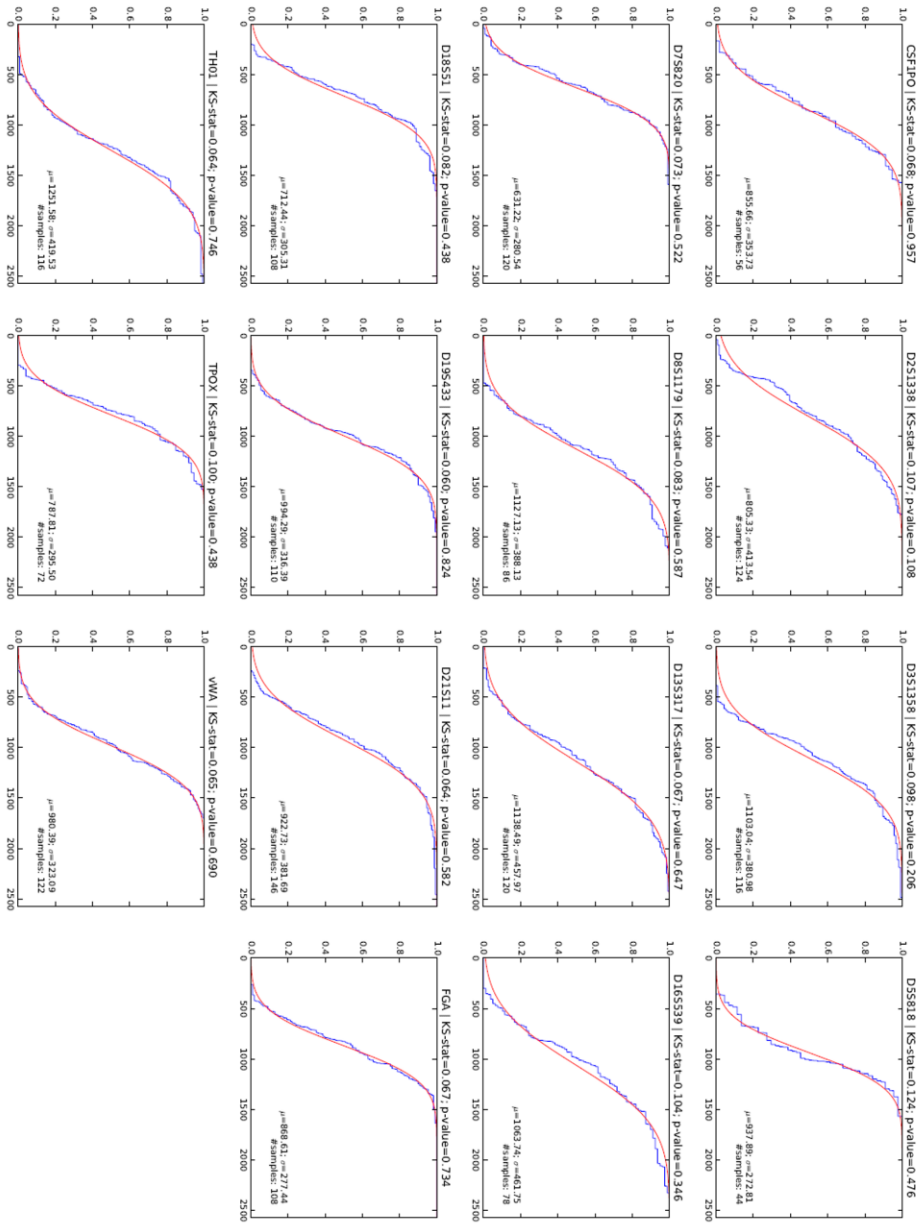


Figure 10. Results from the Kolmogorov-Smirnov test and the cumulative (▬) empirical distribution and (▬) Gaussian distribution for 0.25 ng samples.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

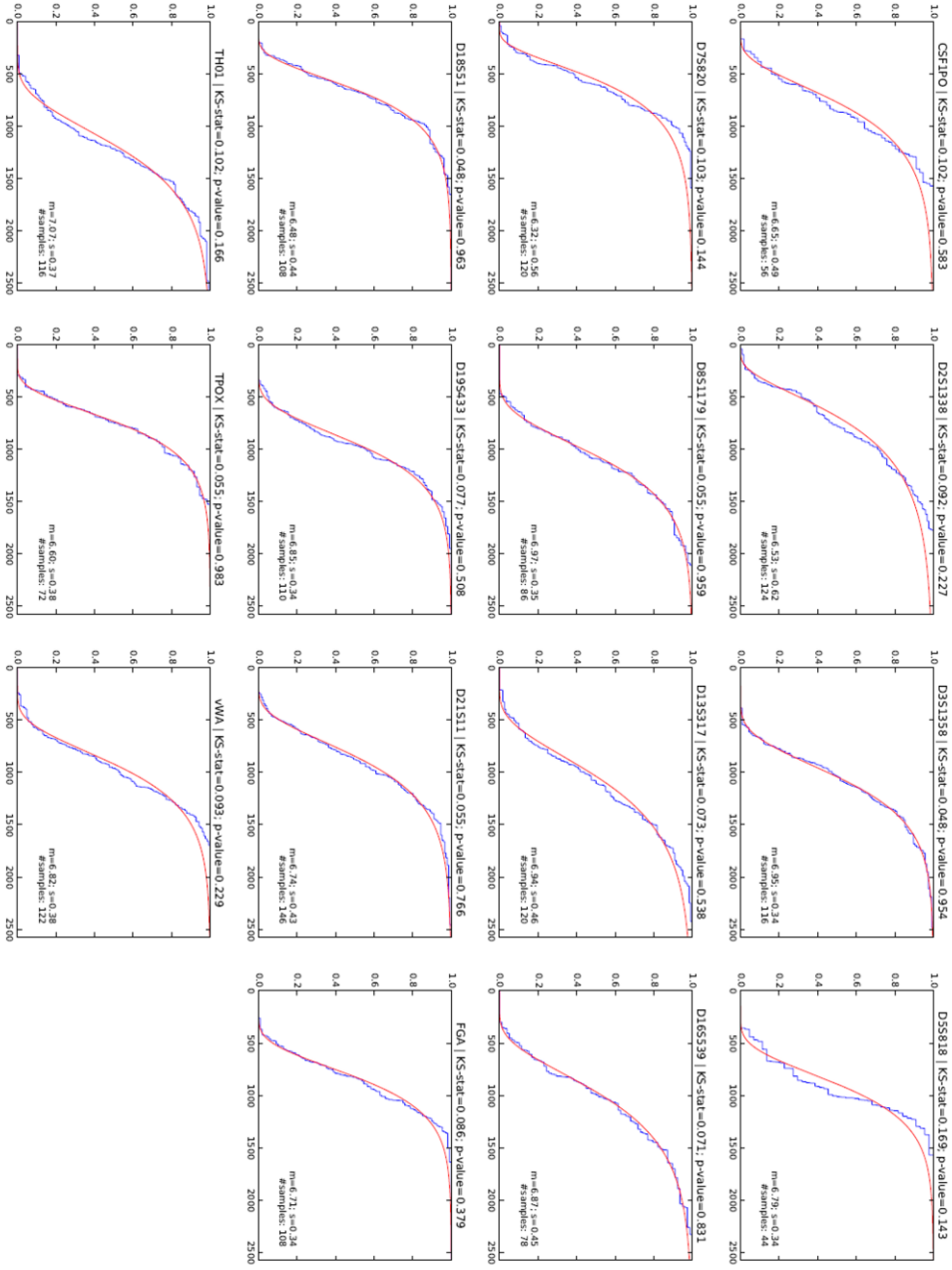


Figure 11. Results from the Kolmogorov-Smirnov test and the cumulative (▬) empirical distribution and (▬) log-normal distribution for 0.25 ng samples.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

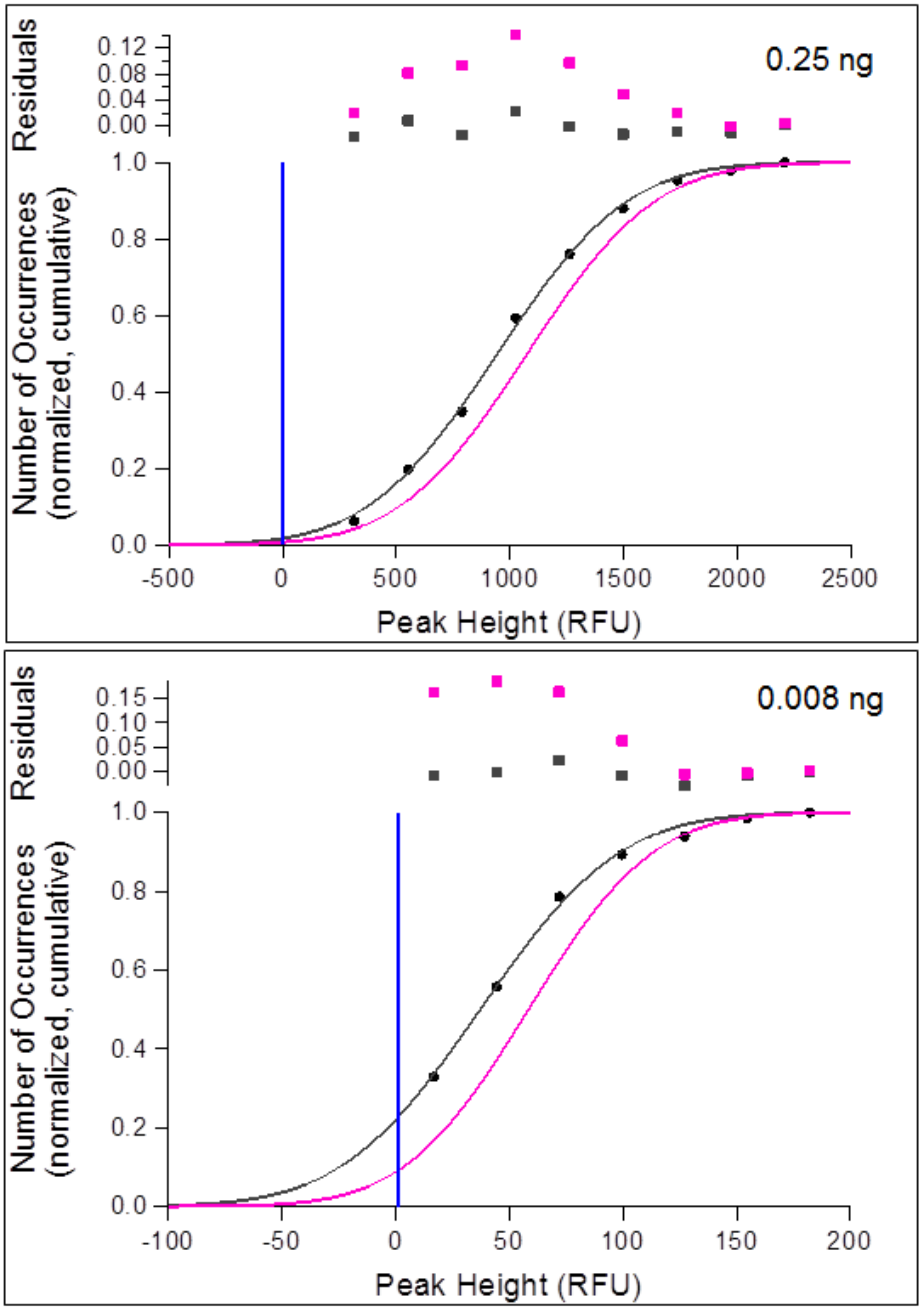


Figure 12. Cumulative peak height (●) at 0.25ng and 0.008 ng with Pr(D) equal to the value at x=1 (▌), showing that *Method 1* (▬) is a better fit of the data compared to *Method 2* (▬) but overestimates dropout at 0.25 ng (*Pr*(*D*)=0.017 versus an observed dropout frequency of 0) and underestimates dropout at 0.008 ng (*Pr*(*D*)=0.226 versus an observed dropout frequency of 0.319).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

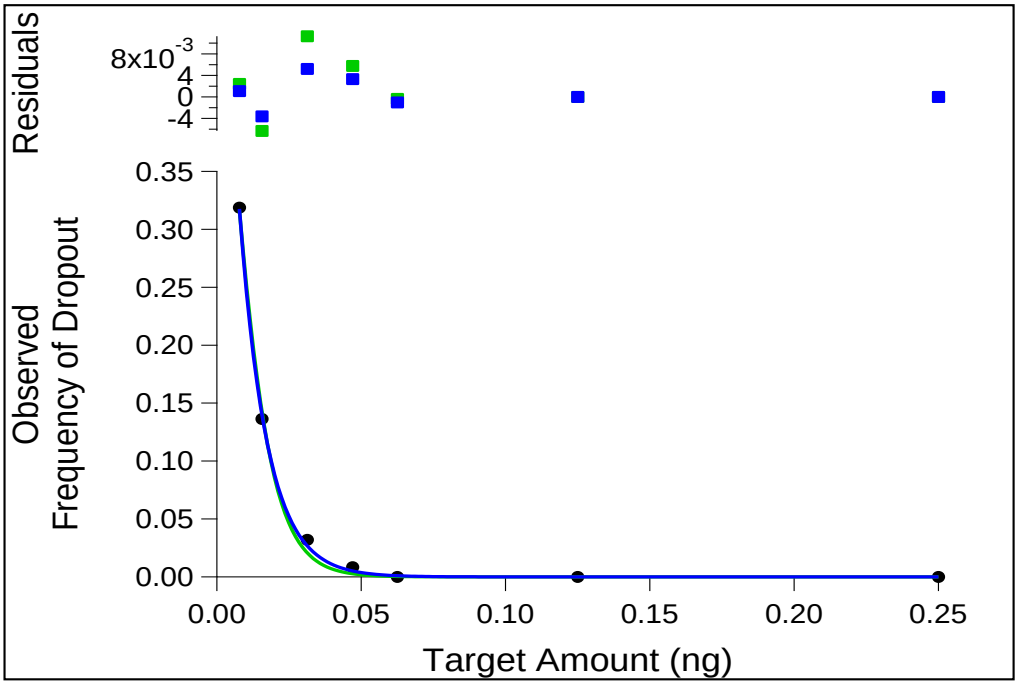


Figure 13. Observed frequencies of dropout (●) for representative locus D16S519 at 7 low-template target amounts, showing that (—) *Methods 3* and (—) *4* result in reasonable fits.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

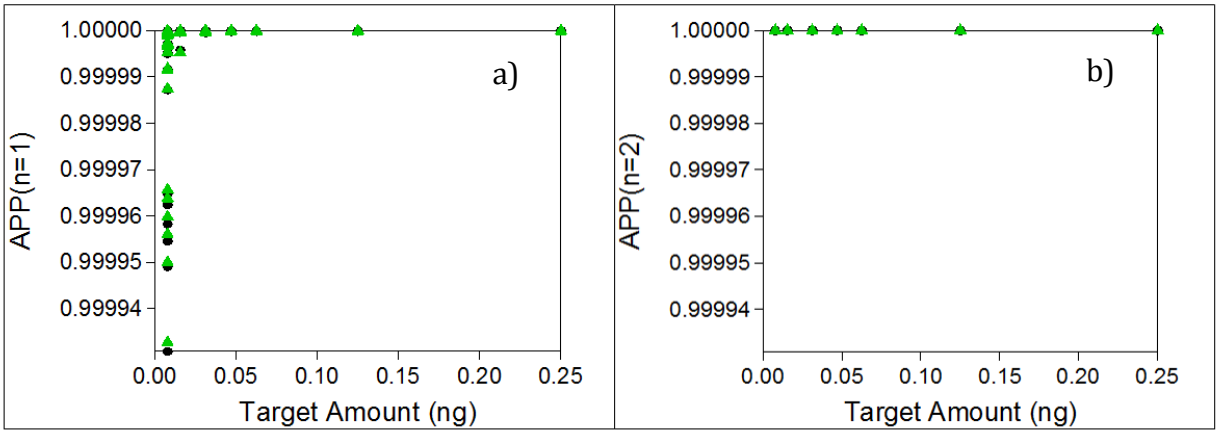


Figure 14. (a) The *APP*(*n*=1) calculated with (●) observed dropout frequencies and (▲) dropout probabilities estimated via Method 4 for 45 single-source samples amplified at 7 targets, showing 100% accuracy and decreasing *APP*s as target amount decreases. (b) The *APP*(*n*=2) calculated with (●) observed dropout frequencies and (▲) dropout probabilities estimated via Method 4 for 12 two-person mixtures, showing 100% accuracy and indistinguishable APPs.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

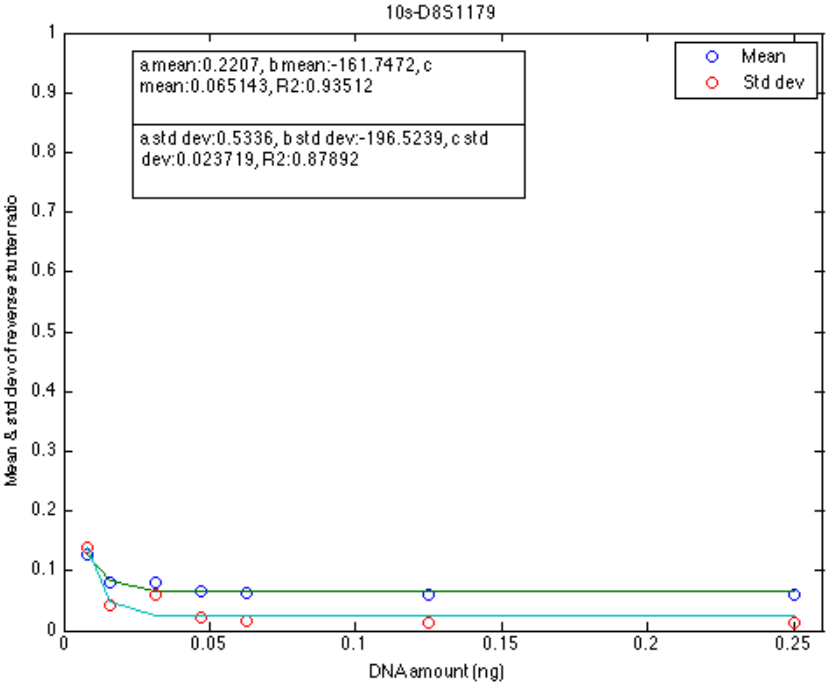


Figure 15. The average and standard deviation of the stutter ratios calculated for 68 single source samples, amplified with the AmpF*l*str® Identifiler® Plus Amplification Kit using target masses from 0.008 to 0.25 ng as determined via qPCR. The exponential fit coefficients for curve $y0+ae^{-bx}$ are shown.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

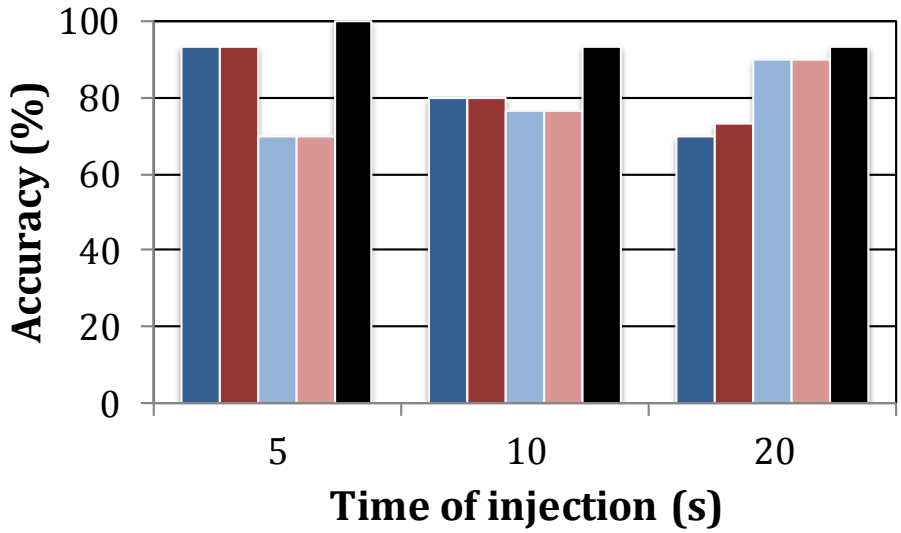


Figure 16. The %accuracy of 1-person samples using the (■)MAC (AT=50), (■)MLE (AT=50), (■)MAC(AT=variable), (■)MLE (AT=variable), (■)NOCIt methods to determine the most likely NOC contributors.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

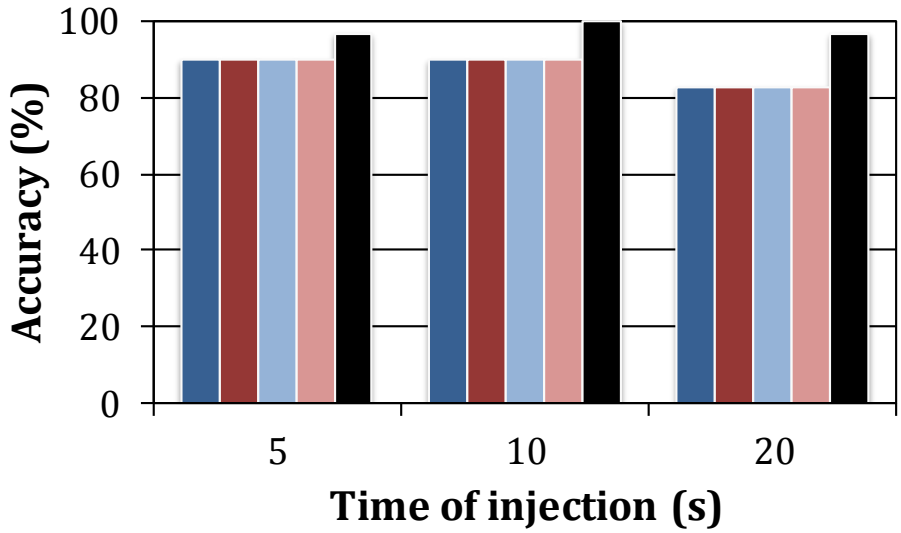


Figure 17. The %accuracy of the 2-person samples using the (■)MAC (AT=50), (■)MLE (AT=50), (■)MAC(AT=variable), (■)MLE (AT=variable), (■)NOCIt methods to determine the most likely NOC contributors.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

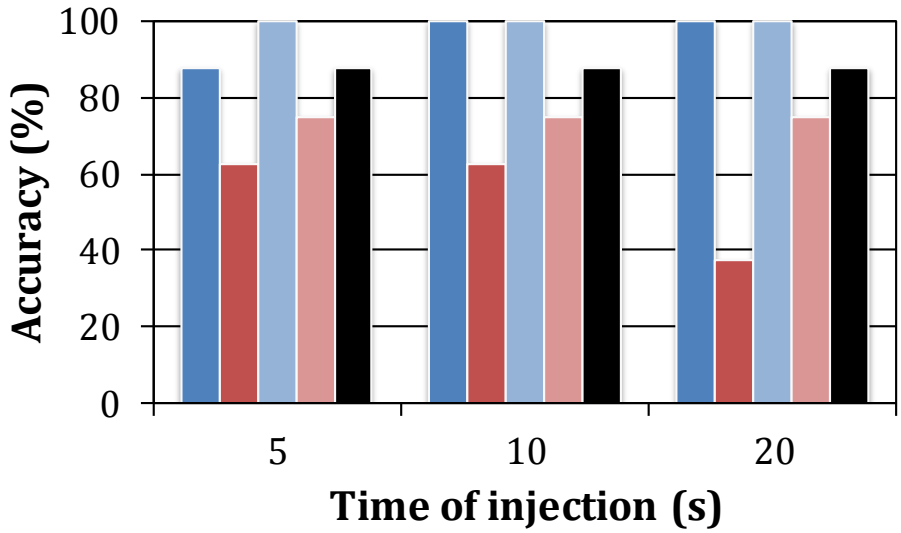


Figure 18. The %accuracy of the 3-person samples using the (■)MAC (AT=50), (■)MLE (AT=50), (■)MAC(AT=variable), (■)MLE (AT=variable), (■)NOCIt methods to determine the most likely NOC contributors.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

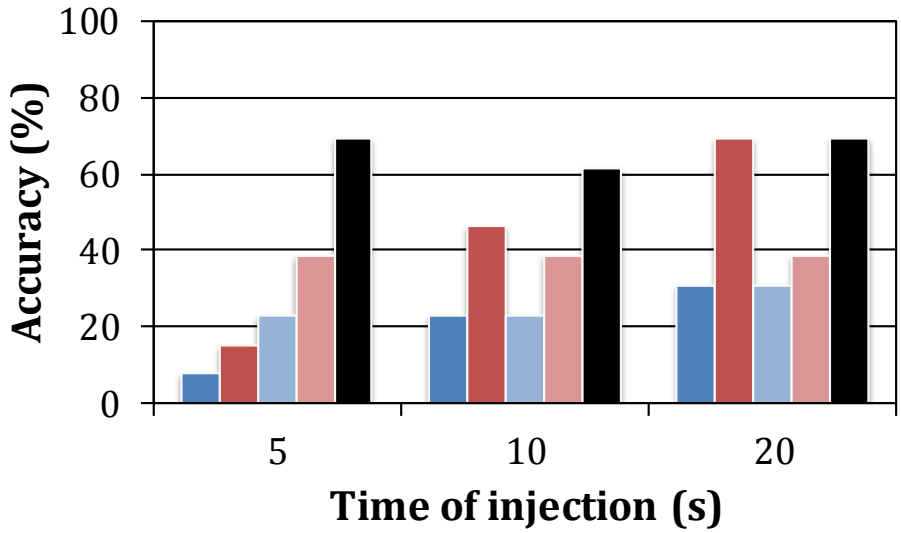


Figure 19. The %accuracy of the 4-person samples using the (■)MAC (AT=50), (■)MLE (AT=50), (■)MAC(AT=variable), (■)MLE (AT=variable), (■)NOCIt methods to determine the most likely NOC contributors.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

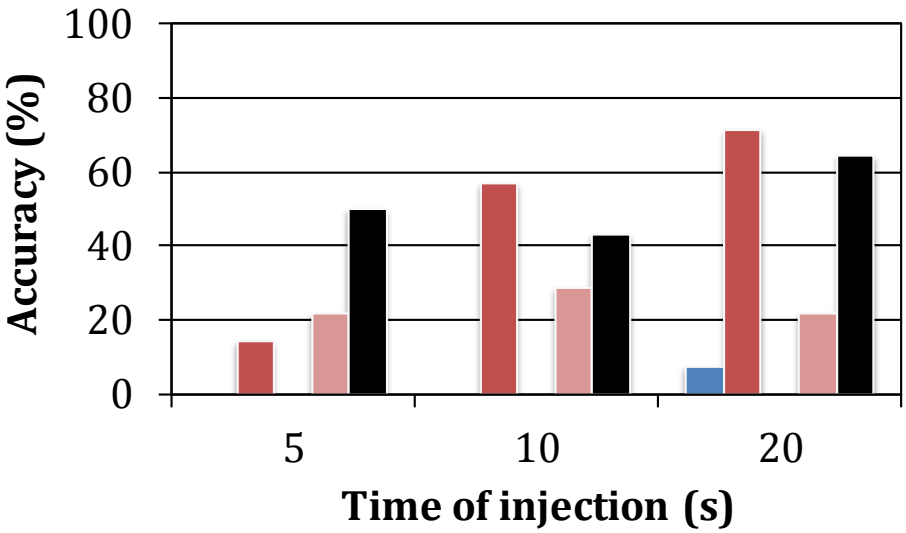


Figure 20. The %accuracy of the 5-person samples using the (■)MAC (AT=50), (■)MLE (AT=50), (■)MAC (AT=variable), (■)MLE (AT=variable), (■)NOCIt methods to determine the most likely NOC contributors.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

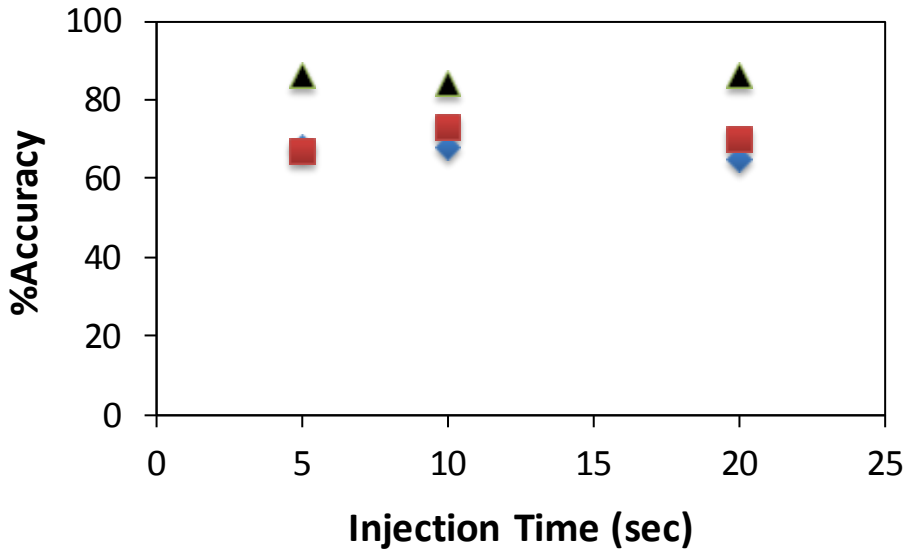
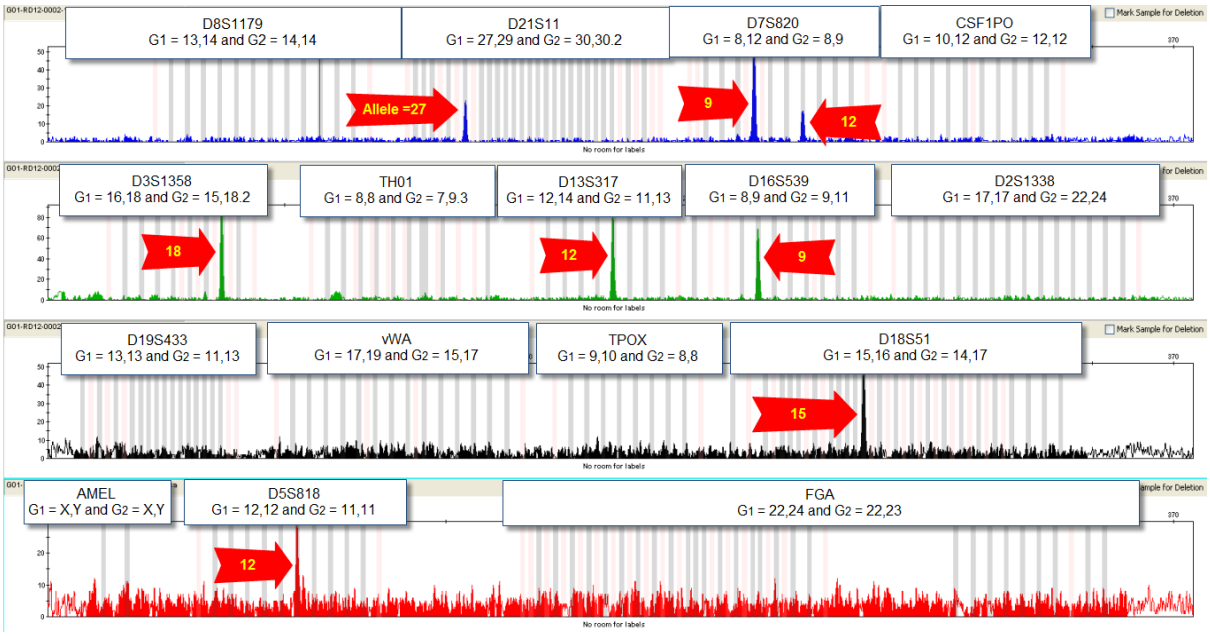


Figure 21. The %accuracy of (♦)MAC, (■)MLE and (▲)NOCIt for all samples/mixtures plotted against injection time.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

2-Person Mixture, 1:1 Ratio, 0.016 ng target




```
Calibration File: C:\Users\cgrgicak\Desktop\calib_10s.csv
Frequency File: C:\Users\cgrgicak\Desktop\Freq_Caucasian.csv
Sample File: C:\Users\cgrgicak\Desktop\RD12-0002-1and2-1to1-
0.0156IP-003-10sec.csv
Sample DNA input: 0.016ng


Number of contributors: 0
Time taken: 0.0m
Probability: 1.46155010416729228947671110773533 7E-429

Number of contributors: 1
Time taken: 0.03m
Probability: 0.91278785845360121008627304 77

Number of contributors: 2
Time taken: 96.56m
Probability: 0.08721214154639878991372695229162374
```

Figure 22. A low-template sample consisting of a 2-person mixture. The loci and known genotypes for contributor one and two are provided in the box above the peaks. Allele designations are in the red arrows. Peak height ranges from baseline to 84 RFU. Corresponding NOCIt results/output is provided below. Note that NOCIt output provides the likelihood and the most likely alleles for each NOC tested. This has been removed for legibility.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

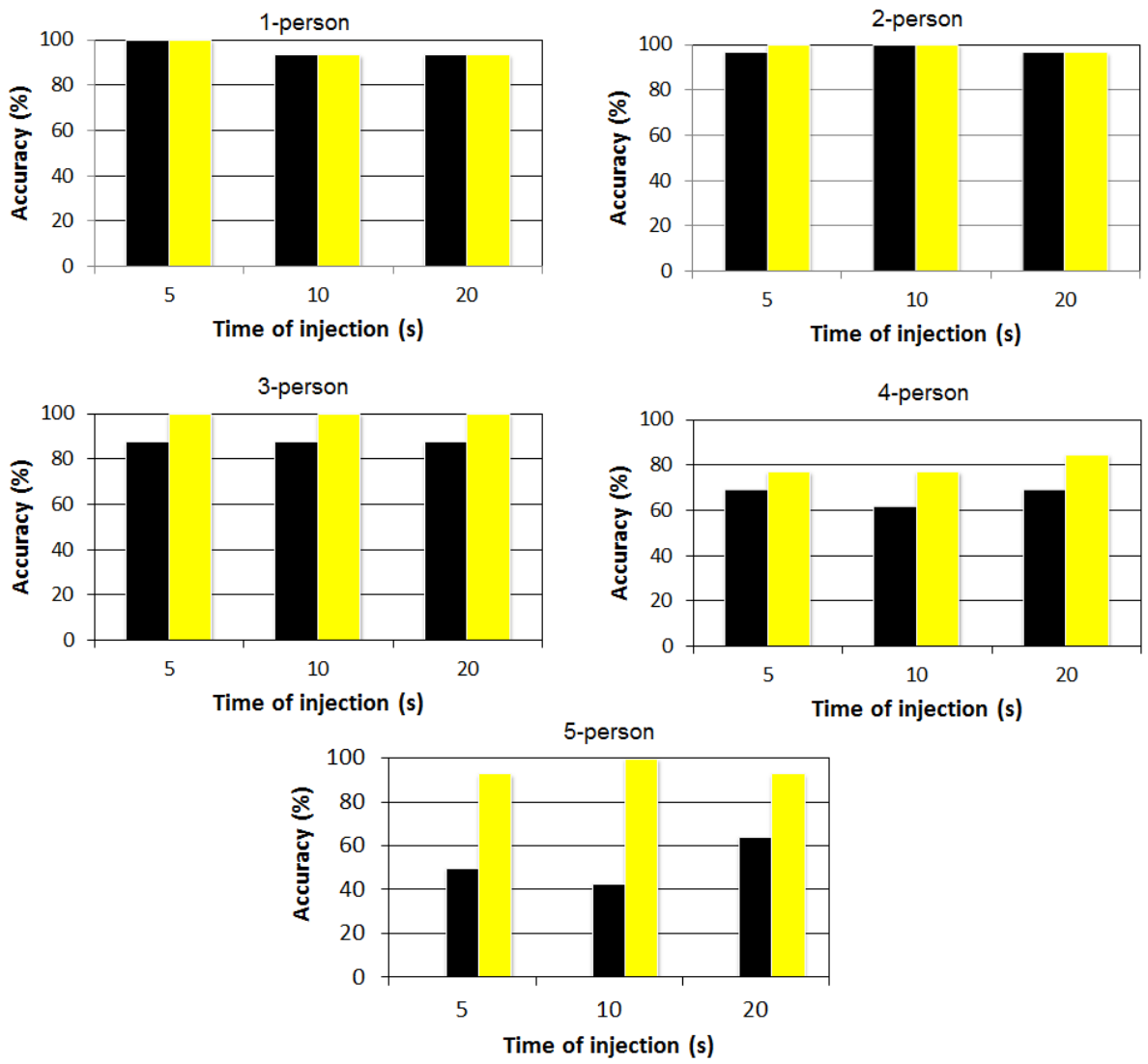


Figure 23. The (■) %accuracy of NOCIt and the (■)%accuracy that NOCIt determines that the APP is at least 1% (i.e. there is a reasonable probability the sample originated from n contributors).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

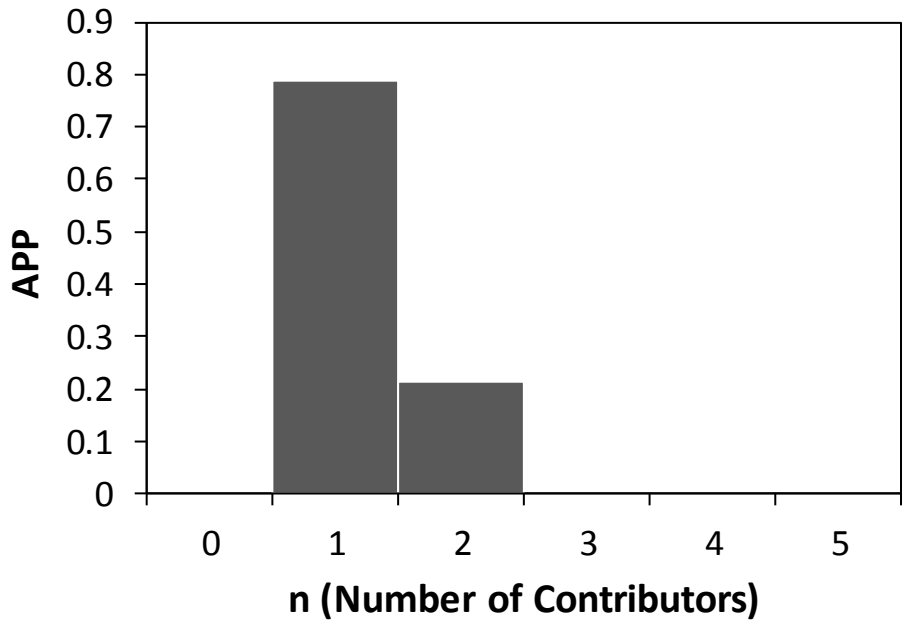


Figure 24. The APP distribution of a low-template single source sample when 10 loci are used during analysis.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

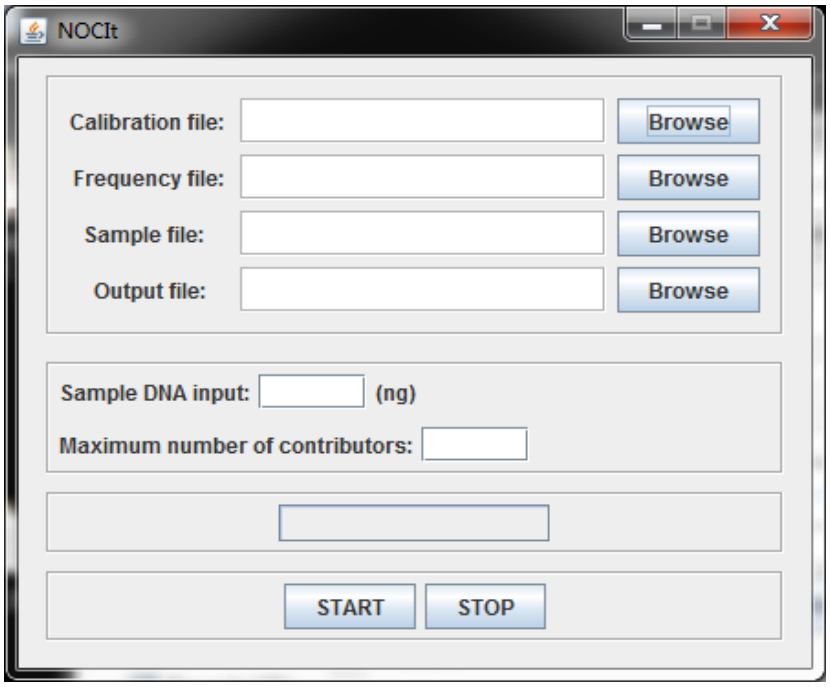


Figure 25. NOCIt software interface.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

```
Calibration File: calib_10s.csv
Frequency File: allele_freq_cauc.csv
Sample File: ex.csv
Sample DNA input: 0.25ng


Number of contributors: 0
Time taken: 0.0m
D16S539 = 0
D18S51 = 0
D5S818 = 0
D2S1338 = 0
D7S820 = 0
vWA = 0
TPOX = 0
TH01 = 0
FGA = 0
D3S1358 = 1.2233355521839572720206874227532 75E-292
CSF1PO = 0
D8S1179 = 0
D13S317 = 0
D21S11 = 0
D19S433 = 0
Likelihood: 0
Probability: 0

Number of contributors: 1
Time taken: 0.15m
D16S539 = 1.3387485456567720400796347762379 48E-16 : [10.0]
D18S51 = 2.9647765115766458780323262070385 06E-27 : [20.0, 18.0]
D5S818 = 8.5689778515514752752074594621037 72E-16 : [12.0, 8.0]
D2S1338 = 6.6061064311288414354869887064764 70E-13 : [23.0, 22.0]
D7S820 = 7.4304839048494730531940888103317 76E-13 : [10.0, 8.0]
vWA = 4.7383425191142755920094763601841 91E-15 : [17.0]
TPOX = 2.5819328501255842981584679549185 86E-15 : [6.0, 10.0]
TH01 = 5.3053903309388698905387853515134 40E-13 : [6.0, 8.0]
FGA = 3.3986587474401109062940596649031 33E-19 : [24.0, 22.0]
D3S1358 = 3.7382789990577888311001044998296 41E-17 : [15.0, 17.0]
CSF1PO = 7.5216156697496601182639793555763 85E-20 : [12.0, 10.0]
D8S1179 = 1.3149622311483071179031922275405 99E-15 : [12.0, 11.0]
D13S317 = 6.8351518575143834623562976080656 62E-19 : [12.0, 9.0]
D21S11 = 5.7607283871672360518249319200675 17E-28 : [28.0, 33.2]
D19S433 = 2.9989323277355248123669141842088 96E-13 : [13.0, 11.0]
Likelihood: 1.6079306553740217844190374611692 94E-250
Probability: 1
```

Figure 26. Example output file for a sample tested with $n_{max} = 1$. The n values of 2 to 5 take a similar form and are not shown for purposes of legibility.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.