The objective of this project was to develop a computational technique that can assess the NOC (number of contributors) to complex DNA samples. The likelihood that \( n \) non-related individuals contributed to a questioned sample is computed via a Monte Carlo approach. During each iteration of the Monte Carlo process, genotypes for the \( n \) contributors at a mixture ratio are randomly chosen based on the allele frequencies and from a uniform distribution, respectively. Modeling of allele and stutter dropout, the means and standard deviations of allele and noise heights and the means and standard deviations of stutter ratios is performed. The modeling is based on a calibration data set consisting of single source samples with known genotypes. The distributions are all assumed to be Gaussian. The likelihood of observing the heights of the peaks given the genotypes of the contributors, the mixture ratio and DNA mass is computed and repeated a number of times. The average of the values computed is the likelihood of observing the evidence at a locus, given \( n \) contributors. The likelihood values at all the loci are multiplied to give the overall likelihood for 1 through 5 contributors.

The performance of NOCIt was tested on 1-, 2-, 3-, 4- and 5- person mixtures, which were amplified with the AmpF/lstr® Identifier® Plus Amplification kit (29 cycles) and injected for 5, 10 and 20 sec. Samples were amplified using 7 DNA amounts (0.25 – 0.008 ng). The performance of NOCIt was compared with the MAC (Maximum Allele Count) and MLE (Maximum Likelihood Estimator) methods. MAC and MLE rely on setting an AT to calculate the number of contributors. Therefore, 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. The second threshold was a variable threshold. 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 utilize analytical thresholds. As a result, when analyzing mixtures with NOCIt an RFU peak threshold of 1 was utilized. 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/lstr® Identifier® Plus manual was used to filter out the stutter peaks at each locus. Allele frequencies from the Caucasian population specified in the AmpF/lstr® Identifier® Plus manual were used to test the NOCIt and MLE methods.

NOCIt resulted in the highest overall accuracy rates for all injection times. The accuracy of NOCIt was 86%, 84% and 86% for the 5, 10 and 20 sec injections, respectively. The overall accuracies of MAC and MLE for the 5, 10 and 20 sec injections were 68%, 68%, 65% and 66%, 73% and 70%, respectively.

When examining the accuracy rates with respect to the actual NOC in the mixture, NOCIt resulted in accuracy rates ≥ 93% for the 1-, and 2- person samples. MLE and MAC gave comparable results for the 5 and 10 sec injections, but significantly lower accuracies, i.e. 70 - 73%, when these samples were injected for 20 sec. NOCIt resulted in an accuracy of 88% for the 3- person mixtures while the MAC and MLE methods resulted in accuracies ≥ 70% and ≥ 73%, respectively. The accuracies for the 4- and 5- person samples fall below 75%, regardless
of the method utilized or injection time. 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. The MLE method also resulted in a two-fold increase in accuracy over MAC. However, for these highly complex samples, MLE was dependent on both injection time and AT, where the highest accuracy was obtained with a 20 sec injection and a constant AT of 50 RFU. 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 value, i.e. 71%, originating from the dataset analyzed using a 20s injection time and a constant AT of 50 RFU. 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 sec, respectively. The accuracy of MLE for the 5-person mixtures was 14%, 57% and 71% for the 5, 10 and 20 sec 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 significantly affected by injection time and were 69%, 61% and 69% (4-person) and 50%, 43% and 64% (5-person) for the 5, 10 and 20 second injections.

The sources of the inaccuracies observed when utilizing NOCIt were examined. Underestimates were obtained for the 4- and 5-person mixtures and may be due to the complexity associated with these samples, i.e. allele sharing, or complete profile drop-out with the low-template, major-minor combinations. Further investigations into whether the underestimates are a result of full profile drop-out at extreme mixture ratios are underway. Overestimates were observed for the 1-, 2-, 3- and 4-person samples, which we suspect is because of the presence of forward stutter. In the instances where NOCIt failed to identify the correct number of contributors, it was able to identify the region in which the number is most likely to lie, demonstrating that NOCIt provides a method that allows the uncertainty in the NOC to be quantified.

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 with 2GHz of processor speed. 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.